

ESTUDIO DE LA REGULACIÓN GÉNICA EN EL  
LOCUS SCN<sub>5A</sub>-SCN<sub>10A</sub> Y SU IMPLICACIÓN EN LA  
ETIOLOGÍA DEL SÍNDROME DE BRUGADA

**Adrián Pérez Agustín**



<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.ca>

Aquesta obra està subjecta a una llicència Creative Commons Reconeixement-NoComercial-SenseObraDerivada

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives licence



TESIS DOCTORAL

**Estudio de la regulación génica en el locus *SCN5A-SCN10A* y su implicación en la etiología del Síndrome de Brugada**

Adrián Pérez Agustín

2023





## TESIS DOCTORAL

**Estudio de la regulación génica en el locus *SCN5A-SCN10A* y su implicación en la etiología del Síndrome de Brugada**

**Adrián Pérez Agustín**

**2023**

Programa de Doctorat en Biologia Molecular, Biomedicina i Salut

Dirigida per la Dra. Sara Pagans Lista

Codirigida per la Dra. Catarina Allegue Toscano

Codirigida pel Dr. Ramon Brugada Terradellas

Memòria presentada per optar al títol de doctor per la Universitat de Girona

3 Annexes





# Universitat de Girona

La Dra. Sara Pagans Lista, professora agregada del Departament de Ciències Mèdiques de la Universitat de Girona;

La Dra. Catarina Allegue Toscano, investigadora associada de la *Universidad de Santiago de Compostela*, Grupo de Medicina Xenómica;

El Dr. Ramon Brugada Terradellas, professor catedràtic del Departament de Ciències Mèdiques de la Universitat de Girona, director del Centre de Genètica Cardiovascular de l'Institut d'Investigació Biomèdica de Girona, responsable del servei de Cardiologia de l'Hospital Trueta de Girona i cardiòleg de l'Hospital Josep Trueta de Girona;

DECLAREM: Que el treball titulat "**Estudio de la regulación génica en el locus *SCN5A-SCN10A* y su implicación en la etiología del Síndrome de Brugada**", que presenta l'Adrián Pérez Agustín per a l'obtenció del títol de doctor, ha estat realitzat sota la nostra direcció.

Girona, 22 de Febrer de 2023



## Agradecimientos

Uno siempre piensa que escribir los agradecimientos será la parte más fácil de la tesis, pero nada más lejos de la realidad. Sentado, enfrente del manuscrito ya finalizado (sí, esta vez sí, ya acabaron las re-re-correcciones), me hago la pregunta: ¿Qué es lo primero que uno hace al mirar la tesis doctoral de otro? Obvio, te vas directamente a leer los agradecimientos (a veces con la esperanza de ser nombrado como alguien especial, otras solamente por puro cotilleo, no nos engañemos). Luego, si hay suerte, miras alguna figura y compruebas cuantas referencias bibliográficas ha puesto (tú también lo has hecho, y lo sabes). Para todos aquellos que os habéis pasado por aquí, entre nosotros, si os interesa saber más sobre las virtudes de la secuenciación de lecturas largas os podéis pasar por la página 238... os va a encantar... Vale, vale, voy ya, ahora en serio. Espero que todo el mundo que quiera encontrar su nombre en estas páginas lo haga, procuraré no dejarme a nadie, pero no prometo nada.

Me gustaría empezar dando las gracias a mi directora de tesis, Sara. Sin lugar a dudas es una de las personas que más se merecen estos agradecimientos. Después de más de 6 años trabajando con ella me ha demostrado que un jefe, mentor, director, llamadlo como queráis, puede ser brillante y perfeccionista en el trabajo y al mismo tiempo tener una calidad humana sobresaliente. Como sabéis, la ciencia es un mundo competitivo y exigente, donde a veces los trabajadores “en formación” son injustamente exigidos bajo condiciones laborales cuestionables. Algunos sabrán a lo que me refiero, pero tener un superior sin escrúpulos puede hacerte creer que no tienes derecho ni a hacer vacaciones, o a tener un bajón anímico. Perdonad que me enrolle, pero los que me conocéis bien sabéis que soy reivindicativo, ¡y que mejor que unos agradecimientos para hacer campaña! Afortunadamente, Sara siempre ha sido un apoyo para mí, y más allá de ser una fantástica directora de tesis, ha sido una persona cercana y en quien confiar a nivel personal. Una persona que, sin excepción, ha sido comprensiva con cualquier problema que me haya podido surgir durante estos años. Además, Sara tiene una habilidad especial: cada vez que la he llamado por teléfono preocupado por el tema que fuese, cuando acababa la llamada, de una forma u otra había conseguido tranquilizarme, y eso vale oro.

Estoy convencido que Catarina, mi co-directora gallega favorita estará muy de acuerdo con los comentarios anteriores. A ti también te tengo que agradecer por haberme acogido con mimo durante varios meses en Santiago, como si fuese mi casa, que, de hecho, lo es. Has estado siempre al pie del cañón con todos los proyectos (demasiados) al mismo tiempo, pero siempre positiva e intentando extraer lo mejor de cada uno de ellos. Siempre es un placer trabajar al lado de científicas tan motivadas y cargadas de energía. Seguro que nos volvemos a encontrar en el futuro “pola terra”.



Quiero darle también las gracias a Ramon por haberme dado la oportunidad de hacer la tesis en Gencardio, y felicitarlo por todo lo que ha logrado últimamente con su equipo de diagnóstico clínico.

Seguidamente, toca hablar de los compis de batalla. Los del día a día, donde se produce el roce, el cariño, las chispas y otras cosas (y no penséis mal). Marta... ya te vale eh, al final me adelantaste y estoy escribiendo esto pocos días después de que te proclamases Doctora con todas las de la ley. Felicidades, te lo mereces mucho, y te mereces un futuro espectacular, futuro que ya has empezado a labrar. Que decir, Marta, has sido una de las personas con quien mejor me lo he pasado en el lab, una amiga que me llevo, y en mi memoria siempre tendré esos días locos de experimentos (exitoso y fallidos), risas, cervezas y mucho más. Las horas de comer con Yajaira nunca estaban exentas de charlas interminables y a veces, terapia grupal... recuerda mis conejos Marta, o de lo contrario iré a por el matraz aforado... Y Yajaira, tu llegada revolucionó el laboratorio, le diste el toque de locura que le faltaba, y promoviste que muchos de nosotros nos conociésemos mejor. Tu paso por Gencardio fue más breve de lo esperado, pero ha sido para elegir un camino mejor y estoy seguro que no te vas a rendir nunca. Èric, *Western Blot master*, hace falta que alguien lo admita, desde que te fuiste de Gencardio el resto de predocs han perdido calidad de vida... solíamos ser los últimos en irnos del laboratorio mientras coincidimos y realmente disfruté todos los ratos que compartimos juntos en el lab. Estoy convencido que estas arrasando por USA, eres un crack y que sepas que por aquí te echamos de menos, especialmente Marcel, que sigue intentando buscarte un sustituto/a... (aunque le va a costar). Aprovecho también para mandarle saludos a Marcel desde Donostia, para que siempre sepas donde estoy, a ver si me acuerdo de mandarte nuestra foto vaciando el tanque de nitrógeno en el campo ;)

No me quiero olvidar de agradecer a Ferran, Mònica (las dos), Anna I, Anna F, Adrià, Marta P y Bernat por ayudarme desde la parte de diagnóstico con cualquier duda que hubiese podido tener. Gracias también a Mireia, Eli C y Eli S por aportar vuestro granito de arena en mi camino.

Me gustaría mandar un agradecimiento especial a Sara y Bet del ICO, las investigadoras más productivas del IDIBGI. No hace falta que diga que la mitad de los viajes al otro lado del laboratorio fueron para pedir asesoramiento. Gente que leáis estos agradecimientos: Bet es la fuente de sabiduría más abundante, simpática y amable que os podáis encontrar. Acudid en masa a ella cuando no sepáis cualquier cosa (problemas en cultivos, con clonajes, en la vida, del amor, que se te ha roto la transmisión del coche, también). Bet, hazme caso, monta una consultoría... Sara, a ti te deseo lo mejor (a ti también, eh, Bet), y más ahora que esperas a tu segundo bebé. Gracias por todas las charlas y ayuda que me has brindado todos estos años y por ser una tía "de puta madre".

A ver, que toca ahora... ¡Ah! Mis queridos/as estudiantes, habéis sido unos cuantos, la verdad. Antes de hablar de vosotros, quiero hablar de cuando yo fui el estudiante (qué tiempos aquellos...)

y agradecer a quien fue mi profesora en mis inicios en Gencardio: Mel·lina (no me olvidaba de ti, eh). Gracias por haber sido una gran mentora y haberme ayudado siempre en todo lo que he necesitado. Estoy convencido que mi etapa inicial contigo fue clave para haber llegado hasta el día de hoy. ¡Gracias!

Ahora sí, Ana T, Ana F, Marc, Natalia (¿la sucesora?) y Hugo, espero que hayáis aprendido tanto de mi como yo he aprendido de vosotros. Haber sido vuestro profe durante vuestros proyectos de máster fue una responsabilidad bonita y me ayudó también a ser mejor investigador, a mejorar aptitudes personales y a desarrollar habilidades de docencia. Con todos vosotros pasé buenos momentos en el lab trabajando codo con codo, pero tengo que hacer una mención especial a Hugo. No solo fui tu “sensei”, sino que ahora también somos amigos. Me enorgullece decir que el “sensei” ha superado al maestro. Ahora tienes una responsabilidad mayor con el pequeño Arán, pero tú y Judith lo estáis bordando. También os quiero dar las gracias a los dos por todas las quedadas con el resto del grupo, que espero sigamos repitiendo (aunque la distancia nos separe).

Doy un salto de más de 1000 kilómetros para aterrizar de nuevo en Santiago. Muchos de vosotros sabéis el amor y conexión que tengo con Galicia, donde además reside gran parte de mi familia. Tuve la suerte de poder combinar esto con mi trabajo y realizar 2 estancias de investigación en el CIMUS. La primera de ella fue más breve, de menos de un mes, pero muy intensa, productiva y exitosa. Quiero agradecer a José Tubío por darme la oportunidad de aprender secuenciación de lecturas largas (¿...os había dicho ya que hablo de ello en la página 238...?) y a todos los miembros de su grupo con los que pude trabajar durante ese tiempo. Me sentí muy acogido por todos, en un ambiente inmejorable, y fueron tres semanas que recuerdo con especial alegría. Alicia, gracias por ayudarme con los protocolos desde el principio y por el buen recibimiento que tuve. JR, el coco, tu tenías las respuestas a todos mis problemas. Eres una gran persona, currante como el que más, y entre los dos nos las arreglamos para tirar el experimento hacia delante (aunque algún día nos fuimos a las tantas de la noche). Ana P, me encantó conocerte y trabajar contigo, me ayudaste muchísimo durante todo el proceso y te deseo lo mejor. Espero que algún día repitamos una comida juntamente con Ana O y Paula O, gracias a las dos también por acogerme. Gracias también al resto de miembros que pude conocer más de cerca: Eva G, Seila D, Jeni B y Martin S.

Mi segunda estancia en el CIMUS fue más extensa, varios meses cargados de momentos. Quiero agradecer con cariño a Angel Carracedo (y de nuevo a Catarina) por hacer posible esa estancia. Debe ser cosa de Galicia, pero de nuevo, la hospitalidad y acogida que recibí fue inmejorable. Son muchas las personas que me ayudaron durante el día a día, haciendo que me adaptase rápidamente, como si fuese mi lab de toda la vida: Elena G, por todos tus trucos y consejos; Alba S, por tus consejos farmacéuticos (nunca volví a ser el mismo después del anginovag) y el apoyo

mutuo en cultivos; Raquel M, Sara D y Cristina G. Quiero también agradecer la amabilidad y compañerismo de Nerea G, Almudena G, Olalla M, Lucía S y Antón L del grupo de BioPharma. A Eduardo le quiero dedicar algunas palabras más. Gracias por aconsejarme y guiarme durante el arduo experimento que pusimos a punto y, a pesar de todas las dificultades, acabamos realizando. El esfuerzo valió la pena, pero quiero valorar la calma y el buen hacer con el que gestionaste los momentos de estrés, que al fin y al cabo, no dejan de ser parte de la vida en el laboratorio, y un aprendizaje más para mí. Muchas gracias por todo.

También quiero agradecer a la gente de Málaga que nos ayudó con los análisis de RNA-seq, especialmente a Pedro y a Elena, les deseo toda la suerte del mundo.

De manera especial, quiero darle las gracias a Sara por ser un gran apoyo durante una parte de este camino. Quiero agradecerte lo buena persona que fuiste siempre conmigo y todos los buenos momentos que compartimos, que hubo muchos. También hubo momentos complicados, exceso de trabajo y la ansiedad y las dudas que todos llegamos a tener en esta etapa profesional. Aun así, supiste sobrellevar mis malos momentos y sacar siempre la parte positiva.

Y obviamente, no podía escribir esto sin dar las gracias a mis amigos. Nunca se agradecieron tanto los planes, viajes y quedadas con ellos como durante esta etapa. A mi gente de toda la vida: David, como no nos veamos pronto te quito de los agradecimientos; Raquel, aunque a veces pasa un mes entre audio y audio (ya nos hemos acostumbrado), ahí seguimos; Lourdes, tú eres de las más antiguas, gracias por el apoyo de estos años, y dile a tu madre que echo de menos sus paninis; y Pau, tu sin más.

Es broma Pau, a ti te hago mención especial. Al pie del cañón día a día, a distancia o en persona. Podríamos decir que eres... ¿mi bro? Cryptobro, tripbro, un poco de todo. Bueno ya sabes, hemos hecho todo tipo de aventuras estos años. Desde subir montañas y perdernos innumerables veces en la bajada y llegar al coche con la última luz del día hasta viajar en solo un año y medio a Burdeos, Praga, Sevilla, Roma, Ibiza, Bélgica, Malta, Oporto, Viena, Bratislava y Budapest (visto así no sé cómo nos lo montamos), y luego a Copenhague, Ámsterdam y California. Espero que nos esperen muchos más, y espero que sigamos teniendo ideas locas como hasta ahora, eso no lo hay que perder nunca.

Quiero dar las gracias a mi grupo nacido durante el doctorado con Marta (a veces +1), Hugo, Iván, Yajaira, Judith y Pau, al cual infiltré entre nosotros, los científicos (y encajó). Hablando de científicos: Marcel (con paso efímero por Gencardio), es un chaval que decidió meterse en un doctorado también. Mucho ánimo Marcel, de todo se sale. Que va, te sobra capacidad para sacártelo mejor que yo. Aún recuerdo como nos conocimos formando equipo como únicos nacionales en nuestro Erasmus por Noruega, desde entonces, eres amigo y persona de confianza, aunque me ha faltado tiempo para que nos viésemos todo lo que me hubiera gustado. Lo mismo

digo para Cristina (aunque ella ya ha salido del agujero, felicidades doctora). Te conocí justo al empezar en Gencardio y fuiste un apoyo muy importante cuando realmente lo necesitaba, muchas gracias por todo!

Finalmente, quiero agradecer a mi familia, especialmente a mis padres y a mi hermana Cristina, quienes al final han sido mi núcleo familiar permanente. No puedo más que agradecerles que por cualquier circunstancia o en cualquier momento puedo contar con su ayuda y apoyo. Ellos no fallan. Con paciencia, año a año han ido viendo cómo me he ido desarrollando, sin exigir nada. Bueno, últimamente ya empezaba a oler mal la preguntita de... “Bueno, ¿y la tesis ya está?” ... Ya casi, falta poco... — contestaba yo, iracundo, mientras pensaba en mis adentros —  
“Sí, poco los \_ \_ \_ \_ \_”.

Llevo diciendo que falta poco los últimos 6 meses, ya sabéis cómo funciona esto. Hasta mi abuelo de 90 años le pregunta a mi madre desde Galicia si ya envié la puñetera tesis, ahí te va abuelo, te la dedico.

No quiero acabar (estos interminables agradecimientos) sin dar las gracias a quien sin duda ha tenido que soportar la peor parte de mi mal humor y mis malos momentos a causa de este bicho que estáis leyendo: Victoria, gracias por todo. Todo es todo. Este tiempo ha sido maravilloso, pero también ha sido un gran reto para los dos. Hemos compartido mucho en poco tiempo, y hemos tomado decisiones valientes. Estoy orgulloso de cada una de ellas y de ti. Gracias por el apoyo incalculable que me has dado durante este período. No hemos podido tener todo el tiempo que necesitábamos, pero al menos ahora mismo tengo una cosa clara: ya falta poco, y esta vez de verdad, porqué, de hecho:

Este es el punto y final.



Para mi familia y seres queridos,  
por su apoyo incondicional durante todo el camino



El misterio de la vida no es un problema a resolver,  
sino una realidad a experimentar

**Frank Herbert**

*-Dune-*





## Abreviaciones

AF	Frecuencia alélica
AV	Arritmia ventricular
Ca <sup>2+</sup>	Ion calcio
ChIP-seq	Inmunoprecipitación de cromatina seguida de secuenciación masiva
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
DHS	Sitios hipersensibles a la Dnasa I
DNA	Ácido desoxirribonucleico
DEGs	Genes diferencialmente expresados
DOX	Doxiciclina
ECG	Electrocardiograma
ES	Células madre embrionarias
FA	Fibrilación auricular
FDR	<i>False xiaemonxiry rate</i>
FT	Factor de transcripción
FTG	Factores de transcripción generales
GWAS	Estudios de asociación del genoma completo
HCM	Cardiomiocitos humanos
HDACs	Histonas desacetilasas
HDR	<i>Homology-directed repair</i>
HEK	Células de riñón embrionario humano
hiPSC	Células madre pluripotentes inducidas
hiPSC-CM	Cardiomiocitos derivados de hiPSC
HPO	Términos de ontología del fenotipo humano
IP	Inmunoprecipitación
K <sup>+</sup>	Ion potasio
KAT	Acetiltransferasas de lisinas
kb	Kilobases
kDa	Kilodaltons
KRAB	<i>Krüppel-associated box</i>
LD	<i>Linkage disequilibrium</i>

<b>LRS</b>	Secuenciación de lecturas largas
<b>MAVD</b>	Miocardiopatía arritmogénica del ventrículo derecho
<b>MCD</b>	Miocardiopatía dilatada
<b>MCH</b>	Miocardiopatía hipertrófica
<b>mRNA</b>	RNA mensajero
<b>MSC</b>	Muerta súbita cardíaca
<b>mV</b>	Milivoltios
<b>NA</b>	Nicotinamida
<b>Na<sup>+</sup></b>	Ion sodio
<b>Nav</b>	Canal de sodio dependiente de voltaje
<b>ncRNA</b>	RNA no codificante
<b>NGS</b>	Secuenciación de nueva generación
<b>NHEJ</b>	<i>Non-homologous DNA end joining</i>
<b>ONT</b>	<i>Oxford Nanopore Technologies</i>
<b>PA</b>	Potencial de acción cardíaco
<b>pb</b>	Pares de bases
<b>PBMCs</b>	Células mononucleares de sangre periférica
<b>PCHi-C</b>	<i>Promotore capture</i> Hi-C
<b>qPCR</b>	PCR cuantitativa en tiempo real
<b>RNPs</b>	Ribonucleoproteínas
<b>rpm</b>	Revoluciones por minuto
<b>RQ</b>	Cuantificación relativa
<b>SBr</b>	Síndrome de Brugada
<b>SE</b>	<i>Superenhancer</i>
<b>SNP</b>	Polimorfismo de un único nucleótido
<b>SNV</b>	Variantes de un único nucleótido
<b>SQTC</b>	Síndrome de QT largo
<b>SRS</b>	Secuenciación de lecturas cortas
<b>ssODN</b>	DNA donador de cadena simple
<b>SV</b>	Variantes estructurales
<b>TAD</b>	Dominio topológico asociado
<b>TFBS</b>	Motivo de unión de factores de transcripción

<b>TSA</b>	Tricostatina A
<b>TSS</b>	Sitio de inicio de la transcripción
<b>WB</b>	<i>Western Blot</i>
<b>WT</b>	<i>Wild-type</i> o fenotipo salvaje

## Índice de figuras

<b>Figura 1.</b> Anatomía del corazón.....	1
<b>Figura 2.</b> Sistema de conducción eléctrica del corazón.....	2
<b>Figura 3.</b> Canales iónicos dependientes de voltaje.....	3
<b>Figura 4.</b> Canal de sodio dependiente de voltaje cardiaco. ....	5
<b>Figura 5.</b> PA y corrientes iónicas del corazón. ....	7
<b>Figura 6.</b> Alteración de la función de Nav1.5 en SBr y SCTL (tipo 3).....	10
<b>Figura 7.</b> Composición del genoma humano. ....	13
<b>Figura 8.</b> Regulación de la expresión génica.....	16
<b>Figura 9.</b> Regulación transcripcional. ....	17
<b>Figura 10.</b> Diferentes tipos de modificaciones postraduccionales.....	18
<b>Figura 11.</b> Representación esquemática de un gen y su región promotora. ....	20
<b>Figura 12.</b> Representación esquemática de la función <i>enhancer</i> .....	21
<b>Figura 13.</b> Modelo de regulación transcripcional asociado a los SE.....	22
<b>Figura 14.</b> Representación esquemática de la función del <i>insulator</i> .....	23
<b>Figura 15.</b> Esquema de la morfogénesis cardíaca. ....	25
<b>Figura 16.</b> Estructura de los FT GATA.....	28
<b>Figura 17.</b> Procesos de acetilación y desacetilación en GATA4.....	31
<b>Figura 18.</b> Estructura de la cromatina.....	32
<b>Figura 19.</b> Alteración en el estado de compactación de la cromatina por modificaciones de histonas. ....	34
<b>Figura 20.</b> Coordinación entre las diferentes modificaciones de histonas. ....	36
<b>Figura 21.</b> Modificación de nucleosomas mediante complejos remodeladores.....	37
<b>Figura 22.</b> Organización jerárquica del genoma en mamíferos.. ....	38
<b>Figura 23.</b> Modelos propuestos para la formación de TADs.....	40
<b>Figura 24.</b> Regiones de cromatina abierta. ....	42
<b>Figura 25.</b> Perfil de modificaciones de histonas por ChIP-seq. ....	43
<b>Figura 26.</b> Análisis de la organización del genoma. ....	44
<b>Figura 27.</b> Representación de diferentes tipos de variantes (SNVs, indels y estructurales) en comparación con el genoma de referencia.....	45
<b>Figura 28.</b> Mecanismos de reparación del DNA para corregir roturas de doble cadena.....	48
<b>Figura 29.</b> Identificación de variantes genéticas en función de la frecuencia del alelo de riesgo y la magnitud del efecto genético. ....	50
<b>Figura 30.</b> Efectos de las variantes reguladoras en <i>cis</i> .....	53
<b>Figura 31.</b> Flujo de trabajo principal del análisis genético con NGS.....	59
<b>Figura 32.</b> Funcionamiento de la tecnología de Oxford Nanopore.....	60

<b>Figura 33.</b> Obtención de hiPSC a partir de células somáticas y diferenciación a otros tipos celulares.....	64
<b>Figura 34.</b> Elementos que constituyen el sistema CRISPR/Cas9 como herramienta de edición genómica.....	66
<b>Figura 35.</b> Sistema dCas9-KRAB para <i>knockdown</i> dirigido.....	68
<b>Figura 36.</b> Efectos potenciales de las variantes no codificantes asociadas con el síndrome de Brugada (SBr).....	70
<b>Figura 37.</b> Ubicación genómica de SNVs comunes relevantes en el promotor de <i>SCN5A</i> .....	72
<b>Figura 38.</b> Procedimiento seguido por Pinsach-Abuin y colaboradores. <sup>351</sup> para la determinación de variantes no codificantes asociadas a SBr.....	75
<b>Figura 39.</b> Funcionamiento de la línea WTB CRISPRi Gen1B.....	89
<b>Figura 40.</b> Mapa del plásmido pCGN (Promega).....	91
<b>Figura 41.</b> Mapa del plásmido pgRNA-CKB.....	93
<b>Figura 42.</b> Amplicones requeridos para los análisis TIDER.....	97
<b>Figura 43.</b> Esquema del protocolo de diferenciación de hiPSCs a hiPSC-CMs.....	104
<b>Figura 44.</b> Esquema de funcionamiento del método <i>QuikChange Lightning site-directed mutagenesis</i> .....	107
<b>Figura 45.</b> Ejemplo de co-transfección de H9c2 en placa de 12 pocillos.....	113
<b>Figura 46.</b> Esquema global del proceso para generar líneas CRISPRi con expresión de una gRNA.....	114
<b>Figura 47.</b> Experimento de <i>knockdown</i> sobre <i>SCN5A</i> .....	115
<b>Figura 48.</b> Ejemplo esquemático de selección de gRNAs en la región reguladora <i>SCN5A</i> .....	116
<b>Figura 49.</b> Esquema del sistema <i>Golden Gate</i> .....	117
<b>Figura 50.</b> Ejemplo de dos electroferogramas de <i>Bioanalyzer</i> .....	121
<b>Figura 51.</b> Análisis de datos de la cuantificación relativa de la qPCR.....	124
<b>Figura 52.</b> Perfiles de excitación y emisión para SBFI-AM y los fluoroforos expresados por la línea WTB CRISPRi Gen1B (mCherry y mKate2).....	125
<b>Figura 53.</b> Sistema Operetta y placa de 384 pocillos utilizada para replaquear los hiPSC-CM y medir el sodio intracelular en el sistema Operetta.....	126
<b>Figura 54.</b> Esquema de los pasos seguidos para la secuenciación mediante ONT.....	126
<b>Figura 55.</b> Dispositivo minION con una <i>flowcell</i> preparada para ser usada.....	128
<b>Figura 56.</b> Pasos seguidos para obtener las muestras para el experimento de RNA-seq.....	129
<b>Figura 57.</b> Esquema del <i>workflow</i> implementado en el análisis de muestras de RNA-seq.....	133
<b>Figura 58.</b> Esquema de las 6 estrategias utilizadas para optimizar la eficiencia de la edición con CRISPR/Cas9.....	134
<b>Figura 59.</b> Pasos seguidos durante el proceso de edición.....	137

<b>Figura 60.</b> Esquema de los péptidos estudiados mediante espectrometría de masas por la Dra. Anna Tarradas.....	142
<b>Figura 61.</b> Sustituciones de lisinas a alaninas en el dominio de unión a DNA de GATA4. ....	143
<b>Figura 62.</b> Niveles de acetilación de GATA4 en presencia o ausencia de inhibidores de HDACs (TSA y NA). ....	144
<b>Figura 63.</b> Niveles de acetilación de GATA4 en mutantes simples de GATA4.....	145
<b>Figura 64.</b> Cuantificación de los niveles de acetilación de los mutantes de GATA4.....	146
<b>Figura 65.</b> Análisis de los niveles de acetilación de mutantes de GATA4 mediante comprobaciones múltiples de Dunnett. ....	147
<b>Figura 66.</b> Niveles de acetilación GATA4 en mutantes simples, dobles y triples de GATA4.....	148
<b>Figura 67.</b> Cuantificación de los niveles de acetilación de los mutantes simples, DM y TM de GATA4.....	149
<b>Figura 68.</b> Análisis de los niveles de acetilación de los mutantes simples, DM y TM de GATA4 mediante comprobaciones múltiples de Dunnett. ....	150
<b>Figura 69.</b> Efecto del incremento de GATA4 sobre la actividad de los promotores <i>SCN5A</i> y <i>PDX1</i> . .....	151
<b>Figura 70.</b> Efecto de los mutantes de GATA4 sobre la actividad del promotor <i>SCN5A</i> .....	152
<b>Figura 71.</b> Efecto de los mutantes de GATA4 sobre la actividad del promotor <i>PDX1</i> . ....	154
<b>Figura 72.</b> Preselección de las regiones reguladoras de estudio en base a los parámetros indicados. .....	156
<b>Figura 73.</b> Región promotora de <i>EXOG</i> (región 1). ....	158
<b>Figura 74.</b> Región del intrón 1 de <i>SCN5A</i> (región 2).....	159
<b>Figura 75.</b> Región del TSS de <i>SCN5A</i> (región 3). ....	160
<b>Figura 76.</b> Regiones 4 y 5 en el 5' <i>SCN5A</i> .....	161
<b>Figura 77.</b> Región intrónica de <i>SCN10A</i> (regiones 6 y 7). ....	163
<b>Figura 78.</b> Esquema general del proceso seguido en los ensayos CRISPRi.....	165
<b>Figura 79.</b> Imagen de hiPSCs y hiPSC-CMs.....	166
<b>Figura 80.</b> Imagen de hiPSC-CMs después del replaqueo.....	166
<b>Figura 81.</b> Imagen de la expresión de mKate2 y mCherry en células hiPSC-CMs en el microscopio de fluorescencia. ....	167
<b>Figura 82.</b> Expresión de los marcadores de pluripotencia <i>NANOG</i> y <i>OCT4</i> . ....	169
<b>Figura 83.</b> Análisis de la expresión de <i>SCN5A</i> en las líneas CRISPRi hiPSC-CM.....	171
<b>Figura 84.</b> Análisis de la expresión de <i>SCN5A</i> en las líneas CRISPRi hiPSC-CM mediante intervalos de confianza. ....	173
<b>Figura 85.</b> Esquema del procedimiento utilizado para analizar el sodio intracelular en las líneas CRISPRi.....	175

<b>Figura 86.</b> Análisis del sodio intracelular en las líneas celulares CRISPRi hiPSC-CMs, en la placa de 384 pocillos.....	176
<b>Figura 87.</b> Análisis del sodio intracelular en las líneas celulares CRISPRi hiPSC-CMs, en la placa de 384 pocillos .....	177
<b>Figura 88.</b> Comparación de la concentración de sodio intracelular entre las condiciones -DOX y +DOX de las rondas analizadas con Operetta. ....	180
<b>Figura 89.</b> Análisis de la señal de mKate2/mCherry en las líneas CRISPRi hiPSC-CMs tratadas con DOX en función del tiempo.....	182
<b>Figura 90.</b> ChIA-PET de CTCF en la región EXOG-SCN5A.....	184
<b>Figura 91.</b> Análisis de las interacciones captadas por PChi-C (hiPSC-CMs) y ChIA-PET (GM12878) en la región SCN5A-SCN10A.....	186
<b>Figura 92.</b> Vista global de las interacciones en el TAD de SCN5A captadas por PChi-C en hiPSC-CM .....	187
<b>Figura 93.</b> Representación esquemática de las interacciones de Hi-C indicadas en la Tabla 18, obtenidas a partir de fibroblastos humanos.....	188
<b>Figura 94.</b> Composición de los haplotipos en el locus SCN5A-SCN10A.....	191
<b>Figura 95.</b> Método de comprobación experimental de los haplotipos mediante secuenciación con ONT.....	192
<b>Figura 96.</b> Amplificación del fragmento de 13 kb correspondiente a la región del enhancer SCN5A-SCN10A que alberga los 7 SNVs que configuran el haplotipo.....	193
<b>Figura 97.</b> Amplificación heterogénea del fragmento de 13 kb en diferentes muestras con <i>long-range</i> PCR.....	194
<b>Figura 98.</b> Productos de <i>nested</i> PCR amplificados a partir de los productos <i>long-range</i> originales.....	194
<b>Figura 99.</b> Frecuencias de los haplotipos en la cohorte de SBr y los grupos control.....	196
<b>Figura 100.</b> Haplotipos validados experimentalmente mediante secuenciación por Oxford Nanopore.....	196
<b>Figura 101.</b> Cariotipo de la línea celular Ctrl BS PBiPS37-Sv4F-1 (GPG2).....	198
<b>Figura 102.</b> Marcadores de pluripotencia de la línea celular Ctrl BS PBiPS37-Sv4F-1 (GPG2) .	200
<b>Figura 103.</b> Diferenciación <i>in vitro</i> de la línea GPG2 a endodermo, ectodermo y mesodermo..	199
<b>Figura 104.</b> Diferenciación <i>in vitro</i> de la línea SBr a endodermo, ectodermo y mesodermo .....	202
<b>Figura 105.</b> Cariotipo de la línea celular BS PBiPS46-Sv4F-10 (SBr).....	201
<b>Figura 106.</b> Resultados de secuenciación por Sanger para confirmar la presencia de las siete SNVs del haplotipo homocigoto Hap <sup>1/1</sup> en las hiPSC BrS.....	200
<b>Figura 107.</b> Marcadores de pluripotencia de la línea celular BS PBiPS46-Sv4F-10 (SBr).....	201
<b>Figura 108.</b> Número total de lecturas antes y después del <i>trimming</i> .....	203
<b>Figura 109.</b> Número total de lecturas <i>post-trimming</i> mapeadas con el genoma humano.....	203



<b>Figura 110.</b> Correlación entre las muestras analizadas en el RNA-seq. ....	204
<b>Figura 111.</b> Clasificación de los genes analizados en el RNA-seq según el tipo de expresión diferencial obtenido.....	205
<b>Figura 112.</b> Detección de DEGs por los paquetes de análisis DESeq2, edgeR, limma y NOISeq.. .....	206
<b>Figura 113.</b> Procesos biológicos GO asociados a los DEGs. ....	208
<b>Figura 114.</b> Matriz de similitud semántica de los DEGs en relación con los términos HPO seleccionados.....	209
<b>Figura 115.</b> Validación de los resultados de RNA-seq mediante qPCR.....	211
<b>Figura 116.</b> Validación de los resultados de RNA-seq mediante qPCR.....	212
<b>Figura 117.</b> Determinación de la concentración de sodio intracelular en las líneas celulares GPG2 y SBr. ....	214
<b>Figura 118.</b> Composición de los genotipos Hap <sup>1/1</sup> y Hap <sup>3/3</sup> del locus SCN5A-SCN10A.....	215
<b>Figura 119.</b> Eficiencia de edición en presencia o ausencia de <i>HDR Enhancer</i> .....	218
<b>Figura 120.</b> Ejemplos de electroferogramas para los diferentes tipos de edición obtenidos. ....	219
<b>Figura 121.</b> Actividad HDR en los clones secuenciados a partir de la línea SBr + <i>HDR Enhancer</i> . .....	220
<b>Figura 122.</b> Haplotipo obtenido después de la edición homocigota de la variante rs6801957. ....	221
<b>Figura 123.</b> Modelo de interacciones del promotor <i>EXOG</i> en el locus SCN5A-SCN10A.....	237
<b>Figura A- 1.</b> Niveles de acetilación de GATA4 en mutantes simples de GATA4. ....	288
<b>Figura A- 2.</b> Niveles de acetilación de GATA4 en mutantes simples, dobles y triples de GATA4.. .....	288

## Índice de tablas

<b>Tabla 1.</b> Cebadores para la mutagénesis dirigida.....	94
<b>Tabla 2.</b> Secuencias de los gRNA diseñados y localización de la diana. ....	95
<b>Tabla 3.</b> Cebadores diseñados para la edición de la variante rs6801957. ....	96
<b>Tabla 4.</b> Cebadores diseñados para el análisis con TIDER.....	96
<b>Tabla 5.</b> Cebadores diseñados para las reacciones de <i>long-range</i> PCR. ....	97
<b>Tabla 6.</b> Cebadores diseñados para validación por secuenciación de Sanger.....	98
<b>Tabla 7.</b> Sondas TaqMan para qPCR de los experimentos CRISPRi.....	98
<b>Tabla 8.</b> Anticuerpos para los experimentos de inmunoprecipitación y Western blot. ....	99
<b>Tabla 9.</b> Composición del medio de diferenciación.....	104
<b>Tabla 10.</b> Suplementos añadidos al medio sin glucosa SILAC.....	105
<b>Tabla 11.</b> Composición del medio de digestión usado en el proceso de replaqueo. ....	106
<b>Tabla 12.</b> Soluciones usadas durante los procesos de IP, electroforesis y Western blot. ....	112
<b>Tabla 13.</b> Composición de los geles de acrilamida.....	112
<b>Tabla 14.</b> Ejemplo de volúmenes totales de DNA para el caso de la Figura 9 .....	113
<b>Tabla 15.</b> Parámetros de calidad registrados en las muestras usadas para RNA-seq. ....	130
<b>Tabla 16.</b> Líneas CRISPRi generadas a partir de la integración del pgRNA-CKB-gRNA especificado. ....	164
<b>Tabla 17.</b> Mediciones del sodio intracelular (+/- DOX) para las líneas celulares CRISPRi hiPSC-CMs.....	179
<b>Tabla 18.</b> Interacciones de Hi-C entre regiones CRISPRi del locus <i>SCN5A-SCN10A</i> . ....	188
<b>Tabla 19.</b> Expresión diferencial en el RNA-seq de los 10 genes con mayor similitud en el análisis de similitud semántica.....	210
<b>Tabla A- 1.</b> Términos HPO utilizados para el análisis de similitud semántica.....	281

# Índice general

Agradecimientos .....	i
Abreviaciones .....	xi
Índice de figuras .....	xiv
Índice de tablas .....	xix
Resumen .....	xxvii
Resumo .....	xxxix
Resum .....	xxxv
Summary .....	xxxix
<b>I. INTRODUCCIÓN .....</b>	
<b>1. El corazón .....</b>	<b>1</b>
1.1. Anatomía y función del corazón .....	1
1.2. Actividad eléctrica cardíaca .....	2
1.3. Canales iónicos dependientes de voltaje .....	3
1.3.1. Canales de sodio .....	4
1.3.2. Canales de calcio y potasio .....	5
1.4. Potencial de acción cardíaco .....	6
<b>2. La muerte súbita cardíaca .....</b>	<b>8</b>
2.1. Factores de riesgo y etiología .....	8
2.2. Enfermedades eléctricas o canalopatías cardíacas .....	9
<b>3. El genoma humano .....</b>	<b>12</b>
3.1. Organización del genoma humano .....	12
3.2. Regulación de la expresión génica .....	16
3.2.1. Elementos reguladores en <i>cis</i> .....	19
3.2.1.1. Promotores .....	19
3.2.1.2. <i>Enhancers</i> .....	20
3.2.1.3. Silenciadores .....	23
3.2.1.4. <i>Insulators</i> o elementos aislantes .....	23
3.3. Los factores de transcripción .....	24
3.3.1. Regulación de la expresión génica cardíaca por factores de transcripción .....	25
3.3.2. El factor de transcripción GATA4 cardíaco .....	27
3.3.2.1. Estructura y función .....	27
3.3.2.2. La acetilación de GATA4 .....	29
3.4. La cromatina .....	31
3.4.1. Estructura de la cromatina .....	31
3.4.2. Dinámica de la cromatina .....	33
3.4.3. Organización tridimensional de la cromatina .....	38
3.4.3.1. <i>Insulated neighborhoods</i> .....	39

3.4.3.2. Dominios topológicos asociados.....	39
3.4.3.3. Compartimientos celulares.....	40
3.4.3.4. Territorios cromosómicos.....	41
3.5. Identificación de elementos reguladores de la transcripción .....	41
3.5.1. Análisis de los lugares de unión a factores de transcripción .....	41
3.5.2. Análisis del patrón de modificaciones de histonas.....	43
3.5.3. Análisis de la organización tridimensional del genoma .....	43
<b>4. La variación genética .....</b>	<b>45</b>
4.1. Tipos de variación genética .....	45
4.2. Fuentes de variación genética .....	46
4.3. El rol de la variación genética en las enfermedades humanas .....	49
4.4. Efectos funcionales de las variantes no-codificantes .....	51
<b>5. Técnicas genómicas para la identificación de nuevas variantes .....</b>	<b>55</b>
5.1. Genómica en patología cardiovascular .....	55
5.2. <i>Next-Generation Sequencing</i> (NGS) .....	56
5.3. Oxford Nanopore Technologies (ONT) .....	60
<b>6. Modelos celulares para el estudio de variantes que afectan a los canales de sodio dependientes de voltaje.....</b>	<b>63</b>
6.1. Sistemas de expresión heterólogos .....	63
6.2. Células específicas de tejido derivadas de células madre pluripotentes inducidas (hiPSC) .....	64
6.2.1. Cardiomiocitos derivados de hiPSC (hiPSC-CM).....	65
6.3. Modelos hiPSC con edición genómica .....	65
6.3.1. CRISPR/Cas9 .....	65
6.3.2. CRISPRi .....	67
<b>7. El Síndrome de Brugada.....</b>	<b>69</b>
7.1. El rol de las variantes no codificantes en el Síndrome de Brugada .....	69
7.1.1. El promotor de <i>SCN5A</i> .....	71
7.1.2. El locus <i>SCN5A-SCN10A</i> .....	73
7.1.3. RNAs no-codificantes.....	74
7.1.4. Haplotipos asociados a Síndrome de Brugada .....	75
<b>II. JUSTIFICACIÓN Y OBJETIVOS.....</b>	<b>79</b>
<b>III. MATERIALES Y MÉTODOS .....</b>	<b>85</b>
<b>1. Materiales.....</b>	<b>87</b>
1.1. Muestras .....	87
1.1.1. Cohorte de pacientes con síndrome de Brugada.....	87
1.1.2. Individuos para la generación de modelos iPSC .....	87
1.2. Modelos celulares experimentales .....	88
1.2.1. HEK293T .....	88
1.2.2. H9c2 .....	88

1.2.3. WTB CRISPRi Gen1B .....	88
1.2.4. Ctrl BS PBIpS37-Sv4F-1 y BS PBIpS46-Sv4F-10.....	89
1.2.5. BS PBIpS Cas9- <i>induced</i> .....	90
1.2.6. Cardiomiocitos derivados de hiPSC (hiPSC-CM).....	90
1.3. Vectores de expresión .....	90
1.3.1. Vectores reporteros luciferasa.....	90
1.3.2. Vectores de expresión de GATA4.....	91
1.3.3. Vectores de expresión de gRNAs.....	92
1.4. Cebadores .....	93
1.4.1. Cebadores para mutagénesis dirigida.....	93
1.4.2. Cebadores para CRISPR/Cas9 y CRISPRi.....	94
1.4.2.1. Cebadores para los gRNAs de ensayos CRISPRi .....	94
1.4.2.2. Cebadores para los experimentos de <i>knock-in</i> con CRISPR/Cas9 .....	96
1.4.3. Cebadores para análisis TIDER.....	96
1.4.4. Cebadores para <i>long-range</i> PCR.....	97
1.4.5. Cebadores para validación por secuenciación de Sanger.....	98
1.5. Sondas <i>TaqMan</i> .....	98
1.6. Anticuerpos .....	99
1.7. Bases de datos públicas y recursos bioinformáticos .....	99
<b>2. Métodos.....</b>	<b>101</b>
2.1. Cultivo de las líneas celulares HEK y H9c2 .....	101
2.1.1. Mantenimiento.....	101
2.1.2. Congelación y descongelación .....	101
2.1.3. Subcultivo.....	102
2.2. Reprogramación celular de células somáticas a hiPSC .....	102
2.3. Cultivo y diferenciación a hiPSC .....	102
2.3.1. Congelación y descongelación .....	102
2.3.2. Mantenimiento.....	103
2.3.3. Subcultivo.....	103
2.3.4. Diferenciación de hiPSC a hiPSC-CMs.....	103
2.3.4.1. Replaqueo.....	105
2.4. Generación de mutantes de GATA4 .....	106
2.4.1. Mutagénesis dirigida.....	106
2.4.2. Transformación de bacterias ultracompetentes .....	107
2.4.3. Mini-preparaciones de DNA plasmídico .....	108
2.4.4. Secuenciación de los plásmidos .....	108
2.4.5. Maxi-preparaciones de DNA plasmídico .....	109
2.4.6. Stocks de glicerol.....	109
2.5. Determinación de los niveles de acetilación de GATA4 .....	109
2.5.1. Transfección de células HEK293T .....	109
2.5.2. Preparación y cuantificación de los lisados celulares.....	110
2.5.3. Inmunoprecipitación.....	110

2.5.4. Electroforesis y Western blot.....	111
2.6. Determinación de la actividad transcripcional de GATA4 acetilada en experimentos luciferasa ....	112
2.6.1. Transfección de H9c2 .....	112
2.6.2. Lectura de la actividad luciferasa y <i>renilla</i> .....	113
2.7. Cribaje de regiones reguladoras de <i>SCN5A</i> utilizando un sistema CRISPRi .....	114
2.7.1. Selección de las regiones de estudio y diseño de guías.....	115
2.7.2. <i>Annealing</i> .....	116
2.7.3. Clonaje de las guías en el vector pgRNA-CKB .....	116
2.7.3.1. Método Golden Gate.....	116
2.7.3.2. Transformación de bacterias DH5 $\alpha$ .....	118
2.7.3.3. Secuenciación y validación del clonaje .....	118
2.7.4. Nucleofección de la línea CRISPRi .....	118
2.7.4.1. Selección de líneas estables .....	119
2.7.5. Rondas de diferenciación .....	119
2.7.6. Extracción de RNA .....	120
2.7.7. Control de calidad de muestras de RNA con <i>Bioanalyzer</i> .....	120
2.7.8. Retrotranscripción.....	121
2.7.9. PCR cuantitativa .....	121
2.7.9.1. Análisis de los datos de qPCR.....	122
2.8. Experimentos funcionales en condiciones de <i>knockdown</i> de <i>SCN5A</i> .....	125
2.8.1. Medición de la concentración de sodio intracelular .....	125
2.9. Secuenciación de haplotipos mediante Oxford Nanopore Sequencing .....	126
2.9.1. <i>Long-range</i> PCR .....	127
2.9.2. Preparación de <i>libraries</i> y secuenciación con el dispositivo minION .....	127
2.9.3. Análisis de la secuenciación de lecturas largas.....	129
2.10. Experimentos de RNA-seq .....	129
2.10.1. Preparación de las muestras y secuenciación .....	129
2.10.2. Análisis del experimento RNA-seq.....	130
2.11. Edición del haplotipo con CRISPR/Cas9 .....	134
2.11.1. Diseño de 6 estrategias de edición diferentes .....	134
2.11.2. Protocolo de edición CRISPR/Cas9 .....	135
2.11.3. Comparativa de las 6 estrategias mediante el software TIDER.....	136
2.12. Análisis estadísticos .....	136
<b>IV. RESULTADOS .....</b>	<b>139</b>
<b>1. CAPÍTULO I. Determinación de las lisinas acetiladas en GATA4 y su efecto sobre el promotor <i>SCN5A</i> .....</b>	<b>141</b>
1.1. Resultados previos a esta tesis y contexto del estudio .....	141
1.2. Obtención de mutantes GATA4 mediante sustitución de lisinas potencialmente acetilables .....	143
1.3. Análisis de los niveles de acetilación de GATA4 con mutantes simples .....	144
1.4. Análisis de los niveles de acetilación de GATA4 con mutantes dobles y triples .....	147
1.5. Efecto de la acetilación de GATA4 sobre la actividad del promotor de <i>SCN5A</i> .....	150

<b>2. CAPÍTULO II. Cribado de regiones reguladoras de SCN5A</b> .....	<b>155</b>
2.1. Selección de las regiones CRISPRi .....	155
2.2. Obtención de cardiomiocitos derivados de las líneas CRISPRi clonadas con las gRNAs .....	164
2.3. Análisis del proceso de diferenciación de hiPSCs a cardiomiocitos por qPCR .....	168
2.4. Análisis del efecto de las gRNAs sobre la expresión de SCN5A .....	169
2.5. Alteraciones del sodio intracelular en experimentos funcionales .....	174
2.6. Estudio comparativo entre las regiones CRISPRi y la información disponible de interacciones de cromatina .....	183
<b>3. CAPÍTULO III. Estudio del haplotipo en el locus SCN5A-SCN10A asociado al Síndrome de Brugada</b> .....	<b>190</b>
3.1. Resultados previos a esta tesis y contexto del estudio .....	190
3.2. Determinación experimental de los haplotipos con nanopore sequencing .....	193
3.2.1. Obtención de fragmentos de 13kb a partir de <i>long-range</i> PCR .....	193
3.2.2. Secuenciación y comprobación experimental de los haplotipos .....	195
3.3. Diferencias en la expresión global asociadas a un paciente con SBr y Hap <sup>1/1</sup> : experimento de RNA-seq.....	197
3.3.1. Caracterización de las líneas celulares SBr y GPG2.....	198
3.3.1.1. Línea GPG2.....	198
3.3.1.2. Línea SBr .....	200
3.3.2. RNA-seq: calidad de las lecturas y el alineamiento.....	202
3.3.3. Análisis de la expresión diferencial .....	204
3.3.4. Implicación funcional de los genes diferencialmente expresados.....	207
3.3.5. Similitud semántica asociada a enfermedades cardiovasculares.....	208
3.3.6. Validación de los resultados de RNA-seq por qPCR.....	211
3.4. Análisis de la concentración de sodio intracelular en hiPSC-CMs derivados de GPG2 y SBr .....	213
3.5. Edición del haplotipo por CRISPR/Cas9 .....	215
3.5.1. Obtención de ediciones HDR homocigotas y heterocigotas con elevada eficiencia .....	215
<b>V. DISCUSIÓN</b> .....	<b>223</b>
<b>1. Acetilación del GATA4 y su efecto sobre la regulación del promotor SCN5A</b> .....	<b>225</b>
<b>2. Identificación y caracterización de regiones genómicas específicas con función reguladora sobre el gen SCN5A</b> .....	<b>229</b>
2.1. Interacción de las regiones CRISPRi con otros elementos del genoma .....	229
2.2. Limitaciones, fortalezas y factores de variabilidad en el estudio CRISPRi .....	233
2.3. Conclusiones generales y futuros estudios .....	236
<b>3. Determinación y estudio del haplotipo asociado al síndrome de Brugada. Avances y futuras investigaciones</b> .....	<b>238</b>
3.1. Comprobación experimental del haplotipo mediante Oxford Nanopore sequencing y comparativa con los grupos control .....	238

3.2. Análisis del perfil de expresión génica en las líneas SBr y GPG2 .....	240
<b>VI. CONCLUSIONES.....</b>	<b>245</b>
<b>VII. BIBLIOGRAFÍA .....</b>	<b>251</b>
<b>Anexo 1.....</b>	<b>275</b>
<b>Anexo 2.....</b>	<b>283</b>
<b>Anexo 3.....</b>	<b>287</b>





## Resumen

El síndrome de Brugada (SBr) es una enfermedad eléctrica cardíaca asociada a una elevada susceptibilidad a arritmias ventriculares y muerte súbita cardíaca. Variantes genéticas en las regiones codificantes del gen *SCN5A*, que codifica para el canal de sodio cardíaco, explican el 30% de los casos de SBr. De forma similar, variantes genéticas en regiones codificantes de otros genes como los que codifican para las subunidades  $\beta$  reguladoras del canal de sodio cardíaco, los canales de calcio u otras proteínas accesorias, explican el 5% de los casos de SBr. En total, las variantes codificantes afectando los genes de canales iónicos, así como sus subunidades reguladoras, explican un 30-35% de los casos de SBr. Por lo tanto, en una elevada proporción de los pacientes diagnosticados con SBr, la etiología de la enfermedad sigue siendo desconocida.

Los estudios realizados durante la última década han demostrado que el locus *SCN5A-SCN10A* juega un papel fundamental en el control de la actividad transcripcional del gen *SCN5A*, además de estar asociado con alteraciones en los parámetros electrocardiográficos, lo que sugiere un papel relevante en la conducción y la función cardíaca. Todos los componentes reguladores que interactúan en el locus son necesarios para la adecuada expresión de *SCN5A* y la conducción cardíaca. Estas evidencias apuntan a que una alteración de la actividad de los factores de transcripción (FT) que regulan la transcripción de *SCN5A*, o bien la presencia de variantes en estas regiones podrían inducir alteraciones en la conducción cardíaca. En este sentido, estudios anteriores del grupo: (1) revelaron la asociación de un haplotipo formado por siete SNVs en el locus *SCN5A-SCN10A* con SBr mediante la secuenciación de las regiones reguladoras de una cohorte de 86 individuos con SBr<sup>351</sup>; y (2) identificaron el FT GATA4 como regulador transcripcional del gen *SCN5A*<sup>157</sup>. Además, estudios preliminares del laboratorio demostraron que la acetilación de GATA4 modula su función como activador transcripcional del promotor de *SCN5A*.

Partiendo de estas observaciones, nuestros objetivos se centraron en: (1) entender con profundidad los mecanismos moleculares que regulan la transcripción de *SCN5A*; (2) identificar posibles variantes reguladoras que puedan alterar la transcripción de *SCN5A*; y (3) relacionar estos factores con posibles alteraciones en las corrientes de sodio i arritmias. Para cumplir con estos propósitos, se desarrollaron tres enfoques distintos.

En primer lugar, se estudió la acetilación del FT cardíaco GATA4 y su efecto en la regulación transcripcional del gen *SCN5A*. Ensayos de acetilación mediante mutagénesis dirigida en células HEK293T permitieron identificar tres lugares de acetilación (K318, K322 y K326) en el dominio de unión al DNA de GATA4. Además, experimentos reporteros en células cardíacas H9c2 demostraron que la acetilación del residuo K318 de GATA4 es importante para su actividad transcripcional sobre el promotor de *SCN5A*.

En segundo lugar, nos propusimos identificar y caracterizar nuevas regiones reguladoras en el locus *SCN5A-SCN10A* utilizando una aproximación sistemática. Para ello, se realizó un cribado en diferentes regiones del locus *SCN5A-SCN10A* mediante ensayos CRISPRi en un modelo de cardiomiocitos derivados de hiPSCs (hiPSC-CMs). Los resultados del ensayo permitieron identificar siete secuencias específicas asociadas a la regulación transcripcional del gen *SCN5A* (*EXOG-CTCF*, Intrón 1-*NKX2-5*, TSS-39pb, TSS+183pb, 5'*SCN5A-GATA4*, *SCN10A-GATA4* y *SCN10A-rs6801957*). Nuestro estudio describe, por primera vez, que la región del promotor del gen *EXOG* actúa como región reguladora de la expresión de *SCN5A*. Además, el bloqueo de la actividad transcripcional de la región *EXOG* altera los niveles de sodio intracelular en el modelo de hiPSC-CMs.

En tercer lugar, se evaluó el efecto de las variantes genéticas en las regiones reguladoras del locus *SCN5A-SCN10A*. Concretamente se profundizó en el estudio del haplotipo de siete SNVs asociado a SBr. En primer lugar, se validó experimentalmente el *phasing* de los haplotipos en la cohorte de 86 individuos con SBr mediante secuenciación de lecturas largas con Oxford Nanopore Sequencing. Los resultados de la secuenciación permitieron confirmar 83 de los 86 haplotipos obtenidos con algoritmos de predicción a partir de datos de lecturas cortas de Illumina. A continuación, se evaluaron las alteraciones en el perfil transcripcional de dos líneas de hiPSC-CMs mediante un experimento de RNA-seq. Se compararon hiPSC-CMs de un individuo con SBr y haplotipo de riesgo y hiPSC-CMs de un individuo control con haplotipo de no-riesgo. Este estudio reveló más de 3000 genes diferencialmente expresados. La aplicación de filtros de priorización determinó que los grupos funcionales más enriquecidos en el análisis fueron genes asociados a la conducción y la diferenciación cardíacas, siendo *SCN5A* el gen más significativo en el análisis. Además, la línea de hiPSC-CMs con el haplotipo asociado al SBr mostró una reducción en los niveles de sodio intracelular de un 36.5% respecto a los hiPSC-CMs control. Por último, mediante CRISPR/Cas9, se editó una de las tres posiciones necesarias para revertir el haplotipo asociado a SBr en la línea hiPSC del paciente con SBr.

En resumen, el presente estudio demuestra que la expresión del gen *SCN5A* está controlada por diferentes mecanismos de regulación transcripcional, tanto a través de la interacción de FT con elementos reguladores o por la presencia de variantes en regiones clave para la regulación de *SCN5A*. El efecto de *GATA4* sobre la expresión de *SCN5A*, el efecto del haplotipo de riesgo en un modelo de cardiomiocitos con SBr y las diferentes regiones del locus *SCN5A-SCN10A* involucradas en la regulación de *SCN5A* son los principales hallazgos que respaldan estas conclusiones. Además, este estudio demuestra la potencial aplicación de la estrategia CRISPRi como herramienta para detectar nuevas regiones que podrían ser relevantes en el diagnóstico futuro de la enfermedad.





## Resumo

A Síndrome de Brugada (SBr) é unha enfermidade eléctrica cardíaca asociada cunha alta susceptibilidade ás arritmias ventriculares e á morte súbita cardíaca. Variantes xenéticas nas rexións codificantes do xene *SCN5A*, que codifica a canle cardíaca de sodio, explican o 30% dos casos de SBr. Do mesmo xeito, variantes xenéticas nas rexións codificantes doutros xenes como os que codifican as subunidades  $\beta$  reguladoras da canle cardíaca de sodio, canles de calcio ou outras proteínas accesorias, explican o 5% dos casos de SBr. En total, as variantes codificantes afectando aos xenes das canles iónicas, así como as súas subunidades reguladoras, explican o 30-35% dos casos de SBr. Polo tanto, nunha alta proporción dos pacientes diagnosticados con SBr, a etioloxía da enfermidade segue sendo descoñecida.

Os estudos realizados durante a última década demostraron que o locus *SCN5A-SCN10A* xoga un papel fundamental no control da actividade transcricional do xene *SCN5A*, ademais de estar asociado a alteracións nos parámetros electrocardiográficos, o que suxire un papel relevante na condución e a función cardíaca. Todos os compoñentes reguladores que interactúan no locus son necesarios para a expresión adecuada de *SCN5A* e a condución cardíaca. Esta evidencia suxire que unha alteración na actividade dos factores de transcrición (FT) que regulan a transcrición de *SCN5A*, ou a presenza de variantes nestas rexións, poderían inducir alteracións na condución cardíaca. Neste sentido, estudos previos do grupo: (1) revelaron a asociación dun haplotipo formado por sete SNVs no locus *SCN5A-SCN10A* con SBr mediante a secuenciación das rexións reguladoras dunha cohorte de 86 individuos con SBr<sup>351</sup>; e (2) identificaron o FT GATA4 como un regulador transcricional do xene *SCN5A*<sup>157</sup>. Ademais, estudos preliminares do laboratorio demostraron que a acetilación de GATA4 modula a súa función como activador transcricional do promotor de *SCN5A*.

A partir destas observacións, os nosos obxectivos centráronse en: (1) comprender en profundidade os mecanismos moleculares que regulan a transcrición de *SCN5A*; (2) identificar posibles variantes reguladoras que poidan alterar a transcrición de *SCN5A*; e (3) relacionar estes factores con posibles alteracións das correntes de sodio e das arritmias. Para cumprir estes propósitos, desenvolvéronse tres enfoques distintos.

En primeiro lugar, estudouse a acetilación do FT GATA4 cardíaco e o seu efecto sobre a regulación transcricional do xene *SCN5A*. Os ensaios de acetilación mediante mutaxénese dirixida en células HEK293T permitiron a identificación de tres sitios de acetilación (K318, K322 e K326) no dominio de unión ao DNA de GATA4. Ademais, experimentos reporteiros en células H9c2 cardíacas demostraron que a acetilación do residuo K318 de GATA4 é importante para a súa actividade transcricional no promotor de *SCN5A*.

En segundo lugar, propuxémonos identificar e caracterizar novas rexións reguladoras no locus *SCN5A-SCN10A* mediante un enfoque sistemático. Para iso, realizouse un cribado en diferentes rexións do locus *SCN5A-SCN10A* mediante ensaios CRISPRi nun modelo de cardiomiocitos derivados de hiPSCs (hiPSC-CMs). Os resultados do ensaio permitiron identificar sete secuencias específicas asociadas á regulación transcricional do xene *SCN5A* (*EXOG-CTCF*, Intron 1-NKX2-5, TSS-39bp, TSS+183bp, 5'*SCN5A-GATA4*, *SCN10A-GATA4* e *SCN10A-rs6801957*). O noso estudo describe, por primeira vez, que a rexión promotora do xene *EXOG* actúa como rexión reguladora da expresión de *SCN5A*. Ademais, o bloqueo da actividade transcricional da rexión *EXOG* altera os niveis de sodio intracelular no modelo de hiPSC-CMs.

En terceiro lugar, avaliouse o efecto das variantes xenéticas sobre as rexións reguladoras do locus *SCN5A-SCN10A*. En concreto, profundouse no estudo do haplotipo de sete SNVs asociado a SBr. En primeiro lugar, validouse experimentalmente o *phasing* dos haplotipos na cohorte de 86 individuos con SBr mediante a secuenciación de lecturas longas con Oxford Nanopore Sequencing. Os resultados da secuenciación confirmaron 83 dos 86 haplotipos obtidos con algoritmos de predición a partir de datos de lectura curta de Illumina. A continuación, avaliáronse as alteracións no perfil transcricional de dúas liñas de hiPSC-CMs mediante un experimento de RNA-seq. Comparáronse os hiPSC-CMs dun individuo con SBr e haplotipo de risco e os hiPSC-CMs dun individuo control con haplotipo sen risco. Este estudo revelou máis de 3 000 xenes expresados de forma diferencial. A aplicación de filtros de priorización determinou que os grupos funcionais máis enriquecidos na análise foron os xenes asociados á condución e diferenciación cardíaca, sendo *SCN5A* o xene máis significativo na análise. Ademais, a liña de hiPSC-CM co haplotipo asociado a SBr mostrou unha redución nos niveis de sodio intracelular do 36.5% respecto aos hiPSC-CMs control. Finalmente, usando CRISPR/Cas9, editouse unha das tres posicións necesarias para reverter o haplotipo asociado a SBr na liña hiPSC do paciente con SBr.

En resumo, o presente estudo demostra que a expresión do xene *SCN5A* está controlada por diferentes mecanismos de regulación transcricional, tanto a través da interacción de FT con elementos reguladores como pola presenza de variantes en rexións clave para a regulación de *SCN5A*. O efecto de *GATA4* na expresión de *SCN5A*, o efecto do haplotipo de risco nun modelo de cardiomiocitos con SBr e as diferentes rexións do locus *SCN5A-SCN10A* implicadas na regulación de *SCN5A* son os principais achados que apoian estas conclusións. Ademais, este estudo demostra a potencial aplicación da estratexia CRISPRi como ferramenta para detectar novas rexións que poidan ser relevantes no diagnóstico futuro da enfermidade.







## Resum

La síndrome de Brugada (SBr) és una malaltia elèctrica cardíaca associada a una elevada susceptibilitat a arrítmies ventriculars i mort sobtada cardíaca. Variants genètiques a les regions codificants del gen *SCN5A*, que codifica per al canal de sodi cardíac, expliquen el 30% dels casos de SBr. De manera similar, variants genètiques en regions codificants d'altres gens com els que codifiquen per a les subunitats  $\beta$  reguladores del canal de sodi cardíac, els canals de calci o altres proteïnes accessòries, expliquen el 5% dels casos de SBr. En total, les variants codificants afectant els gens de canals iònics, així com les subunitats reguladores, expliquen un 30-35% dels casos de SBr. Per tant, en una proporció elevada dels pacients diagnosticats amb SBr, l'etiologia de la malaltia continua sent desconeguda.

Els estudis realitzats durant la darrera dècada han demostrat que el locus *SCN5A-SCN10A* juga un paper fonamental en el control de l'activitat transcripcional del gen *SCN5A*, a més d'estar associat amb alteracions als paràmetres electrocardiogràfics, cosa que suggereix un paper rellevant en la conducció i la funció cardíaca. Tots els components reguladors que interactuen al locus són necessaris per a l'adequada expressió de *SCN5A* i la conducció cardíaca. Aquestes evidències apunten que una alteració de l'activitat dels factors de transcripció (FT) que regulen la transcripció de *SCN5A*, o bé la presència de variants en aquestes regions podria induir alteracions en la conducció cardíaca. En aquest sentit, estudis anteriors del grup: (1) van revelar l'associació d'un haplotip format per set SNVs al locus *SCN5A-SCN10A* amb SBr mitjançant la seqüenciació de les regions reguladores d'una cohort de 86 individus amb SBr<sup>351</sup>; i (2) van identificar el FT GATA4 com a regulador transcripcional del gen *SCN5A*<sup>157</sup>. A més, estudis preliminars del laboratori van demostrar que l'acetilació de GATA4 modula la seva funció com a activador transcripcional del promotor de *SCN5A*.

Partint d'aquestes observacions, els nostres objectius es van centrar a: (1) entendre amb profunditat els mecanismes moleculars que regulen la transcripció de *SCN5A*; (2) identificar possibles variants reguladores que puguin alterar la transcripció de *SCN5A*; i (3) relacionar aquests factors amb possibles alteracions als corrents de sodi i arrítmies. Per complir aquests propòsits, es van desenvolupar tres estratègies diferents.

En primer lloc, es va estudiar l'acetilació del FT cardíac GATA4 i el seu efecte a la regulació transcripcional del gen *SCN5A*. Assajos d'acetilació mitjançant mutagènesi dirigida a cèl·lules HEK293T van permetre identificar tres llocs d'acetilació (K318, K322 i K326) al domini d'unió al DNA de GATA4. A més, experiments reporters en cèl·lules cardíques H9c2 van demostrar que l'acetilació del residu K318 de GATA4 és important per a la seva activitat transcripcional sobre el promotor de *SCN5A*.

En segon lloc, ens vam proposar identificar i caracteritzar noves regions reguladores al locus *SCN5A-SCN10A* utilitzant una aproximació sistemàtica. Per fer-ho, es va realitzar un cribratge a diferents regions del locus *SCN5A-SCN10A* mitjançant assajos CRISPRi en un model de cardiomiòcits derivats de hiPSCs (hiPSC-CMs). Els resultats de l'assaig van permetre identificar set seqüències específiques associades a la regulació transcripcional del gen *SCN5A* (*EXOG-CTCF*, *Intró 1-NKX2-5*, *TSS-39pb*, *TSS+183pb*, *5'SCN5A-GATA4*, *SCN10A-GATA4* i *SCN10A-rs6801957*). El nostre estudi descriu, per primera vegada, que la regió del promotor del gen *EXOG* actua com a regió reguladora de l'expressió de *SCN5A*. A més, el bloqueig de l'activitat transcripcional de la regió *EXOG* altera els nivells de sodi intracel·lular al model de hiPSC-CMs.

En tercer lloc, es va avaluar l'efecte de les variants genètiques a les regions reguladores del locus *SCN5A-SCN10A*. Concretament, es va aprofundir en l'estudi de l'haplotip de set SNVs associat a SBr. En primer lloc, es va validar experimentalment el *phasing* dels haplotips a la cohort de 86 individus amb SBr mitjançant seqüenciació de lectures llargues amb Oxford Nanopore Sequencing. Els resultats de la seqüenciació van permetre confirmar 83 dels 86 haplotips obtinguts amb algorismes de predicció a partir de dades de lectures curtes d'Illumina. A continuació, es van avaluar les alteracions al perfil transcripcional de dues línies de hiPSC-CMs mitjançant un experiment d'RNA-seq. Es van comparar hiPSC-CMs d'un individu amb SBr i haplotip de risc i hiPSC-CMs d'un individu control amb haplotip de no-risc. Aquest estudi va revelar més de 3 000 gens diferencialment expressats. L'aplicació de filtres de prioritització va determinar que els grups funcionals més enriquits a l'anàlisi van ser gens associats a la conducció i la diferenciació cardíaca, sent *SCN5A* el gen més significatiu en l'anàlisi. A més, la línia de hiPSC-CMs amb l'haplotip associat a SBr va mostrar una reducció en els nivells de sodi intracel·lular d'un 36.5% respecte als controls hiPSC-CMs. Finalment, mitjançant CRISPR/Cas9, es va editar una de les tres posicions necessàries per revertir l'haplotip associat a SBr a la línia hiPSC del pacient amb SBr.

En resum, aquest estudi demostra que l'expressió del gen *SCN5A* està controlada per diferents mecanismes de regulació transcripcional, tant a través de la interacció de FT amb elements reguladors o per la presència de variants en regions clau per a la regulació de *SCN5A*. L'efecte de *GATA4* sobre l'expressió de *SCN5A*, l'efecte de l'haplotip de risc en un model de cardiomiòcits amb SBr i les diferents regions del locus *SCN5A-SCN10A* involucrades en la regulació de *SCN5A* són les principals troballes que donen suport a aquestes conclusions. A més, aquest estudi demostra la potencial aplicació de l'estratègia CRISPRi com a eina per detectar noves regions que podrien ser rellevants en el diagnòstic futur de la malaltia.





## Summary

Brugada syndrome (SBr) is a cardiac electrical disease associated with a high susceptibility to ventricular arrhythmias and sudden cardiac death. Genetic variants in the coding regions of the *SCN5A* gene, which encodes for the cardiac sodium channel, explain 30% of SBr cases. Similarly, genetic variants in the coding regions of other genes, such as those that encode for the  $\beta$  regulatory subunits of the cardiac sodium channel, calcium channels, or other accessory proteins, explain 5% of SBr cases. In total, coding variants affecting ion channel genes, as well as their regulatory subunits, explain 30-35% of SBr cases. Therefore, in a high proportion of patients diagnosed with SBr, the etiology of the disease remains unknown.

Studies carried out over the last decade have shown that the *SCN5A-SCN10A* locus plays a fundamental role in the control of the transcriptional activity of the *SCN5A* gene, in addition to being associated with alterations in electrocardiographic parameters, suggesting a relevant role in conduction and heart function. All interacting regulatory components at the locus are required for proper *SCN5A* expression and cardiac conduction. This evidence suggests that an alteration in the activity of the transcription factors (FT) that regulate the transcription of *SCN5A*, or the presence of variants in these regions, could induce alterations in cardiac conduction. In this regard, previous studies by the group: (1) revealed the association of a haplotype formed by seven SNVs at the *SCN5A-SCN10A* locus with SBr by sequencing the regulatory regions of a cohort of 86 individuals with SBr<sup>351</sup>; and (2) identified FT GATA4 as a transcriptional regulator of the *SCN5A* gene<sup>157</sup>. In addition, preliminary laboratory studies demonstrated that GATA4 acetylation modulates its function as a transcriptional activator of the *SCN5A* promoter.

Based on these observations, our objectives were focused on: (1) understanding in depth the molecular mechanisms that regulate the transcription of *SCN5A*; (2) identify potential regulatory variants that may alter *SCN5A* transcription; and (3) relate these factors with possible alterations in sodium currents and arrhythmias. To meet these purposes, three distinct approaches were developed.

First, the acetylation of cardiac FT GATA4 and its effect on the transcriptional regulation of the *SCN5A* gene were studied. Acetylation assays by site-directed mutagenesis in HEK293T cells allowed the identification of three acetylation sites (K318, K322 and K326) in the GATA4 DNA-binding domain. In addition, reporter experiments in cardiac H9c2 cells demonstrated that acetylation of the K318 residue of GATA4 is important for its transcriptional activity on the *SCN5A* promoter.

Second, we aimed to identify and characterize novel regulatory regions at the *SCN5A-SCN10A* locus using a systematic approach. For this, a screening was carried out in different regions of the

*SCN5A-SCN10A* locus using CRISPRi assays in a model of cardiomyocytes derived from hiPSCs (hiPSC-CMs). The results of the assay allowed the identification of seven specific sequences associated with the transcriptional regulation of the *SCN5A* gene (*EXOG*-CTCF, Intron 1-NKX2-5, TSS-39bp, TSS+183bp, 5'*SCN5A*-GATA4, *SCN10A*-GATA4 and *SCN10A*-rs6801957). Our study describes, for the first time, that the promoter region of the *EXOG* gene acts as a regulatory region for *SCN5A* expression. Furthermore, blocking the transcriptional activity of the *EXOG* region alters intracellular sodium levels in the hiPSC-CMs model.

Third, the effect of genetic variants on the regulatory regions of the *SCN5A-SCN10A* locus was evaluated. Specifically, the study of the haplotype of seven SNVs associated with SBr was deepened. First, the phasing of the haplotypes in the cohort of 86 individuals with SBr was experimentally validated by sequencing long reads with Oxford Nanopore Sequencing. The sequencing results confirmed 83 of the 86 haplotypes obtained with prediction algorithms from Illumina short read data. Next, alterations in the transcriptional profile of two lines of hiPSC-CMs were evaluated using an RNA-seq experiment. hiPSC-CMs from an individual with SBr and risk haplotype and hiPSC-CMs from a control individual with non-risk haplotype were compared. This study revealed more than 3000 differentially expressed genes. The application of prioritization filters determined that the most enriched functional groups in the analysis were genes associated with cardiac conduction and differentiation, with *SCN5A* being the most significant gene in the analysis. In addition, the hiPSC-CMs line with the BrS-associated haplotype showed a 36.5% reduction in intracellular sodium levels compared to control hiPSC-CMs. Finally, using CRISPR/Cas9, we edited one of the three positions required to reverse the BrS-associated haplotype in the hiPSC line of the BrS patient.

In summary, the present study demonstrates that the expression of the *SCN5A* gene is controlled by different mechanisms of transcriptional regulation, both through the interaction of TF with regulatory elements or by the presence of variants in key regions for the regulation of *SCN5A*. The effect of GATA4 on *SCN5A* expression, the effect of risk haplotype in a model of cardiomyocytes with SBr and the different regions of the *SCN5A-SCN10A* locus involved in the regulation of *SCN5A* are the main findings supporting these conclusions. Furthermore, this study demonstrates the potential application of the CRISPRi strategy as a tool to detect new regions that could be relevant in the future diagnosis of the disease.









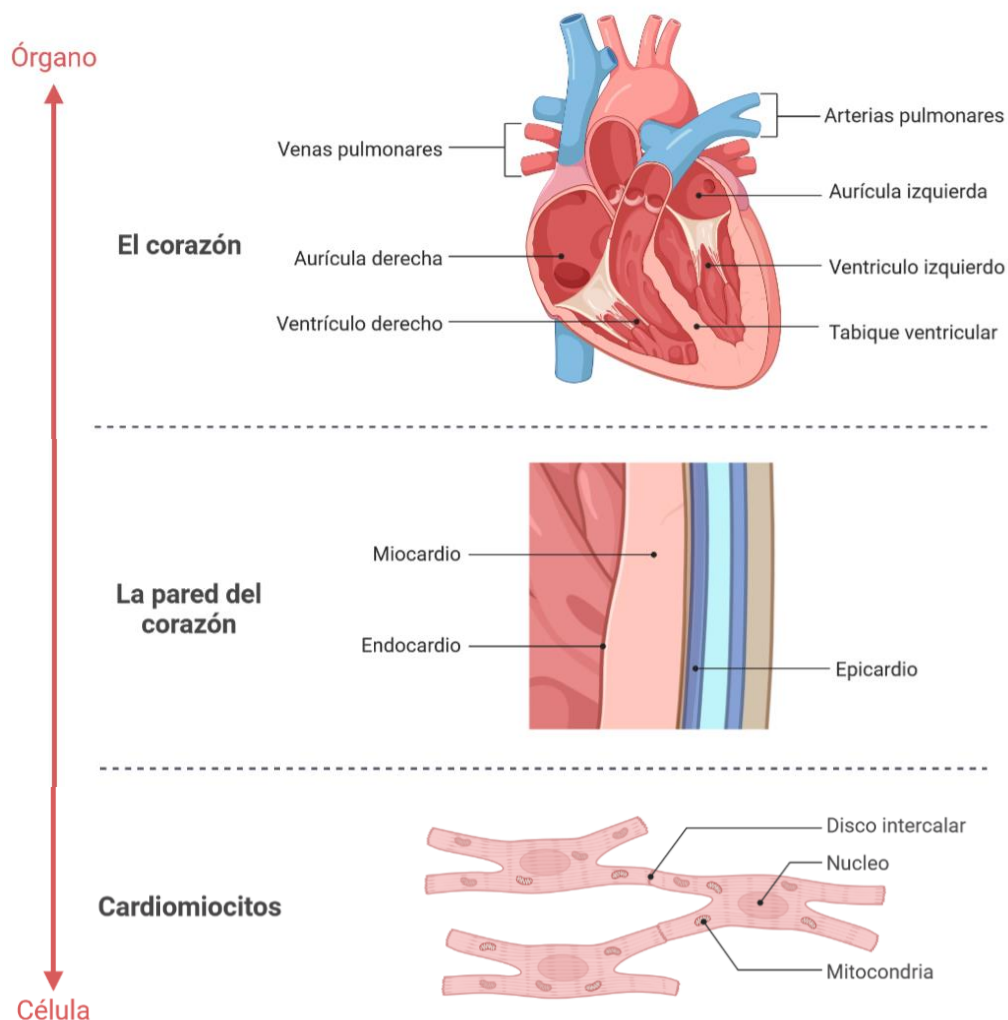
# **I. Introducción**



# 1. El corazón

## 1.1. Anatomía y función del corazón

El corazón es un órgano muscular especializado en bombear la sangre por todo el cuerpo, con el fin de proporcionar oxígeno y nutrientes al resto de órganos. El corazón humano está dividido en cuatro cavidades, dos aurículas en la parte superior y dos ventrículos en la parte inferior (**Figura 1**). Una pared muscular llamada tabique ventricular divide el corazón en dos mitades (izquierda y derecha), evitando así el paso de la sangre entre las dos aurículas y entre los dos ventrículos. La mitad izquierda del corazón recibe la sangre oxigenada, que proviene de los pulmones, y la expulsa mediante el sistema arterial en todo el cuerpo. Sin embargo, a la mitad derecha del corazón llega la sangre desoxigenada, a través del sistema venoso, que proviene de los diferentes órganos del cuerpo, y la envía a los pulmones<sup>1</sup>.

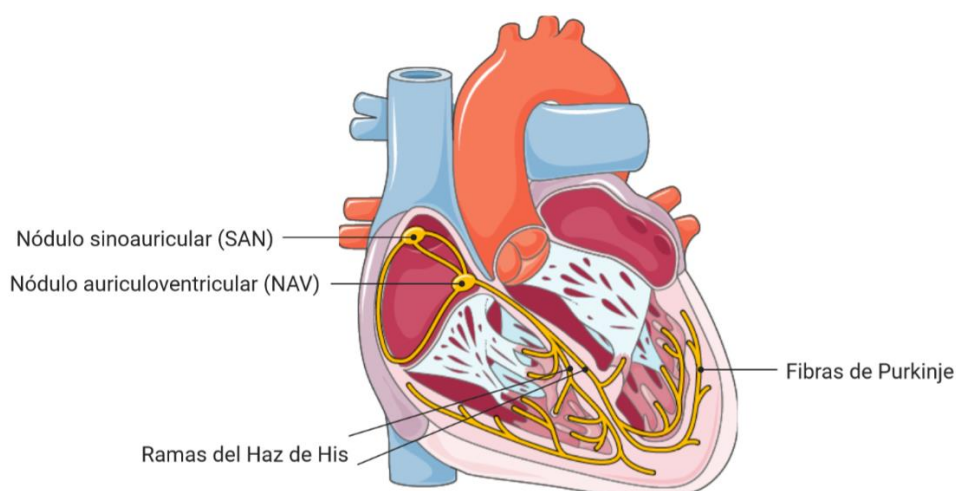


**Figura 1. Anatomía del corazón.** Arriba: Representación de las cuatro cámaras del corazón y el tabique. Centro: Representación de las tres capas que recubren el corazón. Abajo: Cardiomiocitos del tejido muscular cardíaco. Figura creada con BioRender (<https://app.biorender.com/>).

La pared del corazón se compone de tres capas principales denominadas, de la más externa a la más interna: epicardio, miocardio y endocardio (**Figura 1**). El epicardio es una membrana serosa que protege el área externa del corazón. El miocardio es el propio tejido muscular cardíaco y contiene las células cardíacas o cardiomiocitos. Finalmente, el endocardio protege la parte interna de las cavidades y válvulas del corazón<sup>2</sup>.

## 1.2. Actividad eléctrica cardíaca

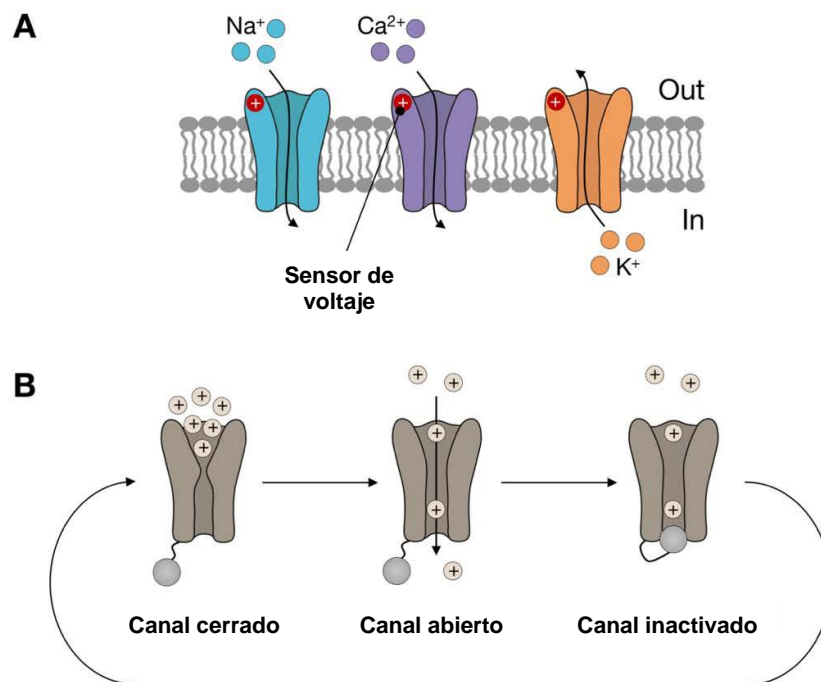
Cada latido del corazón se caracteriza por una relajación (diástole) y una contracción (sístole) perfectamente coordinadas de las aurículas y los ventrículos. Durante la sístole, los ventrículos se contraen para bombear sangre a la arteria pulmonar y la aorta, mientras que las aurículas se relajan y recogen sangre. Durante la diástole, los ventrículos se relajan y las aurículas se contraen para bombear sangre a los ventrículos<sup>3</sup>. La coordinación entre sístole y diástole está mediada por el sistema de conducción eléctrica del corazón, que transmite el impulso eléctrico de las aurículas a los ventrículos (**Figura 2**). La actividad eléctrica cardíaca se origina en las células autónomas del nódulo sinoauricular (SAN), ubicado en la pared de la aurícula derecha. El SAN también se conoce como el marcapasos natural del corazón, ya que inicia de forma continua y autónoma el impulso eléctrico, que se propaga a los cardiomiocitos auriculares, lo que provoca la excitación de las aurículas. La onda de excitación originada en el SAN se propaga al nódulo auriculoventricular (NAV), que conecta eléctricamente las aurículas y los ventrículos. El NAV retrasa el pulso cardíaco para asegurar que las aurículas se hayan contraído y la sangre haya sido expulsada hacia los ventrículos antes de la contracción ventricular<sup>4</sup>. Finalmente, la onda de excitación se propaga desde NAV hacia los ventrículos derecho e izquierdo por las ramas izquierda y derecha del Haz y las fibras de Purkinje<sup>5</sup>.



**Figura 2. Sistema de conducción eléctrica del corazón.** Representación esquemática de los componentes que constituyen el sistema de conducción eléctrica. Figura creada con BioRender (<https://app.biorender.com/>).

### 1.3. Canales iónicos dependientes de voltaje

Los canales iónicos controlados por voltaje son proteínas de membrana integrales que permiten el paso de iones a través de las membranas celulares. Se abren y cierran en respuesta a cambios en el voltaje transmembrana y desempeñan un papel clave en la señalización eléctrica de células excitables como neuronas o cardiomiocitos<sup>6</sup>. Este tipo de canales iónicos contienen un sensor de voltaje, una región que contiene aminoácidos cargados, los cuales se reubican con los cambios en el campo eléctrico de la membrana. El movimiento del sensor inicia un cambio de conformación en el canal, controlando así el flujo de iones. Los canales iónicos activados por voltaje pueden existir en tres estados funcionalmente distintos: cerrado (reposo), abierto (activo) e inactivo (**Figura 3**). Tanto los estados cerrados como los inactivados no son conductores, pero los canales que se han inactivado son refractarios a menos que la célula vuelva al potencial de membrana inicial para permitirles volver al estado cerrado<sup>7</sup>.



**Figura 3. Canales iónicos dependientes de voltaje.** (A) Esquemas de los tres principales canales iónicos cardiacos dependientes de voltaje ( $\text{Na}^+$ ; sodio,  $\text{Ca}^{2+}$ ; calcio y  $\text{K}^+$ ; potasio).  $\text{Na}^+$  y  $\text{Ca}^{2+}$  participan en las corrientes de entrada mientras que  $\text{K}^+$  participa en las corrientes de salida. (B) Estados conformacionales de los canales iónicos dependientes de voltaje. Figura adaptada de M. Pinsach<sup>8</sup>.

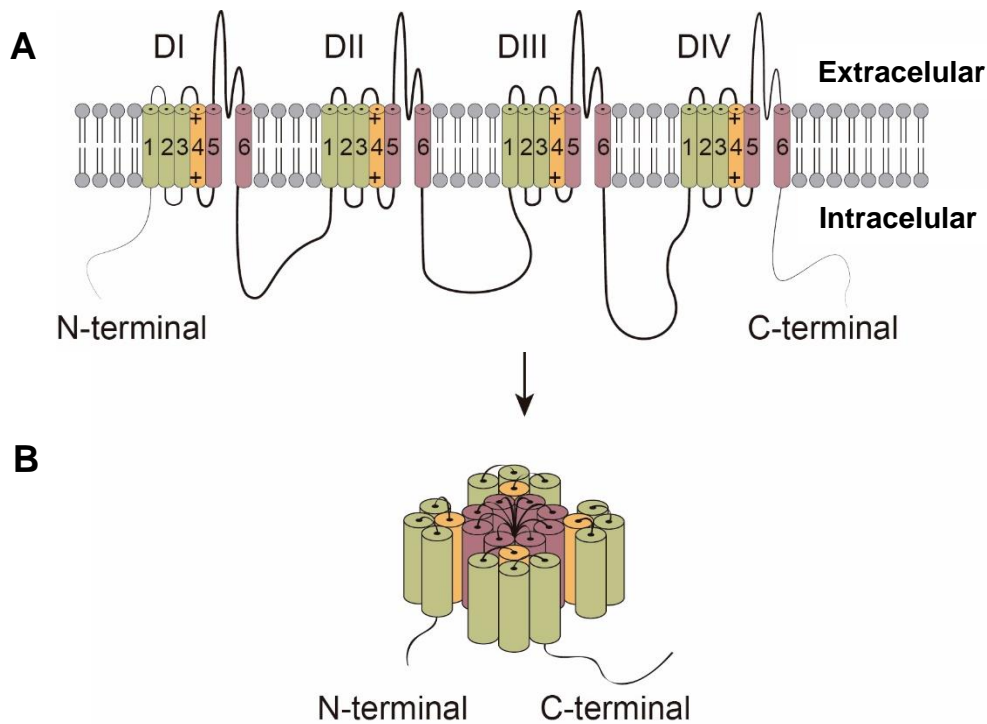
### 1.3.1. Canales de sodio

Los canales de sodio dependientes de voltaje (Nav) son canales iónicos con permeabilidad selectiva a los iones  $\text{Na}^+$ , lo que les permite fluir siguiendo su gradiente electroquímico. Los canales de Nav tienen un papel determinante en la actividad eléctrica de muchas células, ya que inician potenciales de acción en células excitables<sup>9</sup>.

En los mamíferos, *SCN1A-SCN11A* son los genes que codifican una familia de nueve Nav (Nav1.1-Nav1.9), que tiene más del 50 % de homología en la secuencia de aminoácidos<sup>10</sup>. Todas las isoformas de los Nav muestran la misma estructura general, que consta de una subunidad  $\alpha$  formadora de poros y una o más de las cuatro subunidades  $\beta$  auxiliares ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$  y  $\beta 4$ ), codificadas por los genes *SCN1B-SCN4B*, respectivamente<sup>11,12</sup>. Las principales propiedades funcionales de los Nav residen en la subunidad  $\alpha$ , aunque su actividad puede ser modulada por subunidades  $\beta$ .

Históricamente, las subunidades  $\beta$  se describieron como meras subunidades auxiliares de la subunidad  $\alpha$ . Sin embargo, a medida que prosperó la investigación sobre estas, surgieron distintas funciones fuera del contexto de la subunidad  $\alpha$  formadora de poros. Hoy en día es de conocimiento común que las subunidades  $\beta$  son proteínas multifuncionales con funciones conductoras y no conductoras; además, se expresan en células excitables a lo largo del sistema nervioso y del corazón y en células no excitables<sup>13</sup>. Como proteínas que interactúan con la subunidad  $\alpha$ , las subunidades  $\beta$  modulan las propiedades biofísicas del canal y su localización subcelular. Las subunidades  $\beta$  también interactúan con otros componentes del canal iónico, sirviendo como enlazadores y anclas. Dado que todas las subunidades  $\beta$  tienen un dominio extracelular similar a la inmunoglobulina, se consideran moléculas de adhesión celular. Fuera del contexto de la subunidad  $\alpha$ , las subunidades  $\beta$  desempeñan funciones en las vías de señalización extracelular e intracelular, la migración, la proliferación, el crecimiento de axones neuronales, la búsqueda de vías y la fasciculación en el sistema nervioso central. Además, se ha demostrado que sus productos de escisión regulan la expresión génica<sup>12</sup>.

En el corazón, la subunidad  $\alpha$  mayoritaria es Nav1.5, codificada por el gen *SCN5A*<sup>14</sup>. Nav1.5 es un tetrámero de 227 kilo Daltons (kDa) de cuatro dominios homólogos (DI-DIV) unidos por bucles intracelulares (**Figura 4**). Cada dominio consta de seis segmentos transmembrana (S1-S6). S4 está cargado positivamente y participa en la activación del canal dependiente del voltaje, mientras que los bucles entre S5 y S6 de cada dominio se curvan hacia la membrana para formar el poro del canal conductor de sodio<sup>15</sup>. Los cuatro dominios homólogos DI-DIV pliegan y posicionan los segmentos S5 y S6 de cada dominio en la parte interna del tetrámero, que forma el poro final con selectividad por los iones  $\text{Na}^+$ <sup>16,17</sup>.



**Figura 4. Canal de sodio dependiente de voltaje cardíaco.** (A) Subunidad  $\alpha$  de Nav1.5 con los 4 dominios (DI-DIV), cada uno compuesto por seis segmentos transmembrana. (B) Plegamiento de los cuatro dominios Nav1.5 alrededor de un poro conductor de iones. Figura adaptada de A. Tarradas<sup>18</sup>.

### 1.3.2. Canales de calcio y potasio

La familia de canales de calcio dependientes de voltaje sirve como transductor clave de los cambios de potencial de membrana de la superficie celular en el proceso de aumento y disminución del calcio intracelular local que inician muchos eventos fisiológicos diferentes<sup>19</sup>. Todos los canales de calcio dependientes de voltaje están compuestos por una combinación de hasta cuatro subunidades distintas ( $\alpha 1$ ,  $\alpha 2\delta$ ,  $\beta$  y  $\gamma$ ). La subunidad  $\alpha 1$  de 190-250 kDa es la subunidad más grande e incluye el poro de conducción, el sensor de voltaje y el aparato de activación. Hasta la fecha, en mamíferos se han caracterizado diez miembros de la subunidad  $\alpha 1$  (Cav1.1-Cav1.4, Cav2.1-Cav2.3 y Cav3.1-Cav3.3), codificados por los genes *CACNA1A-CACNA1I* y *CACNA1S*<sup>19</sup>.

Los canales de potasio dependientes de voltaje muestran amplias distribuciones en el sistema nervioso y otros tejidos como el músculo cardíaco<sup>20</sup>. Desempeñan un papel importante en la regulación de la excitabilidad del músculo cardíaco al devolver la célula despolarizada a un estado de reposo y al controlar la duración y la frecuencia del potencial de acción<sup>21</sup>. Las variaciones en el nivel de expresión de estos canales son responsables de las diferencias en la configuración del potencial de acción entre las aurículas, los ventrículos y a través de la pared del miocardio.



Cada canal de potasio activado por voltaje es un tetrámero de cuatro subunidades  $\alpha$  formadoras de poros y también puede contener subunidades  $\beta$  auxiliares que pueden modular la función y/o localización del canal<sup>22,23</sup>. Cada subunidad  $\alpha$  formadora de poros contiene seis segmentos transmembrana (S1-S6), con los primeros cuatro segmentos transmembrana (S1-S4) formando el sensor de voltaje y los dos últimos segmentos transmembrana (S5 y S6) formando la estructura de los poros<sup>23</sup>. En humanos, las subunidades se agrupan en 12 subfamilias (Kv1- Kv12) codificadas por 40 genes diferentes, mientras que las subunidades  $\beta$  están compuestas por 12 miembros codificados por 12 genes diferentes<sup>23</sup>.

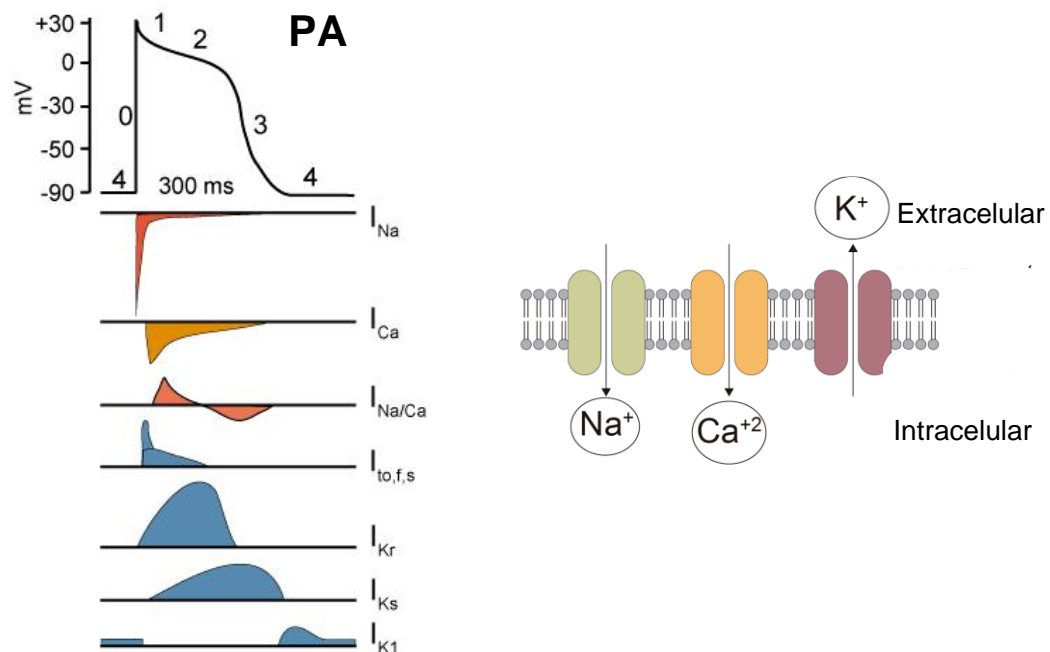
#### 1.4. Potencial de acción cardíaco

A medida que se va propagando el impulso eléctrico, se va produciendo la contracción de las distintas partes del corazón gracias a la generación de un potencial de acción cardíaco (PA). El PA es un balance de corrientes de iones que circulan a favor de su gradiente electroquímico, a través de los canales iónicos presentes en la membrana de los cardiomiocitos. La difusión de iones provoca cambios reversibles en el potencial de membrana, que se define como la diferencia de potencial que existe entre ambos lados de la membrana plasmática. En las células excitables (neuronas, cardiomiocitos...) la membrana plasmática presenta una distribución asimétrica de cargas entre el interior y el exterior de la célula que produce la diferencia de potencial<sup>24</sup>. Los cardiomiocitos en estado de reposo presentan una diferencia de potencial de -90 milivoltios (mV), lo que significa que el interior de las células es 90 mV más negativo que el exterior. La apertura y el cierre de los canales iónicos, discutidos en la sección anterior, permiten las corrientes iónicas transmembrana y, como resultado, la formación de PA<sup>25</sup>. Estos PA son responsables de la conducción de la actividad eléctrica en todo el corazón que termina con la contracción del corazón. El PA consta de 5 fases resumidas a continuación<sup>1,25-27</sup> (**Figura 5**):

- **Fase 0 o de despolarización:** la llegada del impulso eléctrico induce la apertura de los canales de  $\text{Na}^+$  cardíacos dependientes de voltaje (canales  $\text{Nav}1.5$ ). A través de estos canales se produce una entrada rápida y masiva de iones  $\text{Na}^+$  en la célula que genera la corriente de  $\text{Na}^+$  ( $I_{\text{Na}}$ ). La entrada de  $\text{Na}^+$  induce un cambio de potencial de -90 mV (en estado de reposo) a valores positivos (aproximadamente +40 mV).
- **Fase 1 o de repolarización parcial:** esta fase se debe a la apertura de canales de  $\text{K}^+$ , que genera una corriente de salida de  $\text{K}^+$  ( $I_{\text{to}}$ ), y a la inactivación de los canales  $\text{Nav}1.5$ .
- **Fase 2 o de *plateau*:** se produce una corriente entrante de  $\text{Ca}^{+2}$  a través de los canales de calcio tipo L ( $I_{\text{Ca}}$ ). La entrada de  $\text{Ca}^{+2}$  activa los receptores de rianodina, que están presentes en la membrana del retículo sarcoplasmático. El receptor de rianodina libera el calcio del interior del RS en el citoplasma. El aumento de calcio citosólico activa el intercambiador  $\text{Na}^+/\text{Ca}^{+2}$ , que provoca la liberación del  $\text{Ca}^{+2}$  fuera de la célula, y hace entrar tres iones  $\text{Na}^+$  por

cada ion de  $\text{Ca}^{+2}$  que sale. En esta fase se produce también la activación de distintos canales de  $\text{K}^+$ , que generan una corriente saliente de  $\text{K}^+$  ( $I_K$ ).

- **Fase 3 o de repolarización:** se caracteriza por una salida prolongada de iones  $\text{K}^+$ . A lo largo de esta fase, el potencial de membrana alcanza cada vez valores más negativos.
- **Fase 4 o de reposo:** corresponde al retorno al potencial de reposo de membrana de los cardiomiocitos (-90 mV). La membrana se mantiene estable a este potencial gracias a la corriente rectificadora de  $\text{K}^+$  ( $I_{K1}$ ). Los cardiomiocitos permanecen en esta fase hasta la llegada de un nuevo impulso eléctrico.



**Figura 5. PA y corrientes iónicas del corazón.** A la izquierda, se muestra un PA ventricular, y debajo, las corrientes de sodio ( $I_{Na}$ ), de calcio ( $I_{Ca}$ ) y de potasio ( $I_{to}$ ,  $I_{Ks}$ ,  $I_{K1}$ ) responsables de generar el PA. En la parte derecha se representan los tres canales iónicos principales presentes en la membrana de los cardiomiocitos que participan en el PA. Figura adaptada de A. Tarradas<sup>18</sup>.

Dado que la actividad eléctrica de las células en las diferentes regiones del corazón está perfectamente sincronizada, alteraciones en la función de estos canales iónicos debido a variantes genéticas podría alterar el PA y provocar una arritmia. Estas alteraciones pueden dar lugar a patologías asociadas a la muerte súbita cardiaca (MSC). El mecanismo de prolongación del PA implica la expresión alterada de diferentes corrientes despolarizantes e hiperpolarizantes en el miocardio<sup>28</sup>. Por lo tanto, los canales iónicos tienen que estar finamente orquestados para una correcta función cardíaca<sup>29</sup>.

## 2. La muerte súbita cardíaca

La Organización Mundial de la Salud define la MSC como una muerte natural súbita e inesperada por una causa cardiovascular, en un paciente con o sin una enfermedad cardiovascular previamente diagnosticada, que ocurre dentro de la primera hora del inicio de los síntomas cuando se presencia. Si no hay testigos, se considera una MSC cuando se observó que el individuo estaba vivo dentro de las 24 horas anteriores. Un paro cardíaco repentino ocurre cuando el corazón deja de latir, o no late lo suficiente como para mantener la perfusión y la vida, y puede provocar MSC<sup>30</sup>. La MSC es una de las principales causas de mortalidad, responsable del 50-60% de todas las muertes en todo el mundo por enfermedades cardiovasculares<sup>30-32</sup>. Además, representa hasta el 15-20% de todas las muertes naturales en las sociedades occidentales<sup>33</sup>, siendo su incidencia menor en individuos asiáticos<sup>34</sup>. La MSC se debe a dos factores: un sustrato subyacente y un evento transitorio. El sustrato es una afección cardíaca y el evento transitorio puede ser cualquier factor externo que induzca inestabilidad eléctrica en el corazón (por ejemplo, aumento de la actividad simpática)<sup>35-37</sup>. En las últimas décadas se ha avanzado sustancialmente en la prevención de la MSC y se han identificado múltiples factores de riesgo clínicos y genéticos que comprometen el funcionamiento cardíaco<sup>30</sup>.

### 2.1. Factores de riesgo y etiología

Los antecedentes familiares de MSC pueden representar un factor de riesgo<sup>38</sup>. Se ha descubierto que el riesgo de MSC es dos veces mayor si una persona tuvo uno de los progenitores que murió súbitamente y nueve veces si ambos progenitores lo hicieron. Esto apoya la idea de que los factores hereditarios juegan un papel importante en la determinación del riesgo de MSC<sup>30</sup>. Entre los factores de riesgo clínicos, la hipertensión, la diabetes, el tabaquismo y la obesidad están asociados con MSC tanto en hombres como en mujeres<sup>30,34</sup>. Otras comorbilidades clínicas pueden ser la fibrilación auricular (FA), enfermedad renal crónica y apnea obstructiva del sueño<sup>31</sup>. La depresión, la ansiedad y la psicosis son enfermedades mentales que se han relacionado con la MSC, así como factores del estilo de vida como los patrones dietéticos, el estrés emocional y la contaminación del aire<sup>34</sup>. El sexo también juega un papel fundamental, siendo el riesgo en hombres de tres a cuatro veces mayor en comparación con las mujeres<sup>30,39</sup>.

Aunque la etiología de la MSC es muy diversa, la causa más común de MSC es la cardiopatía coronaria, que representa hasta el 75-80% de las MSC<sup>40</sup>. Es un tipo de cardiopatía en la que las arterias del corazón no pueden aportar suficiente cantidad de sangre oxigenada al corazón<sup>41</sup>. Alrededor de 182 millones de estadounidenses adultos tienen arteriopatía coronaria, lo que la convierte en el tipo más frecuente de enfermedad cardíaca en los Estados Unidos<sup>41</sup>. La causa más frecuente de la arteriopatía coronaria es el colesterol, que se acumula en el interior de la

capa que recubre a las arterias coronarias, formando una placa. Esta acumulación puede obstruir el flujo sanguíneo de las grandes arterias del corazón en forma parcial o total<sup>41</sup>.

El 20-25% de las MSC restantes son hereditarias, y por lo tanto tienen una causa genética. De estas, el 10-15% son miocardiopatías, principalmente relacionadas con anomalías estructurales cardíacas<sup>40</sup>. Este tipo de enfermedades están asociadas a alteraciones genéticas en proteínas estructurales, incluidas las de los sarcómeros, los desmosomas y el citoesqueleto<sup>41-43</sup>. La mayoría de variantes asociadas a las miocardiopatías se encuentra en los genes *MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *LMNA*, *TTN*, *PKP2* y *ACTC*<sup>156</sup>.

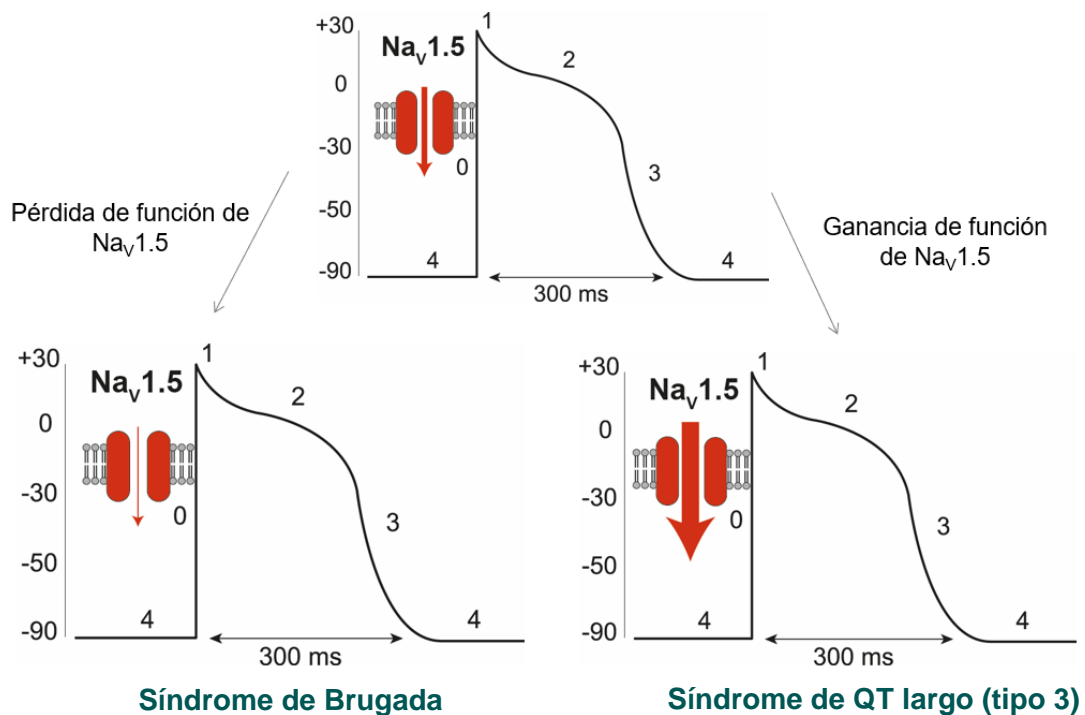
Las miocardiopatías más comunes relacionadas con la MSC incluyen la miocardiopatía hipertrófica (MCH), la miocardiopatía dilatada (MCD), la miocardiopatía arritmogénica del ventrículo derecho (MAVD) y el ventrículo izquierdo no compacto (VINC). El 5-10% restante de las MSC hereditarias son canalopatías, relacionadas con una alteración de la actividad eléctrica en corazones estructuralmente normales.

## 2.2. Enfermedades eléctricas o canalopatías cardíacas

Las canalopatías cardíacas se asocian a alteraciones genéticas en los canales iónicos de membrana o en sus proteínas reguladoras que comprometen su función<sup>44,45</sup>. Las principales canalopatías cardíacas son el síndrome de Brugada (SBr), el síndrome de QT largo (SQTL), el síndrome de QT corto (SQTC) y la taquicardia ventricular polimórfica catecolaminérgica (TVPC)<sup>42,43,45</sup>. De manera general, las canalopatías presentan penetrancia incompleta (probabilidad de que una afección clínica se manifieste cuando está presente un genotipo específico) y expresividad variable (rango de signos y síntomas que pueden ocurrir en diferentes personas con la misma afección genética)<sup>44</sup>.

Las variantes genéticas asociadas a las canalopatías pueden dar lugar a cambios en los niveles de expresión de los canales, o bien a las propiedades biofísicas de los canales iónicos, lo que conlleva a una alteración del equilibrio de las corrientes iónicas, que es esencial para un PA normal<sup>25,45</sup> (**Figura 6**). Estas alteraciones pueden conducir a una heterogeneidad en la amplitud y sincronización temporal del PA y/o al acoplamiento entre los cardiomiocitos de las diferentes regiones del corazón, causando una arritmia que favorece la MSC<sup>24,45,46</sup>.

Una de las causas principales de las canalopatías son las variantes genéticas en los canales iónicos o sus proteínas reguladoras. La información recopilada a través de análisis genéticos, así como el electrocardiograma (ECG) y la información clínica del paciente deben integrarse para el diagnóstico de la enfermedad. Los análisis genéticos son especialmente útiles para apoyar el diagnóstico clínico, definir un tratamiento e identificar portadores de variantes con riesgo de sufrir MSC<sup>47-49</sup>. Se han identificado cientos de variantes asociadas a enfermedades cardíacas que afectan a las corrientes de iones Na<sup>+</sup>, K<sup>+</sup> y Ca<sup>2+</sup>.



**Figura 6. Alteración de la función de  $\text{Na}_v1.5$  en SBr y SQTL (tipo 3).** Consecuencias de la alteración de las corrientes de sodio, donde en SBr hay una pérdida de función de  $\text{Na}_v1.5$  y en SQTL (tipo 3) una ganancia.

Sin embargo, en los últimos años se ha observado que las variantes genéticas que afectan la actividad de los elementos reguladores del DNA que controlan la expresión de los canales iónicos cardíacos también pueden resultar en defectos de conducción cardíaca, aumentando la susceptibilidad a las arritmias cardíacas y a sufrir MSC<sup>50,51</sup>.

En general, las canalopatías constituyen un grupo heterogéneo de desórdenes arritmogénicos. A continuación, se describen las características de las principales canalopatías mencionadas anteriormente:

- **SQTL:** Se caracteriza por la prolongación del intervalo QT en el ECG, debido a un alargamiento del tiempo de repolarización ventricular. Es la canalopatía con mayor prevalencia entre los adultos menores de 35 años, y afecta a 1 de cada 2 000 individuos<sup>52</sup>. Clínicamente se manifiesta con síncope y paros cardíacos (principalmente desencadenados por el estrés emocional o físico), convulsiones, arritmias ventriculares (AV), FA y *torsade de pointes*<sup>53</sup>. Las causas del SQTL pueden ser genéticas (SQTL familiar) o adquiridas (por ejemplo, a partir de un tratamiento con fármacos bloqueadores de los canales de potasio<sup>54</sup>). La forma congénita se asocia a alteraciones genéticas en canales de potasio, por una disminución de las corrientes repolarizadoras, o en canales de sodio, por un retraso en la entrada del flujo de iones en el miocito. El patrón de herencia es autosómico dominante y

aunque se han asociado hasta 15 genes distintos a esta enfermedad, en la inmensa mayoría de casos la causa genética se identifica en variantes raras en *KCNQ1* (40-55%), *KCNH2* (30-45%) y *SCN5A* (5-10%), este último con una ganancia de función del canal (**Figura 6**). El cribado identifica la causa genética de la enfermedad en un 75-85% de los casos<sup>55</sup>.

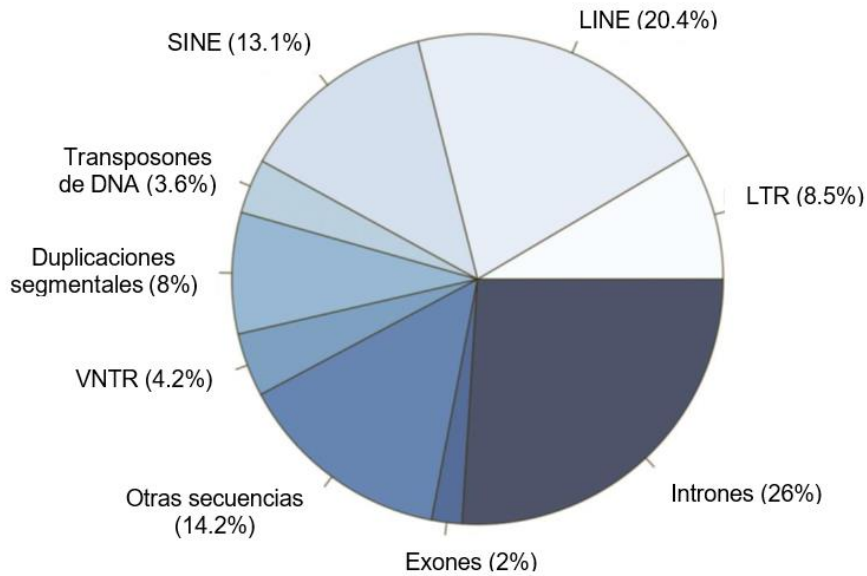
- **SQTC:** El SQTC se caracteriza por presentar un intervalo de QT inferior a los 330 ms, con una onda T alta y terminada en pico. Los fenotipos son muy agresivos y la edad de aparición de las primeras manifestaciones clínicas oscila desde la infancia hasta los 40 años, con síntomas como FA, AV, síncope y MSC<sup>57</sup>. La prevalencia del SQTC es de un 0.02-0.1% en la población adulta y de un 0.05% en población pediátrica, con una mayor prevalencia entre los individuos de sexo masculino<sup>58</sup>. El síndrome presenta un patrón de herencia autosómico dominante y hasta la fecha se han asociado 3 genes que codifican para canales de potasio, también asociados con el SQT: *KCNH2*, *KCNQ1* y *KCNJ2*<sup>59</sup>, y tres genes que codifican para canales de calcio: *CACNA1C*, *CACNB2*, y *CACNA2D1*. Las variantes en estos tres genes conllevan una ganancia de función del canal de potasio<sup>60</sup>. Y permiten explicar la etiología del 15-40% de los casos reportados, aun cuando por la mayoría de los pacientes existe una asociación familiar<sup>61</sup>.
  
- **TVPC:** La TVPC se caracteriza por la aparición de taquicardias ventriculares polimórficas y/o bidireccionales inducidas por catecolaminas durante el estrés emocional o físico, y por poder causar episodios de síncope y MSC en individuos jóvenes<sup>62,63</sup>. La prevalencia estimada de la enfermedad es de 1 entre 10 000 individuos<sup>64</sup>. En un 30% de los casos, la MSC puede ser la primera manifestación clínica de la enfermedad<sup>47</sup>. La canalopatía presenta dos patrones de herencia: la autosómica dominante, causada por alteraciones genéticas en el gen del receptor de rianodina (*RYR2*) en un 50-55% de los casos y el gen codificante de la calmodulina (*CALM1*), en menos de un 1% de los casos; y autosómica recesiva, causada por alteraciones genéticas en *CASQ2* y *TRDN*, en un 5% y 2% de los casos, respectivamente<sup>65</sup>.
  
- **SBr:** Se estima que el SBr es el responsable de entre el 4 y el 12% del total de MSC. La causa genética más conocida de la SBr son las variantes en el gen que codifica para el canal Nav1.5<sup>66</sup>. Estas variantes se relacionan con una pérdida de función del canal (**Figura 6**) que provoca su cierre prematuro, o bien que éste no se active, provocando un acortamiento de la fase 1 del PA y dando lugar a arritmias por reentrada<sup>67</sup>. El SBr se describe más ampliamente en el apartado 7 de la Introducción<sup>68</sup>.

### 3. El genoma humano

#### 3.1. Organización del genoma humano

El genoma humano se define como el contenido genético total que se encuentra dentro de una célula humana. La unidad básica del genoma es el ácido desoxirribonucleico o DNA, que fue descubierto en 1868 por Friedrich Miescher<sup>69</sup> y, posteriormente, en 1953, su estructura fue descrita por Francis Crick y James Watson<sup>70</sup>. El genoma total consta de 3234.83 millones de pares de bases (pb) condensadas y empaquetadas en 23 pares de cromosomas (22 pares somáticos y uno sexual). En el esbozo original de la secuencia se predijo un número aproximado de 30 000 genes codificantes para proteínas<sup>71</sup>, número que ha ido disminuyendo hasta el día de hoy, en el que se consideran 19 969 genes codificantes en la versión más actualizada del genoma (T2T-CHM13), el equivalente a un 2% del genoma, además de 43 525 genes no-codificantes<sup>72</sup>. Este porcentaje de ocupación del genoma para genes codificantes es mucho menor al esperado para un organismo tan complejo como el humano<sup>73</sup>. De hecho, en comparación con casi cualquier otro organismo, el número de genes codificantes de los humanos no parecen ser muchos. La mosca de la fruta *Drosophila melanogaster*, por ejemplo, tiene unos 13 000 genes codificantes<sup>74</sup>. Otro ejemplo sería *Arabidopsis thaliana*, organismo modelo para estudiar la genética de las plantas. *A. thaliana* tiene incluso más genes codificantes que los humanos, aproximadamente 25 500, además de tener uno de los genomas más pequeños del mundo de las plantas<sup>74</sup>. Parecería obvio que los humanos tuviesen más genes codificantes que las plantas, pero ese no es el caso. De hecho, no existe una correlación clara entre el tamaño del genoma y la cantidad de genes que codifican proteínas, otra indicación de que la cantidad de genes en un genoma eucariota revela poco sobre la complejidad del organismo. El motivo y la solución a esta paradoja se atribuye a los complejos mecanismos de regulación del 98% restante del genoma humano<sup>72</sup>.

A partir del conocimiento actual, las regiones del genoma se pueden clasificar según sus propiedades (**Figura 7**). Los **exones** o fragmentos de secuencias codificantes de los genes, aparecen muy separados los unos de los otros para contrarrestar el efecto de posibles variantes que podrían causar un cambio en la pauta de lectura, o para evitar que los procesos de cruce genético afecten a información sensible del genoma<sup>75</sup>. El resto de elementos del genoma corresponden al DNA no codificante, constituido por un amplio abanico de elementos (**Figura 7**), y aunque algunos carecen de una función determinada o conocida, otros desarrollan funciones de gran importancia, como la de proteger información indispensable del genoma:



**Figura 7. Composición del genoma humano.** Más de la mitad de las secuencias del genoma son secuencias repetitivas (57.8% sumando todos los tipos de elementos repetitivo). Los casi 20 000 genes codificantes solamente representan alrededor del 2% del total del genoma, mientras que solamente los intrones ocupan el 26% del genoma. Figura adaptada de Patrushev *et al.*<sup>90</sup>.

A) **RNA no codificantes:** Los RNAs no codificantes (ncRNAs) comprenden aquellos genes que no codifican proteínas. Este grupo de genes se transcriben a moléculas de RNA que son funcionales y que no se traducen a proteínas. Hasta hace unos años, la mayoría de los ncRNAs conocidos cumplían funciones relativamente genéricas en las células, como el RNA ribosomal (rRNA) –que forma los ribosomas–, los RNA de transferencia (tRNA) –implicados en la traducción del RNA mensajero (mRNA)– o los RNA de pequeño tamaño (snRNA) implicados en el splicing<sup>76</sup>. Sin embargo, a medida que se ha estudiado el alcance de las funciones de los ncRNAs, estos parecen incluir una capa oculta de señales internas que controlan varios niveles de expresión génica, incluida la regulación de la arquitectura de la cromatina (*long-ncRNA* o *lncRNA*), la transcripción (*lncRNA*) y la traducción (*micro-RNA* –*miRNA*– y RNA pequeño de interferencia –*siRNA*–)<sup>76</sup>. En el caso de los RNA-*enhancer* (eRNA), su expresión se ha correlacionado con la actividad de su correspondiente *enhancer*<sup>77</sup>. También se ha descrito un sistema de regulación de los propios miRNAs mediante *competing endogenous RNA* (ceRNA). Consiste en un grupo de diferentes ncRNAs que compiten con los mRNAs por un mismo grupo de miRNA. Al hacer esto, regulan la capacidad de los miRNAs para inhibir la traducción del mRNA<sup>78</sup>.

B) Los **intrones** o secuencias que flanquean los exones representan el 26% del genoma y forman parte de la transcripción primaria del RNA, siendo finalmente eliminados del transcrito maduro<sup>79</sup>. A medida que se estudia la función de los intrones, existe una evidencia



cada vez mayor de que los intrones tienen funciones importantes en varios procesos relacionados con la transcripción de genes<sup>79</sup>. En términos generales, se han identificado al menos cuatro clases distintas de intrones: (i) intrones en genes que codifican proteínas y que son eliminados por el *spliceosoma*; (ii) intrones en ncRNA eliminados por proteínas; (iii y iv) intrones con corte y empalme en el que el propio intrón actúa como catalizador en su eliminación, por lo que no se requiere de proteínas (*autosplicing*). Existen dos tipos de intrones que actúan como ribozimas, los intrones del grupo I y los del grupo II<sup>75</sup>.

- C) **Elementos reguladores en cis:** Una función primordial del DNA no codificante es la regulación de la expresión génica. Actualmente hay catalogadas cientos de miles de regiones genómicas funcionales destinadas a esta tarea<sup>80</sup> (apartado 3.2.1. de la Introducción). Las secuencias de unión a factores de transcripción (FT) –proteínas de unión al DNA que controlan la transcripción a RNA– se definen como elementos reguladores en *cis* y son cruciales para la adecuada regulación del genoma ya que orquestan los programas genéticos necesarios para el desarrollo embrionario, así como la diferenciación y función de los diferentes tipos celulares que componen el organismo<sup>81-83</sup>.
- D) **Pseudogenes:** Otros elementos presentes en el DNA no codificante son los pseudogenes, secuencias derivadas de genes del propio genoma que han perdido la capacidad de expresarse. Son subproductos evolutivos aparecidos por retrotransposición o por la duplicación de genes funcionales. El genoma humano contiene más de 14 000 pseudogenes conocidos<sup>84</sup>, y aunque no son codificantes, muchos de ellos han demostrado tener papeles importantes tanto en la fisiología normal como en procesos patológicos<sup>85</sup>. De hecho, los pseudogenes pueden transcribirse en ncRNA funcionales como miRNA o siRNA<sup>86</sup>.
- E) **Secuencias repetitivas:** El genoma humano es el resultado de procesos naturales en evolución constante. Aparte de los elementos descritos anteriormente, el genoma actual está constituido en más de un 50% por fragmentos redundantes o de alta identidad de secuencia con otras regiones genómicas<sup>87</sup>. Las secuencias repetitivas pueden desarrollar diferentes funciones importantes, en función del tipo de elemento:
1. Las repeticiones en tándem, o *Variable Number Tandem Repeats* (VNTR): repeticiones de baja complejidad con cientos de miles de copias a lo largo del genoma, representando un 4.2% del DNA total de la célula. Juegan un rol crítico en el metabolismo del DNA en la célula (como los centrómeros) o en la protección de la integridad de la estructura del cromosoma (los telómeros)<sup>88</sup>.
  2. Duplicaciones segmentarias o *Low copy repeats* (LCRs): consisten en secuencias de entre 10 y 300 kb, con una identidad de secuencia compartida del 95% y con una alta homología

con diversas regiones del genoma eucariota. Aunque son poco frecuentes en los genomas de los mamíferos, las duplicaciones segmentarias representan un 8% del genoma humano<sup>89</sup>.

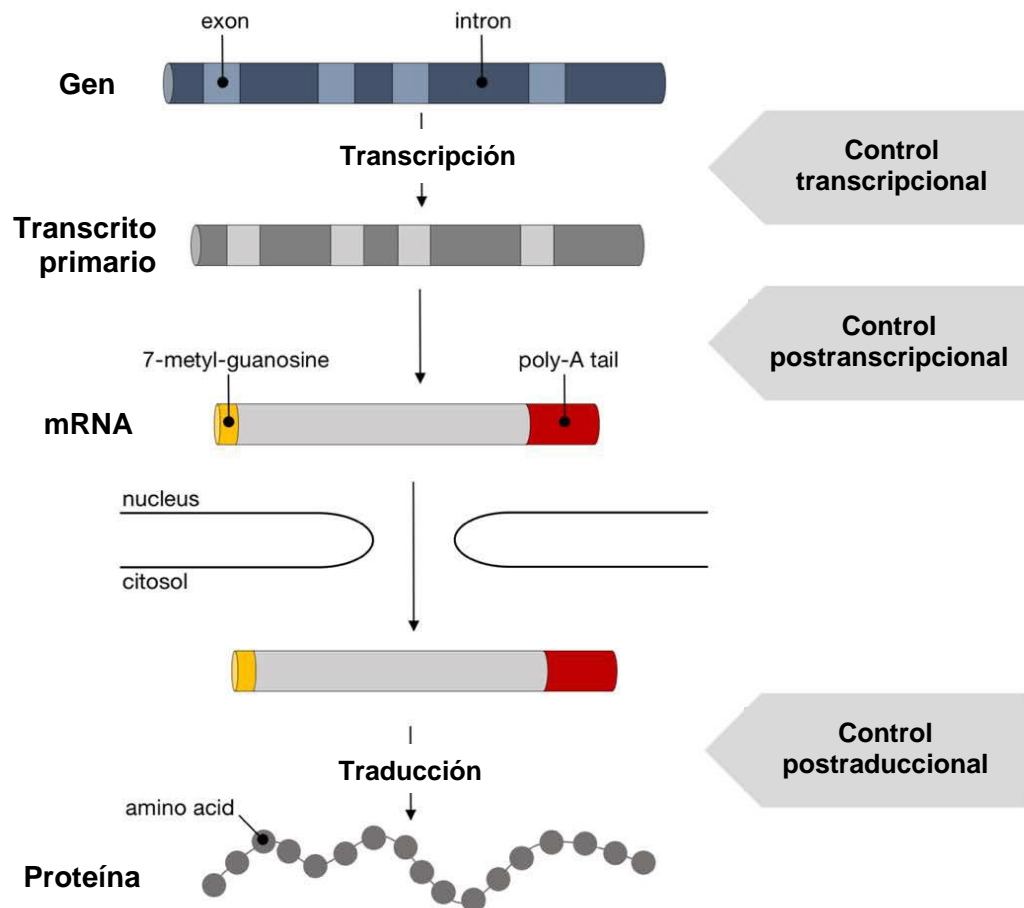
3. Transposones: Son elementos genéticos con capacidad de moverse a través del genoma. Pueden insertarse en regiones concretas y arrastrar con ellos otros elementos funcionales, cambiándolos de localización o incluso alterando su número de copias en el genoma<sup>91</sup>. En los genomas individuales se encuentran unas 300 000 copias de transposones de DNA, equivalente a un 3.6% del genoma<sup>91</sup>. A diferencia de los retrotransposones, éstos pueden propagarse de forma autónoma (sin un intermediario de mRNA), ya que llevan codificada la secuencia de una transposasa.
4. Los retrotransposones, en cambio, son remanentes de retrovirus endógenos acumulados durante miles de años de evolución que han perdido la capacidad de infectar a la célula.
  - Los retrotransposones de tipo LTR (*Long Terminal Repeat*) consisten en largas repeticiones que flanquean diversas secuencias codificantes para proteínas enzimáticas y estructurales. Éstas permiten la transcripción reversa y la integración del DNA complementario dentro del genoma de forma muy similar a la que llevaría a cabo un retrovirus<sup>92</sup>.
  - A su vez, los de tipo no LTR (los LINEs y SINEs (*Long / Short Interspersed Nuclear Elements*)), dependen o bien de las proteínas codificadas por ellos mismos, o de la acción de otros elementos móviles para su propagación, respectivamente. Los LINEs tienen una longitud media de 7 kb; por individuo se encuentran unas 100 000 copias a lo largo del genoma, representando un 20.4% del mismo<sup>92</sup>. Los SINEs son mucho más pequeños y frecuentes, con unas 1 500 000 copias, el equivalente al 13.1% del total de DNA celular<sup>93</sup>.

Dada la composición del genoma y los procesos fisiológicos y patológicos en los que intervienen los elementos que lo forman, surge la duda sobre qué porcentaje del genoma es realmente funcional. Para hacer frente a este reto se han desarrollado técnicas que se basan en la tradicional PCR cuantitativa en tiempo real (qPCR) hasta sistemas más complejos, como la secuenciación de nueva generación (NGS) o la espectrometría de masas<sup>95</sup>. Además, estos tipos de estudios también requieren un análisis bioinformático preciso para obtener resultados fiables. Estos métodos brindan la oportunidad de realizar un análisis funcional completo que involucra varios campos de estudio: genómica, epigenómica, proteómica e interactómica<sup>95</sup>. Además, las técnicas de edición genómica también permiten estudiar los posibles elementos funcionales del genoma a gran velocidad, aunque a un elevado coste<sup>94</sup>. Esto es, precisamente, uno de los pilares de esta tesis.

### 3.2. Regulación de la expresión génica

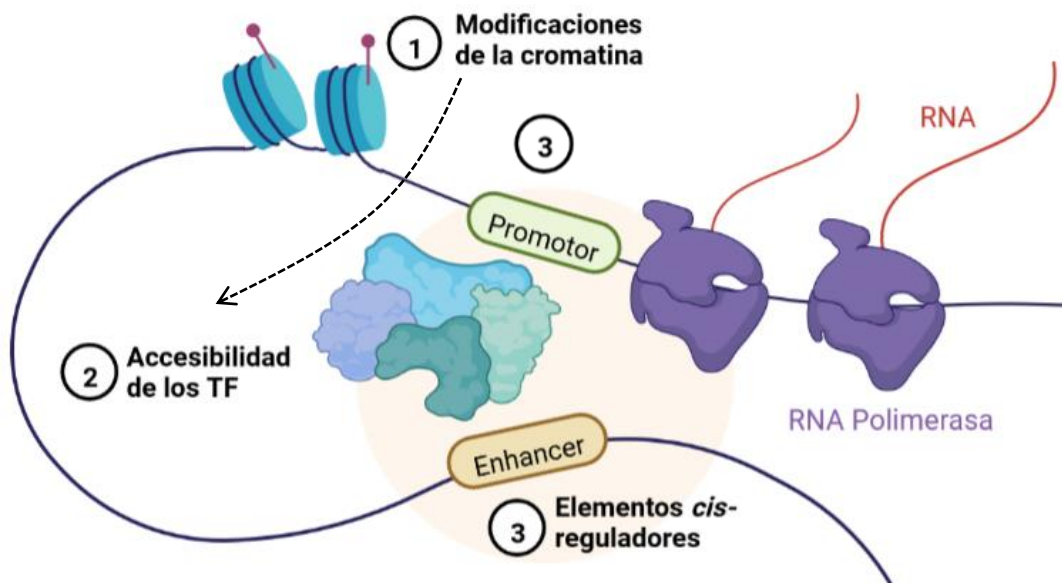
La regulación de la expresión génica incluye una amplia gama de mecanismos destinados a aumentar o disminuir la producción de productos génicos específicos (proteína o ARN). Los programas de expresión génica se regulan de manera muy sofisticada para garantizar una expresión espaciotemporal precisa y estricta de los genes. Así, cualquier alteración en los mecanismos que regulan la expresión génica puede tener graves consecuencias que pueden conducir a la enfermedad<sup>96</sup>.

La expresión génica se puede regular en cualquiera de las etapas que tiene lugar desde la secuencia de DNA al producto génico: transcripción, procesamiento del mRNA y modificaciones postraduccionales de una proteína (**Figura 8**):



**Figura 8. Regulación de la expresión génica.** Representación esquemática de las diferentes etapas en la vía del DNA-mRNA-proteína en las que se puede regular la expresión génica. Figura adaptada de M. Pinsach<sup>8</sup>.

- **Regulación transcripcional:** representa el primer y más importante nivel de regulación. Implica todos los mecanismos que controlan el ensamblaje del complejo RNA pol II para iniciar la transcripción génica (**Figura 9**). La regulación transcripcional depende principalmente de dos factores: (i) la unión de los FT a los elementos reguladores *cis*<sup>97</sup> (apartado 3.3. de la Introducción), y (ii) las modificaciones de la cromatina que permitirán la accesibilidad de los FT al DNA<sup>98</sup> (apartado 3.4.2. de la Introducción). Aparte de estos dos factores, un tercer elemento que también está relacionado con el control de la transcripción génica es la organización tridimensional del genoma. Esta organización espacial del genoma facilita la interacción entre los elementos reguladores en *cis* y sus genes diana, pero también bloquea la interacción de los elementos reguladores en *cis* con las regiones fuera del objetivo<sup>99</sup>.



**Figura 9. Regulación transcripcional.** Se representan los elementos que intervienen en la regulación transcripcional de los genes. (1) Modificaciones de la cromatina. (2) Accesibilidad de los TF al DNA. Depende de las modificaciones de la cromatina. (3) Elementos *cis*-reguladores (promotor y *enhancer*). El promotor también recluta toda la maquinaria basal para el inicio de la transcripción. Figura creada con BioRender (<https://app.biorender.com/>).

- **Regulación postranscripcional:** representa el segundo nivel de regulación. Implica todas las modificaciones que ocurren durante el procesamiento del mRNA antes del proceso de traducción a proteína (**Figura 8**). Estas modificaciones incluyen: (i) adición de una cola poli-A en el extremo 3', (ii) adición de 7-metil-guanosina en el extremo 5' y (iii) *splicing* de intrones<sup>75</sup>. Algunos ncRNAs como los miRNA y los siRNA también intervienen en este nivel de regulación, tal y como se comentó anteriormente.

- **Regulación postraduccional:** representa el tercer nivel de regulación. Se refiere a todas las modificaciones añadidas después de la traducción para que las proteínas sean funcionales. La gran variedad existente de modificaciones y su aparición simultánea en las proteínas aumentan drásticamente el tamaño del proteoma, hasta alcanzar millones de posibles formas de proteínas (**Figura 10**). Estas modificaciones modulan varias propiedades de las proteínas como la activación/desactivación de su función catalítica, su localización en diferentes compartimentos celulares, su degradación o su afinidad de unión al DNA. Las modificaciones químicas son reversibles e incluyen fosforilación, acetilación, metilación y modificaciones de oxidación/reducción (redox). Las modificaciones de polipéptidos también son enzimáticamente reversibles e incluyen ubiquitinación, SUMOilación y ubiquitinación. Las modificaciones por moléculas complejas son reversibles e incluyen glicosilación, unión de lípidos (por ejemplo, acilación o prenilación), ADP-ribosilación y AMPilación. Finalmente, las modificaciones de los aminoácidos o del esqueleto polipeptídico son irreversibles (por ejemplo, la desamidación)<sup>75</sup>.

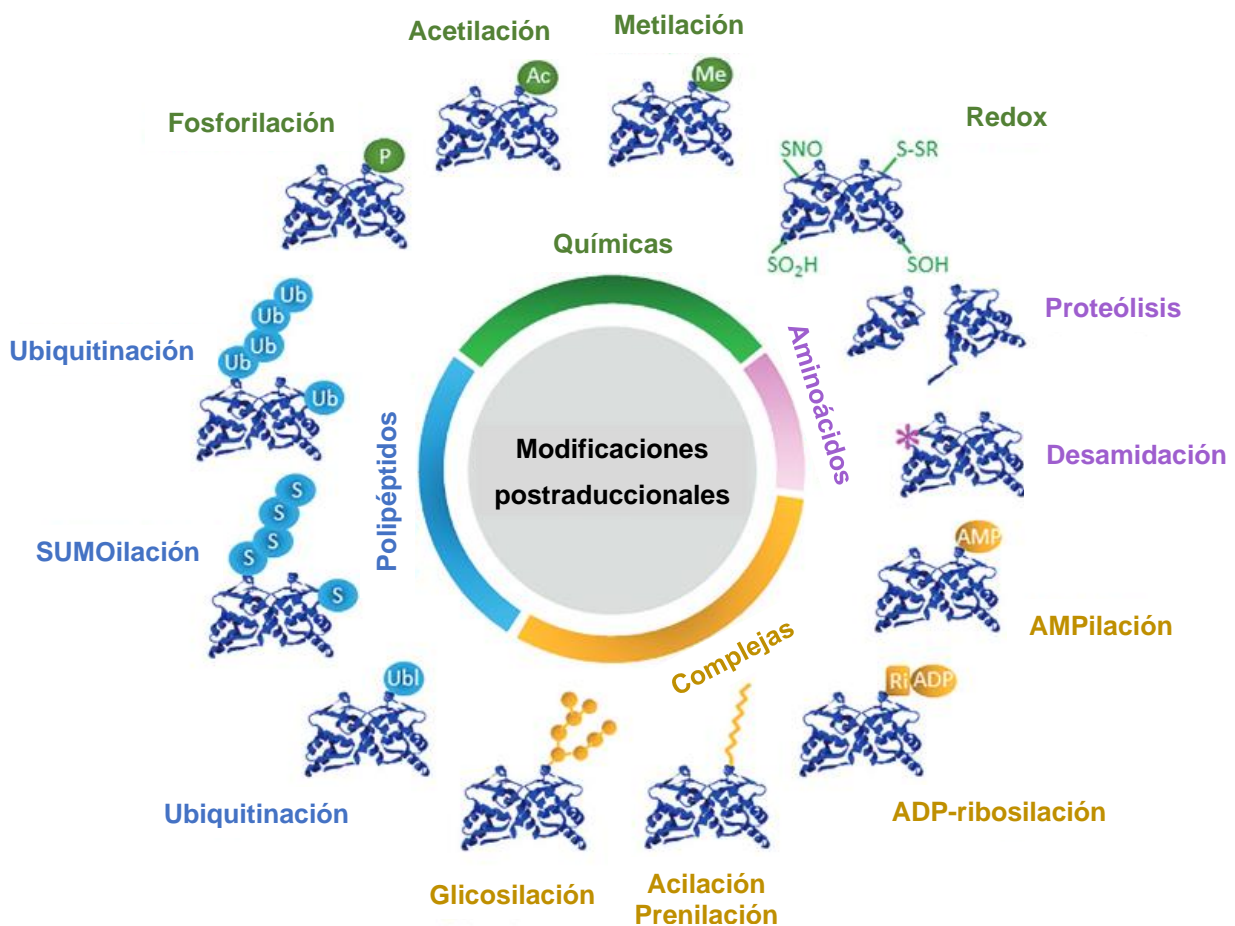


Figura 10. Diferentes tipos de modificaciones postraduccionales. Figura adaptada de Spoel *et al.*<sup>468</sup>.

### 3.2.1. Elementos reguladores en *cis*

Los elementos reguladores en *cis* y los FT se explican en esta tesis en secciones separadas. Sin embargo, su función puede verse como una sola entidad dado que los elementos reguladores en *cis* requieren la unión de los FT para cumplir su función reguladora<sup>100</sup>. En los genomas humanos podemos distinguir cuatro tipos de elementos reguladores en *cis*:

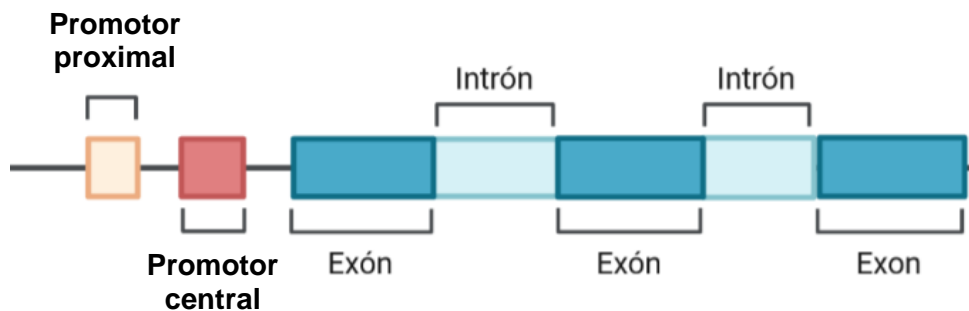
#### 3.2.1.1. Promotores

Los promotores son secuencias de DNA que indican el comienzo de la transcripción de genes que codifican proteínas. Se pueden distinguir dos regiones diferentes dentro de la secuencia del promotor: el promotor central y el promotor proximal (**Figura 11**).

El **promotor central** es la región que sirve como sitio de acoplamiento para que se ensamble la maquinaria transcripcional básica (RNA pol II y FT generales). Define la posición del sitio de inicio de la transcripción (TSS), así como la dirección de la transcripción<sup>101</sup>. El primer elemento promotor central descrito fue la **TATA box**, llamada así por su secuencia de DNA conservada TATAAA. La primera hipótesis tras la identificación de la TATA box fue que los promotores centrales eran universales. Sin embargo, el análisis estadístico de aproximadamente 10 000 predicciones de promotores humanos reveló que los promotores centrales contenían otros tipos de secuencias además de la TATA box<sup>102</sup>. Se cree que esta diversidad en la composición del promotor central tiene un significado funcional, ya que se ha observado que diferentes promotores centrales pueden limitar la cantidad de señales reguladoras a las que responderán<sup>101,103</sup>. Con los datos de secuenciación recientes, se puede confirmar que la TATA box solo está presente en un 24% de los promotores humanos y que, además, sólo un 10% de los promotores con TATA box tienen la secuencia consenso exacta TATAAA. El 76% de los promotores restantes sin TATA box son ricos en CG y tienen un alto contenido de sitios de unión para el FT Sp1. Sp1 es una proteína tipo dedo de zinc que se une a regiones con un elevado contenido de CG, y regula muchos procesos celulares, como la diferenciación y el crecimiento celular, la apoptosis, la reparación del DNA y la remodelación de la cromatina. Otro elemento importante del promotor central es el **iniciador**, que se encuentra en un 46% de los promotores humanos. Se ha observado que la mayoría de promotores que tienen iniciador no contienen la TATA box y, por tanto, podrían ser elementos mutuamente excluyentes. El promotor central también puede incluir otros elementos como el DPE (*downstream promoter element*), donde se unen FT que forman parte del complejo de iniciación de la transcripción, y el BRE (*B recognition element*), que es un sitio de unión de TFIIB (FT que forma parte del complejo de preiniciación de la transcripción)<sup>104,105</sup>.

El **promotor proximal** es una región ubicada inmediatamente *upstream* del TSS, que contiene una alta densidad de sitios de unión para FT con función activadora de la transcripción<sup>102</sup>. Aproximadamente un 60% de los promotores proximales humanos se encuentran cerca de una

isla CpG, que es una secuencia de DNA relativamente corta (típicamente de 500 pb a 2 kb de longitud) enriquecida en nucleótidos GC. Muchas islas CpG dispersas por todo el genoma están metiladas; sin embargo, aquellas islas CpG cercanas a los promotores proximales no lo están. De hecho, existen correlaciones entre la presencia de islas CpG y ciertos elementos promotores centrales. Por ejemplo, las TATA *box* son más comunes en promotores que no tienen una isla CpG cerca<sup>102,104</sup>.



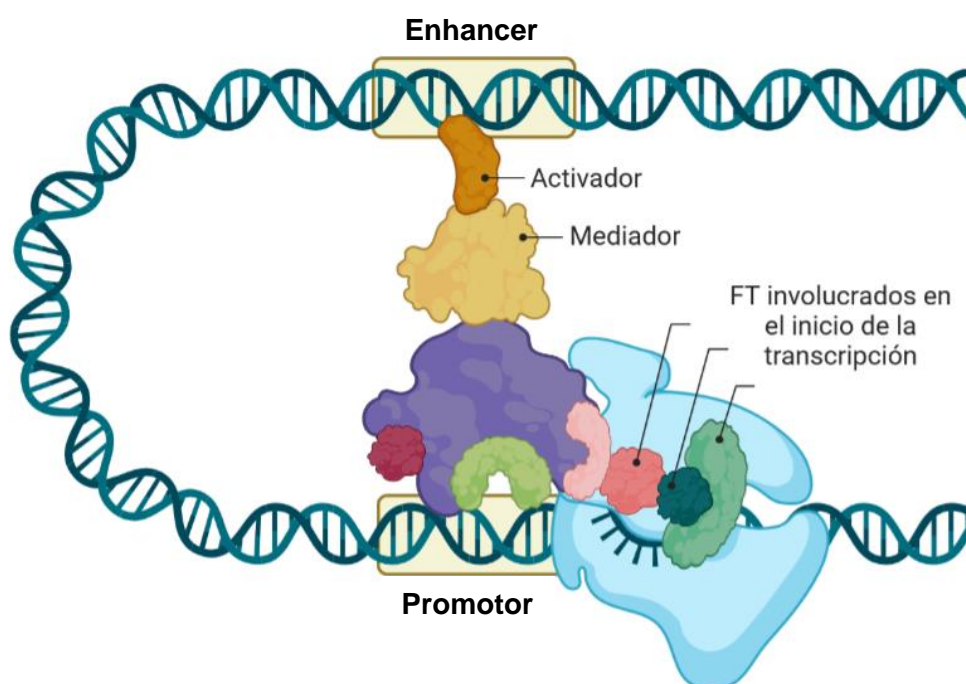
**Figura 11. Representación esquemática de un gen y su región promotora.** El promotor proximal está ubicado *upstream* del TSS y contiene múltiples sitios de unión para los FT. El promotor central contiene el TSS y sirve como sitio de acoplamiento para la RNA pol II. Figura creada con BioRender (<https://app.biorender.com/>).

### 3.2.1.2. *Enhancers*

Los *enhancers* son secuencias cortas de DNA (100-1000 pb) que actúan como una plataforma para el ensamblaje cooperativo de determinados FT. Estos elementos potencian la transcripción génica y organizan con precisión los patrones de expresión de genes necesarios para numerosos procesos, incluido el desarrollo embrionario y la función de los organismos<sup>106</sup>. A diferencia de la mayoría de los promotores, los *enhancers* suelen ser elementos que pueden localizarse varios cientos de kb *upstream* de su gen diana y pueden regular la transcripción de genes independientemente de su orientación, posición y distancia con los promotores<sup>107</sup>. La interacción entre los *enhancers* y sus promotores diana está mediada por un complejo de varios FT y co-factores como, por ejemplo, el *mediator*<sup>108</sup> (**Figura 12**). El complejo *mediator*, compuesto por co-factores unidos por interacciones proteína-proteína, está involucrado en la formación de un bucle de DNA que acerca los *enhancers* y promotores, lo que permite la interacción de los *enhancers* con la maquinaria transcripcional anclada en sus promotores diana.

La identificación de los *enhancers* es un desafío debido a que suelen estar localizados a grandes distancias del gen que regula, en la zona 3', 5' o dentro de intrones. Otra razón que complica su identificación es que tienen unas secuencias muy variables, lo que dificulta su búsqueda computacional<sup>107</sup>. Se han desarrollado varias técnicas para identificar y evaluar la

actividad de elementos reguladores dentro de su contexto endógeno, derivadas de las ya conocidas técnicas de hipersensibilidad a la digestión por enzimas, como la Dnasa I, e inmunoprecipitación de cromatina (ChIP) (apartado 3.5. de la Introducción). Hasta la fecha, se han identificado >100 000 *enhancers* en el genoma humano, superando en número a los genes en aproximadamente un orden de magnitud<sup>109</sup>. Este hallazgo, junto con la dificultad para determinar qué gen o genes son la diana de un *enhancer* particular, ha planteado la cuestión de la funcionalidad del *enhancer*. Se ha sugerido que diferentes *enhancers* podrían regular el mismo promotor, pero sigue sin estar claro si diferentes *enhancers* de mamíferos regulan el mismo promotor en distintos momentos espaciotemporales<sup>110,111</sup>, o si esta complejidad regulatoria da como resultado una redundancia funcional entre *enhancers* asociados con el mismo gen<sup>112,113</sup>.



**Figura 12. Representación esquemática de la función *enhancer*.** Interacción entre un *enhancer* con la maquinaria transcripcional (FT y RNA pol II) anclada en su promotor diana. También se muestra el complejo mediador involucrado en la formación del bucle de DNA. Figura creada con BioRender (<https://app.biorender.com/>).

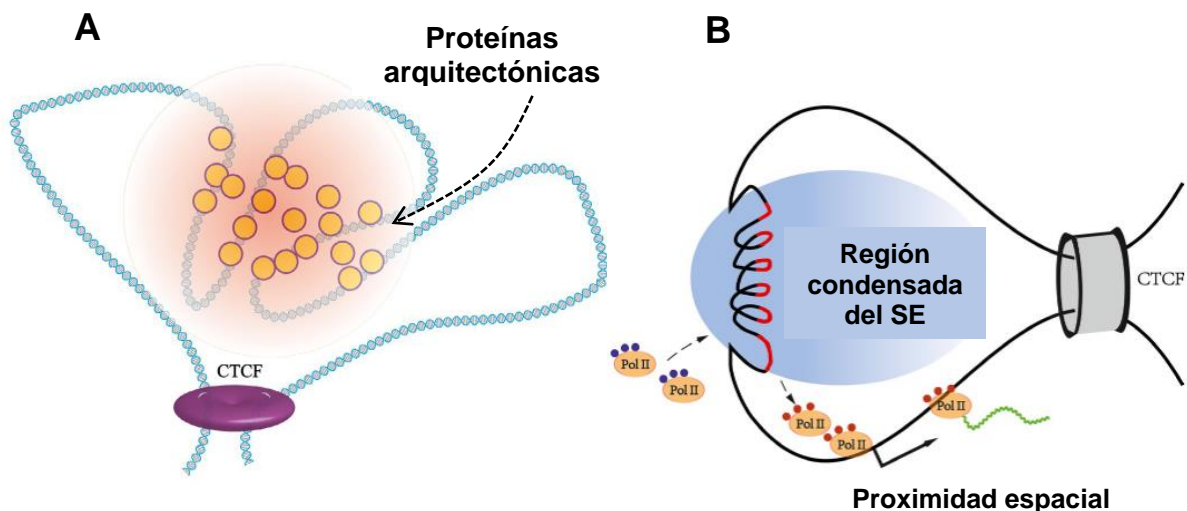
Además de los *enhancers*, recientemente se ha descrito otra clase de elementos reguladores. Estos se denominan ***superenhancers*** (SE) porque están asociados a genes que dictan la identidad celular. Los SE tienden a abarcar grandes regiones genómicas, con un tamaño medio generalmente un orden de magnitud mayor que el de los *enhancers* convencionales<sup>114</sup>. Los SE activos están muy ocupados por FT y coactivadores, especialmente el complejo *mediator*<sup>115</sup>. Al igual que *mediator*, una serie de factores generalmente asociados con la actividad *enhancer* muestran un enriquecimiento en los SE en comparación con los *enhancers* normales, incluida la



RNA pol II, los eRNAs, las histonas acetiltransferasas p300, factores de la cromatina como la cohesina, las modificaciones de histonas H3K27ac, H3K4me2 y H3K4me1, y una mayor accesibilidad a la cromatina<sup>116</sup>. La detección de todos estos elementos en conjunto es una manera para caracterizar los SE.

Tanto los *enhancers* como los SE están enriquecidos en variantes genéticas asociadas a enfermedades, particularmente cuando se han sido caracterizados en tipos celulares relevantes para la enfermedad en cuestión<sup>114</sup>. Por ejemplo, los SE identificados en el tejido cerebral humano estaban enriquecidos en variantes asociadas con la enfermedad de Alzheimer. Asimismo, también se encontró que las variantes asociadas con la artritis reumatoide residen en los SE identificados en las células B y T23<sup>118</sup>.

Los SE dependen de un estado de la cromatina determinado. El genoma se divide en regiones genómicas con la ayuda de proteínas que modulan la arquitectura de la cromatina. En las regiones donde se encuentran los SE, la maquinaria de transcripción forma una región condensada y aislada que mantiene una regulación precisa de los genes<sup>119</sup>. Esta regulación se basa en el reclutamiento de la maquinaria de transcripción por parte de la región condensada y una activación transcripcional de los genes regulados por el SE mediante proximidad espacial con la región condensada<sup>119</sup>, en lugar del modelo de “puente” descrito en la **Figura 12 (Figura 13)**.



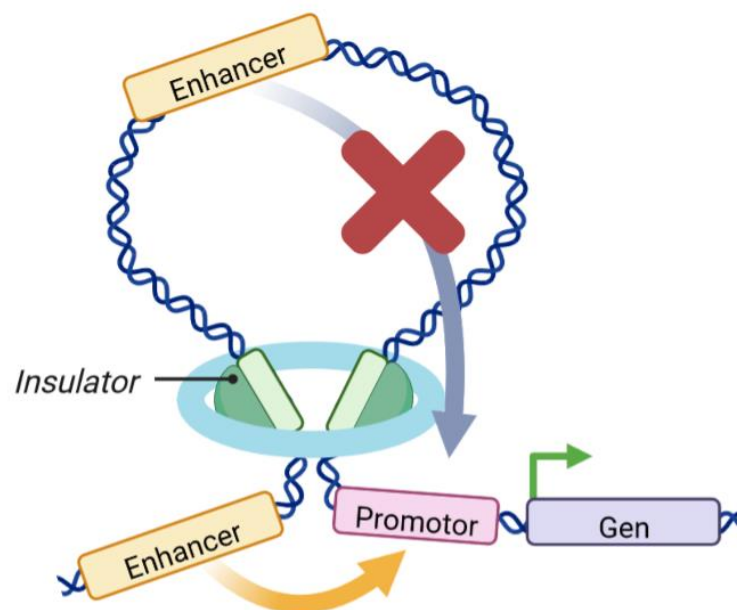
**Figura 13. Modelo de regulación transcripcional asociado a los SE.** (A) Con la ayuda de proteínas arquitectónicas, se configura la estructura de la cromatina que conforma el SE. (B) Las regiones condensadas del SE (regiones en rojo) reclutan la maquinaria de transcripción y se inicia el proceso. A continuación, la maquinaria transcripcional se libera y es reclutada por los genes situados *downstream* de la región condensada. Figura adaptada de Zhang *et al.*<sup>119</sup>.

### 3.2.1.3. Silenciadores

Los silenciadores son secuencias de DNA donde los FT se unen para reprimir la transcripción génica. En este caso, los FT funcionan como represores y sus co-factores, son co-represores. Al igual que los *enhancers*, los silenciadores funcionan independientemente de la distancia y la orientación en relación con sus promotores objetivo. Se ha sugerido que los silenciadores reprimen la transcripción génica utilizando el mismo mecanismo de formación de bucles que los *enhancers*<sup>104</sup>.

### 3.2.1.4. *Insulators* o elementos aislantes

Los *insulators* son secuencias de DNA que restringen los límites funcionales de los *enhancers* y silenciadores (**Figura 14**). Dividen el genoma en áreas de regulación de la expresión génica y juegan un papel crucial en la organización tridimensional del genoma humano<sup>75,83</sup> (apartado 3.4.3. de la Introducción).



**Figura 14. Representación esquemática de la función del *insulator*.** El *insulator* que se muestra bloquea la propagación de la actividad *enhancer* a otros genes que se encuentran fuera de los límites. Figura creada con BioRender (<https://app.biorender.com/>).

### 3.3. Los factores de transcripción

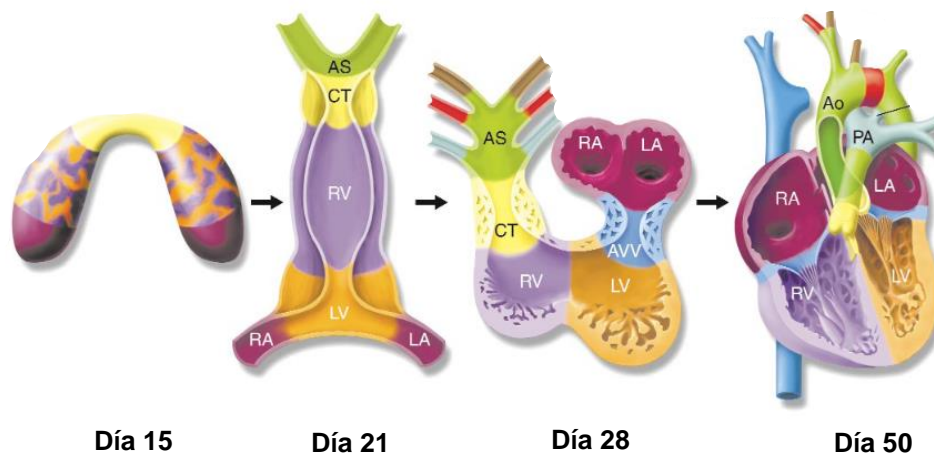
Los FT son proteínas de unión al DNA que participan en la regulación transcripcional. Pueden activar o reprimir la transcripción génica mediante la unión directa al DNA y del reclutamiento de co-factores (co-activadores o co-represores) que no se unen directamente al DNA, pero promueven la activación o represión a través de diferentes mecanismos. Todos los FT tienen como mínimo un dominio de unión al DNA y un dominio de interacción proteína-proteína. Básicamente, se distinguen dos tipos de FT: los FT generales y los FT específicos<sup>120</sup>.

- A) **Factores de transcripción generales (FTG) o basales.** Los FTG (TFIIA, TFIIB, TFIID, TFIIIE y TFIIH) se unen al promotor de los genes y son necesarios para el inicio de la transcripción de cualquier gen. Estas proteínas reclutan la RNA pol II y la sitúan en el TSS. Los FTG y la RNA pol II forman un complejo llamado complejo de preiniciación (CPI). La unión del CPI al promotor por sí sola, además de la fosforilación de la cola C-terminal de RNA pol II, es suficiente para inducir una transcripción basal<sup>121</sup>.
- B) **Factores de transcripción específicos.** Los FT específicos reconocen y se unen a regiones reguladoras del DNA. Pueden ser activadores o represores dependiendo de si activan o reprimen la actividad de la RNA pol II. Sin embargo, un mismo FT puede realizar ambas funciones en contextos distintos. Mientras que los FTG contribuyen a la transcripción basal de los genes, los FT específicos activadores son los responsables de la expresión génica diferencial<sup>120</sup>.

Los activadores se unen a los *enhancers* e interactúan con la maquinaria transcripcional anclada en las regiones promotoras a través de la formación de bucles de DNA (**Figura 12**). Esta interacción es facilitada por el complejo mediator<sup>122</sup>. Los represores se unen a los silenciadores y se cree que puede reprimir la actividad de la maquinaria transcripcional a través de diferentes mecanismos. Por ejemplo, pueden competir con los activadores para interactuar con la maquinaria transcripcional a través de la misma formación de bucles de DNA, o pueden unirse directamente a los activadores a través de interacciones proteína-proteína para evitar la interacción de los activadores con la maquinaria transcripcional, lo que se conoce también como silenciamiento<sup>123</sup>. Además, los activadores y represores pueden influir en la transcripción al cambiar la estructura de la cromatina en los alrededores de sus sitios de unión. Pueden reclutar modificadores de cromatina que, a su vez, aumentarán o disminuirán la accesibilidad del DNA a otros FT<sup>124</sup>.

### 3.3.1. Regulación de la expresión génica cardíaca por factores de transcripción

El primer órgano que se forma en el embrión es el corazón, en un proceso llamado cardiogénesis (**Figura 15**). En una primera etapa tiene lugar la formación de un tubo cardíaco, llamado miocardio embrionario, que conduce el impulso eléctrico a una velocidad muy baja. Este tubo se va elongando y comienzan a diferenciarse regiones específicas que acabarán formando las dos aurículas y los dos ventrículos<sup>125</sup>. A partir de progenitores cardíacos se generan los diferentes tipos de células del corazón (cardiomiocitos de las aurículas, de los ventrículos, del sistema de conducción...). En etapas muy tempranas de la cardiogénesis se observa ya un rápido incremento de la expresión de FT cardíacos. Estos FT regulan la expresión de los genes responsables de la formación del corazón y la morfogénesis de las estructuras cardíacas y, por tanto, son claves para definir el destino de las células cardíacas<sup>126,127</sup>.



**Figura 15. Esquema de la morfogénesis cardíaca.** Las ilustraciones representan el desarrollo cardíaco con codificación de colores de las regiones morfológicamente relacionadas. Los precursores cardíacos forman una media luna (Día 15) que se transforma en un tubo cardíaco primitivo el día 21, con su ventrículo y aurícula primitivos. Para el día 28, el tubo cardíaco lineal hace un *loop* hacia la derecha (*D-loop*) para establecer la futura posición de las regiones cardíacas (aurículas, ventrículos, tracto de salida). Hacia el día 50, se adquiere la estructura del corazón maduro. Abreviaciones: AS, saco aórtico; CT, región conotroncal; Ao, aorta; LA, aurícula izquierda; LV, ventrículo izquierdo; PA, arteria pulmonar; RA, aurícula derecha; RV, ventrículo derecho; AVV, válvula atrioventricular. Figura adaptada de Srivastava & Olson<sup>130</sup>.

Se ha identificado un grupo de FT cardíacos centrales que son claves en la regulación transcripcional cardíaca tanto en el embrión como en el adulto. Estos FT cardíacos centrales incluyen: NKX2-5, GATA4, TBX5, MEF2C y HAND1/2. En la evolución desde organismos más sencillos ( artrópodos, etc) a mamíferos, la red transcripcional cardíaca se ha vuelto cada vez más compleja, con un aumento del número de genes que codifican por diferentes tipos de los FT cardíacos centrales<sup>128</sup>. El incremento de FT cardíacos ha dado lugar a nuevos patrones de expresión génica, lo que se ha traducido en un corazón cada vez más complejo evolutivamente,

que ha pasado de ser un simple tubo cardíaco a ser una potente bomba contráctil formada por cuatro cámaras. A pesar de los cambios a lo largo de la evolución, la conservación de los FT cardíacos centrales en la red de transcripción cardíaca apoya su papel clave para el correcto desarrollo y funcionamiento del corazón. Cabe destacar también que los FT cardíacos centrales a menudo tienen funciones redundantes, y en caso de ausencia de uno de ellos pueden parcialmente compensar su función para activar los genes diana<sup>129</sup>. Otra propiedad de estos FT es el hecho de que a menudo interactúan entre ellos formando complejos multiproteicos para activar cooperativamente la expresión de sus genes diana. Las evidencias para esta acción cooperativa provienen de estudios en los que se han identificado múltiples genes cardíacos con lugares de unión para diversos FT cardíacos centrales<sup>129</sup>. En este sentido, estudios de ChIP-seq sugieren que los *enhancers* cardíacos activos están a menudo definidos por aquellas regiones del genoma co-ocupadas simultáneamente por varios FT cardíacos centrales<sup>131</sup>.

Alteraciones en los FT cardíacos centrales se han asociado con diversas enfermedades, porque provocan un funcionamiento anómalo de la red de transcripción cardíaca, dando lugar a cambios en los patrones de expresión de genes del corazón. En función del momento en que ocurran estas alteraciones, las patologías cardíacas pueden ser congénitas o desarrollarse en el adulto. A continuación, se muestran algunos ejemplos de alteraciones de cada uno de los FT centrales:

- A) **NKX2-5**: tiene un papel clave en la formación y el desarrollo del corazón. En consecuencia, las alteraciones genéticas en NKX2-5 están relacionadas con defectos del desarrollo cardíaco. Los ratones con un *knockdown* de NKX2-5 tiene un corazón defectuoso y una expresión disminuida de marcadores ventriculares<sup>132,133</sup>. En humanos, las variantes genéticas que afectan la capacidad de unión al DNA de NKX2-5 se han relacionado con alteraciones en la conducción eléctrica auriculoventricular<sup>134</sup>. Además, el FT NKX2-5 regula la expresión de varios de los canales iónicos cardíacos más importantes durante el desarrollo del corazón embrionario, incluidos *CACNA1C*, *KCNH2* y *SCN5A*<sup>135,136</sup>. Como se ha comentado anteriormente, estos canales juegan un papel esencial en la correcta generación y propagación del PA del corazón.
- B) **MEF2C**: es un activador de la transcripción que se une específicamente al elemento MEF2 y controla la morfogénesis y la miogénesis cardíacas. Los embriones de ratones con *knockdown* de MEF2C no logran desarrollar adecuadamente un ventrículo derecho y no expresan un subconjunto de genes específicos de cardiomiocitos<sup>137,138</sup>. Los ratones sin MEF2C también muestran alteraciones en la expresión de HAND1/2, lo que respalda el papel de MEF2C en el desarrollo ventricular<sup>137</sup>.

- C) **HAND1/2**: se expresan asimétricamente en las cámaras ventriculares en desarrollo y juegan un papel esencial en la morfogénesis cardíaca. Los ratones sin *HAND1* presentan letalidad embrionaria por la detención del desarrollo cardíaco<sup>139</sup>. Distintas evidencias sugieren que *HAND2* desempeña un papel homeostático en el mantenimiento del sistema de conducción cardíaco, ya que la pérdida de función de *HAND2* dentro del sistema de conducción postnatal en un modelo murino da como resultado anomalías en la conducción auriculoventricular<sup>140</sup>.
- D) **TBX5**: es un FT con una función central en el desarrollo cardíaco y de las extremidades anteriores. Parece ser especialmente importante para la correcta conducción cardíaca y la formación del tabique ventricular<sup>141</sup>. Los ratones sin *TBX5* muestran un desarrollo cardíaco anormal con fallos en la conducción e hipoplasia del ventrículo izquierdo, lo que resulta en letalidad embrionaria<sup>142</sup>. También se ha identificado una relación directa entre *TBX5* y *SCN5A*. Concretamente, se ha demostrado que el equilibrio de las actividades de *TBX5*, *TBX3* y *NKX2-5* es esencial para la regulación de la expresión de *SCN5A* en las distintas regiones del corazón, con una estricta necesidad de *TBX5* para la máxima expresión de *Nav1.5* en el sistema de conducción ventricular<sup>143</sup>.
- E) **GATA4**: es un FT identificado como mediador crucial de la cardiogénesis de vertebrados. Se ha demostrado que regula directamente la actividad transcripcional de numerosos genes específicos del corazón que controlan la diferenciación de células progenitoras cardíacas<sup>126</sup>. *GATA4* se describe con más detalle en el próximo apartado.

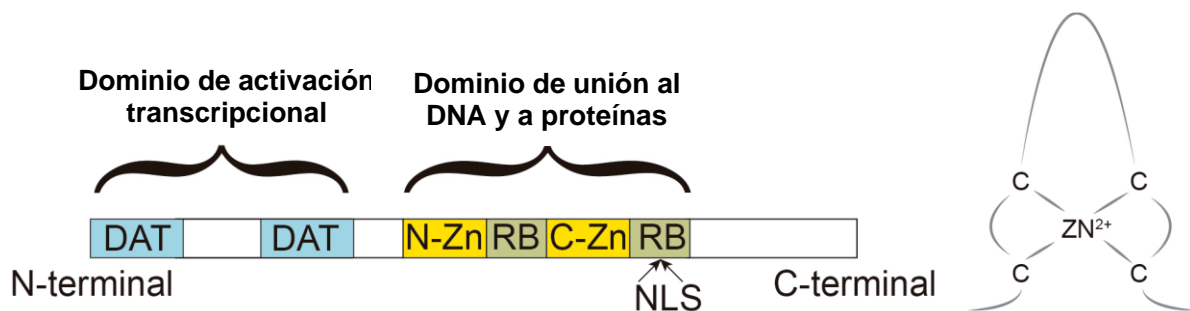
### 3.3.2. El factor de transcripción *GATA4* cardíaco

Los FT *GATA* son una familia de proteínas tipo dedo de zinc que se unen a regiones concretas del DNA para regular la expresión de genes específicos. Tienen un papel importante en la regulación transcripcional durante la diferenciación, el crecimiento y la supervivencia de distintos tipos celulares<sup>144</sup>. El genoma de mamíferos codifica para seis FT *GATA* (*GATA1-GATA6*). Se dividen en dos subgrupos en función de su patrón de expresión y su homología en la secuencia: *GATA1*, *GATA2* y *GATA3* se expresan predominantemente en las células madre hematopoyéticas; en cambio *GATA4*, *GATA5* y *GATA6* se expresan en tejidos derivados de mesodermo y endodermo, como son corazón, hígado, pulmón, gónadas e intestino<sup>145</sup>.

#### 3.3.2.1. Estructura y función

Los factores *GATA* se caracterizan por tener un dominio de unión al DNA y de interacción proteína-proteína formado por dos motivos de dedo de zinc (N-Zn y C-Zn, según si corresponde

al motivo más cercano al extremo N-terminal o C-terminal, respectivamente). Además, cada uno de los motivos de dedo de zinc contiene una región básica adyacente rica en aminoácidos lisina<sup>146</sup> (**Figura 16**). La región básica adyacente del C-Zn contiene más lisinas que la del N-Zn y comprende también una secuencia de localización nuclear<sup>147</sup>. En la región N-terminal, los factores GATA contienen uno o dos dominios de activación transcripcional, que son necesarios para la función activadora de la transcripción de sus genes diana<sup>148</sup>. Basándose en un estudio realizado con una proteína GATA truncada, se observó que sin la región N-terminal tiene lugar igualmente una ligera activación transcripcional, lo que sugiere que la región C-terminal podría también contener algún dominio con una débil actividad transcripcional<sup>149</sup>. El dominio de unión al DNA y de interacción proteína-proteína está muy conservado entre los distintos factores GATA, los cuales comparten un 75% de la secuencia de aminoácidos. Entre los factores GATA cardíacos (GATA4, GATA5 y GATA6) la similitud de este dominio es del 85%<sup>126</sup>. Por lo que respecta al resto de regiones, la homología entre los factores GATA cardíacos es sustancialmente inferior, de un 45% en el dominio N-terminal y un 35% en el dominio C-terminal. En general, cada uno de los factores GATA cardíacos está altamente conservado entre especies, siendo el dominio de unión al DNA y de interacción proteína-proteína el más conservado<sup>150,151</sup>.



**Figura 16. Estructura de los FT GATA.** A la izquierda, estructura general de los FT GATA. DAT: dominio de activación transcripcional; N-Zn: dedo de zinc N-terminal; C-Zn: dedo de zinc C-terminal; RB: región básica; NLS: sitio de localización nuclear. A la derecha, estructura de un motivo de zinc, donde el zinc se une a cuatro residuos de cisteínas. Figura adaptada de A. Tarradas<sup>18</sup>.

Los factores GATA se unen al DNA con elevada afinidad a los lugares de unión GATA (o motivos GATA), que tienen la secuencia consenso WGATAR (W=A/T, R=A/G)<sup>152</sup>. El motivo C-Zn y su región básica adyacente son necesarios y suficientes para que los factores GATA puedan unirse al DNA.

Diversos estudios han demostrado que GATA4 tiene una gran importancia en el proceso de diferenciación de los cardiomiocitos durante la cardiogénesis. Concretamente, se ha observado que la diferenciación de una línea celular en cardiomiocitos está inhibida en condiciones de silenciamiento de GATA4, además de producirse un aumento de la apoptosis. Por el contrario,

la sobreexpresión de GATA4 da lugar a una inducción de la diferenciación a células cardíacas<sup>153</sup>. En un modelo *knockdown* de ratón, se observó que la ausencia de GATA4 es compensada por un incremento de los niveles endógenos de la proteína GATA6, lo cual sugiere que GATA4 y GATA6 tendrían funciones redundantes en la diferenciación celular.

Aparte de su papel durante el desarrollo embrionario, GATA4 también es importante para la actividad eléctrica del corazón. En un estudio del Grupo de Genética Cardiovascular, realizado por Tarradas et al.<sup>157</sup>, se demostró que GATA4 regula la activación transcripcional del promotor del gen *SCN5A*. Concretamente, se observó que GATA4 modula la expresión de *SCN5A*, confirmando que GATA4 es un activador transcripcional de *SCN5A*<sup>157</sup>. Además, este estudio proporcionó evidencias de que: (i) GATA4 y GATA5 co-ocupan el promotor de *SCN5A* y la primera región intrónica del gen, interactúan, y actúan sinérgicamente para regular la expresión del gen *SCN5A* en el tejido cardíaco humano; (ii) dos lugares de unión a GATA4 que se encuentran en el intrón 1 de *SCN5A* son elementos importantes pero no los únicos que contribuyen a esta función reguladora; y (iii) los niveles de mRNA de *GATA4* y *SCN5A* se correlacionan positivamente en corazones humanos adultos<sup>157</sup>. Esto evidencia la importancia de los factores GATA cardíacos en la regulación de la expresión génica en el corazón.

Además, un estudio reciente demostró que IRX5 y GATA4 actúan sinérgicamente para activar el promotor de *SCN5A* en el corazón humano, donde *IRX5* no es capaz de promover la transcripción por sí solo, pero estaría potenciando la inducción de la expresión de *SCN5A* por parte de GATA4<sup>154</sup>. En línea con este estudio, Canac y colaboradores<sup>155</sup>, vincularon los FT IRX3 e IRX5 con tres FT cardíacos claves: GATA4, NKX2-5 y TBX5. De hecho, se sabe que GATA4 actúa en un complejo multiproteico con los FT cardíacos NKX2-5 y TBX5. Las interacciones entre estos 3 FT controlan y remodelan en el tiempo el programa de expresión transcripcional que gobierna el desarrollo del corazón<sup>154-156</sup>.

### 3.3.2.2. La acetilación de GATA4

La actividad de los factores GATA cardíacos está modulada por modificaciones postraduccionales que afectan su unión con el DNA o las interacciones con proteínas. Las modificaciones postraduccionales más estudiadas son las de GATA4, y entre ellas encontramos acetilación, fosforilación, metilación y SUMOilación<sup>158</sup>. No obstante, los estudios en esta tesis están centrados en la acetilación de GATA4, por lo que a continuación se explican los procesos relacionados con esta modificación postraduccional.

La acetilación de lisinas consiste en la adición de un grupo acetil en el grupo  $\epsilon$ -amino de la cadena lateral de la lisina. Esta modificación (y otras modificaciones postraduccionales) pueden regular diversos aspectos de las proteínas como son la activación o inhibición de la actividad



catalítica (excepto en los FT ya que no tiene actividad catalítica), la localización en los diferentes compartimentos celulares y la degradación.

Las enzimas encargadas de introducir los grupos acetil reciben el nombre de acetiltransferasas de lisinas (KAT)<sup>159</sup>. Las KAT se pueden dividir en dos grandes familias: KAT tipo A y KAT tipo B. Las KAT tipo B son principalmente citoplasmáticas y acetilan histonas recién sintetizadas antes de unirse a la cromatina<sup>160</sup>. Las KAT tipo A son una familia de enzimas más numerosa, las más importantes de las cuales son KAT3A (originalmente CBP o *CREB-binding protein*) y KAT3B (originalmente p300), dos co-activadores transcripcionales estructuralmente muy similares y funcionalmente relacionados<sup>161,162</sup>.

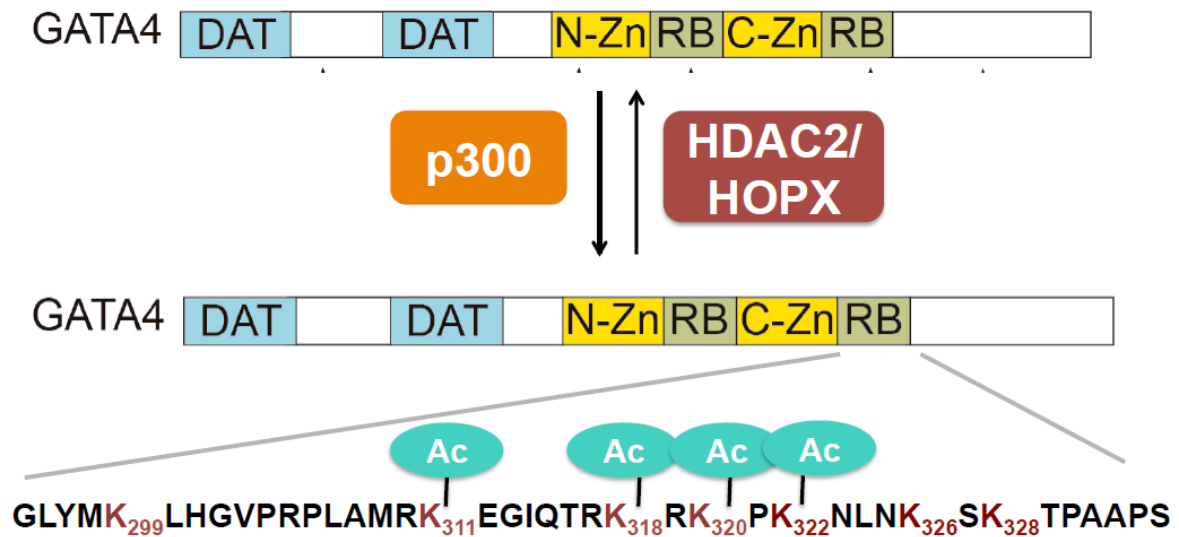
La actividad transcripcional de GATA4 está regulada por un balance de acetilación y desacetilación (**Figura 17**). La forma acetilada de GATA4 presenta una mayor afinidad de unión al DNA y un aumento de la actividad transcripcional. La acetiltransferasa p300 es la encargada de acetilar GATA4, y actúa como co-activador de GATA4<sup>162,163</sup>. Un estudio realizado por Takaya *et al.*<sup>163</sup> en el que se pretendían identificar las dianas de acetilación de GATA4, describió las lisinas K311, K318, K320 y K322, situadas en la región básica del motivo C-Zn de GATA4, como los residuos acetilados por p300<sup>163</sup>. Sin embargo, tal y como se describe con más detalle en el apartado 1.1. de resultados, el diseño del estudio de Takaya *et al.* no permite llegar a estas conclusiones.

Los niveles de acetilación de GATA están regulados por un equilibrio entre la actividad de p300 y la actividad desacetilasa de HDAC2 (**Figura 17**). Experimentos realizados en corazones de ratón han demostrado la interacción entre HDAC2 y GATA4, una interacción que está estabilizada por la proteína HOPX, que se expresa únicamente en cardiomiocitos. Una inhibición de HOPX y HDAC2 causa importantes malformaciones cardíacas y una excesiva proliferación de los cardiomiocitos, debido a una hiperacetilación de GATA 4<sup>166,167</sup>.

He *et al.* demostraron que GATA4 y p300 interaccionan y co-regulan múltiples genes cardíacos en ratones<sup>131</sup>. Concretamente, sus estudios demostraron que un 80% de las regiones del DNA enriquecidas en p300 están también ocupadas por GATA4<sup>131</sup>. En base a estudios (recientes) de ChIP-seq se ha propuesto un mecanismo por el que GATA4 y p300 activan la transcripción en el que GATA4 recluta p300 en las regiones reguladoras de sus genes diana. Una vez reclutado, p300 promueve la acetilación de la histona H3K27, favoreciendo así la activación transcripcional de estos genes<sup>164</sup>.

Además de ser la enzima de acetilar GATA4, p300 también participa en la regulación de la expresión de GATA4. Se ha observado que durante la diferenciación de células madres embrionarias a cardiomiocitos, p300 se une al promotor de GATA4 y estimula su expresión<sup>165</sup>. También se ha observado que durante la diferenciación de los cardiomiocitos aumentan los niveles de p300 y, consecuentemente, también los de GATA4 acetilado.

Cabe mencionar que las lisinas de GATA4 también pueden sufrir metilaciones (K299)<sup>372</sup>, SUMOilaciones (K366) o fosforilaciones<sup>374</sup>. La metilación de K299 (y también la metilación de la arginina R317) bloquean la acetilación por parte de p300, disminuyendo su actividad transcripcional<sup>372/373</sup>. La SUMOilación de K366, por su parte, se asocia a un aumento de su actividad transcripcional<sup>374</sup>.



**Figura 17. Procesos de acetilación y desacetilación en GATA4.** p300 y HDAC2 mantiene el equilibrio en los niveles de acetilación de GATA4. Abajo se muestran los lugares de acetilación que sugiere el estudio de Takaya et al. Figura adaptada de A. Tarradas.

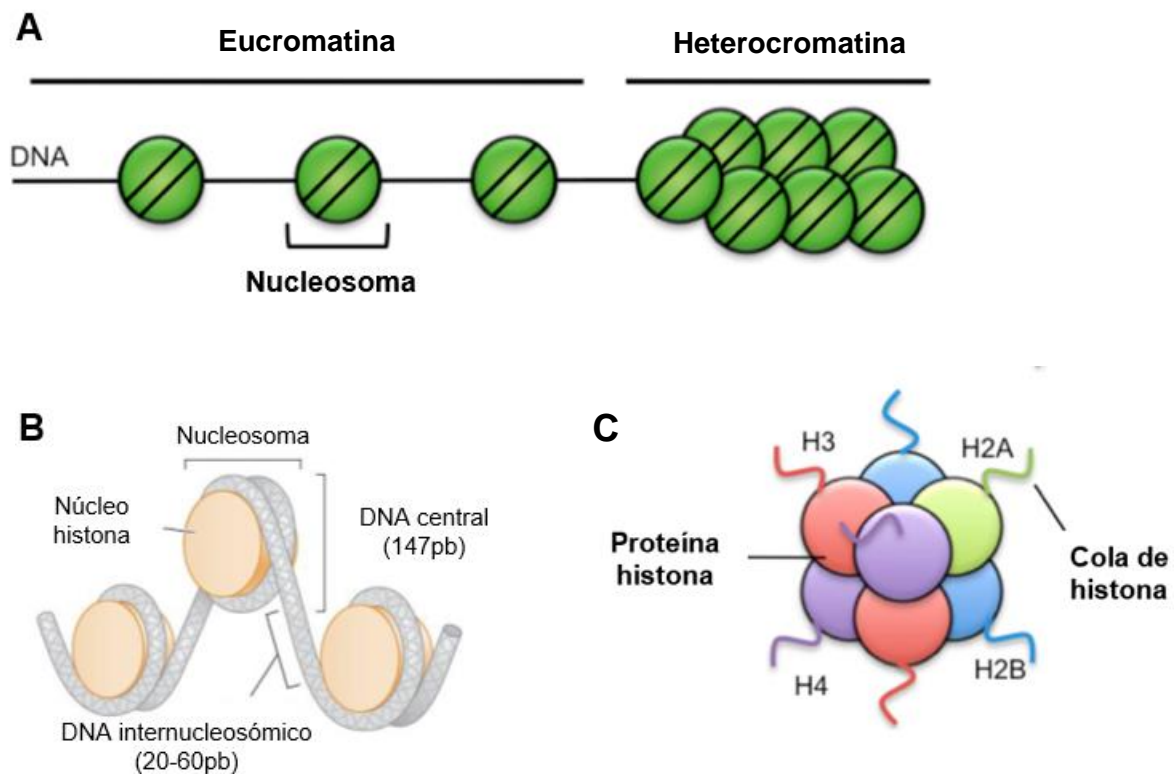
### 3.4. La cromatina

#### 3.4.1. Estructura de la cromatina

En células eucariotas, el DNA se encuentra compactado en el núcleo en forma de cromatina, un complejo formado por la interacción del DNA y unas proteínas asociadas llamadas histonas. La unidad fundamental de la cromatina es el nucleosoma, que consiste en 147 pb de DNA enrollados en un octámero de histonas. Este octámero normalmente está formado por dos unidades de cada una de las histonas H2A, H2B, H3 y H4 (**Figura 18**). La histona H1, situada entre el octámero y el DNA, ayuda a estabilizar la estructura del nucleosoma<sup>168</sup>.

Los nucleosomas están separados por regiones de DNA de longitud variable, llamado DNA internucleosómico. Las histonas tienen un dominio C-terminal de estructura globular muy similar entre sí, que es importante para la formación del nucleosoma ya que está involucrado en la interacción histona-histona e histona-DNA<sup>169</sup>. Las colas N-terminal de las histonas están menos estructuradas y sobresalen de los nucleosomas. Estas colas están sujetas a múltiples

modificaciones postraduccionales reversibles (metilación, acetilación, fosforilación...), que modulan el estado de la cromatina<sup>158</sup>.



**Figura 18. Estructura de la cromatina.** Tipos de cromatina en función de su grado de compactación y ordenados por tamaño de la estructura. **(A)** Estructura de la eucromatina y la heterocromatina. **(B)** Estructura en octámero que forma los nucleosomas. **(C)** Representación de un nucleosoma, formado por 147 pb de DNA envuelto alrededor del octámero de histonas. Figura creada con BioRender (<https://app.biorender.com/>).

Clásicamente, en función del grado de compactación del DNA en los nucleosomas, se han definido dos tipos de cromatina. La **eucromatina** corresponde a la cromatina menos compactada, y al tener una estructura más laxa, permite que el DNA sea más accesible a los FT y maquinaria de la transcripción<sup>168</sup>. Por otra parte, la **heterocromatina** corresponde a un estado más condensado de la cromatina, que se localiza sobre todo en la periferia del núcleo. Existen dos tipos de heterocromatina: la constitutiva y la facultativa. La heterocromatina constitutiva corresponde a regiones silenciadas, es decir, que no se transcriben debido al alto grado de compactación; son regiones iguales en todas las células, como por ejemplo los telómeros y los centrómeros. En cambio, la heterocromatina facultativa corresponde a regiones que se pueden encontrar condensadas o no dependiendo del tipo celular o del estado de la célula<sup>168</sup>.

### 3.4.2. Dinámica de la cromatina

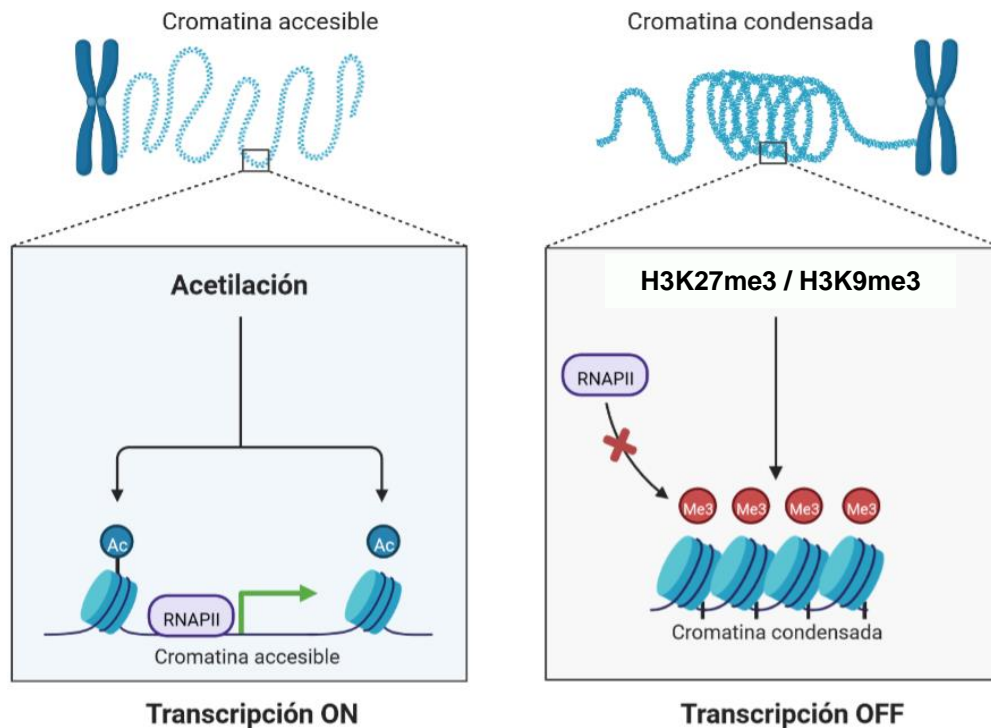
La cromatina es una estructura altamente dinámica modulada por múltiples modificaciones que influyen en su compactación, organización y función. Las modificaciones más estudiadas de la cromatina incluyen: modificaciones de histonas, remodelación de la cromatina por posicionamiento del nucleosoma y metilación del DNA. Es importante destacar que algunas de estas modificaciones afectan la condensación de la cromatina sin alterar la secuencia de nucleótidos del DNA y están involucradas en la regulación de varios procesos como la transcripción, la replicación o la reparación del DNA<sup>158</sup>.

#### A) Modificaciones de las histonas

Como se mencionó anteriormente, las colas N-terminales de las histonas están sujetas a múltiples modificaciones postraduccionales en distintos residuos<sup>158</sup>. Hasta la fecha, se han descrito más de 200 modificaciones diferentes de histonas<sup>170</sup>. Dependiendo del tipo de modificación y del residuo modificado, las modificaciones de histonas estarán involucradas en diferentes procesos como la regulación de la transcripción génica o la replicación del DNA. Las modificaciones de histonas pueden influir en la transcripción de genes mediante dos mecanismos diferentes.

El **primer mecanismo** está relacionado con la interacción de las histonas y el DNA dentro del nucleosoma, que se estabiliza por la diferencia de cargas entre las histonas (cargadas positivamente) y el DNA (cargado negativamente). Varias modificaciones de histonas (normalmente acetilaciones y fosforilaciones) pueden modificar la carga de las histonas, alterando así la interacción entre DNA e histonas. La condensación de cromatina, a su vez, influye en la accesibilidad del DNA a los FT y, en consecuencia, en los niveles de transcripción<sup>158</sup> (**Figura 19**). Además de los cambios en la carga, algunas modificaciones de histonas, como la ubiquitinación o la SUMOilación, pueden aumentar el *bulkiness* de las histonas. Este término se refiere al hecho de que la modificación química sea lo suficientemente grande como para generar distorsiones estructurales y cambios a gran escala en la dinámica de los nucleosomas a causa del cambio en el volumen total de las histonas<sup>171</sup>.

El **segundo mecanismo** implica el reconocimiento de modificaciones específicas de histonas por parte de proteínas efectoras que contienen dominios lectores. El reclutamiento de estas proteínas, que podrían agregar otras modificaciones de histonas o cambiar la posición del nucleosoma a través de complejos de remodelación, conduce a cambios en la transcripción de los genes<sup>172,173</sup>.



**Figura 19. Alteración en el estado de compactación de la cromatina por modificaciones de histonas.** La acetilación de histonas (como H3K27ac) da lugar a una descondensación de la cromatina que permite que la RNA pol II se una a un promotor para inducir la expresión del gen diana. Por el contrario, la trimetilación de las lisinas K9 y K27 de la histona H3 (H3K9me3 o H3K27me3), resultan en una compactación de la cromatina y reprimen la transcripción al evitar la unión de la RNA pol II. Figura creada con BioRender (<https://app.biorender.com/>).

La acetilación y la metilación son 2 de las modificaciones más estudiadas en las histonas. Sus características principales se describen a continuación:

- **Acetilación:** la acetilación neutraliza la carga positiva de las lisinas, provocando una disminución de la interacción entre la cola N-terminal de la histona y la carga negativa de los grupos fosfatos del DNA. Esta desestabilización del complejo DNA-histona induce a que la cromatina adquiera un estado menos compactado y, favoreciendo el acceso de los FT al DNA<sup>158</sup>. Por ello, la acetilación de lisinas de las histonas está asociada con la activación transcripcional y predomina principalmente en regiones eucromáticas. La acetilación de la lisina 27 de la histona 3 (H3K27ac) se asocia a *enhancers* activos y es una marca importante a la hora de identificar estos elementos reguladores<sup>202,203</sup>. Las enzimas que catalizan la reacción de acetilación están descritas en el apartado 3.3.2.2. de la Introducción, siendo las mismas que intervienen en la acetilación de las proteínas no-histonas.
- **Metilación:** los residuos susceptibles a metilación son los residuos de lisina y arginina. A diferencia de la acetilación y la fosforilación, la metilación no produce cambios en la carga de

los aminoácidos. Funcionalmente, la metilación puede promover tanto la activación como represión transcripcional, dependiendo del residuo modificado y el número de grupos metilo añadidos. Por ejemplo, la dimetilación y la trimetilación en H3K4 se asocia con una mayor expresión génica, mientras que la metilación en H3K9, H3K27 y H4K20 se asocia con represión transcripcional<sup>176</sup>. En concreto, H3K4me3 está asociada a promotores activos y es una marca importante a la hora de identificar estos elementos reguladores<sup>204</sup>. Las lisinas pueden ser mono-, di- o trimetiladas, mientras que las argininas pueden ser mono- o dimetiladas, y en este caso puede ser simétrica o asimétrica. Las enzimas responsables de la metilación de las histonas se clasifican en dos familias dependiendo del aminoácido diana: lisina metiltransferasa (KMT) y arginina metiltransferasa (PRMT).

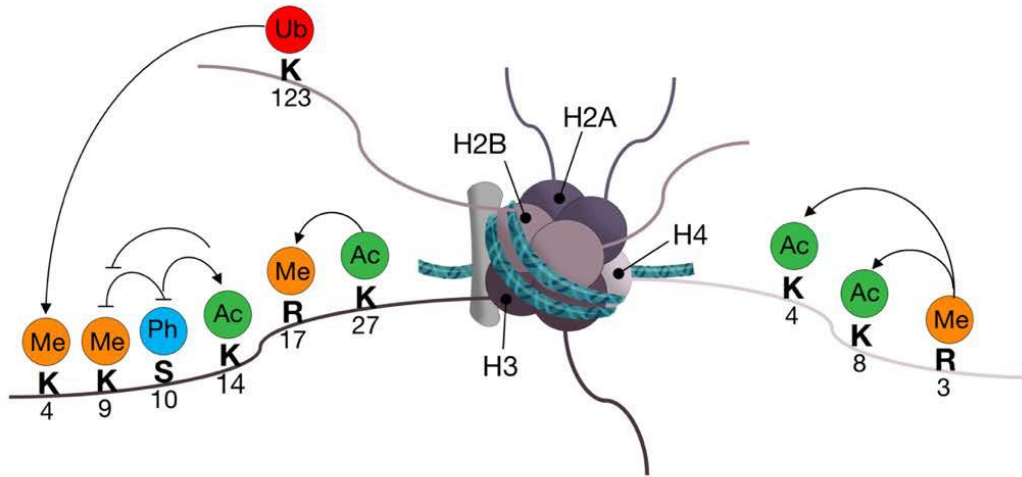
1. Las KMT transfieren un grupo metilo procedente de S-adenosilmetionina (SAM) al grupo amino de la lisina<sup>174</sup>. Se han identificado un gran número de KMT, la mayoría de las cuales comparten el dominio SET, cuya función enzimática es metiltransferasa.
2. Las PRMT transfieren un grupo metilo procedente de SAM en el grupo guanidinio de la arginina<sup>175</sup>. Las PRMT se clasifican en dos grupos: de tipo I (PRMT1-4, 6 y 8) y tipo II (PRMT5 y 7). Ambos tipos dan lugar a argininas mono- y dimetiladas, pero en el caso de las de tipo I la dimetilación es asimétrica, mientras que las de tipo II producen una dimetilación simétrica.

Durante mucho tiempo, la metilación era considerada como una modificación estática. Sin embargo, a partir de 2002 varios estudios demostraron que existían mecanismos de desmetilación de argininas y lisinas<sup>181</sup>. Estos grupos metilo pueden ser eliminados por histonas desmetilasas que se clasifican en dos familias: KDM1 y la familia de desmetilasas que contienen el dominio catalítico Jumonji C (JmJC)<sup>177</sup>.

La familia KDM1 está compuesta por dos miembros, KDM1A y KDM1B. KDM1A, también conocida como desmetilasa específica de lisina 1 (LSD1), fue descrita por primera vez por Shi *et al.* en 2004<sup>178</sup>. LSD1 desmetila las H3 H3K4 y H3K9 mono- y dimetiladas y, en cambio, es incapaz de actuar sobre lisinas trimetiladas<sup>179</sup>. El segundo y mayor grupo de histonas desmetilasas, descrito en 2006, son las que contienen el dominio JmJC<sup>180</sup>. Éstas son específicas de residuos trimetilados, por ejemplo, desmetilan H3K9me3 y H3K36me3<sup>181</sup>.

Existe la hipótesis de que el nucleosoma podría ser considerado como un todo y sería la **combinación de las diferentes modificaciones** postraduccionales en las histonas de un nucleosoma las que regularían la expresión de la propia secuencia de DNA dentro del nucleosoma<sup>182,158,184</sup>. En este sentido, la hipótesis del “código de histonas”, descrita por Strahl & Allis en el año 2000, proponía que combinaciones específicas de estas modificaciones dictan la actividad transcripcional de un determinado promotor<sup>182</sup>. Concretamente, los autores

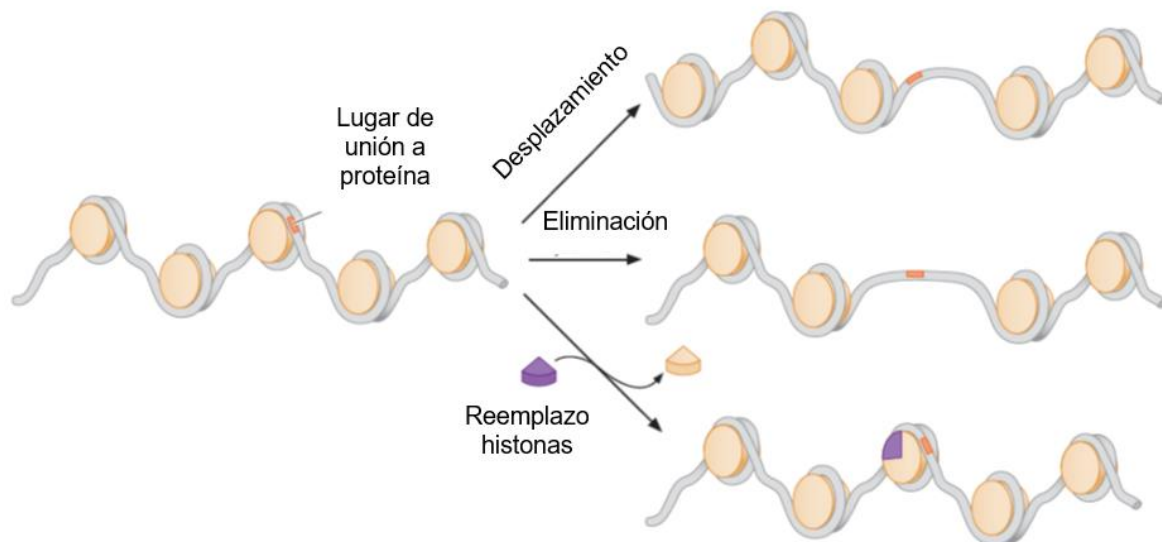
propusieron que la presencia de ciertas modificaciones pueda influir en la generación o no de otras modificaciones, creando patrones específicos de modificaciones de histonas (**Figura 20**)<sup>158,184</sup>.



**Figura 20. Coordinación entre las diferentes modificaciones de histonas.** La influencia positiva de una modificación sobre otra se muestra con una flecha y el efecto negativo con una línea. Figura adaptada de M. Pinsach<sup>8</sup> & Kouzarides *et al.*<sup>158</sup>.

## B) Remodelación de nucleosomas por complejos remodeladores de cromatina

Los complejos de remodelación de cromatina o remodeladores están compuestos por unas 10-15 subunidades, algunas de las cuales contienen dominios lectores de modificaciones de histonas que permiten a los remodeladores unirse a regiones específicas de la cromatina. Una vez reclutados, los remodeladores de la cromatina pueden inducir cambios en la posición de los nucleosomas que aumentan la accesibilidad para los FT, lo que facilita la transcripción de los genes. Sin embargo, la remodelación del nucleosoma también puede reducir la accesibilidad a los FT, bloqueando la transcripción. La posición de los nucleosomas se puede modificar mediante cinco mecanismos diferentes: desplazamiento de nucleosomas, eliminación de nucleosomas, ensamblaje de nucleosomas, espaciado de nucleosomas y reemplazo de histonas<sup>183</sup> (**Figura 21**).



**Figura 21. Modificación de nucleosomas mediante complejos remodeladores.** Se muestran 3 mecanismos por los cuales los complejos alteran los nucleosomas: desplazamiento, eliminación y reemplazo de histonas. Figura creada con BioRender (<https://app.biorender.com/>).

### C) Metilación del DNA

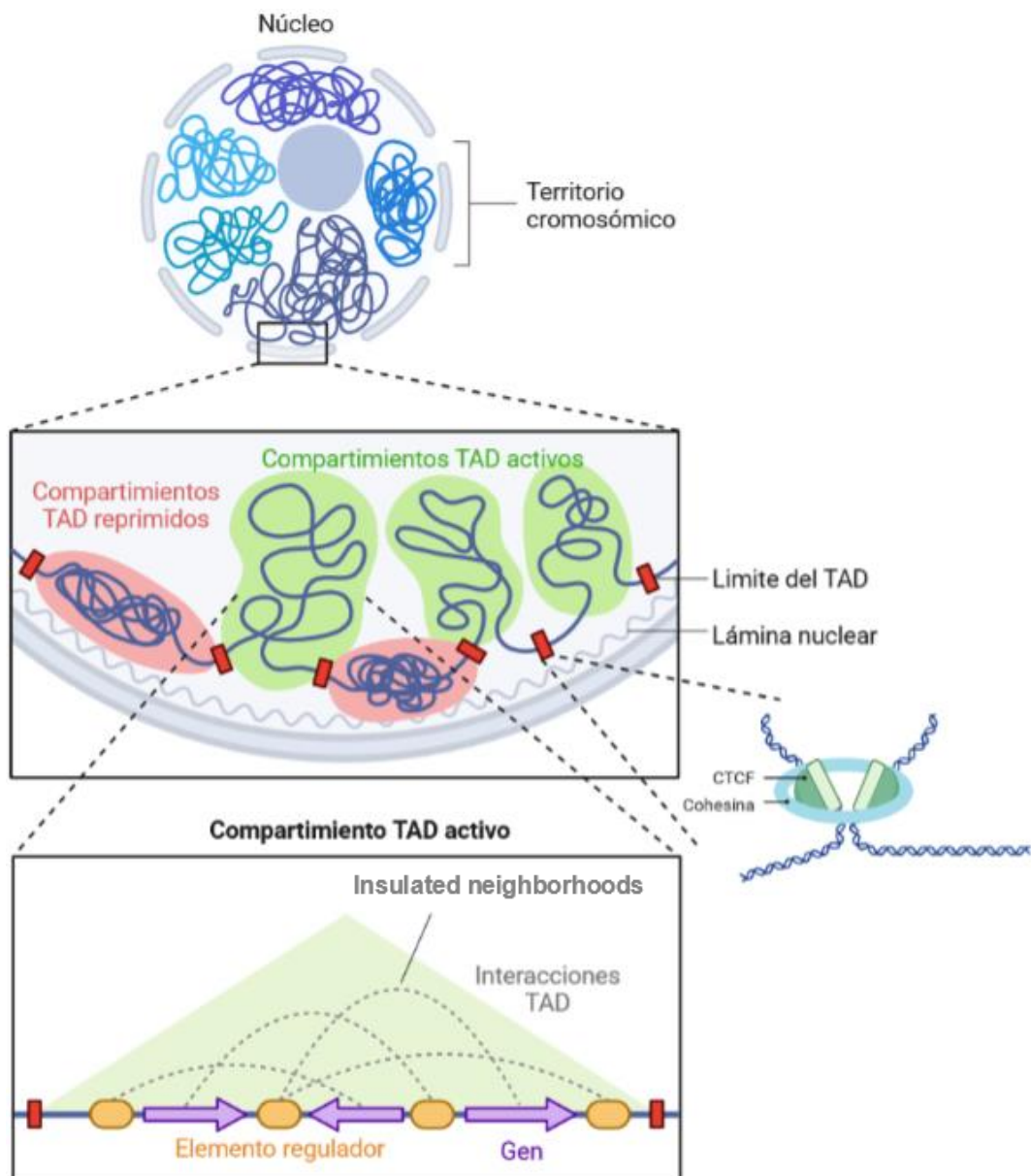
El DNA es metilado por la adición de grupos metilo unidos covalentemente en la posición 5' de las citosinas del DNA. La metilación del DNA es catalizada por las DNA metiltransferasas tipo I y II. Aunque la metilación del DNA no es tan dinámica como las modificaciones de las histonas, es también una reacción reversible, y los grupos metilo pueden ser eliminados por las DNA desmetilasas<sup>185</sup>. La metilación del DNA ocurre principalmente en las islas CpG presentes en el genoma, que son secuencias cortas de DNA intercaladas y enriquecidas en nucleótidos GC<sup>186</sup>. Como se mencionó anteriormente, alrededor del 60% de los promotores humanos contienen islas CpG que están relacionadas con la regulación transcripcional. Los grupos metilo interfieren con la accesibilidad del DNA con los FT y, por lo tanto, las islas CpG metiladas se han asociado con la represión transcripcional, mientras que las islas CpG desmetiladas se han asociado con la activación transcripcional<sup>187-188</sup>.

En resumen, la dinámica de las modificaciones de las histonas, la remodelación de los nucleosomas y la metilación del DNA permite que la célula responda rápidamente a diversos estímulos, modificando la accesibilidad de los FT al DNA e induciendo cambios transcripcionales en el equilibrio activación/represión de los genes.



### 3.4.3. Organización tridimensional de la cromatina

La regulación del genoma humano no solo depende de la secuencia lineal del genoma que incorpora millones de elementos reguladores en *cis*, sino también de la arquitectura tridimensional de la cromatina que orquesta la interacción entre los elementos reguladores en *cis* y sus genes diana<sup>189</sup>. Los diferentes niveles de organización de la cromatina (**Figura 22**) se describen a continuación, de menor a mayor nivel de organización:



**Figura 22. Organización jerárquica del genoma en mamíferos.** De gran a pequeña escala (de arriba a abajo), se muestran territorios cromosómicos, compartimientos TAD e *insulated neighborhoods*. Arriba. Cada territorio cromosómico se denota con diferentes colores. Centro. Se muestran compartimientos TAD tanto activos como reprimidos. En detalle, a la derecha, se muestra cómo se forman los límites de los TAD. Abajo. Se muestran interacciones de largo alcance entre elementos reguladores dentro de un mismo TAD. Figura creada con BioRender (<https://app.biorender.com/>).

### 3.4.3.1. *Insulated neighborhoods*

Como se explicó antes, en un DNA lineal, los elementos reguladores en *cis* se pueden encontrar a varios cientos o miles de pb de distancia de los genes que regulan. De acuerdo con la naturaleza polimérica de las fibras de cromatina, dos loci genómicos distantes contactarían entre sí a muy baja frecuencia debido a una colisión aleatoria<sup>191</sup>. Sin embargo, se ha observado que ciertos loci tienen una frecuencia de contacto significativamente mayor que la esperada por casualidad. Estos contactos de cromatina, conocidos como ***insulated neighborhoods***, son específicos del tipo de célula y están asociados con la formación de bucles de DNA para permitir interacciones *enhancer*-promotor<sup>192,193</sup> (**Figura 22**).

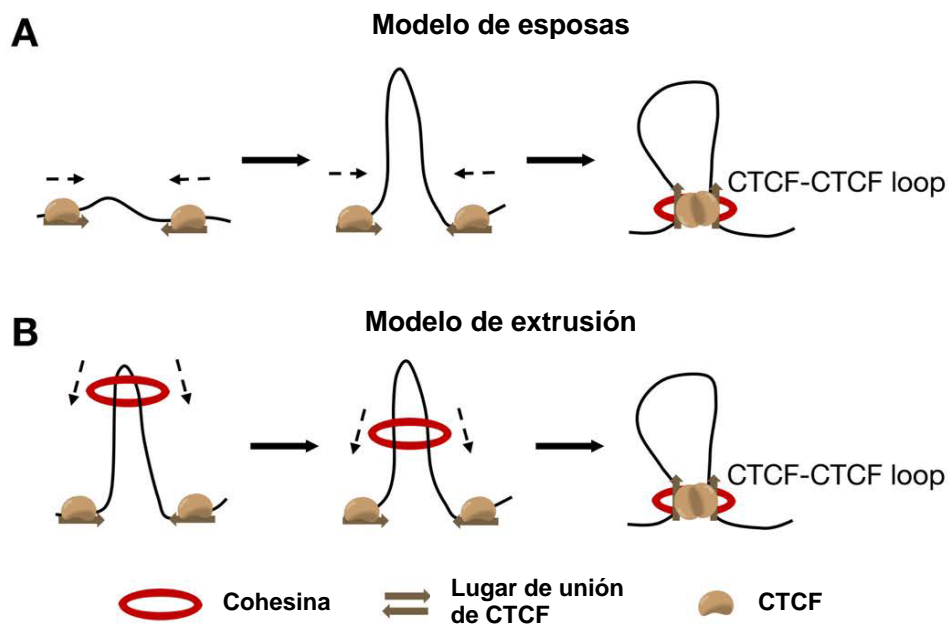
### 3.4.3.2. Dominios topológicos asociados

Los *insulated neighborhoods* se organizan aún más en estructuras de orden superior denominadas dominios topológicos asociados<sup>109</sup> (TAD). Los **TAD** representan unidades de organización del genoma físicamente aisladas que pueden considerarse funcionalmente separadas del resto del genoma por dos razones principales (**Figura 22**). Primero, la frecuencia de contacto entre loci genómicos en el mismo TAD es varias veces mayor que la frecuencia de contacto de loci pertenecientes a diferentes TADs. En segundo lugar, los TADs tienen una propiedad de aislamiento que bloquea la propagación de la actividad de los elementos reguladores en *cis* o la propagación de la cromatina represora a las regiones ubicadas en los TAD vecinos<sup>99</sup>.

Los límites entre TADs adyacentes están definidos por aisladores o ***insulators***. Las regiones límite entre TADs tienen sitios de unión específicos para estos elementos *insulators*. El factor de unión CCCTC (CTCF) es un FT con 11 dedos de zinc involucrado en la formación de las organizaciones de orden superior en el genoma humano, así como en la formación de barreras de cromatina para protegerse de la heterocromatina<sup>194</sup>. A pesar de la evidencia de que alrededor del 80% de los límites entre TADs están unidos por CTCF, el mecanismo exacto por el cual CTCF aísla las interacciones de cromatina entre ellos ha sido una cuestión de intensa investigación.

Actualmente, se han propuesto dos modelos diferentes para explicar la formación de los límites TAD: el “modelo de esposas” y el “modelo de extrusión”. El modelo de esposas postula que los dos extremos de un TAD están unidos por proteínas CTCF que interactúan entre sí para reclutar el complejo cohesina y estabilizar el *loop*<sup>195</sup> (**Figura 23A**). La cohesina es un complejo proteico de múltiples subunidades que está altamente conservado en eucariotas y desempeña funciones esenciales en varios procesos biológicos, incluidos: (i) la segregación cromosómica en células en división, (ii) la reparación del DNA y (iii) el establecimiento de límites TAD<sup>196</sup>. Alternativamente, el modelo de extrusión propone que un complejo motor de cromatina (como el complejo cohesina) se acopla al DNA y luego extruye un bucle progresivamente más grande

hasta que se detiene por la unión del CTCF en orientación convergente, configurando así los TADs<sup>197</sup> (Figura 23B).



**Figura 23. Modelos propuestos para la formación de TADs.** (A) El modelo de esposas propone la formación de TADs mediante CTCF y el complejo cohesina que conecta las dos secuencias. (B) El modelo de extrusión implica un par de proteínas CTCF unidas a motores de cromatina que impulsan la extrusión de la fibra de cromatina, mientras que las dos moléculas de CTCF deslizan la fibra de cromatina en direcciones opuestas antes de detenerse en los lugares de unión a CTCF. Figura adaptada de M. Pinsach<sup>8</sup>.

A diferencia de los *insulated neighborhoods*, los TADs son relativamente estables en todos los tipos de células y parecen ser independientes de la expresión génica específica del tejido o de las modificaciones de las histonas. En este sentido, se ha demostrado que el posicionamiento de los TADs se conserva evolutivamente (50-70% de los límites TAD se comparten entre células embrionarias de humanos y de ratón). Este dato ha llevado a considerar a los TAD como la unidad fundamental de organización del genoma en diferentes especies<sup>198</sup>.

### 3.4.3.3. Compartimientos celulares

Los TADs se organizan además en dos compartimentos diferentes (compartimento A y B), separados en el espacio y organizados de manera espacialmente polarizada. La partición de los compartimentos A y B es específica del tipo de célula y se han observado cambios entre estos dos compartimentos en aproximadamente el 60 % del genoma humano<sup>199</sup>. Las secuencias del compartimento A son de replicación temprana, contienen una alta densidad de genes, exhiben fuertes actividades de expresión de mRNA y están enriquecidas para H3K36me3<sup>191</sup>.

Por el contrario, las secuencias del compartimento B se replican tarde, contienen una baja densidad de genes (que están transcripcionalmente reprimidos o se expresan a bajos niveles) y están enriquecidas para las marcas típicas de histonas de heterocromatina H3K9me2 y H3K9me3<sup>200,201</sup>.

#### 3.4.3.4. Territorios cromosómicos

El mayor nivel de organización tridimensional del genoma humano son los territorios cromosómicos, que son regiones específicas del núcleo ocupadas por cromosomas específicos<sup>205</sup> (**Figura 22**). Los cromosomas pequeños ricos en genes generalmente se ubican cerca del centro del núcleo, mientras que los cromosomas más grandes y pobres en genes se ubican cerca de la periferia nuclear.

En resumen, la regulación de la expresión génica a nivel transcripcional es un proceso combinatorio dinámico que involucra una variedad de elementos y mecanismos que pueden operar solo en tipos de células particulares, en una etapa determinada del desarrollo o en respuesta a factores ambientales. Por lo tanto, cualquier alteración en estos mecanismos puede tener consecuencias funcionales que implican cambios en la expresión génica, que finalmente pueden traducirse en fenotipos de enfermedad.

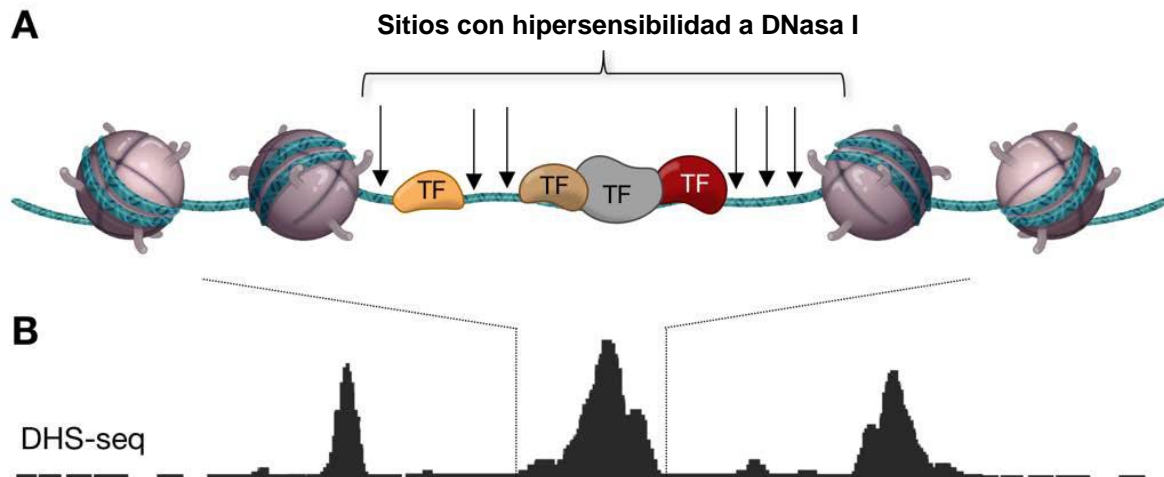
### 3.5. Identificación de elementos reguladores de la transcripción

#### 3.5.1. Análisis de los lugares de unión a factores de transcripción

La unión de los FT a los elementos reguladores en *cis* requiere que el DNA sea accesible dentro de la estructura de la cromatina. Esto se logra mediante cambios en la condensación de la cromatina y el desplazamiento de los nucleosomas. Las regiones de DNA expuestas se denominan regiones de cromatina abierta y se han asociado a regiones reguladoras transcripcionalmente activas del genoma<sup>206,207</sup>. El estudio e identificación de estas regiones se puede llevar a cabo mediante diferentes técnicas:

- A) ***Dnase hypersensitive sites (DHS-seq)***: la exposición al DNA facilita la unión de los FT, pero también hace que el DNA sea más sensible a la escisión enzimática. Esta propiedad se ha utilizado durante mucho tiempo para identificar regiones de cromatina abierta mediante la digestión de la cromatina con la nucleasa Dnasa I. La Dnasa I puede digerir regiones de cromatina abiertas, aunque las regiones abiertas con FT unidos tendrán una leve protección contra la digestión (**Figura 24**). Actualmente, la digestión con Dnasa I suele ir acompañada de una secuenciación masiva del material digerido (DHS-seq), lo que permite la identificación de todas las regiones de cromatina abierta que se encuentran en el genoma

para un tipo de célula o tejido específico<sup>208,209</sup>. DHS-seq identifica regiones reguladoras transcripcionalmente activas, que señalan indirectamente las regiones de unión a FT. Se han desarrollado técnicas más novedosas como **ATAC-seq** (*Assay for Transposase-Accessible Chromatin using sequencing*), que también sirve para evaluar el genoma de toda la accesibilidad de la cromatina de manera más rápida y sensible en comparación con DHS-seq<sup>212</sup>.



**Figura 24. Regiones de cromatina abierta.** (A) Representación esquemática de los sitios de hipersensibilidad a la DNasa I con FT unidos. Las flechas negras indican regiones susceptibles a la digestión con DNasa I. (B) Descripción general de las señales en forma de pico obtenidas después de procesar los datos de DHS-seq. Figura adaptada de M. Pinsach<sup>8</sup> y Ecker *et al.*<sup>214</sup>.

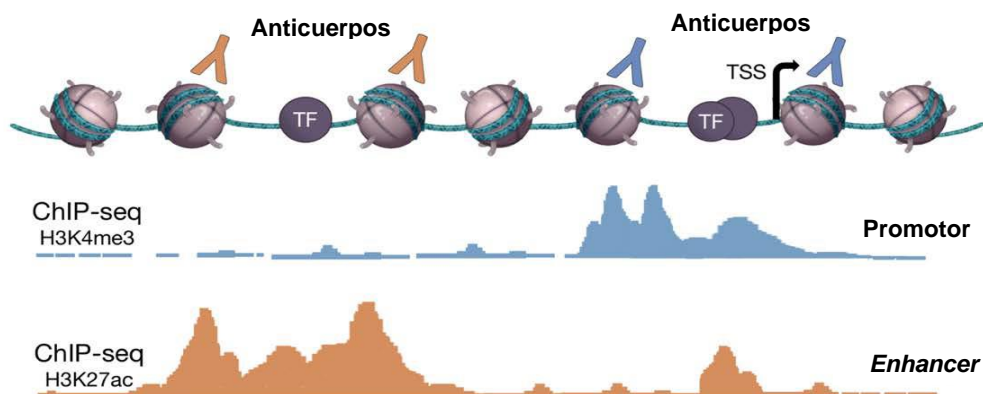
- B) **Algoritmos bioinformáticos:** la técnica DHS-seq, no obstante, no tiene suficiente resolución para detectar qué FT se une o el sitio de unión exacto que se está utilizando. Por lo tanto, la combinación de DHS-seq o ATAC-seq con algoritmos computacionales que realizan análisis de motivos de secuencia como HOMER<sup>210</sup> o JASPAR<sup>211</sup> han demostrado ser muy útiles para identificar motivos de unión de FT (TFBS) en las secuencias probadas.
- C) **ChIP-seq:** si el interés experimental es el perfil completo de unión de un FT en un tipo de célula concreto, el enfoque más utilizado es la inmunoprecipitación de cromatina seguida de secuenciación paralela masiva (ChIP-seq)<sup>213</sup>. De manera resumida, la técnica de ChIP-seq se basa en la fijación de las interacciones proteínas-DNA con formaldehído seguido de una fragmentación de la cromatina. La cromatina fragmentada se incuba con anticuerpos contra el FT de interés para inmunoprecipitar solo aquellas regiones genómicas que están unidas por ese FT. Luego, el material inmunoprecipitado se secuencía y, después del análisis bioinformático posterior, los perfiles de unión del FT distribuidos a lo largo del

genoma se pueden visualizar como picos correspondientes a la cantidad de fragmentos secuenciados en una posición genómica en concreto<sup>214</sup>.

### 3.5.2. Análisis del patrón de modificaciones de histonas

Tal y como se mencionó anteriormente, uno de los mecanismos que influyen en la accesibilidad del DNA a los FT es la presencia de modificaciones de histonas que alteran las interacciones DNA-histonas dentro del nucleosoma<sup>203</sup>.

Estas modificaciones se distribuyen por todo el genoma y se ha observado que patrones particulares de modificaciones de histonas pueden estar asociados a diferentes tipos de elementos reguladores de la transcripción (apartado 3.4.2. de la Introducción). Los patrones de modificaciones de histonas también pueden detectarse mediante ChIP-seq usando anticuerpos contra la modificación de histonas que se pretenda estudiar, y los resultados también se representan como picos correspondientes a la cantidad de regiones secuenciadas en una posición genómica particular<sup>204</sup> (**Figura 25**).

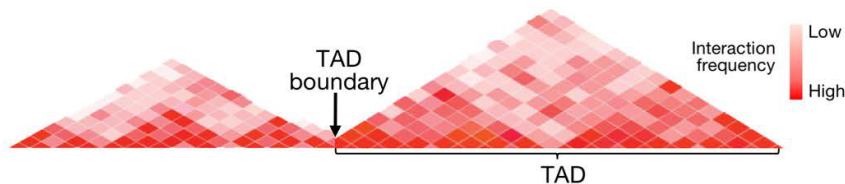


**Figura 25. Perfil de modificaciones de histonas por ChIP-seq.** Arriba: Los anticuerpos que reconocen H3K4me3 se utilizan para identificar regiones promotoras activas, mientras que los anticuerpos que reconocen H3K27ac se utilizan para identificar *enhancers* activos. Abajo: perfiles de H3K4me3 y H3K27ac después de procesar los datos de ChIP-seq. Las zonas con mayores picos representan más concentración de la marca de histona. Figura adaptada de M. Pinsach<sup>8</sup> y Ecker *et al.*<sup>214</sup>.

### 3.5.3. Análisis de la organización tridimensional del genoma

Los elementos reguladores de la transcripción también se pueden identificar mediante técnicas de captura de cromatina de alto rendimiento, también conocidas como tecnologías C. Estas técnicas permiten la identificación de interacciones de cromatina correspondientes a bucles de DNA entre elementos reguladores en *cis*, pero también son útiles para descubrir características generales de la organización del genoma, como los compartimentos celulares y los TAD<sup>215</sup>. Por lo tanto, dependiendo del propósito experimental, existen varias tecnologías C

disponibles. Una técnica muy utilizada, el 4C, se aplica para encontrar todas las regiones del genoma que interactúan con una determinada región de interés<sup>216</sup>. Por el contrario, el Hi-C se aplica para obtener los perfiles de interacción de todas las regiones en todo el genoma, y sus interacciones con todas las demás regiones genómicas<sup>191</sup>. Cuando la evaluación de los contactos de cromatina se realiza a nivel de todo el genoma, como en los experimentos Hi-C, los datos se representan en forma de mapas de contacto de cromatina donde la intensidad del color está relacionada con la frecuencia de interacción (**Figura 26**). Los mapas de contacto de cromatina se pueden complementar aún más con información de ChIP-seq para CTCF, para identificar los límites de TAD; o bien con información de DHS o modificaciones de histonas, en este caso para identificar regiones activas e inactivas dentro de TAD.



**Figura 26. Análisis de la organización del genoma.** Representación esquemática de los mapas de contacto de Hi-C que indican las frecuencias de las interacciones entre las regiones genómicas. Figura adaptada de M. Pinsach<sup>8</sup>.

Se han desarrollado otras técnicas más específicas como por ejemplo PChI-C (*Promoter Capture Hi-C*). Esta técnica, basada en el Hi-C, aumenta en gran medida la potencia para detectar interacciones que involucran secuencias promotoras<sup>217,218</sup>. De manera similar, ChIA-PET es una técnica que combina la inmunoprecipitación de cromatina (ChIP) con el análisis de interacciones de larga distancia. ChIA-PET se puede utilizar para determinar la contribución de una proteína en concreto en la formación de redes de contacto con otras secuencias proximales o distales, como promotores de los genes con los que interactúa<sup>219</sup>.

En resumen, los métodos mencionados anteriormente permiten identificar elementos reguladores de la transcripción y sus interacciones. Su utilización y combinación entre sí dependerá de los fines experimentales.

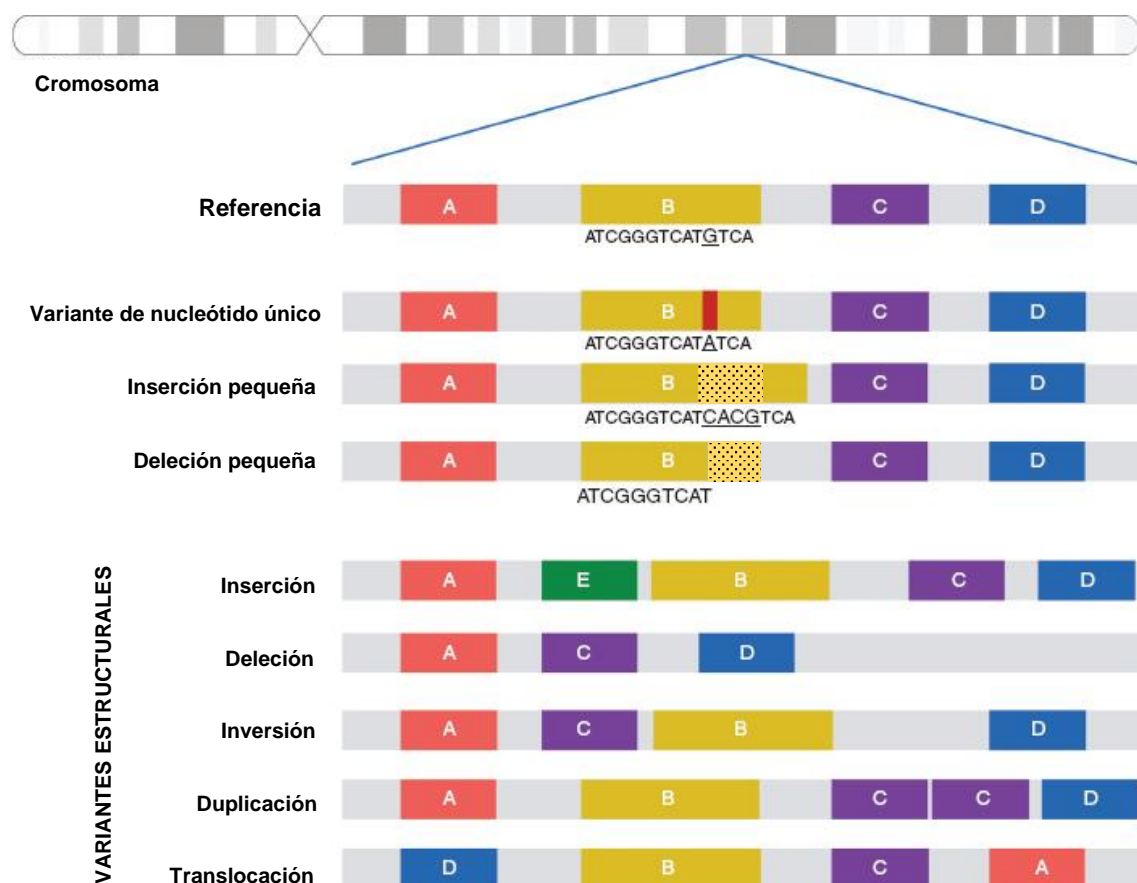
## 4. La variación genética

### 4.1. Tipos de variación genética

La variación genética se refiere a las diferencias en la secuencia de DNA que se encuentran en el genoma de los organismos. Estas variantes genéticas se distribuyen aleatoriamente por todo el genoma y, en el caso de los genomas diploides, se pueden encontrar en forma de heterocigotos (si solo afectan a uno de los dos alelos), o bien homocigotos (si afectan a ambos alelos)<sup>221</sup>.

En términos generales, las variantes genéticas se clasifican en tres grupos principales según la extensión de los nucleótidos que se ven afectados:

- **Variantes de un único nucleótido (SNV):** las SNVs son el tipo más común de variante genética observada en el genoma<sup>222</sup>. Estas variantes tienen lugar cuando un solo nucleótido es sustituido por otro en una posición específica en la secuencia de DNA (**Figura 27**).



**Figura 27. Representación de diferentes tipos de variantes (SNVs, indels y estructurales) en comparación con el genoma de referencia.** Cada cuadro diferente representa un elemento genómico diferente. Figura adaptada de E. Ruzzo *et al.*<sup>227</sup>.



- **Pequeñas inserciones y deleciones (indels):** los indels consisten en una inserción o deleción de uno o más nucleótidos de DNA en el genoma (**Figura 27**). Representan el segundo tipo más común de variación genética, aunque la frecuencia estimada en el genoma humano varía de un estudio a otro<sup>223</sup>.
- **Variantes estructurales (SVs):** el término SV se usa para describir la variación genética que tiene lugar en una secuencia larga de DNA, desde 50 pb hasta miles de pb. Son menos frecuentes que las SNV o los indels pero afectan a un mayor número de nucleótidos<sup>224</sup>. Los SVs se pueden encontrar como cambios en el número de copias (deleciones, inserciones y duplicaciones), orientación (inversiones) o ubicación cromosómica (translocaciones) entre individuos<sup>225</sup> (**Figura 27**).

## 4.2. Fuentes de variación genética

Las variantes genéticas pueden ser originadas por: (i) fuentes exógenas como factores ambientales o (ii) fuentes endógenas como productos del metabolismo celular, errores en la replicación del DNA y fallas en los mecanismos de reparación del DNA<sup>221</sup>:

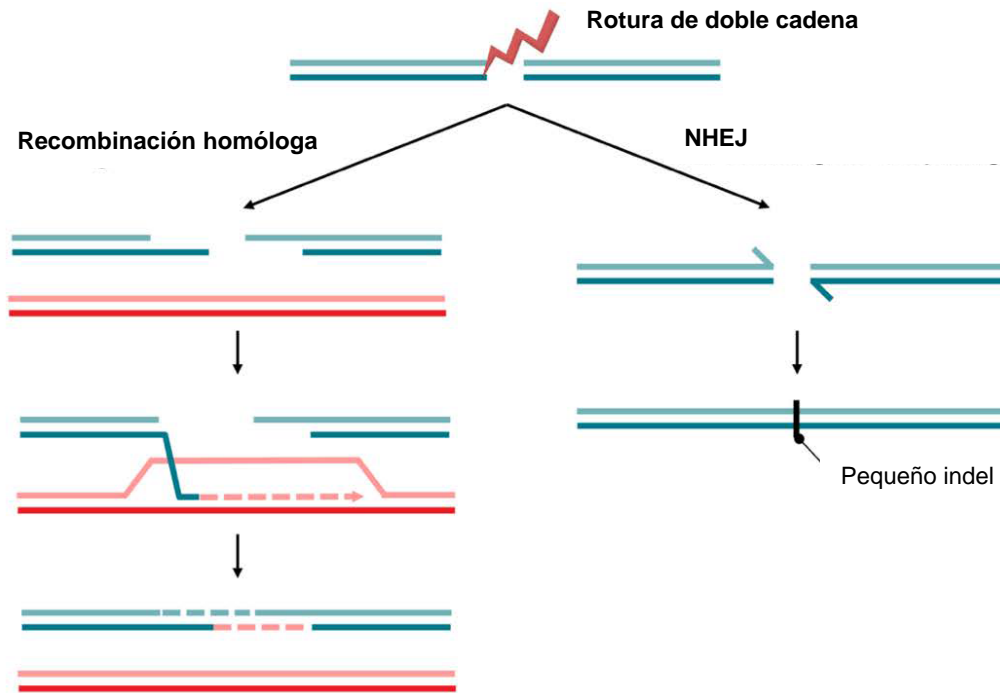
- **Factores ambientales:** a grandes rasgos, existen dos grandes grupos de factores ambientales que pueden generar variantes genéticas. Estos incluyen agentes químicos y radiación. En el caso de los productos químicos, existe una cantidad casi infinita de sustancias químicas que pueden alterar la secuencia de DNA a través de diferentes mecanismos<sup>226</sup>. El daño del DNA inducido por radiación puede ser de dos tipos: daño causado por la radiación ultravioleta (luz UV) y daño causado por la radiación ionizante.
- **Metabolismo celular:** las variantes genéticas también pueden generarse por hidrólisis, exposición a especies de oxígeno reactivo (ROS) y otros metabolitos reactivos que se producen en el metabolismo de rutina de la célula<sup>221</sup>. Estos mecanismos dañinos por los que se introducirán variantes genéticas son diferentes según el tipo de metabolito generado.
- **Errores de replicación del DNA:** en las células humanas, las DNA polimerasas involucradas en la replicación son muy precisas, con una tasa de error de  $10^{-6}$  a  $10^{-7}$  errores por pb. Esta tasa de error se ve disminuida aún más por los mecanismos de reparación del DNA, que reducen la tasa neta a aproximadamente  $10^{-10}$  errores por pb<sup>221</sup>. Sin embargo, incluso con este alto nivel de fidelidad en la replicación, algunos errores en la replicación del DNA se incorporan en la secuencia final de DNA, lo que origina variantes genéticas. En el caso de las SNVs, los errores de replicación se producen por la adición de un par de bases no complementarias por parte de la DNA polimerasa. Por el contrario, los indels pueden

originarse por dos errores de replicación diferentes: (i) deslizamiento de la polimerasa y (ii) formación de estructuras secundarias durante la replicación<sup>228</sup>.

- **Mecanismos de reparación del DNA:** en las células eucariotas, los mecanismos de reparación del DNA se activan para corregir alteraciones en la secuencia de DNA y mantener su integridad. Estos mecanismos de reparación se incluyen como fuente de variación genética porque, si bien buscan restaurar la secuencia de DNA cuando se introduce una alteración, su actividad también puede resultar en la inserción de variantes genéticas. Los mecanismos de reparación del DNA pueden clasificarse en dos grupos principales. El primer grupo incluye todos aquellos mecanismos destinados a corregir errores de replicación del DNA (corrección de errores o *proofreading* y el sistema de reparación del DNA por apareamiento erróneo o *mismatch repair*). El segundo grupo incluye todos aquellos mecanismos destinados a corregir el daño en el DNA inducido por factores ambientales o metabolitos celulares (reparación por escisión de bases, reparación por escisión de nucleótidos y reparación por rotura de la doble cadena)<sup>229</sup>. De estos mecanismos mencionados, solamente se va a describir la reparación por rotura de la doble cadena, ya que presenta un mayor interés para esta tesis debido a su implicación en los procesos de edición por *Clustered Regulatory Interspaced Short Palindromic Repeats* (CRISPR):

Reparación por rotura de la doble cadena: algunos tipos de factores ambientales, como la radiación de alta energía, pueden causar roturas de doble cadena en el DNA. Las roturas de doble cadena son peligrosas porque grandes segmentos de cromosomas y los genes que contienen pueden perderse si no se repara la rotura. Hay dos vías diferentes involucradas en la reparación de roturas de DNA de doble cadena: la unión de extremos no homólogos (*non-homologous DNA end joining* o **NHEJ**) y la **recombinación homóloga (Figura 28)**.

En la recombinación homóloga, el DNA que falta es copiado a partir de la secuencia homóloga, dando como resultado una secuencia idéntica a la original<sup>230</sup>. Este último mecanismo puede aprovecharse en las técnicas de edición genética, introduciendo una secuencia homóloga artificial con el cambio nucleotídico que se busca en la edición. De esta manera, habrá eventos de reparación donde la maquinaria celular use el fragmento artificial para reparar la rotura (reparación por **HDR** o *homologous directed repair*)<sup>231</sup>. Por el contrario, en la NHEJ, los dos extremos rotos se vuelven a unir directamente. Este mecanismo evita la pérdida de material genético, pero normalmente implica la pérdida o adición de algunos nucleótidos en el punto de ruptura, lo que da lugar a pequeños indels<sup>230</sup>.



**Figura 28. Mecanismos de reparación del DNA para corregir roturas de doble cadena.** Representación de los dos mecanismos que pueden activarse para reparar roturas bicatenarias producidas por radiaciones ionizantes: la recombinación homóloga (izquierda) y la unión de extremos no homólogos o NHEJ (derecha). Figura adaptada de M. Pinsach<sup>8</sup>.

- **Elementos móviles del DNA:** los elementos móviles dispersos en el genoma (transposones y retrotransposones) se consideran una fuente importante de SV, pero también pueden contribuir a la formación de pequeños indels. Estos pequeños indels se forman porque, cuando los elementos móviles se separan de su posición genómica, dejan una estructura similar a la de las roturas de doble cadena. Por lo tanto, el mecanismo de reparación de roturas de doble cadena se activará y, si la rotura se repara mediante la unión de extremos no homólogos, puede dar lugar a pequeños indels<sup>88</sup>.
- **Recombinación meiótica:** En organismos diploides como los humanos, en los que las células tienen dos copias de cada cromosoma, las células germinales se dividen mediante el proceso de la meiosis para dar lugar a los gametos haploides. Antes de que comience la meiosis, durante la fase S del ciclo celular, se replica el DNA de cada cromosoma. Inmediatamente después de la replicación del DNA, las células meióticas entran en una etapa conocida como profase meiótica I. Durante este período, los cromosomas homólogos se aparean entre sí y se someten a una recombinación genética, un proceso denominado *crossover*, en el que se produce un intercambio de información del DNA entre los dos cromosomas homólogos<sup>75</sup>.

En general, la recombinación meiótica no implica la introducción de SNVs o indels; sin embargo, cuando la recombinación meiótica involucra cromosomas homólogos desalineados, pueden aparecer pequeños indels<sup>228</sup>.

### 4.3. El rol de la variación genética en las enfermedades humanas

Las enfermedades humanas pueden ser causadas por variantes genéticas que abarcan la gama completa de tipos de variantes, desde SNVs hasta SVs. Estas variantes incluyen un amplio espectro de frecuencias, desde las variantes *de novo* muy raras hasta las comunes<sup>232</sup>.

Las **variantes comunes** se refieren a aquellas variantes que se encuentran con una frecuencia alta en una población ( $AF > 1\%$ ). Algunos autores han propuesto que tienen su origen en variantes antiguas surgidas en un ancestro común y que han sido seleccionadas hasta alcanzar una frecuencia significativa entre la población<sup>233</sup>.

Por el contrario, las **variantes raras** se refieren a aquellas variantes que se encuentran en una frecuencia relativamente baja en una población ( $MAF < 1\%$ ). Estas se han asociado generalmente a variantes recientes en las que la selección no ha tenido tiempo de actuar, o bien a variantes en las que la selección está intentando eliminarlas por su carácter deletéreo<sup>233</sup>.

Cuando se estudian enfermedades genéticas humanas, los estudios de asociación del genoma completo (GWAS) proporcionan información muy valiosa sobre el potencial papel de la variación común en patologías humanas. Los GWAS tienen como objetivo identificar variantes asociadas a una enfermedad o rasgo específico mediante el estudio de las frecuencias alélicas de estas variantes comunes en individuos enfermos, en relación a la encontrada en individuos sanos. Hasta la fecha, se han realizado más de 5 700 GWAS para más de 3 300 rasgos distintos<sup>234</sup>, superando el millón de individuos estudiados<sup>234</sup>. Los resultados de estos estudios han permitido identificar una enorme cantidad de variantes genéticas asociadas a esos rasgos o enfermedades. Sin embargo, incluso con esta información obtenida de GWAS, seguimos conociendo solo una pequeña parte de la variación genética asociada a las enfermedades humanas<sup>235</sup>. Estas observaciones podrían estar relacionadas con el hecho de que los GWAS se centran en la detección de variantes comunes, lo que sugiere que parte de esta heredabilidad faltante podría dilucidarse, en cierta medida, por variantes raras<sup>235</sup>.

Sin embargo, la detección de variantes raras ha sido especialmente complicado dado que, para determinar adecuadamente la frecuencia y la distribución en la población de estas variantes genéticas, es necesario generar datos de secuenciación de alta precisión a partir de muchas muestras<sup>232</sup>. Por ejemplo, el proyecto HapMap<sup>236,237</sup> que proporcionó un estudio inicial de SNVs en las principales poblaciones humanas, solo catalogó una fracción de la variación genética por encima de una AF de 0.05. Incluso los estudios piloto del Proyecto 1 000 Genomas solo pudieron capturar exhaustivamente la variación en más de una AF de 0.01<sup>238</sup>.

A medida que estos y otros proyectos han ido aumentando su tamaño muestral y se han secuenciado más genomas, se hace evidente que la variación genética entre individuos es mayor de lo que se esperaba<sup>239-242</sup>. Como ejemplo, la secuenciación de más de 2500 individuos de 5 poblaciones diferentes realizada por el Proyecto 1000 Genomas reveló que, en comparación con el genoma de referencia, cada genoma individual contiene una media de 3.5 millones de SNVs y > 2100 SVs. Además, muchas de estas variantes parecen ser raras en la población de la que se tomó la muestra del individuo, y la tasa de identificación de variantes que no han sido descritas previamente sigue creciendo con cada nuevo individuo secuenciado<sup>243</sup>.

En global, la alta proporción de variantes presentes en el genoma y el hecho de que algunas de las variantes son particulares para cada genoma individual, sugiere que el aspecto más importante cuando se estudian enfermedades genéticas humanas es no centrarse de manera excesiva en variantes específicas, sino más bien **integrar todas las clases de variantes asociadas al riesgo** (Figura 29).

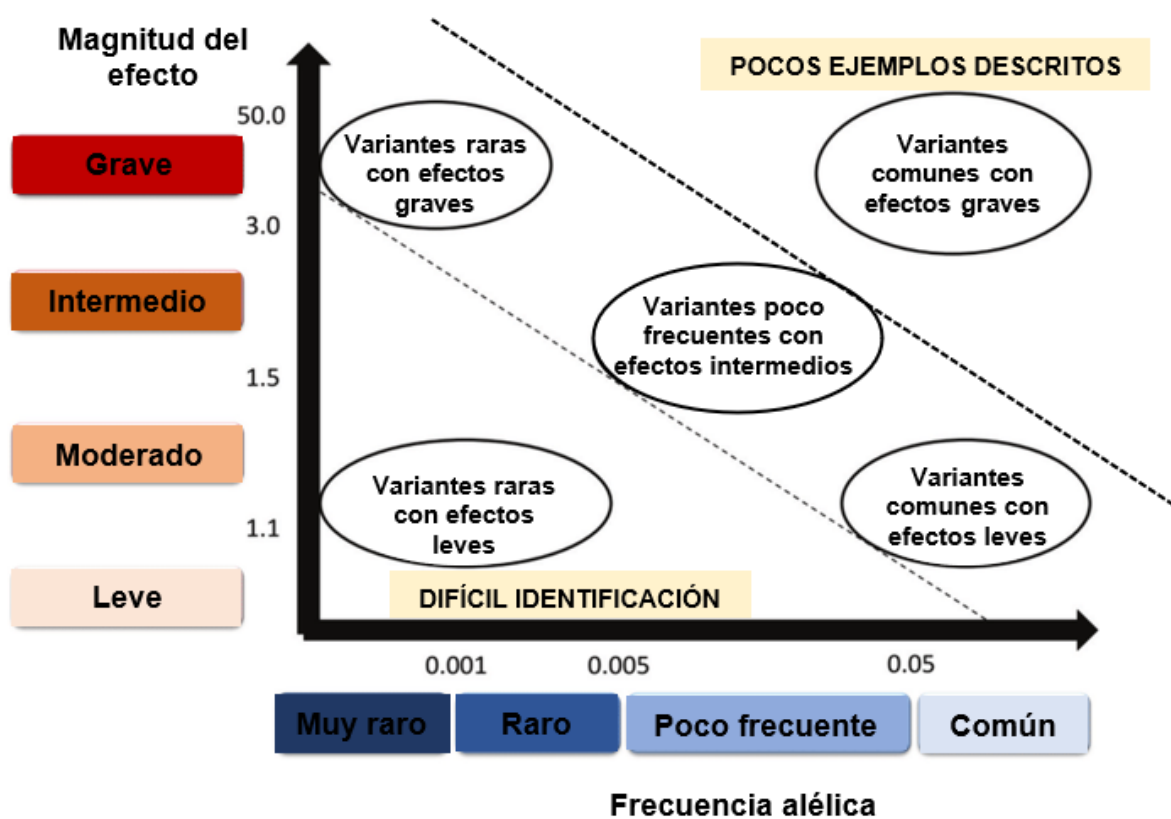


Figura 29. Identificación de variantes genéticas en función de la frecuencia del alelo de riesgo y la magnitud del efecto genético. Relación entre la frecuencia del alelo de riesgo y sus efectos en la enfermedad humana. Figura adaptada de Manolio *et al.*<sup>469</sup>.

Históricamente, la mayoría de los estudios genéticos se han centrado en estudiar variantes en regiones codificantes. Estas variantes son estudiadas para describir su función o predecir su efecto con herramientas bioinformáticas. Polyphen, por ejemplo, es una herramienta utilizada para predecir el posible impacto de una sustitución de aminoácidos en la estructura y función de una proteína humana utilizando consideraciones físicas y comparativas sencillas<sup>244</sup>.

Sin embargo, durante los últimos quince años, los GWAS han descubierto que la mayoría de las variantes asociadas a enfermedad se encuentran dentro de las regiones no codificantes, especialmente en las regiones reguladoras *cis*<sup>117</sup>. Esto ha llevado a un creciente interés en examinar el papel de las variantes reguladoras en las enfermedades hereditarias. En este sentido, un análisis de 150 millones de SNVs en 10 000 genomas humanos reveló que la variación genética se distribuye de manera desigual en el genoma humano y que los promotores, los TFBS, las regiones de cromatina abierta y el factor de unión de CCCTC (CTCF), en este orden, acumulan la mayor cantidad de SNVs en elementos reguladores en todo el genoma humano<sup>245</sup>.

Las conclusiones principales de los GWAS junto a otros estudios genéticos son que a menudo existe una correlación entre la gravedad y la frecuencia de las variantes, donde las variantes con un impacto mayor son raras y aquellas con un impacto limitado son comunes<sup>246</sup>. Como resultado, Cerrone *et al.*, Walsh *et al.* & Monaksy *et al.* han sugerido que las variantes comunes podrían exacerbar o atenuar la manifestación de la enfermedad junto con un defecto genético primario, que generalmente se describe en una región codificante, pero estas variantes también pueden ser responsables del fenotipo de la enfermedad en ausencia de variantes codificantes<sup>247-249</sup>.

Otra fuente de complejidad es que una variante común en una población en particular puede ser rara en otra<sup>250</sup>. Autores como Lupski *et al.* y Corradin *et al.* sostienen que, en algunos casos, el fenotipo de la enfermedad podría modularse a través de una combinación anormal de variantes comunes<sup>232,251</sup>.

#### 4.4. Efectos funcionales de las variantes no-codificantes

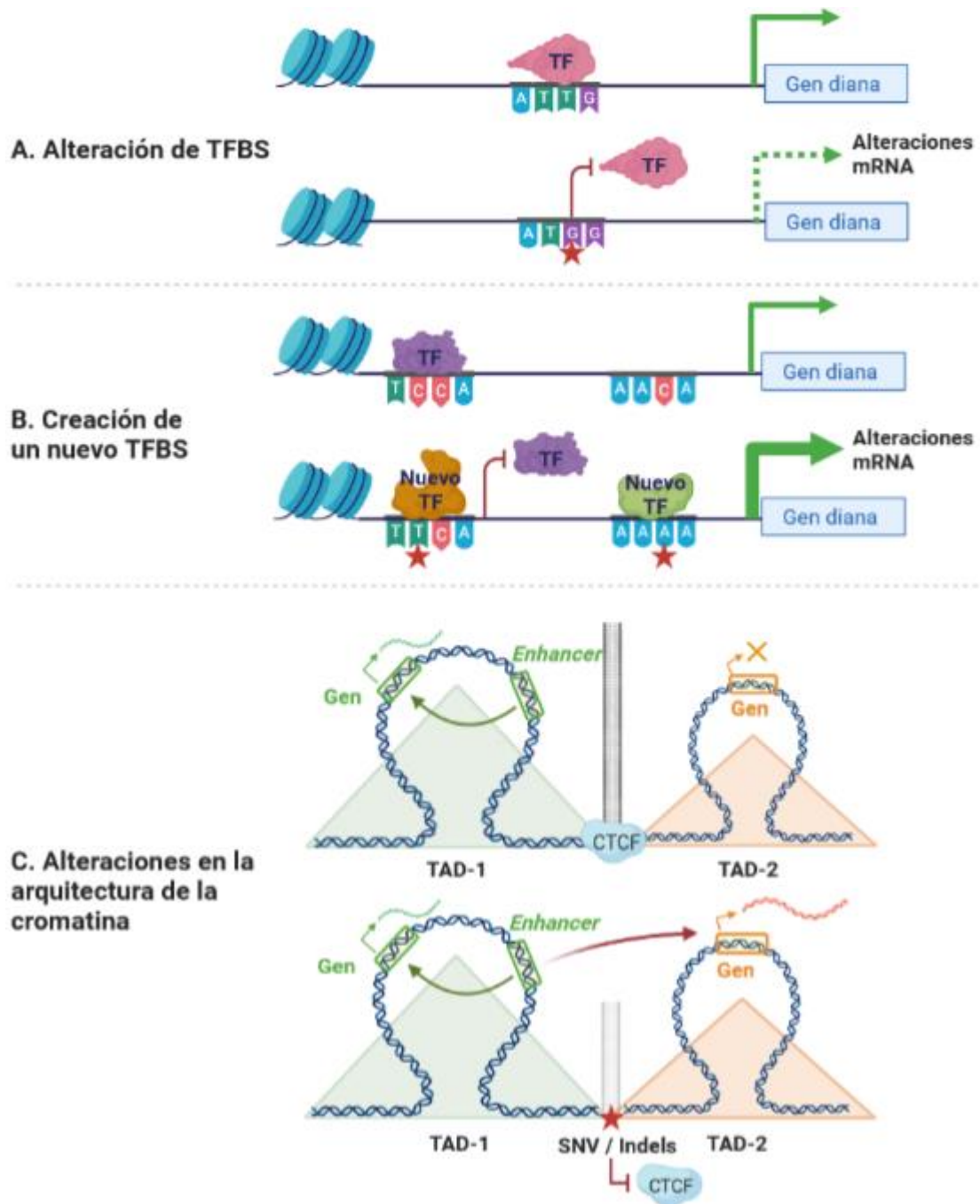
El efecto funcional de las variantes en las regiones codificantes es fácil de predecir, ya que se conoce el funcionamiento del código genético y existe una cantidad importante de datos disponibles del exoma, así como un buen conocimiento de los trastornos mendelianos<sup>252</sup>. Sin embargo, el efecto de las variantes en las regiones no codificantes es difícil de evaluar. Estas variantes no codificantes se pueden encontrar dentro de elementos reguladores *cis*, como *enhancers*, promotores y *insulators*, así como en regiones que codifican ncRNAs<sup>253</sup>.

Las variantes en estos elementos reguladores pueden tener distintos efectos según el cambio de nucleótido y la posición de la variante:

- **Alteraciones en la accesibilidad del DNA:** una variante no codificante podría interrumpir la unión de los FT a los elementos reguladores *cis*, cambiando la accesibilidad del DNA a los FT<sup>254,255</sup>. Además, las variantes no codificantes que afectan la accesibilidad del DNA se han relacionado con alteraciones en la unión de las proteínas de unión a la cromatina, lo que puede provocar alteraciones en la expresión génica<sup>256,257</sup>.
- **Alteraciones en la unión de los TF al DNA:** aunque este podría ser un efecto derivado de las alteraciones en la accesibilidad del DNA, una variante no codificante también puede afectar directamente el TFBS de los TF, disminuyendo o aumentando la afinidad de los TF para unirse a su TFBS<sup>258</sup> (**Figura 30A**).
- **Creación de un nuevo TFBS:** una variante no codificante también podría crear un nuevo TFBS, alterando así la red regulatoria existente y generando vías regulatorias alternativas<sup>258</sup> (**Figura 30B**).
- **Alteraciones en la organización espacial del genoma:** las variantes dentro de los TFBS de CTCF pueden afectar la organización espacial del genoma, lo que lleva a alteraciones de las redes transcripcionales<sup>259,260</sup> (**Figura 30C**).

Sin embargo, en la mayoría de los casos, las alteraciones en la unión de los FT (ya sea por variantes en el TFBS o en el dominio de unión al DNA del FT) no tienen grandes efectos en los niveles de transcripción, y por esta razón se cree que actúan como modificadores genéticos<sup>261</sup>. Un modificador genético puede ejercer su efecto sobre el fenotipo establecido por variantes genéticas de varias maneras<sup>262</sup>. Dependiendo de la naturaleza del modificador, esta interacción puede resultar en:

- Penetrancia incompleta del fenotipo.
- Expresividad variable, pudiendo presentar fenotipos exacerbados (edad de inicio más temprana y progresión más rápida de la enfermedad).
- Una presentación clínica más leve (edad de inicio más tardía, progresión de la enfermedad atenuada y ausencia/protección de fenotipos específicos).
- Modificadores que potencian la manifestación de fenotipos específicos que dan como resultado lo que se describe como nuevas formas de enfermedad.



**Figura 30. Efectos de las variantes reguladoras en *cis*.** (A) Las variantes dentro de las regiones reguladoras *cis* pueden alterar los TFBS, lo que lleva a una disminución de la expresión del mRNA. (B) Las variantes dentro de las regiones reguladoras en *cis* pueden crear nuevos TFBS y generar nuevas vías reguladoras. (C) Las variantes que afectan la unión de CTCF pueden interrumpir la actividad aislante entre dos TADs, generando nuevas interacciones promotor-*enhancer* que pueden conducir a una expresión alterada de los genes dentro de estos TAD. La estrella roja indica la posición de una SNV o indel. La X naranja indica inhibición de la expresión génica. Figura propia extraída de Pérez-Agustín *et al.*<sup>352</sup>.



Los nucleótidos de los TFBS tienen distinta relevancia para la unión de los FT. Los nucleótidos más conservados, también denominados nucleótidos centrales, son esenciales para la unión de FT. Por lo tanto, las variantes genéticas que afecten a los nucleótidos centrales tendrán un mayor impacto en la unión del FT. Por el contrario, se cree que los nucleótidos menos conservados confieren cierta tolerancia a la variación genética, lo que podría explicar por qué muchos cambios en TFBS no dan como resultado alteraciones medibles en los niveles de mRNA<sup>263</sup>. Además, los leves efectos observados en las variantes reguladoras en *cis*, especialmente en los elementos *enhancer*, también podrían explicarse por el concepto de redundancia reguladora<sup>264</sup>.

En un estudio reciente, Osterwalder y colaboradores observaron que la eliminación de un solo *enhancer* en genes rodeados de múltiples *enhancers* no producía cambios fenotípicos, mientras que la eliminación de más de uno producía alteraciones medibles. Estos hallazgos sugieren que los *enhancers* presentan una robustez frente a la variación genética cuando actúan juntos, y que su contribución a la regulación génica es probablemente redundante en lugar de aditiva<sup>265</sup>. Además, otros estudios han propuesto que la regulación no solo depende de la afinidad del FT hacia una secuencia de DNA, sino también de un equilibrio entre la afinidad del FT y la sintaxis regulatoria. El concepto de sintaxis regulatoria se refiere a factores como el orden, espaciado y orientación de los sitios de unión. Es decir, una sintaxis óptima puede compensar un TFBS débil, y viceversa<sup>266</sup>.

En resumen, los mecanismos de unión a FT y la función promotora/*enhancer* reducen la probabilidad de que se produzcan grandes efectos causados por una única variante. Esto sería consistente con múltiples estudios que sugieren que las variantes reguladoras comunes deben estudiarse en combinación y no aisladas<sup>232,249</sup>.

## 5. Técnicas genómicas para la identificación de nuevas variantes

### 5.1. Genómica en patología cardiovascular

La mayoría de enfermedades cardíacas arritmogénicas de etiología genética son consideradas de herencia mendeliana. Sin embargo, la evidencia acumulada en la última década (a partir de la secuenciación masiva y el cribado de miles de individuos) sugiere patrones de herencia complejos para algunas de ellas<sup>267,268</sup>. Estas enfermedades comparten características clínicas y genéticas comunes: la heterogeneidad genética (la mayoría son consideradas oligogénicas o poligénicas) y fenotípica<sup>269</sup>, la expresividad variable y la penetrancia incompleta. En algunos casos, el solapamiento fenotípico y de los genes asociados entre los distintos desórdenes ensombrece a los patrones de herencia y dificulta la diagnosis y el consejo genético<sup>270</sup>.

Por tanto, la identificación de variantes causales en un paciente es crucial para: (1) la confirmación del diagnóstico en los casos límite; (2) el manejo precoz de los familiares en situación de riesgo y (3) el asesoramiento genético y preimplantacional, así como para planificación familiar<sup>271,272</sup>.

En este contexto, las directrices clínicas actuales recomiendan realizar un cribado genético en pacientes que cumplan los criterios establecidos por los distintos desórdenes, secuenciando las regiones codificantes de los principales genes asociados a cada una de las enfermedades a estudio<sup>272</sup>. Por ejemplo, en el caso del SBr, se recomienda la secuenciación de *SCN5A* (aunque se han asociado unos 20 genes más a la patología). De forma comparable, en el caso de la MCH, se recomienda la secuenciación de las regiones codificantes de los principales 5 genes sarcoméricos (*MYBPC3*, *MYH7*, *TNNI3*, *TNNT2* y *TPM1*), aunque el número total de genes asociados a la MCH (con diferentes grados de evidencia) sean aproximadamente unos 25. La situación es extrapolable a todas las enfermedades arritmogénicas asociadas a la MSC, ya sean canalopatías o miocardiopatías<sup>271</sup>.

Este cribado de genes en la práctica clínica se debe a la necesidad de optimizar y transformar la inversión destinada a la identificación de la causa genética en un resultado útil derivado de la investigación. El cribado genético tiene como resultado un listado de variantes identificadas en los pacientes de interés. Este listado será resumido en un informe que se hará llegar al cardiólogo y/o al forense para la toma de decisión clínica o diagnóstica. Por este motivo, la identificación de variantes no reportadas previamente debe ir acompañada de evidencia sólida, para ser asociada a la enfermedad del paciente en cuestión. Esta evidencia se obtiene mediante las predicciones de patogenicidad de algoritmos *in silico*, o por la asociación de la variante a la enfermedad, concluida a partir de estudios funcionales *in vitro* (en modelos celulares) o *in vivo* (en modelos animales).

En este sentido, la llegada de las tecnologías de secuenciación de alto rendimiento ha mejorado indiscutiblemente el campo del diagnóstico genético. Se ha reducido el tiempo y la intervención de personal cualificado capaz de llevar a cabo la labor de secuenciación, así como la inversión en reactivos e infraestructuras para secuenciar a demanda del médico. Sin embargo, hay que tener en cuenta que un experimento de secuenciación de alto rendimiento proporciona una enorme cantidad de datos que requieren su procesado bioinformático posterior. Además, para la mayoría de variantes identificadas, la información sobre su efecto fenotípico es generalmente limitada<sup>273</sup>: si se realizara un estudio funcional para cada una de las variantes candidatas a tener un rol patogénico en la enfermedad del paciente, la inversión económica en el diagnóstico genético sería inalcanzable y los largos tiempos de espera (de varios meses hasta años) harían que el sistema fuera insostenible. Este problema remarca la necesidad de encontrar un equilibrio entre el diagnóstico clínico y la investigación biomédica de forma que la investigación básica se coordine con la praxis clínica hacia una medicina de precisión, personalizada y traslacional: del paciente al laboratorio y de vuelta al paciente. Habitualmente, este equilibrio se encuentra realizando una secuenciación de paneles de genes asociados (y candidatos) y con la anotación exhaustiva de toda la información presente en las bases de datos de las variantes identificadas. Es importante tener en cuenta que este análisis debe seguir unos criterios de clasificación de variantes previamente establecidos y comunes por la comunidad médica y científica<sup>272</sup>. El reto actual es lograr un sistema de salud en que la validación de la variación genómica se coordine de forma eficiente para llegar de nuevo al paciente.

Con el desarrollo de las tecnologías de secuenciación de alto rendimiento, el diagnóstico genético ha mejorado considerablemente. Sin embargo, el porcentaje de casos en los cuales la etiología permanece desconocida tras el cribado de variantes en los principales genes asociados a las enfermedades arritmogénicas cardíacas es todavía muy alto<sup>55,274</sup>. Estos porcentajes varían en función de lo completo que sea el cribado en cuanto a número de genes revisados. Los casos no resueltos pueden explicarse por variantes patogénicas en genes no asociados a la enfermedad en el momento del cribado y, por tanto, no incluidos en el estudio, pero también puede haber influencia de variantes en regiones reguladoras<sup>275</sup>, o alteraciones epigenéticas<sup>276</sup>. Estas regiones no se incluyen en el estudio genético por motivos de coste-efectividad y por la complejidad en la interpretación y asociación con la enfermedad. Otro tipo de variantes potencialmente patogénicas en los casos negativos son las variantes estructurales, no detectables por la secuenciación Sanger capilar convencional.

## **5.2. Next-Generation Sequencing (NGS)**

Desde año 1977, la secuenciación Sanger se ha considerado tiempo la técnica de referencia para la identificación de variantes genéticas. Si bien es una técnica sólida y produce resultados

muy fiables, el rendimiento y la cobertura genómica (número de veces que una porción del genoma es secuenciada) son limitados y, por lo general, no se pueden detectar variantes con frecuencias alélicas bajas<sup>277</sup>. Las tecnologías NGS ofrecen un rendimiento y una sensibilidad mucho mayores que la técnica Sanger, lo que permite el análisis de genomas, exomas o transcriptomas completos en una secuencia<sup>278</sup>. Todo ello, gracias a la capacidad de generar una mayor cantidad de lecturas del genoma que permiten obtener un resultado más sólido, a precios más asequibles y en un menor período de tiempo. Existen distintas plataformas NGS que se pueden clasificar según la longitud de las lecturas que obtienen: secuenciación de lecturas cortas (*short-read sequencing*; SRS) o lecturas largas (*long-read sequencing*; LRS)<sup>279</sup>. Ambos métodos de secuenciación tienen sus beneficios y defectos, y debemos escoger la que sea más adecuada según el objetivo del experimento.

La **secuenciación de lecturas cortas** es una herramienta muy potente para generar datos genómicos. Con la SRS, el DNA o el RNA se pueden secuenciar de forma rápida y a un coste menor que los métodos tradicionales. Esta tecnología ha revolucionado la investigación biomédica y ha dado lugar a importantes descubrimientos en genómica, evolución y enfermedades. También se ha utilizado para ensamblar genomas completos, proporcionando información valiosa sobre la estructura y función de los genes<sup>279</sup>. Las lecturas cortas son efectivas para aplicaciones destinadas a cuantificar la cantidad de una secuencia específica, identificar variantes en secuencias bien conservadas o para definir la expresión de transcritos<sup>279</sup>. Las principales plataformas para secuenciación de lecturas cortas son:

- **454 pyrosequencing**: esta estrategia se fundamenta en la pirosecuenciación, una metodología que se basa en la detección quimioluminiscente del pirofosfato liberado durante la elongación de la cadena complementaria de DNA, lo que permite la rápida determinación de secuencias a tiempo real. A día de hoy es una plataforma prácticamente en desuso, ya que el resto de tecnologías disponibles ofrecen resultados de mejor calidad y a un coste menor<sup>280</sup>.
- **SOLID**: es la única estrategia que lleva a cabo una secuenciación por ligación. Debido a su alta tasa de error y la generación de secuencias muy cortas, de unas 50 pb, actualmente, como en el caso de 454, es una de las estrategias menos utilizadas<sup>281</sup>.
- **Ion Torrent**: es un método de secuenciación basado en la detección de protones liberados durante el proceso de polimerización del DNA. Por lo tanto, este método se guía a través de la adición de nucleótidos que se llevan a cabo en las cadenas simples del DNA de estudio. La principal ventaja de esta tecnología es la velocidad de secuenciación, así como los bajos costes operativos. Esto ha sido posible gracias a que no se utilizan nucleótidos modificados y a que la señal no es óptica<sup>281</sup>. Dado que el sistema registra las incorporaciones naturales

de nucleótidos mediadas por polimerasa, la secuenciación ocurre a tiempo real. No obstante, existen algunas limitaciones. Entre ellas, la más importante es la repetición de algunas bases iguales en la hebra molde. En este caso se añadirán más dNTPs en el mismo momento, generando un cambio de pH más pronunciado y como consecuencia una señal eléctrica mayor, lo cual actualmente no se detecta con mucha precisión<sup>281</sup>.

- **Illumina:** es la tecnología de secuenciación más utilizada en la actualidad, debido a que ofrece plataformas efectivas capaces de ofrecer resultados rápidos y precisos a precios razonables. Esta metodología se basa en la secuenciación por síntesis y la utilización de nucleótidos marcados con fluoróforos que bloquean de forma reversible la elongación de la cadena<sup>282</sup>. De este modo, tras la detección de la incorporación del fluoróforo, y la eliminación del mismo, es posible continuar con un nuevo ciclo de adición de un nuevo nucleótido. Illumina ha permitido abaratar los costes y ofrecer lecturas más largas (hasta 300 pb), sin embargo, se requiere de una gran inversión inicial y de mantenimiento, para poder adquirir y mantener uno de los equipos que ofrece<sup>282</sup>.

La **secuenciación de lecturas largas** permite la obtención de lecturas de secuenciación más largas (>10 000 pb) que los sistemas SRS más comúnmente utilizados (75-300 pb). Algunas plataformas de LRS han producido lecturas de secuencias de 882 000 pb<sup>283</sup>, y algunos grupos han reportado más de 2 000 000 pb<sup>284</sup>; sin embargo, las longitudes de lectura de entre 10 000 – 100 000 pb son más comunes<sup>285</sup>. Las tecnologías de LRS secuencian directamente moléculas individuales de DNA en tiempo real, a menudo sin necesidad de amplificación. Actualmente, las 2 principales plataformas de LRS son Pacific Biosciences (PacBio) y Oxford Nanopore Technologies (ONT):

- **PacBio:** adopta una estrategia de secuenciación por síntesis similar a la secuenciación de Illumina, con la excepción de que PacBio captura una sola molécula de DNA e Illumina detecta señales amplificadas de una población clonal de fragmentos de DNA amplificados<sup>286</sup>. PacBio opera dentro de un chip de silicio (celda SMRT), que contiene un gran número de pozos microscópicos llamados ZMW (zero-mode waveguides). Dentro de cada pozo, la secuenciación se realiza en modo de molécula circularizada llamada SMRTbell.
- **ONT:** es una tecnología de secuenciación basada en nanoporos (canal de una proteína sintética con un diámetro interior de 1 nanómetro). En comparación con otras tecnologías de secuenciación que utilizan la incorporación o hibridación de nucleótidos, ONT secuencian directamente una molécula de DNA monocatenario (ssDNA) midiendo los cambios de corriente que se producen a medida que las bases pasan a través de un nanoporo<sup>286</sup>. Debido al uso de ONT en esta tesis, se explica su tecnología con mayor detalle en el apartado 5.3. de la introducción.

Actualmente, las técnicas de NGS ya están presentes en la práctica clínica para la detección de variantes somáticas de relevancia diagnóstica y pronóstica y para facilitar el diagnóstico de enfermedades hereditarias<sup>287</sup>. Aunque las técnicas son las mismas que se utilizan en proyectos de investigación, la implementación de NGS en ensayos clínicos requiere el cumplimiento de varios criterios adicionales. De manera general, los pasos a seguir son los siguientes (**Figura 31**):



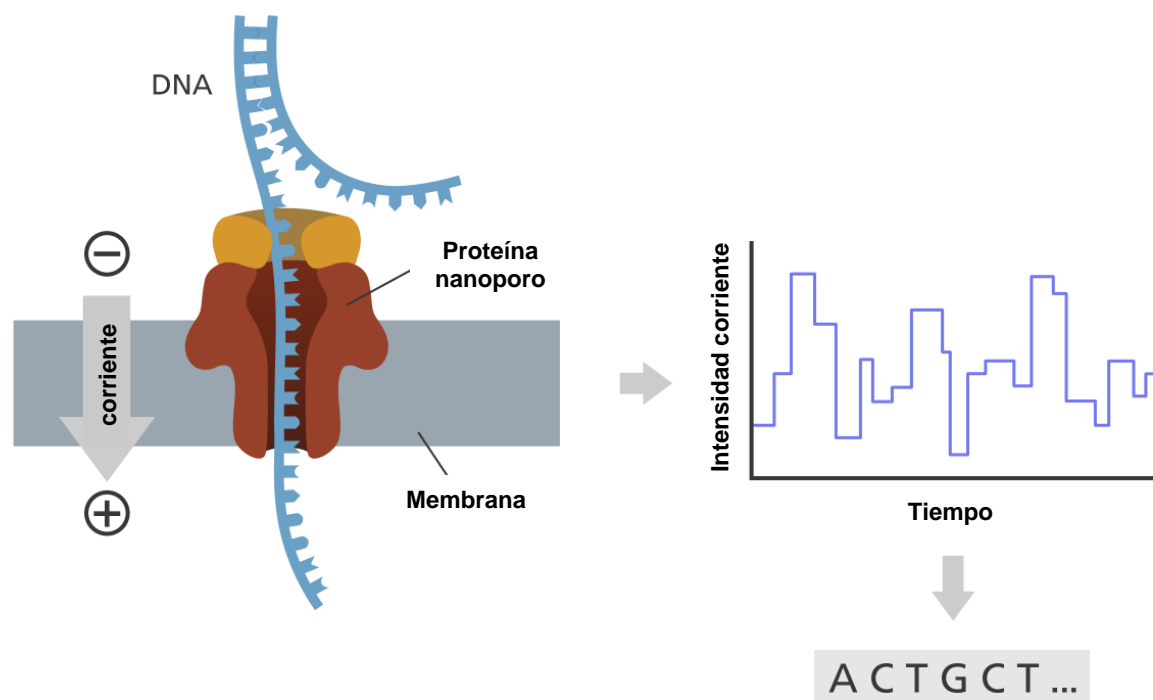
**Figura 31. Flujo de trabajo principal del análisis genético con NGS.** Las flechas indican los pasos a seguir: recogida de la muestra, secuenciación, alineamiento de las lecturas con un genoma de referencia, detección de las variantes, anotación y selección de las variantes de impacto clínico y finalmente, transmisión de los resultados al médico y al paciente. Figura adaptada de V. Ljungström<sup>291</sup>.

1. Extracción y optimización de la muestra (controles de calidad, requerimientos mínimos).
2. Preparación y secuenciación de la muestra del paciente
3. Obtención de los archivos FASTQ que contienen los datos crudos de las lecturas de secuenciación sin procesar.
4. Alineamiento de las lecturas con el genoma humano de referencia.
5. Identificación de las variantes en la muestra secuenciada mediante el *variant calling* (proceso mediante el cual se identifican variantes a partir de datos de secuenciación). El *variant calling* se realiza mediante herramientas bioinformáticas como *GATK*<sup>288</sup>.
6. Anotación de las variantes (*variant annotation*) con información relevante (datos funcionales, poblacionales, predicción del impacto funcional, etc), procedente de múltiples bases de datos como RefSeq, 1 000 Genomas, gnomAD, etc.

7. Selección de las variantes con mayor utilidad clínica mediante filtros específicos (se descartan, por ejemplo, variantes en regiones codificantes que no suponen un cambio aminoacídico).
8. Se informa de los resultados al médico, que transmitirá la información al paciente y tomará las medidas pertinentes.

### 5.3. Oxford Nanopore Technologies (ONT)

El sistema de secuenciación de ONT, comercializado en 2015, utiliza un pequeño dispositivo de secuenciación portátil llamado MinION. Su tecnología se basa en la medición de los cambios en la corriente eléctrica que provoca una molécula de DNA de cadena única al pasar a través de un nanoporo proteico a una velocidad de 450 bases por segundo<sup>290</sup>. Una enzima helicasa situada encima del nanoporo desenrolla la doble hélice de DNA y una de las hebras pasa a través del nanoporo. El canal tiene justo el tamaño adecuado para que pase la hebra de DNA. Durante el proceso de secuenciación, se aplica un voltaje a través del nanoporo, generando un movimiento de los iones de la solución a través del poro y generando una corriente iónica<sup>290</sup> (**Figura 32**).



**Figura 32. Funcionamiento de la tecnología de Oxford Nanopore.** Una de las cadenas de la doble hélice de DNA pasa por el nanoporo y el sistema detecta las variaciones en la corriente iónica a medida que van pasando las bases nitrogenadas del DNA. Figura adaptada de YourGenome (<https://www.yourgenome.org>).

Al pasar la hebra de DNA, como cada una de las cuatro bases del DNA tiene un tamaño distinto, bloquean una proporción diferente del nanoporo en atravesarlo. Esto significa que hay más o menos espacio para el paso de los iones a través del nanoporo y, por lo tanto, más o menos corriente puede fluir por el hueco (**Figura 32**). El aparato mide la corriente iónica y los datos se transfieren al ordenador y se analizan con el software minKNOW, que convierte las mediciones en tiempo real en una secuencia de DNA<sup>292</sup>.

Esta tecnología tiene algunas ventajas en comparación con las plataformas de SRS para el análisis clínico del genoma:

- **Ensamblaje de regiones del genoma:** el genoma humano tiene una longitud de más de tres mil millones de pares de bases de DNA y contiene muchos tramos repetitivos de código genético. Ensamblar el genoma a partir de lecturas cortas puede ser un reto, ya que muchos fragmentos tienen una elevada similitud (sin un contexto adicional). En cambio, los datos de lectura larga pueden simplificar esta tarea, ya que es más probable que las lecturas resulten distintas, lo que permite ensamblarlas con menos ambigüedad y error<sup>285</sup>.
- **Detección de variantes:** algunas características de los genomas individuales son particularmente difíciles de detectar y cuantificar con tecnologías SRS. Algunos ejemplos son los reordenamientos complejos, grandes inserciones o deleciones, regiones repetitivas, regiones altamente polimórficas o regiones con baja diversidad de nucleótidos de DNA<sup>285</sup>. El análisis de lecturas largas permite visualizar y detectar estos elementos con mayor precisión y fiabilidad.
- **Phasing de haplotipos:** en áreas como la medicina reproductiva puede ser útil saber si existen variantes genéticas en la misma copia del cromosoma. Esta utilidad reside en el hecho de que estas variantes estén formando un haplotipo con un determinado interés biológico. Esto se puede determinar mediante un proceso conocido como *phasing*<sup>293</sup>. El *phasing* consiste en identificar las copias de cada cromosoma heredadas por vía materna y paterna en haplotipos para determinar qué variantes son de la misma copia de un cromosoma y cuáles de copias diferentes. Las lecturas largas pueden proporcionar información de largo alcance para resolver haplotipos sin inferencia estadística adicional, secuenciación materna/paterna o preparación de muestras, como se requiere para una aproximación del *phasing* mediante SRS.
- **Portabilidad:** a diferencia de otras plataformas de secuenciación, los dispositivos de Nanopore se basan en la detección de señales electrónicas en lugar de ópticas. Esto les permite diseñar dispositivos tan pequeños como una memoria USB, haciéndolos altamente portátiles<sup>285</sup>.



No obstante, este dispositivo también conlleva algunas desventajas respecto a las tecnologías de SRS:

- **Tasa de error más elevada:** aunque esta tecnología ha mejorado en los últimos años, todavía muestra una tasa de error relativamente alta en las secuencias sin procesar en comparación con los dispositivos estándar NGS como Illumina. La tasa de error global media actual en lecturas sin procesar es de aproximadamente el 6%, mientras que en Illumina es inferior al 0.5%<sup>294,295</sup>.
- **Regiones de baja complejidad o repetitivas:** los secuenciadores de ONT tienden a tener dificultades para secuenciar regiones repetitivas con precisión, debido a una variación menor en la señal eléctrica del nanoporo cuando la base que atraviesa el nanoporo es la misma<sup>294</sup>.
- **Longitud de los fragmentos:** la determinación de la longitud exacta de los homopolímeros puede ser difícil de determinar debido a que la velocidad de translocación del DNA (velocidad de paso por el nanoporo) no es constante<sup>296</sup> (por ejemplo, una secuencia con 6 repeticiones del mismo nucleótido). Aproximadamente la mitad de los errores de secuenciación se deben a los homopolímeros.
- **Alta sensibilidad a regiones GC:** la secuenciación ONT es muy sensible al contenido de lecturas de GC. En consecuencia, las lecturas de alto contenido de GC tienen menor precisión<sup>296</sup>.

Aunque la tecnología está constantemente en fase de perfeccionamiento, tiene características que la convierten en una opción muy atractiva. La sencillez en la preparación de las librerías y su pequeño tamaño hacen de este método una opción adecuada para secuenciar diferentes tipos de organismos en prácticamente cualquier lugar<sup>297</sup>. También existen plataformas diseñadas específicamente para sumar la capacidad de los dispositivos individuales minION. Así se puede encontrar el GridION, con capacidad para albergar 5 MinIONs, y el PromethION, con capacidad para 48 celdas de flujo independientes de 300 nanoporos cada una (con un total de 144 000 canales)<sup>297</sup>.

## 6. Modelos celulares para el estudio de variantes que afectan a los canales de sodio dependientes de voltaje

Se han realizado numerosos estudios para comprender los mecanismos subyacentes a las diferentes canalopatías. Debido a limitaciones metodológicas y éticas, estos estudios se han realizado en modelos animales y celulares. Los modelos murinos se usan comúnmente para estudiar distintas canalopatías, incluidas las arritmias cardíacas y neuronales<sup>298,299</sup>. No obstante, los modelos murinos no son los más adecuados para el estudio de canalopatías cardíacas ya que tienen un ritmo cardíaco más rápido, con lo cual es difícil extrapolar los resultados a los humanos<sup>300</sup>.

Otros modelos, como el porcino y el canino, también se han utilizado para el estudio de las canalopatías cardíacas<sup>300-302</sup>. En el caso de canalopatías neuronales, se han utilizado modelos animales de pez cebra y mosca del vinagre<sup>299,303</sup>. Además de los modelos animales, también se utilizan ampliamente los modelos celulares. Estos modelos incluyen sistemas de expresión heterólogos y células específicas de tejido derivadas de células madre pluripotentes inducidas (hiPSC).

### 6.1. Sistemas de expresión heterólogos

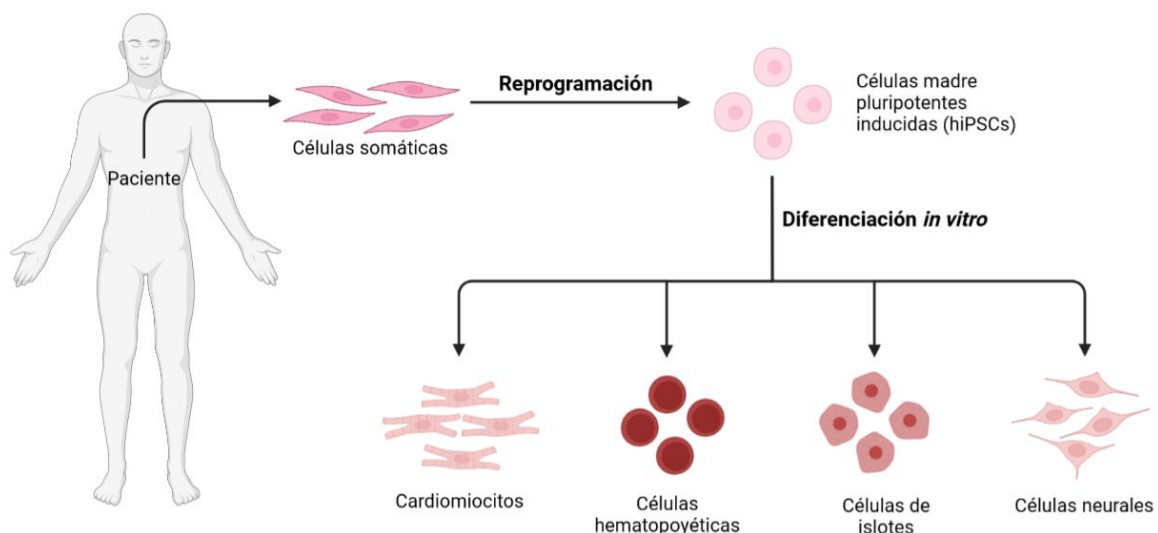
Los vectores de expresión portadores de las variantes en genes que codifican para los canales de sodio regulados por voltaje normalmente se transfectan en sistemas de expresión heterólogos. Para este tipo de aproximación, se utilizan líneas celulares inmortalizadas como las células de riñón embrionario humano (HEK), células de ovario de hámster chino (CHO) y ovocitos de *Xenopus laevis*<sup>299,300,304</sup>.

Las principales ventajas de usar sistemas de expresión heterólogos son que tienen un coste económico asequible, son fáciles de mantener y de manipular. Sin embargo, este modelo carece del *background* genético específico del paciente y, por lo tanto, no aporta información sobre la penetrancia incompleta que muestran algunas enfermedades arritmogénicas. Los sistemas de expresión heterólogos no muestran el entorno nativo de los tipos de células afectadas por las variantes, como los cardiomiocitos o las neuronas. Además, el hecho de que la expresión se produzca a través de un plásmido (debido a que la variante no está insertada en el genoma) tiene consecuencias como: (i) niveles de expresión anormales y (ii) dificultad para estudiar variantes en homocigosis/heterocigosis. Para superar estas desventajas, en los últimos años se han desarrollado otros sistemas modelo para el estudio de las variantes que afectan a los canales de sodio dependientes de voltaje.

## 6.2. Células específicas de tejido derivadas de células madre pluripotentes inducidas (hiPSC)

Desde 1958, distintos laboratorios han trabajado en el desarrollo de métodos de reprogramación celular. La primera técnica que se desarrolló fue la transferencia nuclear de células somáticas, en la cual se transfiere el núcleo de una célula somática a un óvulo enucleado<sup>305</sup>. En este momento se descubrieron las células madre embrionarias (ES), las cuales se caracterizan por su pluripotencia, ya que tienen capacidad de poder dar lugar a todo tipo de células<sup>306,307</sup>. Algunos estudios también descubrieron que algunos tipos de células somáticas, como los fibroblastos y los linfocitos T, podían reprogramarse mediante la fusión celular con ES. Las células reprogramadas resultantes expresaron genes asociados con la pluripotencia, como *Oct3/4*<sup>308</sup>. Estos resultados sugirieron la existencia de factores de reprogramación que podrían modificar la memoria de las células somáticas.

A partir de estos estudios previos, Takahashi y Yamanaka generaron, células madre pluripotentes inducidas (iPSC) a partir de embriones de ratón y fibroblastos adultos, utilizando cuatro factores de transcripción (*Oct3/4*, *Sox2*, *Klf4* y *c-Myc*)<sup>309</sup>. Estas iPSC se derivaron de células somáticas y compartían un parecido funcional con las células madre embrionarias. Poco después, el mismo grupo descubrió la obtención de hiPSC a partir de células somáticas humanas usando los mismos cuatro TF<sup>310</sup>. Estudios posteriores demostraron que era factible obtener hiPSC usando diferentes conjuntos de FT<sup>311</sup> (**Figura 33**).



**Figura 33. Obtención de hiPSC a partir de células somáticas y diferenciación a otros tipos celulares.** Figura creada con BioRender (<https://app.biorender.com/>).

Las hiPSC tiene la propiedad de ser pluripotentes, y son capaces de diferenciarse en todos los tipos de células somáticas. Las células que portan un perfil genético específico pueden obtenerse de los pacientes y reprogramarse a hiPSC. Además, estas hiPSC pueden ser

diferenciadas a un tipo de célula específico para imitar los rasgos característicos de una determinada enfermedad. El uso de hiPSC permite el estudio de proteínas mutantes en su entorno nativo, y se han convertido en un modelo valioso para superar varias limitaciones presentes en el modelo explicado anteriormente.

Hasta la fecha, las hiPSC suponen una estrategia prometedora para la determinación de la acción de los fármacos, el modelado de enfermedades y la medicina regenerativa, debido a su capacidad para diferenciarse en cualquier tipo de célula, incluidas las neuronas y los cardiomiocitos (**Figura 33**).

### 6.2.1. Cardiomiocitos derivados de hiPSC (hiPSC-CM)

En los últimos años se han establecido varios protocolos para la diferenciación de células hiPSC en cardiomiocitos. Una de las principales ventajas de este modelo es que los hiPSC-CM específicos del paciente tienen el *background* genético del paciente. Además, estas células reproducen un entorno similar al de los cardiomiocitos. Por ejemplo, este modelo celular ofrece la posibilidad de detección de cardiotoxicidad en fármacos probando sus efectos tóxicos en los cardiomiocitos antes de entrar en ensayos clínicos<sup>300</sup>.

Sin embargo, los cultivos de hiPSC-CM presentan heterogeneidad de tipos de células cardíacas y un fenotipo inmaduro en comparación con los cardiomiocitos adultos. De hecho, se ha demostrado que su perfil de expresión génica es más similar al de los cardiomiocitos fetales<sup>300,312</sup>. Presentan latidos espontáneos, su potencial de membrana está más despolarizado que los cardiomiocitos adultos y su organización estructural es deficiente<sup>300</sup>.

Para aumentar la pureza de los cardiomiocitos y excluir o reducir los no cardiomiocitos dentro de la población celular, distintos estudios han desarrollado una selección metabólica de supervivencia dependiente del lactato libre de glucosa<sup>313,314</sup>. Se han realizado varios esfuerzos para promover la maduración de hiPSC-CM, incluida la duración prolongada del cultivo, la estimulación eléctrica, el estiramiento mecánico, la estimulación con hormonas adrenérgicas y tiroideas<sup>315</sup>, la sobreexpresión forzada de miRNAs expresados en cardiomiocitos adultos y tejidos cardíacos vascularizados diseñados en 3D. Las revisiones publicadas por Liang *et al.* y, más recientemente, Reilly *et al.* resumen todos estos enfoques<sup>316,317</sup>.

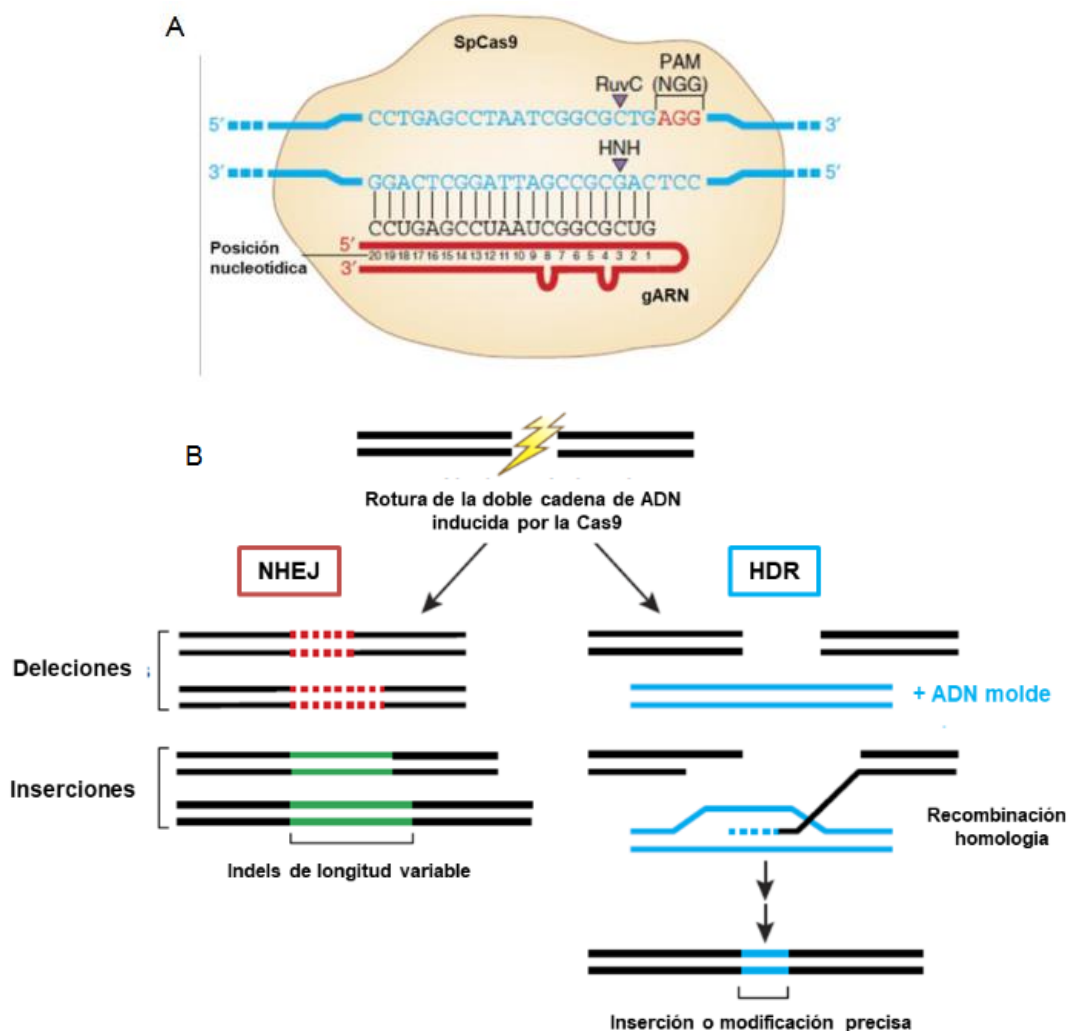
## 6.3. Modelos hiPSC con edición genómica

### 6.3.1. CRISPR/Cas9

Una de las ventajas principales de las hiPSC es que ofrecen la posibilidad de introducir la variante de interés mediante la técnica de edición genómica. En los últimos años, múltiples estudios han seguido esta estrategia para analizar el efecto de variantes asociadas a canalopatías cardíacas. Por ejemplo, Liang y colaboradores utilizaron este modelo en un estudio

en el que compararon las corrientes de sodio, los potenciales de acción y la dinámica del calcio en hiPSC-CM derivados de dos pacientes con SBr (portadores de una variante en *SCN5A*) y dos controles sanos<sup>316</sup>. Los autores encontraron una corriente de sodio reducida, un aumento de la actividad del potencial de acción y transitorios anormales de  $\text{Ca}^{2+}$  en las células derivadas de pacientes con SBr. A continuación, revirtieron la variante en cuestión (c.4190delA) al genotipo WT utilizando CRISPR/Cas9. Esta reversión dio lugar a valores normales en las propiedades eléctricas de las células derivadas del paciente, lo que indica que la variante fue suficiente para causar los defectos. Este estudio muestra que una aproximación de edición del genoma en hiPSC-CM puede ser valiosa para identificar asociaciones genotipo-fenotipo.

De manera resumida, la técnica de edición con CRISPR y Cas9 se basa en la capacidad de producir un corte en la doble cadena del DNA con especificidad de secuencia (**Figura 34**).



**Figura 34. Elementos que constituyen el sistema CRISPR/Cas9 como herramienta de edición genómica. (A)** Complejo Cas9 con el gRNA hibridado. **(B)** La rotura de la doble cadena de DNA activa las vías de reparación NHEJ o HDR (en presencia de un DNA molde). Figura adaptada de P. Rodríguez<sup>289</sup>.

Esta especificidad se debe a la secuencia de gRNA diseñada de forma complementaria a la secuencia diana de DNA en la que, más tarde, se producirá el corte por la endonucleasa Cas9<sup>318</sup>. El corte de la doble cadena de DNA desencadena los mecanismos de reparación celular. En ausencia de una secuencia de DNA que sirva como molde para la reparación, se activa la vía de reparación NHEJ, produciendo aleatoriamente inserciones, deleciones y/o sustituciones en el sitio de corte<sup>318</sup>. Por otro lado, la presencia de una secuencia de DNA que sirva como molde activa la vía de reparación por homología directa o HDR, basada en la recombinación homóloga del DNA genómico y el DNA molde o donador, que permite la edición precisa y controlada del genoma<sup>318</sup>.

### 6.3.2. CRISPRi

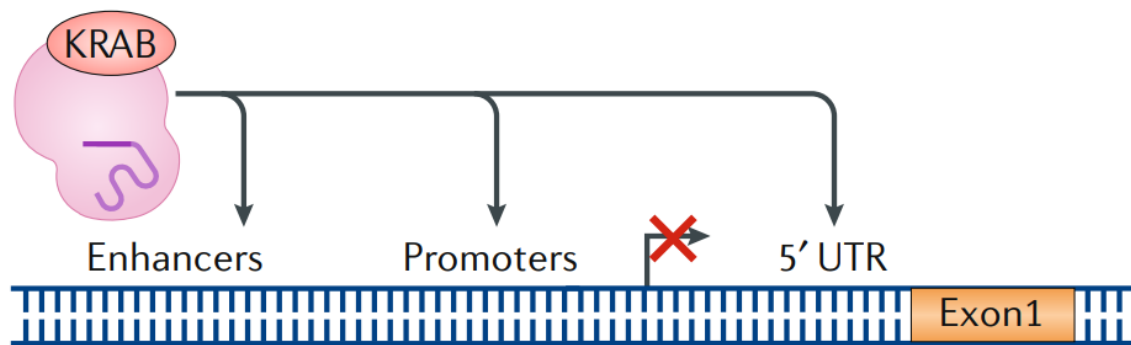
El sistema CRISPRi (CRISPR de interferencia), es una variedad de CRISPR/Cas9 que se ha desarrollado para estudios de expresión génica. Este tipo de estudios permiten analizar de manera precisa el efecto de una región concreta del genoma sobre la expresión de un gen determinado.

CRISPRi presenta importantes ventajas: (1) se puede estudiar el genoma con gran precisión, ya que el sistema se dirige por medio de gRNAs para reprimir la expresión en regiones específicas, (2) permite realizar estudios de pérdida de función a gran escala genómica, (3) puede silenciar alelos únicos en presencia de polimorfismos, y, (4) produce una represión génica más eficiente y homogénea. Además, es considerablemente más económico que otros sistemas como los TALENs<sup>220,319,320</sup>.

A nivel técnico, CRISPRi, es utilizado para estudiar los efectos de la represión de la expresión génica a través del uso de la versión desactivada de la nucleasa Cas9 (dCas9). Este sistema, dirige a dCas9 de manera específica, mediante un gRNA, a una secuencia diana del genoma. Una vez en la diana, dCas9 tiene la particularidad de que no corta la doble cadena de DNA, sino que se mantiene unida a esta, interfiriendo con cualquier otro complejo que interactúe de manera normal con la secuencia diana (como FT o *enhancers*), impidiendo así la expresión del gen de una forma reversible<sup>320</sup>.

La fusión de la dCas9 con el dominio de represión KRAB (*Krüppel-associated box*), permite aumentar la eficiencia de CRISPRi (**Figura 35**), generando fenotipos *knockdown* en células hiPSC<sup>220,319</sup>. A diferencia de los fenotipos *knockout*, en los que se elimina la secuencia de un gen o parte de la misma, en los fenotipos *knockdown* no se altera la secuencia de DNA, sino que se pretende inhibir la actividad transcripcional del gen.

## Represión transcripcional



**Figura 35. Sistema dCas9-KRAB para *knockdown* dirigido.** El complejo dCas9-KRAB se unirá a la región específica marcada por la gRNA. Figura adaptada de Pickar-Oliver & Gersbach<sup>367</sup>.

## 7. El Síndrome de Brugada

El SBr se describió por primera vez en 1992 como un trastorno cardíaco hereditario raro. Se caracteriza por un patrón electrocardiográfico distintivo con una elevación del segmento ST de tipo cóncavo en las derivaciones precordiales derechas<sup>321</sup>. Este trastorno se asocia con un alto riesgo de MSC entre los individuos, y es responsable del 15-20% de las arritmias supraventriculares y los casos de FA en corazones estructuralmente normales<sup>322</sup>.

Hasta la fecha, la causa más conocida de SBr es la presencia de variantes en las regiones codificantes del gen *SCN5A*, que representan aproximadamente el 30% de los casos de SBr<sup>323,324</sup>. Estas variantes en *SCN5A* incluyen SNVs y indels, y se han asociado a la pérdida de función de Nav1.5, ya sea afectando la actividad de su canal o reduciendo el número de canales en la membrana plasmática<sup>323</sup>. Hasta 42 genes más, como *SCN1B*, *SCN2B*, *SCN3B*, *CACNA1C*, *CACNB2*, *CACNA2D1*, *HEY2*, *SCN10A*, *SEMA3A* o *TBX*, también se han asociado con la enfermedad con menor incidencia, representando menos del 5% de los casos de SBr<sup>324</sup>. Sin embargo, su relevancia aún se cuestiona debido a la falta de pruebas definitivas de causalidad<sup>325</sup>. Por lo tanto, en más del 65% de los pacientes diagnosticados de SBr, la etiología de la enfermedad aún es desconocida.

El SBr se ha considerado tradicionalmente una enfermedad monogénica con un modo de transmisión autosómico dominante<sup>326</sup>. Sin embargo, en los últimos años, varias observaciones han proporcionado evidencia que sugiere que el SBr tiene una herencia más compleja y una base genética heterogénea<sup>327-329</sup>. Además, está bien establecido que la edad, el género, los factores ambientales y las comorbilidades contribuyen al fenotipo del SBr<sup>330</sup>. A medida que aparecen nuevos estudios genéticos, se hace más evidente que tanto las regiones codificantes como las no codificantes desempeñan un papel fundamental en la fisiopatología de las enfermedades humanas hereditarias, incluido el SBr<sup>327</sup>. Los retos alrededor de la genética compleja del SBr, así como la evidencia que sugiere que este trastorno puede no ser causado por variantes genéticas únicas también han sido discutidos recientemente por Monasky y colaboradores<sup>249</sup>.

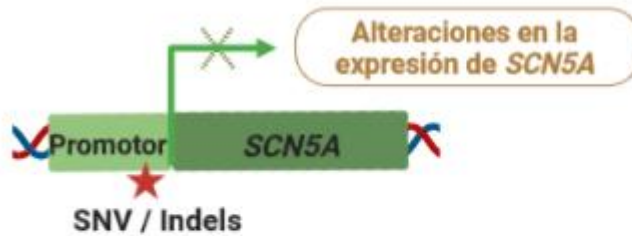
### 7.1. El rol de las variantes no codificantes en el Síndrome de Brugada

Desde la descripción clínica del SBr hace 30 años, se han llevado a cabo múltiples estudios para dilucidar la base genética y la fisiopatología del SBr. A pesar de los grandes avances en este campo, todavía existe una brecha de conocimiento que debe abordarse para poder tener una comprensión robusta del tipo de herencia y la importancia de las variantes asociadas a SBr<sup>352</sup>. Durante los últimos años han surgido distintas evidencias sugiriendo que las variantes reguladoras no codificantes podrían desempeñar un papel importante en la susceptibilidad al SBr, al modular su fenotipo<sup>352</sup>. Muchos de estos estudios han identificado variantes intrónicas en



el locus *SCN5A–SCN10A* y han sugerido que podrían estar afectando los niveles de expresión de *SCN5A* y/o *SCN10A* en el corazón<sup>50,51</sup> (Figura 36 A-B). Sin embargo, aún falta una mayor

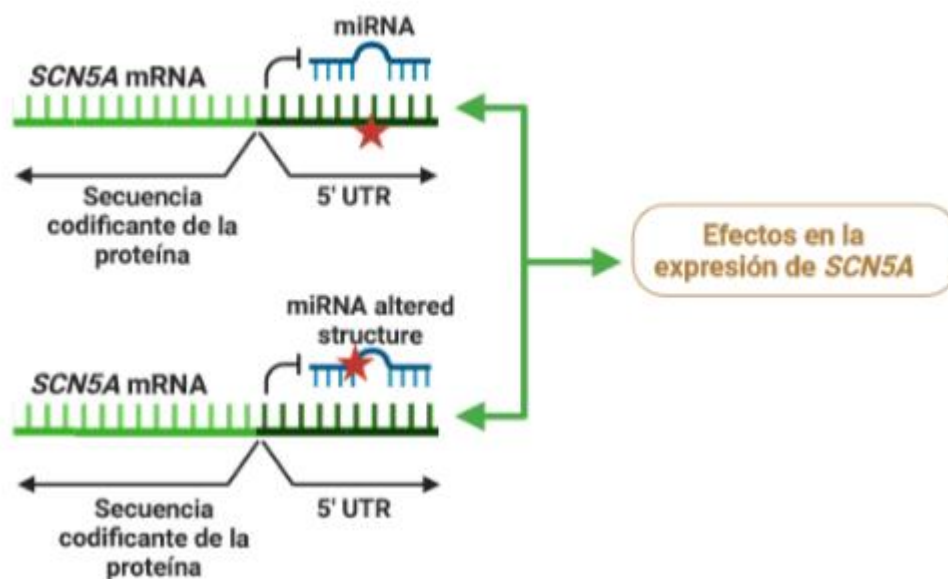
### A. Variantes promotor *SCN5A*



### B. Variantes *SCN10A*



### C. Variantes miRNA



**Figura 36. Efectos potenciales de las variantes no codificantes asociadas con el síndrome de Brugada (SBr).** (A) Las variantes en el promotor de *SCN5A* pueden provocar alteraciones en los niveles de expresión del mRNA. (B) Las variantes intrónicas en *SCN10A* podrían estar afectando la expresión de *SCN5A* a través de un mecanismo mediado por un *enhancer* o afectando la actividad de *SCN10A*. (C) Las variantes dentro de la región 5'UTR de *SCN5A*, o que afecten la función de los miRNAs relacionados con SBr, podrían provocar una desregulación de la expresión de *SCN5A*. La estrella roja indica la posición de una nueva variante de un solo nucleótido (SNV) o una inserción/delección (indel). La X verde indica posibles alteraciones de la expresión génica. Figura propia extraída de Pérez-Agustín *et al.*<sup>352</sup>.

comprensión de los mecanismos que explican cómo la desregulación de la expresión de *SCN5A* y/o *SCN10A* podría estar relacionada con alteraciones en el ECG y dar lugar a arritmias cardíacas.

Por otro lado, algunos estudios se han centrado en examinar el papel potencial de los miRNAs en la patogénesis del SBr, aunque sus hallazgos sugieren que las variantes dentro de *SCN5A* 5'UTR o los miRNA podrían estar asociadas con alteraciones en la expresión de *SCN5A* y la patogénesis del SBr<sup>329</sup> (**Figura 36C**).

En este escenario, es fundamental descifrar el efecto de las variantes no codificantes asociadas al SBr para comprender mejor la base genética de este trastorno cardíaco.

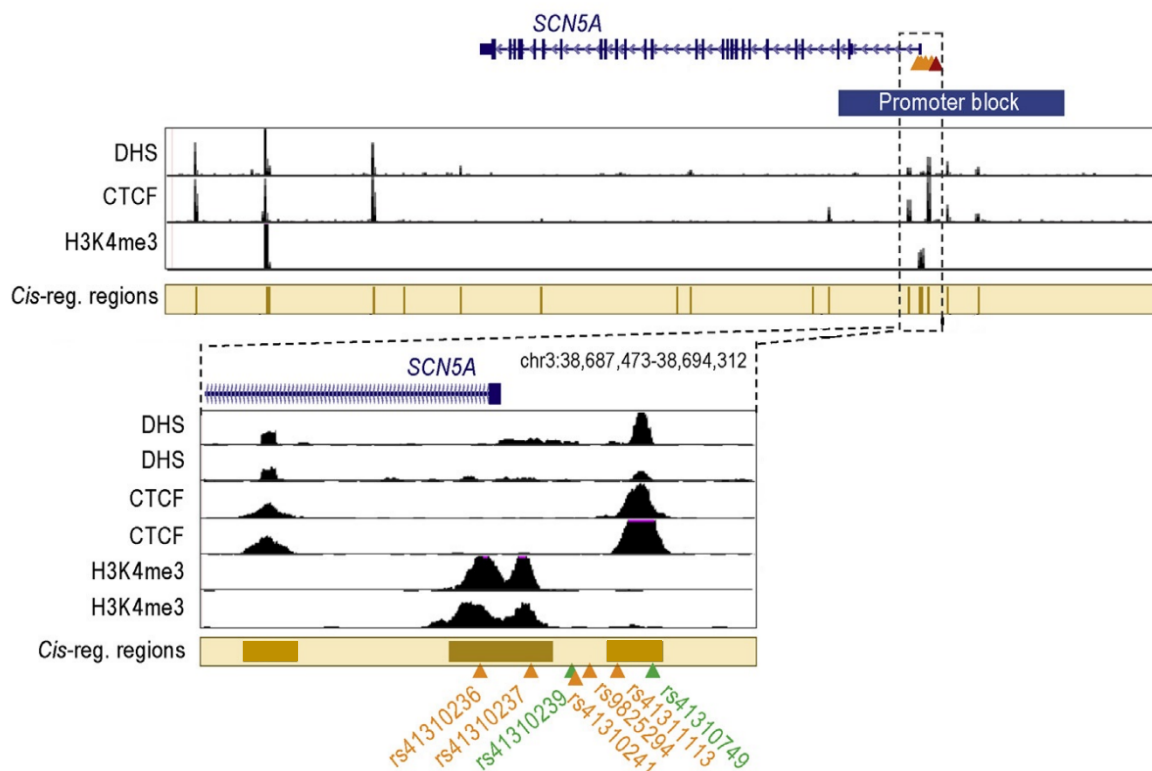
### 7.1.1. El promotor de *SCN5A*

El promotor central del gen *SCN5A* humano abarca desde los nucleótidos -261 a +140 en relación con el TSS, por lo que incluye parte del intrón 1 de *SCN5A*<sup>371</sup>. Nuestro laboratorio ha proporcionado información sobre los mecanismos de regulación de la expresión de *SCN5A* a nivel transcripcional. Los estudios realizados demostraron que los FT GATA4 y GATA5 coocupan el promotor de *SCN5A* y la primera región intrónica del gen, y lo hacen de manera sinérgica para regular la expresión del gen *SCN5A* en el corazón humano<sup>157</sup>. De forma similar, el estudio de Al Sayed *et al.* demostró que IRX5 y GATA4 actúan de forma sinérgica para activar el promotor de *SCN5A*<sup>154,155</sup>.

Otros estudios han identificado SNVs comunes en el promotor central de *SCN5A* que se asocian con alteraciones en la actividad transcripcional de *SCN5A* (**Figura 37**). Bezzina y colaboradores describieron tres haplotipos (HapA-C) dentro de la región promotora de *SCN5A*, que consisten en diferentes combinaciones de seis SNVs comunes en desequilibrio de ligamiento (en inglés, *linkage disequilibrium* (LD)) casi completo (rs41311113, rs9825294, rs41310241, rs41310239, rs41310237 y rs41310236)<sup>332</sup> (**Figura 37**). Los loci están en LD cuando la frecuencia de asociación de sus diferentes alelos es mayor o menor de lo que se esperaría si los loci fueran independientes y se asociaran al azar. Ensayos reporteros luciferasa mostraron que HapB, común entre la población japonesa, está asociado con una disminución significativa de la actividad del promotor. Curiosamente, aunque el haplotipo está presente en el 22% de la cohorte japonesa, es raro o está ausente en otras poblaciones, lo que sugiere que los efectos asociados con este haplotipo podrían ser específicos de la población<sup>332</sup>. Posteriormente, Yang y colaboradores describieron un haplotipo en el promotor de *SCN5A* que consta de dos variantes comunes, rs41310749 y rs41310239<sup>333</sup> (**Figura 37**). Observaron que pacientes portadores del haplotipo en heterocigosis con una variante codificante en *SCN5A* con pérdida de función mostraban un fenotipo de SBr más grave<sup>334</sup>, lo que respalda la noción de un efecto modulador de las variantes no codificantes sobre la gravedad de la enfermedad. En la misma línea de evidencia, Yang y colaboradores identificaron SNVs dentro de la zona 5' UTR, el exón

1 y el intrón 1 de *SCN5A* en una cohorte de SBr caucásico. Estas variantes se encuentran dentro de TFBS y afectan la unión de FT específicos. De manera similar, un estudio reciente de Yagihara *et al.* identificó nuevas variantes raras dentro de la región central del promotor *SCN5A* en pacientes con SBr de origen japonés y observó que tres de estas variantes están asociadas con una actividad promotora reducida<sup>335</sup>.

Un ejemplo interesante de una variante del promotor de *SCN5A* es rs1805124, que se ha descrito como un modulador genético protector. Se ha demostrado que esta variante común disminuye la tasa de metilación del promotor, lo que aumenta los niveles de expresión de *SCN5A*. Como resultado, los portadores de rs1805124 presentan patrones de ECG con menor gravedad, mayor actividad de los canales de sodio y fibrilación ventricular menos frecuente<sup>336</sup>. En conjunto, estos estudios respaldan la idea de que las variantes en el promotor *SCN5A*, y especialmente aquellas que afectan a los TFBS, pueden tener un impacto en la expresión del gen *SCN5A* y los parámetros del ECG y, por lo tanto, modular la patogénesis del SBr.



**Figura 37. Ubicación genómica de SNVs comunes relevantes en el promotor de *SCN5A*.** Arriba se muestra la región genómica donde se localiza *SCN5A* con los picos correspondientes a DHS, CTCF y H3K4me3, además de representarse las regiones reguladoras en *cis*. Abajo se amplía la región promotora de *SCN5A* donde se encuentran las SNVs de los haplotipos descritos por Bezzina *et al.*<sup>332</sup> y Yang *et al.*<sup>333</sup>. Las SNVs pintadas de color amarillo están asociadas a cambios en la expresión de *SCN5A*, mientras que las verdes están asociadas a cambios en la expresión de *SCN5A* y también a SBr. Figura adaptada de Pinsach-Abuin y colaboradores<sup>351</sup>.

### 7.1.2. El locus *SCN5A-SCN10A*

Estudios de GWAS realizados con características de los ECGs y los trastornos de la conducción cardíaca han demostrado la asociación del locus *SCN5A-SCN10A* con la duración del QRS y el intervalo PR. Estos resultados demuestran un papel relevante de este locus en la conducción cardíaca y la función cardíaca<sup>337,338,352</sup>. Sorprendentemente, la mayoría de las variantes identificadas en estos estudios se encuentran en regiones no codificantes. Si bien se sabe que las variantes de *SCN5A* están asociadas a arritmias cardíacas, la identificación de *SCN10A* como una región de riesgo importante para los rasgos de ECG fue sorprendente.

El gen *SCN10A*, que codifica la subunidad  $\alpha$  del canal de sodio Nav1.8, fue originalmente descrito por su expresión en neuronas sensoriales nociceptivas y su papel importante en la percepción del dolor<sup>339-341</sup>. Estudios posteriores han demostrado que Nav1.8 contribuye a la electrofisiología cardíaca, pero el mecanismo molecular aún está en debate, especialmente en lo que respecta a su expresión en cardiomiocitos. Utilizando un modelo murino, Nav1.8 se ha localizado en neuronas intracardíacas, pero no en cardiomiocitos ventriculares<sup>342</sup>. Otros estudios han afirmado que Nav1.8 contribuye a la corriente de sodio cardíaca tardía<sup>343,344</sup> y han sugerido que Nav1.8 podría estar involucrado en la arritmogénesis<sup>345</sup>. Sin embargo, un estudio reciente mostró la ausencia de canales Nav1.8 funcionales en células cardíacas de conejo y hiPSC-CM<sup>345</sup>.

Bezzina y colaboradores proporcionaron más evidencia de la relevancia del locus *SCN5A-SCN10A* con un GWAS realizado en una cohorte de SBr, con 312 pacientes con SBr de ascendencia europea. En su estudio, identificaron tres variantes comunes independientes (rs11708996, **rs10428132**, rs9388451) en tres loci diferentes (*SCN5A*, *SCN10A* y *HEY2*). La asociación más fuerte corresponde a rs10428132, una variante ubicada dentro del intrón 14 del gen *SCN10A*<sup>346</sup>. La segunda asociación más fuerte, rs9388451, se detectó *downstream* de *HEY2*, que codifica por un FT cardíaco que regula la actividad eléctrica cardíaca y posiblemente esté involucrado en la patogenia de SBr<sup>346,347</sup>. La tercera asociación más fuerte, rs11708996, está ubicada dentro de una región intrónica de *SCN5A*. Esta variante ya se había relacionado previamente con alteraciones en los intervalos QRS y PR y, por tanto, con la conducción cardíaca<sup>333,348</sup>. Bezzina y colaboradores demostraron que estos tres loci tienen un gran efecto acumulativo sobre el riesgo de SBr, ya que los pacientes con más de cuatro alelos de riesgo presentaban una asociación más fuerte con el SBr en comparación con los pacientes que portaban menos de dos alelos de riesgo. Los autores también encontraron que el 1.5% de la población europea alberga más de cuatro alelos de riesgo, lo que sugiere que estas tres variantes comunes actúan como moduladores de la susceptibilidad a la enfermedad en lugar de ser variantes causales<sup>346</sup>.

Los estudios realizados por van den Boogaard y colaboradores revelaron otra posible función de las variantes no codificantes de *SCN10A* en las alteraciones de la conducción. En particular, demostraron que la variante intrónica de *SCN10A* **rs6801957**, identificada en un GWAS asociado

con la conducción cardíaca, se ubica en una región *enhancer* que modula la expresión del gen *SCN5A*. También comprobaron que rs6801957 altera la unión de los FT TBX3 y TBX5 a este *enhancer* intrónico en *SCN10A*, lo que da como resultado niveles reducidos de expresión de *SCN5A*. Además, también observaron que esta variante en homocigosis está asociada con una expresión disminuida de *SCN5A* en muestras del ventrículo izquierdo<sup>50,51,143</sup>.

El estudio de Man *et al.* ha reafirmado recientemente el papel crucial del locus *SCN5A-SCN10A* para el control de la actividad y la organización tridimensional del gen, a través del estudio de un *superenhancer* situado *downstream* de *SCN5A* (región intergénica entre el gen *EXOG* y *SCN5A*). Descubrieron que todos los componentes reguladores que interactúan en el locus son necesarios para la expresión adecuada de *SCN5A* y la conducción cardíaca normal. Asimismo, los resultados de este estudio sugieren que las variantes que afectan al *superenhancer* pueden afectar la arquitectura de la cromatina del locus *SCN5A-SCN10A* y, en consecuencia, la expresión de *SCN5A*<sup>349</sup>.

En resumen, los estudios descritos demuestran que la región intrónica de *SCN10A* contribuye a la electrofisiología cardíaca y que las variantes no codificantes dentro de esta región pueden asociarse con SBr u otras características del ECG<sup>50</sup>. Sin embargo, la base molecular que subyace a los efectos de las variantes comunes en *SCN10A* aún no está clara, ya que el papel de Nav1.8 en la función cardíaca y su expresión en los cardiomiocitos sigue siendo controvertido<sup>344,350,352</sup>.

### 7.1.3. RNAs no-codificantes

Durante la última década, se ha demostrado que los miRNAs desempeñan funciones importantes en el desarrollo y la función del corazón. Entre estos, miR-1, miR-25, miR-26, miR-29 y miR-328 se han asociado específicamente a arritmias cardíacas<sup>353-357</sup>. Los estudios funcionales han demostrado que miR-24, miR-98, miR-106, miR-200, miR-219 y miR-1270 regulan la expresión de *SCN5A*<sup>329,358-360</sup>, mientras que los miRNAs como miR-125 y miR-153 han sido descritos como reguladores indirectos de *SCN5A*<sup>358</sup>.

Actualmente solo hay unos pocos estudios que hayan identificado variantes que afectan la función de los miRNAs que se dirigen a *SCN5A* y que, por lo tanto, podrían estar asociados con SBr. Daimi y colaboradores identificaron dos variantes comunes (rs4073796 y rs4073797) dentro de la región 5' UTR de *SCN5A* en una familia tunecina con SBr, que es, hasta la fecha, el principal estudio que relaciona variantes asociadas a miRNAs con SBr<sup>329</sup>. En particular, demostraron que juntos, rs4073797 y rs4073796 crean un sitio de unión adicional para miR-1270, lo que conduce a una regulación a la baja de la expresión de *SCN5A*.

En conjunto, estas observaciones sugieren que las variantes no codificantes que afectan la función de los miRNAs también podrían explicar la base molecular de las arritmias cardíacas en las que aún se desconoce la etiología, incluido el SBr.

### 7.1.4. Haplotipos asociados a Síndrome de Brugada

Desde hace más de diez años, nuestro grupo está interesado en identificar variantes no codificantes asociadas al SBr en regiones reguladoras. El estudio de Pinsach-Abuin *et al.* sirvió para descubrir varios haplotipos asociados a la enfermedad y evidenciar la importancia de estudiar como la presencia de determinadas variantes comunes podría modular el riesgo asociado a la enfermedad<sup>351</sup>.

Para llevar a cabo el estudio (**Figura 38**), se seleccionó una cohorte de 86 pacientes que fueron diagnosticados con SBr en base al ECG y eran negativos para variantes en el gen *SCN5A*, con el propósito de enriquecer los casos de SBr con causas genéticas desconocidas. A continuación, se utilizó un método de secuenciación de alto rendimiento (*nextera rapid capture-resequencing* o captura de las regiones preseleccionadas seguida de secuenciación) para secuenciar de manera específica regiones no codificantes preseleccionadas cercanas a los genes asociados con SBr. Después de un extenso análisis bioinformático de los datos de

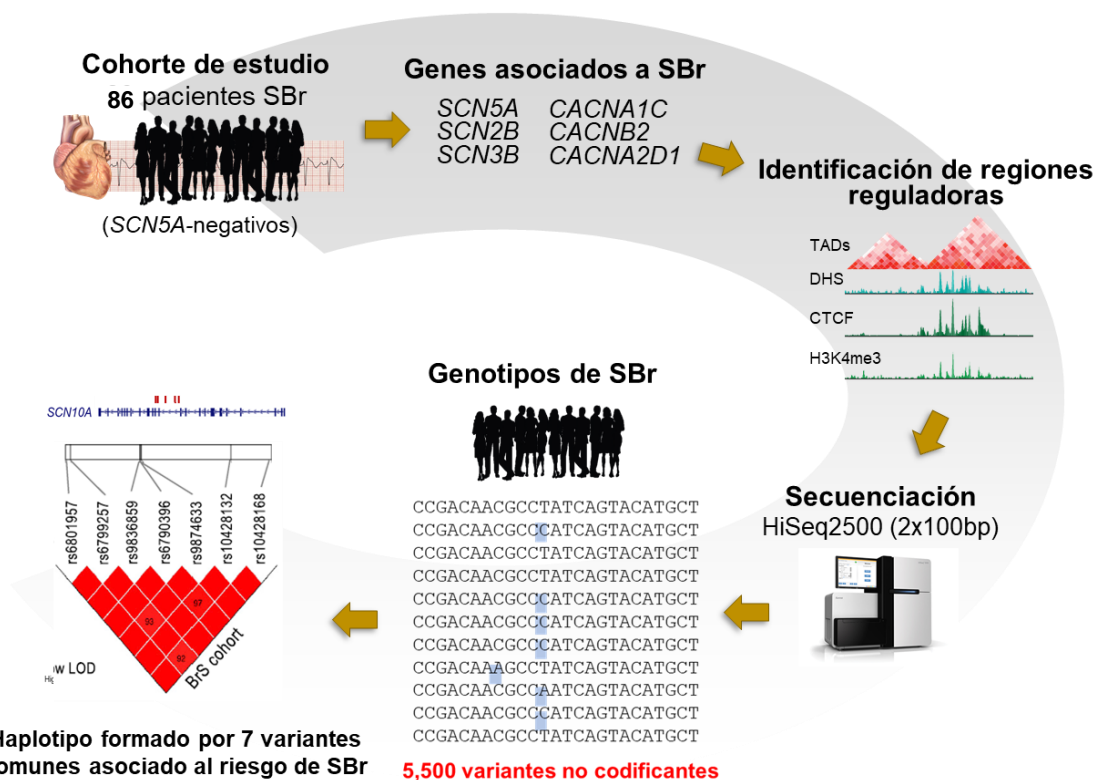


Figura 38. Procedimiento seguido por Pinsach-Abuin y colaboradores.<sup>351</sup> para la determinación de variantes no codificantes asociadas a SBr.

secuenciación, se identificaron un total de 5 500 variantes no codificantes en los pacientes con SBr<sup>351</sup>.

El análisis de estas variantes por parte de Pinsach-Abuin y colaboradores reveló 5 haplotipos comunes de 7 SNVs (Hap1-5) con 2 combinaciones asociadas con arritmia potencialmente mortal: SBr (los genotipos Hap<sup>1/1</sup> de riesgo y Hap<sup>2/3</sup> protector)<sup>351</sup>. Las 7 SNVs se distribuyen a lo largo de una región genómica de 12 477 pb. Por este motivo, para confirmar experimentalmente el *phasing* de estos haplotipos fue necesario secuenciar los 86 pacientes mediante LRS. Para ello, en esta tesis se realizaron estas validaciones mediante el uso de secuenciación de ONT. La información relacionada con estos experimentos se amplía en el apartado 3.1. de Resultados.









## **II. Justificación y objetivos**



## Justificación

El SBr es una enfermedad cardíaca hereditaria con elevada susceptibilidad a las arritmias ventriculares y MSC. El gen *SCN5A*, que codifica la subunidad  $\alpha$  del canal de sodio cardíaco, fue el primer gen asociado al SBr y sigue siendo el principal gen relacionado con la patogénesis de la enfermedad. Las variantes genéticas en las regiones codificantes del gen *SCN5A* se han asociado en aproximadamente el 30% de los casos de SBr<sup>323,324</sup>. Sin embargo, más del 65% de los casos diagnosticados no pueden explicarse por variantes en *SCN5A* u otros genes asociados. Por lo tanto, en un número importante de casos, los mecanismos subyacentes del síndrome siguen siendo desconocidos. Actualmente, distintos autores hipotetizan que las regiones no codificantes podrían estar involucradas en la etiología del SBr en todos estos casos, principalmente mediante las regiones con funciones de regulación de la expresión génica<sup>352</sup>.

La implicación de las regiones reguladoras en la enfermedad puede abordarse desde diferentes perspectivas/estrategias:

- **La interacción de FT con los elementos reguladores y sus efectos sobre la expresión génica.** En esta tesis se pretende continuar con los estudios previos realizados por la Dra. Tarradas<sup>18</sup> en relación al FT cardíaco GATA4. Concretamente, se pretende identificar los lugares específicos de acetilación de GATA4, así como sus efectos. Desde la publicación del artículo de Takaya *et al.*<sup>163</sup>, en el que se apuntaba a que p300 acetila las lisinas K311, K318, K320 y K322 de GATA4, la comunidad científica que estudia GATA4 ha dado como asumidos estos cuatro residuos como los sitios específicos de acetilación de GATA4. Sin embargo, Takaya *et al.* se basaba solamente en estudios funcionales realizados con mutantes de GATA4. Gracias a los experimentos de acetilación *in vitro* de la Dra. Tarradas (apartado 1.1. de Resultados) sabemos que GATA4 presenta tres lugares de acetilación en la región básica adyacente a C-Zn, del dominio de unión al DNA de GATA4, y no cuatro como sugería Takaya *et al.* Concretamente, los estudios de la Dra. Tarradas demostraron que existen tres lugares de acetilación que estarían entre las lisinas K318, K320, K322, K326 y K328: dos se encontrarían entre los residuos K318, K320 y K322, mientras el otro residuo acetilado sería K326. Por tanto, falta resolver qué dos residuos de entre K318, K320 y K322 son acetilados por p300.

Dada la importancia de GATA4 en la regulación del gen *SCN5A*, y partiendo del conocimiento actual, esta tesis tiene como objetivo principal identificar las lisinas de GATA4 que son acetiladas por p300, así como determinar el efecto de la acetilación de GATA4 en su función de activador transcripcional.

- **La caracterización de regiones reguladoras involucradas en la expresión de *SCN5A*.** Este trabajo parte de la hipótesis de que entender los mecanismos de regulación transcripcionales de *SCN5A* permitirá descubrir nuevos elementos determinantes de las

enfermedades cardíacas relacionadas con el canal de sodio. Para evaluar esta hipótesis, este trabajo pretende identificar y caracterizar nuevos elementos reguladores en *cis* del gen *SCN5A*. A pesar de que ya se conocen algunos elementos reguladores de *SCN5A*, la red transcripcional de este gen aún no se conoce con detalle. La caracterización de nuevas regiones reguladoras en el locus *SCN5A-SCN10A* puede contribuir a entender mejor las bases moleculares de la regulación de *SCN5A*, además de convertirse en regiones con potencial diagnóstico para variantes patogénicas.

- **El efecto de las variantes genéticas en estas regiones reguladoras.** Las variantes comunes, en su mayoría ubicadas en regiones no codificantes, han surgido como moduladores potenciales de la enfermedad al afectar diferentes mecanismos reguladores, incluidos los FT o la organización tridimensional del genoma. Pinsach-Abuin *et al.*<sup>351</sup> identificó un bloque de haplotipos con siete SNVs que se localizan en el locus *SCN5A-SCN10A* y están asociadas a distintas susceptibilidades al SBr. Además, también se ha descrito que estas SNVs se localizan en elementos reguladores en *cis* que modulan la expresión de *SCN5A*. En esta tesis se pretende confirmar experimentalmente el *phasing* de los haplotipos descubiertos por la Dra. Pinsach. Además, con el objetivo de examinar las posibles implicaciones de este haplotipo en la patogénesis SBr, se pretende generar un modelo de hiPSCs a partir de un paciente con SBr y haplotipo de riesgo. Con este modelo se estudiaría el perfil transcripcional de este paciente en relación con un control sano. Además, el modelo celular ofrece la oportunidad de estudiar el efecto de las variantes del haplotipo en un contexto isogénico. Para ello, se propone editar la línea con Hap<sup>1/1</sup> mediante la tecnología CRISPR/Cas9, con el objetivo de obtener una línea con idéntico *background* genético, que sólo se diferencie en los cambios nucleotídicos asociados al haplotipo.

## Objetivos

En el presente trabajo, nuestro objetivo general/principal es entender los mecanismos de regulación transcripcionales de *SCN5A* y sus posibles implicaciones en la etiología del SBr. Con este fin, nos proponemos desarrollar los siguientes objetivos:

1. Identificar las lisinas de GATA4 que son acetiladas por p300, y determinar el efecto de la acetilación de GATA4 en su función de activador transcripcional del gen *SCN5A*.
2. Identificar y caracterizar nuevos elementos reguladores en *cis* del gen *SCN5A* en el locus *SCN5A-SCN10A* utilizando un modelo celular de hiPSC-CMs con el sistema CRISPRi integrado.
3. Validar experimentalmente los haplotipos en pacientes con SBr con secuenciación de lecturas largas (Oxford Nanopore Sequencing) y estudiar su efecto sobre el transcriptoma.
4. Estudiar el efecto del haplotipo de riesgo al SBr sobre la expresión génica y el sodio intracelular en un modelo de hiPSC-CMs.
5. Diseñar una estrategia de edición genética para la reversión del haplotipo de riesgo a SBr a un haplotipo no de riesgo mediante la tecnología CRISPR/Cas9.





### **III. Materiales y métodos**





## 1. Materiales

### 1.1. Muestras

#### 1.1.1. Cohorte de pacientes con síndrome de Brugada

La cohorte de pacientes de estudio está formada por 96 individuos no emparentados diagnosticados en base a un ECG positivo para SBr (patrón de ECG tipo I), ya sea espontáneo o inducido por la administración de fármacos como la flecainida o la ajmalina. Todos los pacientes seleccionados son *SCN5A*-negativos, por lo que carecen de variantes genéticas en el gen *SCN5A*, que codifica por Nav1.5. El análisis genético para la selección de pacientes *SCN5A*-negativos fue previo a esta tesis y se realizó en el Centro de Genética Cardiovascular (CGC; Girona), mediante secuenciación Sanger convencional o secuenciación de última generación (del inglés, Next Generation Sequencing, NGS).

Este estudio se realizó de acuerdo con los Principios de la Declaración de Helsinki de 2008. Todos los pacientes firmaron un consentimiento informado por escrito para participar en el estudio (Anexo 1) y todos los procedimientos fueron aprobados por el comité ético del Hospital Universitario Dr. Josep Trueta de Girona (España).

#### 1.1.2. Individuos para la generación de modelos iPSC

Con la finalidad de obtener modelos celulares de hiPSC para el haplotipo asociado a SBr y para un caso control se tomaron muestras sanguíneas de 2 individuos:

- Hombre de 68 años diagnosticado con SBr (de la cohorte de 96 pacientes), específicamente homocigoto para el haplotipo 1 (Hap<sup>1/1</sup>).
- Hombre de 55 años sin antecedentes de ninguna patología, con haplotipo heterocigoto (Hap<sup>1/3</sup>).

Los haplotipos de los individuos se determinaron mediante secuenciación de ONT (apartado 2.9.2. de Métodos). El Centro de Medicina Regenerativa de Barcelona (CMRB) se encargó de recibir las muestras y obtuvo hiPSC mediante el método de reprogramación celular (apartado 2.2. de Métodos).

## 1.2. Modelos celulares experimentales

### 1.2.1. HEK293T

La línea celular HEK293T (*Human Embryonic Kidney*) deriva de células de riñón de embrión humano. Expresan de forma estable el antígeno-T SV40, el cual se puede unir al origen de replicación SV40 de los vectores transfectados, aumentando así su replicación. Se obtuvieron de la *Health Protection Agency Culture Collections-Salisbury*, UK. Estas células se utilizaron para el análisis de los niveles de acetilación de GATA4.

### 1.2.2. H9c2

Esta línea celular es un subclon derivado del clon original BDXI establecido por Kimes y Brandt en 1976, que se originó a partir de células de ventrículo embrionario de rata. Se obtuvieron células cardíacas derivadas de ventrículo de rata embrionaria (células H9c2) de la Colección Americana de Cultivos Tipo (ATCC). Esta línea se utilizó para evaluar el efecto de la acetilación de GATA4 sobre la actividad del promotor de *SCN5A*.

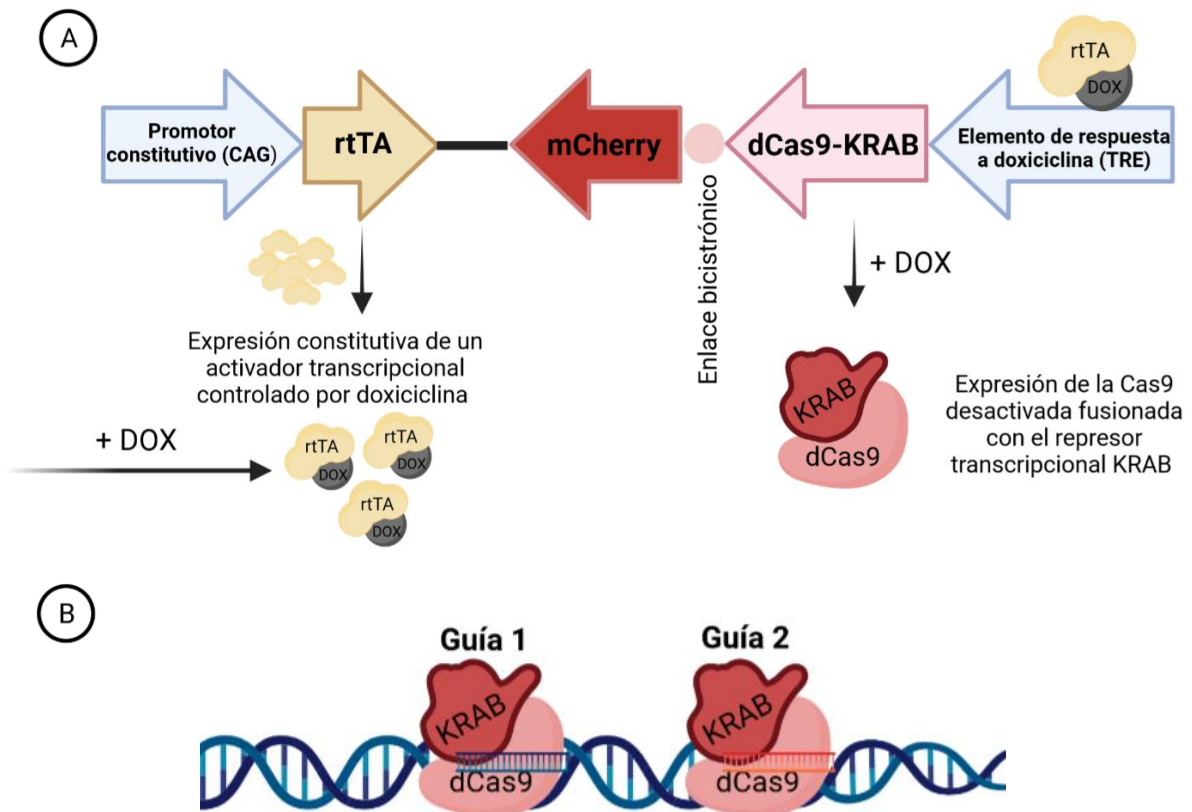
### 1.2.3. WTB CRISPRi Gen1B

CRISPRi WTB Gen1B es una línea celular de hiPSC diseñada por Mandegar *et al.* con el fin de obtener una herramienta funcional, versátil y modulable para la realización de cribados basados en CRISPRi<sup>220</sup>. Esta línea fue proporcionada por el Dr. Bruce Conklin (Gladstone Institutes, San Francisco, CA, EE. UU).

Esta línea celular tiene integrado un constructo que regula la expresión de dCas9-KRAB (Cas9 desactivada y fusionada al dominio de represión transcripcional KRAB) en un locus seguro (AAVS1). A diferencia de la Cas9 salvaje o *wild type* (WT), que genera un corte de doble cadena en el DNA, dCas9 solamente tiene la capacidad de unirse a la secuencia a donde se dirija. La expresión del transgén está bajo el control de un sistema inducible regulado por el activador transcripcional de respuesta a doxiciclina (rtTA), gracias al cual se mantiene silenciada en el tiempo hasta añadir doxiciclina (DOX) en el medio de cultivo (**Figura 39A**). La activación del elemento de respuesta a DOX (TRE) mediante el rtTA fusionado a DOX da lugar a la expresión bicistrónica de dCas9-KRAB y la proteína fluorescente mCherry. Se ha comprobado que la DOX no tiene efectos citotóxicos en esta línea celular, ni disminución en su proliferación o cambios en su morfología, por lo que se considera una potente herramienta para reprimir la expresión génica<sup>220</sup>.

La ventaja de esta herramienta es que dCas9-KRAB puede ser dirigida a cualquier zona del genoma mediante guías (gRNAs) y ejercer su función de represor transcripcional en la región donde se haya unido (**Figura 39B**). En esta tesis se ha usado esta herramienta para hacer un

cribaje de nuevas zonas reguladores del genoma que podrían alterar la expresión de *SCN5A* (experimento de *knockdown*).



**Figura 39. Funcionamiento de la línea WTB CRISPRi Gen1B.** (A) Constructo integrado en el locus AAVS1 de la línea WTB CRISPRi Gen1B. La doxiciclina se une al activador transcripcional (rtTA), permitiéndole a este activar el elemento de respuesta a doxiciclina (TRE), que a su vez promueve la expresión bicistónica de dCas9-KRAB y mCherry. (B) Representación esquemática de la unión de dos gRNAs a la proteína dCas9-KRAB, que se une a dos regiones diferentes del genoma. Figura creada con BioRender (<https://app.biorender.com/>).

#### 1.2.4. Ctrl BS PbiPS37-Sv4F-1 y BS PbiPS46-Sv4F-10

- Ctrl BS PbiPS37-Sv4F-1 (GpG2) es la línea celular de hiPSC generada a partir de PBMCs de un donante sano y con haplotipo heterocigoto Hap<sup>1/3</sup>.
- BS PbiPS46-Sv4F-10 (BS hiPSC) es la línea celular de hiPSC generada a partir de PBMCs del paciente con SBr y haplotipo Hap<sup>1/1</sup>.

Ambas líneas fueron generadas por el CMRB mediante el método del virus Sendai y depositadas en el banco nacional de líneas celulares a fecha de 8 de febrero del 2021.

### 1.2.5. BS PbiPS Cas9-*induced*

BS PbiPS Cas9-*induced* es una línea celular modificada a partir de la línea BS PbiPS46-Sv4F-10. Esta línea fue creada y proporcionada por Sara Veiga (Laboratorio de Genómica y Bioinformática del CIMUS, Santiago de Compostela, España).

Tiene las mismas características descritas para BS PbiPS46-Sv4F-10, pero se le añadió un constructo que le permite expresar Cas9 en presencia de doxiciclina. Esta línea fue creada para probar su eficacia en ediciones del haplotipo asociado a SBr mediante CRISPR/Cas9.

### 1.2.6. Cardiomiocitos derivados de hiPSC (hiPSC-CM)

Los cardiomiocitos con los que hemos trabajado se generan en nuestro laboratorio a partir de las líneas de hiPSC disponibles. Mediante un proceso de diferenciación de 30 días se obtienen cardiomiocitos con cierto grado de maduración y con capacidad de latir por sí mismas (apartado 2.3.4. de Métodos). Los hiPSC-CM generados presentan expresión de marcadores cardíacos e inhiben la expresión de los marcadores de pluripotencia detectables en las hiPSC (apartado 2.3. de Resultados).

Estas células se han utilizado para realizar un estudio de *RNA-sequencing* (RNA-seq) con los hiPSC-CM generadas a partir de las líneas Ctrl hiPSC y SBr hiPSC. También se generaron hiPSC-CM a partir de la línea WTB CRISPRi Gen1B para los experimentos de *knockdown* de *SCN5A*.

## 1.3. Vectores de expresión

### 1.3.1. Vectores reporteros luciferasa

Para los ensayos reporteros luciferasa se utilizaron los genes reporteros luciferasa *firefly* y *renilla*. Dependiendo del diseño del ensayo luciferasa se utilizaron distintas combinaciones de los vectores descritos a continuación:

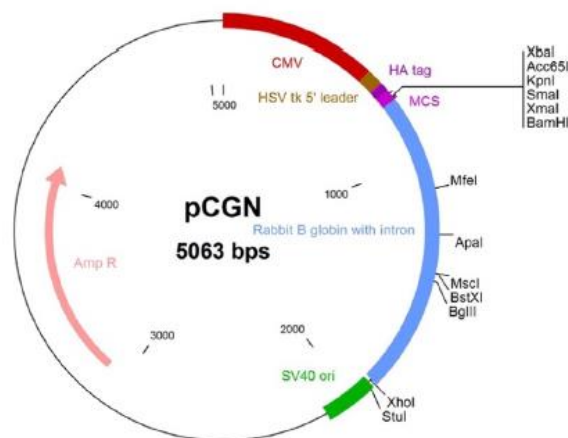
- **pGL3-SCN5A-luciferasa (SCN5A-*firefly*)**: vector reportero con el gen de la luciferasa *firefly* controlado por la región promotora del gen *SCN5A* humano correspondiente a los nucleótidos -260 a +612 pb en relación al sitio de inicio de la transcripción (del inglés, TSS). Este es el promotor que en el trabajo de Tarradas *et al*<sup>157</sup>. Se denomina "*Promoter B*". El vector confiere resistencia a ampicilina. Fue proporcionado por el Dr. Dan Roden (Vanderbilt University, Nashville, TN, EE.UU.).
- **pGL3-PDX1-luciferasa (PDX1-*firefly*)**: vector reportero con el gen de la luciferasa *firefly* controlado por el promotor del gen homeobox 1 pancreático y duodenal (*PDX1*) de ratón. El

vector confiere resistencia a ampicilina. Fue proporcionado por la Dra. Anabel Rojas (CABIMER, Sevilla, España).

- **pEF1- $\alpha$ -renilla** vector reportero con el gen de la luciferasa *renilla* controlado por el promotor constitutivo del factor de elongación 1 humano (*EF1 $\alpha$* ). El vector confiere resistencia a ampicilina. El vector fue proporcionado por la Dra. Melanie Ott (Gladstone Institutes, San Francisco, EE.UU.).

### 1.3.2. Vectores de expresión de GATA4

- **pCGN-GATA4/HA**: vector de expresión de la proteína GATA4 de ratón fusionada con la proteína hemaglutinina (HA), que permite la detección y/o purificación de la proteína fusionada. Este vector es de tipo pCGN (**Figura 40**) que contiene el promotor del citomegalovirus (CMV) para controlar la expresión del transgén, y el origen de replicación SV40, ideal para la replicación y aumento de DNA plasmídico en cultivos bacterianos. Además, otorga resistencia a la ampicilina con el gen Amp R. El vector pCGN-GATA4/HA se utilizó como molde para generar versiones con mutaciones puntuales en GATA4 mediante mutagénesis dirigida. Fue proporcionado por el Dr. Jonathan A. Epstein (University of Pennsylvania, Philadelphia, PA, EE.UU.).



**Figura 40. Mapa del plásmido pCGN (Promega).** Figura extraída de Promega (<https://www.promega.es>).

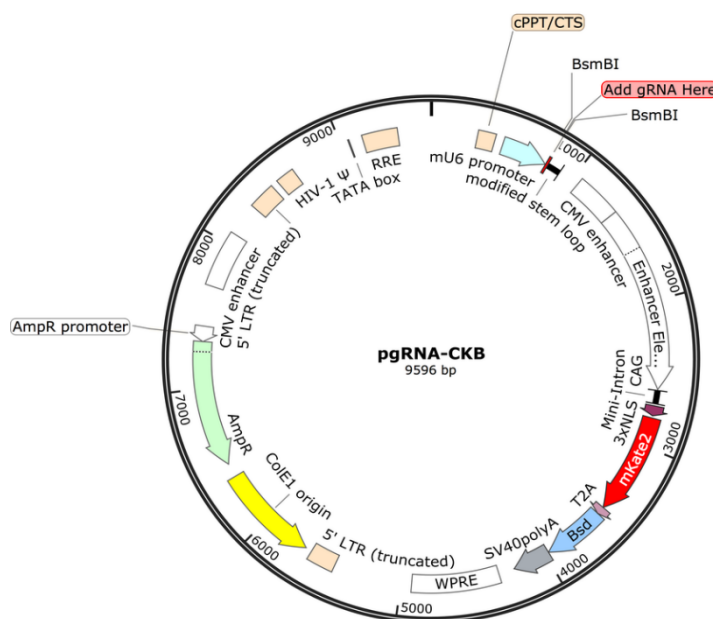
- **pCGN-GATA4-K318A/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4/HA. La lisina K318 de la proteína GATA4 está mutada a alanina.

- **pCGN-GATA4-K320A/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4/HA. La lisina K320 de la proteína GATA4 está mutada a alanina.
- **pCGN-GATA4-K322A/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4/HA. La lisina K322 de la proteína GATA4 está mutada a alanina.
- **pCGN-GATA4-K326A/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4/HA. La lisina K326 de la proteína GATA4 está mutada a alanina.
- **pCGN-GATA4-K328A/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4/HA. La lisina K328 de la proteína GATA4 está mutada a alanina.
- **pCGN-GATA4-DM(K322A/K326A)/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4-K326A/HA. Las lisinas K322 y K326 de la proteína GATA4 están mutadas a alaninas.
- **pCGN-GATA4-TM(K318A/K322A/K326A)/HA**: vector de expresión de GATA4 generado por mutagénesis dirigida a partir del vector pCGN-GATA4-DM(K322A/K326A)/HA. Las lisinas K318, K322 y K326 de la proteína GATA4 están mutadas a alaninas.
- **pCGN-GATA4-KA**: vector de expresión de GATA4 que presenta los residuos K311, K318, K320 y K322 mutados a alanina. Proporcionado por el Dr. Jonathan A. Epstein (University of Pennsylvania, Philadelphia, PA, EE.UU.).
- **pcDNA 3.1 (+)**: vector básico diseñado para producir altos niveles de expresión del fragmento de cDNA clonado, gracias a la presencia del promotor de CMV. Confiere resistencia a ampicilina. En este trabajo se ha utilizado como vector vacío en experimentos de transfección para igualar la cantidad total de DNA transfectado en las distintas condiciones. Este vector se adquirió en Invitrogen (Carlsbad, CA, EE.UU.).

### 1.3.3. Vectores de expresión de gRNAs

Para la expresión de gRNAs específicos se utilizó el plásmido pgRNA-CKB (Addgene plasmid #73501) (**Figura 41**). Este plásmido de 9596 pb, contiene una secuencia de 16 nucleótidos

(GGAGACGGACGTCTCC) con 2 sitios de restricción BsmBI entre los cuales se clonan los gRNAs. El plásmido contiene también dos secuencias de resistencia a antibióticos: (1) una a Amp R, útil para seleccionar aquellas bacterias que hayan incorporado el plásmido durante el proceso de transformación (apartado 2.7.3.2. de Métodos) y, (2) una de resistencia a blasticidina, utilizada para seleccionar las hiPSCs que hayan incorporado el plásmido durante el proceso de nucleofección (apartado 2.7.4 de Métodos).



**Figura 41. Mapa del plásmido pgRNA-CKB.** Se señalan los sitios de restricción BsmBI entre los que se inserta la gRNA (marcado en rojo como “Add gRNA Here”). Figura extraída de Addgene (<https://www.addgene.org>).

## 1.4. Cebadores

Todos los cebadores utilizados en esta tesis fueron sintetizados en CONDA labs (Madrid).

### 1.4.1. Cebadores para mutagénesis dirigida

Los cebadores para realizar la mutagénesis dirigida se diseñaron con el programa *QuikChange Primer Design* (Agilent Technologies). Por cada mutación que se quería introducir, se utilizó una pareja de cebadores, que contenían la mutación en cuestión y se hibridaban en la misma secuencia de DNA, pero en cadenas opuestas del vector (**Tabla 1**). La mutación correspondía al cambio de una lisina (AAA, AAG) a alanina (GCA, GCG), impidiendo la acetilación del residuo modificado.



**Tabla 1. Cebadores para la mutagénesis dirigida.** Secuencias sentido (Fw) y antisentido (Rv) de las parejas de cebadores empleadas para mutar las lisinas a alaninas. Subrayados los nucleótidos mutados.

Región mutada	Pareja de cebadores	Vector resultante
K318A del vector pCGN-GATA4/HA	(Fw) 5' GGAAGGAGGGGATTCAAACCAGA <u>GCACGGAAGCCCAA</u> -3' (Rv) 5' TTGGGCTTCCG <u>TGCTCTGGTTTG</u> AATCCCCTCCTTCC-3'	pCGN-GATA4-K318A/HA
K320A del vector pCGN-GATA4/HA	(Fw) 5' GGGATTCAAACCAGAAAACGGGC <u>GCCCAAGAATCTGAATAAATC</u> -3' (Rv) 5' GATTTATTCAGATTCTTGGG <u>CGC</u> CCGTTTTCTGGTTTGAATCCC-3'	pCGN-GATA4-K320A/HA
K322A del vector pCGN-GATA4/HA	(Fw) 5' TTCAAACCAGAAAACGGAAGCCC <u>GCGAATCTGAATAAATCTAAGACAC</u> -3' (Rv) 5' GTGTCTTAGATTTATTCAGATTC <u>GCGGGCTTCCGTTTTCTGGTTTGAA</u> -3'	pCGN-GATA4-K322A/HA
K326A del vector pCGN-GATA4/HA	(Fw) 5' GGAAGCCCAAGAATCTGAAT <u>GCA</u> TCTAAGACACCAGCAGGT-3' (Rv) 5' ACCTGCTGGTGTCTTAGAT <u>GCA</u> T TCAGATTCTTGGGCTTCC-3'	pCGN-GATA4-K326A/HA
K328A del vector pCGN-GATA4/HA	(Fw) 5' GGAAGCCCAAGAATCTGAATAAA <u>TCTGCAACACCAGCAGGT</u> -3' (Rv) 5' ACCTGCTGGTGT <u>TGCAGATTTAT</u> TCAGATTCTTGGGCTTCC-3'	pCGN-GATA4-K328A/HA
K322A del vector pCGN-GATA4-K326A/HA	(Fw) 5' CAAACCAGAAAACGGAAGCCCGC <u>GAATCTGAATGCATCTAAGAC</u> -3' (Rv) 5' GTCTTAGATGCATTCAGATTC <u>CGC</u> GGGCTTCCGTTTTCTGGTTTG-3'	pCGN-GATA4-DM(K322A/K326A)/HA
K318A del vector pCGN-GATA4-DM(K322A/K326A)/HA	(Fw) 5' GGAAGGAGGGGATTCAAACCAGA <u>GCACGGAAGCCCGC</u> -3' (Rv) 5' GCGGGCTTCCG <u>TGCTCTGGTTTG</u> AATCCCCTCCTTCC-3'	pCGN-GATA4-TM(K318A/K322A/K326A)/HA

## 1.4.2. Cebadores para CRISPR/Cas9 y CRISPRi

### 1.4.2.1. Cebadores para los gRNAs de ensayos CRISPRi

Se diseñaron gRNAs de 20 nucleótidos de una sola hebra para dirigir el constructo dCas9-KRAB a las regiones de interés. En total, se diseñaron 11 gRNAs nombrados según la región a

la que dirigió la dCas9-KRAB (**Tabla 2**). Se añadió un extremo adicional de 4 nucleótidos a las gRNAs para formar extremos cohesivos compatibles con el plásmido pgRNA-CKB. Los gRNAs se hibridaron para generar gRNAs de doble hebra que posteriormente se clonaron en el plásmido.

**Tabla 2. Secuencias de los gRNA diseñados y localización de la diana.** En verde, las secuencias añadidas a cada hebra de gRNA.

Nombre gRNA	Secuencias (20pb)		Diana	Coordenadas de la diana
	5'-Cebador Fw-3'	5'-Cebador Rv-3'		
G1	TTGGGCATCCCGGTTTGTCAACCAG AAACTGGTGACAAACCGGGATGC		EXOG Promotor – Pico CTCF	chr3:38537735
G2	TTGGTCTTCTCATGGTGCGCCAGG AAACCTGGCGCACCATGAGAAGA		SCN5A Intrón 1 – Pico CTCF	chr3:38688255
G3	TTGGGAGTGATTTCGTTGGGCCCCA AAACTGGGGCCCCAACGAATCACTC		SCN5A Intrón 1 – Pico NKX2-5	chr3:38688796
G4	TTGGATGCAGGGATCGCTCCCCCG AAACCGGGGAGCGATCCCTGCAT		TSS SCN5A	chr3:38691125
G5	TTGGGGTGCTCCGCCCGCTCGGAG AAACCTCCGAGCGGGCGGAGCACC		TSS SCN5A	chr3:38691347
G6	TTGGGCTCCCACTATCGGCCAGAA AAACTTCTGGCCGATAGTGGGAGC		SCN5A 5'UTR – Pico CTCF	chr3:38692885
G7	TTGGCTGATACACTTGCTGACAAG AAACTTGTGTCAGCAAGTGTATCAG		SCN5A 5'UTR – Pico GATA4	chr3:38700496
G8	TTGGTGAAGGGCTTGATAACTTTG AAACCAAAGTTATCAAGCCCTTCA		SCN10A – Pico GATA4	chr3:38765801
G9	TTGGGCCAACACTTGTGAGGACCT AAACAGGTCTGACAAGTGTGGC		SCN10A – rs6801957	chr3:38767323
G10	TTGGGCACTTGTACCTCCTTAGA AAACTCTAAGGAGGTGACAAGTGC		SCN10A – rs10428132	chr3:38777554
G11	TTGGGCCAACACTTGTGAGGACCT AAACAGGTCTGACAAGTGTGGC		SCN10A – Pico CTCF	chr3:38780322

### 1.4.2.2. Cebadores para los experimentos de *knock-in* con CRISPR/Cas9

Para editar de manera precisa un SNV mediante HDR diseñamos una sgRNA (20 pb) para dirigir la Cas9 al lugar de corte, y un DNA donador de cadena simple (ssODN) como molde para introducir la variante deseada (**Tabla 3**). El ssODN tiene una longitud de 88 pb con dos brazos de homología de 40 pb cada uno. En la siguiente tabla se muestran los cebadores usados para editar la variante rs6801957:

**Tabla 3. Cebadores diseñados para la edición de la variante rs6801957.** En verde se marca la variante que se introducirá en la edición.

Nombre	Cebadores (5' – 3')
sgRNA	CTCTGCTGTTACCTAGACAA
ssODN	AGTGGCAGACAGCTGTACAGAGTTCATGCTCTCTGCTGT <b>C</b> ACC TAGACAAAGGCATGTAAAATATGGCAGCTGATAACAGCAGCTGC

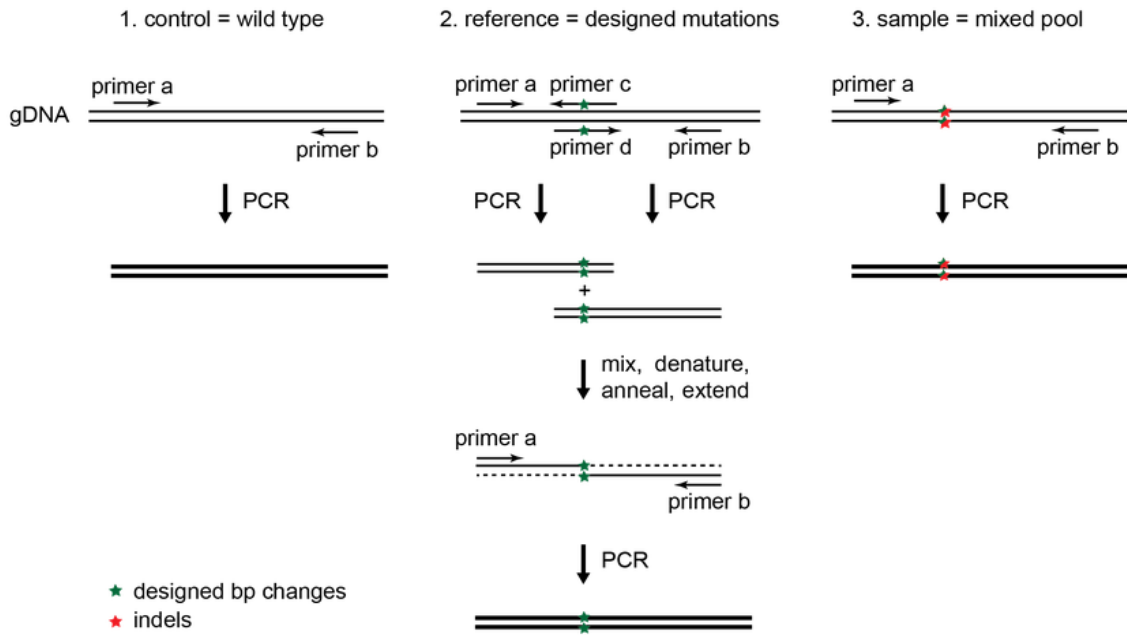
### 1.4.3. Cebadores para análisis TIDER

Para los análisis de TIDER se necesitan 3 amplicones de PCR (**Figura 42**), todos generados a partir del mismo conjunto de cebadores (**Tabla 4**):

- **Control** (DNA de células WT, no editadas): amplicón generado con los cebadores A + B.
- **Referencia** (DNA que contenga las mutaciones diseñadas como en la plantilla del oligo del donante): 2 amplicones generados (cebadores A + C y cebadores D + B) para introducir la variante de interés. Posteriormente se anillan los dos amplicones (**Figura 42**).
- **Muestra experimental** (DNA del grupo de células tratadas con Cas9, sgRNA y ssODN): amplicón generado con los cebadores A + B.

**Tabla 4. Cebadores diseñados para el análisis con TIDER.** En verde se marca la variante de interés.

Nombre	Cebadores (5' – 3')
Cebador A	CGAAGTGCTTGGCTCAATAGA
Cebador B	CTGGGACTCAGGCGGGTCACG
Cebador C	CTGCTGT <b>C</b> ACCTAGACAAAG
Cebador D	CTTTGTCTAGG <b>T</b> GACAGCAG



**Figura 42. Amplicones requeridos para los análisis TIDER.** (1) DNA control no editado. (2) DNA referencia con la variante de interés (introducida mediante PCR). (3) DNA del *pool* de células usado en CRISPR/Cas9. Figura extraída de la web de TIDER (<http://shinyapps.datacurators.nl/tider/>).

#### 1.4.4. Cebadores para *long-range* PCR

Para obtener fragmentos de 13 kb que contengan las 7 variantes del haplotipo se realizó una PCR de largo alcance (*long-range* PCR), seguido de una *97aemon* PCR para mejorar la pureza del fragmento amplificado. Los cebadores diseñados para estas reacciones contienen adaptadores universales que permiten añadir *barcodes* para bibliotecas de Oxford Nanopore Sequencing en posteriores reacciones de PCR (Tabla 5).

**Tabla 5. Cebadores diseñados para las reacciones de *long-range* PCR.** En verde los adaptadores universales para *barcodes* de Oxford Nanopore.

Reacción	5'-Cebador Fw-3'	5'-Cebador Rv-3'
<i>Long-range</i> PCR	<b>TTCTGTTGGTGC</b> TGATAT <b>TGCGCCATGACCATTGTTATTTGTCCAGA</b>	<b>ACTTGCCTGTCGCTCTATCTTCCCTGAAGAAATGTCACGGCTTGTAG</b>
<i>Nested Long-range</i> PCR	<b>TTCTGTTGGTGC</b> TGATAT <b>TGCAGTAACTGAAAATGCTTCTGAGTGGC</b>	<b>ACTTGCCTGTCGCTCTATCTTCTTGAAGAGAGGTGACAAAACAAAGGG</b>

### 1.4.5. Cebadores para validación por secuenciación de Sanger

Todos los cebadores utilizados para validaciones realizadas mediante secuenciación por Sanger se detallan en la **Tabla 6**:

**Tabla 6. Cebadores diseñados para validación por secuenciación de Sanger.**

Experimento	Secuencia (5' – 3')	Lugar de hibridación
Clonaje de gRNAs en el vector pgRNA-CKB	GAGATCCAGTTTGGTTAGTACCGGG	<i>Upstream</i> del lugar de restricción BsmBI en pgRNA-CKB
Mutagénesis dirigida en el vector pCGN	CTGTGCCAACTGCCAGACTA	<i>Upstream</i> de la posición K318A en el vector pCGN
Validación de las variantes del haplotipo en la línea BS hiPSC	rs6801957: CGAAGTGCTTGGCTCAATAGA rs6799257: GGAACACAAGTGACGACTGC rs9836859 } rs6790396 } GACGCTGAGCATTGAAACA rs9874633 } rs10428132: GTGCTAGGCCCTTGGGATAA rs10428168: CACCCCTAATCCAGGTACGC	<i>Upstream</i> de cada una de las variantes secuenciadas

### 1.5. Sondas *TaqMan*

Para evaluar los efectos de los ensayos CRISPRi sobre la expresión génica se tuvieron que realizar experimentos de qPCR. Para ello, se usaron sondas *TaqMan* (Thermo Fisher) (**Tabla 7**).

**Tabla 7. Sondas *TaqMan* para qPCR de los experimentos CRISPRi.**

Gen	Código de la sonda
<i>SCN5A</i>	Hs00165693_m1
<i>PP1A</i>	Hs04194521_s1
<i>GAPDH</i>	Hs02758991_g1
<i>CACNA1C</i>	Hs00167681_m1
<i>NKX2-5</i>	Hs00231763_m1
<i>SCNN1A</i>	Hs00168906_m1
<i>KCNQ1</i>	Hs00923522_m1

## 1.6. Anticuerpos

Los anticuerpos utilizados por los ensayos de inmunoprecipitación y *Western Blot* (WB) se detallan en la **Tabla 8**:

**Tabla 8.** Anticuerpos para los experimentos de inmunoprecipitación y Western blot.

Anticuerpo	Origen	Proveedor (número de catálogo)	Aplicación	Dilución
$\alpha$ -lisinas acetiladas (policlonal)	Conejo	Cell Signaling Technology (9441)	Inmunoprecipitación	1:200
$\alpha$ -GATA4 (monoclonal)	Ratón	Santa Cruz Biotechnology (sc-25310)	Western blot (Anticuerpo primario)	1:100
$\alpha$ -ratón (monoclonal) HRP	Cabra	Thermo Scientific (32430)	Western blot (Anticuerpo secundario)	1:10 000

## 1.7. Bases de datos públicas y recursos bioinformáticos

- **Bases de datos genómicas:** Para definir las regiones no codificantes sobre las que hicimos los experimentos de CRISPRi, se utilizó información de cardiomiocitos humanos (HCM) de varios conjuntos de datos disponibles públicamente. Mediante la plataforma *Genome Browser* de la Universidad de California Santa Cruz (UCSC) (<https://genome.ucsc.edu/>) se utilizaron diferentes recursos basados en datos ENCODE y ChIP-seq de FT cardiacos. La plataforma *WashU Epigenome Browser* de la Universidad de Washington se utilizó para obtener y visualizar datos de interacciones entre regiones genómicas (Hi-C, PCHi-C, etc). Las bases de datos de *Ensembl* también fueron utilizadas para consultar información relacionada con las variantes de estudio (<https://www.ensembl.org/index.html>).
- **Diseño de secuencias de DNA:** Para diseñar las gRNAs de los ensayos CRISPRi se utilizaron los programas CHOPCHOP y Benchling<sup>361</sup>. En el caso de la sgRNA y el ssODN para CRISPR/Cas9 se utilizó la herramienta de diseño web de *IDT Technologies* (<https://www.idtdna.com/pages/tools/alt-r-crispr-hdr-design-tool>).
- **Análisis CRISPR/Cas9:** Para realizar la cuantificación de la edición por CRISPR/Cas9 con el fin de comparar los diferentes métodos usados, se utilizó el software TIDER<sup>362</sup>. Para medir la intensidad de las bandas en los WB se pudo utilizar el software ImageJ.
- **Análisis RNA-seq:** El centro de Supercomputación y Bioinnovación de la Universidad de Málaga (UMA) nos proporcionó soporte y recursos científicos para llevar a cabo el proceso

de análisis de RNA-seq. El procedimiento bioinformático de análisis se llevó a cabo con MobaXterm, una herramienta de computación remota para poder acceder al superordenador de la UMA a distancia.

## 2. Métodos

### 2.1. Cultivo de las líneas celulares HEK y H9c2

#### 2.1.1. Mantenimiento

Las células HEK293T y H9c2 se cultivaron en medio Dulbecco's Modified Eagle Medium (DMEM, Sigma) suplementado con un 10% de suero fetal bovino (Sigma), un 1% de penicilina/estreptomicina (Sigma) y un 1% de Glutamax (Invitrogen), manteniéndose en un incubador a 37°C con un 5% de CO<sub>2</sub>. Ambos tipos celulares tienen un crecimiento en monocapa sembradas en frascos de cultivo de 75 cm<sup>2</sup> de superficie (Falcon Corning). Para conservar las células HEK293T y H9c2 a largo plazo se mantuvieron stocks congelados en nitrógeno líquido.

#### 2.1.2. Congelación y descongelación

Para la congelación, en primer lugar, se aspiró el medio celular de un frasco donde se mantenía el cultivo de células HEK293T o H9c2, y se hizo un lavado de las células con 10 mL de tampón salino fosfato o PBS (del inglés, *Phosphate buffer saline*) (Sigma-Aldrich). Seguidamente se tripsinizaron las células con 1 mL de Tripsina-EDTA 0.05% (Gibco) durante 3 minutos a 37°C. A continuación, se resuspendieron las células en 10 mL de medio DMEM suplementado para detener la tripsinización, y se determinó la concentración de células de la suspensión con un contador automático de células (*Countess II Cell Counter*, Invitrogen). Se sedimentaron las células centrifugándolas 5 minutos a 1500 rpm, y el sedimento obtenido se resuspendió en medio de congelación (medio DMEM suplementado con 10% de DMSO, Sigma) para conseguir una concentración de aproximadamente 1x10<sup>6</sup> células/mL. La suspensión de células se alicuotó en criotubos de 1 mL y se mantuvo 24 horas a -80°C en un recipiente de congelación que contenía isopropanol (*Nalgene® Mr. Frosty, Thermo Scientific*). Este método permite una congelación lenta y evita la formación de cristales intracelulares que podrían dañar las células. Transcurridas las 24 horas, los crioviales se introdujeron en el tanque de nitrógeno líquido.

La descongelación de los viales de células HEK293T o H9c2, conservados en nitrógeno líquido, se hizo rápidamente en un baño a 37°C para prevenir la aparición de cristales. Una vez descongeladas, las células se resuspendieron en un tubo de 15 mL con un volumen de 5 mL de medio DMEM suplementado. Seguidamente, se centrifugaron durante 5 minutos a 300 g y se descartó el sobrenadante. El sedimento celular se resuspendió en 5 mL de medio DMEM suplementado, y las células se sembraron en un frasco de 75 cm<sup>2</sup> con 15 mL de medio DMEM suplementado. Se dejaron crecer en incubación a 37°C.



### 2.1.3. Subcultivo

Las células HEK293T y H9c2 normalmente se cambiaron de frasco cada tres días, cuando la confluencia de los cultivos alcanzaba un 70-80%. En primer lugar, se aspiró el medio, se realizó un lavado con 10 mL de PBS, y a continuación se tripsinizaron las células como se ha descrito en el apartado anterior. Después de ese tiempo, se resuspendieron las células en 10 mL de medio DMEM suplementado. El subcultivo se hizo en un nuevo frasco de 75 cm<sup>2</sup>, previamente acondicionado con 15 mL de medio DMEM suplementado, a una proporción de 1:10 (suspensión celular: medio) en el caso de las células HEK293T y de 1:4 en el caso de las células H9c2. Esta diferencia entre ambos tipos celulares se debe a que las células HEK293T tienen una tasa de crecimiento más rápido.

## 2.2. Reprogramación celular de células somáticas a hiPSC

El Centro de Medicina Regenerativa de Barcelona (CMRB) se encargó de realizar el proceso de reprogramación a partir de las muestras que recibió de los 2 individuos descritos en el apartado 1.1.2. de materiales. Allí purificaron células mononucleares de sangre periférica (PBMCs) mediante Lympholyte-H kit (Cedarlane, Art.No.CL5016) y centrifugación.

El método de reprogramación se realizó por transducción de los factores de Yamanaka en las PBMCs mediante un virus de RNA no integrativo (virus Sendai). Las PBMCs se infectaron en solución con virus Sendai (Life Technologies, *CytoTune-iPS 2.0 Reprogramming Kit A16517*), siguiendo las instrucciones del fabricante y fueron sembradas sobre feeders. A los 15-20 días aparecieron las primeras colonias de hiPSC que se picaron, se expandieron y se caracterizaron.

## 2.3. Cultivo y diferenciación de hiPSC

### 2.3.1. Congelación y descongelación

Las células hiPSCs procedentes del laboratorio del Dr. Bruce Conklin (Gladstone Institutes, San Francisco, CA, EE. UU) llegaron en viales de 1 mL (1 millón de células/vial) en medio *Knock Out Serum Replacement* (KOSR) con DMSO, que se mantuvieron en nitrógeno líquido hasta su uso. Se descongeló 1 vial en el baño maría a 37°C y se inoculó en un tubo con 6 mL de medio mTeSR1 (STEMCELL Technologies), suplementado con 6 µL del inhibidor de las vías RHO/ROCK Y-27632 (Ri, STEMCELL Technologies) a una concentración de 10 µM. El inhibidor Ri, se utiliza para promover la supervivencia de las células cuando están disociadas, ya que previene la apoptosis inducida por la disociación celular. Los tubos inoculados se centrifugaron 5 minutos a 200 g y se eliminó el sobrenadante. El sedimento de células resultante se resuspendió en 3 mL de mTeSR1 con 3 µL de Ri y se sembró en placas de 6 pocillos con Matrigel (Corning), incubándose a 37°C, 5% CO<sub>2</sub>. El proceso de congelación es el mismo que en el

apartado 2.1.2. salvo por el uso de mTeSR1 + Ri como medio y *accutase* (STEMCELL Technologies) como agente de disociación celular.

### 2.3.2. Mantenimiento

Las hiPSC se mantuvieron de forma rutinaria en placas de 6 pocillos (Falcon Corning) recubiertos con matrigel y con medio mTeSR1 (STEMCELL Technologies). El medio de cultivo se cambió diariamente agregando 2 mL de mTeSR1.

### 2.3.3. Subcultivo

Para realizar los pases se retiró el medio de cultivo y se hizo un lavado con PBS. Posteriormente las hiPSCs se disociaron con 1 mL de *accutase*, incubando 3 minutos a 37°C. Se agregaron 3 mL de mTeSR1 para neutralizar la acción de la *accutase* y se transfirieron las células a un tubo de centrifuga de 50 mL. Se centrifugaron las células 3 minutos a 200 g, se descartó el sobrenadante, y se resuspendieron en 7 mL de mTeSR1 + 7 µL de Ri. Finalmente se contaron las células para calcular el volumen necesario para inocular entre 125 000 y 270 000 hiPSCs por pocillo en una placa nueva con matrigel y 2 mL de mTeSR1 + 2 µL de Ri.

### 2.3.4. Diferenciación de hiPSC a hiPSC-CMs

Para diferenciar las hiPSCs a cardiomiocitos, se siguió un protocolo de diferenciación optimizado en nuestro laboratorio, basado en los métodos propuestos por Lin *et al.*, Parikh *et al.* y Sharma *et al.*, con variaciones<sup>315,363,364</sup>. El protocolo de diferenciación incluye 4 etapas principales (**Figura 43**):

- (1) Etapas de crecimiento: en esta etapa, se dejan crecer las hiPSCs hasta que adquieren una confluencia del 90%.
- (2) Etapas de diferenciación: una vez las hiPSCs han crecido, se induce la diferenciación a cardiomiocitos mediante la adición de diferentes reactivos.
- (3) Etapas de selección por estrés metabólico: para determinar que hiPSCs se han diferenciado a cardiomiocitos y cuales no lo han hecho correctamente se añade un medio sin glucosa, por lo que solo aquellas células que se han diferenciado sobrevivirán.
- (4) Etapas de maduración hormonal: durante 15 días los cardiomiocitos maduran gracias a la adición de 2 hormonas.

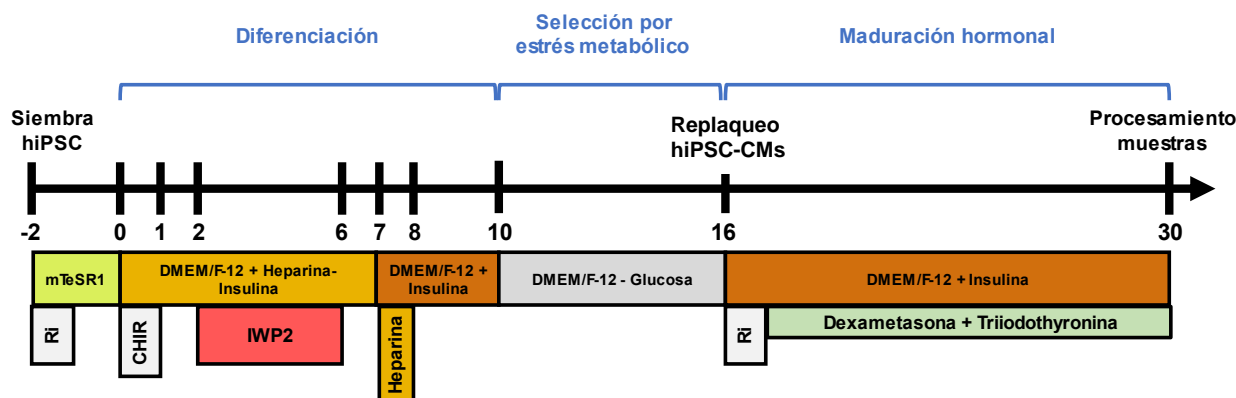


Figura 43. Esquema del protocolo de diferenciación de hiPSCs a hiPSC-CMs.

Al iniciarse una ronda de diferenciación a partir de un cultivo de hiPSC, las células se contaron y, según la línea celular, se sembraron entre 600 000 – 1 000 000 de células por pocillo en placas de 12 pocillos. El día de siembra se consideró el día -2 de diferenciación y se mantuvieron las células en matrigel con 2 mL de mTeSR1 + 2  $\mu$ L de Ri. El día -1 se cambió el medio de cultivo por mTesR1 para dejar crecer a las células.

Cuando se cambió el medio de cultivo a medio de diferenciación (**Tabla 9**) con CHIR99021 (Selleckchem), se consideró como el día 0. El CHIR99021 se añade en una concentración de entre 4-6  $\mu$ M, dependiendo de la línea celular. Este parámetro debe ajustarse de manera óptima para cada línea ya que influye de manera drástica al resto del proceso de diferenciación.

**Tabla 9. Composición del medio de diferenciación.** Se usa DMEM/F-12 y se suplementa con el resto de compuestos.

Compuesto	Concentración
DMEM/F-12 (Thermo Fisher)	Solvente
Ácido L-ascórbico (Sigma)	0.064 mg/mL
Selenito de sodio (Sigma)	0.01344 $\mu$ g/mL
Holo-transferrina (Sigma)	10 $\mu$ g/mL
Concentrado de lípidos definido químicamente (Gibco)	1%
GlutaMAX (Thermo Fisher)	1%
Penicilina/Estreptomina (Gibco)	1%

Desde el día 1 hasta el 7, el cambio de medio de las células se realizó cada 24 horas exactas. El día 1 se cambió por medio de diferenciación con heparina (Sigma; 3  $\mu$ g/mL). Los días 2, 3, 4 y 5, por medio de diferenciación con heparina (3  $\mu$ g/mL), inhibidor de las vías Wnt o IWP2 (Tocris; 3  $\mu$ M) y Ri (10  $\mu$ M). El día 6 se utilizó medio de diferenciación con heparina (3  $\mu$ g/ml). El día 7

se cambió por medio de diferenciación con heparina (3 µg/mL) e insulina (Sigma; 10 µg/mL). A partir del día 8 el medio de cultivo se cambió cada 2 días y sin restricción horaria. El día 8 se utilizó medio de diferenciación con insulina (10 µg/mL). Los días 10, 12 y 14 se cambió por medio de diferenciación sin glucosa (SILAC). El medio SILAC se suplementó con los compuestos de la **Tabla 10**.

El día 16, los cardiomiocitos se disgregaron y se transfirieron a una nueva placa en un proceso llamado “replaqueo”. Entre los 16 a 30, se cambió el medio diariamente, utilizando medio de diferenciación con insulina (10 µg/mL) + dexametasona (DEX; 0.001 mM) + Triyodo-L-Tironina (T3; 0.0001 mM).

**Tabla 10. Suplementos añadidos al medio sin glucosa SILAC.**

Compuesto	Concentración
SILAC (Thermo Fisher)	Solvente
Ácido L-ascórbico (Sigma)	0.0615 mg/mL
Selenito de sodio (Sigma)	0.00844 µg/mL
Holo-transferrina (Sigma)	2.5 µg/mL
Concentrado de lípidos definido químicamente (Thermo Fisher)	1%
GlutaMAX (Thermo Fisher)	1%
Penicilina/Estreptomicina (Thermo Fisher)	1%
L-Arginina (Thermo Fisher)	147.5 mg/L
L-Lisina (Thermo Fisher)	91.25 mg/L

### 2.3.4.1. Replaqueo

Los cardiomiocitos se trataron con 0.5 mL de colagenasa B (1.818 mg/mL) por pocillo, un máximo de 45 minutos para separarlos de la placa. La colagenasa B se preparó en solución con medio DMEM suplementado + Ri (10 µM) + insulina (10 µg/mL). Los cardiomiocitos se transfirieron a tubos de 15 mL una vez tratados con colagenasa B y se centrifugaron 5 minutos a 200g. Se descartó el sobrenadante y se resuspendieron las células con 4 mL de tripsina 0.25% (Corning). Se mantuvieron en reposo 4 minutos a 37°C, luego se agitaron vigorosamente y se dejaron reposar 7 minutos más. Se agregó 4 mL de medio de digestión (**Tabla 11**) y se removió la solución suavemente con una pipeta serológica de 10 mL. Se centrifugó 5 minutos a 200 g y se descartó el sobrenadante. El sedimento celular se resuspendió en 1-2 mL de medio de digestión y se pasó a un tubo de 50 mL pasando a través de un filtro celular de 70 µm (Falcon). Se lavó el filtro con 1-2 mL más de medio de digestión y, finalmente, se inocularon entre 500 000 y 2 000 000 de células por pocillo en una placa de 12 pocillos, según la cantidad obtenida.

Tabla 11. Composición del medio de digestión usado en el proceso de replaqueo.

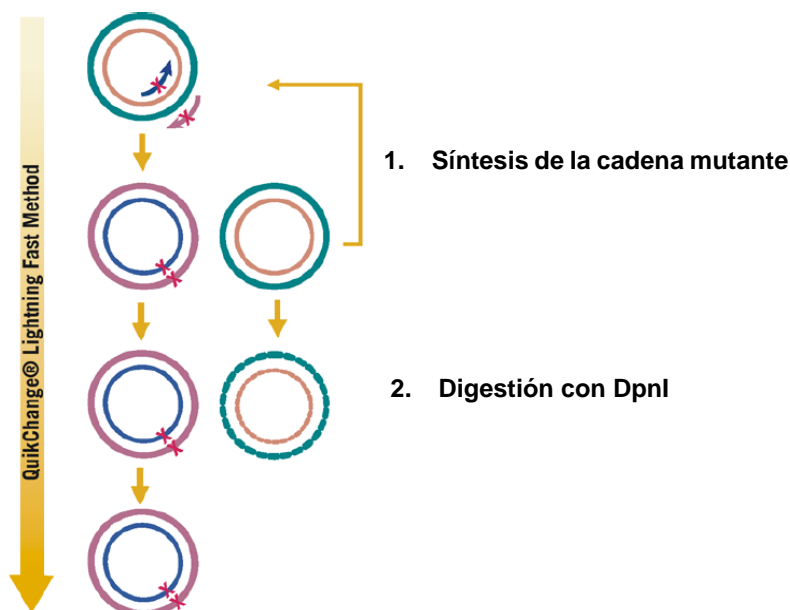
Compuesto	Concentración
Medio de diferenciación (Tabla 7)	80%
FBS (Gibco)	20%
Ri (STEMCELL Technologies)	10 $\mu$ M
Insulina (Sigma)	10 $\mu$ g/mL
T3 (Sigma)	0.0001 mM
DEX (Tocris)	0.01

## 2.4. Generación de mutantes de GATA4

### 2.4.1. Mutagénesis dirigida

La introducción de mutaciones puntuales en el motivo GATA4 del vector pCGN-GATA4/HA para generar los diferentes mutantes simples, dobles y triples, se realizó por mutagénesis dirigida a partir del kit *QuikChange Lightning Site-Directed Mutagenesis* (Agilent Technologies), siguiendo las instrucciones del fabricante. Este sistema permite introducir mutaciones puntuales en vectores de doble cadena.

Esta técnica consiste en que cada uno de los cebadores hibrida en una cadena del vector y elonga el fragmento incorporando la mutación de interés, hasta replicar el vector completamente y formando una molécula de DNA bicatenaria. La enzima DpnI digiere específicamente la cadena de DNA metilado del vector de origen bacteriano. De esta forma, se elimina el vector WT, quedando únicamente el vector con la mutación introducida (**Figura 44**). Con el fin de generar todos los vectores mutados necesarios para esta tesis se repitió el proceso de mutagénesis con los cebadores específicos por cada mutación deseada (**Tabla 1**).



**Figura 44.** Esquema de funcionamiento del método *QuikChange Lightning site-directed mutagenesis*. Figura adaptada de Agilent Technologies (<https://www.agilent.com>).

En primer lugar, se realizó una PCR que se preparó de la siguiente manera: 5  $\mu\text{L}$  de buffer de reacción 10X, 10-100 ng de dsDNA molde (vector pCGN), 125 ng de cada cebador (Fw y Rv), 1  $\mu\text{L}$  de mezcla de dNTPs, 1,5  $\mu\text{L}$  de reactivo *QuikSolution*, 1  $\mu\text{L}$  de enzima *QuikChange Lightning* y agua *nuclease-free* hasta 50  $\mu\text{L}$ . La reacción de PCR se realizó en un termociclador bajo las siguientes condiciones: (i) 2 minutos a 95°C; (ii) 18 ciclos de 20 segundos a 95°C, 10 segundos a 60°C y 30 segundos/kb de vector a 68°C; (iii) 5 minutos a 68°C; y (iv) mantener a 4°C. Una vez finalizada la PCR, se realizó un tratamiento con la endonucleasa DpnI durante 1 hora a 37°C.

#### 2.4.2. Transformación de bacterias ultracompetentes

Una vez finalizada la reacción con DpnI, los vectores mutados se transformaron en bacterias ultracompetentes *E. coli* XL10-Gold (Agilent Technologies), siguiendo las instrucciones del fabricante. Este proceso consiste en exponer las bacterias a un choque térmico para que incorporen los vectores de interés al interior de las células bacterianas. Para cada transformación, se mezclaron 45  $\mu\text{L}$  de bacterias con 2  $\mu\text{L}$  de  $\beta$ -mercaptoetanol, y se dejaron incubando 2 minutos en hielo. A continuación, se añadieron 2  $\mu\text{L}$  de vector mutado y se hizo una incubación de 30 minutos en hielo. Pasado ese tiempo se hizo el choque térmico de 30 segundos a 42°C en un baño, seguido de una incubación de 2 minutos en hielo. Posteriormente, se añadieron 0.5 mL de medio NZY+ (NZYTech) atemperado a 42°C, dejando un tiempo de recuperación de 1 hora a 37°C en agitación. Finalmente, se sembraron entre 20-100  $\mu\text{L}$  de las

bacterias en placas de agar LB (Luria-Bertani) autoclavado y con 100 µg/ml de ampicilina. Para crecer las colonias se incubaron las placas durante unas 16 horas a 37°C.

### 2.4.3. Mini-preparaciones de DNA plasmídico

Las mini-preparaciones de DNA plasmídico (mini-prep) permiten la obtención de pequeñas cantidades de DNA de una forma fácil y rápida. Las mini-prep correspondientes a los plásmidos mutados se realizaron después de la transformación bacteriana. Por cada uno de los vectores de interés se picaron 4 colonias (réplicas biológicas) en tubos de 15 mL con 4 mL de medio LB autoclavado, y 4 µL de ampicilina (100 µg/mL). Estos cultivos se dejaron crecer 16 horas a 37°C y 220 rpm de agitación. De estos cultivos, 3 mL fueron utilizados para realizar las mini-preps y el mL restante se conservó a 4°C. Para realizar la extracción de DNA plasmídico se siguió el protocolo *NucleoSpin Plasmid/Plasmid* (Macherey-Nagel) según las instrucciones del fabricante. En primer lugar, se consiguió el sedimento bacteriano centrifugando el cultivo durante 30 segundos a 11 000 g. Después de eliminar el sobrenadante, el sedimento se resuspendió en tampón de lisis, durante 5 minutos, y se detuvo la reacción con tampón de neutralización. A continuación, se lavó el lisado por centrifugación durante 5 minutos a 11 000 g, precipitando los restos celulares. El DNA plasmídico quedó en suspensión en el sobrenadante, el cual se trasladó a las columnas del kit comercial para que el DNA se uniera a la membrana de la columna. Finalmente se limpió la membrana con tampón de lavado, se secó y se eluyó el DNA en un tubo de 1.5 mL. Una vez purificadas, las preparaciones de DNA plasmídico se cuantificaron con un espectrofotómetro NanoDrop 1000 (Thermo Scientific).

### 2.4.4. Secuenciación de los plásmidos

Para secuenciar y validar que la mutagénesis funcionó correctamente, se utilizó el kit de secuenciación *BigDye™ Terminator v3.1* (Thermo Fisher). Las reacciones de *BigDye* se prepararon en una placa de 96 pocillos y cada reacción contenía: 1 µL de muestra de DNA plasmídico (100 ng/µL), 0.5 µL de cebador de secuenciación (**Tabla 6**), 0.5 µL de reactivo *BigDye*, 1 µL de tampón de secuenciación y 2 µL de agua *nuclease-free*. La placa se tapó con film metálico y se colocó en el termociclador en las siguientes condiciones: (i) 1 minuto a 96°C; (ii) 25 ciclos de 10 segundos a 96°C, 5 segundos a 50°C y 4 minutos a 60°C; (iii) indefinidamente a 4°C.

Una vez finalizada la reacción de *BigDye*, el DNA se precipitó para purificarlo y desnaturalizarlo en fragmentos de cadena sencilla que permitieran realizar la secuenciación. Brevemente, se agregaron 80 µL de etanol-acetato a cada reacción de *BigDye* y se incubó la placa durante 10 minutos en oscuridad. Se centrifugó 45 minutos a 2000 g, se eliminó el sobrenadante y se dejó secar la placa 10 minutos. Se agregaron 150 µL de etanol (70%), y se

centrifugó 10 minutos a 2 000 g, se eliminó el sobrenadante y se volvió a dejar secar la placa 10 minutos. Por último, se agregaron 10  $\mu$ L de formamida a cada reacción y se colocó la placa en el termociclador 7 minutos a 95°C. El producto de la precipitación se entregó al personal autorizado del laboratorio (Gencardio) que se encarga de la manipulación del secuenciador Genetic Analyzer 3130XL (Applied Biosystems). Finalmente, se analizaron los resultados de la secuenciación con el software *Sequencing Analysis* (Thermo Fisher).

### 2.4.5. Maxi-preparaciones de DNA plasmídico

Las maxi-preparaciones de DNA plasmídico (maxi-preps) permiten la extracción y purificación de DNA plasmídico y, a diferencia de las minis que solo extraen cantidades suficientes para secuenciar, se obtienen grandes cantidades de DNA. Este procedimiento tiene el objetivo de obtener una cantidad suficiente de plásmido para realizar todos los procedimientos subsiguientes necesarios. Experimentalmente se partió del cultivo que se reservó en el apartado 2.4.3., que se hizo crecer de nuevo en un volumen de 100 mL de medio LB. La extracción se realizó con el *QIAGEN® Plasmid Maxi Kit*, siguiendo las instrucciones del fabricante. La filosofía y los pasos del protocolo son los mismos que los descritos en el apartado 2.4.3. pero con mayores volúmenes.

### 2.4.6. Stocks de glicerol

Para conservar una reserva de las bacterias transformadas se mezcló 750  $\mu$ L del cultivo bacteriano transformado con 250  $\mu$ L de glicerol autoclavado al 60%. Estos stocks se guardaron en criotubos y se conservaron a -80°C.

## 2.5. Determinación de los niveles de acetilación de GATA4

### 2.5.1. Transfección de células HEK293T

Para estudiar los niveles de acetilación de las lisinas de GATA4 en experimentos de inmunoprecipitación, las células HEK293T se sembraron en placas de 10 cm<sup>2</sup> a una concentración de 10<sup>6</sup> células/mL. Al cabo de 24 horas, se realizó la transfección de los plásmidos de interés con Lipofectamina® 2000. La lipofectamina, al mezclarla con el DNA, forma complejos de liposomas que son endocitados dentro de la célula y, por tanto, permiten la incorporación del material genético de interés. Para formar los liposomas se utilizó una proporción de 1:2 ( $\mu$ g DNA:  $\mu$ L de lipofectamina) que se incubó en medio Opti-MEM (Sigma-Aldrich) durante 20 minutos. Tras la incubación, se distribuyó la mezcla gota a gota sobre el cultivo celular, mantenido en medio DMEM sin antibiótico. El medio sin antibiótico evita la intoxicación de las células más



permeables debido a la transfección. Al cabo de 6 horas se cambió el medio de transfección por medio DMEM suplementado con antibiótico.

Por cada placa, se transfectaron 10 µg del vector de expresión WT (GATA4/HA) o de los vectores de GATA4 mutados (apartado 1.3.2.). Al cabo de 24 horas de la transfección, las células se trataron con inhibidores de desacetilasas: Trichostatina A (TSA; Sigma) a 400 nM y Nicotinamida (NA; Sigma) a 5 mM, y se dejaron incubando 24 horas más a 37°C.

### 2.5.2. Preparación y cuantificación de los lisados celulares

Al cabo de 48 horas de la transfección, se lisaron las células. De forma previa a la lisis, se realizaron dos lavados de las placas con PBS frío. Las células se lisaron incubándolas en 800 µL de solución de lisis (**Tabla 12**) suplementado con inhibidor de proteasas (*Protease Inhibitor Cocktail*, Sigma-Aldrich) durante 30 minutos a 4°C en agitación, para evitar la degradación de las proteínas. Seguidamente se utilizó un raspador o *scraper* para recuperar el lisado celular, que se transfirió a tubos de 1.5 mL. Tras centrifugar los lisados durante 15 minutos a 4°C y 15 000 rpm, el sobrenadante, que contenía las proteínas de interés, se transfirió a otro tubo nuevo. La concentración de proteína total de cada lisado se cuantificó con el método BCA (*Pierce BCA Protein Assay Kit*, Thermo Fisher). Esta cuantificación permitió calcular el volumen de muestra necesario para inmunoprecipitar 1.5 mg de proteína para cada condición. De ese volumen calculado, un 1% fue transferido a un nuevo tubo para utilizarlo como *input* (muestra sin inmunoprecipitar). Se añadieron 10 µL de tampón de carga 2X (**Tabla 12**) al *input*, y se congeló hasta ser requerido como control en el WB.

### 2.5.3. Inmunoprecipitación

La inmunoprecipitación (IP) se realizó con la proteína A-sepharose (GE Healthcare). Primeramente, se hidrató la proteína A incubándola con PBS durante 15 minutos a temperatura ambiente; seguidamente se centrifugó durante 1 minuto a 1 300 rpm y se descartó el sobrenadante. A continuación, se resuspendió el sedimento de proteína A con 3 volúmenes de PBS, y de esta suspensión se utilizaron 80 µL por cada condición de IP. Antes de la inmunoprecipitación, las bolitas de proteína A correspondientes a cada condición se lavaron 5 veces en 500 µL de tampón de lisis (**Tabla 12**), centrifugando durante 5 minutos a 300 g y 4°C, eliminando el sobrenadante después de cada lavado. El último lavado se realizó con tampón de lisis suplementado con inhibidores de proteasas. Finalizados los lavados de la proteína A, se le añadió el lisado celular preparado en el apartado anterior (equivalente a 1.5 mg de proteína) y se ajustaron todas las condiciones con el mismo volumen añadiendo el tampón de lisis correspondiente. Seguidamente, se añadió la cantidad de anticuerpo anti lisina-acetilada correspondiente según las indicaciones del fabricante para inmunoprecipitación (**Tabla 12**), y

dejando incubar 16 horas a 4°C en agitación. Pasado este tiempo, se volvieron a realizar cinco lavados por condición en 500 µL de tampón de lisis. Los dos últimos lavados se realizaron sin detergente (Nonidet P-40). Finalizados los lavados, se añadieron 40 µL de tampón de carga 6X a cada muestra, congelándose a -20°C, hasta ser utilizados en el WB.

#### 2.5.4. Electroforesis y Western blot

La electroforesis en geles de poliacrilamida-SDS permite separar las proteínas cargadas en función de su peso molecular bajo la acción de un campo eléctrico. El gel de poliacrilamida-SDS consta de dos fases (**Tabla 13**): el gel separador (fase inferior), permite la migración proteica en función del peso molecular; y el gel concentrador (fase superior), facilita que las proteínas cargadas se concentren en una banda estrecha y entren todas al mismo tiempo en el gel separador. Para los experimentos con GATA4, debido a que tiene un peso molecular aproximado de 55 kDa, el gel separador se hizo a un 10% de acrilamida y el gel concentrador al 4% acrilamida. Antes de cargar los geles polimerizados, se descongelaron las muestras (IP e inputs, con sus respectivos volúmenes de tampón de carga), y se incubaron durante 5 minutos a 95°C en un termobloque con el fin de desnaturalizar las proteínas. Una vez cargadas las muestras en el gel, se activó el campo eléctrico: primero se dejó 10 minutos a 100 V para concentrar todas las muestras en una banda estrecha sobre el hielo separador, y entonces se aumentó a 150V hasta que las muestras llegaron al final del gel.

Una vez terminado, se realizó la transferencia de las proteínas del gel a una membrana de fluoruro de polivinilideno (PVDF) con tampón de transferencia (**Tabla 13**) durante 2 horas a 90V y 4°C. Una vez hecha la transferencia, la membrana se bloqueó con solución de bloqueo (**Tabla 12**) durante una hora a temperatura ambiente y en agitación. La membrana se lavó con PBS, y entonces se incubó con el anticuerpo primario (**Tabla 8**) durante 16 horas. El día siguiente, la membrana se lavó dos veces en solución PBS-Tween (**Tabla 12**) y otras dos con solución PBS, y entonces se incubó con el anticuerpo secundario (**Tabla 8**) conjugado con peroxidasa, diluido en solución de bloqueo al 3% de leche en polvo y sin Tween, durante 1 hora a temperatura ambiente y en agitación. El exceso de anticuerpo secundario se eliminó haciendo dos lavados con solución PBS-Tween y otros dos con solución PBS. Para detectar la peroxidasa conjugada en el anticuerpo secundario se utilizó el sustrato quimioluminiscente *Pierce® ECL Western Blotting Substrate* (Thermo Scientific). Por último, se desveló las bandas mediante la exposición de la membrana en una película de autoradiografía.

Tabla 12. Soluciones usadas durante los procesos de IP, electroforesis y Western blot.

Solución	Composición
Tampón de lisis	50 mM Tris/HCl, pH 7.4; 150 mM NaCl; 1mM EDTA, 1% Nonidet P-40
Tampón de carga 6x	0.5 M tris/HCl, pH 6.8; 30% glicerol; 10% SDS; 0.6 M DTT; 0.012% azul de bromofenol
Tampón de electroforesis	1 g/L SDS; 28.8 g/L glicina; 6 g/L Tris/HCl pH 8.8
Tampón de transferencia	1 g/L SDS; 11.13 g/L glicina; 2.42 g/L Tris/HCl pH 8.8; metanol 20%
Solución de bloqueo	5% leche desnatada en polvo en una solución de PBS con 0.1% de Tween-20
Solución de PBS	8 g/L NaCl; 0.2 g/L KCl; 1.44 g/L Na <sub>2</sub> HPO <sub>4</sub> ; 0.34 g/L KH <sub>2</sub> PO <sub>4</sub>
Solución de PBS-Tween	Solución de PSB y 0.1% de Tween

Tabla 13. Composición de los geles de acrilamida.

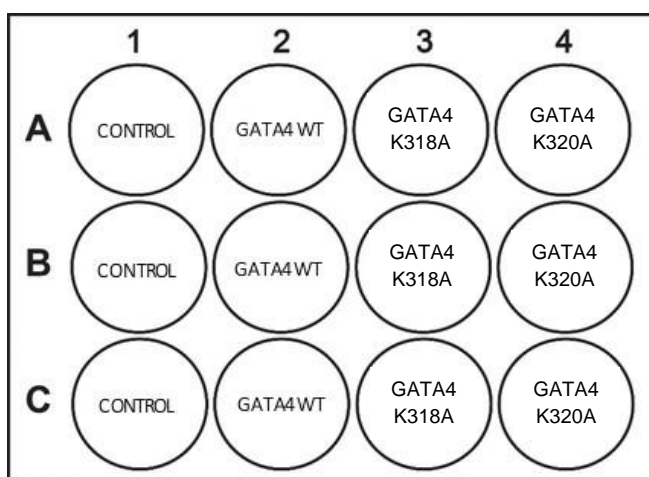
Gel separador	Gel concentrador
10% Acrilamida:Bisacrilamida 37, 5:1	4% Acrilamida:Bisacrilamida 37, 5:1
1.5 M Tris/HCl pH 8,8	0.5 M Tris/HCl pH 8.8
50 µL SDS 10%	25 µL SDS 10%
23 µL PSA solución al 10%	15 µL PSA solución al 10%
7.7 µL TEMED solución (Sigma)	5 µL TEMED solución (Sigma)
Agua milli-Q (MQ) 5 mL	Agua MQ 2.5 mL

## 2.6. Determinación de la actividad transcripcional de GATA4 acetilada en experimentos luciferasa

### 2.6.1. Transfección de H9c2

Con el fin de evaluar cómo la acetilación de GATA4 en diferentes lisinas modula su actividad transcripcional, se realizaron experimentos luciferasa. Para ello, se co-transfectaron los vectores reporteros duales (*SCN5A-firefly* o *PDX1-firefly*) y pEF-1 $\alpha$ -*renilla* (actúa como normalizador) junto con los vectores de GATA4 mutados en distintas lisinas de la región C-Zn (apartado 1.3.2.). El día antes de la transfección, se realizó la siembra de 50 000 células por pocillo utilizando placas de 12 pocillos. Al cabo de 24 horas, se realizó la transfección con Lipofectamina-2000®, siguiendo el protocolo previamente descrito en el apartado 2.5.1. Se transfectaron 100 ng de vector reportero, 10 ng del vector pEF-1 $\alpha$ -*renilla* y 900 ng de los vectores de expresión GATA4

WT y mutantes. Para llegar a esta cantidad de vector de expresión de GATA4 óptima, primero se realizó un ensayo transfectando cantidades crecientes de DNA (300, 600 o 900 ng). En cada experimento, la cantidad de DNA total se ajustó con pcDNA3.1 (plásmido control). En la **Figura 45** se muestra un ejemplo de una transfección en placas de 12 pocillos. Se realizaron triplicados técnicos para cada una de las condiciones experimentales. En la **Tabla 14** se muestra un ejemplo de las cantidades de DNA y lipofectamina preparadas para un experimento.



**Figura 45. Ejemplo de co-transfección de H9c2 en placa de 12 pocillos.** El control corresponde a la condición transfectada con el pcDNA3.1. Para cada condición se hicieron triplicados técnicos.

**Tabla 14. Ejemplo de volúmenes totales de DNA para el caso de la Figura 9.** X3.5 volúmenes para cada triplicado.

Condición	Vector reportero 100 ng/μL	Ef1α – Ren 25 ng/μL	GATA4 200 ng/μL	pcDNA3.1 200 ng/μL	DNA total
CONTROL	3.5 μL	0.7 μL	0	15.75 μL	3.52 μg
GATA4 WT	3.5 μL	0.7 μL	15.75 μL	0	3.52 μg
GATA4 K318A	3.5 μL	0.7 μL	15.75 μL	0	3.52 μg
GATA4 K320A	3.5 μL	0.7 μL	15.75 μL	0	3.52 μg

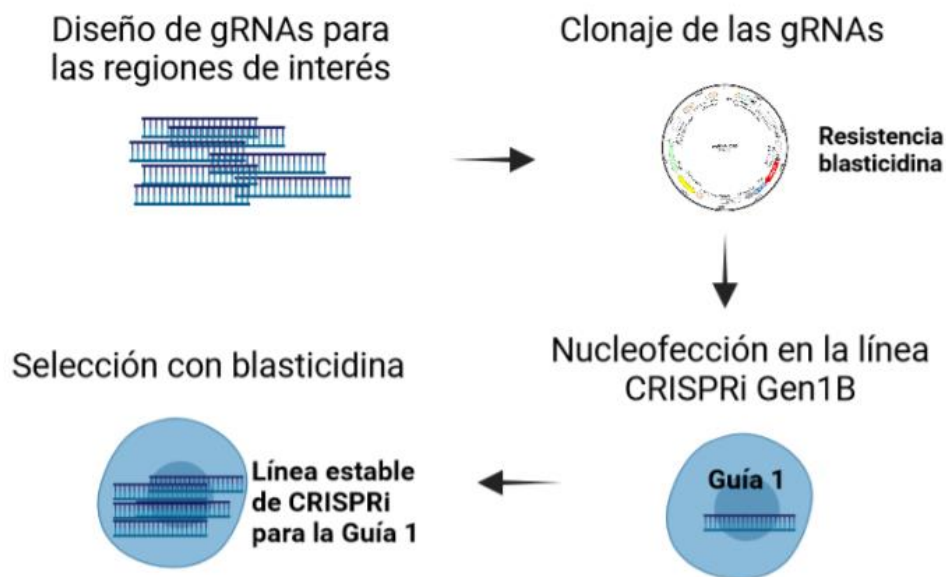
### 2.6.2. Lectura de la actividad luciferasa y renilla

Transcurridas 24 horas de la transfección, las células se lavaron dos veces en PBS frío. A continuación, se lisaron las células añadiendo 100 μL de tampón de lisis 1X (*Passive Lysis Buffer*, Promega) por pocillo durante 15 minutos a temperatura ambiente y en agitación suave. Se rasparon los pocillos para recoger el lisado celular, que se transfirió a tubos de 1.5 mL. El

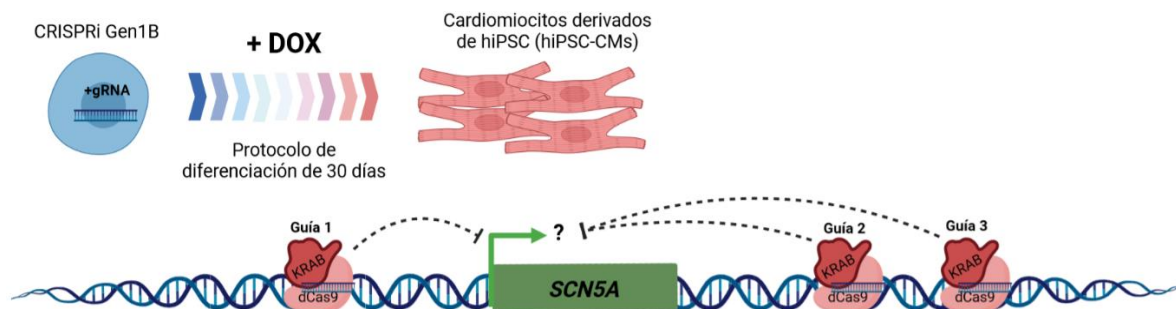
lisado celular se centrifugó durante 2 minutos a 16 000 rpm y 4°C, y el sobrenadante se transfirió a nuevos tubos. Las muestras se mantuvieron en hielo en todo momento. La actividad luciferasa *firefly* y *renilla* se determinó usando el kit *Dual-Luciferase® Reporter Assay System* (Promega) y el luminómetro GloMax-96 (Promega) con dos inyectores. La lectura se realizó en placas de 96 pocillos donde previamente se habían añadido 10 µL de la muestra en cada pocillo. El luminómetro permitió inyectar automáticamente, dentro de cada pocillo, 50 µL del reactivo *Luciferase Assay Reagent* (Promega), para detectar la actividad luciferasa *firefly*, y una vez tomada la medida, inyectó 50 µL del reactivo *Stop&Glo Reagent* (Promega), para detectar la actividad *renilla*, que permitió normalizar los valores de la luciferasa *firefly*. La normalización consiste en dividir los valores de luciferasa *firefly* por los de *renilla*, eliminando la variabilidad debido a la densidad celular.

## 2.7. Cribaje de regiones reguladoras de *SCN5A* utilizando un sistema CRISPRi

En los siguientes apartados se describen detalladamente los pasos seguidos para realizar los ensayos CRISPRi, resumidos y esquematizados en las **Figuras 46 y 47**.



**Figura 46. Esquema global del proceso para generar líneas CRISPRi con expresión de una gRNA.** (1) Diseño de gRNAs para las regiones de interés. (2) Clonaje de las gRNAs en el vector pgRNA-CKB con resistencia a blasticidina. (3) Nucleofección de los vectores pgRNA-CKB clonados en la línea CRISPRi Gen 1B. (4) Selección de células que expresan la gRNA mediante blasticidina para obtener una línea estable para esa determinada gRNA. Figura creada con BioRender (<https://app.biorender.com/>).



**Figura 47. Experimento de *knockdown* sobre *SCN5A*.** Las líneas estables obtenidas en la **Figura 44** se diferencian a hiPSC-CMs. Durante el proceso de diferenciación, cada gRNA unirá dCas9-KRAB a su región potencialmente reguladora, en consecuencia, pudiendo alterar la expresión de *SCN5A*. Figura creada con

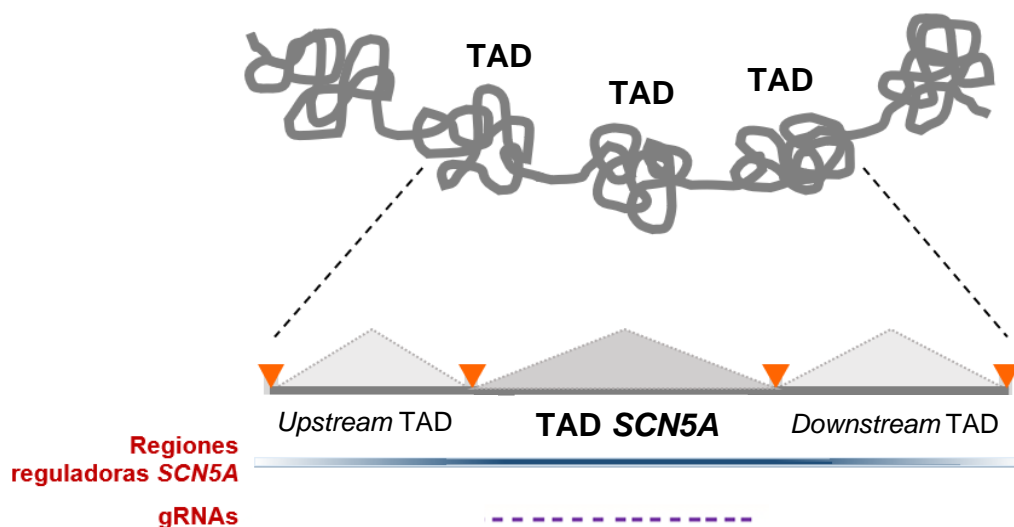
### 2.7.1. Selección de las regiones de estudio y diseño de guías

Con el fin de realizar un cribaje de regiones reguladoras de *SCN5A*, como primer paso se hizo una preselección de elementos potencialmente reguladores de la transcripción de *SCN5A* (**Figura 48**). Para ello, definimos como región de estudio el dominio topológico de *SCN5A*. Para la selección de las regiones de interés se recopilaron datos del proyecto ENCODE: DHS-seq, ChIP-seq para FT cardíacos y marcas de histonas como H3K4me3 y H3K27ac.

Una vez seleccionadas las regiones potencialmente reguladoras de *SCN5A*, se diseñaron las gRNAs dirigidas a estas regiones. Para cada una de ellas se extrajo la secuencia consenso de la versión del genoma humano GRCh37.p13 que se obtuvo en formato fasta del UCSC Genome browser. Cada secuencia se introdujo en la herramienta de diseño de guías CHOPCHOP de la página web de Zhang Lab<sup>361</sup> para analizar las mejores opciones de gRNAs para esta región. En la herramienta CHOPCHOP, se indicó que las guías se utilizarían para experimentos de represión con un sistema basado en CRISPRi. En CHOPCHOP, se emplean algoritmos rigurosos que buscan secuencias a las que la gRNA diseñada pueda unirse de manera no deseada (off-targets), para predecir la especificidad que cada gRNA que se diseña tendrá con su diana en el genoma<sup>361,365</sup>.

De entre las gRNAs diseñados por CHOPCHOP, se seleccionaron aquellas que presentaron el menor número de off-targets. Además de contar con una secuencia PAM “NGG”, en el diseño se agregó una secuencia extra de 4 nucleótidos al extremo 5' de la gRNA Fw y otra al extremo 5' de la gRNA Rv. Estos extremos eran compatibles con el plásmido pgRNA-CKB digerido, haciendo que el proceso de ligación fuera factible. Un aspecto que también se tuvo en cuenta es que las gRNAs no tuviesen similitud con las secuencias de reconocimiento de la enzima que

se utiliza para digerir pgRNA-CKB (BsmBI en este caso), para evitar digestión del plásmido después de la ligación y la formación de productos no deseados.



**Figura 48. Ejemplo esquemático de selección de gRNAs en la región reguladora de SCN5A.** Figura adaptada y cedida por S. Pagans.

### 2.7.2. Annealing

Las gRNAs se compraron en forma de oligonucleótidos que llegaron al laboratorio liofilizados. Las gRNAs se hidrataron con tampón TE (10 mM Tris-HCl, 1 mM EDTA) diluido al 50% en agua *nuclease-free* a una concentración final de 100  $\mu$ M.

Para cada una de las regiones a estudiar, se realizó una reacción de *annealing* que contenía: 6.5  $\mu$ L de agua *nuclease-free*, 1  $\mu$ L del cebador Fw (20 – 26.5  $\mu$ M), 1  $\mu$ L del cebador Rv (15.7 – 22.5  $\mu$ M), 1  $\mu$ L de tampón T4 PNK con ATP (BioLabs) y 0.5  $\mu$ L de la enzima T4 *Polynucleotide Kinase* (PNK, BioLabs). Cada reacción de *annealing* se llevó a cabo en un pocillo de una tira de PCR en un termociclador con las siguientes condiciones: (i) 30 minutos a 37°C, (ii) 5 minutos a 95°C, y (iii) incubaciones de 12 segundos comenzando a 95°C y disminuyendo 1°C hasta los 25°C. Al finalizar la reacción, cada producto del *annealing* se diluyó en proporción 1/100 en agua *nuclease-free* y se conservó a -20°C.

### 2.7.3. Clonaje de las guías en el vector pgRNA-CKB

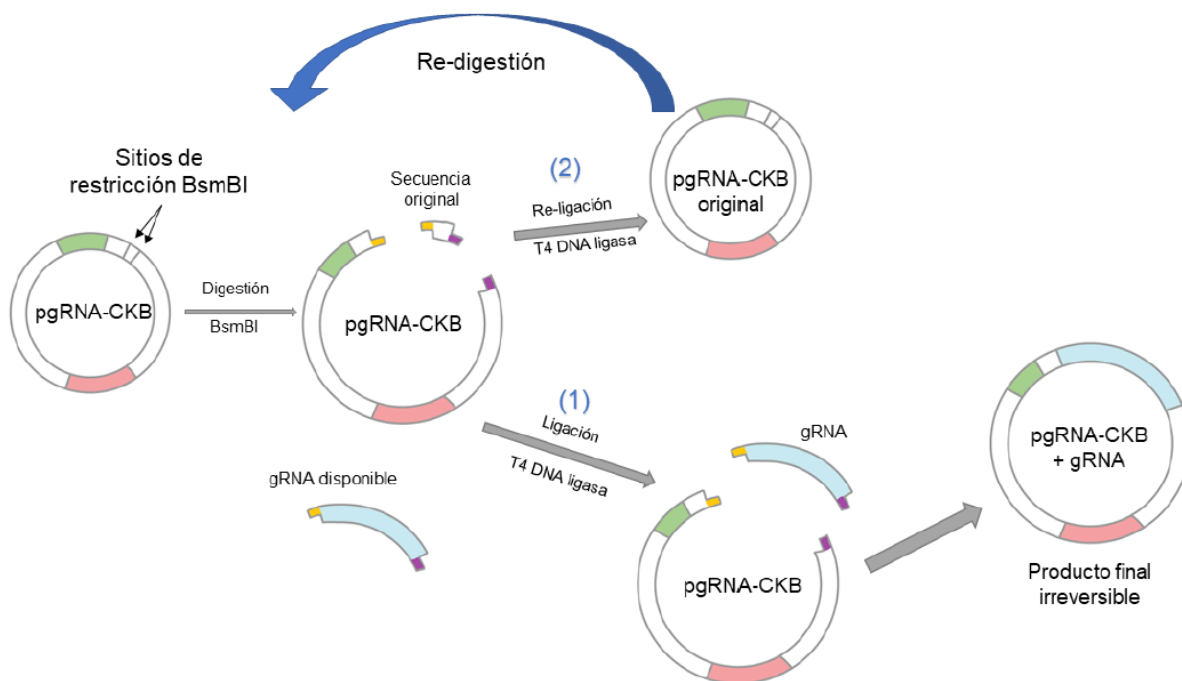
#### 2.7.3.1. Método Golden Gate

Cada gRNA se clonó en el plásmido pgRNA-CKB mediante la técnica de clonación “*Golden Gate*”, que permite integrar el gRNA en el plásmido en una sola reacción. Esta tecnología se basa en el uso de enzimas de restricción Tipo IIS, las cuales se caracterizan por reconocer

secuencias de DNA asimétricas y por “cortar” fuera de sus secuencias de reconocimiento. En este caso, se usó la enzima BsmBI (BioLabs) que forma extremos cohesivos de cuatro nucleótidos.

Mediante el *Golden Gate*, se consiguió que después de la digestión del pgRNA-CKB con la enzima BsmBI, los extremos cohesivos quedaran disponibles. Según los fundamentos de la técnica, estos extremos pueden: (1) unirse con la gRNA introducido en la reacción, de manera irreversible, ya que los sitios de restricción y reconocimiento de BsmBI se pierden una vez ligado el gRNA o, (2) unirse nuevamente con la secuencia original del plásmido, en cuyo caso, BsmBI volverá a digerir la secuencia en ciclos posteriores, promoviendo la ligación con los gRNA para obtener el producto deseado (**Figura 49**).

Para cada uno de los cuatro gRNAs, se realizó una reacción de clonaje *Golden Gate* que contenía: 13  $\mu\text{L}$  de agua *nuclease-free*, 2  $\mu\text{L}$  de tampón T4 DNA ligasa 10X, 2  $\mu\text{L}$  de tampón 3.1 10X (BioLabs), 100 ng (1  $\mu\text{L}$ ) de plásmido pgRNA-CKB, 1  $\mu\text{L}$  del gRNA correspondiente (producto del *annealing*), y 1  $\mu\text{L}$  de enzima BsmBI (10 000 U/ $\mu\text{L}$ ). Cada reacción de *Golden Gate* se llevó a cabo en un pocillo de una tira de PCR en un termociclador con las siguientes condiciones: (i) 30 minutos a 42°C; (ii) 30 minutos a 16°C, después de agregar 1  $\mu\text{L}$  (400 000 U/mL) de la enzima T4 DNA ligasa a cada reacción; (iii) 25 ciclos de 5 minutos a 42°C, 5 minutos a 16°C, 20 minutos a 55°C y 10 min a 80°C; (iv) indefinidamente a 4°C. Cada plásmido generado se nombró igual que la guía que le fue insertada.



**Figura 49. Esquema del sistema *Golden Gate*.** Los extremos cohesivos (amarillo y púrpura) formados por BsmBI, permiten que el plásmido se una al gRNA añadido en la reacción (1). Si el plásmido se liga con la secuencia que contenía originalmente (2), será digerido en ciclos posteriores de la reacción, hasta alcanzar una eficiencia cercana al 100% de ligación con el gRNA. Figura cedida por H. Gálvez.



### 2.7.3.2. Transformación de bacterias DH5 $\alpha$

Después de la reacción con el sistema *Golden Gate*, las bacterias *One Shot<sup>TM</sup> Stbl3<sup>TM</sup>* (Thermo Fisher), una cepa de *E. coli* químicamente competente, fueron transformadas con los productos de ligación. Estas bacterias están especialmente diseñadas para utilizarse con plásmidos integrativos o inestables (de gran tamaño), como los lentivirales o el usado en este caso (pgRNA-CKB), ya que evitan recombinaciones típicas al utilizar este tipo de plásmidos (como las LTR).

Para realizar el protocolo de transformación, se utilizaron 16  $\mu$ L de bacterias que fueron descongeladas en hielo. A continuación, se agregaron 2  $\mu$ L del producto del *Golden Gate* y se dejaron incubar durante 30 minutos en hielo. Posteriormente, se aplicó un choque térmico de 45 segundos a 42°C y se incubaron las bacterias 2 minutos más en hielo. Las bacterias transformadas fueron recuperadas añadiendo 250  $\mu$ L de medio SOC (Thermo Fisher) e incubándolas durante 1 hora a 37°C y 225 rpm. El volumen total de la reacción de transformación se sembró en placas LB agar con ampicilina (100  $\mu$ g/mL) y se incubaron 16 horas a 30°C (los eventos de recombinación disminuyen a 30°C en comparación con hacerlo a 37°C).

### 2.7.3.3. Secuenciación y validación del clonaje

Para validar que el plásmido pgRNA-CKB introdujo correctamente las gRNAs, se secuenciaron todos los clonajes realizados utilizando cebadores específicos para la región del plásmido donde se inserta la gRNA (apartado 1.3.3.). Para hacerlo se siguieron los pasos de secuenciación descritos en el apartado 2.4.4.

### 2.7.4. Nucleofección de la línea CRISPRi

La nucleofección, es un método de transfección basado en la electroporación, que permite insertar DNA directamente al núcleo de las células a través de la aplicación externa de corrientes eléctricas y el uso de reactivos específicos<sup>366</sup>. Para llevar a cabo este procedimiento con la línea celular WTB CRISPRi Gen1B se utilizó el kit *Human Stem Cell Nucleofector<sup>TM</sup> Kit 1* (Lonza), diseñado para el dispositivo Nucleofector<sup>TM</sup> 2b (Lonza), utilizando el programa A-23 específico para hiPSCs.

Para la reacción, se prepararon 100  $\mu$ L de solución de nucleofección con el suplemento del kit a una proporción 4.5:1, respectivamente. Se retiró el medio de la placa con las hiPSCs y se lavó cada pocillo con DPBS. Las células se disociaron con *accutase* y se mantuvieron en mTeSR1 + Ri para contarlas. Se calculó el volumen necesario para obtener 2 millones de células, las cuales se traspasaron a tubos de microcentrífuga de 1.5 mL que se centrifugaron 3 minutos a 115 g. Se descartó el sobrenadante y se resuspendieron las células con la solución de nucleofección suplementada. A cada condición se le agregaron 5  $\mu$ g de plásmido pgRNA-CKB

clonado con la gRNA correspondiente. La suspensión celular se transfirió a las cubetas de nucleofección y se colocaron en el nucleofector.

Una vez realizada la nucleofección (con el programa A-23), se agregaron 500  $\mu\text{L}$  de mTeSR1 + 0.5  $\mu\text{L}$  de Ri a cada cubeta y el contenido de cada una se sembró en un pocillo de una placa de 6 pocillos, los cuales se habían preparado previamente con matrigel y 2 mL de mTeSR1 + 2  $\mu\text{L}$  de Ri.

#### 2.7.4.1. Selección de líneas estables

Las células nucleofectadas se mantuvieron en la misma placa hasta que alcanzaron una confluencia de aproximadamente el 80%, cambiando el medio de cultivo diariamente con 2 mL de mTeSR1 con blasticidina (10  $\mu\text{g}/\text{mL}$ ).

El plásmido pgRNA-CKB incorpora el gen de resistencia a la blasticidina. De esta manera, cada vez que se añadía blasticidina en el medio, las células que no habían integrado correctamente el plásmido pgRNA-CKB o lo hicieron en un lugar transcripcionalmente inactivo, fueron eliminadas, promoviendo el crecimiento de las células que expresaban la gRNA en mayor medida. Cuando las células alcanzaron la confluencia, se hizo el subcultivo a una nueva placa de 6 pocillos y se mantuvieron en las mismas condiciones de cultivo con blasticidina hasta realizarse 2 subcultivos más y obtener colonias con expresión estable de la gRNA (**Figura 46**).

El plásmido pgRNA-CKB también contiene un transgén que codifica por la proteína fluorescente mKate2 (**Figura 41**), por lo que la expresión de la gRNA es visible mediante el uso de microscopía de fluorescencia.

#### 2.7.5. Rondas de diferenciación

Para diferenciar las hiPSCs a cardiomiocitos, se siguió el protocolo de diferenciación descrito en el apartado 2.3.4. Para realizar el experimento de *knockdown* (**Figura 47**) se añadió DOX al medio (2  $\mu\text{M}$ ) a partir del día 16 del proceso de diferenciación (justo después del replaqueo) y se mantuvo hasta el día 30. La DOX se añadió en la mitad de los pocillos obtenidos después del replaqueo, mientras que la otra mitad sin DOX sirvió como control del experimento. Se eligió añadir el antibiótico después del replaqueo debido a que en ese paso se obtienen pocillos idénticos entre sí en pureza y confluencia. Hacerlo antes del replaqueo supondría aumentar la variabilidad biológica entre la condición con y sin DOX.

Una vez finalizado el proceso de diferenciación (día 30 de diferenciación) se procedió a realizar la extracción de RNA de cada muestra.

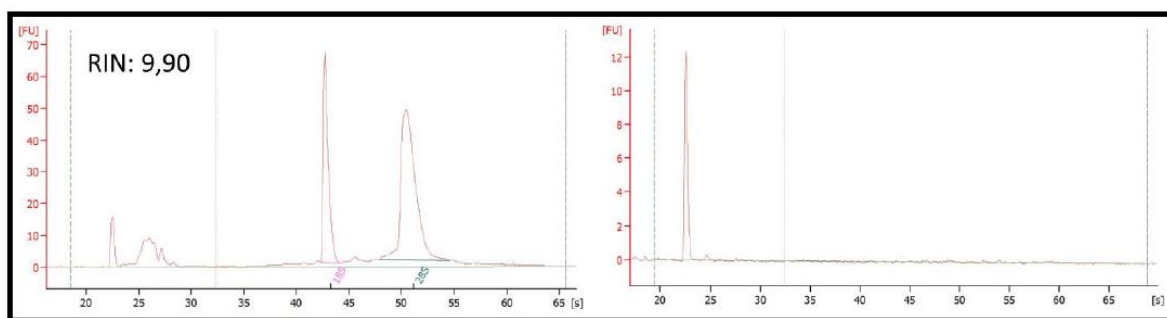
### 2.7.6. Extracción de RNA

Para realizar la extracción de RNA se utilizó el kit *Direct-zol RNA Miniprep Kit* (Zymo Research) siguiendo las instrucciones del fabricante, con pequeñas variaciones. En primer lugar, se aspiró el medio de cultivo de las placas en que se mantenían las células y se realizaron 2 lavados con PBS, luego a cada muestra se le agregó 1 mL de Qiazol (QIAGEN), y se trituraron, primero con una aguja 19G y enseguida con una aguja 21G hasta que no quedó ningún fragmento visible. A continuación, las muestras se centrifugaron 2 minutos a 15 000 g y se recuperó el sobrenadante. Teniendo en cuenta el volumen total de muestra obtenido, se añadió etanol a una proporción 1:1, se traspasaron las muestras a las columnas de extracción del kit y se centrifugaron 30 segundos a 15 000 g. Seguidamente se añadió 400  $\mu$ L de *Wash Buffer* a cada una de las muestras y se centrifugó 1 minuto a 15 000 g. A continuación, se preparó una *mix* que contenía 5  $\mu$ L de DNAasa I y 75  $\mu$ L de *Digestion Buffer*. Se añadieron 80  $\mu$ L de *mix* a las muestras y se dejaron incubar durante 15 minutos a temperatura ambiente. Seguidamente, se añadieron 400  $\mu$ L de *Direct-zol* y se centrifugó 1 minuto a 15 000 g, un paso que se repitió dos veces. Antes de eluir las muestras se lavaron con 700  $\mu$ L de *Wash Buffer* y se centrifugaron 1.5 minutos a 15 000 g. Finalmente, el contenido de las columnas de extracción se eluyó en 30  $\mu$ L de *Elution Buffer* centrifugando 1 minuto a 15 000g y se recogió el eluido en tubos de 1.5 mL. Las muestras se conservaron a  $-80^{\circ}\text{C}$ .

### 2.7.7. Control de calidad de muestras de RNA con *Bioanalyzer*

La calidad y cantidad de las muestras de RNA se analizó a través del *Bioanalyzer 2100* (Agilent Technologies). Se utilizó el kit comercial *Bioanalyzer RNA Nano 6000* (Agilent Technologies) siguiendo las indicaciones del fabricante.

El *Bioanalyzer* utiliza la electroforesis capilar para analizar las moléculas de ácidos nucleicos, las cuales son separadas por tamaño a través de la aplicación de una corriente eléctrica. Durante el proceso se intercala un colorante entre las moléculas de RNA que permite seguir su migración. Posteriormente, las moléculas son detectadas mediante fluorescencia inducida por un láser, que es captada y convertida en un electroferograma al analizarse con el software del *Bioanalyzer*. De esta manera, el *Bioanalyzer* permitió cuantificar las muestras y determinar el Número de Integridad de RNA (RIN) en las mismas. El RIN tiene una escala de 1 a 10, donde 1 representa una muestra muy degradada y 10 una muestra en su mayoría intacta. Este valor se obtiene a partir de la *ratio* de rRNA 28S y 18S en las muestras (**Figura 50**).



**Figura 50. Ejemplo de dos electroferogramas de *Bioanalyzer*.** A la izquierda, electroferograma de una muestra de RNA con alta integridad para la que se calculó un valor RIN de 9.90. A la derecha, electroferograma de una muestra de RNA tan degradada que no fue posible calcular un valor RIN.

### 2.7.8. Retrotranscripción

Para sintetizar cDNA a partir de del RNA de cada una de las muestras obtenidas, se utilizó el Kit *QuantiTect Reverse Transcription* (QIAGEN). En primer lugar, se formó un RNA *mix* para cada muestra a partir de 1  $\mu\text{g}$  de RNA disuelto en 12  $\mu\text{L}$  de agua *nuclease-free* a los que se agregaron 2  $\mu\text{L}$  de tampón *gDNA wipeout*, para eliminar restos de DNA genómico que pudieran contaminar la muestra. Cada *mix* se incubó 2 minutos a 42°C y luego se conservaron en hielo. A cada tubo de reacción se le agregaron 6  $\mu\text{L}$  de *master mix*, compuesta por 1  $\mu\text{L}$  de enzima transcriptasa reversa, 1  $\mu\text{L}$  del *mix* de oligonucleótidos RT y 4  $\mu\text{L}$  de tampón *Quantiscript RT 5X*, obteniendo un volumen final de reacción de 20  $\mu\text{L}$  por cada muestra. Finalmente, se incubaron 15 minutos a 42°C para producir la reacción de retrotranscripción (RT) y luego 3 minutos a 95°C para inactivarla. Los productos se conservaron a -20°C hasta utilizarse en la qPCR.

### 2.7.9. PCR cuantitativa

Para analizar los niveles de expresión de *SCN5A*, se realizó una qPCR utilizando sondas TaqMan. El fluoróforo en el extremo 5' de las sondas fue FAM y el motivo *quencher* en el extremo 3' estuvo formado por NFQ-MGB.

Para cada muestra, se hicieron 3 réplicas técnicas a partir del producto de la retrotranscripción. La qPCR se realizó en una placa de reacción *MicroAmp™ Optical 384-Well* (Thermo Fisher) que contenía los siguientes componentes: 5  $\mu\text{L}$  de tampón TaqMan, 0.5  $\mu\text{L}$  de la sonda TaqMan correspondiente, 3.5  $\mu\text{L}$  de agua libre de nucleasas y 1  $\mu\text{L}$  del cDNA correspondiente, para un volumen total de 10  $\mu\text{L}$  por reacción. Las reacciones se hicieron en el sistema *QuantStudio™ 7 Flex Real-Time PCR* (Thermo Fisher) utilizando el *software QuantStudio™ Real-Time PCR* (v1.2), indicando que se realizara un análisis comparativo Ct ( $\Delta\Delta\text{Ct}$ ) de una placa de 384 pocillos con reactivos TaqMan. En una primera etapa se llevaron las muestras de temperatura ambiente a 95°C subiendo 1.9°C/segundo. En la segunda etapa, las muestras se sometieron a 40 ciclos

en los que primero se mantienen 1 segundo a 95°C y luego se llevaron a 60°C bajando 1.6°C/segundo, manteniéndose a esta temperatura 20 segundos (en este punto se recogen los datos), para luego comenzar un nuevo ciclo.

### 2.7.9.1. Análisis de los datos de qPCR

El análisis de los datos obtenidos de las qPCR se hizo siguiendo las pautas y los métodos de la revisión de Taylor et al.<sup>368</sup>, con el objetivo de obtener datos de calidad de manera rigurosa (**Figura 51**). El análisis de datos y los cálculos asociados para la cuantificación relativa implican varios pasos (**Figura 51**). Estos cálculos son complejos porque los datos de qPCR son relativos por naturaleza entre muestras o grupos de muestras, normalizados con genes de referencia. Además, las mediciones de qPCR se realizan en una escala logarítmica (valores Cq) realizando el análisis estadístico usando datos asociados a los valores de Cq (por ejemplo, valores  $\Delta\Delta Cq$  o usando expresión normalizada relativa transformada logarítmicamente), mientras que los niveles de expresión se informan después de la transformación lineal de los resultados de  $\Delta\Delta Cq$  ( $2^{\Delta\Delta Cq}$ ).

Los resultados de la qPCR se presentan en grupos biológicos formados por 3 réplicas biológicas. Por ejemplo, 3 rondas de diferenciación de la línea EXOG-CTCF -DOX conforman un grupo biológico, el cual se comparará con otro grupo biológico que consiste en las 3 rondas de diferenciación de EXOG-CTCF +DOX. Al tratar los datos de esta manera se evitó que la significancia estadística en los cambios de expresión recayera solamente en los replicados técnicos de los que se dispondría al comparar -DOX/+DOX en cada ronda de manera individual. Al contrario, lo se obtuvo es una respuesta apoyada por triplicados biológicos a la pregunta de que gRNAs alteran significativamente la expresión de *SCN5A*.

En el tratamiento de los datos de qPCR es importante entender que las muestras son pareadas. Es decir, cada ronda de diferenciación tiene su pareja. Al comparar grupos biológicos no sería correcto hacerlo usando directamente las medias de Cq de las 3 réplicas biológicas. Podrían existir diferencias entre las réplicas biológicas que no son relevantes para el estudio, y la única diferencia que representa el objeto del estudio es la que se observe entre la condición -DOX y +DOX de una misma ronda. Por lo tanto, la ronda 1 (R1) -DOX siempre deberá ser comparada con la R1 + DOX.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Gen	Ronda de diferenciación	Media Cq	ΔCq (Muestras pareadas)	Cuantifi. relativa (RQ) 2 <sup>ΔCq</sup>	Factor norm. GEOMEAN [RQ (HPRT+GAPDH)]	Expresión norm. (A)	Log <sub>2</sub> Expresión norm. (B)	Expresión grupo biológico (C)	log <sub>2</sub> (C) (D)	SD (B) (E)	Error estándar de la media (SEM) (F)	Intervalo de confianza: límite inferior	Intervalo de confianza: límite superior
PP1A	EXOG -DOX R1	24.49	0	1		$\frac{RQ (SCN5A)}{\text{Factor norm.}}$	$\log_2 (A)$	$\text{GEOMEAN}(A)$	$\text{SD}(B)$	$\frac{\text{SD}(B)}{\sqrt{n}}$	$2^{D-t*F}$	$2^{D+t*F}$	
PP1A	EXOG -DOX R2	23.57	0	1									
PP1A	EXOG -DOX R3	23.97	0	1									
PP1A	EXOG +DOX R1	24.24	0.25	1.19									
PP1A	EXOG +DOX R2	23.58	-0.01	1.01									
PP1A	EXOG +DOX R3	24.04	-0.07	0.95									
GAPDH	EXOG -DOX R1	21.65	0	1									
GAPDH	EXOG -DOX R2	21.99	0	1									
GAPDH	EXOG -DOX R3	21.12	0	1									
GAPDH	EXOG +DOX R1	21.35	0.30	1.23									
GAPDH	EXOG +DOX R2	21.79	0.19	1.14									
GAPDH	EXOG +DOX R3	21.10	0.02	1.01									
SCN5A	EXOG -DOX R1	25.86	0	1	1	1	0	} 1.00	0	0	0	1	1
SCN5A	EXOG -DOX R2	25.08	0	1	1	1	0						
SCN5A	EXOG -DOX R3	25.78	0	1	1	1	0						
SCN5A	EXOG +DOX R1	26.62	-0.75	0.59	1.21	0.49	-1.03	} 0.57	-0.82	0.18	0.11	0.41	0.78
SCN5A	EXOG +DOX R2	25.68	-0.59	0.66	1.07	0.62	-0.69						
SCN5A	EXOG +DOX R3	26.56	-0.78	0.58	0.98	0.60	-0.74						

(pie de figura en la siguiente página)

**Figura 51. Análisis de datos de la cuantificación relativa de la qPCR.** Tratamiento riguroso de los datos. Figura adaptada de Taylor *et al*<sup>668</sup>. En este ejemplo se muestran los resultados de la qPCR para la línea EXOG-CTCF, donde la expresión de *SCN5A* se mide entre grupos biológicos tratados y no tratados (+DOX vs -DOX), cada uno de los cuales contiene tres réplicas biológicas y tres réplicas técnicas (no se muestran los datos de las réplicas técnicas). *PP1A* y *GAPDH* se utilizan como genes *housekeeping*. En primer lugar, se calculó el ciclo cuantitativo medio (Cq medio; columna 3) de las réplicas técnicas para cada ronda de diferenciación de cada gen [por ejemplo, para el gen *PP1A* de la ronda EXOG -DOX R1 el Cq medio es 24.49]. Para calcular la diferencia relativa ( $\Delta Cq$ ) se restó el Cq de la ronda -DOX por el Cq de la ronda +DOX (columna 4). Es esencial tener en cuenta que las muestras son pareadas, por lo que debe compararse cada muestra individualmente, y no a partir de la media de las 3 rondas (muestras agrupadas). Es decir, para calcular la diferencia relativa de la primera ronda de diferenciación (R1), se comparó R1 + DOX con R1 -DOX. Así pues,  $\Delta Cq$  fue calculado restando la Cq media de R1 -DOX a la Cq media de R1 + DOX. Al ser muestras pareadas las rondas -DOX siempre obtienen una cuantificación relativa de 1, ya que se comparan consigo mismas.

La cuantificación relativa (RQ) se calculó a partir de  $\Delta Cq$  ( $2^{\Delta Cq}$ ) donde, en este caso, se asume una eficiencia de la reacción del 100% y, por lo tanto, se aplicó una base de dos (columna 5). RQ representa numéricamente el cambio entre los grupos biológicos para cada combinación de ronda y gen, antes de la normalización con los *housekeeping*. Para cada ronda de diferenciación, se determinó un factor de normalización a partir de la media geométrica de las RQ de los genes *housekeeping* asociados (columna 6). A continuación, la expresión normalizada relativa para el gen a evaluar (*SCN5A* en este caso) se calculó dividiendo la RQ por el factor de normalización (columna 7) seguido de la transformación logarítmica (columna 8). También se calculó la expresión media relativa normalizada (media del grupo biológico) calculando la media geométrica (columna 9), seguido de la transformación logarítmica (columna 10). La desviación típica (SD), el error estándar de la media (SEM) y el intervalo de confianza del 95% [en este caso, se utilizó una t-estadística de 4.3 basada en tres muestras por grupo biológico (2 grados de libertad)] se calcularon para cada grupo a partir de la expresión normalizada logarítmica (columnas 11–14).

## 2.8. Experimentos funcionales en condiciones de *knockdown* de SCN5A

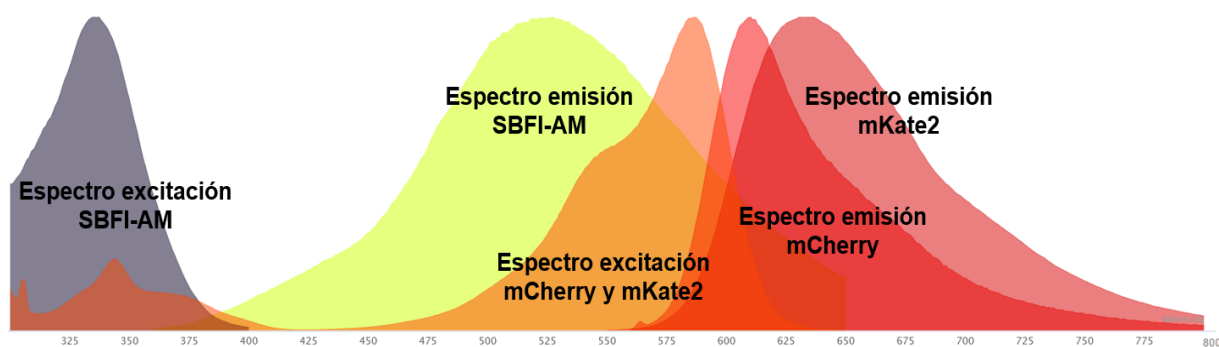
### 2.8.1. Medición de la concentración de sodio intracelular

Para determinar si los modelos *knockdown* de SCN5A en hiPSC-CMs presentan alteraciones en las corrientes de sodio, se realizaron ensayos para medir la concentración de sodio intracelular en aquellas guías que resultaron en una disminución de la expresión de SCN5A.

Se realizaron rondas de diferenciación siguiendo el protocolo del apartado 2.3.4., excepto durante el proceso de replaqueo, donde los cardiomiocitos se sembraron en placas de 384 pocillos de fondo transparente (Perkin-Elmer) previamente tratadas con 30 µg/mL de poli-D-lisina a una densidad de 32 000 células por pocillo (**Figura 53**). Se sembraron 12 replicados técnicos para cada condición, de los cuales 6 fueron tratados con DOX.

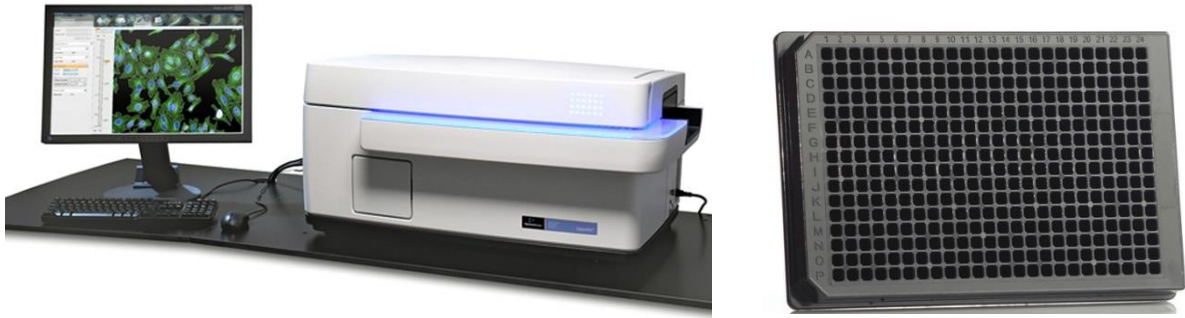
El colorante fluorescente de unión específica a sodio (SBFI-AM; Abcam) se reconstituyó con DMSO y se diluyó en medio SILAC que contenía Pluronic F127 al 0.02 % (Thermo Fisher). El día 30 de diferenciación los pocillos se lavaron con SILAC y se añadieron 50 µL por pocillo de solución SBFI a una concentración final de 3.5 µM. Se utilizó el sistema Operetta de imágenes de alto contenido (Perkin-Elmer; **Figura 53**) para adquirir señales de fluorescencia con una longitud de onda de excitación de 340 nm y una longitud de onda de emisión de 540 nm a las 4 horas después de la adición de SBFI. 30 minutos antes de la medición se añadió un colorante fluorescente de DNA (Hoechst; Thermo Fisher) a cada pocillo a una concentración 1:4 000. También se tomaron imágenes digitales de contraste de fase para delimitar el contorno de las células. El software de análisis e imágenes de alto contenido Harmony (Perkin-Elmer) permitió evaluar la intensidad de la fluorescencia en el interior de las células.

La elección de SBFI-AM como colorante fluorescente fue determinada por su espectro de emisión y excitación, debido a que consigue evitar el solapamiento con la emisión de mCherry y mKate2 (**Figura 52**), que inevitablemente eran expresados durante el experimento.



**Figura 52.** Perfiles de excitación y emisión para SBFI-AM y los fluoróforos expresados por la línea WTB CRISPRi Gen1B (mCherry y mKate2). Figura creada con FPbase (<https://www.fpbase.org/spectra>).

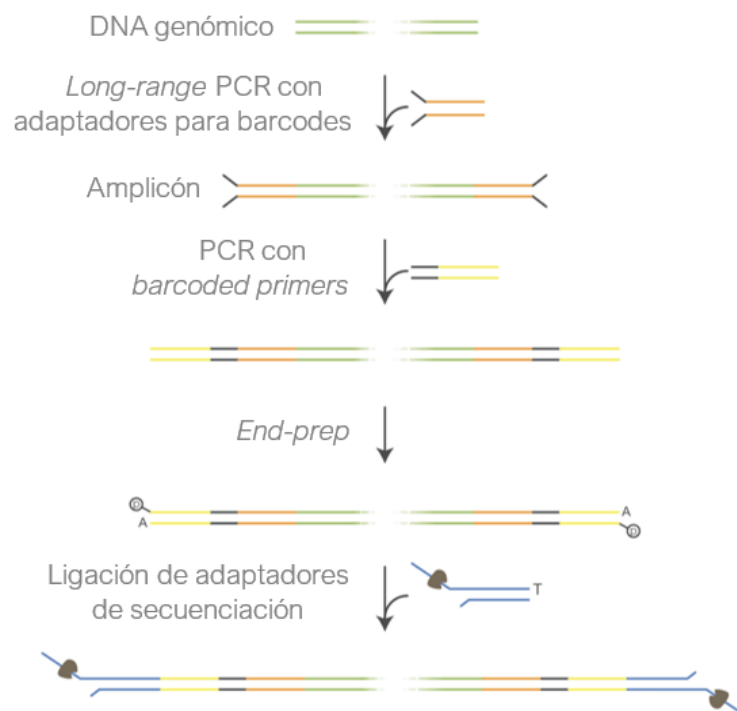




**Figura 53. Sistema Operetta y placa de 384 pocillos utilizada para replaquear los hiPSC-CM y medir el sodio intracelular en el sistema Operetta.** Imágenes obtenidas de Perkin Elmer (<https://www.perkinelmer.com>).

## 2.9. Secuenciación de haplotipos mediante Oxford Nanopore Sequencing

En los siguientes apartados se describen los pasos seguidos para realizar las secuenciaciones con ONT, resumidos y esquematizados en la **Figura 54**:



**Figura 54. Esquema de los pasos seguidos para la secuenciación mediante ONT.** Primero se amplificó la zona del haplotipo con *long-range* PCR. Luego, se añadieron *barcodes* a los amplicones para diferenciar a cada paciente. A continuación, se prepararon las *DNA libraries* juntando todas las muestras y realizando una reacción de *end-prep* y la ligación de los adaptadores de secuenciación. Finalmente, las muestras se cargaron en el secuenciador. Figura adaptada de ONT (<https://nanoporetech.com>).

### 2.9.1. Long-range PCR

De cada muestra de DNA genómico se amplificó por *long-range* PCR la región del haplotipo que contiene las 7 variantes. Se utilizó el protocolo del kit *Supreme NZYLong polymerase* (NZYTech): 5  $\mu$ l de *buffer* de reacción 10X, 5  $\mu$ l de mezcla de dNTPs (2.5 mM cada uno), 1.75  $\mu$ l de cebador Fw (10 nM), 1.75  $\mu$ l de cebador Rv (10 nM), 300 ng de DNA genómico, 1  $\mu$ l de *Supreme NZYLong DNA polymerase* y agua PCR *grade* hasta 50  $\mu$ l. Las condiciones de la PCR fueron: (i) 5 minutos a 94°C, (ii) 30 ciclos de 20 segundos a 94°C, 30 segundos a 60.5°C y 14 minutos a 68°C, (iii) 21 minutos a 68°C, (iv) mantener a 4°C. Los productos de PCR que mostraban una baja concentración de DNA se limpiaron con ExoSap-IT (ThermoFisher) y se sometieron a una segunda ronda de amplificación (*nested* PCR). La *nested* PCR se realizó con 1  $\mu$ l de producto de PCR limpio (a una dilución de 1/500) siguiendo el mismo protocolo de PCR que se ha descrito anteriormente. Todas las secuencias de cebadores se diseñaron para incluir adaptadores universales necesarios para el *barcoding* posterior de las muestras durante la preparación de la *library* de Oxford Nanopore. La primera PCR produjo un amplicón de 13 kb mientras que la PCR anidada produjo un amplicón de 12.84 kb.

### 2.9.2. Preparación de *libraries* y secuenciación con el dispositivo minION

Para preparar la DNA *library* se utilizaron las regiones amplificadas por *long-range* PCR del apartado anterior (en total 86 muestras correspondientes a 86 pacientes con SBr de la cohorte de pacientes, apartado 1.1.1. de Materiales) usando el *PCR barcoding (96) amplicons kit SQK-LSK109*.

La *library* se preparó siguiendo las recomendaciones del fabricante, aunque se incluyeron algunas optimizaciones relevantes. En primer lugar, los productos de la *long-range* PCR se purificaron con el kit *Agencourt AMPure XP beads* (Beckman-Coulter) usando un volumen de 0.41X de *magnetic beads*, seguido de limpiezas con etanol al 75%. Todas las muestras fueron cuantificadas con Qubit después de la purificación. A continuación, se prosiguió con la adición de los *barcodes* mediante una PCR con *barcoded primers*. Cada una de las 86 muestras recibió 86 *barcodes* diferenciados. La condición requerida por la PCR de *barcoding* es tener una concentración de DNA inicial de 0.5 nM en 50  $\mu$ l. En nuestro caso, para un amplicón de 13kb, se requirieron 200 ng para cada reacción. Se utilizó el protocolo del kit *SQK-LSK109*: 200 ng de producto de PCR en un volumen de 24  $\mu$ l de agua, 1  $\mu$ l de *barcode* a 10  $\mu$ M, 25  $\mu$ l de *LongAmp Taq 2x master mix*. Las condiciones de la PCR fueron: (i) 3 minutos a 95°C, (ii) 14 ciclos de 15 segundos a 95°C, 15 segundos a 62°C y 13 minutos a 65°C, (iii) 20 minutos a 65°C, (iv) mantener a 4°C. Después de la PCR de *barcoding* las muestras se purificaron con *Agencourt AMPure XP beads* con las mismas condiciones que anteriormente.

A partir de este punto, las 86 muestras se juntaron en un solo *pool* añadiendo cantidades equimolares de cada una de ellas. Antes de proseguir con el paso de *end-prep*, necesario para

garantizar que cada molécula esté libre de extremos cohesivos y que contenga grupos 5' fosfato y 3' hidroxilo, se hizo una purificación adicional del *pool* no especificada en el protocolo. Este paso fue clave para eliminar productos inespecíficos de la muestra y facilitar el posterior análisis de la secuenciación. La purificación consistió en usar el producto *Short Read Eliminator XS* (Circulomics), siguiendo las instrucciones del fabricante, el cual elimina de manera progresiva todos los fragmentos de DNA de menos de 10 kb y casi al completo todos los fragmentos de menos de 5kb. Esta purificación se realizó 2 veces consecutivas para mejorar la pureza de nuestro producto.

La reacción de *end-prep* se realizó siguiendo las instrucciones de *NEBNext Companion module ligation*: 1  $\mu\text{L}$  de DNA CS, 1.5  $\mu\text{g}$  de *pool* en 47  $\mu\text{L}$  de agua *nuclease-free*, 3.5  $\mu\text{L}$  de *NEBNext FFPE DNA Repair buffer*, 2  $\mu\text{L}$  de *NEBNext FFPE DNA Repair Mix*, 3.5  $\mu\text{L}$  de *Ultrall End-prep reaction buffer* y 3  $\mu\text{L}$  *Ultrall End-prep enzyme mix*. La reacción se mantuvo 5 minutos a 20°C y después 5 minutos a 65°C.

Después de la reacción de *end-prep*, el *pool* se purificó una vez más con *Agencourt AMPure XP beads* y se resuspendió en 60  $\mu\text{L}$ . A continuación, se realizó la ligación de los adaptadores de secuenciación con los siguientes reactivos: 60  $\mu\text{L}$  del *pool*, 25  $\mu\text{L}$  de *ligation buffer*, 10  $\mu\text{L}$  de *NEBNext Quick T4 DNA Ligase*, 5  $\mu\text{L}$  de *Adapter Mix*. La reacción se mantuvo 10 minutos a temperatura ambiente. Se hizo un paso final de purificación con *AMPure XP beads* usando *long fragment buffer* en lugar de etanol para los lavados. Este *buffer* elimina fragmentos de DNA con menos de 3 kb. La biblioteca se resuspendió en 15  $\mu\text{L}$  de *buffer EB*. Finalmente, se preparó la mezcla final para cargar la *library* en el secuenciador: 37.5  $\mu\text{L}$  de *Sequencing buffer*, 25.5  $\mu\text{L}$  de *loading beads* y 12  $\mu\text{L}$  de la biblioteca de DNA (300-400 ng).

La DNA *library* obtenida se secuenció en un minION (*Flow cell R9.4*) en el Centro de Medicina Molecular e Investigación de Enfermedades Crónicas (CiMUS, Santiago de Compostela, España). Se procedió a cargar todo el volumen de la biblioteca a través del *SpotON* del minION (**Figura 55**). Se inició la secuenciación en el software *minKnow* indicando que se hiciese el *base-calling* a la vez que iba secuenciando y que se realizara el *trimming* de los *barcodes*. Se partió de una *flow cell* con unos 1500 nanoporos disponibles.

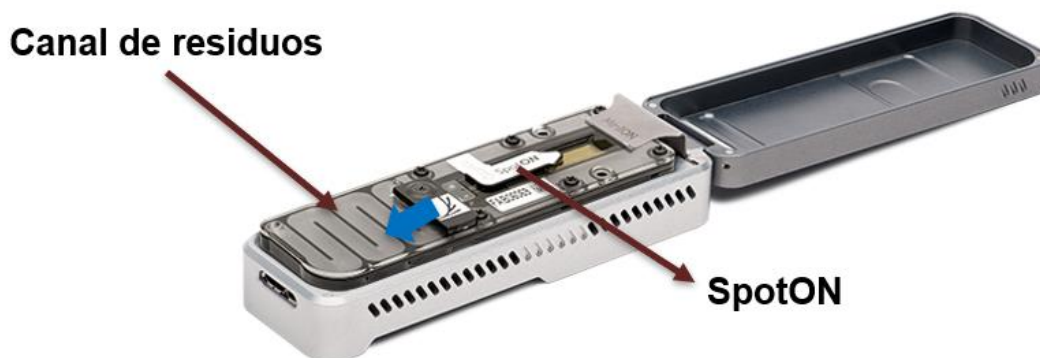


Figura 55. Dispositivo minION con una *flowcell* preparada para ser usada.

### 2.9.3. Análisis de la secuenciación de lecturas largas

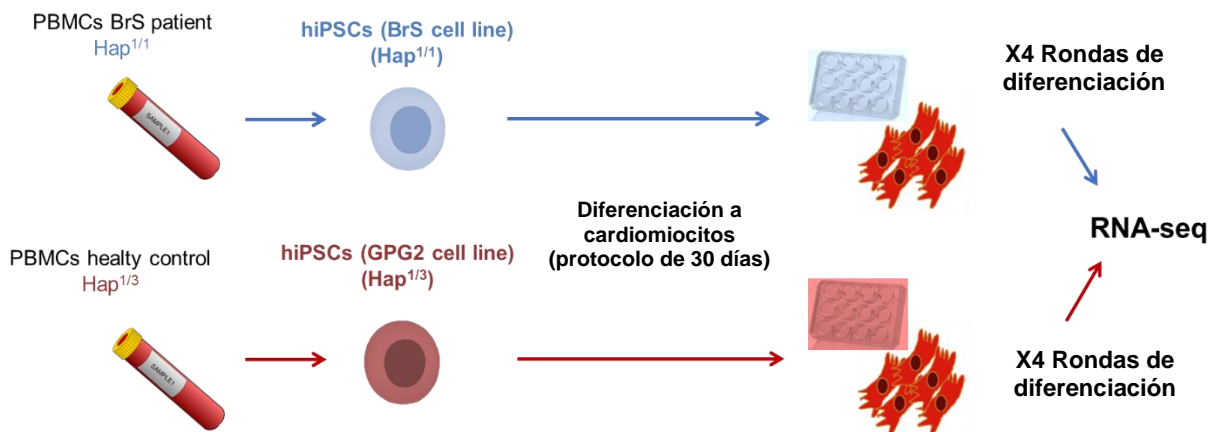
El *base calling* de las lecturas en crudo se realizó utilizando el *base caller* Guppy (v3.2.10) de Oxford Nanopore para generar archivos FASTQ que contenían los archivos de lecturas de secuenciación de cada muestra de SBr. Las lecturas se alinearon con el genoma de referencia humano (GRCh37/hg19) usando minimap2 (v2.17) con el preajuste -ont. Se usó SAMtools para convertir las alineaciones de minimap2 al formato BAM y se eliminaron las lecturas de mapeo múltiple y las alineaciones complementarias. SAMtools también se usó para eliminar lecturas con *soft-clipping* y lecturas de menos de 8 kb (longitud mínima que estimamos necesaria para la determinación del *phasing*).

## 2.10. Experimentos de RNA-seq

### 2.10.1. Preparación de las muestras y secuenciación

Para comparar el perfil de expresión cardíaco de un paciente con SBr que presenta el haplotipo homocigoto Hap<sup>1/1</sup> con un individuo sano (Hap<sup>1/3</sup>), se tuvo que lograr RNA de cardiomiocitos para cada uno de ellos. En los apartados anteriores se explicó las características de estos individuos (1.1.2. de Materiales), las características de las líneas celulares (1.2.4. de Materiales), el protocolo de diferenciación para obtener los hiPSC-CMs (apartado 2.3.4. de Métodos y **Figura 43**) y el protocolo de extracción de RNA (apartado 2.7.6. de Métodos).

Para cada caso se realizaron 4 rondas de diferenciación independientes (réplicas biológicas), para minimizar el impacto de la variabilidad experimental entre rondas (**Figura 56**).



**Figura 56. Pasos seguidos para obtener las muestras para el experimento de RNA-seq.** Los PBMCs recogidos de cada individuo fueron reprogramados a hiPSCs, para posteriormente ser diferenciados a cardiomiocitos (4 rondas de diferenciación para cada individuo).

El RNA extraído de las rondas de diferenciación pasó por controles de calidad antes de ser enviado al Centro de Regulación Genómica (CRG, Barcelona) para realizar el RNA-seq. En primer lugar, se midió el RNA en el Nanodrop para determinar la concentración y las *ratios* de pureza. También se utilizó el *Bioanalyzer* tal y como se describe en el apartado 2.7.7. para determinar el RIN de las muestras (**Tabla 15**). Se consideró que un RIN superior a 9 era aceptable para el experimento.

**Tabla 15.** Parámetros de calidad registrados en las muestras usadas para RNA-seq.

Nombre muestra	Concentración (ng/μL)	Ratio Nanodrop 260/280	Ratio Nanodrop 260/230	RIN
Control Ronda 1	100	1.97	1.71	9.8
Control Ronda 2	130	2.01	1.95	10
Control Ronda 3	129	1.96	1.87	10
Control Ronda 4	137	2.00	1.95	10
SBr Ronda 1	162	1.99	1.84	10
SBr Ronda 2	221	2.00	1.98	10
SBr Ronda 3	218	2.00	1.86	10
SBr Ronda 4	209	1.99	2.04	10

Una vez el CRG recibió las muestras de RNA, la secuenciación se realizó en el dispositivo de secuenciación masiva HiSeq3000 (Illumina). La secuenciación se puede realizar desde un único extremo de las moléculas de DNA (*single-end*) o desde ambos (*paired-end*). En cuanto a las lecturas, se pueden elegir diferentes longitudes de lectura según el sistema y la *flowcell*. En este caso la secuenciación se realizó *paire-end* con una longitud de lecturas de 75 pb. Los experimentos de RNA-seq tienden a beneficiarse de lecturas *paired-end* más largas (como 75 pb o 100 pb) para permitir una cobertura más completa de los transcritos y la identificación de nuevas variantes. El centro de secuenciación (CRG) se comprometió a proporcionar 50 millones de lecturas para cada una de las muestras secuenciadas. Los resultados se obtuvieron en archivos con formato FASTQ, conteniendo las lecturas de secuenciación sin procesar.

### 2.10.2. Análisis del experimento RNA-seq

Se utilizó un *workflow* de análisis bioinformático (**Figura 57**) con el propósito de analizar los datos de RNA-seq. El análisis pretendía comparar los valores de expresión entre los dos grupos (paciente SBr y control) con el fin de obtener todas las diferencias en los patrones de expresión entre ambos. Para ello, se trabajó con el clúster *Picasso* perteneciente al Servicio de

Supercomputación y Bioinformática de la Universidad de Málaga (SCBI), realizando una conexión remota mediante cliente el software MobaXterm.

Para realizar el análisis de expresión de las muestras de RNA-seq se utilizó un *workflow* acorde al objetivo del análisis, llamado DEG *Workflow* y que tiene como núcleo de análisis la herramienta *ExpHunter Suite*<sup>369,412</sup>. La metodología de *ExpHunter Suite* está implementada como dos módulos principales de análisis (**Figura 57**). Se pueden ejecutar de forma interactiva como paquetes de R convencionales o, como en el caso de esta tesis, directamente desde una línea de comandos como *scripts*, ejecutados a través de MobaXterm. El **módulo Degenes Hunter** realiza un control de calidad, un filtrado basado en la expresión, un análisis de expresión diferencial y un análisis de co-expresión. El **módulo funcional Hunter** realiza análisis de enriquecimiento funcional, utilizando como entrada el resultado del módulo Degenes Hunter. Ambos producen múltiples archivos e informes como se muestra en la **Figura 57**. El código de *ExpHunter Suite* está disponible en la base de datos bioinformática Bioconductor a través de: <https://bioconductor.org/package/ExpHunterSuite>. La última versión del código se puede encontrar en <https://github.com/seoanezonjic/ExpHunterSuite>.

De manera resumida, DEG *workflow* toma muestras de RNA-seq de Illumina sin preprocesar y devuelve un informe de composición de las muestras (cómo ha salido el experimento y la secuenciación), y los resultados del programa DEGenesHunter. Este programa es capaz de (i) encontrar genes que cambian su expresión entre dos grupos de muestras de una forma significativa, (ii) establecer módulos de co-expresión, es decir, definir conjuntos de genes que se expresan de la misma forma a lo largo de las muestras y (iii) analizar la composición funcional de los genes con expresión diferencial y de los módulos de co-expresión.

DEG *Workflow* se compone de varias fases que permitió realizar un análisis de RNA-seq al completo, con el único input de las muestras de secuenciación (archivos FASTQ) e información adicional del experimento. Estas fases se llevaron a cabo en orden a partir del ejecutable *daemon.sh*. y se explican con más detalle a continuación:

La **fase 0** consistió en la configuración manual del *workflow* antes de ejecutar las fases posteriores. En primer lugar, se debió crear el sistema de carpetas que guarda los datos y las ejecuciones correspondientes al proyecto. Para empezar a trabajar, se necesitaron los datos de secuencias (archivos FASTQ). Los archivos FASTQ se descargaron en la carpeta "*original\_data*". Las muestras originales tenían un nombre por defecto del servicio de secuenciación y generalmente no suele ser interpretable. El siguiente paso fue renombrar los archivos para que se pudieran interpretar fácilmente los resultados. Se escogió un nombre corto identificativo del experimento para cada archivo.

La **fase 1** consistió en la descarga e indexado del transcriptoma o genoma que se usó como referencia para el mapeo (GRCh38 *assembly*). Esta fase fue preparatoria para las siguientes. Se

encargó de descargar todos los archivos necesarios para ejecutar el mapeo. En este caso se descargaron las referencias del genoma humano ejecutando el *script* “./daemon.sh 1a”.

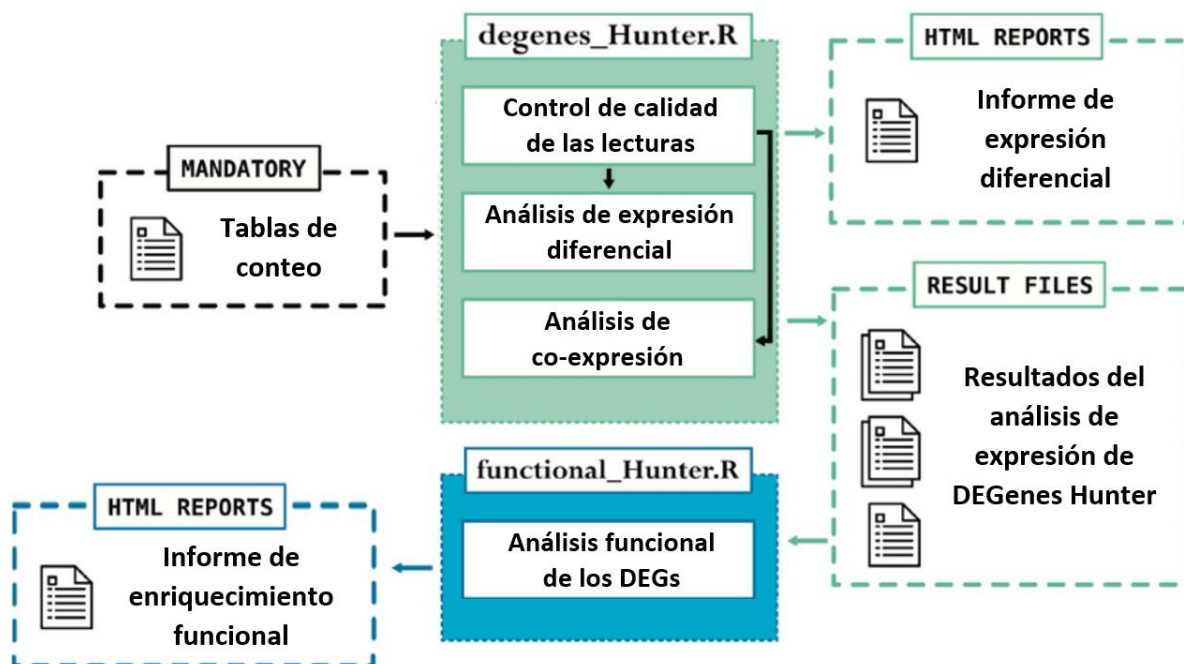
La **fase 2** consistió en el protocolo de pre-procesamiento (limpiar las muestras de lecturas de poca calidad, contaminantes y artefactos de secuenciación) y mapeo (clasificar las lecturas en función del gen al que pertenecen). Después se transformaron los datos de mapeo a un formato llamado “matriz de conteos”, que es un archivo de texto plano tabulado que indica cuantas lecturas de cada gen hay en cada muestra de secuenciación. Esta matriz de conteos, fue el input principal de DEGenesHunter (**Figura 57**). La forma de ejecutar esta sección fue ejecutando el *script* “./daemon.sh 2”, que envió todos los trabajos al sistema de colas.

La **fase 3** consistió en la ejecución del análisis de expresión diferencial con DEGenesHunter. Tras conseguir las tablas de conteo para cada muestra hubo que combinarlas y ejecutar DEGenesHunter. Al ejecutar el *script* “./daemon.sh 3” se hicieron 3 cosas:

1. Creación de un informe de mapeo, que describió el estado de las muestras, pudiendo encontrar errores en la secuenciación, etiquetado y en la elección de parámetros de pre-procesamiento y mapeo.
2. Combinación de las tablas de conteo en una matriz de conteo.
3. Ejecución de DEGenesHunter. Para cada comparación se creó una carpeta dentro de las cuales se ejecutó DEGenesHunter. Al ejecutar “./daemon.sh 3” se envió un solo trabajo al sistema de colas que lo llevó todo a cabo. Al finalizar ya estaban disponibles los resultados del análisis de expresión diferencial.

DEGenesHunter, con la configuración por defecto de *ExpHunter Suite*, realiza un análisis de expresión diferencial, es decir, busca el conjunto de genes que se expresan diferente y de manera significativa en los tratamientos con respecto a los controles. Además, realiza un análisis de co-expresión que devuelve una serie de módulos o clústeres de co-expresión (**Figura 57**). Dentro de estos módulos, todos los genes cambian su expresión con la misma tendencia a lo largo de las muestras, con la excepción del módulo 0, que es donde se guardan los genes que no se parecen estadísticamente a ningún otro. La detección de genes expresados diferencialmente (DEGs) se realizó con EdgeR, Limma, DESeq2 y NOIseq, que son los 4 paquetes de detección usados por DEGenesHunter. Se utilizó un valor p ajustado (o FDR) de <0.01 para determinar los DEG significativos.

La **fase 4** consistió en analizar la composición funcional de los resultados obtenidos en la fase 3. Esta herramienta usó los resultados del análisis de expresión de DEGenes Hunter para buscar



**Figura 57. Esquema del workflow implementado en el análisis de las muestras de RNA-seq.** La caja verde representa el módulo DEgenes\_Hunter relacionado con el análisis de expresión diferencial. La caja azul representa el módulo funcional\_Hunter relacionado con el análisis funcional. Las cajas con bordes discontinuos representan archivos de entrada o salida. Figura adaptada de Jabato *et al.*<sup>369</sup>.

qué funciones representan los genes con expresión diferencial y los genes de cada módulo o clúster. Esta herramienta se ejecutó en el sistema de colas mediante el *script* “./daemon.sh 4a\_”.

Por último, con el objetivo de reducir el número de genes expresados de manera diferencial y filtrar aquellos que realmente estuviesen alterados por la condición del estudio (SBr/haplotipos), se realizó un análisis de similitud semántica. Para ello se utilizaron términos de ontología del fenotipo humano (HPO), centrándose en calcular la similitud semántica entre SBr y la similitud funcional entre genes. Se siguieron estos pasos:

1. Se utilizó el perfil de HPO de la enfermedad (archivo con la lista de términos; **Tabla A-1** del Anexo 2).
2. Se descargaron las relaciones HPO-gen del consorcio *Monarch initiative*. Este consorcio las extrae de distintas bases de datos que, a su vez, las referencian de la curación manual de bibliografía, mayormente.
3. De todos los genes, se seleccionaron solo los que son DEGs en el análisis diferencial del RNA-seq.
4. Se computó la similitud semántica de los HPO de cada gen respecto a los HPO de la enfermedad.



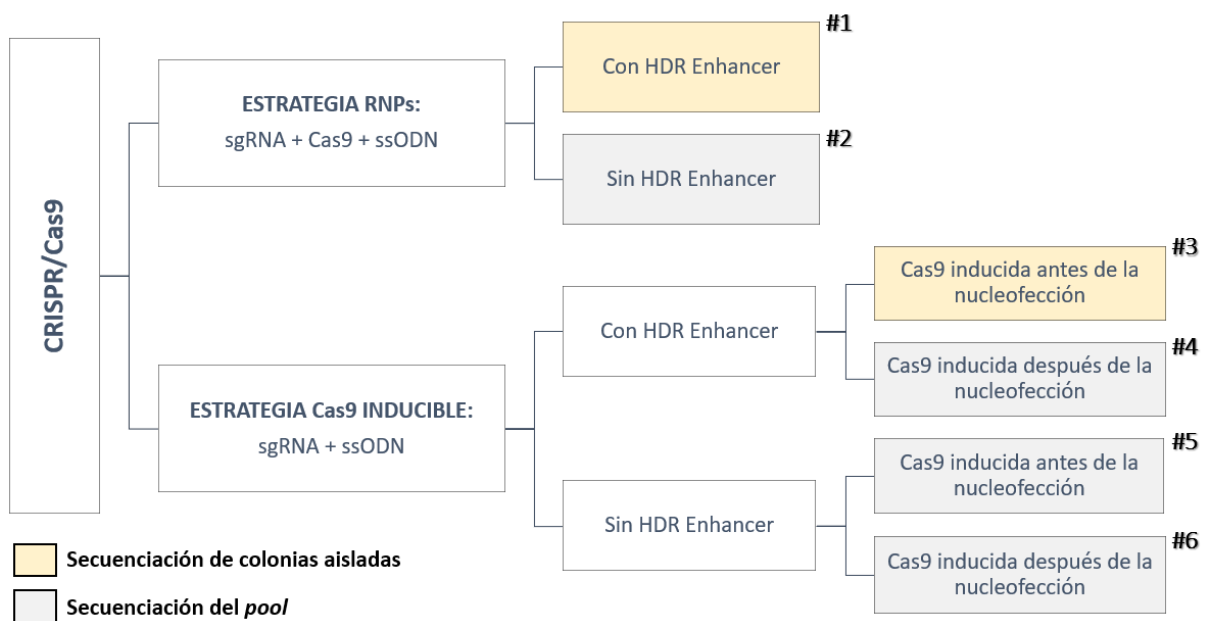
## 2.11. Edición del haplotipo con CRISPR/Cas9

### 2.11.1. Diseño de 6 estrategias de edición diferentes

Con el objetivo de estudiar el haplotipo asociado a SBr en un contexto isogénico, se decidió editar la línea de SBr Hap<sup>1/1</sup> con CRISPR/Cas9. Para esto, se decidió pasar de Hap<sup>1/1</sup> a Hap<sup>3/3</sup>, con el objetivo de obtener una línea idéntica a la SBr con un haplotipo sin riesgo asociado a SBr.

Para obtener Hap<sup>3/3</sup> es necesario editar 3 de las 7 variantes de manera homocigota. La estrategia de edición se basó en el sistema CRISPR/Cas9 y, dado las bajas eficiencias de la técnica en hiPSC, se optó por editar cada variante una por una, hasta lograr la edición homocigota de esta para luego pasar a la siguiente. La línea celular usada fue la BS hiPSC, ya que contiene el haplotipo 1 en homocigosis (Hap<sup>1/1</sup>).

Para encontrar la manera más eficaz de generar los *knock-in*, se diseñaron 6 estrategias de edición (**Figura 58**). La estrategia más común es con RNPs (ribonucleoproteínas), una estrategia que consiste en introducir directamente en la célula la guía (sgRNA), el donador (ssODN) y Cas9, teniendo actividad desde el momento en que entra en la célula. En nuestro caso se usó la HiFi-Cas9, una versión más eficaz de la Cas9 WT.



**Figura 58. Esquema de las 6 estrategias utilizadas para optimizar la eficiencia de la edición con CRISPR/Cas9.** En el caso de la estrategia con RNPs, solamente hay la variante con o sin HDR Enhancer debido a que la Cas9 se añade en el momento de la nucleofección. En el otro caso Cas9 debe ser expresada por la célula con anterioridad o posteriormente a la nucleofección.

Se probaron distintas variantes:

- HiFi-Cas9 con presencia o no de HDR *Enhancer*, un componente que aumenta la eficiencia de edición por HDR (**Figura 58**; estrategias #1 y #2).
- Uso de una versión modificada de la línea BS hiPSC con un sistema de Cas9 inducible por DOX (apartado 1.2.5. de Materiales). En este caso se evitó nucleofectar con Cas9, y solo se introdujo la sgRNA y el ssODN. Por tanto, Cas9 se tuvo que inducir para que la célula tuviese la proteína disponible para la edición. Según el momento de inducción:
  1. DOX 24h antes de la nucleofección de sgRNA y ssODN. Se probó con y sin HDR *enhancer* (**Figura 58**; estrategias #3 y #5)
  2. DOX en el momento de nucleofectar sgRNA y ssODN. Se probó con y sin HDR *enhancer* (**Figura 58**; estrategias #4 y #6).

### 2.11.2. Protocolo de edición CRISPR/Cas9

El proceso de edición para todas las estrategias explicadas en el apartado anterior siguió los mismos pasos (**Figura 59**), aunque con diferencias en la manera de llevarlos a cabo.

En primer lugar, se hidrataron los sgRNA que llegaron liofilizados con *buffer* TE a una concentración de 100  $\mu$ M. Los sgRNA se diluyeron a una concentración de trabajo de 75  $\mu$ M. La hidratación de los ssODN se realizó de la misma manera.

Para la nucleofección se usó el aparato *Amaxa Nucleofector 2b device* (Lonza) con el programa A-023. Se tuvieron en cuenta las siguientes variaciones:

- En las estrategias con uso de HDR *enhancer*, este se añadió a la mezcla de nucleofección para entrar a la célula con el resto de componentes.
- En todos los casos se añadió Ri a la mezcla para aumentar la supervivencia de las células.
- En las estrategias con la línea Cas9-inducible se añadió DOX a 8  $\mu$ g/mL (en la mezcla de nucleofección o bien 24 horas antes de esta).

Por lo tanto, para cada estrategia se nucleofectaron los siguientes componentes:

- **Estrategia #1:** sgRNA + HiFi-Cas9 + ssODN + HDR + Ri
- **Estrategia #2:** sgRNA + HiFi-Cas9 + ssODN + Ri
- **Estrategia #3:** sgRNA + ssODN + HDR + Ri (+ DOX 24 horas antes)
- **Estrategia #4:** sgRNA + DOX + ssODN + HDR + Ri
- **Estrategia #5:** sgRNA + ssODN + Ri (+ DOX 24 horas antes)
- **Estrategia #6:** sgRNA + DOX + ssODN + Ri

Después del proceso de nucleofección las células se incubaron durante 24 horas a 32°C (shock térmico para aumentar la eficiencia de la edición por HDR<sup>370</sup>). Pasado este tiempo se cambió el medio a mTeSR1 y se esperó a obtener una confluencia del 70% aproximadamente.

En el caso de las estrategias #1 y #3, se realizaron sendos pases mediante diluciones para obtener una concentración de aproximadamente 10 células/cm<sup>2</sup> y conseguir células individuales. Las células sobrantes se congelaron para ser secuenciadas y analizadas con TIDER. En las estrategias #2, #4, #5 y #6 no se hizo subclonaje, por lo que se congelaron directamente las células del *pool* editado para ser posteriormente secuenciado y analizado con TIDER.

Después del subclonaje se cambió el medio mTeSR1 cada 2 días mientras hubiese poca confluencia. Al cabo de 7-10 días, se empezaron a obtener colonias a partir de las células individuales de un tamaño apto para ser *picadas*. Para la estrategia #1 se picaron 200 clones y se sembraron en placas de 96 pocillos. Para la estrategia #2 se picaron 72 clones. Una vez las células alcanzaron la confluencia se usaron la mitad de los clones para secuenciarse y la otra mitad para congelarse.

### 2.11.3. Comparativa de las 6 estrategias mediante el software TIDER

TIDER proporciona una evaluación rápida y fiable de la edición del genoma dirigida con ssODNs (*knock-in*). Cuantifica la eficiencia de HDR en una muestra editada al descomponer los datos de las secuencias de tres reacciones simples de Sanger.

Así pues, para calcular la eficiencia de HDR en las 6 estrategias de edición usadas, se proporcionaron las 3 secuenciaciones por Sanger que requiere el programa TIDER (<http://shinyapps.datacurators.nl/tider/>). Se realizaron las siguientes secuenciaciones:

- **Control:** DNA de células WT, no editadas.
- **Referencia:** DNA que contenga la mutación introducida en el ssODN.
- **Referencia:** secuenciación de los *pools* de células de cada estrategia.

Para llevarlas a cabo se usó la estrategia de cebadores explicada en el apartado 1.4.3. de Materiales.

## 2.12. Análisis estadísticos

Todos los análisis estadísticos se realizaron con el programa GraphPad Prism 8. Las diferencias se consideraron significativas en  $p \leq 0.05$  (\*),  $p \leq 0.01$  (\*\*) y  $p \leq 0.001$  (\*\*\*).

Los datos se expresan como media  $\pm$  SD o SEM, como se indica en las leyendas de las figuras y tablas. La significación estadística se calculó mediante la prueba t de Student de dos colas, o mediante *one-way* ANOVA seguido del test de comparaciones múltiples de Dunnett cuando se querían comparar más de 2 grupos al mismo tiempo.

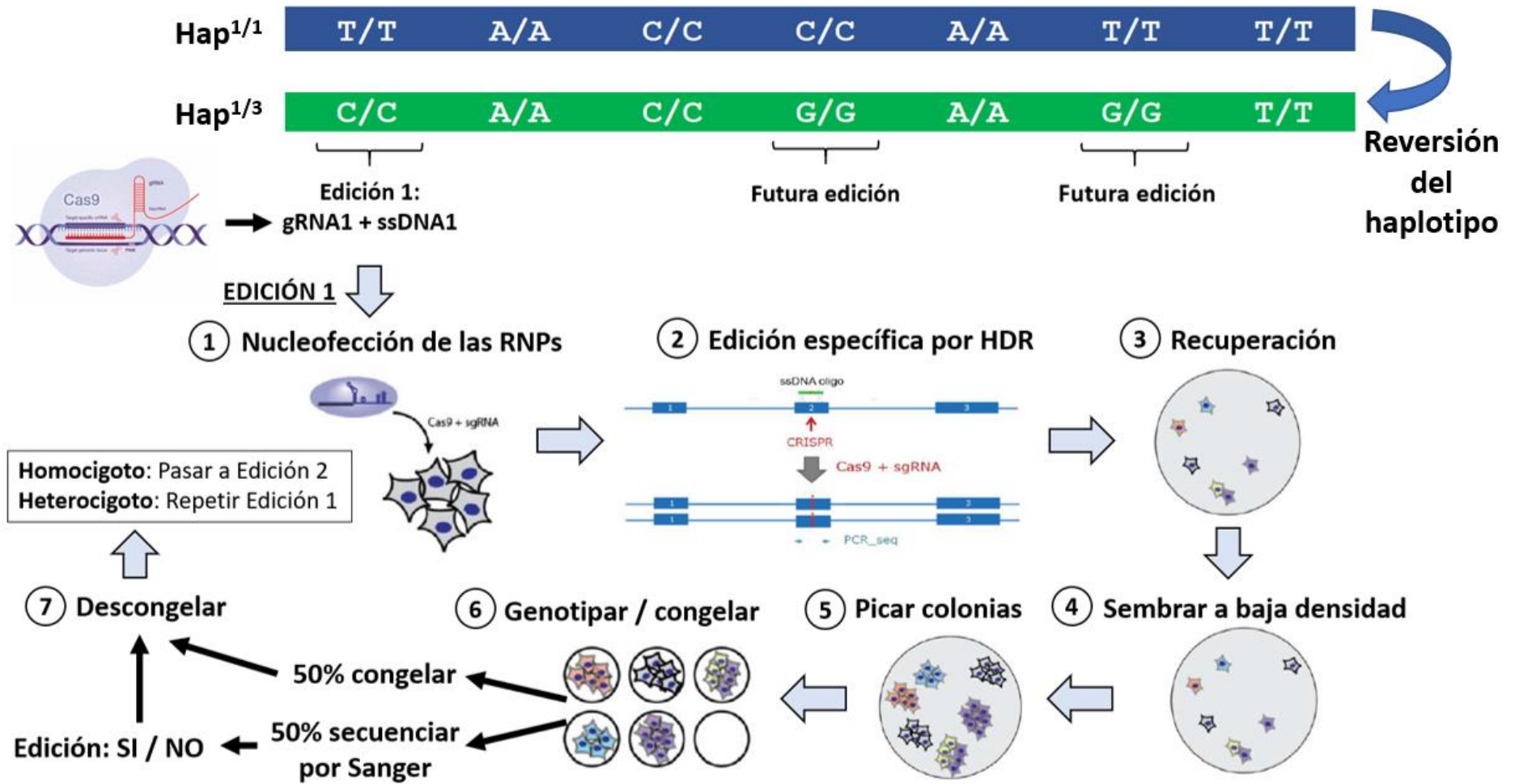
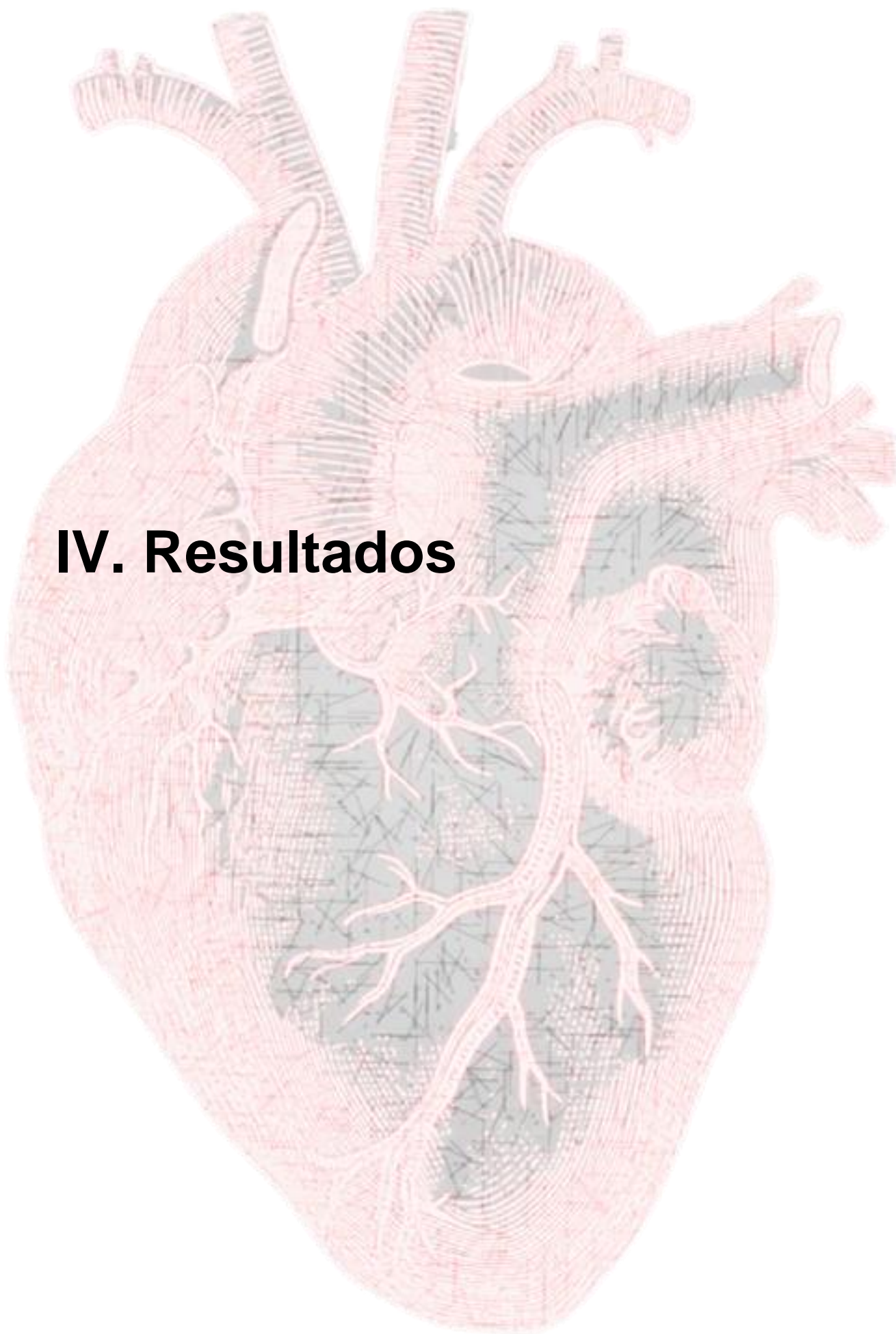


Figura 59. Pasos seguidos durante el proceso de edición. Etapas en común que siguen todas las estrategias de edición usadas, a pesar de que en algunos pasos hubo variaciones entre métodos.



## **IV. Resultados**





## CAPÍTULO I. Determinación de las lisinas acetiladas en GATA4 y su efecto sobre el promotor *SCN5A*

### 1.1. Resultados previos a esta tesis y contexto del estudio

La transcripción del gen *SCN5A* está modulada por un balance de acetilación y desacetilación, tanto a nivel de las histonas de las regiones reguladoras de *SCN5A*, como de los FT que regulan su expresión. Estudios anteriores del grupo han demostrado que el FT GATA4 es un activador transcripcional de la expresión del gen *SCN5A*<sup>157</sup>. Esta función depende de la unión de GATA4 a la región promotora de *SCN5A*, específicamente a TFBS específicos para GATA4 que se hayan a lo largo del promotor, con mayor relevancia en la región del intrón 1<sup>157</sup>.

La función del FT GATA4 está modulada por diversas modificaciones postraduccionales, de las cuales, la acetilación de lisinas es una de las más estudiadas. La forma acetilada de GATA4 tiene más afinidad de unión al DNA y una mayor actividad transcripcional que la forma sin acetilar. Diversos estudios han demostrado que el balance de acetilación/desacetilación de GATA4 está modulado por la actividad acetiltransferasa de p300 y la actividad deacetilasa de HDAC2<sup>166,167</sup>. En la misma línea, estudios preliminares del laboratorio sugieren que la acetilación de GATA4 está asociada con un aumento en la actividad del promotor *SCN5A*, mientras que la desacetilación de GATA4 disminuye su actividad<sup>18</sup>.

El primer estudio cuyo objetivo era dilucidar los lugares de acetilación de GATA4 en el dominio de unión al DNA (región básica adyacente a C-Zn) fue publicado en 2008 por Takaya y colaboradores. Su aproximación experimental se basaba en ensayos luciferasa para analizar la actividad transcripcional de mutantes de GATA4. A partir de estos estudios, propusieron las lisinas K311, K318, K320 y K322 de GATA4 como residuos diana de p300<sup>163</sup>. A pesar de que este estudio ha sido referenciado en múltiples ocasiones en la literatura, no establece la posición exacta de las lisinas de GATA4 acetiladas. Esto es debido a 2 motivos principales, discutidos más detalladamente en el apartado 1 de la discusión: (i) algunas de las lisinas mutadas fueron analizadas simultáneamente, por lo que no se pudo evaluar correctamente cada mutante de manera individual; (ii) se basan solamente en los resultados del ensayo luciferasa para determinar las posiciones acetiladas. No se puede afirmar categóricamente que la pérdida de actividad transcripcional demuestre que la lisina mutada es acetilada por p300, ya que puede haber otros factores involucrados en ese efecto (apartado 1 de la Discusión).

Estudios recientes de nuestro grupo de espectrometría de masas por MALDI-TOF sugieren que, únicamente, tres de las lisinas presentes en esta región se acetilan<sup>18</sup>. Estos estudios, llevados a cabo por la Dra. Anna Tarradas, consistían en ensayos de acetilación y desacetilación *in vitro* de 3 péptidos sintéticos de la región básica del C-Zn, con el posterior análisis de cada péptido por espectrometría de masas (**Figura 60**). Estos análisis indican que sólo uno de los

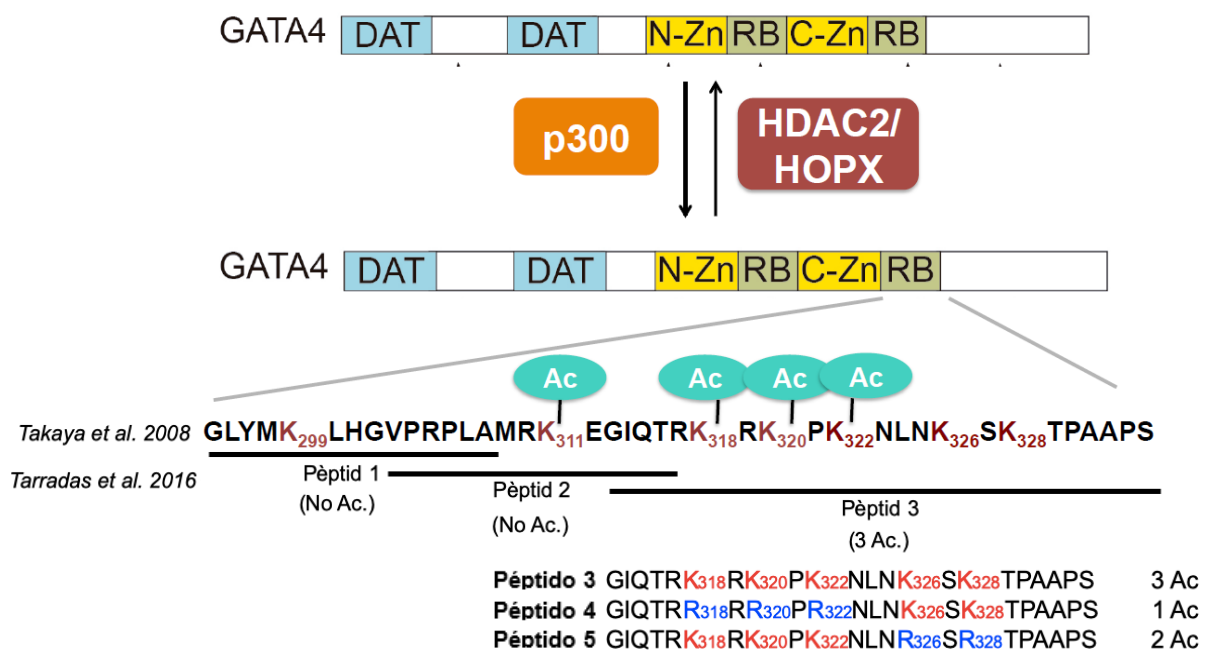


péptidos se acetila, el péptido 3, descartando K299 y K311 como lugares de acetilación y contradiciendo el estudio de Takaya y colaboradores.

A partir del péptido 3, se crearon otros 2 péptidos modificados, también estudiados mediante MALDI-TOF (**Figura 60**):

- Péptido 4: con K318, K320 y K322 mutadas a argininas (K326 y K328 intactas).
- Péptido 5: con K326 y K328 mutadas a argininas (K318, K320 y K322 intactas).

Como resultado, se determinó que 1 de las posiciones del péptido 4 se acetila (K326 o K328) y 2 de las posiciones en el péptido 5 se acetilan (K318, K320 o K322). Este último resultado también contradice el estudio de Takaya y colaboradores. Que afirmaba que las 3 lisinas eran acetiladas. Además, se realizó espectrometría de masas en tándem acoplada a un sistema de cromatografía en fase líquida (LC-MS/MS) para el péptido 3. Esta técnica permite detectar la posición exacta de las modificaciones postraduccionales presentes en un péptido. El experimento no tuvo la suficiente resolución para discernir las 2 posiciones acetiladas del grupo K318-K320-K322, sin embargo, se pudo determinar que K326 sería uno de los lugares de acetilación de GATA4. Posteriormente, experimentos de WB confirmaron que K326 es un sitio de acetilación de GATA4.



**Figura 60. Esquema de los péptidos estudiados mediante espectrometría de masas por la Dra. Anna Tarradas.** C-Zn: Dedo de zinc C-terminal; DAT: Dominio de activación transcripcional; DCT: dominio C-terminal; N-Zn: Dedo de zinc N-terminal; Rb: Región básica; RLN: Región de localización nuclear. Figura adaptada y cedida por Anna Fàbrega.

A partir de todos estos datos, en esta tesis se pretende esclarecer cuáles son las lisinas que se acetilan en GATA4, así como el efecto de la acetilación sobre la actividad del promotor de *SCN5A*.

## 1.2. Obtención de mutantes GATA4 mediante sustitución de lisinas potencialmente acetilables

Para identificar cuáles son las lisinas de GATA4 que se acetilan en la región básica adyacente a C-Zn, se diseñaron mutantes de GATA4 en los que se sustituyeron los residuos de lisinas por alaninas mediante mutagénesis dirigida (**Figura 61**). Debido a que los residuos de alaninas no se pueden acetilar, este cambio permite deducir los residuos implicados en la acetilación de GATA4. Se crearon:

- 5 mutantes simples, los cuales sólo presentaban una lisina mutada.
- 1 mutante doble, teniendo en cuenta que la Dra. Tarradas observó que K326 es un lugar de acetilación de GATA4 mediante LC-MS/MS y WB.
- 1 mutante triple, en el que se utilizó el doble mutante como molde para mutar la posición K318.

Esta batería de mutantes se utilizó para analizar los niveles de acetilación de GATA4 mediante inmunoprecipitación y WB, a partir de la transfección de los vectores de expresión de GATA4 en células HEK293T; y también para evaluar la actividad transcripcional de GATA4 en ensayos reportero luciferasa, a partir de la transfección de células H9c2.

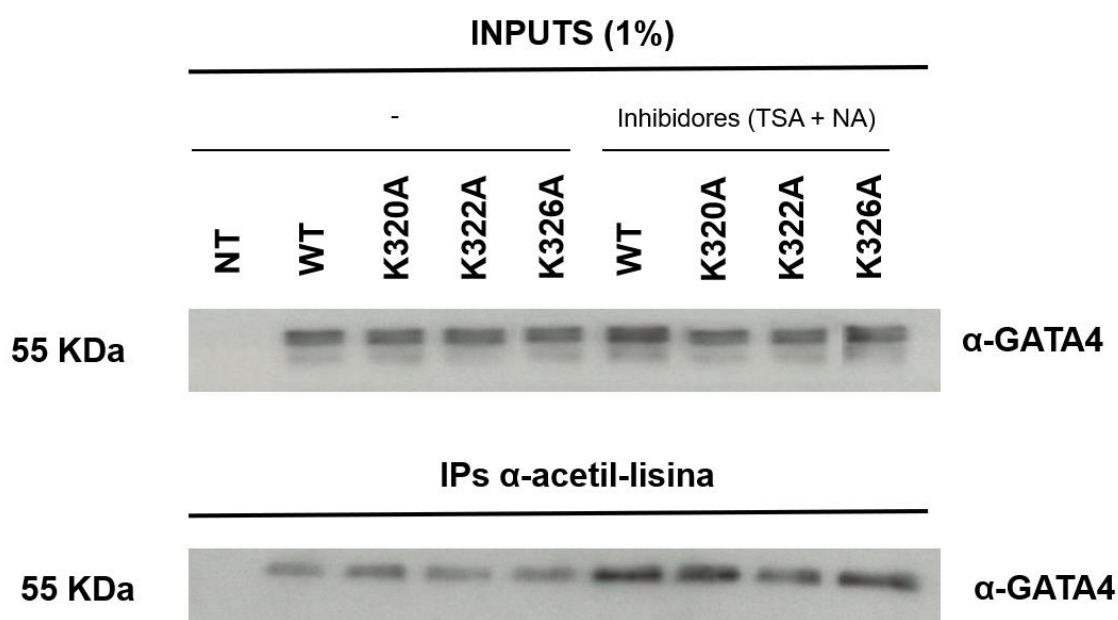
1. GATA4/HA <b>K318</b>	} Mutantes simples	TR <b>A</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>K</b> <sub>322</sub> NLN <b>K</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA
2. GATA4/HA <b>K320</b>		TR <b>K</b> <sub>318</sub> R <b>A</b> <sub>320</sub> P <b>K</b> <sub>322</sub> NLN <b>K</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA
3. GATA4/HA <b>K322</b>		TR <b>K</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>A</b> <sub>322</sub> NLN <b>K</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA
4. GATA4/HA <b>K326</b>		TR <b>K</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>K</b> <sub>322</sub> NLN <b>A</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA
5. GATA4/HA <b>K328</b>		TR <b>K</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>K</b> <sub>322</sub> NLN <b>K</b> <sub>326</sub> S <b>A</b> <sub>328</sub> TPA
6. GATA4/HA Doble mutante ( <b>K322 + K326</b> )		TR <b>K</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>A</b> <sub>322</sub> NLN <b>A</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA
7. GATA4/HA Triple mutante ( <b>K318 + K322 + K326</b> )		TR <b>A</b> <sub>318</sub> R <b>K</b> <sub>320</sub> P <b>A</b> <sub>322</sub> NLN <b>A</b> <sub>326</sub> S <b>K</b> <sub>328</sub> TPA

**Figura 61. Sustituciones de lisinas a alaninas en el dominio de unión a DNA de GATA4.** En azul los residuos mutados a alaninas, y en rojo las lisinas de la secuencia no modificadas.

### 1.3. Análisis de los niveles de acetilación de GATA4 con mutantes simples

La acetilación de GATA4 es un cambio postraduccional reversible, sujeto al equilibrio entre la acetiltransferasa p300 y HDAC2. La actividad deacetilasa puede influir en los niveles de acetilación de GATA4, reduciendo su número, y dificultando su evaluación. Los inhibidores de las desacetilasas como la tricostatina A y la nicotinamida (TSA y NA, respectivamente) aumentan los niveles de acetilación de GATA4, siendo el TSA un inhibidor de las HDACs tipo I, II y IV y la NA un inhibidor de las HDACs tipo III. Para maximizar los niveles de acetilación de GATA4 y evitar la influencia de la actividad de las HDACs se incluyó TSA y NA en los experimentos de acetilación en HEK293T.

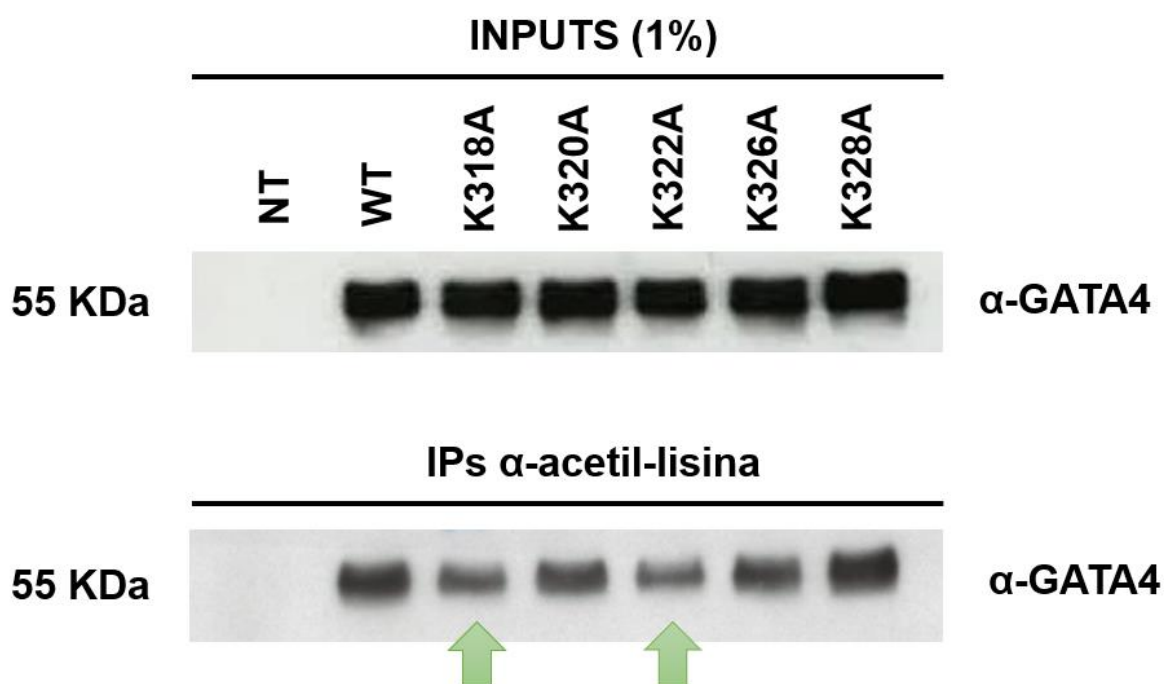
Las células HEK293T fueron transfectadas con el vector GATA4 WT y los mutantes simples K320A, K322A y K326A como prueba para testar los inhibidores de HDACs. Veinticuatro horas después de la transfección se trataron la mitad de las células con TSA y NA para cada condición. Al cabo de 48 horas de la transfección, los lisados celulares se inmunoprecipitaron con el anticuerpo  $\alpha$ -lisinas acetiladas y se analizaron por WB con  $\alpha$ -GATA4 (**Figura 62**). Tal y como se esperaba resultados obtenidos mostraron, por un lado, que GATA4 inmunoprecipitada presenta mayores niveles de acetilación cuando está tratada con TSA y NA. Habiendo comprobado que la adición de TSA y NA permite maximizar los niveles de acetilación de GATA4, todos los siguientes experimentos por WB se realizaron en presencia de TSA y NA. Por otro lado, este



**Figura 62. Niveles de acetilación de GATA4 en presencia o ausencia de inhibidores de HDACs (TSA y NA).** GATA4 WT, K320A, K322A y K326A fueron expresados en células HEK293 en presencia o ausencia de TSA y NA, inmunoprecipitados con  $\alpha$ -lisinas acetiladas y analizados por WB con anticuerpos  $\alpha$ -GATA4; el input corresponde a un 1% del volumen utilizado por la IP.

experimento también reveló una leve disminución en los niveles de acetilación del mutante K322A respecto GATA WT.

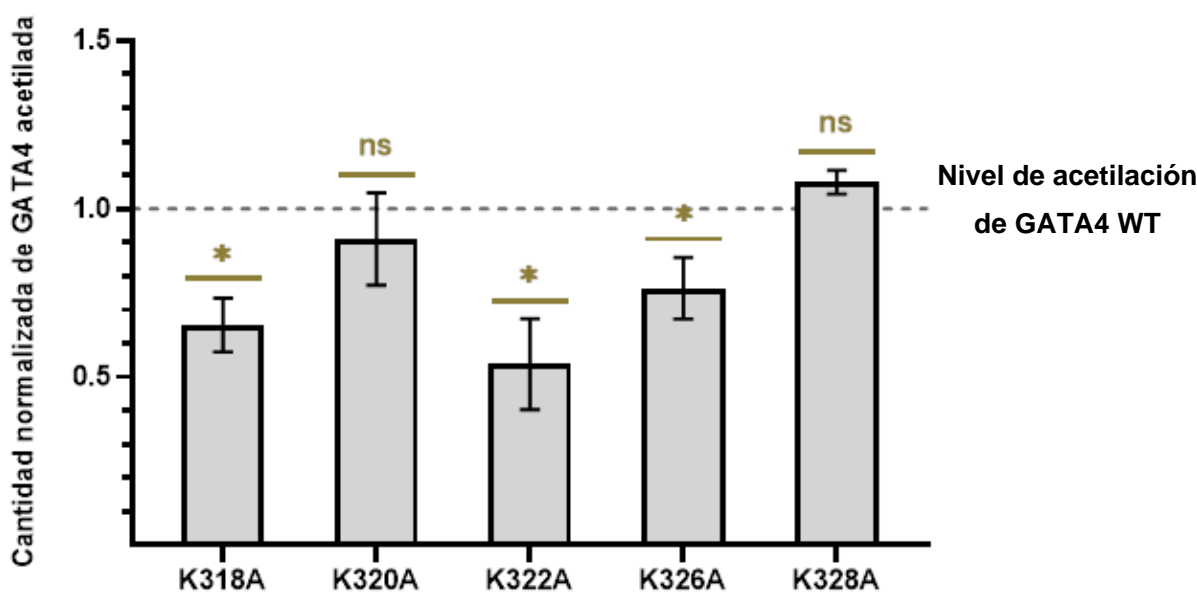
Para identificar cuáles de las 5 lisinas mutadas (K318, K320, K322, K326 y K326) son acetiladas, se realizó un experimento en el que células HEK293T se transfectaron con el vector GATA4 WT y los 5 mutantes simples, en presencia de TSA y NA (**Figura 63**). Según la intensidad de la banda se deduce que los residuos con menor grosor de banda son los que tienen un mayor efecto en los niveles de acetilación de GATA4. Teniendo en cuenta nuestros datos de MALDI-TOF, que indican que en esta región únicamente se acetilan 3 residuos<sup>18</sup>, los resultados de IP-WB sugieren que K318 y K322 serían las lisinas que más contribuyen a la acetilación de GATA4, ya que la mutación de estas lisinas produce una reducción significativa de los niveles de acetilación (**Figura 63**). Los estudios previos realizados sugieren que el tercer residuo acetilado en esta región es la lisina K326. No obstante, en este experimento con mutantes simples, no se observa una disminución clara de los niveles de acetilación con el mutante K326A.



**Figura 63. Niveles de acetilación de GATA4 en mutantes simples de GATA4.** WB α-GATA4 de los lisados de HEK293T transfectados con mutantes simples de GATA4, tratados previamente con inhibidores de HDAC (TSA y NA), e inmunoprecipitados (IP) para lisinas acetiladas; el input corresponde a un 1% del volumen utilizado por la IP; la condición NT corresponde al control sin transfectar.

El experimento descrito en la **Figura 61** se repitió un total de 3 veces para validar el resultado (Anexo 3; **Figura A-1**). De todos los experimentos, el presentado en la **Figura 61** fue el WB más representativo. No obstante, se utilizaron las otras dos réplicas (en total, 3 WB independientes)

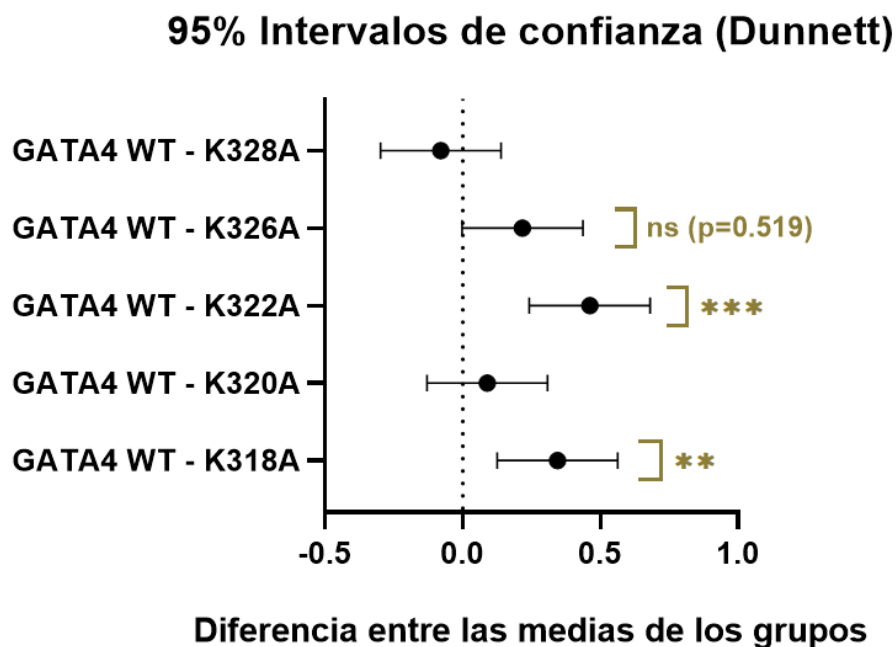
para analizar la intensidad de las bandas con el software Image J. Este análisis permitió obtener un resultado numérico que reflejara las variaciones en los niveles de acetilación. El WB del input representa el control de carga de los experimentos ya que revela la cantidad total de GATA4 acetilada o sin acetilar de la que se parte en cada muestra antes del WB con  $\alpha$ -acetil-lisina. Las variaciones del input entre muestras deben ser las mínimas posibles para que los resultados sean fiables, tal y como se aprecia en la **Figura 63**. Estas mediciones del input con Image J se utilizaron para normalizar las mediciones del WB de las Ips con  $\alpha$ -acetil-lisina, otorgando de esta manera una mayor fiabilidad a los datos.



**Figura 64. Cuantificación de los niveles de acetilación de los mutantes de GATA4.** Las mediciones se realizaron con el software Image J y los niveles de acetilación se normalizaron con los inputs de cada WB (n=3) y con GATA4 WT. La significación estadística de cada mutante se examinó mediante la prueba t de Student en relación con el control WT (\* $p \leq 0.05$ )

El análisis de la cuantificación de las bandas de los 3 WB, presentados en la **Figura 64**, refuerzan los resultados de la **Figura 63**, siendo los mutantes K318A y K322A los que presentan una mayor disminución en los niveles de acetilación. Sin embargo, con estos datos también se observa una disminución significativa de los niveles de acetilación en el mutante K326A cuando se realiza la prueba t de Student. Para dar una mayor robustez estadística al análisis numérico se hizo también un ANOVA seguido de una prueba de comparaciones múltiples de Dunnett (**Figura 65**). El método de Dunnett se utiliza en ANOVA para crear intervalos de confianza para las diferencias entre la media de cada nivel de factor y la media de un grupo de control, en nuestro caso GATA4 WT<sup>375</sup>. El test de ANOVA confirmó que hay diferencias entre nuestros grupos (GATA4 WT y cada uno de los mutantes), y el método de Dunnett determinó los niveles de confianza para cada comparación individual (**Figura 65**). Los mutantes K318A y K322A son

los que presentan las diferencias más importantes. Sin embargo, según este análisis el mutante K326A no reduce los niveles de acetilación de forma estadísticamente significativa, aunque presenta una  $p=0.519$  muy cercana a la significancia.

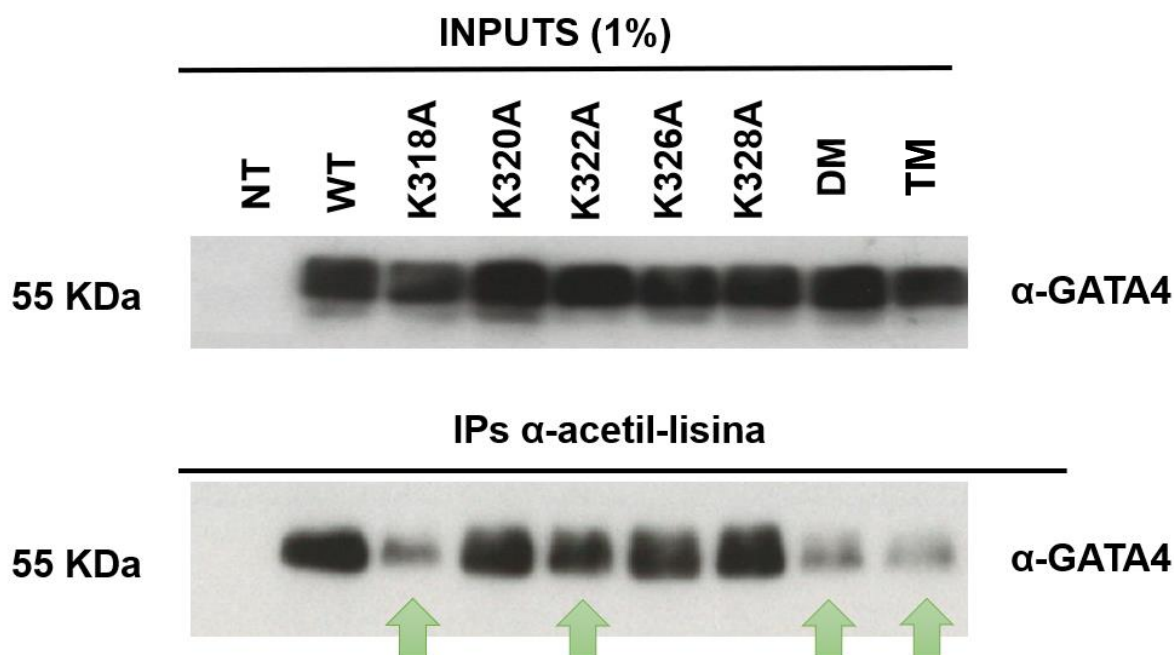


**Figura 65. Análisis de los niveles de acetilación de mutantes de GATA4 mediante comprobaciones múltiples de Dunnett.** Intervalos de confianza para cada mutante contra GATA4 WT. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , and \*\*\* $p \leq 0.001$ .

#### 1.4. Análisis de los niveles de acetilación de GATA4 con mutantes dobles y triples

Para confirmar los resultados con mutantes de GATA4 simples, se estudiaron los niveles de acetilación de GATA4 introduciendo mutantes de lisinas dobles y triples. Para ello, se realizaron experimentos similares a los descritos en la sección 1.3 introduciendo también el mutante doble K322A+K326A y el mutante triple K318A+K322A+K326A (ver **Figura 61**). De forma similar a lo observado en los experimentos anteriores, los resultados obtenidos indican que los mutantes K318A y K322A, en especial K318, muestran unos niveles de acetilación menores que GATA4 WT (**Figura 66**). Sin embargo, la sustitución de dos lisinas (K322 y K326) en el doble mutante da lugar a una reducción de los niveles de acetilación sustancialmente mayor que en cualquiera de las versiones simples de K322A y K326A. Este resultado sugiere que K326A está contribuyendo a la disminución de los niveles de acetilación, conjuntamente con K322A, y apoya los resultados anteriores que sugerían que K326 podría ser un lugar de acetilación. Por otro lado, el triple mutante (K318A+K322A+K326A), revela una disminución aún mayor de los niveles de

acetilación de GATA4, lo que indicaría un efecto acumulativo en la eliminación de los 3 lugares de acetilación.

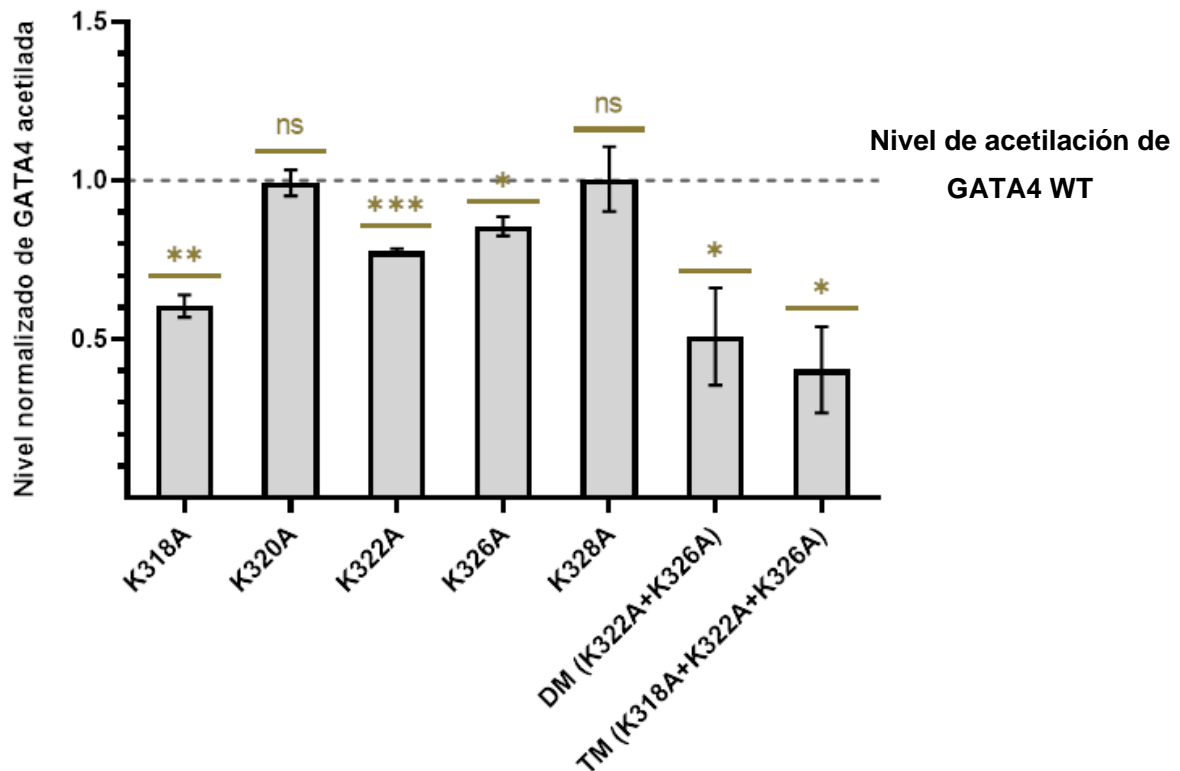


**Figura 66. Niveles de acetilación de GATA4 en mutantes simples, dobles y triples de GATA4.** WB α-GATA4 de los lisados de HEK293T transfectados con mutantes simples, doble (DM; K322A+K326A) y triple (TM; K318A+K322A+K326A) de GATA4, tratados previamente con inhibidores de HDAC (TSA, y NA), e inmunoprecipitados (IP) para lisinas acetiladas; el input corresponde a un 1% del volumen utilizado por la IP.

Debido a que el WB de la **Figura 66** sólo refleja una variación evidente de los niveles de acetilación para los mutantes K318A, K322A, DM (K322A+K326A) y TM (K318A+K322A+K326) (marcados con flechas verdes) y que la forma de la banda de GATA4 WT dificulta apreciar las diferencias con algunos mutantes, se realizó un análisis cuantitativo con Image J para poder detectar variaciones en los niveles de acetilación inapreciables visualmente, nuevamente utilizando datos de 3 réplicas independientes (Anexo 3; **Figura A-2**).

Los resultados de la cuantificación confirman las observaciones descritas hasta el momento (**Figura 67**). Por un lado, se observa una disminución significativa de los niveles de acetilación en los mutantes simples K318A, K320A y K326A, este último siempre en menor medida. Se evidencia de forma clara que el DM y el TM dan lugar a cambios bastante más drásticos que cualquier mutante simple. También se observa la misma tendencia al realizar la prueba de comparaciones múltiples de Dunnett (**Figura 68**). Una vez más, el mutante K326A no muestra una disminución significativa en la acetilación de GATA4 cuando se utiliza este test estadístico. No obstante, hay que tener en cuenta el conjunto de todos los experimentos y pruebas realizadas antes y durante esta tesis, incluyendo: (i) experimentos de MALDI-TOF, que indicaban una

acetilación de GATA4 en K326 o K328; (ii) los resultados de LC-MS/MS, que indicaban acetilación de K326; (iii) experimentos de WB de la Dra. Tarradas, que mostraba una menor acetilación con el mutante K326A, y (iv) los resultados presentados en esta tesis, en los cuales K326A resulta en una menor acetilación en algunos experimentos, quedando cerca de la significación estadística en el resto. Teniendo en cuenta todas las evidencias presentadas, proponemos a K326 como el tercer residuo acetilado en GATA4, junto con K318 y K320.



**Figura 67. Cuantificación de los niveles de acetilación de los mutantes simples, DM y TM de GATA4.** Las mediciones se realizaron con el software Image J y los niveles de acetilación se normalizaron con los inputs de cada WB (n=3). La significación estadística de cada mutante se examinó mediante la prueba t de Student en relación con el control WT (\* $p \leq 0.05$ , \*\* $p \leq 0.01$ , and \*\*\* $p \leq 0.001$ ).



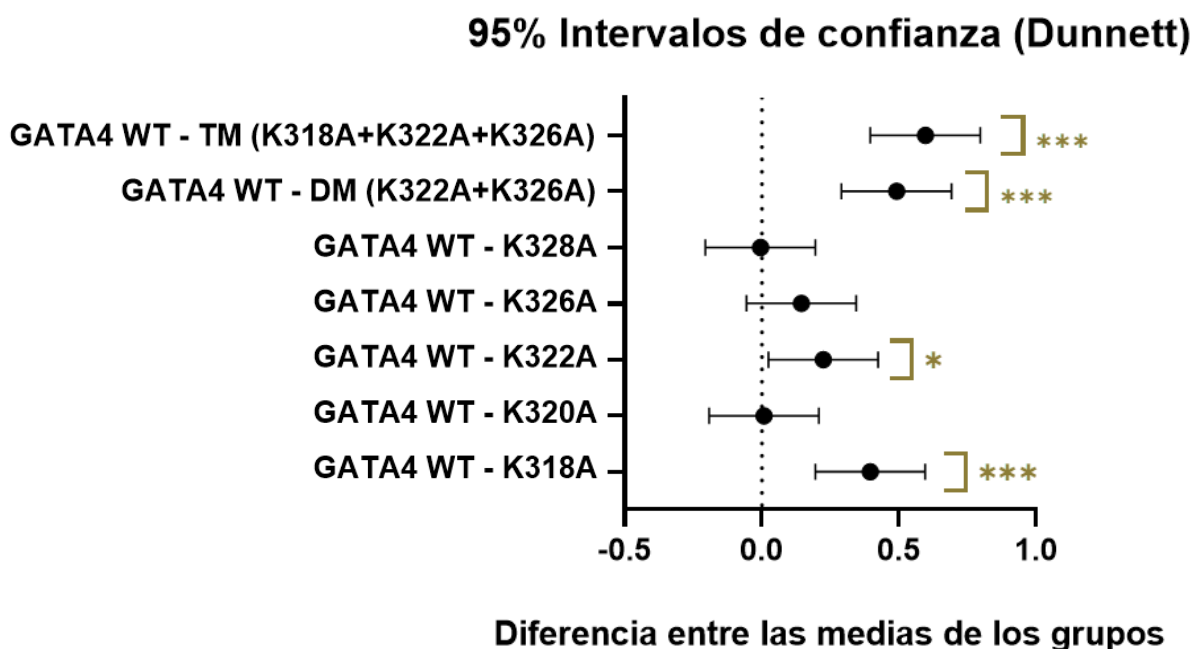


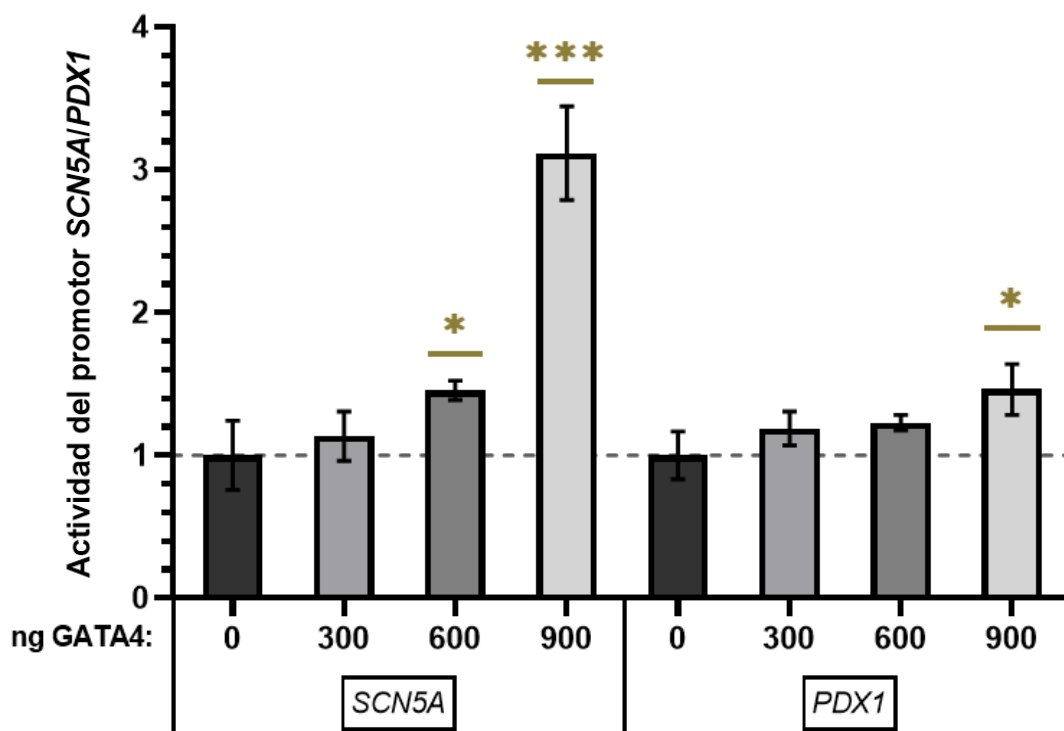
Figura 68. Análisis de los niveles de acetilación de los mutantes simples, DM y TM de GATA4 mediante comprobaciones múltiples de Dunnett. Intervalos de confianza para cada mutante contra GATA4 WT. \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , and \*\*\* $p \leq 0.001$ .

### 1.5. Efecto de la acetilación de GATA4 sobre la actividad del promotor de *SCN5A*

Los estudios iniciales sobre la acetilación de GATA4 sugerían que esta modificación está asociada a un incremento de la afinidad de unión de GATA4 al DNA y un aumento de su actividad transcripcional *in vitro*<sup>161,163</sup>. Con el fin de evaluar el efecto de la acetilación de GATA4 sobre la actividad del promotor del gen *SCN5A*, se realizaron ensayos reporteros luciferasa de los diferentes mutantes de GATA4 en el modelo celular H9c2.

En primer lugar, se realizaron ensayos para determinar la cantidad óptima de plásmido de expresión de GATA4 para poder evaluar cambios en su actividad transcripcional sobre el promotor de *SCN5A*. En estos experimentos, se utilizó la versión del promotor (-260 a +612 pb en relación al TSS de *SCN5A*) descrito en el apartado 1.3.1. de Materiales. Los plásmidos con el promotor *SCN5A*-luciferasa y con el promotor *EF1 $\alpha$* -renilla, se co-transfectaron conjuntamente con el plásmido de expresión GATA4 WT (300, 600 o 900 ng) en células H9c2. En paralelo, se realizó el mismo experimento utilizando como reportero un plásmido con el promotor pancreático *PDX1* para comparar la actividad de GATA4 en diferentes promotores<sup>157</sup>. En ambos experimentos, se observó que la transfección de 900 ng de plásmido de expresión de GATA4 WT aumentó significativamente la actividad de los promotores. Concretamente, la actividad de

*SCN5A* resultó 3 veces superior al control (0 ng de GATA4 WT) y 1.5 veces superior en el caso del promotor *PDX1* (Figura 69).

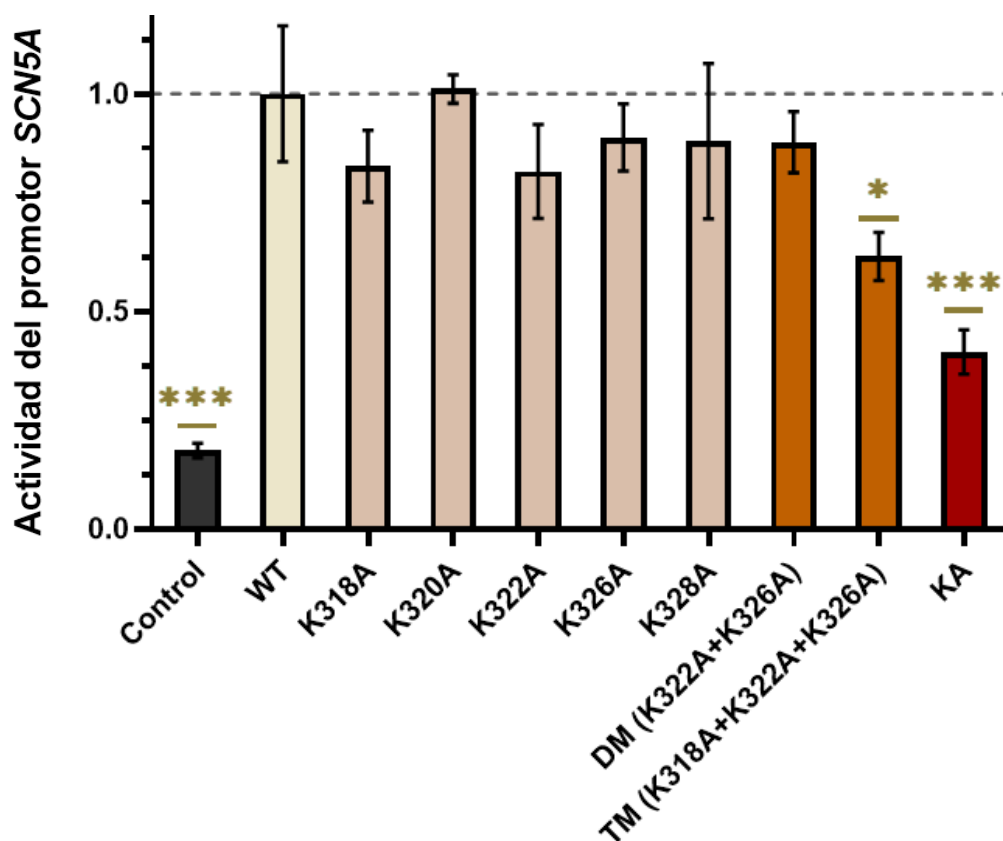


**Figura 69. Efecto del incremento de GATA4 sobre la actividad de los promotores *SCN5A* y *PDX1*.** Ensayos luciferasa en células H9c2 transfectadas con el promotor *SCN5A* (100 ng) o el promotor *PDX1* (100 ng), junto con el promotor *EF1 $\alpha$ -RL* (10 ng) y el vector GATA4 WT (300, 600, o 900 ng). Los valores de luciferasa representados están normalizados por los valores de renilla y se muestran relativos al control (condición en ausencia de GATA4) (n=3). Para el análisis estadístico se realizó una prueba t-Student relativa al control. \*p $\leq$ 0,05, \*\*p $\leq$ 0,01 y \*\*\*p $\leq$ 0,001.

El uso de 900 ng de GATA4 WT resultó en el mayor incremento de la actividad en ambos promotores, siendo el resultado más óptimo para evaluar los cambios en la actividad promotora en presencia de mutantes de GATA4. Por lo tanto, se realizó un ensayo reportero luciferasa para evaluar el efecto de las sustituciones de lisinas a alaninas en la actividad transcripcional de GATA4 sobre el promotor *SCN5A*. En este ensayo se añadió como control el vector de expresión GATA4-KA (KA), el cual tiene cuatro lisinas mutadas a alaninas (K311, K318, K320 y K322). Estudios previos han reportado que el mutante GATA4-KA no es acetilado por p300, no se une al DNA y pierde su función de activador transcripcional sobre los promotores *ANF* y *ET-1*<sup>166</sup>.

El resultado del ensayo mostró que la actividad transcripcional de GATA4 sobre el promotor de *SCN5A* no se ve afectada de forma significativa cuando se sustituyen las distintas lisinas de forma individual ni tampoco en el DM K322A+K326A (Figura 70). Aun así, la tendencia de los mutantes K318A y K322A es una reducción de la capacidad de GATA4 de activar la transcripción

del promotor *SCN5A*, con una disminución en la actividad del 16 y 18%, respectivamente. Es importante resaltar que, aunque en este ensayo la disminución de la actividad transcripcional de los mutantes K318A y K322A respecto el WT no resulta significativa, estos dos mutantes son los que previamente habían mostrado una reducción más significativa de niveles de acetilación (Figura 64-67).



**Figura 70. Efecto de los mutantes de GATA4 sobre la actividad del promotor *SCN5A*.** Ensayos luciferasa de células H9c2 transfectadas con el promotor de *SCN5A*, el promotor constitutivo *EF1 $\alpha$*  y los vectores de expresión GATA4 WT, K318A, K320A, K322A, K326A, DM, TM y KA (900 ng). Los valores luciferasa están normalizados por los valores *renilla* y se muestran relativos a GATA4 WT (n=3). El control corresponde a la actividad basal del promotor *SCN5A* (en ausencia del FT GATA4). Para el análisis estadístico se realizó una prueba t-Student relativa a GATA4 WT. \*p $\leq$ 0,05, \*\*p $\leq$ 0,01 y \*\*\*p $\leq$ 0,001.

Los resultados de este ensayo también indican que la actividad transcripcional del TM K318A+K322A+K326A es un 40% inferior que GATA4 WT. Esto indica que GATA4 requiere de las lisinas K318, K322 y K326 intactas para realizar su función activadora. Dado que estos residuos concuerdan con que K318A es el mutante que muestra una mayor reducción de los niveles de acetilación (y de la actividad transcripcional, juntamente con K322A), esto sugiere que la acetilación de K318 es clave para la actividad transcripcional de GATA4. También concuerda con el hecho de que el DM K322A+K326A no presenta una disminución significativa en la

actividad del promotor, siendo el resultado del TM un posible efecto acumulativo de las 3 mutaciones. Los resultados concuerdan con los análisis de acetilación por WB en los que se observa que la sustitución simultánea de estos tres residuos da lugar a una reducción de los niveles de acetilación de GATA4, y también lo hace en mayor medida que el resto de mutantes. No obstante, estos resultados no coinciden con los ensayos luciferasa realizados por Takaya y colaboradores, donde obtienen disminuciones significativas en la actividad transcripcional en los mutantes para K311A, K318A+K320A, y K322A. Sin embargo, hay que tener en cuenta que el doble mutante K318A+K320A solamente lo evalúan conjuntamente y no realizan un análisis con mutantes simples para K318 y K320. Otra consideración a tener en cuenta es que el estudio de Takaya analiza el efecto de los mutantes de GATA4 en los promotores *ANP ET-1* en un modelo de células de riñón de chimpancé (COS7), mientras que en esta tesis se ha estudiado el efecto sobre el promotor de *SCN5A* en un modelo de cardiomiocitos embrionicos de rata (H9c2).

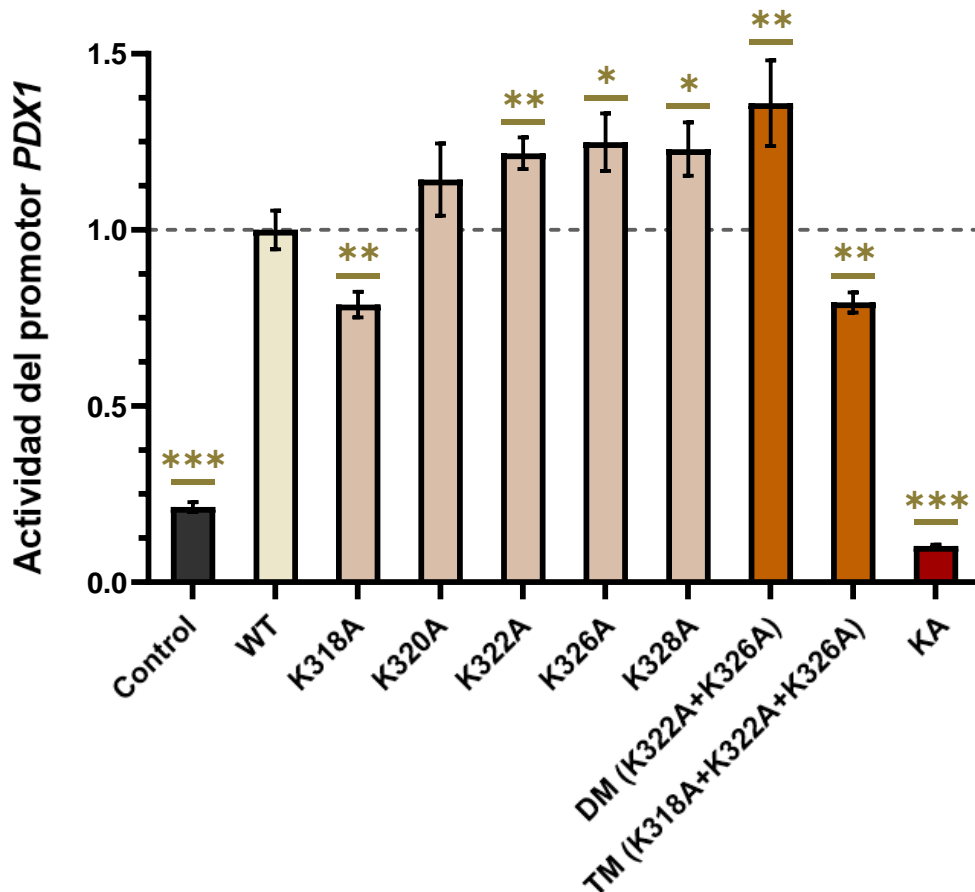
De acuerdo con los resultados previos de Takaya *et al*<sup>63</sup>. y Trivedi *et al*<sup>66</sup>., nuestros experimentos también demuestran que la actividad transcripcional de GATA4-KA sobre el promotor *SCN5A* es un 60% menor que GATA4 WT. Esta disminución es un poco mayor a la observada en el TM, con el que comparte 2 lisinas mutadas (K318 y K322). Los resultados de esta tesis sugieren que las otras dos lisinas de GATA4-KA (K311 y K320) no se acetilan, pero es posible que la disminución de la actividad transcripcional en KA sea independiente de la acetilación.

El efecto de las mutaciones de GATA4 en su activación transcripcional se evaluó también sobre el promotor *PDX1* (promotor del gen homeobox 1 pancreático y duodenal), el cual ya se ha descrito previamente por ser regulado transcripcionalmente por GATA4 y GATA6<sup>376</sup>.

En este ensayo se observó una mayor cantidad de mutantes con alteraciones en la actividad transcripcional de *PDX1* que en el ensayo sobre el promotor *SCN5A* (**Figura 71**). De forma similar a los ensayos con el promotor *SCN5A*, observamos una disminución significativa de la actividad transcripcional de los mutantes K318A, TM y GATA4-KA de un 19, 21 y 91% respectivamente. Este resultado sugiere que los residuos K318, K322 y K326 de GATA4 afectan directamente la actividad transcripcional de GATA4 independientemente del promotor diana. Además, los resultados de este ensayo denotan que todos los mutantes que presentaban la sustitución K318 (K318A y TM) muestran una actividad significativamente menor que GATA4 WT, sugiriendo que, de la misma manera que en la **Figura 70**, el residuo K318 es clave para la actividad transcripcional de GATA4.

Por otra parte, sorprendentemente, los resultados de este ensayo indican que los mutantes K322A, K326A, K328A, y DM presentan una actividad transcripcional un 20% más elevada, que GATA4 WT. Aunque este efecto estaría en desacuerdo con la premisa de que la acetilación de GATA4 aumenta su actividad transcripcional, se tiene que tener en cuenta que este resultado sólo se ha observado con el promotor *PDX1* (apartado 1 de la Discusión).

En relación al mutante KA, los resultados obtenidos revelan que se comporta de manera similar al control en ausencia de GATA4. Tanto para el promotor de *SCN5A* como para el de *PDX1*, el mutante KA ha resultado ser con diferencia el que menor actividad transcripcional presenta. Estos resultados dan más fuerza a la hipótesis de que la combinación de estos 4 residuos mutados pueda estar afectando de manera relevante la unión de GATA4 al DNA y, en general, su capacidad como activador transcripcional, más allá de la disminución en los niveles de acetilación derivados de mutar K318 o K322.



**Figura 71. Efecto de los mutantes de GATA4 sobre la actividad del promotor *PDX1*.** Ensayos luciferasa de células H9c2 transfectadas con el promotor de *PDX1*, el promotor constitutivo *EF1α* y los distintos vectores de expresión de GATA4 WT, K318A, K320A, K322A, K326A, DM, TM y KA (900 ng). Los valores luciferasa están normalizados por los valores *renilla* y se muestran relativos a GATA4 WT (n=3). El control corresponde a la actividad basal del promotor *PDX1* (en ausencia del FT GATA4). Para el análisis estadístico se realizó una prueba t-Student relativa a GATA4 WT. \*p≤0,05, \*\*p≤0,01 y \*\*\*p≤0,001.

## CAPÍTULO II. Cribado de regiones reguladoras de *SCN5A*

### 2.1. Selección de las regiones CRISPRi

En los últimos años, distintos laboratorios han podido identificar regiones promotoras o *enhancers* que regulan la expresión de genes asociados al SBr. Sin embargo, hasta la fecha no ha habido ningún trabajo que haya determinado, con una aproximación sistemática, la ubicación exacta de las regiones que regulan la expresión del gen *SCN5A*. En este contexto, uno de los objetivos de esta tesis ha sido desarrollar una estrategia para la identificación de regiones reguladoras de *SCN5A*. Para ello, se ha realizado un cribado de regiones reguladoras con la tecnología de CRISPRi. Este cribado requirió de una etapa previa, una preselección de elementos potencialmente reguladores de la transcripción de *SCN5A* realizada a partir de datos de ENCODE y otras bases de datos públicas.

Los análisis recientes de datos de Hi-C han demostrado que el genoma humano está organizado en TADs que están involucrados tanto en la co-regulación de genes como en el bloqueo de las interacciones entre regiones TAD adyacentes<sup>99</sup>. Por lo tanto, es más probable que los elementos reguladores que se encuentran en el mismo TAD de un gen en particular estén implicados en la regulación de este gen. Teniendo en cuenta estos factores, con el fin de establecer las regiones de estudio para los ensayos CRISPRi dentro del TAD de *SCN5A* se utilizaron datos del proyecto ENCODE<sup>109</sup> y de experimentos ChIP-seq en cardiomiocitos humanos<sup>378</sup> (**Figura 72**). La preselección de estas regiones potencialmente reguladoras se basó en los siguientes parámetros obtenidos de las bases de datos mencionadas (a partir de cardiomiocitos humanos):

- Perfiles de hipersensibilidad a la Dnasa I (DHS), los cuales marcan regiones de la cromatina transcripcionalmente activas. Las regiones DHS son regiones del genoma menos condensadas y en las que el ADN está más accesible. Estas regiones corresponden a regiones transcripcionalmente activas, ya que su menor compactación es necesaria para la unión de FT.
- Mapas de marcas de histonas H3K4me3 (que indica regiones promotoras transcripcionalmente activas) y H3K27ac (que indica promotores o *enhancers* transcripcionalmente activos).
- Presencia de factores de transcripción cardíacos como TBX5, GATA4 o NKX2-5, conocidos por su papel en la regulación de genes cardíacos, a partir de datos de ChIP-seq.
- Mapas de CTCF asociados a promotores y límites de los TADs. Las alteraciones en la unión de este FT a lugares específicos podrían afectar a la expresión génica, por ejemplo, alterando las interacciones entre promotores o *enhancers* con otros elementos reguladores.



A continuación, se analizaron los parámetros de unión a FT y marcas epigenéticas dentro de la región TAD del *SCN5A*. Las regiones que presentaron un mayor solapamiento de estos parámetros fueron consideradas regiones con más probabilidades de ser reguladoras de *SCN5A* (regiones 1-7, **Figura 72**). A partir de estas 7 regiones, se definieron las guías de estudio (gRNAs) que servirían para llevar a cabo el cribado con el sistema CRISPRi, permitiendo dirigir dCas9-KRAB a las regiones potencialmente reguladoras del gen *SCN5A*.

Durante el proceso de preselección, se identificó también la región intergénica entre *EXOG* y *SCN5A* (**Figura 72**). Esta región ha estado previamente estudiada por Man y colaboradores como un SE involucrado en la arquitectura de la cromatina del locus *SCN5A-SCN10A*<sup>349</sup>. Esto confirma que la estrategia de preselección de regiones reguladoras utilizada en esta tesis es adecuada. Sin embargo, debido a que la función reguladora de esta región ya había estado confirmada, se omitió de este estudio.

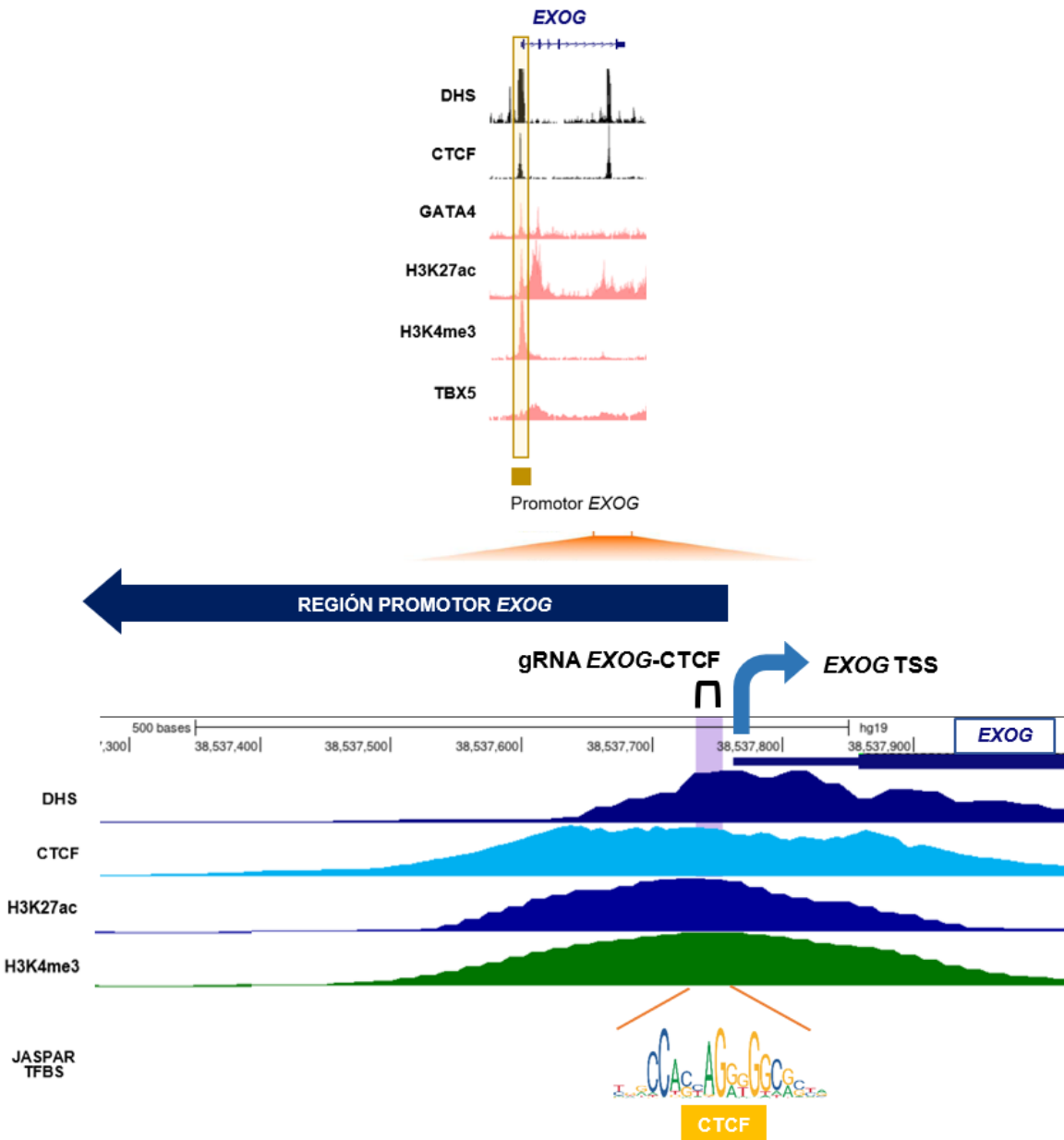
En total, mediante las herramientas descritas en el apartado 2.7.1. de Métodos, se definieron un total de 11 gRNAs comprendidas a lo largo de las 7 regiones seleccionadas en la **Figura 72**. Para la mayoría de casos, la localización exacta de cada guía (**Figura 73-77**) se definió en base a la presencia de lugares de unión (basados en predicciones o datos de CHIP-seq) para factores de *SCN5A* o en regiones donde se han identificado SNVs asociados al SBR<sup>332,346,351</sup>.

- **Región en el promotor de *EXOG* (región 1)**

La primera zona de estudio se localiza en la región promotora del gen *EXOG* situado inmediatamente *upstream* del gen *SCN5A*. Esta zona es relevante por el solapamiento de múltiples criterios establecidos para la selección de regiones reguladoras. En primer lugar, destaca una fuerte señal de H3K4me3, lo cual indica actividad promotora relacionada con el inicio de la transcripción. La proximidad con el TSS de *EXOG* sugiere que esta señal promovería la transcripción de este gen. No obstante, también encontramos señal H3K27ac, lo cual podría indicar la presencia de actividad *enhancer* además de la actividad promotora. Adicionalmente, es una región del DNA con alta accesibilidad a la Dnasa I tal como se muestra en los datos de DHS, sugiriendo la unión de FT como GATA4, CTCF y TBX5 en cardiomiocitos (datos de CHIP-seq).

Con el objetivo de diseñar una gRNA en un motivo de unión a FT en esta región se utilizó una base de datos del *Genome Browser* de la UCSC (colección CORE de la base de datos JASPAR) que representa una predicción de los motivos de unión a FT para todo el genoma. La gRNA utilizada para los ensayos CRISPRi (***EXOG-CTCF***) se diseñó para que solapase con el lugar de unión predicho para CTCF, según la información aportada por la base de datos JASPAR (**Figura 73**). Se seleccionó este lugar en específico para dirigir la gRNA ya que era la secuencia con mayor probabilidad de unión para CTCF según JASPAR.



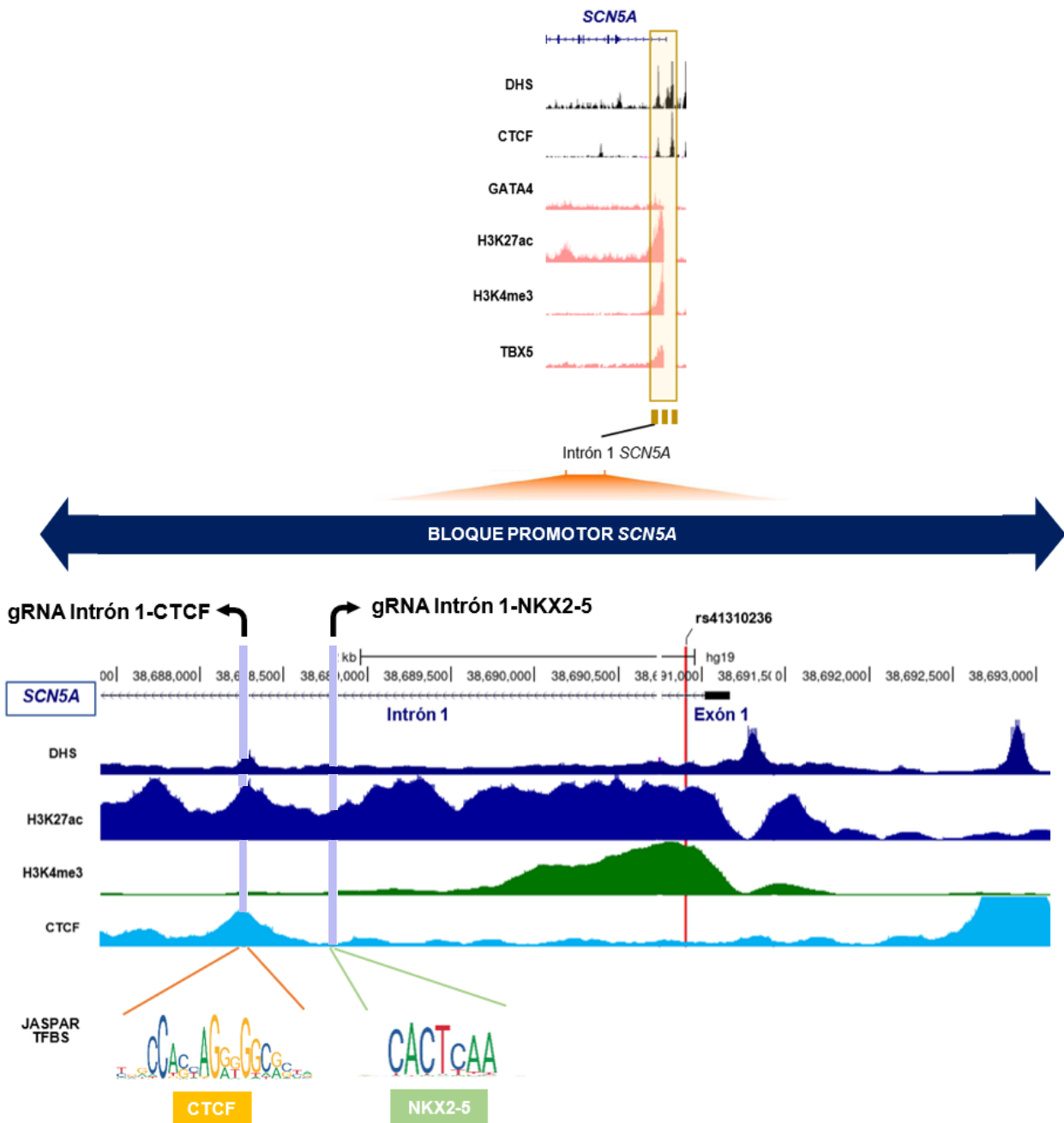


**Figura 73. Región promotora de *EXOG* (región 1).** Arriba, región preseleccionada con los picos correspondientes a DHS, CTCF, GATA4, H3K27ac, H3K4me3 y TBX5. Abajo, región ampliada del promotor de *EXOG* con la región diana de la gRNA (en violeta) y las marcas más relevantes para esa zona. En JASPAR TFBS se muestra el motivo de unión a CTCF predicho.

- **Región en el intrón 1 de *SCN5A* (región 2)**

La segunda zona de estudio se sitúa en el intrón 1 de *SCN5A*. En esta región predomina la marca de activación transcripcional H3K27ac, además de picos puntuales de DHS y CTCF. Las dos gRNAs seleccionadas se solapan con los motivos de unión a los FT cardíacos CTCF y NKX2-5, respectivamente (**Figura 74**). La región también solapa con una señal de ChIP-seq para el FT TBX5. TBX5 actúa sinérgicamente con NKX2-5 en la activación de numerosos promotores de genes cardíacos, y está involucrado en el correcto funcionamiento del sistema de

conducción cardíaco<sup>380</sup>. Ambas gRNAs (**Intrón 1-CTCF** e **Intrón 1-NKX2-5**) están dirigidas a una zona con actividad *enhancer* activa debido a la presencia de una fuerte señal de H3K27ac. Además, esta zona también forma parte de la región promotora de *SCN5A* (bloque promotor), que se define como un segmento de DNA de 2.8 kb que se extiende hasta el intrón 1 de *SCN5A* e incluye 2.1 kb de la secuencia 5' *upstream* del exón 1<sup>334,351,371</sup>.

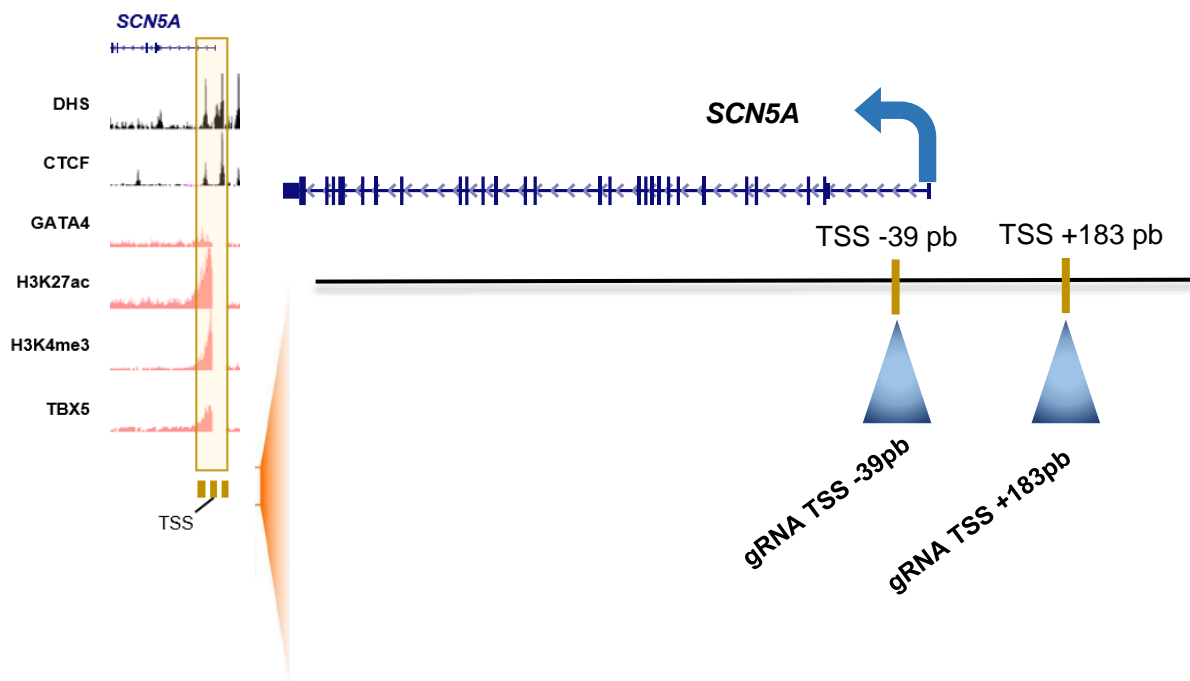


**Figura 74. Región del intrón 1 de *SCN5A* (región 2).** Arriba, región preseleccionada con los picos correspondientes a DHS, CTCF, GATA4, H3K27ac, H3K4me3 y TBX5. Abajo, ampliación de la región del intrón 1 de *SCN5A* con las regiones dianas de las gRNAs (en violeta) y las marcas más relevantes para esa zona. Se muestran también los motivos de unión predichos para para CTCF y NKX2-5 con el programa JASPAR TFBS Motivos de unión o lugares de unión.

Es interesante destacar que el intrón 1 comprende la variante rs41310236. La combinación de esta variante junto con otras 5 variantes en LD, forman parte de un haplotipo descrito en el bloque promotor de *SCN5A*. Este haplotipo está asociado con variaciones en la actividad transcripcional de *SCN5A* y el ECG<sup>332</sup>.

- **Región en el TSS de *SCN5A* (región 3)**

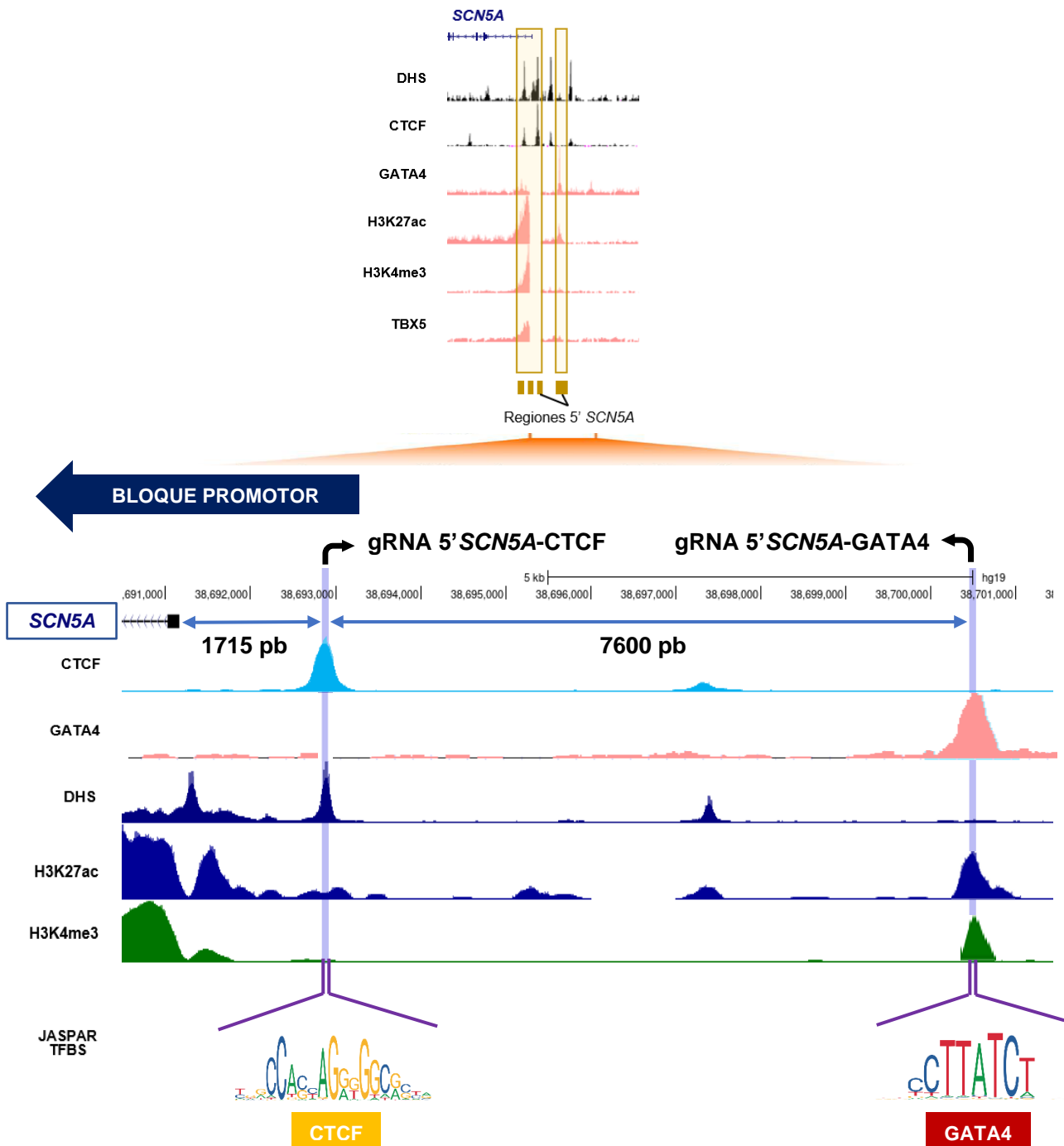
La tercera zona de estudio se definió alrededor del TSS de *SCN5A* y fue elegida como control positivo del ensayo CRISPRi. Estudios previos que han utilizado la misma metodología CRISPRi han observado que se obtiene una fuerte actividad CRISPRi dirigiendo dCas9-KRAB en una ventana entre -50 a +300 pb en relación con el TSS de un gen<sup>220,319</sup>. Estos resultados sugieren que este sería el rango óptimo para dirigir la gRNA y obtener una disminución significativa en la expresión del gen asociado al TSS en cuestión. Siguiendo estas premisas, se diseñaron dos gRNAs (**TSS-39pb** y **TSS+183pb**) alrededor del TSS de *SCN5A*: una 39 pb *downstream* del TSS y la otra 183 pb *upstream* del TSS (**Figura 75**).



**Figura 75. Región del TSS de *SCN5A* (región 3).** Izquierda, región preseleccionada con los picos correspondientes a DHS, CTCF, GATA4, H3K27ac, H3K4me3 y TBX5. Derecha, ampliación de la región del TSS de *SCN5A*, con las regiones diana (indicadas por los triángulos azules). Para las gRNAs diseñadas en el TSS de *SCN5A* no se utilizaron datos de marcas de histonas o FT, solamente se tuvo en cuenta la proximidad con el TSS.

- **Regiones en 5' *SCN5A* (regiones 4 y 5)**

Las dos gRNAs diseñadas en la zona 5' de *SCN5A* están localizadas en las regiones preseleccionadas 4 y 5, respectivamente (**5' *SCN5A*-CTCF** y **5' *SCN5A*-GATA4**). Las gRNAs están separadas por 7.6 kb y solapan con lugares de unión a los FT cardíacos CTCF y GATA4, respectivamente (**Figura 76**).



**Figura 76. Regiones 4 y 5 en el 5' *SCN5A*.** Arriba, región preseleccionada con los picos correspondientes a DHS, CTCF, GATA4, H3K27ac, H3K4me3 y TBX5. Abajo, ampliación de la región 5' *SCN5A* donde están dirigidas las gRNAs (en violeta). También se muestran las marcas más relevantes para esta zona. Se indican también los motivos de unión predichos para CTCF y GATA4 con el programa JASPAR TFBS.

El TFBS de CTCF forma parte del bloque promotor de *SCN5A* y solapa con una fuerte señal DHS y ChIP-seq para CTCF. En cambio, el TFBS de GATA4 está en una región más alejada y no forma parte del bloque promotor. En este caso, aparte de la señal de ChIP-seq para GATA4, también hay actividad *enhancer* representada por la señal H3K27ac.

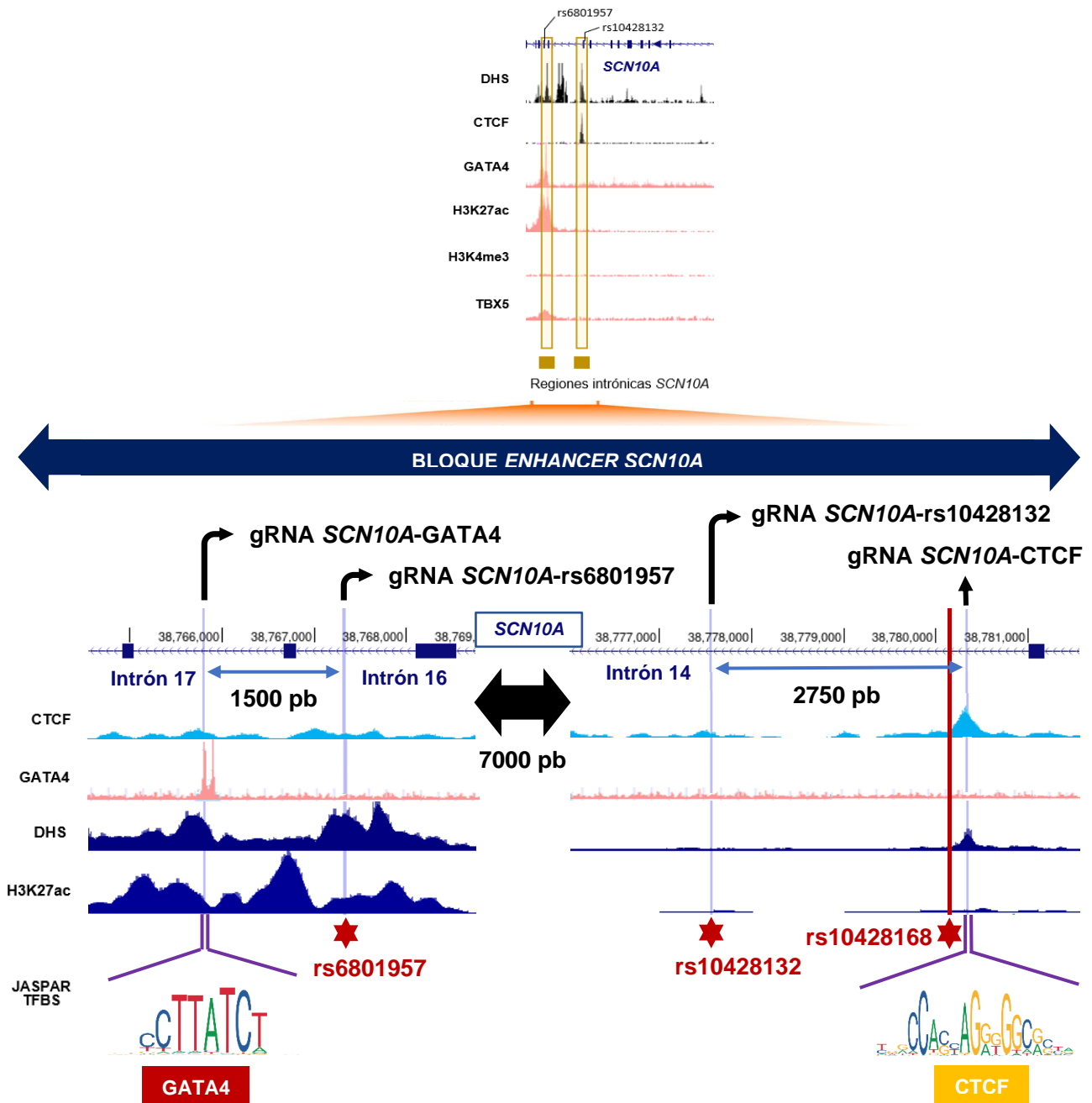
- **Regiones intrónicas en *SCN10A* (regiones 6 y 7)**

Las regiones de estudio 6 y 7 están situadas en el intrón 14, 16 y 17 del gen *SCN10A*, (**Figura 77**). En estas regiones, concretamente en el intrón 14, se ha descrito la presencia de un *enhancer* previamente descrito que actuaría como modulador de la expresión de *SCN5A*<sup>50</sup>. Asimismo, esta región también incluye el haplotipo de riesgo asociado a SBr descrito en el apartado 7.1.4. de la Introducción<sup>351</sup>.

En la región del intrón 14 de *SCN10A*, se diseñaron dos gRNAs. Una de ellas tiene su diana en un lugar de unión a CTCF predicho por JASPAR (***SCN10A-CTCF***). La segunda tiene su diana en la posición genómica donde aparece la variante rs10428132, una de las integrantes del haplotipo asociado a riesgo de SBr (***SCN10A-rs10428132***). Además, el TFBS para CTCF se localiza a solo 270 pb de la variante rs10428168, que también forma parte del haplotipo.

En el intrón 16 de *SCN10A* se diseñó otra gRNA (***SCN10A-rs6801957***) para dirigirla a las coordenadas de la variante rs6801957 del haplotipo. Se describió que el mecanismo por el cual rs6801957 se asocia a estos efectos sería la alteración en la unión de los FT TBX3/TBX5 con la región *enhancer*. La relación entre la variante y la unión de FT concuerda con la presencia de señales DHS y H3K27ac (**Figura 77**), así como señal ChIP-seq para TBX5 (**Figura 72**).

Por último, en el intrón 17, se diseñó una gRNA (***SCN10A-GATA4***) dirigida a un lugar de unión para GATA4 predicho por JASPAR. Esta gRNA está situada a 1 500 pb de la gRNA *SCN10A-rs6801957*.



**Figura 77. Región intrónica de *SCN10A* (regiones 6 y 7).** Arriba, región preseleccionada con los picos correspondientes a DHS, CTCF, GATA4, H3K27ac, H3K4me3 y TBX5. Abajo, ampliación de la región intrónica de *SCN10A* donde están dirigidas las gRNA (en violeta). También se muestran las marcas más relevantes para esa zona. En JASPAR TFBS se muestran los motivos de unión predichos para GATA4 y CTCF. El bloque *enhancer* se refiere a la región que contiene las variantes del haplotipo descritas por Pinsach-Abuin y colaboradores<sup>351</sup>.

Las once gRNAs seleccionadas fueron clonadas en plásmidos pgRNA-CKB siguiendo el protocolo descrito en el apartado 2.7.3. de Métodos.

## 2.2. Obtención de cardiomiocitos derivados de las líneas CRISPRi clonadas con las gRNAs

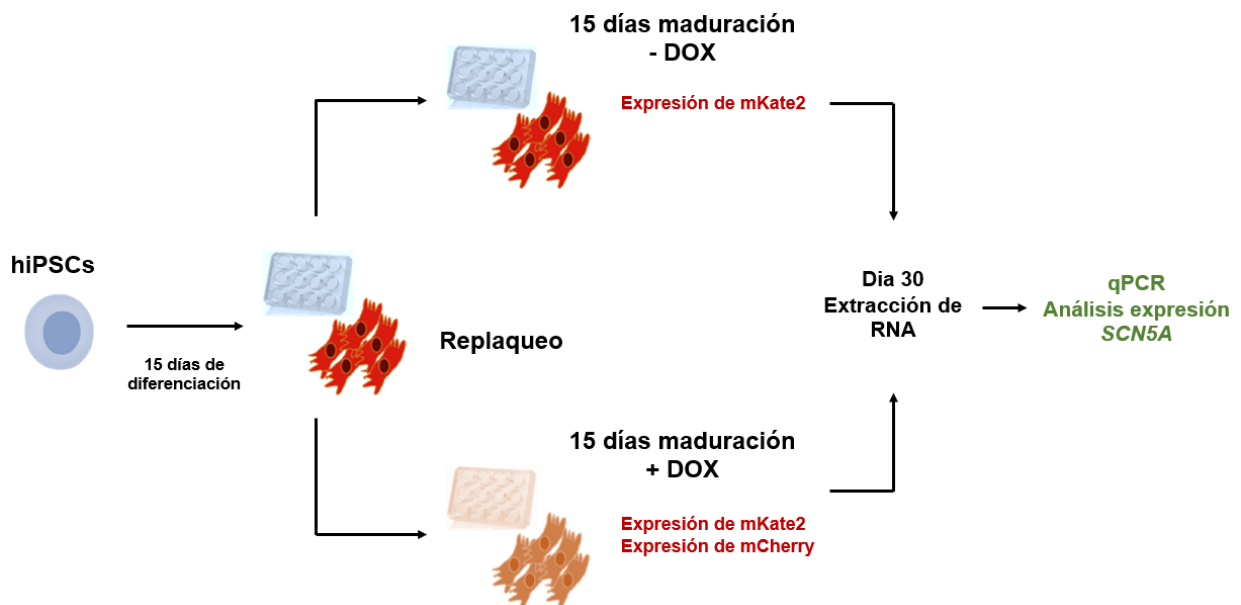
Una vez diseñadas las gRNAs de interés, nuestro objetivo fue generar líneas CRISPRi estables que tuviesen integradas las gRNAs y las expresasen de manera constitutiva. Estas líneas serían posteriormente diferenciadas a cardiomiocitos para la realización del ensayo CRISPRi. En primer lugar, se descongeló la línea celular de hiPSC CRISPRi WTB Gen1B y se mantuvo hasta el momento de nucleofectar los plásmidos con las gRNAs clonadas, siguiendo el protocolo descrito en el apartado 2.7.4 de Métodos. Durante los primeros 2-3 días después de la nucleofección, se observó una elevada muerte celular provocada por la selección con blasticidina. La blasticidina permitió seleccionar aquellas células que habían integrado el plásmido pgRNA-CKB en el genoma, mientras que el resto de células murieron. A partir del día 5-7, el crecimiento de las células fue en aumentando, debido a una disminución de la muerte celular por blasticidina. Aproximadamente 14 días después de la nucleofección, las células alcanzaron el 80% de la máxima confluencia. Con este procedimiento se obtuvieron 11 líneas celulares (**Tabla 16**), cada una de ellas expresando individualmente una gRNA.

**Tabla 16.** Líneas CRISPRi generadas a partir de la integración del pgRNA-CKB-gRNA especificado.

Línea celular	Abreviación	Región genómica	Diana
CRISPRi Gen1B gRNA <i>EXOG-CTCF</i>	(G1) <i>EXOG-CTCF</i>	Promotor <i>EXOG</i>	Lugar de unión CTCF
CRISPRi Gen1B gRNA Intrón 1-CTCF	(G2) Intrón 1-CTCF	Intrón 1 <i>SCN5A</i>	Lugar de unión CTCF
CRISPRi Gen1B gRNA Intrón 1-NKX2-5	(G3) Intrón 1-NKX2-5	Intrón 1 <i>SCN5A</i>	Lugar de unión NKX2-5
CRISPRi Gen1B gRNA TSS- 39pb	(G4) TSS-39pb	TSS <i>SCN5A</i>	39 pb <i>downstream</i> del TSS
CRISPRi Gen1B gRNA TSS+183pb	(G5) TSS+183pb	TSS <i>SCN5A</i>	183 pb <i>upstream</i> del TSS
CRISPRi Gen1B gRNA 5' <i>SCN5A-CTCF</i>	(G6) 5' <i>SCN5A-CTCF</i>	Región 5' <i>SCN5A</i>	Lugar de unión CTCF
CRISPRi Gen1B gRNA 5' <i>SCN5A-GATA4</i>	(G7) 5' <i>SCN5A-GATA4</i>	Región 5' <i>SCN5A</i>	Lugar de unión GATA4
CRISPRi Gen1B gRNA <i>SCN10A-GATA4</i>	(G8) <i>SCN10A-GATA4</i>	Intrón 17 <i>SCN10A</i>	Lugar de unión GATA4
CRISPRi Gen1B gRNA <i>SCN10A-rs6801957</i>	(G9) <i>SCN10A-rs6801957</i>	Intrón 16 <i>SCN10A</i>	Variante rs6801957
CRISPRi Gen1B gRNA <i>SCN10A-rs10428132</i>	(G10) <i>SCN10A-rs10428132</i>	Intrón 14 <i>SCN10A</i>	Variante rs10428132
CRISPRi Gen1B gRNA <i>SCN10A-CTCF</i>	(G11) <i>SCN10A-CTCF</i>	Intrón 14 <i>SCN10A</i>	Lugar de unión CTCF

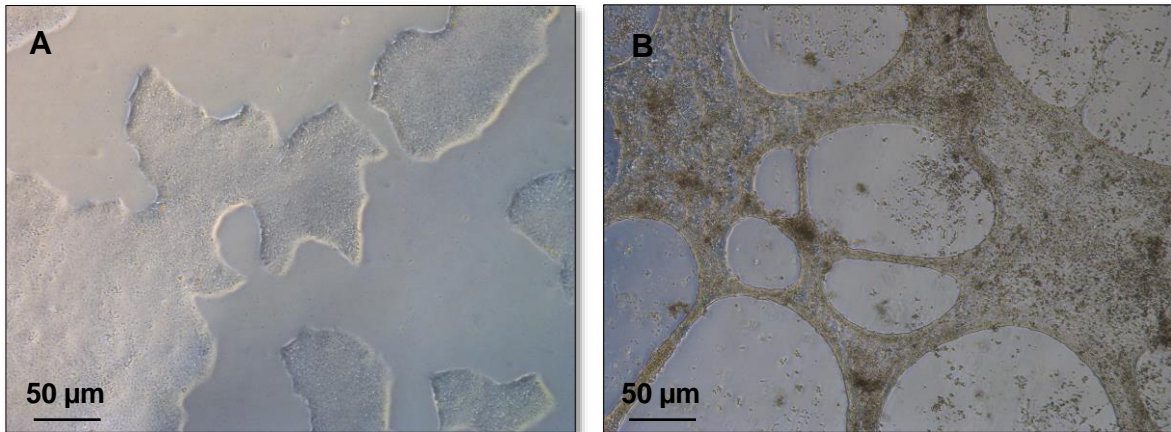
Para realizar el ensayo CRISPRi, las líneas de hiPSCs obtenidas fueron diferenciadas a cardiomiocitos. Durante la diferenciación, las células se incubaron con DOX para inducir la expresión de dCas9-KRAB, que posteriormente sería dirigida a la región de estudio por cada una de las gRNAs de cada línea celular (**Figura 78**).

La diferenciación de las hiPSCs a cardiomiocitos se realizó siguiendo el protocolo de diferenciación detallado en el apartado 2.3.4. de Métodos. Se realizaron tres rondas de diferenciación individuales (réplicas biológicas) para cada una de las 11 líneas. A nivel microscópico, a lo largo de los primeros 5-6 días se observó un fenotipo celular similar en todas las líneas. De forma similar a experimentos de diferenciación previos realizados con este protocolo, todas las células empezaron a latir entre los días 7 y 10 (**Figura 79**). Hasta el día del replaqueo (día 15), el crecimiento celular fue bastante homogéneo entre las diferentes líneas y las pocas variaciones que se observaron estaban relacionadas con la morfología, la formación de estructuras multicelulares características en cardiomiocitos y la intensidad del latido.



**Figura 78. Esquema general del proceso seguido en los ensayos CRISPRi.** Después del replaqueo en el día 15 de diferenciación, las células se dividen en pocillos idénticos, y unos son tratados y los otros no (+DOX/-DOX). Después de los 30 días de diferenciación, se extrae el RNA de ambas condiciones y se analiza la expresión de *SCN5A* mediante qPCR.

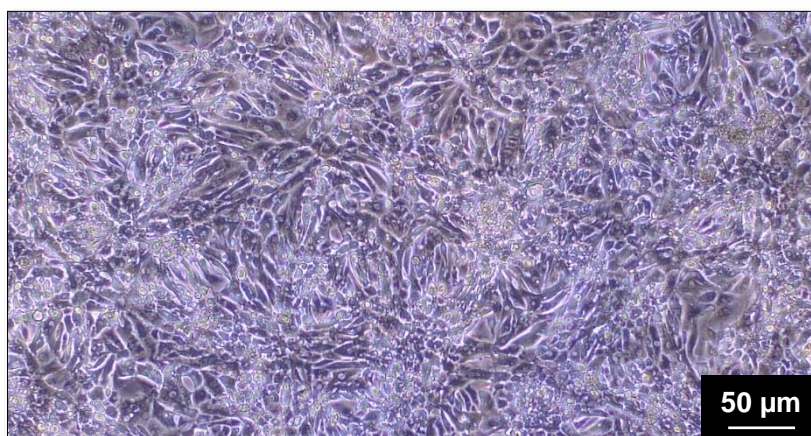




**Figura 79. Imagen de hiPSCs y hiPSC-CMs.** (A) se observan colonias de hiPSCs en perfecto estado (a una confluencia aproximadamente del 50%). Los bordes de las colonias son definidos y brillantes y no se observan diferenciaciones inespecíficas. (B) se observan hiPSC-CMs en el día 8 de diferenciación. Muestran la estructura propia en esa fase de la diferenciación, donde las células se despegan en gran parte del pocillo y laten acompasadamente.

Debido a la variabilidad inherente en todos los protocolos con hiPSCs, una vez realizado el replaqueo no se pudo disponer del mismo número de células para cada una de las rondas de diferenciación. Mientras que en algunas de ellas pudieron sembrarse pocillos con 2 millones de células, en otras líneas sólo se obtuvo material para sembrar pocillos con 1 millón de células.

Sin embargo, el hecho de comenzar el tratamiento con DOX justo después del replaqueo (**Figura 80**) evitó que esta variabilidad tuviese influencia en el análisis de la expresión de *SCN5A*. Para cada ronda de diferenciación, se replaquearon un mínimo de 2 pocillos idénticos, de los cuales un 50% fueron tratados con DOX, por lo que cada comparación (tratados vs. no tratados) se realizó entre condiciones idénticas (**Figura 78**). A partir de este momento, las muestras tratadas con doxiciclina se denominan +DOX, y las muestras no tratadas, -DOX.

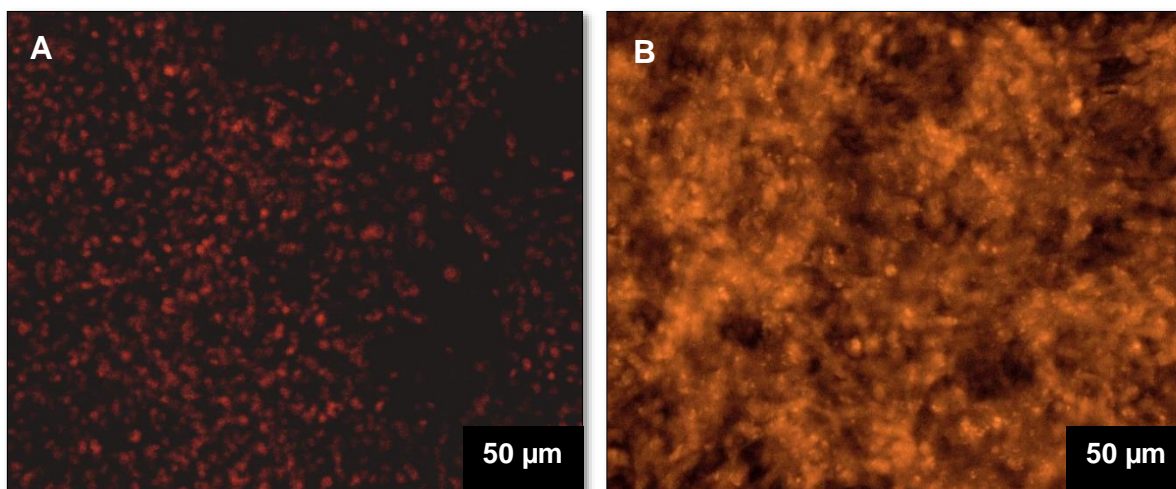


**Figura 80. Imagen de hiPSC-CMs después del replaqueo.** Aspecto de los hiPSC-CMs entre el día 16 y 30 de diferenciación. Las células laten de manera coordinada, sin generar tanto movimiento como antes del replaqueo.

La línea CRISPRi clonadas se caracteriza por tener integrados dos genes que codifican dos proteínas fluorescentes distintas y sirven como marcadores:

- **mKate2**: el gen de esta proteína fluorescente está presente en el plásmido pgRNA-CKB (**Figura 41**) nucleofectado en las células. La expresión de las gRNAs va acompañada de la expresión de mKate2, y sus niveles de expresión son directamente proporcionales entre ellos.
- **mCherry**: el constructo integrado en la línea CRISPRi que contiene dCas9-KRAB, incluye el gen de mCherry (**Figura 39**), de manera que aquellas células tratadas con DOX, expresarán mCherry.

Tanto mKate2 como mCherry pueden visualizarse mediante microscopía de fluorescencia. Debido a que mKate2 y mCherry emiten en el mismo espectro (600-630 nm), no es posible diferenciarlas en el microscopio de fluorescencia. No obstante, sí fue posible diferenciar las células +DOX de las -DOX, ya que la expresión de mCherry otorga una intensidad de color característica (**Figura 81**). La visualización de mCherry sirvió también como validación de que la expresión de dCas9-KRAB se estaba produciendo.



**Figura 81. Imagen de la expresión de mKate2 y mCherry en células hiPSC-CMs en el microscopio de fluorescencia.** (A) hiPSC-CMs -DOX, donde solamente hay expresión de mKate2, que indica que la gRNA correspondiente también se está expresando en las células fluorescentes. (B) hiPSC-CMs +DOX, en las cuales la expresión de mKate2 y mCherry se solapa y no pueden diferenciarse. Aun así, la intensidad de la señal es mucho mayor debido a la expresión de mCherry, inducida por DOX.

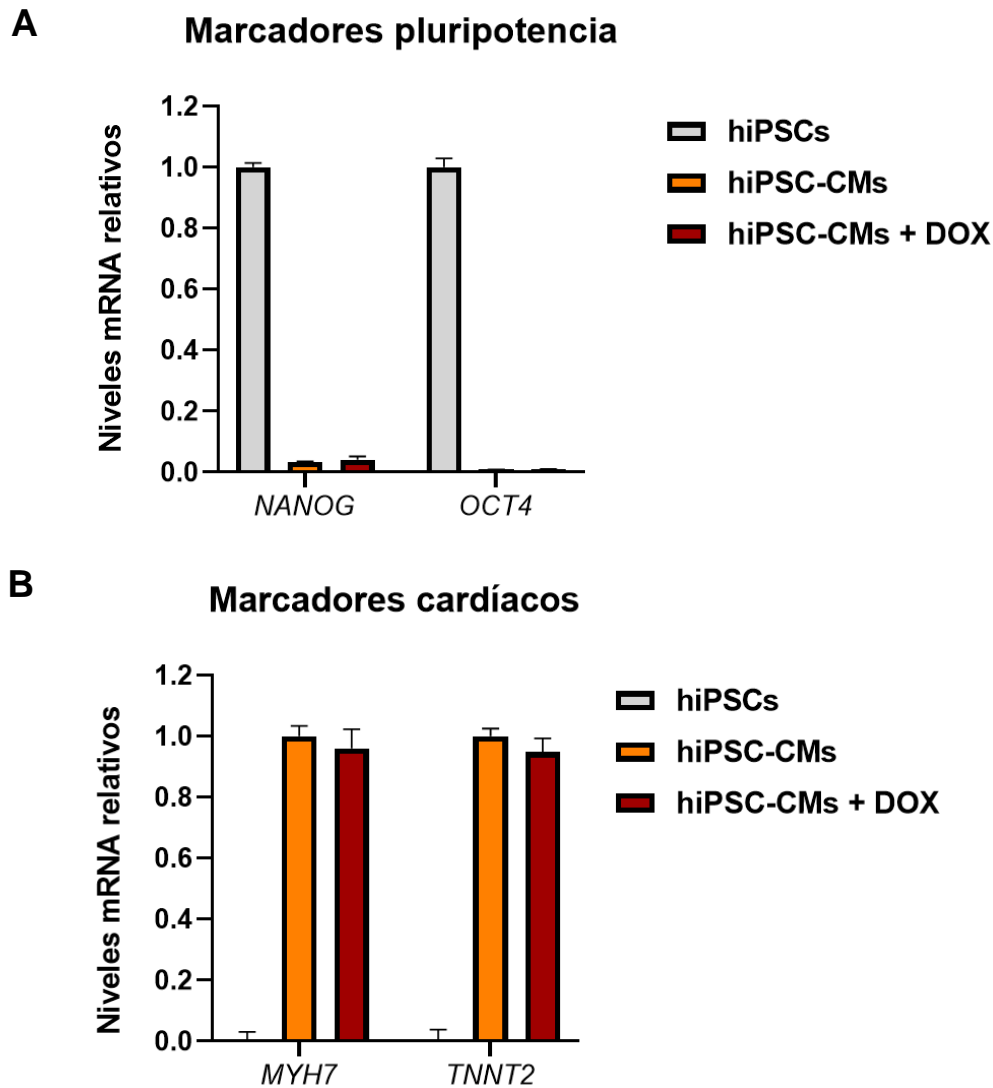
### 2.3. Análisis del proceso de diferenciación de hiPSCs a cardiomiocitos por qPCR

Con el fin de confirmar que el proceso de diferenciación a cardiomiocitos había funcionado correctamente, se realizó una extracción de RNA total de los hiPSC-CMs en los días 0 y 30 de diferenciación. Este RNA se utilizó para analizar la expresión de marcadores de pluripotencia y de cardiomiocitos mediante qPCR. Las ratios de pureza (OD 260/280), obtenidos mediante la medición del RNA con Nanodrop oscilaron entre 1.8 y 1.98, unos valores adecuados para realizar la qPCR<sup>385</sup>. Como marcadores de pluripotencia, se utilizaron los genes *NANOG* y *OCT4*, los cuáles actúan como reguladores clave del potencial de diferenciación de células madre y controlan la diferenciación embrionaria de células pluripotentes en tejidos endodérmicos y mesodérmicos<sup>382</sup>. Como marcadores de cardiomiocitos, se utilizaron los genes *MYH7* y *TNNT2*, que codifican para la cadena pesada de la  $\beta$ -miosina y la troponina T, respectivamente. Ambos genes se expresan de manera específica en el tejido muscular cardíaco, por lo que son adecuados para evaluar la transición de célula pluripotente a célula con fenotipo cardíaco<sup>383</sup>. Finalmente, como normalizadores para todas las muestras, se utilizaron los genes *housekeeping* *PP1A* y *GAPDH*, que codifican para las proteínas fosfatasa 1 y gliceraldehído-3-fosfato deshidrogenasa, respectivamente, y que se expresan de forma estable<sup>384</sup>.

El análisis de la expresión de los marcadores de pluripotencia *NANOG* y *OCT4* por qPCR mostró una expresión 31 veces inferior en los hiPSC-CMs respecto a las hiPSCs para *NANOG*, y 125 veces inferior para *OCT4* (**Figura 82A**). Por lo tanto, los niveles de expresión para los dos marcadores de pluripotencia fueron prácticamente indetectables en las células hiPSC-CMs. Además, la expresión en hiPSC-CMs + DOX no presento cambios significativos respecto a hiPSC-CMs.

Estos resultados demuestran, en primer lugar, que las células hiPSCs de la línea CRISPRi son pluripotentes. Asimismo, muestran que la diferenciación celular a cardiomiocitos conlleva la pérdida de esta característica, ya que estos no expresan dichos marcadores de pluripotencia.

De manera opuesta, en el análisis de expresión de los marcadores cardíacos *MYH7* y *TNNT2*, los hiPSC-CMs presentaron hasta 10 000 veces mayor expresión de estos genes en comparación con las hiPSCs, donde son indetectables (**Figura 82B**). Conjuntamente, estos resultados confirman que el proceso de diferenciación cardíaca en la línea CRISPRi fue exitoso.



**Figura 82. Expresión de los marcadores de pluripotencia *NANOG* y *OCT4*. (A) y marcadores cardíacos *MYH7* y *TNNT2* (B) en la línea CRISPRi WTB Gen1B para las condiciones de hiPSCs, hiPSC-CMs y hiPSC-CMs + DOX. Las barras de error para cada muestra fueron calculadas a partir de la desviación típica de los valores  $\Delta Ct$ .**

## 2.4. Análisis del efecto de las gRNAs sobre la expresión de *SCN5A*

A partir de la diferenciación cardíaca de las 11 líneas (**Tabla 16**) se obtuvo el RNA para evaluar las posibles alteraciones en la expresión de *SCN5A* inducidas por cada gRNA. Partiendo de las muestras celulares disgregadas el día 30 de diferenciación, se realizó la extracción de RNA: para cada gRNA y ronda de diferenciación individual se realizó la extracción del RNA para +DOX y -DOX. A continuación, se examinó la concentración de muestras de RNA obtenidas, obteniendo valores entre 50 y 300 ng/ $\mu$ L. Las ratios de pureza (OD 260/280), obtenidos mediante la medición

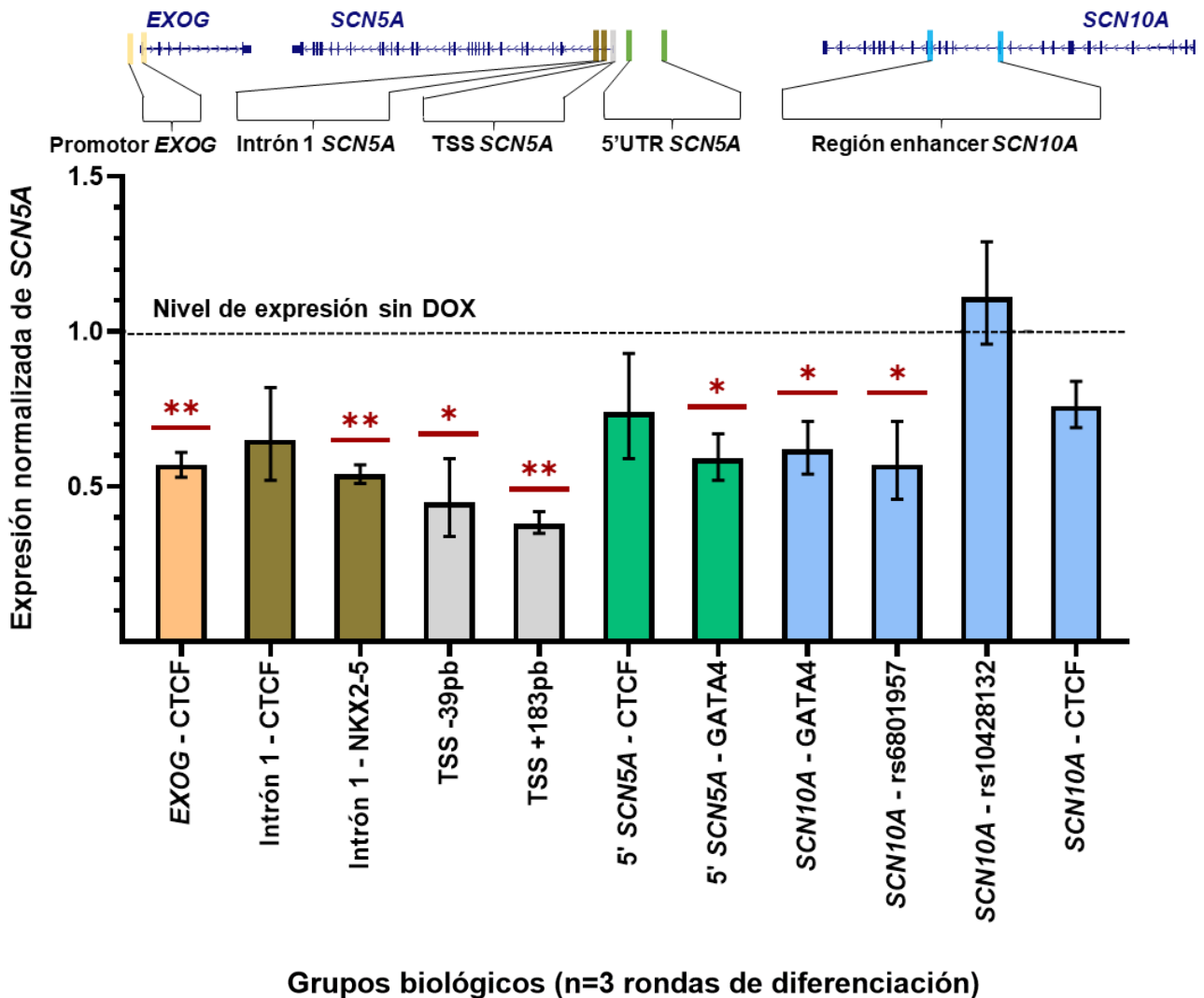
del RNA con Nanodrop, oscilaron entre 1.7 y 2.04 unos valores considerados adecuados para realizar la qPCR<sup>385</sup>.

Para evaluar la expresión de *SCN5A* se realizó una qPCR utilizando sondas para los genes *PP1A* y *GAPDH* como genes normalizadores o *housekeeping*. El análisis de los datos se realizó siguiendo las pautas de la revisión de Taylor *et al.*<sup>368</sup> (revisar apartado 2.7.9.1. de Métodos).

De manera global, los resultados de la qPCR muestran una disminución generalizada en la expresión de *SCN5A* en todas las líneas celulares excepto en una (**SCN10A-rs10428132**) (**Figura 83**). Para determinar la significación estadística de estas disminuciones en la expresión se realizó una prueba-t pareada usando los datos de expresión normalizados para cada réplica biológica. Se comprobó mediante el test de normalidad Shapiro-Wilk que los datos eran normales y se podía realizar un test paramétrico. Como resultado, de las 11 líneas CRISPRi, las que presentan una disminución estadísticamente significativa son siete, concretamente **EXOG-CTCF**, **Intrón 1-NKX2-5**, **TSS-39pb**, **TSS+183pb**, **5'SCN5A-GATA4**, **SCN10A-GATA4** y **SCN10A-rs6801957**.

Un aspecto interesante a destacar es que, de las regiones genómicas a estudiar, definidas anteriormente en la **Figura 72** (promotor *EXOG*, intrón 1 *SCN5A*, TSS *SCN5A*, 5' *SCN5A* y región *enhancer* *SCN10A*), en todas hubo al menos una línea CRISPRi que indujo la *downregulación* de *SCN5A* en presencia de DOX. Esto sugiere que las 5 regiones identificadas como potencialmente reguladoras contienen secuencias que participan en la regulación de *SCN5A*.

Los resultados con las dos líneas TSS que se utilizaron como control (**TSS-39pb** y **TSS+183pb**) indican que el sistema CRISPRi funciona correctamente para inhibir la expresión génica. De hecho, las dos líneas creadas para hacer *knockdown* en esta región son las que muestran una mayor disminución porcentual de la expresión de *SCN5A* con un 55% de disminución para **TSS-39pb** y un 62% para **TSS+183pb** (respecto la condición sin DOX). Las líneas con mayor significancia estadística después del TSS fueron **EXOG-CTCF** e **Intrón 1-NKX2-5**, ambas con una  $p < 0.01$ . Por su parte, **EXOG-CTCF** disminuyó la expresión de *SCN5A* en un 43%, mientras que **Intrón 1-NKX2-5** lo hizo en un 46%.



**Figura 83. Análisis de la expresión de SCN5A en las líneas CRISPRi hiPSC-CM.** La figura muestra los niveles de mRNA de SCN5A en la condición +DOX respecto a la expresión basal (-DOX), representada con la línea discontinua en el 1. Los colores de las barras corresponden a la región genómica donde la gRNA está dirigida. Los valores de expresión relativos fueron calculados siguiendo el ejemplo de la **Figura 49** (columna 9) para cada una de las líneas. Las barras de error representadas fueron calculadas siguiendo el ejemplo de la **Figura 49**, de la siguiente manera:  $2^{\text{columna 10} + \text{columna 12}}$  para la barra superior y  $2^{\text{columna 10} - \text{columna 12}}$  para la barra inferior. \* $p \leq 0,05$ , \*\* $p \leq 0,01$  y \*\*\* $p \leq 0,001$ .

Estos niveles de inhibición transcripcionales son muy relevantes teniendo en cuenta que esta técnica no genera *knockouts* y, además, en el caso de **EXOG-CTCF**, se interfiere con una secuencia relativamente lejana a SCN5A.

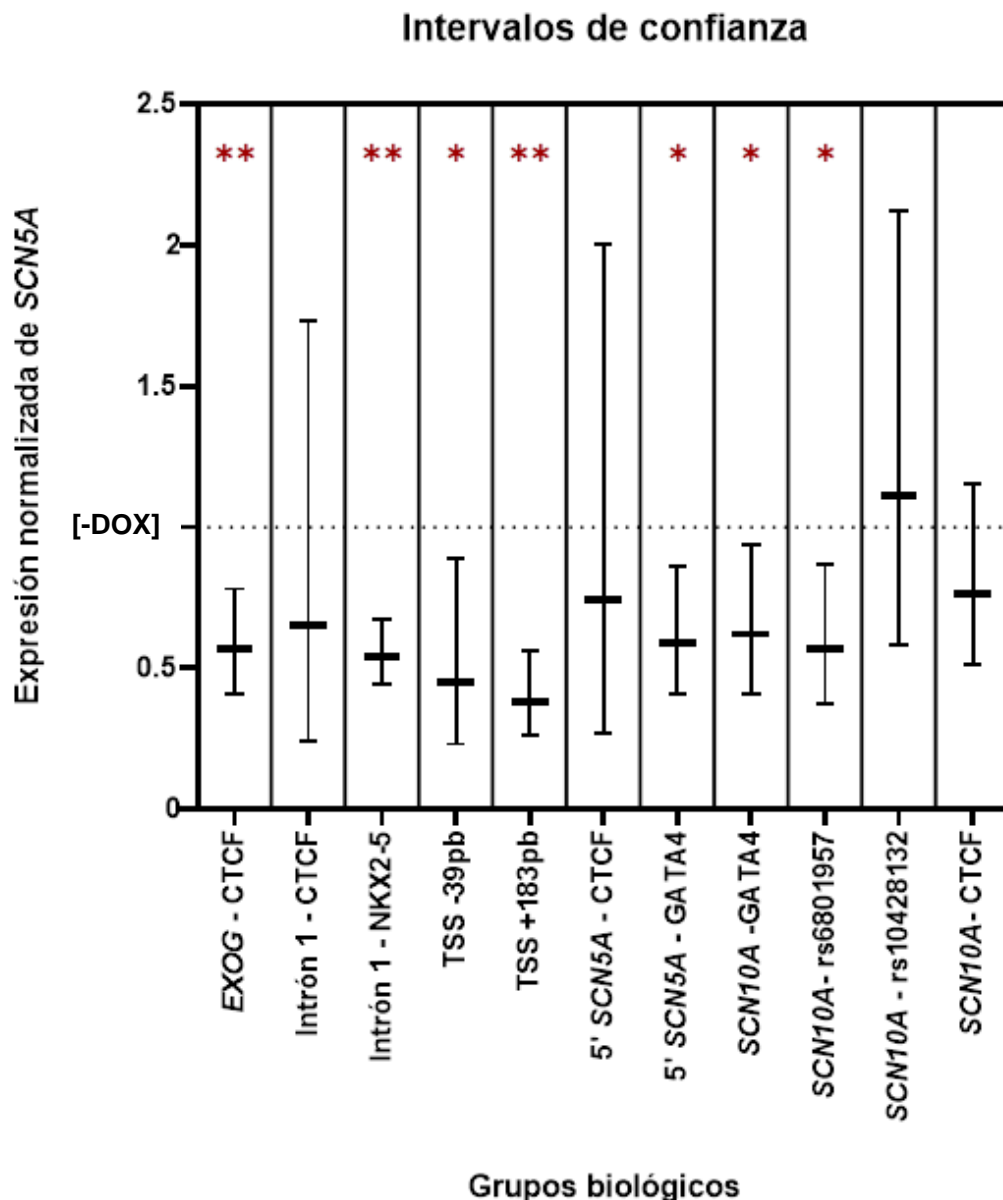
Salvo la línea **EXOG-CTCF**, el resto de líneas con CTCF como diana no disminuyeron la expresión de SCN5A de forma significativa, resultando solo en 1 de 4 líneas dirigidas a CTCF con efectos sobre la regulación del gen. En el caso de **Intrón 1-CTCF**, la expresión de SCN5A disminuyó un 35% (aunque con una SD elevada entre réplicas), para **5' SCN5A-CTCF** un 26% y

para **SCN10A-CTCF** un 24%. Ninguna de las reducciones fue estadísticamente significativa, aunque en el apartado 2.1. de la discusión se discute la posible significación clínica de **SCN10A-CTCF**. Se puede hipotetizar que la presencia de diversos elementos CTCF a lo largo del locus **SCN5A-SCN10A** pudiesen estar actuando conjuntamente en la regulación de **SCN5A**, sugiriendo que la alteración de uno de esos elementos pueda no ser suficiente para alterar la expresión génica, o que solamente suceda en las secuencias con mayor relevancia (en este caso se sugeriría **EXOG-CTCF**). Esta hipótesis estaría apoyada por diversos estudios que han observado un efecto de robustez frente a la variación genética en *clusters* de CTCF que, al actuar juntos, su contribución a la regulación génica es probablemente redundante en lugar de aditiva<sup>386,387</sup>.

En cambio, en las dos líneas con la diana GATA4, nuestros resultados muestran una disminución significativa de la expresión de **SCN5A**: **5'SCN5A-GATA4** con una disminución de un 41% y **SCN10A-GATA4**, de un 38%.

Por otro lado, de las 4 líneas con diana en la región intrónica de **SCN10A**, 2 de ellas presentaron disminuciones significativas en la expresión del **SCN5A** (**SCN10A-GATA4** y **SCN10A-rs6801957**). Este hecho resulta especialmente interesante, ya que estaría demostrando que el sistema CRISPRi funciona incluso en regiones muy distantes al gen de interés. Las 4 líneas expresan gRNAs dirigidas a distintos intrones de **SCN10A**: la línea **SCN10A-GATA4** sobre el intrón 16 y la línea **SCN10A-rs6801957** el intrón 17, en este caso disminuyendo la expresión de **SCN5A** en un 43%. **SCN10A-rs10428132** y **SCN10A-CTCF** tenían el intrón 14 como diana en ambos casos, siendo **SCN10A-rs10428132** la única línea con un aumento en la expresión de **SCN5A**, concretamente un 11%, aumento que no resultó significativo.

La representación de los mismos resultados en términos de intervalos de confianza (**Figura 84**) se realizó siguiendo el ejemplo de los cálculos efectuados en la **Figura 51** (columnas 13 y 14). En los estudios científicos, los p-valores se utilizan para determinar si una hipótesis nula formulada antes de la realización del estudio debe ser aceptada o rechazada. En los estudios exploratorios, los p-valores permiten el reconocimiento de cualquier hallazgo estadísticamente significativo. Los intervalos de confianza nos proporcionan información sobre un rango en el que el valor real se encuentra con un cierto grado de probabilidad, así como sobre la dirección y la fuerza del efecto demostrado<sup>388</sup>. Esto permite sacar conclusiones sobre la plausibilidad estadística y la relevancia clínica de los hallazgos del estudio. A menudo es útil que los artículos científicos proporcionen ambas medidas estadísticas, porque proporcionan tipos de información complementarios. Para nuestros datos de qPCR, se obtuvo un rango para cada línea de estudio donde se puede afirmar con un 95% de confianza que el efecto de dCas9-KRAB sobre una determinada región se encuentra en el rango representado. Así, en el caso de la línea **SCN10A-CTCF**, a pesar de que su disminución no es estadísticamente significativa, su intervalo de confianza podría revelar un resultado clínicamente significativo y con potencial para ser



**Figura 84. Análisis de la expresión de SCN5A en las líneas CRISPRi hiPSC-CM mediante intervalos de confianza.** La línea punteada en el número 1 [-DOX], representa la expresión de SCN5A en la condición -DOX. Cada intervalo de confianza representa la media del grupo biológico y el rango de expresión en el que cada línea celular + DOX se encontraría con un 95% de confianza.

estudiado más ampliamente (más desarrollado en el apartado 2 de la discusión). Esta hipótesis tiene sentido con el hecho de que los resultados con una  $p \leq 0.01$  son aquellos con un intervalo de confianza más estrecho, pero que al mismo tiempo es un factor altamente influido por el número de muestras o la dispersión entre ellas. En concordancia con estas observaciones, el efecto en las líneas con mayor SD entre réplicas biológicas (**Intrón 1-CTCF**, **5'SCN5A-CTCF** o **SCN10A-rs10428132**) no resultó estadísticamente significativo. Por el contrario, también cabe



destacar la robustez de resultados como el de la línea **TSS-39 pb**, donde aún con una SD considerable, el intervalo de confianza muestra una disminución significativa clara.

En resumen, a pesar de que los p-valores son generalmente indicadores robustos para confirmar o descartar hipótesis, es importante valorar cada caso individualmente antes de llegar a una conclusión definitiva sobre el significado clínico del resultado.

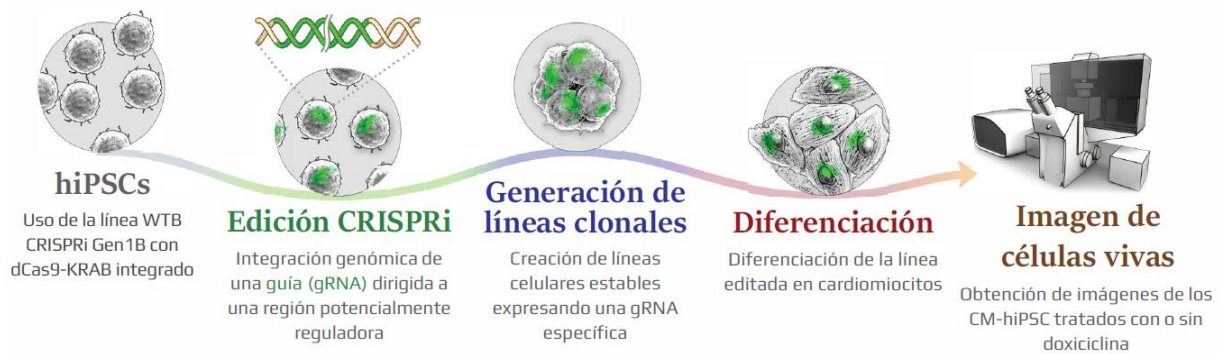
## 2.5. Alteraciones del sodio intracelular en experimentos funcionales

Los resultados obtenidos con las 11 líneas CRISPRi fueron prometedores de caras a identificar nuevas vías de regulación de la expresión del gen *SCN5A*. La observación de una disminución en los niveles de expresión de *SCN5A* para algunas regiones sugiere que funcionarían como regiones reguladoras de *SCN5A*. Con el objetivo de confirmar estos resultados utilizando otra aproximación experimental, se estudió si las alteraciones en la expresión de *SCN5A* se traducen en alteraciones funcionales en la electrofisiología celular, concretamente en los niveles de sodio intracelular. Nuestra hipótesis es que los niveles de sodio intracelular podrían reflejar cambios en la actividad de los canales de sodio de los cardiomiocitos de estudio.

Para llevar a cabo estos experimentos funcionales, se realizaron nuevas rondas de diferenciación para las mismas líneas celulares usadas en los análisis de expresión (**Figura 85**). Se siguió el mismo protocolo a excepción del replaqueo, el cual se realizó en placas de 384 pocillos adecuadas para este procedimiento. Una vez más, la mitad de los pocillos de cada ronda fueron tratados con DOX. Al llegar al día 30 de diferenciación, las células se mantuvieron vivas para realizar las mediciones del sodio intracelular con un sistema de imagen celular automático de alta definición (Operetta). Se utilizaron 2 indicadores fluorescentes para medir diferentes parámetros:

- (i) indicador SBFI-AM. Se caracteriza por unir-se al  $\text{Na}^+$  y emitir señal en el espectro del color verde. La intensidad de la fluorescencia aumenta (proporcionalmente) con la unión a  $\text{Na}^+$  (**Figura 86**).
- (ii) indicador Hoechst, utilizado para visualizar los núcleos celulares, utilizado para normalizar la cuantificación del indicador de sodio. Emite en el espectro del color azul (**Figura 86**).

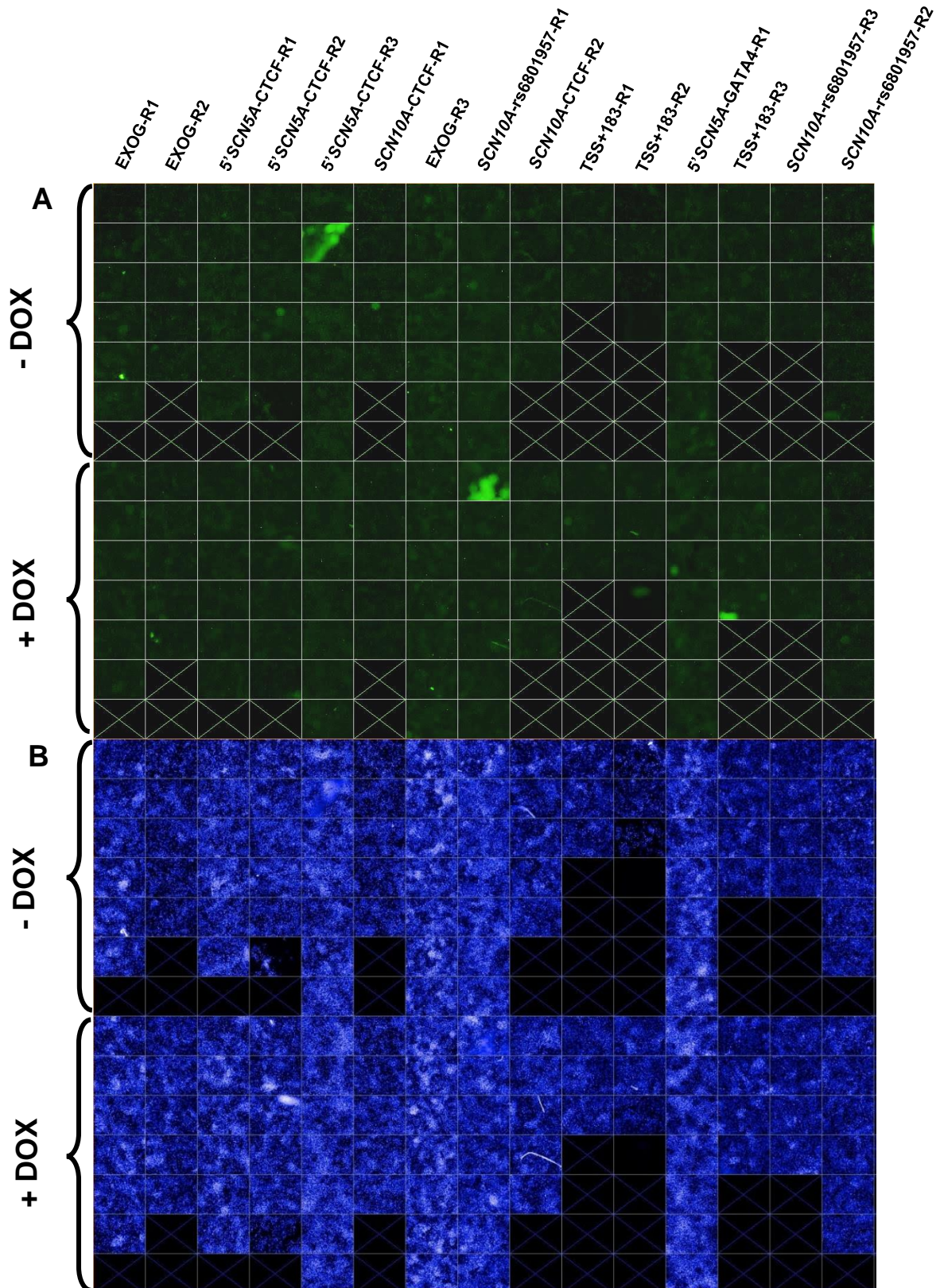
Además de estos dos indicadores, las células expresan mKate2 en paralelo a la expresión de la gRNA y también mCherry en aquellas tratadas con DOX, ambos emitiendo en el espectro del color rojo (**Figura 87**). Todos los indicadores fueron medidos simultáneamente, por lo que se eligieron cuidadosamente para evitar que sus rangos de excitación y emisión se solapasen (**Figura 52**).



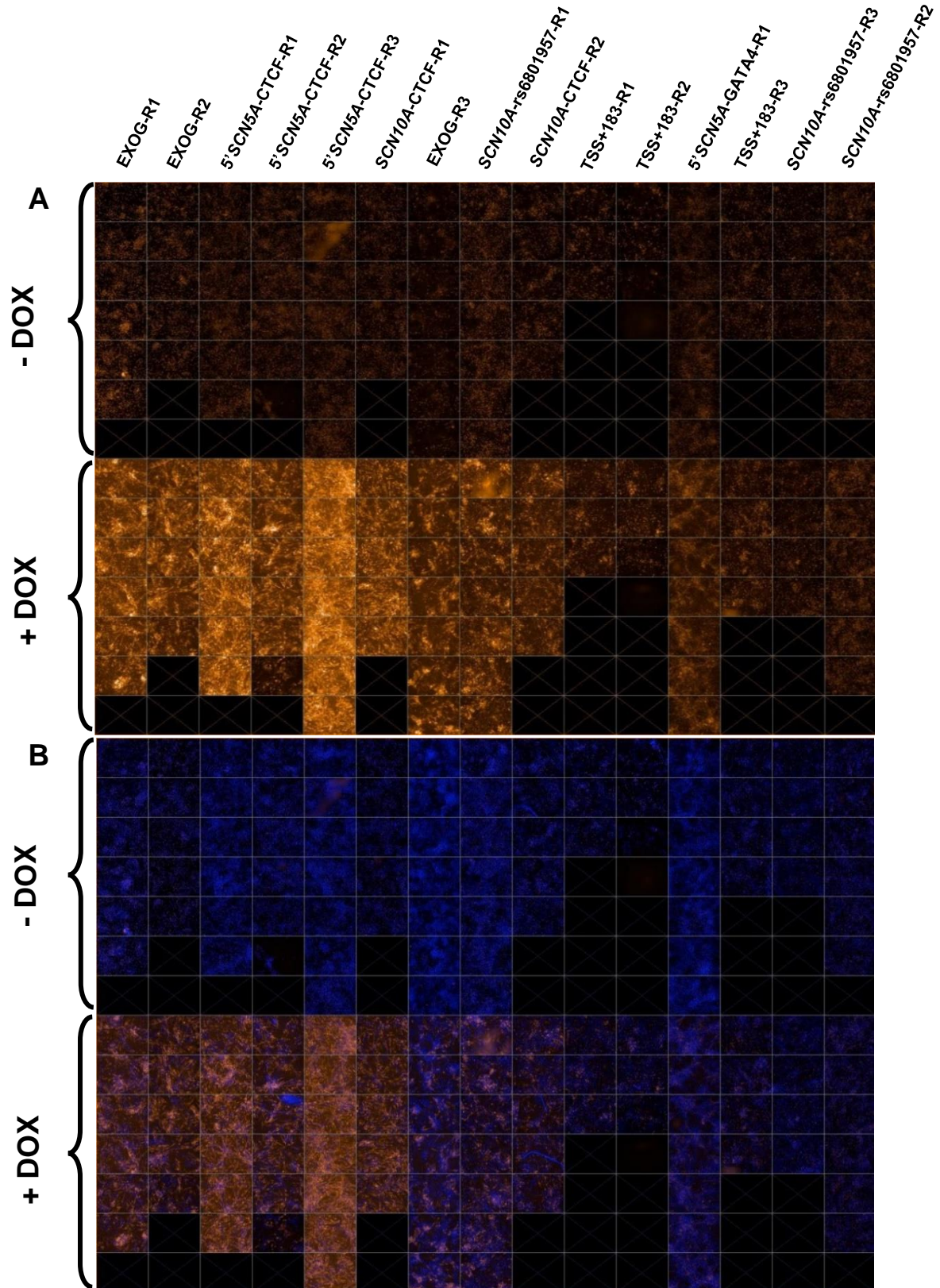
**Figura 85. Esquema del procedimiento utilizado para analizar el sodio intracelular en las líneas CRISPRi.**

Estos experimentos fueron realizados en el laboratorio de Genómica y Bioinformática del CiMUS (Santiago de Compostela), donde pudimos tener acceso al sistema Operetta, necesario para hacer las mediciones. El hecho de realizar los experimentos en otro laboratorio y restringidos a un tiempo limitado añadió una serie de dificultades en el desarrollo de la metodología que son importantes a tener en cuenta. Para algunas de las líneas celulares se obtuvieron resultados fiables y representativos, mientras que otras tuvieron que ser excluidas por los siguientes motivos:

- Algunas rondas de diferenciación tuvieron un mal desarrollo (exceso de mortalidad, formación incorrecta de las estructuras típicas de los hiPSC-CMs o ausencia de latido) antes del replaqueo y tuvieron que desecharse. Esta circunstancia es relativamente normal y sucede ocasionalmente, aunque se aplique siempre el mismo protocolo de manera idéntica. En estos casos, la solución es hacer una nueva ronda diferenciación, pero dadas las limitaciones temporales, en nuestro caso no fue posible.
- Los cardiomiocitos cultivados en placas con 384 pocillos mostraron, en algunos casos, un tiempo de supervivencia inferior a lo visto con el formato de 12 pocillos. Concretamente, se pudo observar que algunas rondas empezaban a tener mortalidad y empeoramiento del fenotipo antes de transcurrir los 15 días de maduración hormonal. Debido a esta circunstancia, se tuvieron que desechar varias rondas de diferenciación. A pesar de que todas las rondas se replaquearon en la misma placa de 384 pocillos, no todas fueron replaqueadas al mismo tiempo, de manera que en una misma placa había rondas en distintos momentos temporales. Después de excluir las rondas no viables y cuando la mayoría del resto de rondas viables estaban entre el día 28-30 de diferenciación, se finalizó el proceso de diferenciación para toda la placa. De esta forma, se pudieron realizar las lecturas en Operetta para garantizar la integridad de todas las células aptas para el análisis.



**Figura 86. Análisis del sodio intracelular en las líneas celulares CRISPRi hiPSC-CMs, en la placa de 384 pocillos. (A) Señal captada para el  $\text{Na}^+$  intracelular con la sonda fluorescente SBF1-AM. (B) Señal de los núcleos celulares captada con la sonda Hoechst. Los pocillos con una cruz (o negros) representan pocillos sin usar.**



**Figura 87. Análisis del sodio intracelular en las líneas celulares CRISPRi hiPSC-CMs, en la placa de 384 pocillos. (A)** La fluorescencia detectada en las células sin DOX corresponde a la señal de mKate2, mientras que en las células con DOX la señal es la suma de mKate2 y mCherry. **(B)** Se muestra la señal de mKate2 y mCherry solapada con Hoechst.

- El efecto colateral de la decisión anterior fue que algunas rondas aún se encontraban entre el día 18-22 de diferenciación, con lo cual también estuvieron expuestas menos días al tratamiento con DOX (aplicado a partir del día 15). Este hecho se tiene en cuenta al analizar los resultados.

Debido a los motivos expuestos anteriormente, las líneas **Intrón 1-NKX2-5** y **SCN10A-CTCF**, así como una o dos rondas de las líneas **5'SCN5A-GATA4** y **SCN10A-GATA4**, se tuvieron que excluir del análisis. Las líneas **Intrón 1-CTCF**, **SCN10A-rs10428132** y **TSS-39 pb** no habían sido incluidas en el estudio desde un principio. Los criterios de exclusión fueron la ausencia de alteraciones en la expresión de *SCN5A* en las dos primeras líneas y la selección del control positivo que había dado una mayor disminución en la expresión de *SCN5A* (**TSS+183 pb**).

Las líneas celulares de las cuales se pudieron obtener una o más rondas de diferenciación válidas y sin inconvenientes fueron analizadas en el sistema Operetta para determinar las variaciones en el sodio intracelular entre las condiciones -DOX y +DOX. Para cada ronda de diferenciación se obtuvo un valor numérico para la señal de Hoechst (número de núcleos detectados), un valor numérico para mKate2/mCherry (que detecta la expresión de las gRNAs/dCas9-KRAB, respectivamente) y un valor numérico para SBFI-AM (sodio intracelular detectado). Los valores de mKate2/mCherry y SBFI-AM se normalizaron por el valor de la señal Hoechst, eliminando así la variabilidad celular de los resultados (**Tabla 17**). También se calculó la variación de la señal mKate2/mCherry entre las condiciones -DOX y +DOX de cada ronda de diferenciación, obteniendo la variación de esta señal en +DOX respecto a -DOX (**Tabla 17**). Aquellas rondas que recibieron un tratamiento con DOX más corto (de 4 a 7 días, en lugar de los 15 originales), también fueron analizadas (datos marcados en naranja, **Tabla 17**).

De las 15 rondas de diferenciación analizadas, 4 de ellas tuvieron un tratamiento con DOX acortado, de las cuales ninguna dio lugar a alteraciones significativas del sodio intracelular. Los valores de la variación mKate2/mCherry en estas 4 rondas oscilan entre x1.043 y x1.249, lo cual significa que la inducción de dCas9-KRAB (y mCherry en paralelo) supuso un aumento de la fluorescencia debido a mCherry de entre el 4.3% y el 24.9% (+DOX), respecto a los niveles observados solamente con mKate2 (-DOX). En cambio, en las rondas con un ciclo de 15 días con DOX se observaron aumentos en la señal mKate2/mCherry de hasta 4 veces más en las condiciones +DOX respecto a -DOX (valores de variación mKate2/mCherry entre x1.684 y x4.014; **Tabla 17**). La variación de SBFI-AM se calculó de la misma manera, obteniendo el aumento o disminución del sodio intracelular detectado en +DOX respecto a -DOX.

**Tabla 17. Mediciones del sodio intracelular (+/- DOX) para las líneas celulares CRISPRi hiPSC-CMs.** Para cada ronda de diferenciación y condición (-DOX/+DOX) se obtuvieron valores de fluorescencia de Hoechst, mKate2/mCherry y SBFI-AM. Los valores se normalizaron a partir de Hoechst y se calcularon las variaciones de mKate2/mCherry y SBFI-AM entre -DOX y +DOX. En naranja: menor tiempo de tratamiento con DOX. En verde: rondas de diferenciación con una disminución estadísticamente significativa del sodio intracelular en la condición +DOX (**Figura 88**).

Ronda diferenciación (réplicas técnicas)	Señal Hoechst	Señal mKate2/mCherry	Señal SBFI-AM	mKate2/mCherry Norm.	Variación mKate2/mCherry	SBFI-AM Norm.	Variación SBFI-AM
EXOG-CTCF R1 -DOX (n=6)	787	80	226	0.102	x4.014	0.288	<b>x0.799</b>
EXOG-CTCF R1 +DOX (n=6)	799	326	183	0.408		0.23	
EXOG-CTCF R2 -DOX (n=5)	597	90	185	0.151	x2.140	0.311	<b>x0.846</b>
EXOG-CTCF R2 +DOX (n=5)	654	211	171	0.323		0.263	
EXOG-CTCF R3 -DOX (n=7)	777	66	181	0.085	<b>x1.249</b>	0.233	x1.017
EXOG-CTCF R3 +DOX (n=7)	754	80	182	0.106		0.237	
TSS+183 R1 -DOX (n=3)	472	85	177	0.180	x1.731	0.379	<b>x0.894</b>
TSS+183 R1 +DOX (n=3)	507	158	172	0.312		0.339	
TSS+183 R2 -DOX (n=3)	344	72	172	0.209	x1.851	0.507	<b>x0.722</b>
TSS+183 R2 +DOX (n=3)	444	172	169	0.387		0.366	
TSS+183 R3 -DOX (n=4)	478	60	164	0.126	<b>x1.143</b>	0.346	x1.003
TSS+183 R3 +DOX (n=4)	502	72	173	0.143		0.347	
5' SCN5A-CTCF R1 -DOX (n=6)	580	127	187	0.219	x1.936	0.322	x0.968
5' SCN5A-CTCF R1 +DOX (n=6)	592	251	183	0.424		0.312	
5' SCN5A-CTCF R2 -DOX (n=5)	502	126	184	0.251	x1.684	0.369	x0.919
5' SCN5A-CTCF R2 +DOX (n=5)	530	224	179	0.423		0.339	
5' SCN5A-CTCF R3 -DOX (n=6)	603	73	168	0.121	<b>x1.043</b>	0.279	x0.978
5' SCN5A-CTCF R3 +DOX (n=6)	602	76	164	0.126		0.273	
5' SCN5A-GATA4 R1 -DOX (n=7)	792	47	194	0.059	x2.624	0.247	<b>x0.955</b>
5' SCN5A-GATA4 R1 +DOX (n=7)	790	123	186	0.156		0.236	
SCN10A-GATA4 R1 -DOX (n=5)	704	83	189	0.118	x3.608	0.269	<b>x0.926</b>
SCN10A-GATA4 R1 +DOX (n=5)	717	305	178	0.425		0.249	
SCN10A-GATA4 R2 -DOX (n=5)	682	95	187	0.139	x2.255	0.274	x1.018
SCN10A-GATA4 R2 +DOX (n=5)	640	201	178	0.314		0.279	
SCN10A-rs6801957 R1 -DOX (n=6)	709	87	193	0.123	x1.888	0.272	x1.033
SCN10A-rs6801957 R1 +DOX (n=6)	643	149	181	0.232		0.281	
SCN10A-rs6801957 R2 -DOX (n=5)	530	90	187	0.170	x1.897	0.353	<b>x0.898</b>
SCN10A-rs6801957 R2 +DOX (n=4)	534	172	180	0.322		0.317	
SCN10A-rs6801957 R3 -DOX (n=4)	528	67	174	0.127	<b>x1.268</b>	0.33	x1.064
SCN10A-rs6801957 R3 +DOX (n=3)	522	84	182	0.161		0.351	

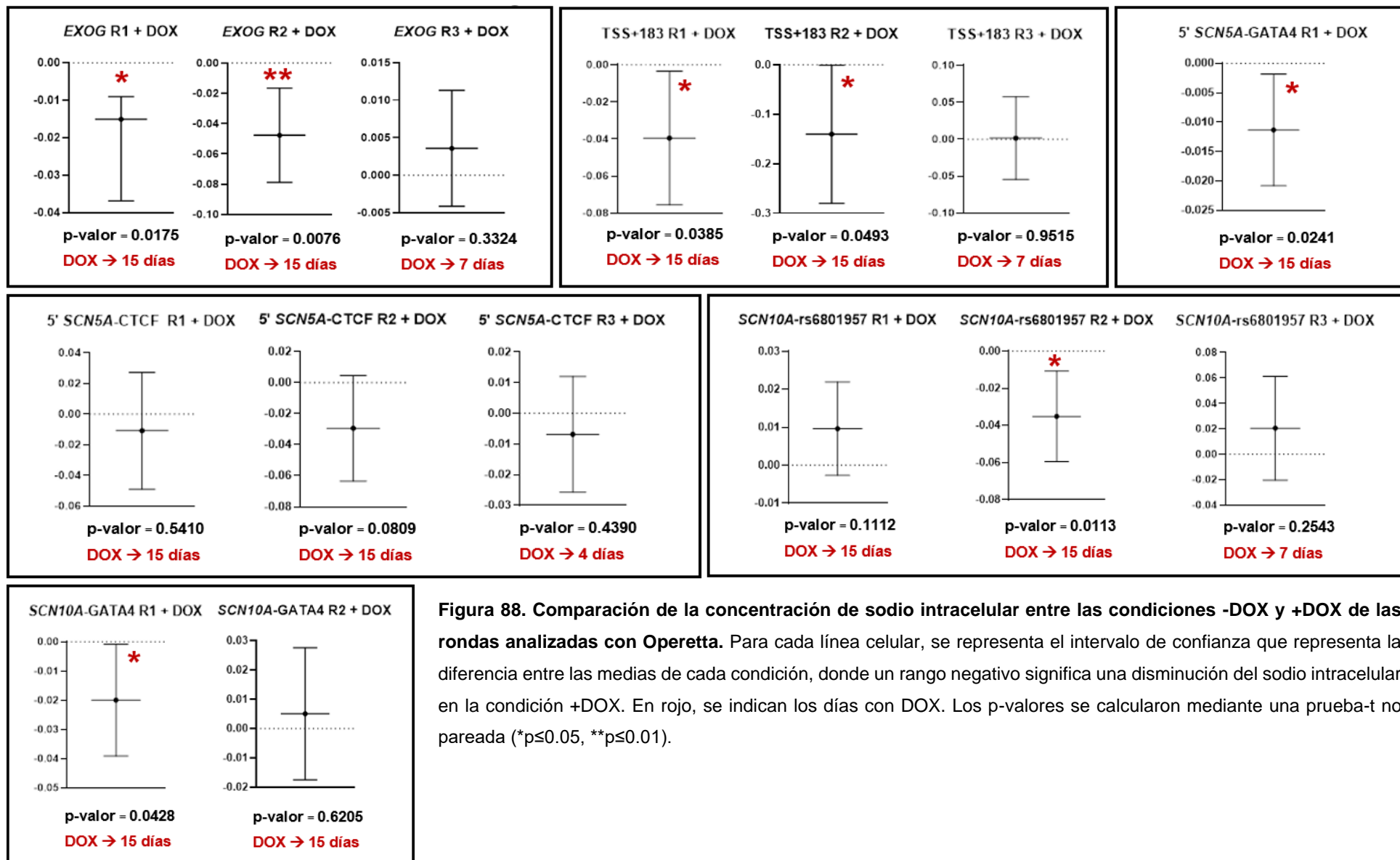


Figura 88. Comparación de la concentración de sodio intracelular entre las condiciones -DOX y +DOX de las rondas analizadas con Operetta. Para cada línea celular, se representa el intervalo de confianza que representa la diferencia entre las medias de cada condición, donde un rango negativo significa una disminución del sodio intracelular en la condición +DOX. En rojo, se indican los días con DOX. Los p-valores se calcularon mediante una prueba-t no pareada (\*p≤0.05, \*\*p≤0.01).

De manera general, los resultados de los experimentos de medición del sodio intracelular siguen una tendencia similar a las diferencias de expresión de *SCN5A* observadas en los experimentos de qPCR. La reducción de los niveles de sodio intracelular en condiciones +DOX no es tan marcada como la disminución de los niveles de mRNA de *SCN5A*. Sin embargo, los resultados muestran una reducción significativa del sodio intracelular en 7 rondas de diferenciación: dos rondas en la línea **EXOG-CTCF**, dos rondas en la línea **TSS+183**, una ronda en la línea **5' SCN5A-GATA4**, una ronda en la línea **SCN10A-GATA4** y una ronda en la línea **SCN10A-rs6801957**. En todas ellas, se detectó una disminución del sodio intracelular de entre el 4.5% y el 27.8% (+DOX), respecto a los niveles observados sin DOX. Es importante destacar que todas las líneas en las que en al menos una ronda de diferenciación se detectaron alteraciones en el sodio intracelular, también se observó una disminución de la expresión de *SCN5A*.

Para determinar la significancia en las variaciones del sodio intracelular, se realizó una prueba t no pareada usando los valores normalizados SBF1-AM de las réplicas entre las condiciones -DOX y +DOX (**Figura 88**). Mediante los test de normalidad Shapiro-Wilk y Kolmogorov-Smirnov se comprobó que los datos eran normales y se podía realizar un test paramétrico. En las rondas **5' SCN5A-GATA4 R1** y **SCN10A-GATA4 R1**, se observó una disminución del sodio intracelular relativamente discreta en la condición +DOX (del 4.5% y 7.4%, respectivamente; **Tabla 17**), pero en ambos casos estadísticamente significativa. Debido al número más elevado de réplicas en comparación con las qPCR (hasta 7 réplicas) y unos valores de SD pequeños, en estos casos se pudieron detectar variaciones significativas a pesar de ser diferencias discretas.

La representación de los intervalos de confianza para la diferencia entre medias indica el rango de valores que contiene la diferencia entre las dos medias con un 95% de confianza, además de indicar los p-valores para cada comparación (**Figura 88**).

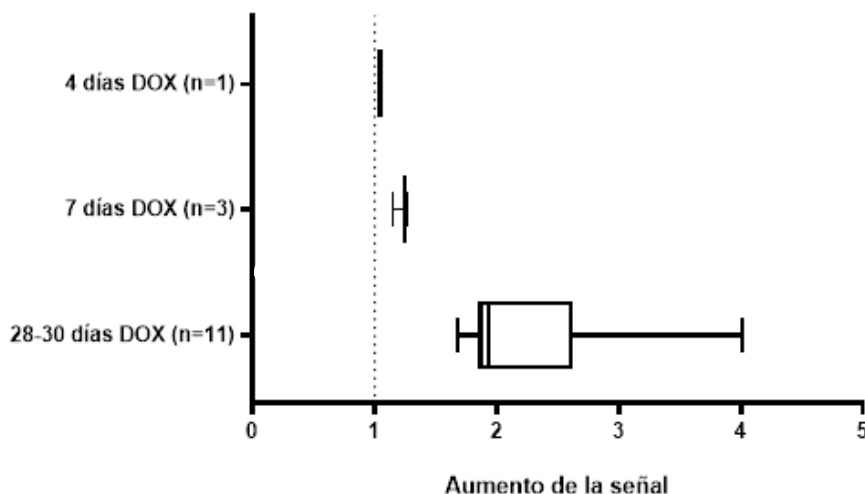
Para las rondas de diferenciación de la línea **EXOG-CTCF**, se observó una disminución significativa del sodio intracelular en las rondas R1 y R2 tratadas con DOX (20% y 15% de reducción, respectivamente), mientras que en la ronda R3 la diferencia no fue significativa. Es relevante que R3 solo recibió 7 días de tratamiento con DOX, lo cual resultó en un escaso aumento de la señal mKate2/mCherry (x1.25 veces más) en comparación con las otras 2 rondas (x4 y x2.1, respectivamente). En el caso de la línea **TSS+183**, se observó el mismo patrón que en la línea **EXOG-CTCF**. Las rondas R1 y R2 tratadas con DOX presentan una disminución significativa del sodio intracelular (11% y 28% de reducción, respectivamente), mientras que la ronda R3 recibió menos tiempo de tratamiento con DOX y, en consecuencia, un menor aumento de mKate2/mCherry. Es importante destacar que, aunque la línea **TSS+183** fue tratada como un control positivo en los experimentos de qPCR, no se podía asumir que esa condición fuera a replicarse con el sodio intracelular. Sin embargo, la ronda R2 de esta línea obtuvo la mayor reducción en el sodio intracelular de todas las líneas estudiadas. Los resultados obtenidos tanto



para *EXOG-CTCF* como para TSS+183 coinciden con los observados en las qPCR, siendo 2 de las líneas con mayor porcentaje de disminución y mayor significancia estadística. Las 3 rondas analizadas para la línea **SCN10A-rs6801957** presentan resultados dispares. Solo la ronda R2 mostró una disminución significativa del sodio intracelular. En las rondas R1 y R3 no se observaron cambios relevantes, aunque R3 también estuvo tratada menos días con DOX. Contrariamente, en esta línea también se observó una disminución significativa de un 43% de la expresión de *SCN5A* en las qPCR. Para la línea **5'SCN5A-CTCF**, no se observaron alteraciones significativas en el sodio intracelular para ninguna de las 3 rondas tratadas con DOX, en línea con lo observado en los resultados de qPCR.

Para las líneas **SCN10A-GATA4** y **5'SCN5A-GATA4** no se pudo disponer de 3 rondas de diferenciación para el análisis. En el primer caso, se dispuso de 2 rondas donde solo una de ellas mostró alteraciones de sodio intracelular en presencia de DOX. En los resultados de qPCR, **SCN10A-GATA4** presentó una disminución de la expresión de *SCN5A*. Con los datos obtenidos en los experimentos funcionales no se puede afirmar que los resultados de la qPCR se trasladen a los niveles de proteína intracelular. En el caso de **5'SCN5A-GATA4**, solo se dispuso de una sola ronda, en la cual se observó una disminución significativa del sodio intracelular en la condición +DOX. Aunque estos resultados estarían en línea con los resultados de expresión en qPCR, la falta de réplicas es insuficiente para confirmar los resultados.

#### Señal de mKate2/mCherry según tiempo de exposición con DOX



**Figura 89. Análisis de la señal de mKate2/mCherry en las líneas CRISPRi hiPSC-CMs tratadas con DOX en función del tiempo.** Diagrama de caja que representa la mediana, cuartiles y valores atípicos de los incrementos en la señal mKate2/mCherry en condiciones +DOX, respecto a -DOX, para cada tiempo de tratamiento (4, 7 y 11 días).

El análisis de la señal mKate2/mCherry en presencia de DOX a lo largo del tiempo en todas las rondas mostró que, a medida que disminuye el tiempo de tratamiento con DOX, también lo hace la expresión de mCherry, que es proporcional a la expresión de dCas9-KRAB (**Figura 89**). Aunque no es posible medir mCherry de manera individual debido a su solapamiento con mKate2, se puede deducir que el aumento de la señal mKate2/mCherry obtenida al tratar con DOX es debida a la expresión de dCas9-KRAB. En base a experimentos previos realizados por Mandegar y colaboradores<sup>220</sup> con la misma línea CRISPRi utilizada en nuestro estudio, se esperaba que 7 días de tratamiento con DOX (2  $\mu$ M) fuesen suficientes para tener una expresión máxima de dCas9-KRAB (saturación). Sin embargo, las mediciones de mKate2/mCherry realizadas en las rondas con 7 o menos días de tratamiento demuestran una diferencia significativa con las rondas tratadas 15 días (**Figura 89**). Las rondas con menos tiempo de tratamiento apenas presentan un aumento de la señal de un 25% respecto a la señal basal de mKate2, en el mejor de los casos, muy inferior a las rondas con el tratamiento completo. Por este motivo, estas rondas deben quedar excluidas como significativas para el estudio ya que podría ser que la cantidad de dCas9-KRAB expresada no fuese suficiente para determinar si, en realidad, se produciría un efecto funcional o no.

En conclusión, de todas las líneas CRISPRi que pudieron ser estudiadas funcionalmente, todas excepto **5'SCN5A-CTCF** presentaron disminuciones en el sodio intracelular. Este resultado está de acuerdo con los experimentos de qPCR, en los cuales se detectó una disminución de la expresión de *SCN5A* para todas las líneas excepto **5'SCN5A-CTCF**. En el resto de casos, solo **EXO-CTCF** y **TSS+183** presentan una evidencia robusta de que las alteraciones observadas a nivel de expresión de *SCN5A* se traducen en una alteración funcional de los niveles de sodio intracelular, siendo las dos líneas con una mayor disminución de estos niveles.

## 2.6. Estudio comparativo entre las regiones CRISPRi y la información disponible de interacciones de cromatina

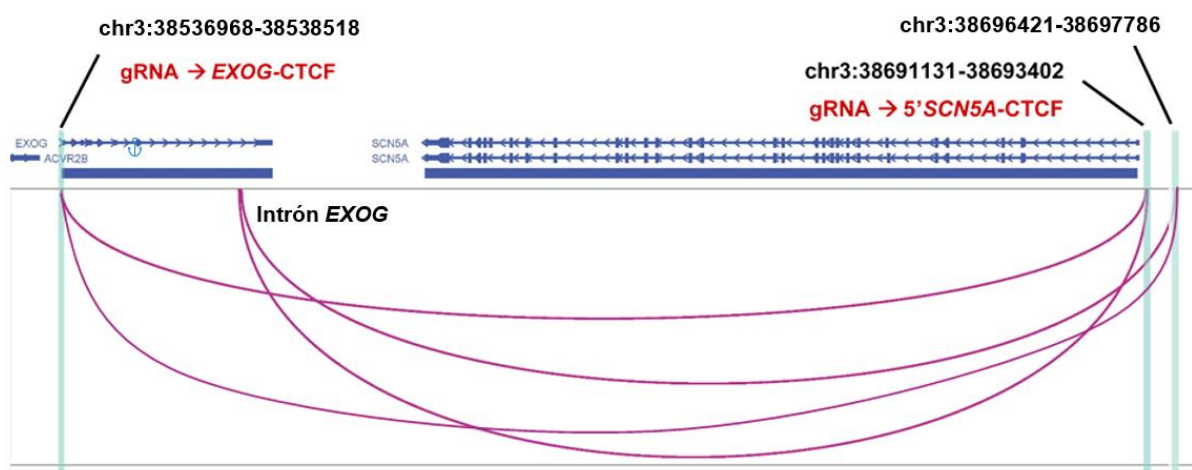
En los apartados anteriores se determinó el efecto de la inhibición de las regiones preseleccionadas como potencialmente reguladoras de *SCN5A* sobre la expresión del gen *SCN5A* i la concentración de sodio intracelular. Para complementar estos resultados, realizamos un análisis de la posible vinculación de esas regiones con las interacciones de la cromatina descritas en la bibliografía.

En primer lugar, se utilizaron datos experimentales del estudio de Tang *et al.*<sup>430</sup>, quienes identificaron más de 42000 interacciones mediadas por CTCF usando ChIA-PET en una línea celular humana de linfocitos B (GM12878). La técnica ChIA-PET se puede utilizar para determinar la contribución de una proteína en concreto en la formación de redes de contacto con

otras secuencias proximales o distales, como promotores de los genes con los que interactúa. Los datos obtenidos por Tang *et al.* se utilizaron para visualizar la unión de CTCF en nuestra región de estudio mediante el *WashU Epigenome Browser* (apartado 1.7. de Materiales). Se seleccionaron las regiones de interacción específicas que coincidieran con las secuencias de unión de los gRNAs de CRISPRi (**Figura 90**):

- Región de interacción en el promotor *EXOG*: chr3:38536968-38538518
- Regiones de interacción en 5' *SCN5A*:
  1. chr3:38691131-38693402
  2. chr3:38696421-38697786
  3. chr3:38704305-38704526 (no mostrado en **Figura 90**)

Es importante destacar que dentro del rango de interacción del promotor *EXOG* se encuentra el lugar donde estaba dirigida la gRNA juntamente con dCas9-KRAB en la línea ***EXOG-CTCF*** (chr3:38537735; **Tabla 2** del apartado 1.4.2.1. de Materiales). La gRNA para la línea ***5'SCN5A-CTCF*** se dirigió específicamente dentro del rango de la región de interacción 1 en 5' *SCN5A* (chr3:38692885; **Tabla 2** del apartado 1.4.2.1. de Materiales). La gRNA para la línea ***5'SCN5A-GATA4*** se dirigió unas 4.4 kb más allá del límite de la segunda zona de interacción en 5' *SCN5A*, a 3.8 kb del inicio de la tercera zona de interacción (chr3:38700496; **Tabla 2** del apartado 1.4.2.1. de Materiales). Estos datos sugieren una posible vinculación entre los CTCF-BS del promotor *EXOG* y de 5' *SCN5A-CTCF* debido a las interacciones observadas (apartado 3 de la Discusión).



**Figura 90. ChIA-PET de CTCF en la región *EXOG-SCN5A*.** Se muestran las interacciones entre regiones con los arcos en morado. En rojo se muestran las gRNA cuya diana coincide con una región de interacción. La información de las interacciones se realizó en la línea celular GM12878. Figura extraída del *WashU Epigenome Browser* a partir de los datos de Tang *et al.*<sup>430</sup>.

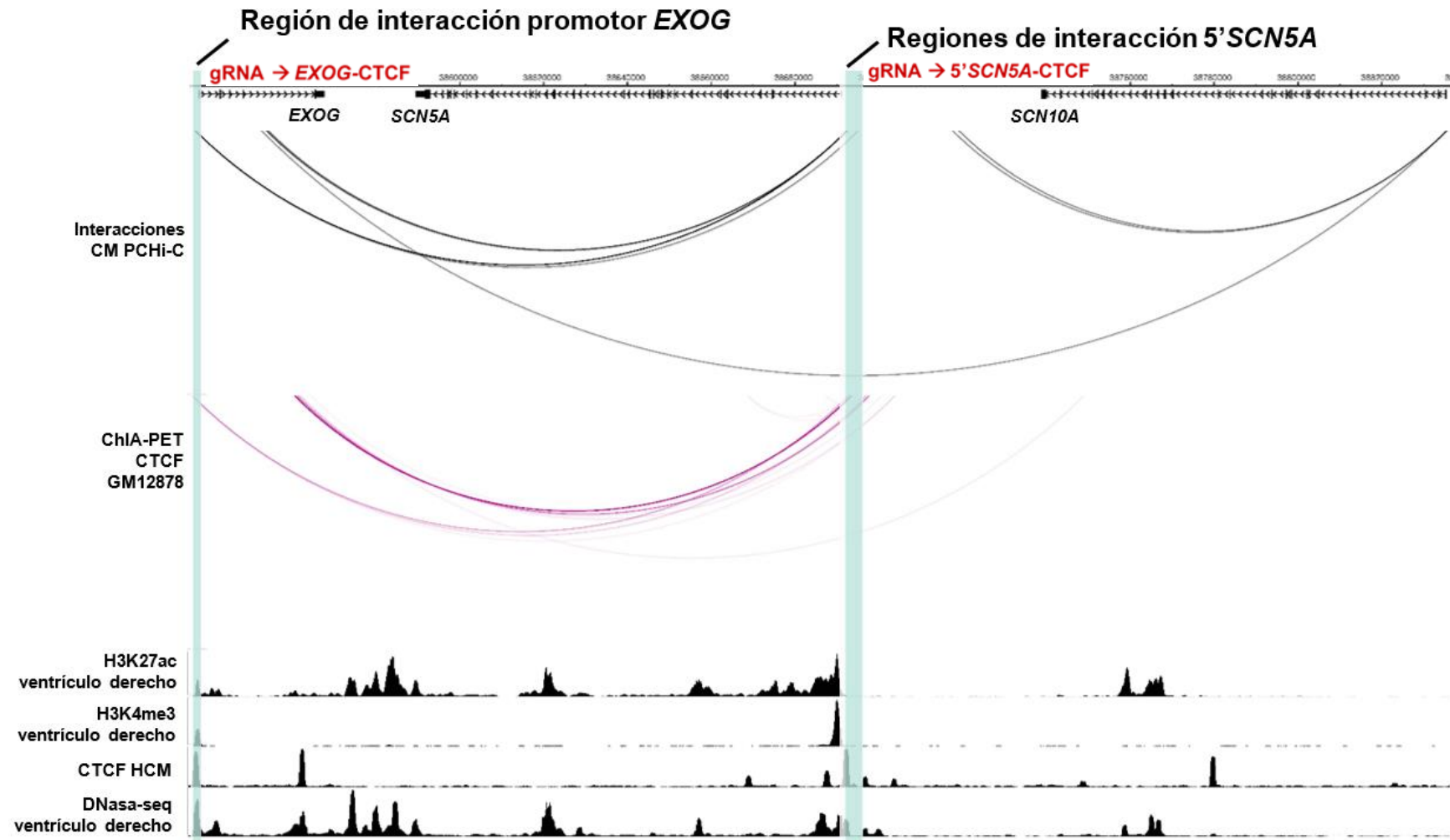
Los datos de ChIA-PET de Tang *et al.* también indican que la región de ***5'SCN5A-GATA4*** no está interaccionando con la región promotora de *EXOG*. Sin embargo, sí se pudo identificar

interacción entre una región intrónica de *EXOG* con la región donde se dirigió la gRNA de **5'SCN5A-GATA4**. Al mismo tiempo, esta región intrónica también mostró interacciones con la región de **5'SCN5A-CTCF (Figura 90)**, sugiriendo que el conjunto de estas regiones podría estar coordinado en la regulación de la expresión de *SCN5A*.

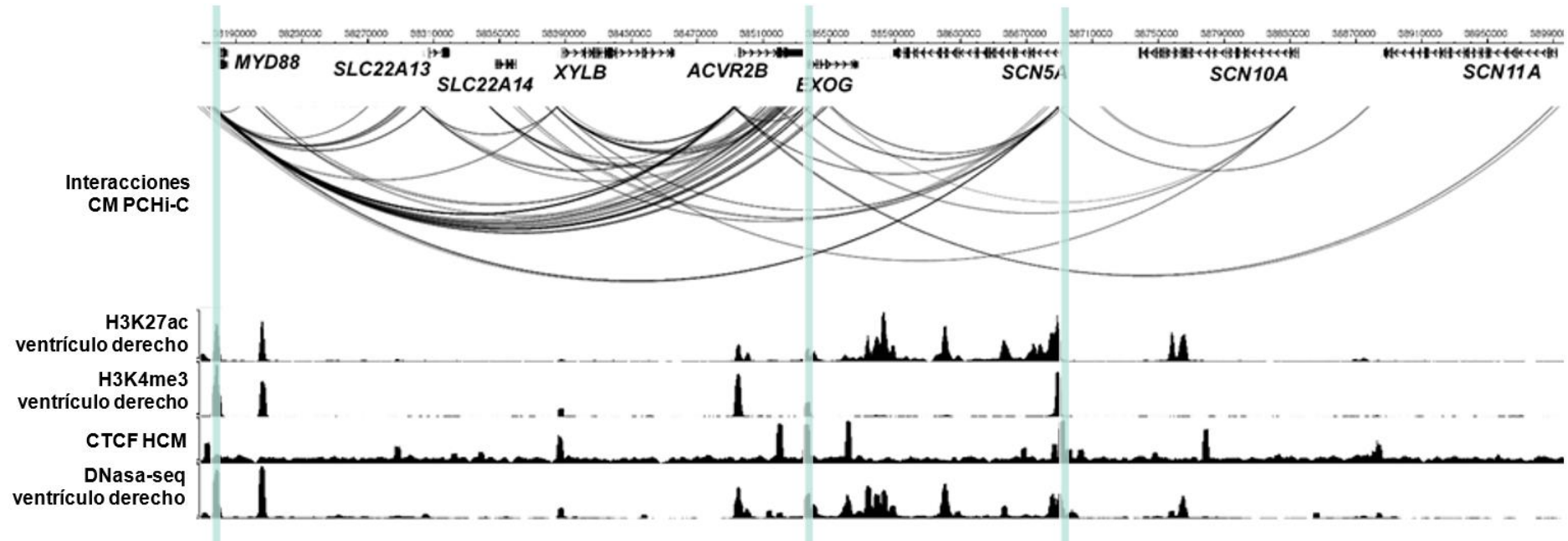
Es importante tener en cuenta que los datos de Tang *et al.* corresponden a un modelo celular no cardíaco (GM12878). Por este motivo, se incluyeron también en el análisis los datos del estudio de Montefiori *et al.*, donde capturaron interacciones en hiPSC-CMs mediante la técnica PCHi-C (*promotore capture Hi-C*)<sup>431</sup>. Esta técnica, basada en la técnica Hi-C, aumenta en gran medida la potencia para detectar interacciones que involucran secuencias promotoras<sup>217,218</sup>. En la bibliografía, la técnica de PCHi-C ha permitido identificar miles de contactos *enhancer-promotor* en diferentes tipos de células y ha revelado amplias diferencias en la arquitectura de los promotores entre tipos celulares y durante la diferenciación<sup>217,218,432-435</sup>. Estos estudios han demostrado colectivamente que la arquitectura del genoma refleja la identidad celular, lo que sugiere que la tipología celular es fundamental para interrogar con éxito los mecanismos reguladores asociados a una enfermedad.

El estudio llevado a cabo por Montefiori *et al.*, resultó apropiado para relacionar estos datos con los resultados obtenidos en nuestro modelo de hiPSC-CM. En relación a las interacciones entre el promotor de *EXOG* y las regiones en 5'*SCN5A*, nuestros datos confirman las observaciones de Montefiori *et al.*, con regiones de interacción prácticamente idénticas a las descritas anteriormente con el estudio de Tang *et al.*, y de nuevo integrando las dianas de las gRNAs usadas en este estudio (**Figura 91**).

Los datos de Montefiori *et al.* muestran, además, que existe otra región de interacción definida en la región promotora de *EXOG* que interactúa con el gen distal *MYD88*, también en su región promotora (**Figura 92**). También se observan marcas de H3K27ac, H3K4me3 y DNasa I, indicando que se trataría de una región con actividad *enhancer*, además de promotora. Esta nueva región de interacción en el promotor de *EXOG* (chr3:38537476-38539120) con *MYD88*, está inmediatamente adyacente a la definida anteriormente con interacción hacia 5'*SCN5A*. Sin embargo, esta nueva región continúa incluyendo el lugar de unión de la gRNA para la línea ***EXOG-CTCF*** (chr3:38537735) (apartado 3 de la Discusión).



**Figura 91. Análisis de las interacciones captadas por PChi-C (hiPSC-CMs) y ChIA-PET (GM12878) en la región *SCN5A-SCN10A*.** En verde se muestran la región promotora de *EXOG* y la región 5' *SCN5A* con la cual interactúa. En rojo se muestran las gRNA cuya diana coincide con una región de interacción. También muestran marcas de H3K27ac, H3K4me3 y DNase I procedentes de ventrículo derecho. Los datos de CTCF provienen de datos de cardiomiocitos humanos. Figura extraída del *WashU Epigenome Browser* a partir de los datos de Tang *et al*<sup>430</sup>. y Montefiori *et al*<sup>431</sup>.

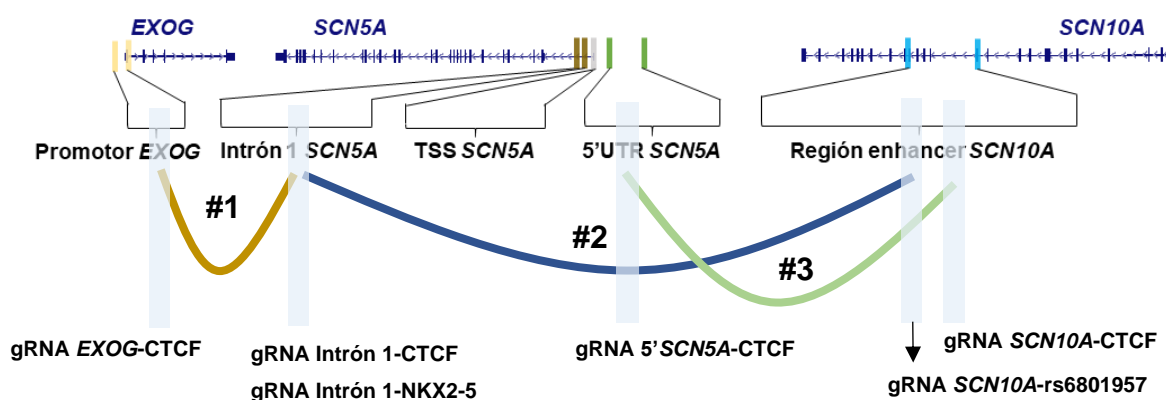


**Figura 92.** Vista global de las interacciones en el TAD de *SCN5A* captadas por PChi-C en hiPSC-CM. En verde se muestran la región promotora de los genes *MYD88*, *EXOG* y *SCN5A*. Se muestran marcas de H3K27ac, H3K4me3 y DNase-seq procedentes de ventrículo derecho. Los datos de CTCF provienen de datos de cardiomiocitos humanos. Figura extraída del *WashU Epigenome Browser* a partir de los datos de Montefiori *et al*<sup>431</sup>.

Además de los datos obtenidos a partir de los estudios de Tang *et al.* (ChIA-PET) y Montefiori *et al.* (PChI-C), se estudiaron los datos disponibles en el repositorio 4DGenome (<https://4dgenome.research.chop.edu>). 4DGenome es una base de datos pública que archiva y difunde datos de los estudios de interacción de la cromatina. Las interacciones más relevantes en relación a los resultados de esta tesis provienen de Jin *et al.*<sup>444</sup>, el cual generó un mapa de interacciones de la cromatina mediante experimentos Hi-C en fibroblastos humanos (**Tabla 18**, **Figura 93**).

**Tabla 18.** Interacciones de Hi-C entre regiones CRISPRi del locus *SCN5A-SCN10A*.

Indicador	Interacción 1	Interacción 2	Referencia
#1	chr3:38536968-38538518 Región gRNA <i>EXOG</i> -CTCF	chr3:38687944-38690219 Región gRNA Intrón 1-CTCF e Intrón 1-NKX2-5	
#2	chr3:38769962-38773440 Región gRNA <i>SCN10A</i> -rs6801957	chr3:38688215-38694080 Región gRNA Intrón 1-CTCF e Intrón 1-NKX2-5	Jin <i>et al.</i> <sup>444</sup>
#3	chr3:38780216-38786079 Región gRNA <i>SCN10A</i> -CTCF	chr3:38690220-38698069 Región gRNA 5' <i>SCN5A</i> -CTCF	



**Figura 93.** Representación esquemática de las interacciones de Hi-C indicadas en la **Tabla 18**, obtenidas a partir de fibroblastos humanos. La figura muestra las 3 interacciones (#1, #2 y #3). Los rectángulos azules muestran las regiones de interacción y las gRNAs que se dirigen a esas regiones.

La región a la que se dirige la gRNA de ***EXOG*-CTCF** interacciona con la región a la que se dirige tanto la gRNA de **Intrón 1-CTCF** como la de **Intrón 1-NKX2-5**, aunque no lo hace con ninguna región en *SCN10A*. Por el contrario, dos regiones intrónicas de *SCN10A* presentan interacciones (apartado 3 de la Discusión). La región a la que se dirige la gRNA de ***SCN10A*-rs6801957** interacciona con las dos regiones a las que se dirigen las gRNAs de las líneas **Intrón 1 (CTCF y NKX2-5)**. La región a la que se dirige la gRNA de ***SCN10A*-CTCF** solamente

interacciona con la región de la gRNA de **5' SCN5A-CTCF**. Esta última interacción CTCF-CTCF no se observó en los datos de Tang *et al*<sup>30</sup>. En el apartado 3 de la Discusión se discute con mayor detalle la relación de las interacciones descritas en este apartado con nuestros resultados y la bibliografía disponible.



## CAPÍTULO III. Estudio del haplotipo en el locus *SCN5A-SCN10A* asociado al Síndrome de Brugada

### 3.1. Resultados previos a esta tesis y contexto del estudio

Tal y como se ha comentado previamente, nuestro grupo ha identificado recientemente un haplotipo de riesgo asociado al SBr<sup>351</sup> (ver apartado 7.1.2.1. de la Introducción). De las siete variantes comunes que conforman el haplotipo (rs1-rs7; **Figura 94A**), sólo tres habían sido asociadas con anterioridad al SBr mediante GWAS (rs9874633, rs10428132 y rs10428168)<sup>346</sup>. De los cuatro SNVs noveles restantes, rs6801957 corresponde al SNV líder (la variante con mayor nivel de asociación) del haplotipo. Esta variante también se había asociado a alteraciones en la expresión de *SCN5A*<sup>51</sup>. Las SNVs del haplotipo se compararon con controles independientes de población europea. En particular, se comparó la cohorte de SBr (n=86) con 3 grupos de individuos sanos:

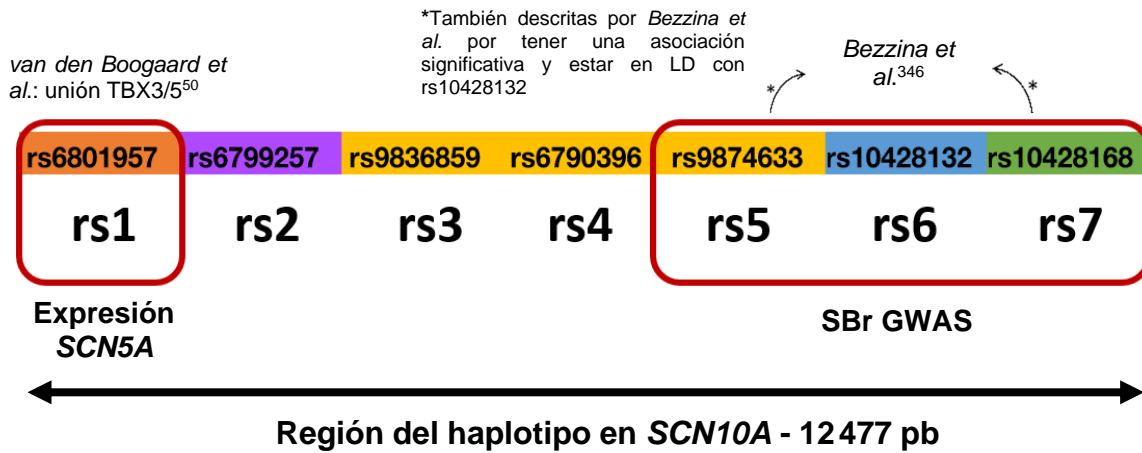
- a) n=404 individuos del proyecto 1000 Genomas (1KG).
- b) n=196 individuos de una cohorte de envejecimiento saludable conocida como Welllderly<sup>389</sup>.
- c) n=355 individuos de la base de datos GTEx (*Genotype-Tissue Expression*).

Suponiendo permutaciones de las siete SNVs, existen n=128 combinaciones haplotípicas posibles. Sin embargo, debido a que rs1-7 están en fuerte LD solo se esperaba un conjunto de 5 o 6 combinaciones comunes.

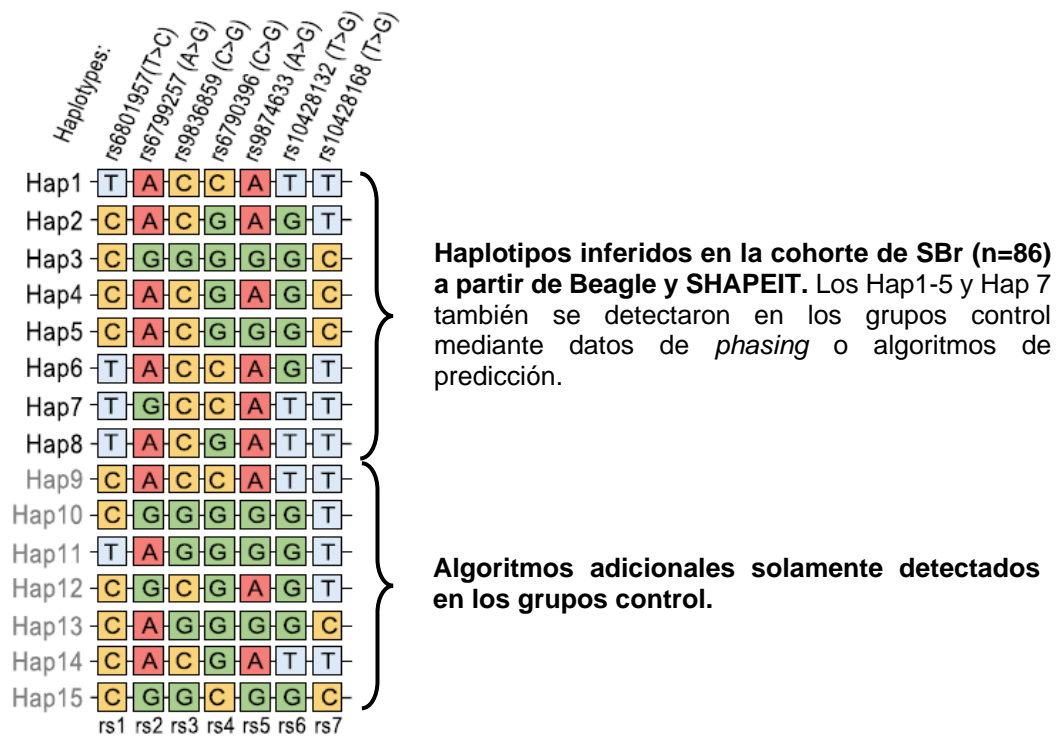
Los haplotipos de la cohorte de SBr identificados a partir de los datos de secuenciación de Illumina solo pueden considerarse predicciones con algoritmos, ya que las lecturas son demasiado cortas para determinar empíricamente el *phasing* de los haplotipos, donde las 7 SNVs se distribuyen a lo largo de una región de 12477 pb (**Figura 94A**). Concretamente, para poder inferir la diversidad de haplotipos presente en la cohorte de SBr se usaron algoritmos bioinformáticos de *phasing*, concretamente Beagle<sup>413</sup> y SHAPEIT<sup>414</sup>. Estos algoritmos de *phasing* utilizan un enfoque basado en el modelo estadístico *Hidden Markov Model* (HMM) para estimar los haplotipos de un individuo en función de los datos genotípicos de una población. Ambos algoritmos infirieron 8 haplotipos (Hap 1-8) en la cohorte de SBr (**Figura 94B**).

Asimismo, se caracterizó la composición haplotípica en los grupos control. Para los datos de 1KG y GTEx, la información de *phasing* ya estaba disponible (obtenida a partir de 1KG *Phase 3 datasets* en <https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502> i GTEx *dataset*<sup>415</sup>, respectivamente). Para el grupo control Welllderly, los haplotipos se infirieron utilizando los algoritmos de predicción Beagle y SHAPEIT. Añadiendo los tres grupos control, se pudieron inferir los Hap1-5 y 7 configuraciones adicionales, Hap 9-15 (**Figura 94B**).

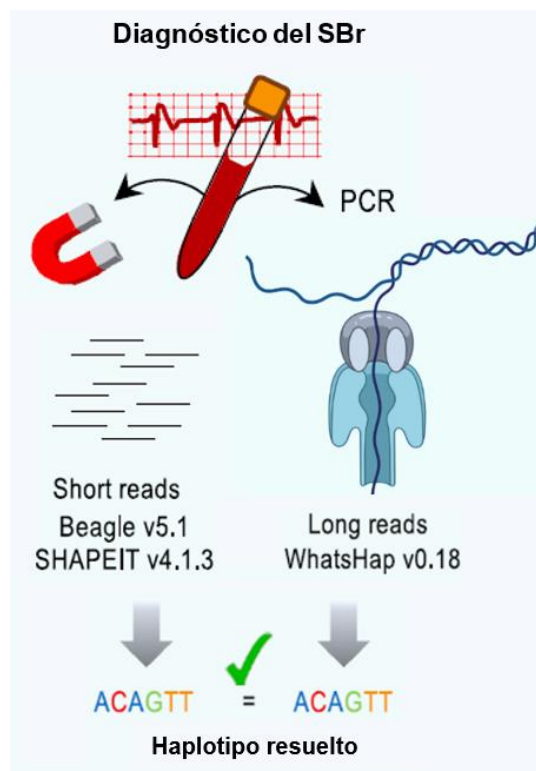
**A**



**B**



**Figura 94. Composición de los haplotipos en el locus SCN5A-SCN10A.** (A) Variantes que conforman los haplotipos, las cuales se distribuyen en el intrón X del gen SCN10A, a lo largo de 12 477 pb. (B) Conjunto de los 15 haplotipos inferidos con los algoritmos Beagle y SHAPEIT. Hap1-8 fueron inferidos para la cohorte de SBr. Hap9-15 fueron solamente inferidos en los grupos control.



**Figura 95. Método de comprobación experimental de los haplotipos mediante secuenciación con ONT.** Los haplotipos obtenidos a partir de ONT se compararon con los predichos por Beagle y SHAPEIT. En caso de coincidir, los haplotipos se consideraron resueltos.

En base a los resultados preliminares, en esta tesis se pretendía confirmar experimentalmente la predicción de los haplotipos en la cohorte de SBr utilizando la tecnología de secuenciación Oxford Nanopore. Utilizando esta tecnología se obtienen lecturas largas que permiten el *phasing* de los haplotipos. Para este procedimiento, se amplificó por PCR un fragmento de DNA de 13 kb de longitud que contiene rs1-7 para cada uno de los  $n = 86$  casos de SBr (**Figura 95**). Después de la secuenciación con ONT y posterior análisis de los datos, se determinaron de los haplotipos mediante el algoritmo *WhatsHap*. Aquellos haplotipos que coincidieron con las predicciones de los algoritmos de *phasing*, se consideraron validados (**Figura 95**).

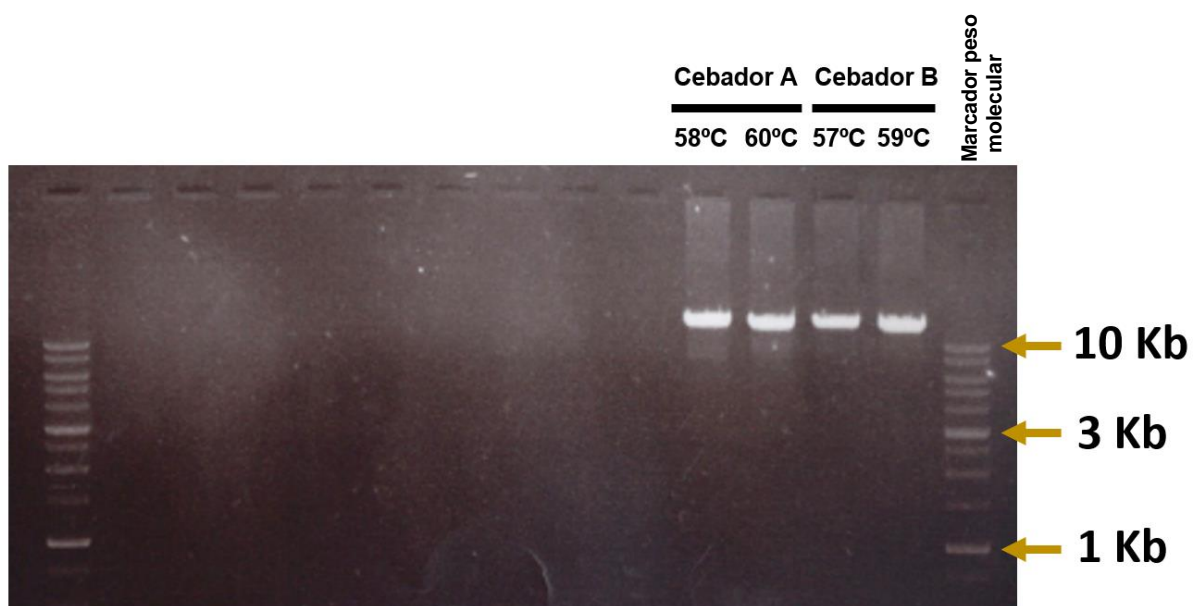
Una vez analizados los datos de predicción a partir de la secuenciación Illuminay de los resultados de secuenciación por ONT de manera conjunta, se realizó un análisis de frecuencias haplotípicas comparando la cohorte de SBr con los grupos control (apartado 3.2.2. de Resultados y apartado 3 de la Discusión).

## 3.2. Determinación experimental de los haplotipos con nanopore sequencing

### 3.2.1. Obtención de fragmentos de 13kb a partir de *long-range* PCR

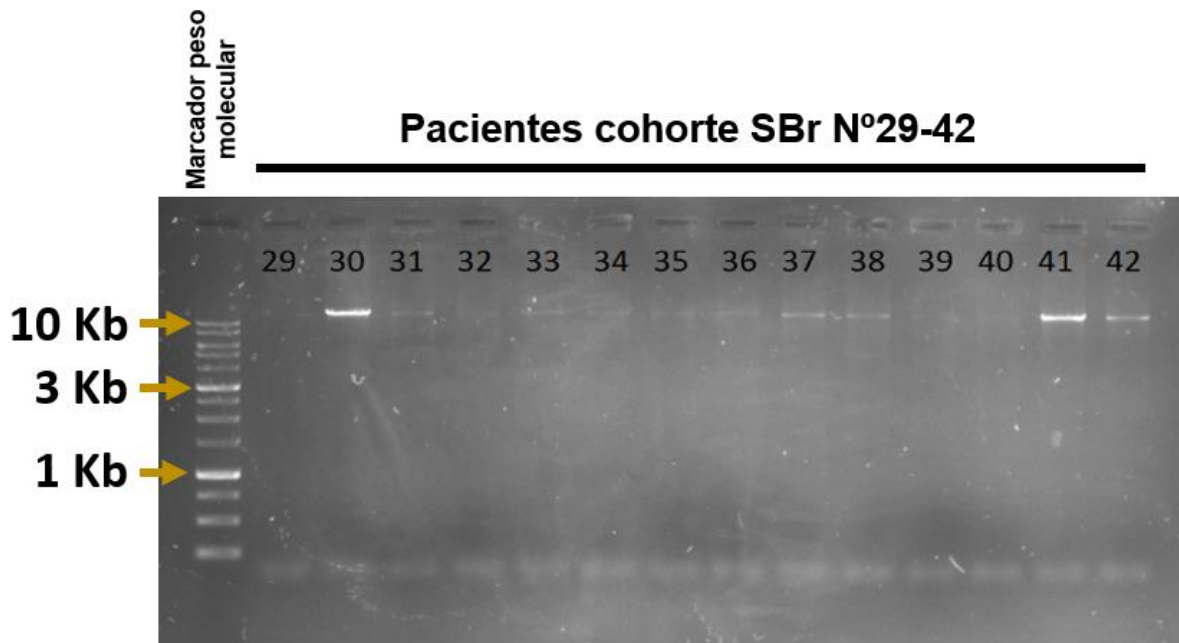
El primer paso para realizar la secuenciación de lecturas largas es la amplificación la secuencia de interés a partir del DNA genómico de cada paciente. En nuestro caso consistió en amplificar, para los 86 casos de SBr, un fragmento con una longitud de 13 kb que contenía los siete SNVs del haplotipo.

Para una PCR común, un fragmento de estas características es inusualmente largo, por lo que se tuvieron que usar kits específicos para PCR de largo alcance (*long-range*). Obtener el fragmento amplificado con la suficiente intensidad y de manera específica fue un proceso complicado. Para poder optimizar la *long-range* PCR se probaron kits de diferentes empresas y se probaron distintas configuraciones en los parámetros de la PCR hasta poder obtener de forma satisfactoria un fragmento correspondiente a la región de interés (**Figura 96**).



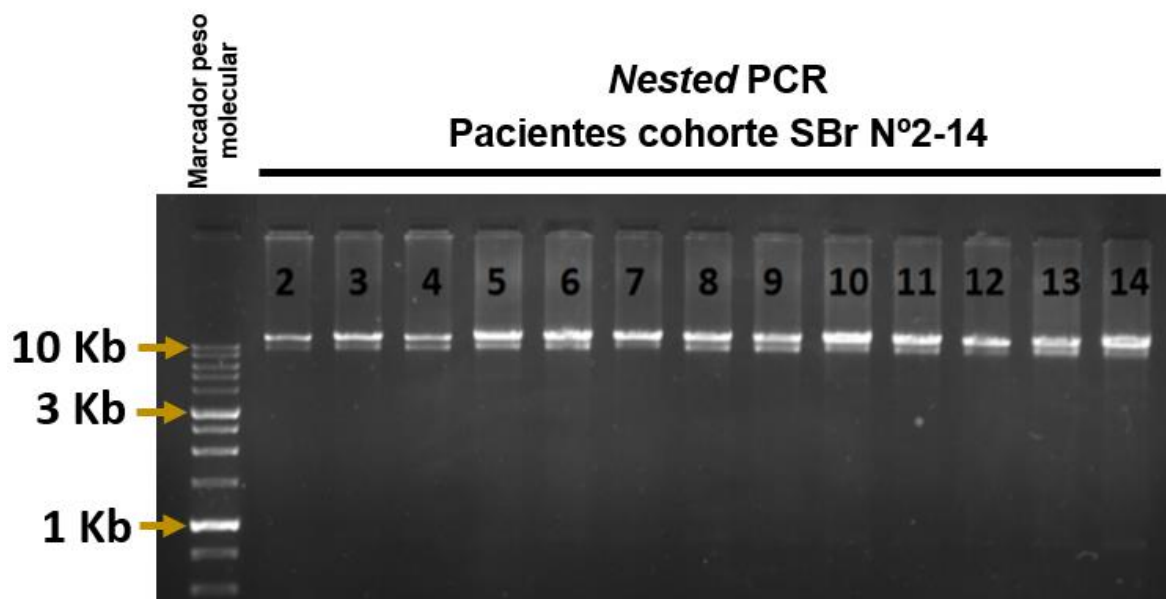
**Figura 96. Amplificación del fragmento de 13 kb correspondiente a la región del enhancer *SCN5A-SCN10A* que alberga los 7 SNVs que configuran el haplotipo.** El fragmento de interés fue amplificado con los cebadores A y B y usando el kit y parámetros indicados en el apartado 2.9.1. de métodos. En este caso la amplificación funcionó con ambos cebadores y dos temperaturas de *melting* distintas (58°C y 60°C para el cebador A; 57°C y 59°C para el cebador B).

Posteriormente, se aplicaron estos parámetros de la PCR a la totalidad de las 86 muestras. Sin embargo, los resultados de las *long-range* PCR fueron heterogéneos, no obteniendo una cantidad suficiente de producto de amplificación en muchas de las muestras (**Figura 97**). No se pudieron determinar las razones de este resultado, pero es posible que el nivel de fragmentación del DNA genómico o la pureza de este pudiera afectar la eficacia de la PCR<sup>416</sup>.



**Figura 97. Amplificación heterogénea del fragmento de 13 kb en diferentes muestras con *long-range* PCR.** Se observa una banda de amplificación en la mayoría de muestras de la figura, aunque muchas de ellas no son lo suficientemente intensas para utilizarlas para secuenciación.

Para solventar este problema, se realizó una *nested* PCR de todas las muestras con una amplificación insuficiente del producto de PCR original. Con la *nested* PCR se consiguió amplificar por segunda vez el producto original de la primera PCR utilizando cebadores que hibridaban con el amplicón original, por lo que se obtuvo un segundo amplicón ligeramente más pequeño (12.8 kb).



**Figura 98. Productos de *nested* PCR amplificados a partir de los productos *long-range* originales.** Todas las muestras presentan una banda clara y visible del producto amplificado, acompañadas de otra banda inespecífica inmediatamente debajo.

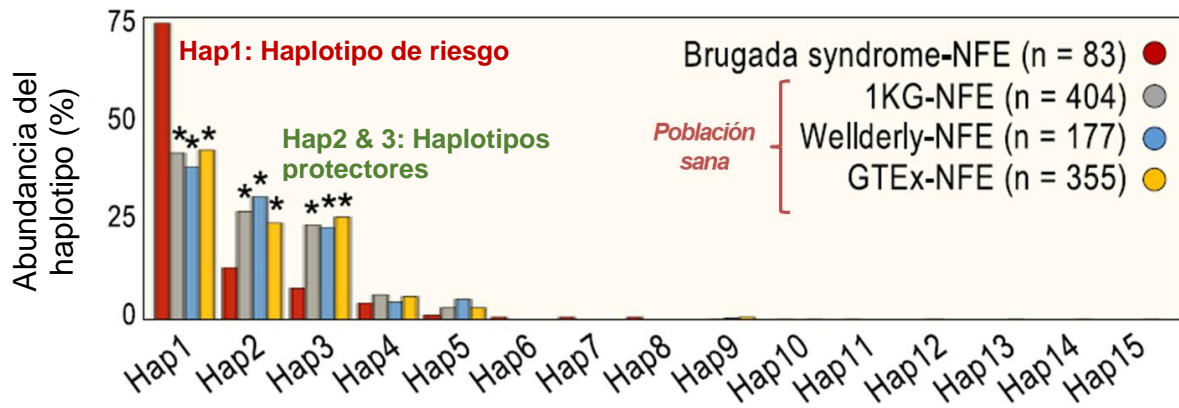
Como resultado, se obtuvieron cantidades de producto amplificado suficientes para realizar el experimento en todas las muestras. Por contra, esta segunda PCR generó cierta cantidad de productos inespecíficos justo debajo de la banda de interés (**Figura 98**). En general, estos productos fueron poco predominantes y se sospechó que probablemente fuesen la propia secuencia de interés en una versión ligeramente más corta o bien compactada de diferente manera, por lo cual aparecería en una altura diferente en el gel. Con todo, se valoró que estas bandas no deberían interferir en la secuenciación del fragmento de interés.

### 3.2.2. Secuenciación y comprobación experimental de los haplotipos

A partir de los productos de amplificación se procedió a realizar la secuenciación con Oxford nanopore siguiendo el protocolo descrito en el apartado 2.9.2. de Métodos. Los resultados de secuenciación se analizaron siguiendo los pasos descritos en el apartado 2.9.3 de Métodos y se utilizó *WhatsHap* para determinar los haplotipos. Utilizando este método, se pudieron secuenciar enteros los fragmentos de 13 kb o 12.8 kb y se validaron empíricamente los haplotipos de 83 de los 86 pacientes de la cohorte de SBr (**Figura 99**). Cabe destacar que las secuenciaciones de lecturas largas, a pesar de permitir el *phasing* de los haplotipos, también generaron otros problemas relacionados con la precisión de los resultados<sup>285</sup>. Su tasa de error es mayor en comparación con las lecturas cortas de Illumina, lo cual se tuvo en cuenta en el análisis de los resultados.

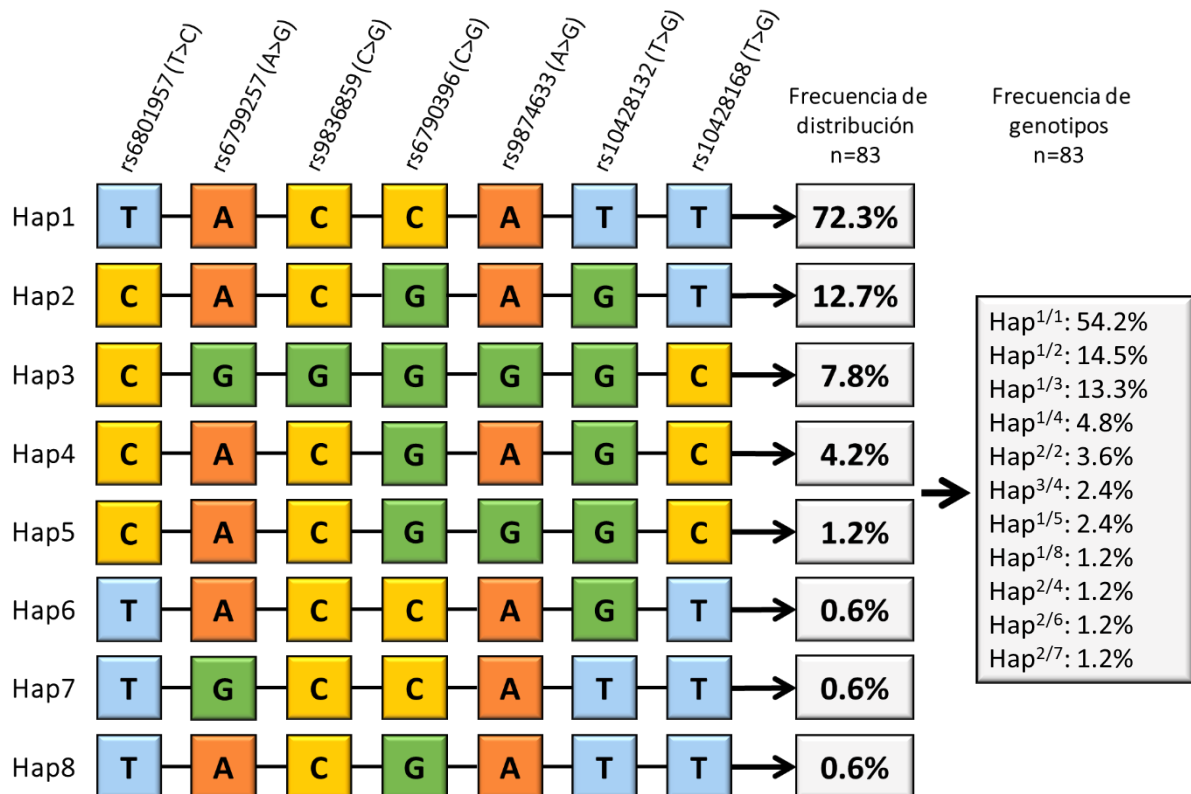
Los resultados de la secuenciación de lecturas largas sólo mostraron discrepancias con las predicciones de Beagle y SHAPEIT en las 3 muestras mencionadas. Estas 3 muestras fueron eliminadas para análisis posteriores, y los cálculos de las frecuencias haplotípicas se realizaron con las 83 muestras restantes. De esta manera, se validaron el 96.5% de los haplotipos predichos. Este resultado demuestra que los algoritmos de predicción del *phasing* funcionan correctamente y que la secuenciación con Oxford nanopore es válida para la dilucidación experimental del *phasing* de los haplotipos.

Los datos obtenidos muestran que Hap1 es el predominante en la cohorte de SBr, con una frecuencia alélica del 72.3%, seguido de Hap2 y Hap3, con unas frecuencias alélicas del 12.3% y 7.8%, respectivamente (**Figura 99**). El resto de haplotipos es minoritario, con frecuencias por debajo del 5%.



**Figura 99. Frecuencias de los haplotipos en la cohorte de SBr y los grupos control.** Se observan diferencias significativas en la abundancia del Hap 1, 2 y 3 entre SBr y controles. Figura adaptada de Pinsach-Abuin *et al.*<sup>351</sup>.

A partir de los resultados obtenidos de las secuenciaciones de lecturas largas y los datos obtenidos previamente de los grupos control, Pinsach-Abuin y colaboradores realizaron el análisis de las frecuencias alélicas de estos datos (**Figura 100**). Hap 1 es significativamente más abundante en la cohorte de SBr que en cualquiera de los tres grupos control, con una frecuencia



**Figura 100. Haplotipos validados experimentalmente mediante secuenciación por Oxford Nanopore.** Composición de los ocho haplotipos validados experimentalmente. Se muestra la frecuencia alélica de cada haplotipo, así como la frecuencia de cada genotipo en la cohorte de SBr.

del 75% de los pacientes con SBr y en menos del 40% en los controles. En cambio, Hap 2 y 3 son significativamente más abundantes en las poblaciones control que en la cohorte de SBr<sup>351</sup>. En el análisis de los datos genotípicos, Hap1 está presente en el 90.4% de las combinaciones existentes (Hap<sup>1/1</sup>, Hap<sup>1/2</sup>, Hap<sup>1/3</sup>, Hap<sup>1/4</sup>, Hap<sup>1/5</sup>, Hap<sup>1/8</sup>) (**Figura 99-100**). Además, el 54.2% del total de pacientes de la cohorte son homocigotos para Hap1 (Hap<sup>1/1</sup>).

Posteriormente, Pinsach-Abuin y colaboradores, mediante modelos de regresión logística y asumiendo diferentes modelos de transmisión hereditaria (recesivo, dominante y multiplicativo), realizaron estimaciones de *odds ratio* (una medida de asociación entre dos variables que indica la fortaleza de la relación entre ellas<sup>391</sup>). Sus resultados sugieren efectos de riesgo y protección para estos haplotipos a nivel genotípico, determinando que:

- El genotipo homocigoto Hap<sup>1/1</sup> está asociado con riesgo a SBr.
- Los heterocigotos Hap<sup>1/-</sup> están asociados con ausencia de riesgo a SBr.
- El genotipo Hap<sup>2/3</sup> está asociado a protección a SBr.

En el apartado de 3 de la Discusión se exponen con mayor detalle los hallazgos derivados de este trabajo<sup>351</sup>.

### 3.3. Diferencias en la expresión global asociadas a un paciente con SBr y Hap<sup>1/1</sup>: experimento de RNA-seq

Ante los resultados del apartado anterior y las conclusiones del estudio de Pinsach-Abuin *et al.*, en esta tesis nos propusimos caracterizar funcionalmente el haplotipo en *SCN5A-SCN10A*. Para ello, uno de los objetivos fue analizar si el haplotipo de riesgo Hap<sup>1/1</sup> tiene algún efecto sobre la expresión génica en los pacientes con SBr. Para estudiarlo, se utilizó un modelo de hiPSC-CMs derivados de un paciente con SBr y haplotipo de riesgo (Hap<sup>1/1</sup>), i como control se utilizaron hiPSC-CMs derivados de un individuo sano con el haplotipo Hap<sup>1/3</sup>.

Se identificó un paciente con SBr de nuestra cohorte de estudio con Hap<sup>1/1</sup>y, como control, se utilizó un individuo sano que en el análisis haplotípico mediante secuenciación por ONT presentó el Hap<sup>1/3</sup>. A ambos individuos se les realizó una extracción de sangre y se purificaron PBMCs. A partir de los PBMCs se generaron hiPSCs mediante reprogramación celular en el Centro de Medicina Regenerativa de Barcelona (CMRB) (apartado 2.2. de Métodos). Se obtuvieron las líneas de hiPSCs Ctrl BS PBiPS37-Sv4F-1 (a partir de ahora **línea GPG2**) y BS PBiPS46-Sv4F-10 (a partir de ahora **línea SBr**) (apartado 1.2.4. de Materiales). Ambas líneas fueron diferenciadas a cardiomiocitos siguiendo el protocolo descrito en el apartado 2.3.4. de Métodos. Se realizaron cuatro rondas de diferenciación a partir de las hiPSC control (línea GPG2) y cuatro rondas de diferenciación a partir de las hiPSC SBr (**Figura 56**; apartado 2.10.1. de Métodos). El



día 30 de diferenciación, se hizo una extracción de RNA, y se midió la pureza obteniendo valores óptimos para realizar el RNA-seq (**Tabla 13**; apartado 2.9.1. de Métodos).

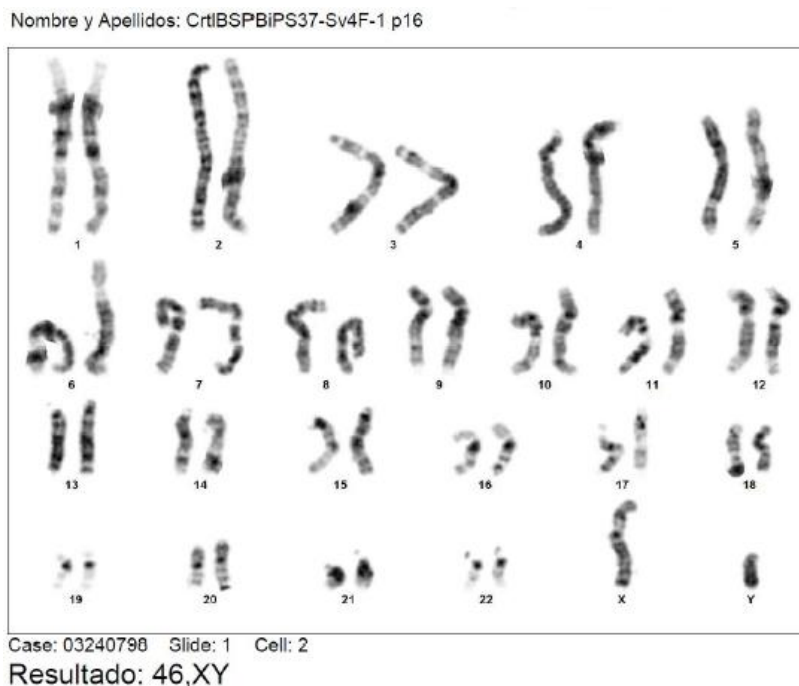
A partir de este RNA, se realizó el análisis de RNA-seq en el Centro de Regulación Genómica (CRG, Barcelona) con un dispositivo HiSeq3000. La secuenciación se realizó en un formato *paired-end* con una longitud de lecturas de 75 pb y con un mínimo de 50 millones de lecturas por muestra (apartado 2.10.1. de Métodos).

Los resultados de este estudio RNA-seq permitieron comparar el perfil de expresión de hiPSCs-CM derivados de un paciente con SBr y con el haplotipo de riesgo con hiPSC-CMs derivados de un individuo control con un haplotipo sin riesgo. El análisis de los datos de RNA-seq, con el objetivo de obtener los genes diferencialmente expresados entre ambas muestras, se realizó a partir del *workflow* descrito en el apartado 2.10.2. de Métodos, en colaboración con Servicio de Supercomputación y Bioinformática de la Universidad de Málaga.

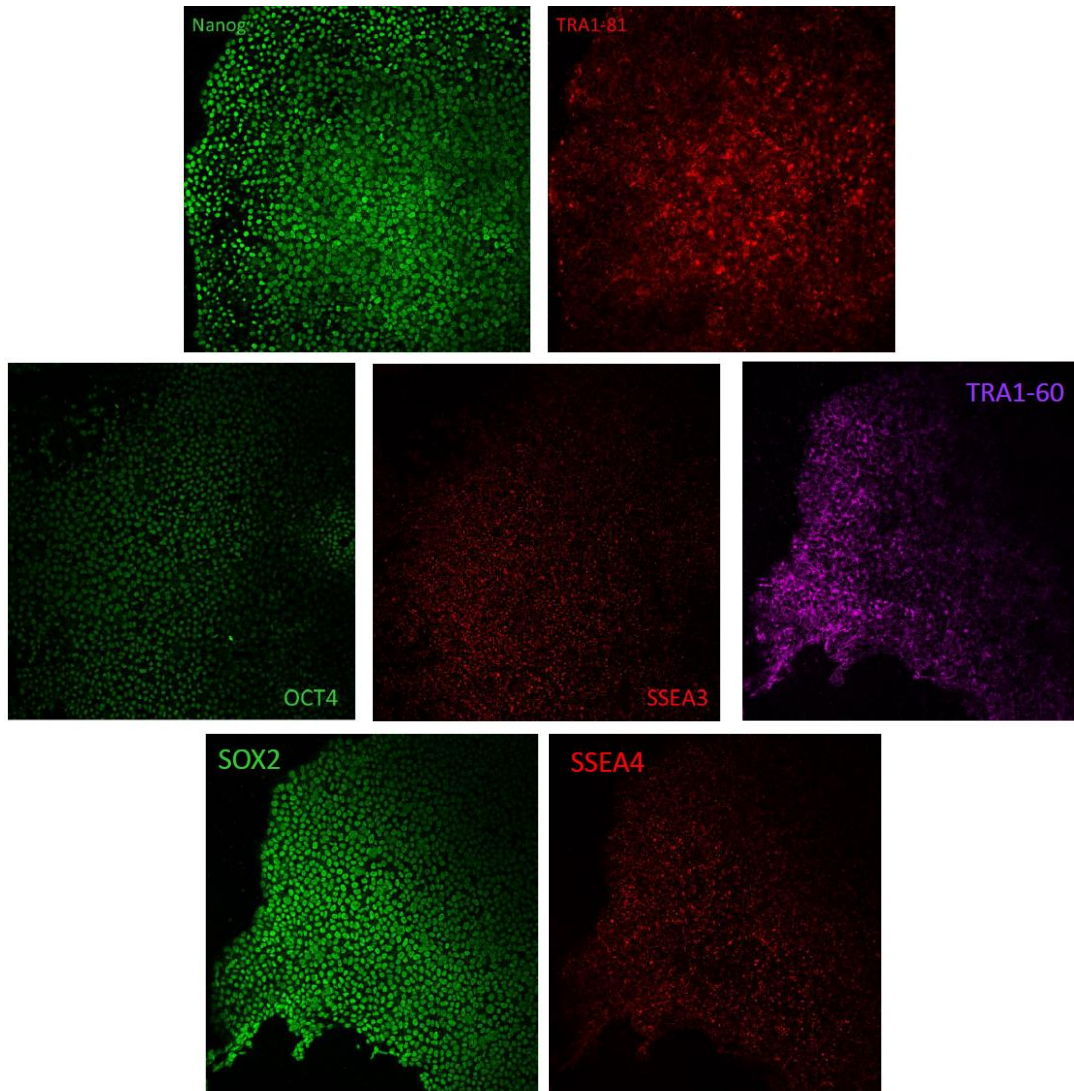
### 3.3.1. Caracterización de las líneas celulares SBr y GPG2

#### 3.3.1.1. Línea GPG2

La caracterización de la línea GPG2 incluye la evaluación del cariotipo. Se observa que no presenta alteraciones y los marcadores de microsatélites coinciden con los de la muestra inicial (PBMCs) (**Figura 101**). También incluye un test de pluripotencia mediante inmunocitoquímica con resultado positivo para todos los marcadores analizados (OCT4, NANOG, SOX2, SSEA3, SSEA4, TRA-1-60, TRA-1-81) (**Figura 102**). Asimismo, se confirmó que línea tiene capacidad

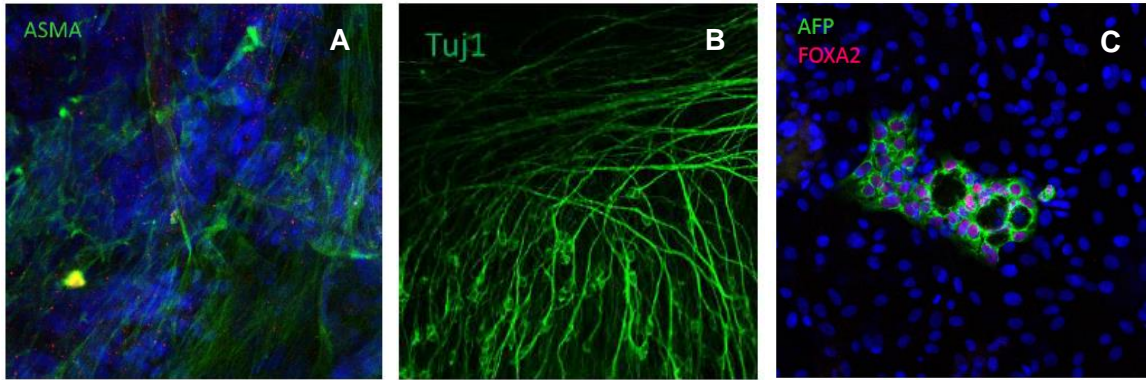


**Figura 101. Cariotipo de la línea celular Ctrl BS PBiPS37-Sv4F-1 (GPG2).** Morfología y cantidad normal de cromosomas (46 cromosomas, siendo XY los cromosomas sexuales).



**Figura 102. Marcadores de pluripotencia de la línea celular Ctrl BS PBiPS37-Sv4F-1 (GPG2).** Resultado positivo por inmunocitoquímica para las proteínas de pluripotencia NANOG, TRA1-81, OCT4 y SSEA3, TRA1-60, SOX2 y SSEA4.

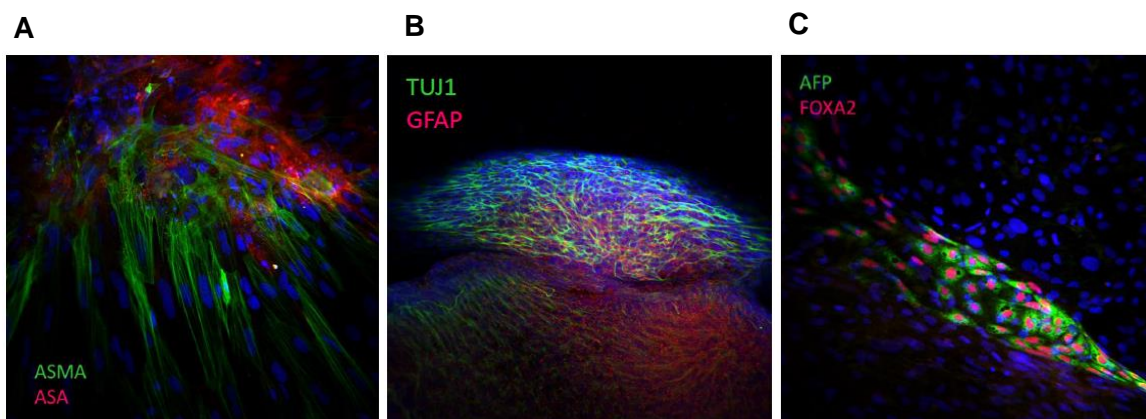
para diferenciarse *in vitro* a mesodermo, ectodermo y endodermo (**Figura 103**). El análisis mediante RT-PCR mostró la ausencia de mRNA derivado de virus Sendai en la línea de hiPSC tras una semana de transducción.



**Figura 103. Diferenciación *in vitro* de la línea GPG2 a endodermo, ectodermo y mesodermo. (A)** Diferenciación a mesodermo, confirmada por la presencia el marcador **ASMA**. **(B)** Diferenciación a ectodermo: Células positivas para **TUJ1** y **GFAP**. **(C)** Diferenciación a endodermo: Células positivas para **AFP** y **FOXA2**.

### 3.3.1.2. Línea SBr

La caracterización de la línea SBr mostró que tiene capacidad para diferenciarse *in vitro* a mesodermo, ectodermo y endodermo (**Figura 104**). Los resultados del análisis también confirmaron que el cariotipo no presenta alteraciones y los marcadores de microsatélites coinciden con los de la muestra inicial (PBMCs) (**Figura 105**). Se confirmó la presencia de las variantes de las células de origen (referidas al haplotipo Hap<sup>1/1</sup>) mediante secuenciación por Sanger (**Figura 106**). El test de pluripotencia mediante inmunocitoquímica (**Figura 107**) resultó positivo para todos los marcadores analizados (OCT4, NANOG, SOX2, SSEA3, SSEA4, TRA-1-60, TRA-1-81). El análisis mediante RT-PCR mostró la ausencia de mRNA derivado de virus Sendai en la línea de hiPSC tras una semana de transducción.



**Figura 104. Diferenciación *in vitro* de la línea SBr a endodermo, ectodermo y mesodermo. (A)** Diferenciación a mesodermo, confirmada por la presencia el marcador **ASMA** y **ASA**. **(B)** Diferenciación a ectodermo: Células positivas para **TUJ1** y **GFAP**. **(C)** Diferenciación a endodermo: Células positivas para **AFP** y **FOXA2**.

Nombre y Apellidos: BS PBiPS46 Sv4F-10 p14

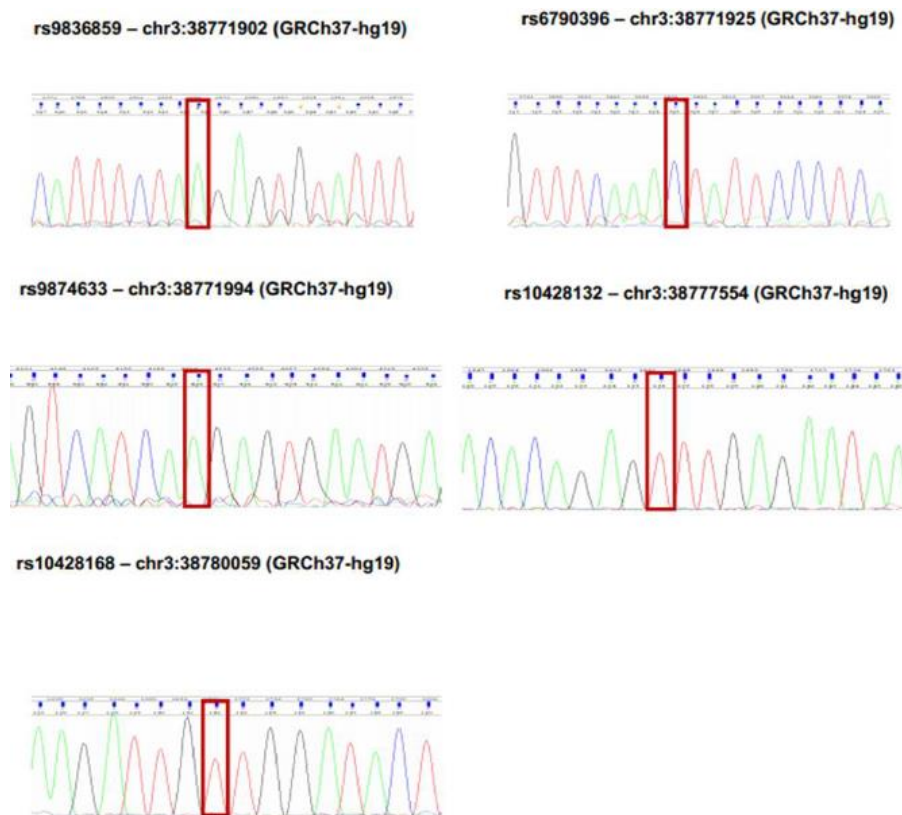
Servicio: CMRB



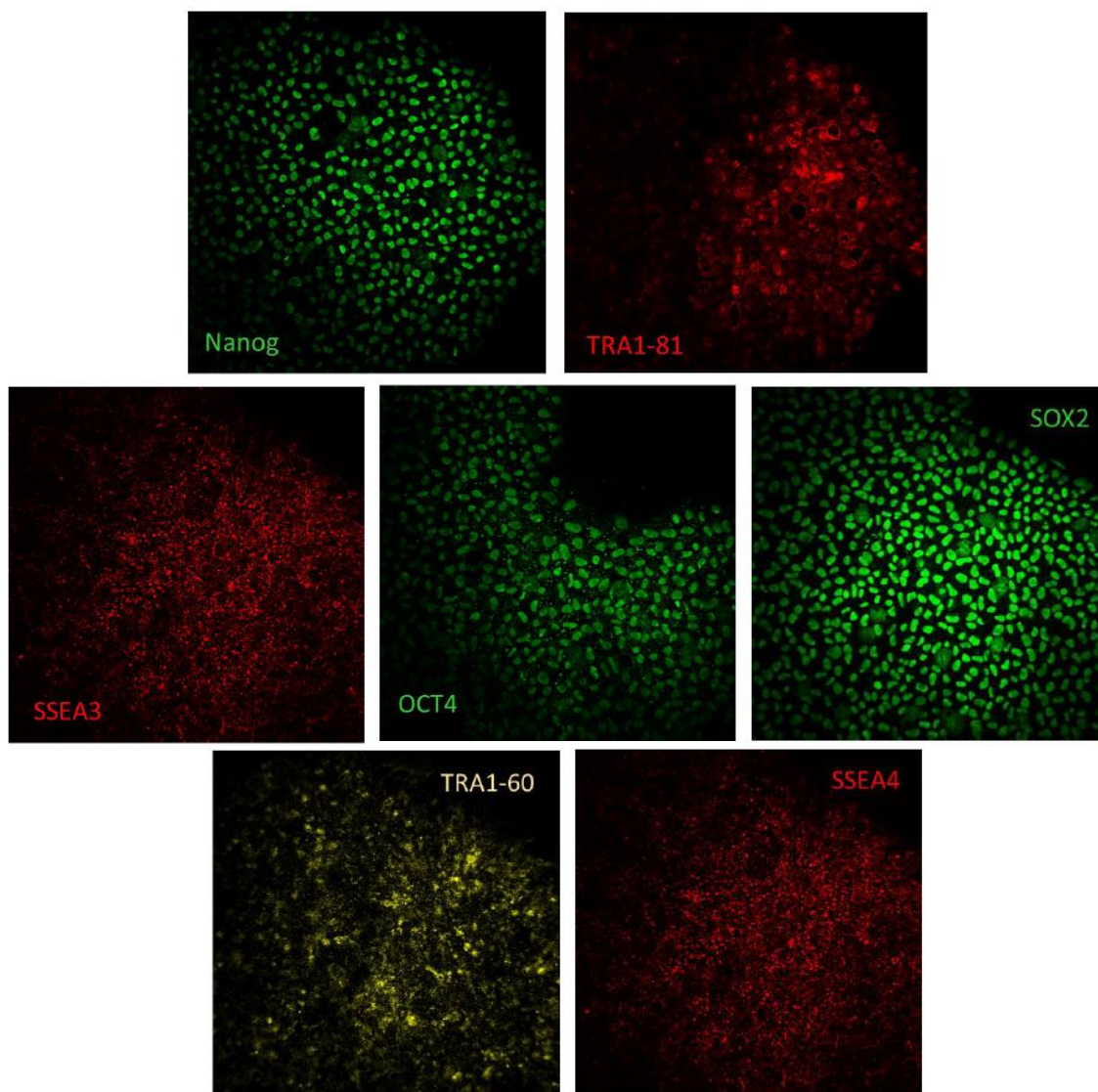
Case: 92410481 Slide: 1 Cell: 6

Resultado: 46,XY

**Figura 105. Cariotipo de la línea celular BS PBiPS46-Sv4F-10 (SBr).** Morfología y cantidad normal de cromosomas (46 cromosomas, siendo XY los cromosomas sexuales).



**Figura 106. Resultados de secuenciación por Sanger para confirmar la presencia de las siete SNVs del haplotipo homocigoto Hap<sup>1/1</sup> en las hiPSC BrS.** En rojo, se indica la posición de cada una de las SNVs en el electroferograma.



**Figura 107. Marcadores de pluripotencia de la línea celular BS PBiPS46-Sv4F-10 (SBr).** Resultado positivo por inmunocitoquímica para las proteínas de pluripotencia NANOG, TRA1-81, OCT4 y SSEA3, TRA1-60, SOX2 y SSEA4.

### 3.3.2. RNA-seq: calidad de las lecturas y el alineamiento

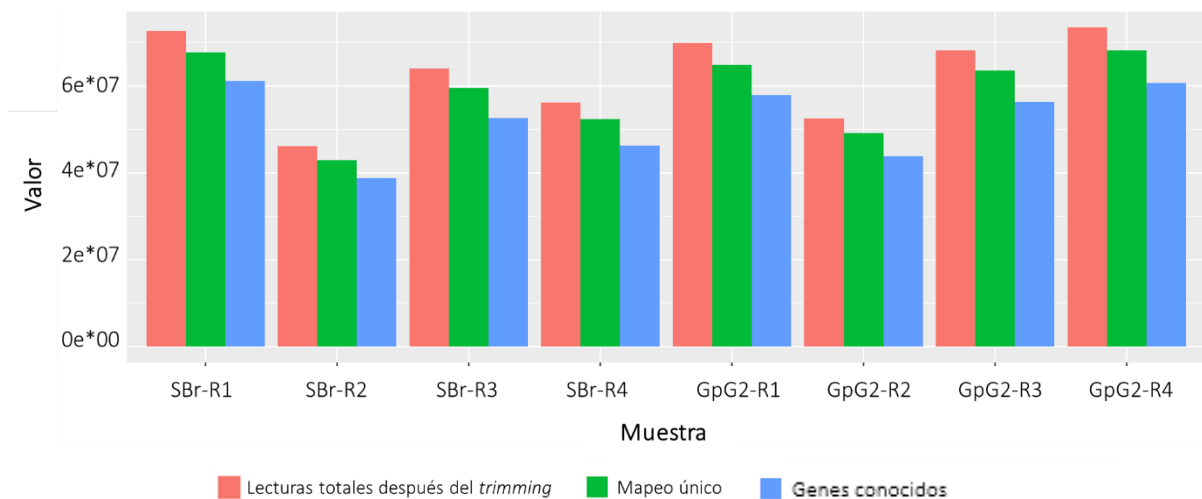
Después de realizarse el RNA-seq, se recibieron los archivos FASTQ con los datos crudos de las lecturas de secuenciación. En un primer paso, los FASTQ se procesaron (*trimming*) para limpiar las muestras de lecturas de poca calidad, contaminantes y artefactos de secuenciación (**Figura 108**).

La calidad de las lecturas se consideró muy buena, ya que después del *trimming* solamente se perdió de media un 5-10% de ellas. A continuación, las lecturas obtenidas después del *trimming* fueron mapeadas con el genoma humano (GRCh38 *assembly*) mediante la herramienta DEGenes Hunter (apartado 2.10.2 de Métodos) para clasificar las lecturas en función del gen al



**Figura 108. Número total de lecturas antes y después del *trimming*.** En el eje de las X se representan las cuatro rondas de diferenciación para las líneas de SBr y GPG2, y en el eje de las Y el número de lecturas. Las barras rojas representan las lecturas antes de procesarse y en azul después del *trimming*. R1-4 se refiere a cada una de las réplicas biológicas (rondas de diferenciación) realizadas en cada grupo.

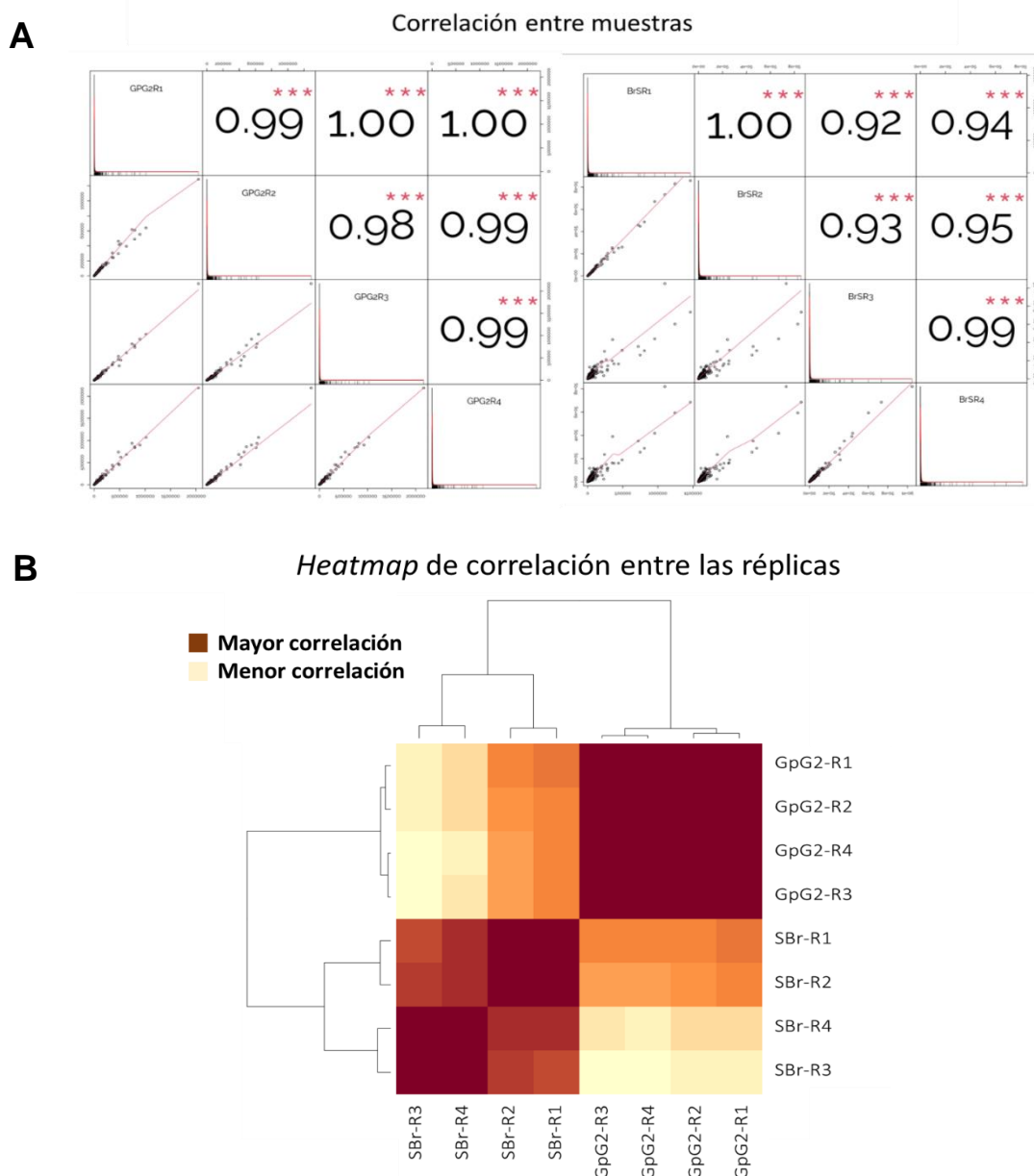
que pertenecen (**Figura 109**). Los resultados del mapeo fueron satisfactorios ya que más del 90% de las lecturas se alinean de manera específica con una única región del genoma. Además, gran parte de estas lecturas están alineadas con una característica genómica concreta (genes conocidos). Por lo tanto, se obtuvo más de un 80% de lecturas caracterizadas genómicamente.



**Figura 109. Número total de lecturas *post-trimming* mapeadas con el genoma humano.** En el eje de las X se representan las rondas de diferenciación para los hiPSC-CMs derivados del paciente con SBr (SBr-R1-4) o control (GPG2-R1-4). En el eje de las Y se representan el número de lecturas. Rojo: número de lecturas después del *trimming*; verde: número de lecturas que solamente se alinean con una secuencia del genoma; azul: número de lecturas que se alinean con genes conocidos.

### 3.3.3. Análisis de la expresión diferencial

Una vez determinados los parámetros de calidad del experimento de RNA-seq, se ejecutó el análisis de expresión diferencial con el programa DEGenesHunter. Este programa permite identificar el conjunto de genes que se expresan diferencialmente y de manera significativa en los tratamientos con respecto a los controles, generando un informe con los resultados. En primer lugar, el informe aportó la correlación entre las muestras analizadas (coeficiente de correlación de Pearson) entre las réplicas de cada grupo (**Figura 110A**). Se observó una correlación positiva y significativa (coeficiente de Pearson 1-0.93) entre las rondas de diferenciación de cada grupo



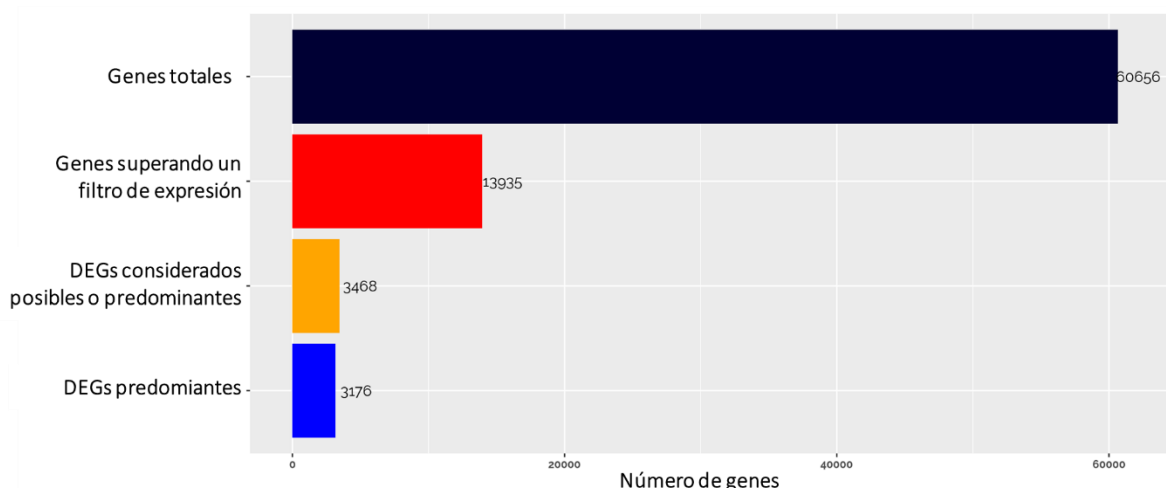
**Figura 110. Correlación entre las muestras analizadas en el RNA-seq.** El análisis se realizó con el programa DEGenes Hunter. **(A)** Coeficiente de correlación de Pearson para las réplicas biológicas de cada grupo de estudio. **(B)** Heatmap mostrando la correlación entre las muestras de GPG2 y SBr. Un color más intenso indica mayor correlación.

(GPG2 o SBr). Este era el resultado deseado, ya que idealmente las réplicas biológicas de un mismo grupo deben ser lo más parecidas posibles para evitar la detección de alteraciones en la expresión debidas a la variabilidad entre réplicas.

En cambio, la comparación entre las réplicas de GPG2 con las de SBr mostró una menor correlación (coeficiente de Pearson de 0.87-0.68) (**Figura 110B**), reafirmando que hay mayores diferencias entre estos dos grupos que entre los replicados biológicos de cada grupo.

El informe creado por DEGenes Hunter mostró el conjunto de genes que se expresan de manera diferencial y de manera significativa en las muestras de SBr con respecto a los controles (**Figura 111**). Se utilizaron cuatro paquetes de detección de expresión diferencial para analizar todos los genes (DESeq2, edgeR, limma y NOISeq) obteniendo una clasificación en base a la cantidad de paquetes que detectan a un gen como diferencialmente expresado:

- **Descartados:** genes descartados durante el proceso de filtrado que muestran una expresión nula o muy baja.
- **DEGs predominantes:** genes considerados diferencialmente expresados (DEGs) por los cuatro paquetes utilizados.
- **Posibles DEGs:** genes considerados DEGs por al menos uno de los paquetes de detección de expresión diferencial.
- **No DEGs:** genes no considerados DEGs en ninguno de los paquetes.

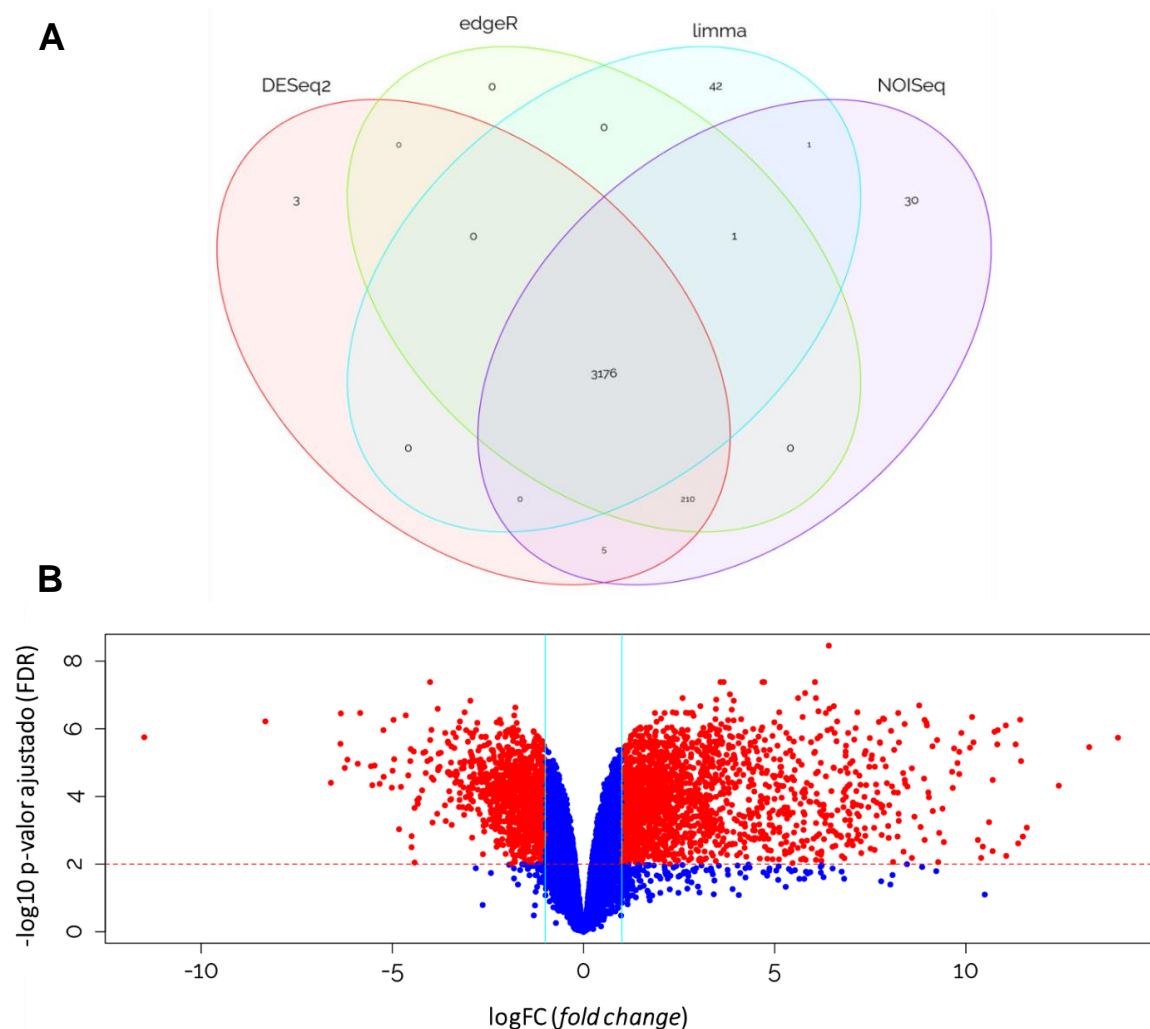


**Figura 111. Clasificación de los genes analizados en el RNA-seq según el tipo de expresión diferencial obtenido.** El número total de genes (60656) representa todos los genes analizados, incluyendo los genes descartados posteriormente por no superar el filtro de expresión (expresión nula o muy baja). Los genes que superaron un filtro de expresión (13935) son aquellos genes en los que se detecta expresión significativa, independientemente de si se expresan diferencialmente o no. Los DEGs posibles (3468) fueron considerados DEGs por al menos un paquete. Los DEGs predominantes fueron considerados DEGs por los cuatro paquetes.



La **Figura 111** muestra el número total de genes que pasaron por cada etapa del análisis. El número total de genes analizados fue de 60656, un número superior al total de genes conocidos en el genoma humano. Esto es debido a que este número también incluye genes de ncRNAs, siendo estos muy numerosos. Una vez se aplica el filtro de expresión integrado en el paquete *ExpHunter Suite*<sup>412</sup>, este número se reduce hasta 13935 genes con expresión significativa. Finalmente, los genes considerados como DEGs fueron un total de 3176, una cifra que sigue siendo difícil de interpretar y muy elevada respecto al número de genes conocidos que están asociados a SBr.

La detección de los DEGs tuvo muy pocas discrepancias entre los cuatro paquetes utilizados con solamente un 8.5% de los genes considerados DEGs por menos de cuatro paquetes (**Figura 112A**). Esto confirma la robustez de los datos, así como la consistencia de los paquetes, aunque utilizan distintos algoritmos.



**Figura 112. Detección de DEGs por los paquetes de análisis DESeq2, edgeR, limma y NOISeq. (A)** Diagrama de Venn mostrando el número de DEGs detectados por cada paquete y de manera combinada. **(B)** Ejemplo de *volcano plot* para el paquete *limma*. En rojo los genes que han superado el filtro de FDR a 0.01. En azul los genes que no se consideraron DEGs al no superar el filtro FDR.

Para seleccionar los genes considerados DEGs en cada paquete, se utilizó un p-valor ajustado o *false discovery rate* (FDR) de 0.01 (**Figura 112B**). El FDR se utilizó para limitar los resultados falsos positivos. En general, cuando se trabaja con cientos y miles de genes, se obtendrán falsos positivos por casualidad, algo que es inherente a la definición de p-valor. Por lo cual, FDR ofrece más confianza que solamente el p-valor. Por lo general, los resultados por debajo de 0.05 se consideran reales, aunque en este caso se optó por poner el límite en 0.01, debido a la gran cantidad de DEGs encontrados.

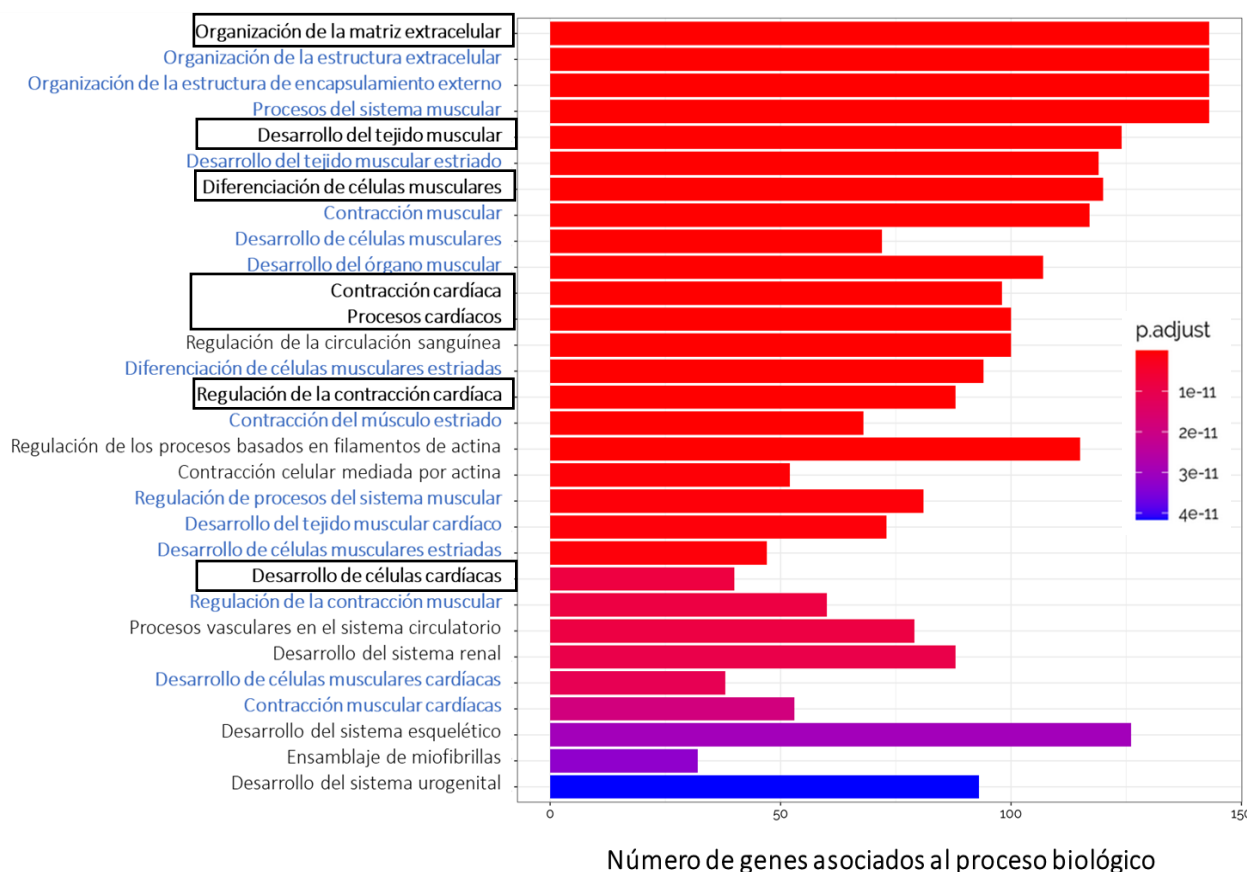
En la comparación entre un individuo con SBr (Hap<sup>1/1</sup>) y un individuo sano (Hap<sup>1/3</sup>) se encontró una cantidad de DEGs que podría estar sobreestimada debido a diversos factores. De los 3176 DEGs, es probable que no todos lo sean a causa de la enfermedad o del haplotipo, sino que factores como la variabilidad interindividual estén presentes en este análisis dado que solamente se ha dispuesto de un individuo para cada grupo de estudio (apartado 3 de la Discusión).

### 3.3.4. Implicación funcional de los genes diferencialmente expresados

Después de obtener los resultados de expresión diferencial, se realizó una predicción funcional de los genes seleccionados. Los DEGs se utilizaron para el análisis de enriquecimiento funcional mediante la detección de grupos de genes sobrerrepresentados entre los DEG en comparación con los que no son DEG, usando la anotación de Gene Ontology (GO) y usando un valor FDR de 0.01. Los términos GO proveen un vocabulario controlado que describe el gen y los atributos del producto génico en cualquier organismo, por ejemplo, su función en los procesos biológicos. Por ello, el uso de estos términos permite clasificar los DEGs en grupos de genes clasificados según sus procesos biológicos.

El informe generado a partir del *workflow* DEGenes Hunter, concretamente ejecutando el “modulo funcional Hunter” (apartado 2.10.2. de Métodos), reveló los procesos biológicos asociados a los DEGs (**Figura 113**). Muchos de los términos que describen los procesos biológicos son redundantes o simplemente subgrupos muy específicos de otro proceso biológico (**Figura 113**). Para entender de manera más simple cuáles son los procesos biológicos mayoritarios en los que se engloban los DEGs de este estudio, se pueden obviar todos los términos redundantes y tomar en cuenta grupos generales. En este caso, se consideraron términos redundantes todos aquellos que se pueden considerar subgrupos de otros términos, como por ejemplo “contracción cardíaca” y “contracción muscular cardíaca”.

Teniendo esto en cuenta, los procesos biológicos asociados a los DEGs más relevantes para el objeto del estudio fueron: organización de la matriz extracelular, desarrollo del tejido muscular, diferenciación de células musculares, contracción cardíaca, procesos cardíacos, regulación de la contracción cardíaca y desarrollo de células cardíacas. Todos estos términos y sus términos redundantes son, además, los procesos biológicos con mayores valores de asociación FDR (**Figura 113**). De acuerdo con la enfermedad de estudio y el modelo celular de hiPSC-CMs, la



**Figura 113. Procesos biológicos GO asociados a los DEGs.** El eje X representa el número de genes asociados a un proceso biológico determinado que se identificaron en los datos de expresión. Los procesos biológicos se clasifican en orden ascendente por el valor FDR. El color representa el valor FDR asociado a cada término (más rojo, mayor asociación). Los términos encuadrados son los seleccionados como más relevantes. Los términos en azul son procesos biológicos redundantes.

mayoría de términos asociados a los DEGs están relacionados con el sistema cardiovascular y el desarrollo celular cardíaco. Otros términos no asociados al sistema cardiovascular, como el sistema renal, circulatorio o urogenital presentan las menores asociaciones de la lista.

### 3.3.5. Similitud semántica asociada a enfermedades cardiovasculares

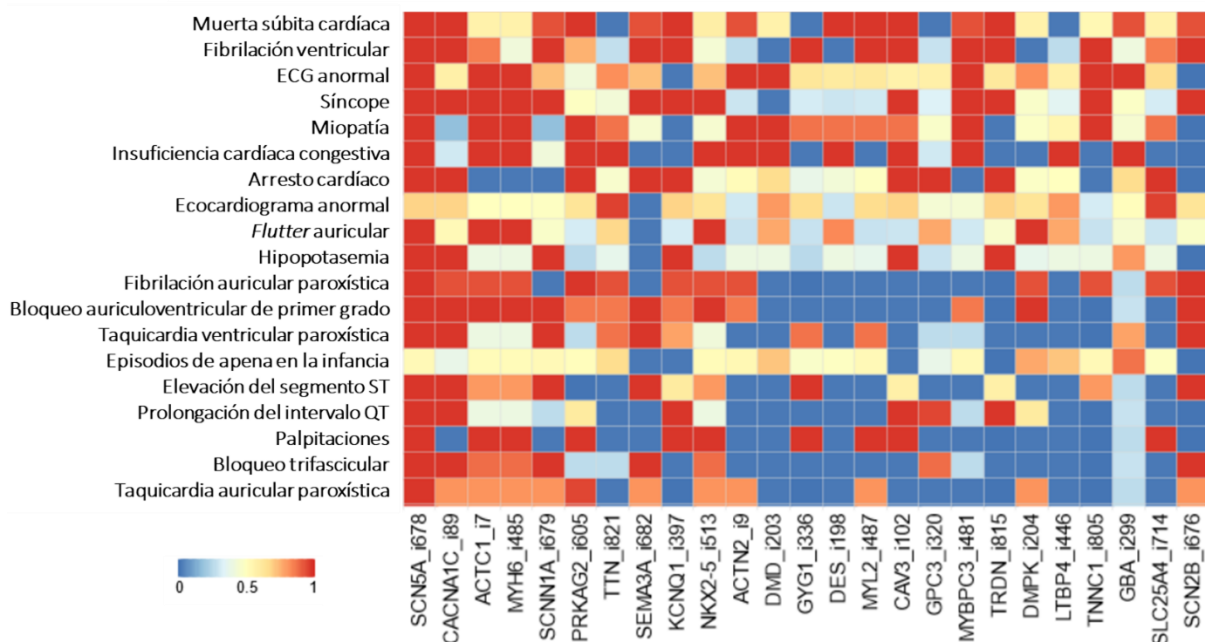
Con los resultados de expresión diferencial y su posterior predicción funcional se obtuvo una visión general sobre cuáles los procesos asociados con el perfil de expresión diferencial entre las muestras de SBr (Hap<sup>1/1</sup>) y el control GPG2 (Hap<sup>1/3</sup>). Sin embargo, debido al elevado número de DEGs, estos análisis resultaron limitados para tener conclusiones determinantes sobre el efecto global de Hap<sup>1/1</sup> en la expresión génica en las células SBr.

Con el objetivo de delimitar los genes que puedan estar relacionados con el SBr y/o los canales iónicos, se realizó un análisis de similitud semántica. Los métodos basados en la semántica

aprovechan las asociaciones de ontologías y el número de genes relacionados con la enfermedad para calcular la similitud con la enfermedad<sup>417,418</sup>.

En primer lugar, y para reducir el número de categorías funcionales y facilitar la visualización de los resultados, se seleccionó una lista de términos HPO (*human phenotype ontology*), que consiste en un vocabulario estandarizado de anomalías fenotípicas en enfermedades humanas. El criterio de selección de estos términos fue que estuvieran relacionados con la enfermedad estudiada. La selección de los términos se realizó en colaboración con el Dr. Moisés Rodríguez del Hospital Universitario de Santiago de Compostela (**Tabla A-1, Anexo 2**). Luego, con la ayuda del grupo de Bioinformática de la Universidad de Málaga, se ejecutó un análisis para calcular la similitud semántica entre las categorías seleccionadas y los resultados del RNA-seq (apartado 2.10.2. de Métodos). En este análisis se computó la similitud semántica de los HPO de cada gen respecto a los HPO de la enfermedad.

Como resultado, se generó una matriz de similitud con aquellos DEGs que están relacionados con los HPO seleccionados. En total, se obtuvo una lista de 859 genes DEGs asociados a los HPO seleccionados, representando un 27% de los 3176 genes iniciales. Los genes con mayor asociación en el análisis de similitud semántica resultaron ser mucho más afines al SBr por distintos motivos: (i) asociación conocida directa o indirecta con SBr (*SCN5A*, *CACNA1C*, *TTN*, *SEMA3A* y *SCNN1A*); (ii) genes cardíacos estructurales (*ACTC1* y *MYH6*), (iii) genes de canales iónicos cardíacos (*KCNQ1*); y (iv) genes de FT claves para la función cardíaca (*NKX2-5*) (**Figura 114**). Es importante remarcar que el objetivo de este análisis fue eliminar, en la medida de lo



**Figura 114. Matriz de similitud semántica de los DEGs en relación con los términos HPO seleccionados.** Se representan los 25 DEGs con mayor asociación en el análisis de similitud. El eje X representa los genes ordenados de mayor (rojo) a menor (azul) similitud (escala del 0 al 1, donde el 1 representa la mayor similitud). El eje Y representa los términos HPO seleccionados. Los términos HPO con similitudes más elevadas se representan en la parte superior.

posible, los DEGs que no tienen relación con SBr o el haplotipo de riesgo. De hecho, el gen *top1* en el análisis fue *SCN5A*, lo cual nos otorga una garantía de que la selección de términos se realizó en la dirección correcta (**Figura 114**).

Entre los 10 genes con mayor similitud del análisis de similitud semántica, el único que aumentó su expresión en el paciente con SBr fue *SCNN1A*, el cuál codifica para para la subunidad  $\alpha$  del canal de sodio epitelial y se ha asociado a SBr<sup>419</sup> (**Tabla 19**). Además del gen principal asociado a SBr, *SCN5A*, cabe destacar *CACNA1C*, *TTN* y *SEMA3A*, presentes en la lista de los 10 genes con mayor similitud (**Tabla 19**). Se han descrito más de 40 genes con posible asociación a SBr, entre ellos *CACNA1C*, *TTN* y *SEMA3A*, aunque las variantes encontradas en esos genes no tienen estudios funcionales convincentes que determinen su asociación con la patogenicidad del SBr en la mayoría de casos<sup>419</sup>.

En resumen, de los 10 DEGs más relevantes después de la selección por similitud semántica, en la mitad de ellos se ha descrito su asociación con SBr. Sin embargo, este es el primer estudio que sugiere que los genes *ACTC1*, *MYH6*, *PRKAG2*, *KCNQ1* y *NKX2-5* estarían asociados con el SBr (apartado 3 de la Discusión).

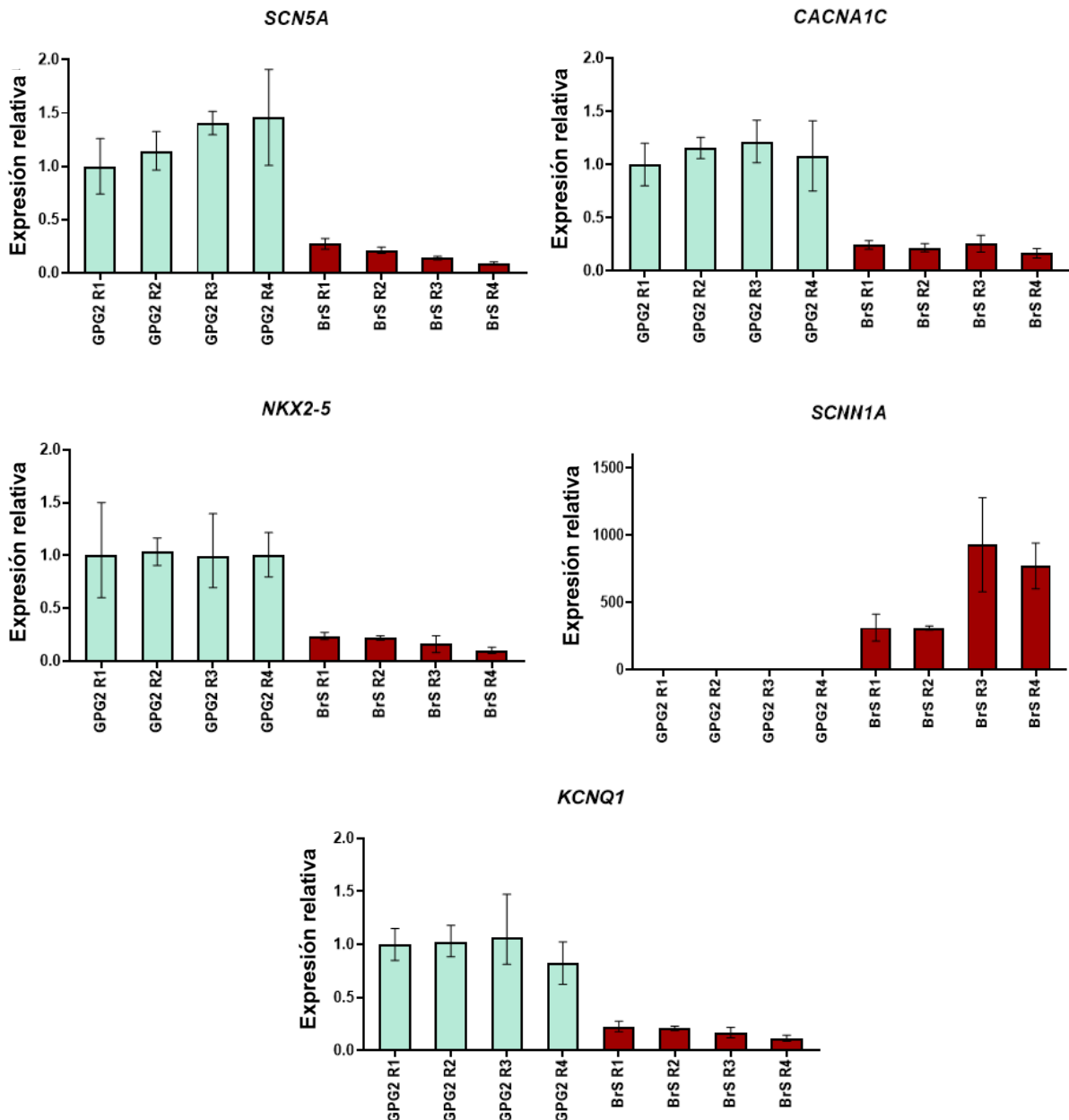
Una cuestión importante en este análisis es si la expresión diferencial de estos genes está relacionada con el SBr o si es resultado de que el paciente tenga el haplotipo homocigoto Hap<sup>1/1</sup>. Para dirimir esta cuestión, sería necesario añadir al estudio pacientes de SBr con o sin él haplotipo de riesgo, o bien editar el haplotipo de la línea SBr, para estudiar el haplotipo en condiciones isogénicas.

**Tabla 19. Expresión diferencial en el RNA-seq de los 10 genes con mayor similitud en el análisis de similitud semántica.** Ordenandos de mayor a menor asociación en el análisis de similitud semántico.

Gen	Valor de similitud semántica (0-1)	Expresión diferencial
<i>SCN5A</i>	0.91	Disminución
<i>CACNA1C</i>	0.79	Disminución
<i>ACTC1</i>	0.77	Disminución
<i>MYH6</i>	0.75	Disminución
<i>SCNN1A</i>	0.74	Aumento
<i>PRKAG2</i>	0.73	Disminución
<i>TTN</i>	0.73	Disminución
<i>SEMA3A</i>	0.73	Disminución
<i>KCNQ1</i>	0.73	Disminución
<i>NKX2-5</i>	0.72	Aumento

### 3.3.6. Validación de los resultados de RNA-seq por qPCR

Con el objetivo de validar los resultados obtenidos en el RNA-seq con otra técnica, se realizó un análisis de expresión por qPCR. Para ello, se seleccionaron algunos de los genes con mayor similitud en el análisis de similitud semántica, concretamente *SCN5A*, *CACNA1C*, *NKX2-5*, *SCNN1A* y *KCNQ1* (**Figura 115**), y como normalizadores *GAPDH* y *PP1A*. Se analizaron las

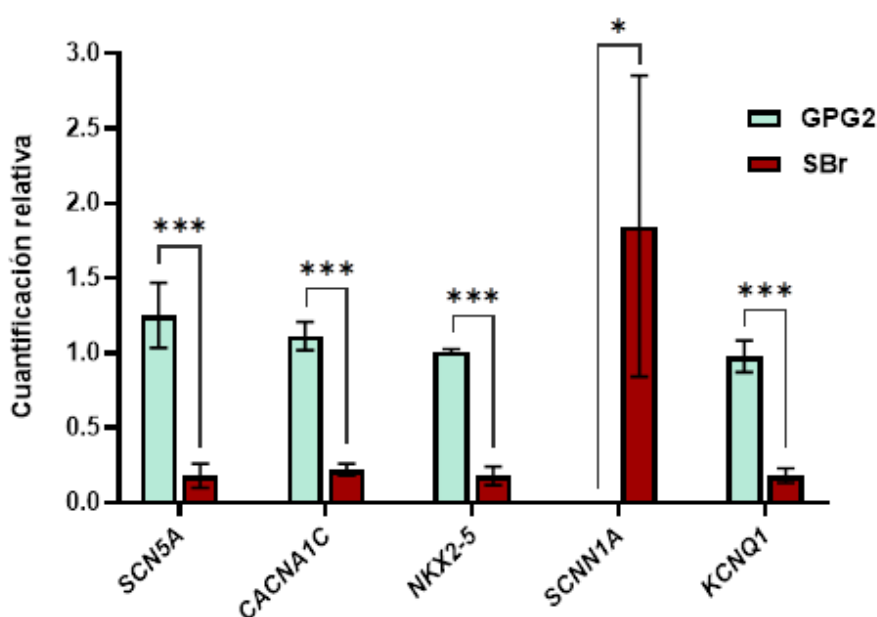


**Figura 115. Validación de los resultados de RNA-seq mediante qPCR.** Análisis de la expresión de los genes *SCN5A*, *CACNA1C*, *NKX2-5*, *SCNN1A* y *KCNQ1* en los hiPSC-CMs de las líneas GPG2 y SBr mediante qPCR utilizando sondas Taqman. Expresión relativa de todas las rondas de diferenciación de cada grupo y para cada gen estudiado. Para relativizar la expresión se utilizaron los genes *housekeeping* *GAPDH* y *PP1A*. Las barras de desviación muestran la SD de los triplicados realizados en las qPCR para cada ronda de diferenciación.

cuatro rondas de diferenciación para las líneas GPG2 y SBr, utilizando el mismo RNA usado en el experimento de RNA-seq.

Los resultados de la qPCR mostraron, en primer lugar, un patrón de expresión similar y bastante homogéneo para las cuatro rondas de diferenciación (**Figura 115**). También se observó, para todos los genes estudiados, la misma tendencia de expresión diferencial observada en el RNA-seq, con una disminución de expresión de los genes *SCN5A*, *CACNA1C*, *NKX2-5* y *KCNQ1* en las células SBr respecto al control, y un aumento de *SCNN1A* en las rondas de SBr. Además, la expresión fue bastante homogénea entre los cuatro replicados biológicos de cada línea celular.

Finalmente, se agruparon los replicados biológicos y se compararon los dos grupos (**Figura 116**). Todas las alteraciones en la expresión génica fueron estadísticamente significativas, confirmando los resultados observados en el análisis de RNA-seq. En los genes *downregulados*, se observó una disminución de la expresión de *SCN5A*, *CACNA1C*, *NKX2-5* y *KCNQ1* en la línea SBr de un 89%, 83%, 85% y 84%, respectivamente. En el caso de *SCNN1A*, se observó un aumento de expresión en la línea SBr de un 99%.



**Figura 116. Validación de los resultados de RNA-seq mediante qPCR.** Análisis de la expresión de los genes *SCN5A*, *CACNA1C*, *NKX2-5*, *SCNN1A* y *KCNQ1* en los hiPSC-CMs de las líneas GPG2 y SBr mediante qPCR utilizando sondas Taqman. Las rondas de diferenciación se representan en un solo grupo biológico (GPG2 o SBr). Los p-valores se calcularon mediante una prueba-t no pareada (\* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ ).

### 3.4. Análisis de la concentración de sodio intracelular en hiPSC-CMs derivados de GPG2 y SBr

Con el fin de estudiar posibles diferencias en la concentración de sodio intracelular entre las dos líneas de GPG2 y SBr, se realizaron experimentos funcionales siguiendo el mismo protocolo descrito en el apartado 2.8.1. de Métodos, con la excepción de los tratamientos con DOX, los cuales no eran necesarios. En total, se pudieron realizar dos rondas de diferenciación independientes (replicados biológicos) para la línea GPG2 y tres rondas para la línea SBr. El sodio intracelular se midió con el indicador fluorescente SBFI-AM en el sistema Operetta y los resultados se normalizaron mediante la cuantificación de núcleos celulares con el indicador fluorescente Hoechst.

Dado que no se pudieron analizar el mismo número de rondas de diferenciación en los dos grupos, se comparó cada ronda individual con todas las demás (**Figura 117A**). No se detectaron diferencias significativas en las mediciones entre las dos rondas de GPG2 ni entre las tres rondas de SBr. Sin embargo, las tres rondas de diferenciación de la línea SBr mostraron una disminución significativa del 36.5% de la concentración de sodio intracelular en comparación con las dos rondas de GPG2.

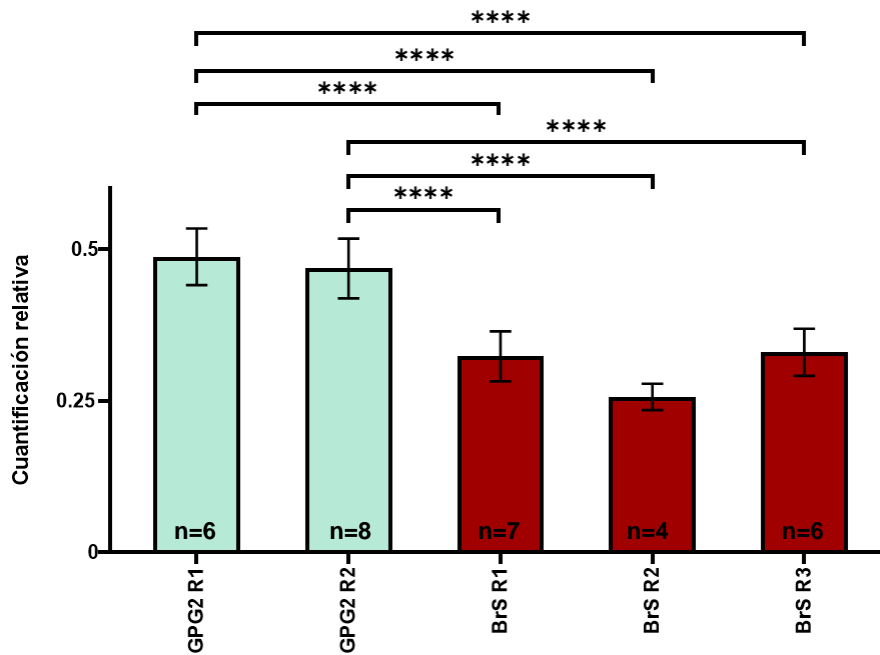
El análisis estadístico se completó con un test ANOVA seguido de una prueba de comparaciones múltiples de Dunnett (**Figura 117B**). Con este análisis solamente se observaron diferencias significativas al comparar las rondas de GPG2 con las de SBr. Este resultado demuestra la robustez de los experimentos y corrobora la observación que los hiPSC-CMs derivados del paciente de SBr y con el haplotipo de riesgo tienen menor concentración de sodio intracelular que los hiPSC-CMs control.

El significado biológico de estos resultados es relevante para apoyar todas las observaciones previas asociadas a estas dos líneas celulares. Detectar la detección de alteraciones en los niveles de sodio intracelular sugiere que las células con *background* genético relacionado con SBr y, en este caso, también con el Hap<sup>1/1</sup>, podrían estar expresando menos cantidad de canal de sodio cardíaco. En consecuencia, la entrada de sodio a la célula sería menor, así como su detección experimental. Esta hipótesis reforzaría el hecho de que *SCN5A* sea uno de los DEGs con mayor relevancia en el análisis de RNA-seq de estas mismas líneas.



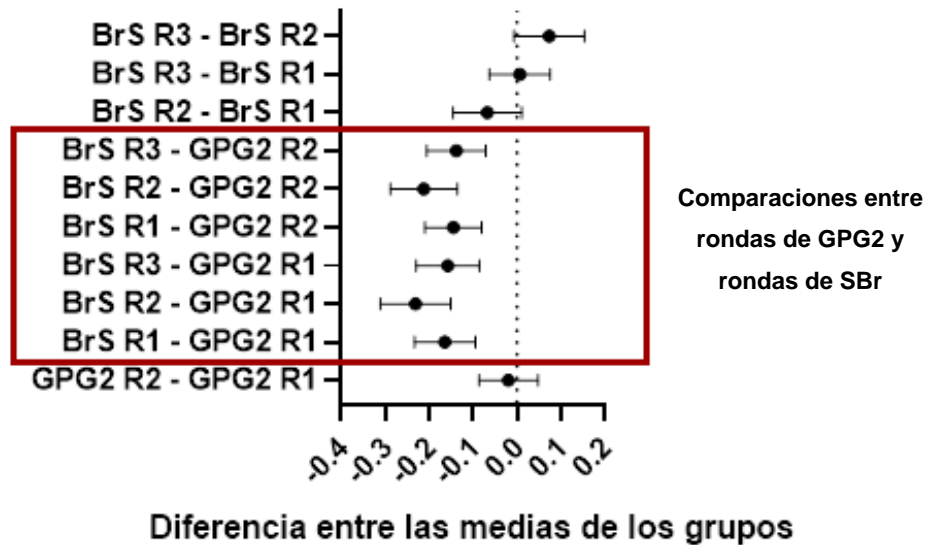
A

## Medición sodio intracelular



B

## Intervalos de confianza



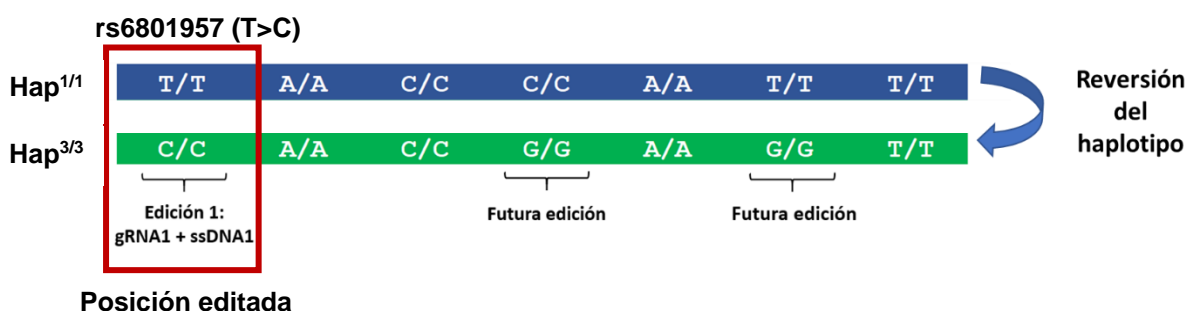
**Figura 117. Determinación de la concentración de sodio intracelular en las líneas celulares GPG2 y SBr.** (A) Cuantificación relativa de la concentración de sodio intracelular. La n representa el número de replicados técnicos en cada ronda de diferenciación. Los p-valores se calcularon mediante una prueba-t no pareada (\*\*\*\* $p \leq 0.0001$ ). (B) Test ANOVA y prueba de comparaciones múltiples de Dunnett. Intervalos de confianza para la diferencia entre medias de todas las comparaciones entre rondas de diferenciación. En rojo, las comparaciones entre rondas de GPG2 y SBr.

### 3.5. Edición del haplotipo por CRISPR-Cas9

#### 3.5.1. Obtención de ediciones HDR homocigotas y heterocigotas con elevada eficiencia

El principal problema que surge al comparar un paciente SBr y haplotipo de riesgo con un control sin haplotipo de riesgo es que la única manera de poder atribuir los resultados obtenidos al haplotipo de riesgo es cambiando el haplotipo. Por lo tanto, para evaluar los efectos de Hap<sup>1/1</sup> en la expresión génica, nos pusimos como objetivo editar el haplotipo de riesgo para cambiarlo a un haplotipo protector (Hap<sup>1/1</sup> → Hap<sup>3/3</sup>). De esta manera, podríamos estudiar líneas isogénicas con un *background* genéticamente idéntico en las que la única diferencia sería la composición del haplotipo, con lo que las alteraciones que pudiesen observarse entre las dos líneas serían explicadas por el haplotipo modificado.

Hap1 y Hap3 sólo difieren en tres posiciones (rs1, rs4, rs6). En consecuencia, para pasar de un genotipo Hap<sup>1/1</sup> a un genotipo Hap<sup>3/3</sup> es necesario añadir tres cambios del total de las siete posiciones que conforman el haplotipo. En esta tesis se realizó el cambio correspondiente a la primera posición del haplotipo (rs6801957) en las hiPSCs de la línea de SBr, que implicó un cambio de una timina por una citosina en esa posición (**Figura 118**). Además, se logró realizar el cambio en los dos alelos (edición homocigota).



**Figura 118. Composición de los genotipos Hap<sup>1/1</sup> y Hap<sup>3/3</sup> del locus SCN5A-SCN10A.** La posición editada corresponde a la variante rs6801957.

La consecución de ediciones de nucleótidos puntuales o *knock-in* con buenas eficiencias es un proceso complejo, especialmente si la edición se hace en hiPSCs, en las cuales la eficiencia y supervivencia al proceso es aún menor<sup>427</sup>. Por este motivo, los esfuerzos en esta tesis se focalizaron en optimizar el protocolo de CRISPR-Cas9 para que sea eficiente, con el objetivo de evitar la secuenciación de cientos de clones para obtener un solo clon editado. Esto fue especialmente relevante teniendo en cuenta que para cambiar el haplotipo se deben hacer diversas ediciones de manera secuencial. El hecho de que las variantes estén distribuidas en una región de más de 12 kb dificulta el protocolo de edición. Las variantes están demasiado lejos

como para editar el haplotipo con una inserción, y dirigir diversas sgRNAs generaría problemas de toxicidad celular. Por todo esto, se optó por realizar las ediciones una a una.

Para realizar la edición genómica con CRISPR-Cas9, se testaron un total de seis estrategias distintas (apartado 2.11.1. de Métodos). Dos de estas estrategias se realizaron con la línea original de SBr (estrategia #1 con HDR *enhancer*, estrategia #2 sin HDR *enhancer*). Las otras cuatro estrategias se realizaron con una versión modificada de la línea de SBr, donde Cas9 se inducía con doxiciclina, evitando tener que nucleofectar la Cas9 (apartado 1.2.5. de Materiales). Las estrategias #3 y #4 incluyeron HDR *enhancer*, diferenciándose la una de la otra según el momento de inducción de la Cas9 (antes o después de la nucleofección). Las estrategias #5 y #6 se realizaron siguiendo la misma aproximación que con #3 y #4, pero sin HDR *enhancer*. Debido al coste experimental de picar colonias aisladas, solamente se realizó secuenciación de colonias aisladas (clones) en dos de las seis estrategias: la #1, línea original de SBr + HDR *enhancer* y #3, línea modificada + HDR *enhancer* y Cas9 inducida antes de la nucleofección. No obstante, se secuenció el *pool* de células editadas de todas las estrategias para calcular el porcentaje de células correctamente editadas (eventos HDR) mediante el programa bioinformático TIDER. De esta forma se pudo evaluar y comparar la eficacia de todas las estrategias sin la necesidad de aislar colonias en todos los casos. El procedimiento seguido con cada línea se detalla a continuación:

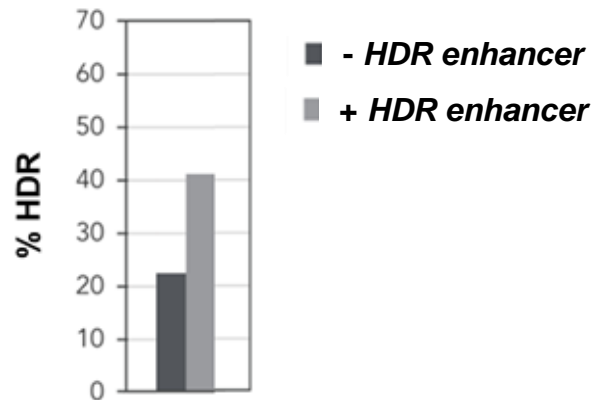
- **Línea original de SBr (#1 y #2):** en la línea original, la edición se realizó con RNPs, la cual tiene diversas ventajas sobre las estrategias de DNA/RNA<sup>421-425</sup>. La función de las RNPs (complejo Cas9/sgRNA) es inmediata al entrar en las células, no teniendo que esperar a los procesos de transcripción y traducción, además de degradarse en poco tiempo, reduciendo la aparición de *off-targets*. Además, el complejo RNP Cas9/sgRNA preformado puede evadir la inhibición por los mRNA celulares<sup>426</sup>. La nucleofección del complejo Cas9/sgRNA debe hacerse junto con el DNA donador (ssODN) para permitir que pueda tener lugar la vía de reparación celular HDR. Las dos estrategias realizadas con este método (#1 y #2) solamente se diferenciaron en que se añadió el producto *HDR enhancer* en una de ellas (#1), con el objetivo de aumentar la frecuencia de edición por la vía HDR. Como se mencionó anteriormente, solamente se aislaron clones para secuenciar en la estrategia #1 (con *HDR enhancer*), bajo la hipótesis de que este producto aumenta las posibilidades de *knock-in*. No obstante, ambas estrategias pudieron ser comparadas a partir de los datos de TIDER (**Figura 119**).
- **Línea modificada de SBr (#3, #4, #5, #6):** la idea de usar la línea modificada de SBr se basó en que la carga de nucleofección es un factor limitante en el proceso de edición. En la estrategia tradicional, sgRNA y ssODN son nucleofectados junto la Cas9, lo cual genera un

compromiso en las cantidades utilizadas de cada componente, pudiendo esto afectar la eficiencia de nucleofección y edición. Por este motivo, el uso de una línea con Cas9 inducible permitió realizar la nucleofección sin necesidad de añadir Cas9. Por contra, fue necesario inducir la expresión de Cas9 con DOX, generando un estrés adicional a las células. Este paso debía hacerse antes o inmediatamente después de la nucleofección. Para validar todos los métodos, se llevaron a cabo las dos opciones, además de la adición/no adición de *HDR enhancer* en ambos casos, dando lugar a las cuatro estrategias usadas en la línea con la Cas9 inducible. Como en el caso anterior, sólo se aislaron clones para secuenciar en la estrategia #3 (*HDR enhancer* + inducción de Cas9 previa a la nucleofección), bajo la hipótesis de que la presencia de Cas9 en el momento de introducir sgRNA y ssODN acelera el proceso de edición. De nuevo, las cuatro estrategias pudieron ser comparadas mediante secuenciación del *pool* y análisis con TIDER.

En general, en las cuatro estrategias con Cas9 inducible se observó una mayor mortalidad celular debido a la inducción con DOX. Además, en ninguna de las estrategias con Cas9 inducible se detectaron ediciones exitosas, ni en las secuenciaciones de los clones de la estrategia #3 ni en los resultados de TIDER en ninguna de las 4 estrategias (#3 a #6). Algunas de las hipótesis que podrían explicar este resultado podrían ser:

- Exceso de mortalidad observado durante el proceso de inducción y nucleofección.
- Toxicidad por parte de DOX, afectando el proceso de edición.
- Falta de optimización de las cantidades de Cas9 inducidas y las cantidades de sgRNA + ssODN nucleofectadas.

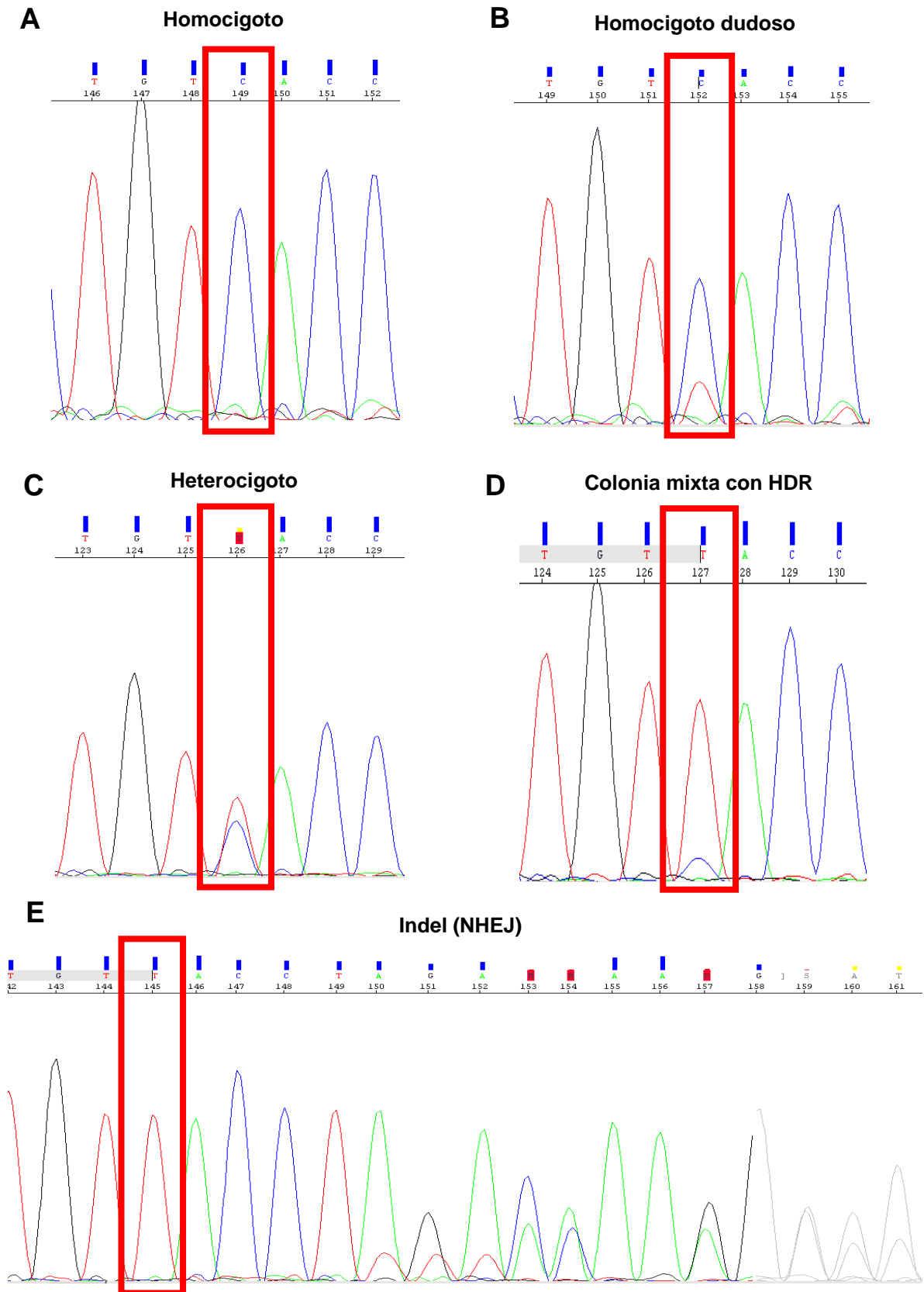
Sin embargo, en las dos estrategias realizadas con la línea original de SBr (#1 y #2) sí se obtuvieron células editadas satisfactoriamente. Los resultados de TIDER indican, en primer lugar, que el HDR enhancer tuvo un papel relevante en aumentar la cantidad de eventos HDR (**Figura 119**). El número de eventos HDR prácticamente se duplicó en presencia de *HDR enhancer*, pasando de un 22% a un 41%. En consecuencia, si el número de eventos HDR aumentó, el número de indels (productos derivados de la vía de reparación NHEJ) también disminuyó.



**Figura 119. Eficiencia de edición en presencia o ausencia de *HDR Enhancer*.** Eventos HDR analizados con TIDER en las células SBr nucleofectadas con HiFiCas9/gRNA + ssODN, tratadas o no con el *HDR enhancer*.

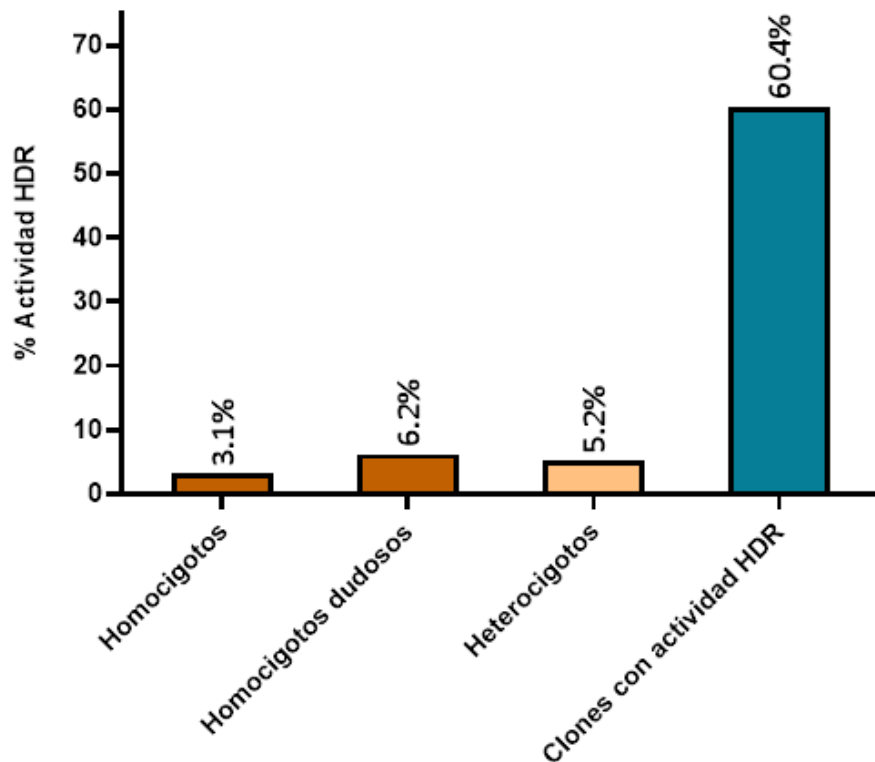
La secuenciación de los clones de la estrategia #1 permitió analizar con detalle todas las secuencias y, en línea con los resultados de TIDER, se obtuvieron diversos clones editados. Se encontraron diferentes tipos de productos de edición:

- **Homocigotos:** clones con ambos alelos editados (T>C). Se observa un solo pico de C en el electroferograma, substituyendo por completo la T original (**Figura 120A**).
- **Homocigotos dudosos:** secuencia limpia que el programa detecta como edición homocigota, pero cuando se analiza el electroferograma todavía queda cierto porcentaje de base T en la secuencia (**Figura 120B**). Esto puede suceder debido a que la colonia secuenciada no se originara de una sola célula, sino de dos o más, pudiendo encontrar una mezcla de ediciones. El hecho de hacer la siembra a baja densidad con cálculo de diluciones (no se disponía de un *sorter*) hace que sea inevitable la presencia de este tipo de colonias.
- **Heterocigotos:** clones con un solo alelo editado (T>C). Se observan dos picos aproximadamente del mismo tamaño en la posición editada (**Figura 120C**).
- **Clones con actividad HDR:** se consideraron clones con actividad HDR todos aquellos que presentaron algún pico significativo de citosina en la posición de edición. No obstante, salvo los casos descritos anteriormente, estos clones sufrieron algún proceso de NHEJ, dando lugar a indels en uno de los alelos (**Figura 120E**). También se pudo dar el caso de clones originados a partir de 2 o más células, donde una de ellas sufriese procesos de NHEJ i la otra HDR (**Figura 120D**).



**Figura 120. Ejemplos de electroferogramas para los diferentes tipos de edición obtenidos. El rectángulo rojo marca la posición de edición. (A) Homocigoto, 100% C. (B) Homocigoto dudoso, pico pequeño de T. (C) Heterocigoto, 50% T, 50% C. (D) Colonia que proviene de 2 o más células con actividad HDR en alguna de ellas. (E) Clon con un indel, producto del NHEJ.**

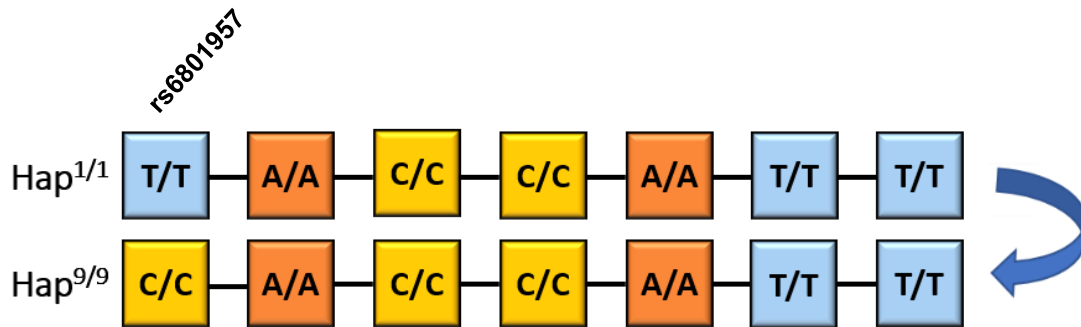
En total se secuenciaron 96 clones, de los cuales el 3.1% (3 clones) fueron homocigotos (**Figura 121**). Este dato fue sorprendente dada la baja eficiencia de la edición por HDR en hiPSCs descrita hasta la fecha<sup>427,428</sup>. De hecho, el protocolo original (**Figura 59**; apartado 2.11.2. de Métodos) contemplaba tener que hacer una segunda ronda de edición para pasar de un producto heterocigoto intermedio al homocigoto. Los homocigotos dudosos representaron un 6.2% del total (**Figura 121**), siendo también un resultado muy valioso, aunque se decidió que estos clones no se usarían más en futuros experimentos para asegurar que la edición fue 100% homocigota. El 5.2% (5 clones) fueron heterocigotos, aunque hubo bastantes más clones cerca de lograrlo, probablemente siendo una mezcla de células, ya que presentaban proporciones de edición del 25%, aproximadamente, en lugar del 50% buscado (**Figura 121**). Por último, un alto porcentaje de los clones presentó algún tipo de evento HDR (60.4%), apoyando el hecho de que el protocolo usado en esta tesis resultó ser altamente eficiente y útil para generar ediciones puntuales en hiPSCs.



**Figura 121. Actividad HDR en los clones secuenciados a partir de la línea SBr + HDR Enhancer.** Porcentaje de los 96 clones para cada tipo de edición. Los clones con actividad HDR engloban a todos aquellos que presentan algún tipo de actividad HDR, independientemente de otros eventos NHEJ.

Con estos resultados se dispone de células hiPCS procedentes del paciente con SBr (Hap<sup>1/1</sup>) con uno de los tres cambios necesarios para cambiarlo a Hap<sup>3/3</sup>. De hecho, solamente con este cambio se ha pasado del genotipo homocigoto Hap<sup>1/1</sup> al homocigoto Hap<sup>9/9</sup> (**Figura 122**). Hap9

no se identificó en la cohorte de SBr, solamente en las poblaciones control, aunque en heterocigosis, en ningún caso el genotipo Hap<sup>9/9</sup>. Las perspectivas de futuro pasan por realizar las dos ediciones restantes, pero también en valorar si el estudio de haplotipos intermedios (como el obtenido aquí) podría aportar un valor añadido en esta investigación.

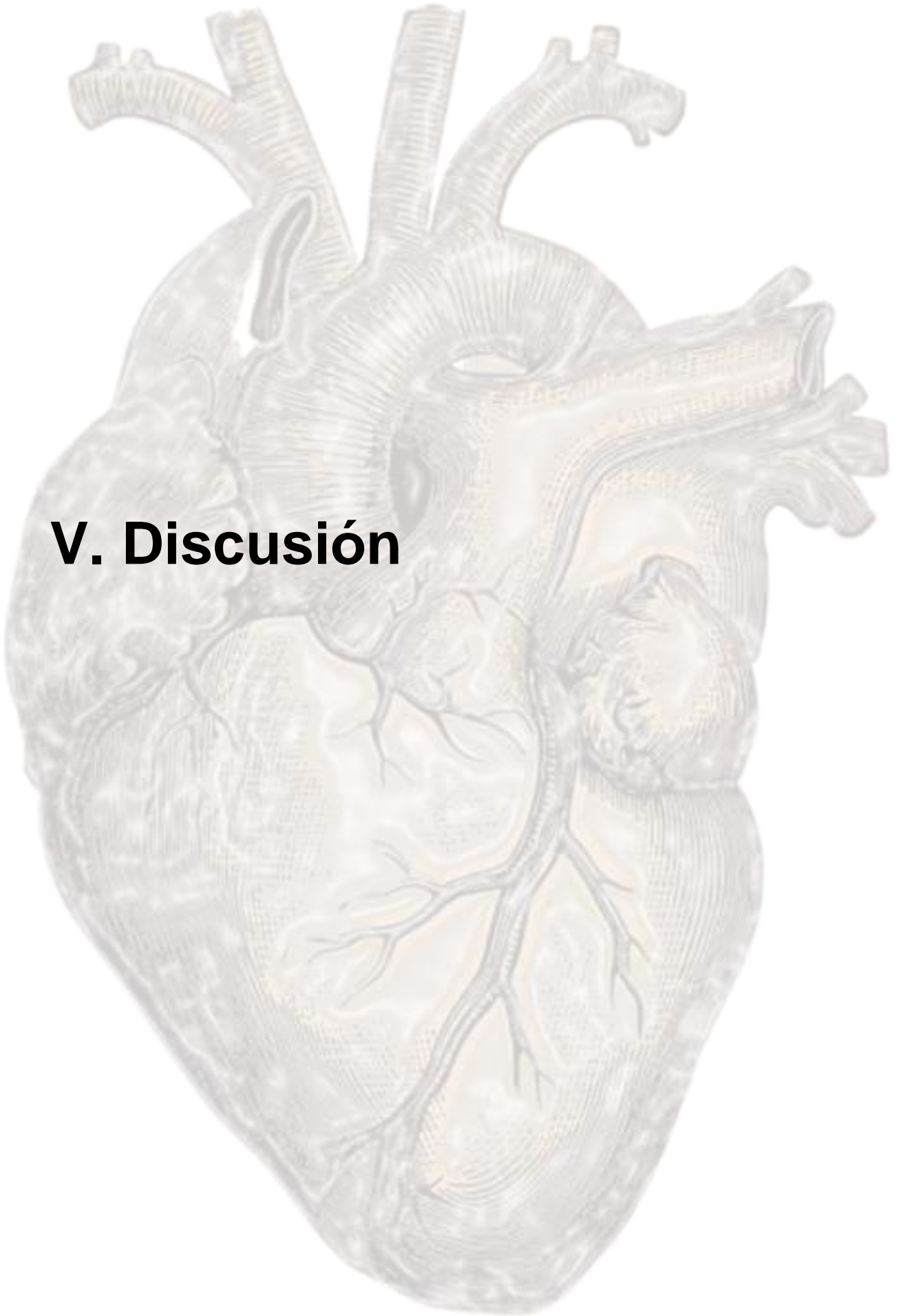


**Figura 122. Haplotipo obtenido después de la edición homocigota de la variante rs6801957.** Con el cambio homocigoto en la posición rs1 del haplotipo Hap<sup>1/1</sup> (rs6801957, T>C), se obtiene Hap<sup>9/9</sup> en la línea hiPSC de SBr.





## **V. Discusión**





## 1. Acetilación del GATA4 y su efecto sobre la regulación del promotor *SCN5A*

El estudio de Takaya *et al.* fue el primero que proporcionó evidencias de los sitios de acetilación de GATA4 por p300. Este estudio definió cuatro lisinas de GATA4 (K311, K318, K320 y K322), localizadas en la región básica del motivo C-Zn, como dianas de p300<sup>163</sup>. La determinación de estos sitios de acetilación se realizó mediante ensayos luciferasa en células HEK293T. En estos ensayos analizaron el efecto de GATA4 WT o con lisinas mutadas sobre la actividad de los promotores *ANP* y *ET-1*, en ausencia o presencia de p300. Concretamente, analizaron la actividad transcripcional de los siguientes mutantes: M1 (K245A) y M2 (K255A), localizados en el motivo N-Zn; M3 (K299A), M4 (K311A), M5 (K318A/K320A), M6 (K322A), y M7 (K326A/K328A) localizados en el motivo C-Zn. En presencia de p300, observaron que los mutantes M4, M5 y M6 eran los que disminuían más su actividad transcripcional en relación a GATA4 WT. A partir de este resultado, así como la no detección de actividad transcripcional al co-transfectar un mutante M456 (equivalente a GATA4-KA utilizado en esta tesis) concluyeron que las 4 lisinas que están mutadas son los sitios de acetilación de p300. Sin embargo, cabe remarcar que este estudio no es concluyente para establecer la posición exacta de las lisinas de GATA4 acetiladas. Al margen de que algunas de las lisinas mutadas fueron analizadas de forma simultánea y no individualmente, los autores no examinaron la correcta expresión de todos los mutantes analizados en el ensayo. Además, se tiene que tener en cuenta que la pérdida de actividad transcripcional de un FT al mutar una lisina no es una prueba irrefutable de que el sitio está acetilado. Una posibilidad es que otras modificaciones postraduccionales que pueden ocurrir en lisinas y que estarían afectadas en estos mutantes, como fosforilaciones o SUMOilaciones<sup>372,374</sup>, se vean afectadas e interfieran en la actividad transcripcional de GATA4. Otra posibilidad es que el cambio de lisina a alanina cambie la conformación del dedo de zinc (C-Zn) y afecte la capacidad de unión al DNA u otras proteínas.

En la tesis de Anna Tarradas, desarrollada en nuestro grupo de investigación, se realizaron ensayos de acetilación y desacetilación *in vitro* de péptidos sintéticos de GATA4. El posterior análisis por espectrometría de masas demostró que tres de las lisinas de GATA4 presentes en el dominio de unión al DNA son dianas de p300 y de HDAC2. Aunque previamente se habían descrito las lisinas K311, K318, K320 y K322 como los residuos dianas de p300<sup>163</sup>, los resultados obtenidos en la tesis de la Dra. Tarradas demuestran que sólo dos de estas lisinas se acetilan. De manera más detallada y a partir de los péptidos sintéticos se determinó que, de los tres lugares de acetilación, dos se hallan entre las lisinas K138, K320 y K322, y el último entre las lisinas K326 y K328, desconociéndose la posición exacta en el primer grupo. En cambio, este mismo trabajo identificó la lisina K326 como un nuevo sitio de acetilación por p300 siguiendo la misma metodología que nuestro trabajo y también mediante LC-MS/MS.

Los resultados de esta tesis sugieren que los residuos que contribuyen en mayor medida a la acetilación de GATA4 serían K318, K322 y K326. Estos resultados se han obtenido en experimentos de inmunoprecipitación seguido de WB en los que se han analizado los niveles de acetilación del FT GATA4 (WT o mutado) expresado en células HEK293T. Hay que tener en cuenta que el análisis por WB es un método semicuantitativo y puede dar lugar a que la interpretación de los resultados sufra un sesgo de objetividad. Sin embargo, los replicados de nuestros experimentos muestran el mismo patrón de acetilación, apoyando los resultados obtenidos. No obstante, no se puede descartar la posibilidad de que existan otros residuos implicados en la acetilación de GATA4, por ejemplo, fuera del dominio estudiado de C-Zn. De acuerdo con esta hipótesis, el mutante TM (K318, K322, K326) mostró niveles de acetilación detectables, sugiriendo que estos residuos no pueden ser los únicos acetilados. Por otra parte, p300 es una proteína dinámica y la sustitución consecutiva de más de un residuo de GATA4 podría provocar una alteración de los residuos diana de p300, dando lugar a la acetilación de otras lisinas que en condiciones WT de GATA4 no son acetiladas.

Por otro lado, hay que tener en cuenta los estudios de acetilación *in vitro* y espectrometría de masas realizados por la Dra. Tarradas. Estos estudios se realizaron con péptidos de aproximadamente 12-15 residuos y el dominio catalítico de p300 recombinante. La acetilación con péptidos pequeños, en lugar de toda la proteína entera, podría alterar la especificidad de la enzima p300 al añadir grupos acetiles. Por tanto, se podría estar omitiendo algún posible lugar de acetilación en esta región, o bien que se acetilen residuos que en células no son acetilados. También cabe destacar que las reacciones de acetilación *in vitro* se realizaron usando el dominio catalítico (HAT) de p300. Por último, es importante también tener en cuenta que tanto los estudios de acetilación analizados por espectrometría como los de esta tesis se han realizado en condiciones *in vitro*, por lo que debería evaluarse si los resultados obtenidos son reproducibles a modelos *in vivo*.

La tesis de Anna Tarradas también demostró que GATA4 y la acetiltransferasa p300 activan sinérgicamente el promotor de *SCN5A*, sugiriendo que el efecto de GATA4 sobre *SCN5A* está modulado por acetilación<sup>18</sup>. Además, en experimentos de ChIP se observó que GATA4 y p300 se unen a la misma región del promotor de *SCN5A* en las muestras de ventrículo humano adulto. De acuerdo con la actividad desacetilasa de HDAC2 sobre GATA4, se observó que la sobreexpresión de HDAC2 reduce la actividad transcripcional de GATA4 sobre el promotor de *SCN5A*<sup>18</sup>.

Los resultados con los mutantes de GATA4 estudiados en esta tesis sugieren que la lisina K318 es la que tiene un papel más importante en modular la función de GATA4. Concretamente, nuestros experimentos luciferasa demuestran que los mutantes que presentaban la sustitución K318 (K318A y TM) muestran una actividad transcripcional significativamente menor que GATA4 WT. Estos resultados sugieren que la imposibilidad de acetilar este residuo afectaría a la afinidad

de unión al DNA, disminuyendo la actividad transcripcional. Sin embargo, serían necesarios más estudios para entender el rol del residuo K318 en la actividad transcripcional de GATA4.

En nuestros resultados luciferasa también observamos una disminución de la actividad transcripcional de los promotores de *SCN5A* y *PDX1* en co-expresar GATA4<sup>TM</sup> (K318A+K322A+K326A). A partir de este resultado, se pueden hipotetizar distintos escenarios:

- (1) Que las lisinas mutadas participan en la actividad transcripcional de GATA4.
- (2) Que la acetilación de las lisinas K318, K322 y K326 aumenta la afinidad de unión de GATA4 al DNA.
- (3) Que las lisinas K318, K322 y K326 son importantes para la interacción de GATA4 con co-activadores.

En referencia a los escenarios planteados y tal como se comentó anteriormente, estudios con las desacetilasas HDAC2 han demostrado que la eliminación de los grupos acetiles de GATA4 reduce significativamente su actividad transcripcional<sup>18,166</sup>. Además, la observación de que el efecto de los mutantes de GATA4 sobre el promotor *PDX1* es similar al promotor *SCN5A* sugiere que estos residuos de GATA4 afectan directamente la actividad transcripcional de GATA4 independientemente del promotor diana. Por otra parte, teniendo en cuenta que GATA4 interacciona con diferentes FT y co-factores para activar la transcripción, se puede hipotetizar que la sustitución de las lisinas podría alterar sitios de interacción de GATA4 con otras proteínas. Sin embargo, estudios de co-inmunoprecipitación apuntan a que GATA4 conserva la capacidad de unión a p300 en presencia de mutaciones en este dominio de unión del DNA<sup>163</sup>. Aun así, habría que realizar más estudios para evaluar si estos mutantes alteran la interacción de GATA4 con otros co-factores, o bien con otros FT. En este sentido, es importante recordar que trabajos de nuestro laboratorio y otros han demostrado que GATA4 actúa sinérgicamente con otros FT cardíacos<sup>18,129,155</sup>.

Nuestros experimentos con el promotor *PDX1*, mostraron un aumento de la actividad transcripcional en co-expresar algunos de los mutantes de GATA4 (K322A, K326A y K328A y DM, que contiene K322A y K326A). La observación de este patrón, sólo para *PDX1*, dificulta hipotetizar un mecanismo general para estos mutantes de GATA4. Sin embargo, se pueden plantear algunas hipótesis asumiendo que los efectos resultantes no son iguales en ambos promotores. Una posibilidad sería que la sustitución de una lisina (aminoácido cargado positivamente) por una alanina (apolar) estuviera alterando la polaridad de la región de acetilación de GATA4, afectando a su capacidad de interacción con otros FT o con el DNA, y por tanto influyendo en su actividad transcripcional. Por otro lado, podríamos hipotetizar que haya una regulación cruzada (*cross-regulation*) entre modificaciones postraduccionales que modulan la actividad transcripcional de GATA4. Las lisinas, aparte de acetilarse, también pueden metilarse, ubiquitinarse o SUMOilarse. Dependiendo del contexto, las modificaciones

postraduccionales pueden afectar a otras modificaciones postraduccionales de forma agonista o antagonista<sup>377</sup>. Por ejemplo, se ha demostrado que la metilación de K299, K265 y K317 inhibe la interacción de p300 con GATA4, afectando a sus niveles de acetilación y consecuentemente a la actividad transcripcional<sup>372,373</sup>. De manera similar, la sustitución los residuos en cuestión (K322A, K326A y K328A) podría producir la alteración de otras modificaciones postraduccionales involucradas en la regulación transcripcional de GATA4.

Otro aspecto a tener en cuenta en esta discusión es el efecto del mutante KA en ambos promotores. Aunque Takaya y colaboradores, así como estudios posteriores, sugirieron que este efecto se debe a la importancia de la acetilación de estos cuatro residuos<sup>163</sup>, los resultados de MALDI-TOF y co-inmunoprecipitación junto con nuestros resultados, indican que de los cuatro residuos sólo K318 y K322 serían acetilados<sup>18</sup>. Por tanto, el efecto de las mutaciones en los residuos K311 y K320 podría estar relacionado con otros mecanismos como, por ejemplo, cambios conformacionales en este dominio de unión impidiendo la interacción de otros FT necesarios para la unión al DNA. Por otra parte, las mutaciones de los residuos K311 y K320 podrían afectar a la adición de otras modificaciones postraduccionales o, al contrario, promover la adición de modificaciones postraduccionales que impidan la acetilación de GATA4.

Algunos estudios han reportado que el mutante KA, además de no unirse al DNA, en presencia de GATA4 WT endógena actúa como inhibidor de su actividad transcripcional (dominante negativo)<sup>163,470</sup>. Estos estudios sugieren que KA compite por los co-factores (como p300) que permiten la actividad transcripcional de GATA4 WT, impidiendo su sinergismo e interacción con otros FT. En base a esta hipótesis, el mutante KA estaría actuando secuestrando p300, lo que podría impedir la acetilación de histonas u otros factores, contribuyendo así a una conformación cerrada de la cromatina, que dificultaría la actividad transcripcional de GATA4.

En resumen, para corroborar los resultados obtenidos en esta tesis, sería necesario estudiar más profundamente las interacciones entre GATA4 acetilado y el DNA. Una posibilidad sería realizar estudios de simulación para evaluar si la alteración de estos residuos sólo está relacionada con una disminución de los niveles de acetilación (y por tanto de la afinidad al DNA), o bien induce una modificación de la estructura proteica de GATA4 debido a los cambios aminoacídicos, dificultando así su actividad transcripcional. En este sentido, realizar estudios de la función de GATA4 realizando mutaciones de lisinas a argininas, en lugar de lisinas a alaninas, conservaría la carga del aminoácido, dando más indicaciones sobre el efecto de estas sustituciones en la actividad transcripcional de GATA4.

Por último, recalcar la importancia del residuo K318 tanto en la acetilación como en la actividad transcripcional de GATA4. Este residuo podría convertirse en una posible diana de modulación de la actividad de GATA4, reduciendo su actividad de forma moderada.

## 2. Identificación y caracterización de regiones genómicas específicas con función reguladora sobre el gen *SCN5A*

Las regiones no codificantes cubren aproximadamente el 98% del total del genoma e incluyen un amplio abanico de elementos (apartado 3.1. de la Introducción), entre ellos, los elementos reguladores en *cis* implicados en la regulación de la transcripción génica. Teniendo en cuenta que sólo un 25% de los casos de SBr son causados por alteraciones en las regiones codificantes del *SCN5A* y alrededor del 5% en regiones codificantes de otros genes asociados, en los últimos años han aumentado las evidencias que apoyan la hipótesis de variantes genéticas en regiones reguladoras como posible causa de los casos en los que la etiología es desconocida. En este contexto, diferentes autores han identificado la necesidad de estudiar las regiones no codificantes para entender los mecanismos y bases genéticas que caracterizan al SBr<sup>249,352</sup>. Todo esto ha supuesto un cambio en la investigación del SBr, que ha detectado la necesidad de estudiar de manera mucho más extensa y detallada las regiones reguladoras del genoma y, en particular, las regiones reguladoras del gen *SCN5A*, para entender la base genética del SBr, especialmente de los casos con etiología desconocida. Dada la complejidad asociada a la regulación de *SCN5A* y la cantidad de elementos reguladores dispersos a lo largo de todo el locus *SCN5A-SCN10A* (apartado 7.1.2. de la Introducción), en esta tesis se han identificado y caracterizado de forma exhaustiva regiones específicas del locus *SCN5A-SCN10A* en base a su efecto en la expresión de *SCN5A*. Hasta la fecha, este es el primer estudio en el que se realiza un cribaje de regiones reguladoras del gen *SCN5A* utilizando una estrategia basada en CRISPRi en un modelo celular de cardiomiocitos derivados de hiPSC.

Los resultados de esta tesis demuestran que el promotor de *EXOG*, el intrón 1 de *SCN5A*, la región 5' *SCN5A* y regiones intrónicas en *SCN10A* están implicados en la regulación de la transcripción I de *SCN5A*. Estos resultados se obtuvieron a partir del análisis de los niveles de mRNA por qPCR, derivado de los ensayos CRISPRi con dCas9-CRAB, i a partir del análisis de sodio intracelular.

### 2.1. Interacción de las regiones CRISPRi con otros elementos del genoma

- Discusión en base a los datos de interacción de Tang *et al.*<sup>430</sup>

Nuestros resultados sugieren que la región del promotor *EXOG*, situado a 5' del gen *SCN5A*, modula la transcripción del gen *SCN5A*. El gen *EXOG* codifica para la endo/exonucleasa G con actividad exonucleasa 5'-3'. *EXOG* se expresa de manera ubicua a nivel de mRNA en todos los tejidos<sup>471</sup>, no obstante, los datos de GTEx indican una sobreexpresión de *EXOG* en el ventrículo izquierdo (x4.1). Aunque la función de la endo/exonucleasa G en cardiomiocitos no está bien



estudiada, algunos autores han sugerido que la depleción de este gen podría dar lugar a hipertrofia en cardiomiocitos neonatales<sup>429</sup>. Además, la base de datos del catálogo de GWAS incluye SNVs en *EXOG* asociados a fenotipos relacionados con el ECG (duración onda P y QRS) con el gen *EXOG* (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=EXOG>).

Nuestros resultados muestran que una disminución significativa de la expresión de *SCN5A* siendo en la región promotora de *EXOG* (línea ***EXOG-CTCF***). En efecto, esta es una de las regiones en las que se observa un mayor efecto de todas las estudiadas (además de Intrón 1-NKX2-5 y TSS+183pb). Este resultado es interesante ya que hasta la fecha no se ha documentado ninguna región reguladora del gen *SCN5A* en el promotor *EXOG*. Además, la gRNA usada fue dirigida específicamente a un TFBS de CTCF, en una región enriquecida en picos de CTCF, lo que la convierte en una región potencialmente involucrada en la organización tridimensional del genoma. Este resultado también estaría respaldado por el estudio de van den Boogaard *et al.* en el cual se describió, mediante experimentos de 4C, que la región promotora del gen *EXOG*, a nivel tridimensional, se encuentra cercana a las regiones promotoras del gen *SCN5A* y la región *enhancer* del gen *SCN10A*<sup>51</sup>. Asimismo, los análisis de interacción de la cromatina a partir de datos del estudio de Tang *et al.* (ChIA-PET de CTCF) muestran que existe una vinculación entre la región promotora de *EXOG* y la región 5'*SCN5A*, coincidiendo las interacciones de manera específica con los lugares donde están dirigidas las gRNAs de las líneas *EXOG-CTCF* y 5'*SCN5A-CTCF*. A partir de estas observaciones, podemos hipotetizar que la interacción entre ambas regiones esté involucrada en la regulación de la expresión de *SCN5A*. A pesar de que la disminución en la expresión de *SCN5A* no fue estadísticamente significativa para los experimentos de CRISPRi realizados con la línea 5'*SCN5A-CTCF*, cabe destacar que esta línea fue la que presentó mayor desviación estándar en el análisis de datos y, dada su vinculación con *EXOG-CTCF*, merece ser tenida en cuenta para futuros estudios.

La técnica CRISPRi se basa en el diseño de gRNAs dirigidas a regiones muy específicas del genoma, pudiendo tener diferentes eficiencias dependiendo del diseño de la gRNA y su accesibilidad a esa zona concreta del genoma. Con el objetivo de confirmar nuestra hipótesis, este estudio se podría complementar con el diseño de gRNAs que amplíen la región de estudio para 5'*SCN5A*.

En el caso de la línea **5'*SCN5A-GATA4***, se observó una disminución significativa en los niveles de expresión de *SCN5A* en los análisis por qPCR, así como en las mediciones del sodio intracelular (aunque solamente se pudo analizar una ronda de diferenciación). Este resultado podría estar asociado a la función de GATA4 como activador transcripcional de *SCN5A*<sup>157</sup>. De forma similar a la línea 5'*SCN5A-CTCF*, la región donde se dirigió la gRNA de 5'*SCN5A-GATA4* muestra una interacción con *EXOG*, aunque en este caso en una región intrónica de *EXOG*, en lugar de en el promotor. Este dato podría sugerir la presencia de lugares de unión a CTCF en la misma zona de 5'*SCN5A-GATA4*, ya que los datos de interacciones usados para este caso captan específicamente las interacciones de CTCF<sup>430</sup>.

- Discusión en base a los datos de interacción de Montefiori *et al.*<sup>431</sup>

A partir del estudio de Montefiori *et al.*<sup>431</sup> (datos de PChi-C; apartado 2.6. de Resultados) se detectaron interacciones adicionales entre la región del promotor *EXOG* donde se dirige la gRNA de ***EXOG-CTCF*** y regiones más distales, como la región promotora del gen ***MYD88***. Esta interacción resulta interesante ya que existe una asociación que vincula *MYD88* y SBr según las anotaciones de variantes de la base de datos ClinVar (<https://rgd.mcg.edu/rgdweb/report/annotation/main.html?term=DOID:0050451&id=735043>), en base a tres estudios<sup>66,323,436</sup>.

*MYD88* codifica para una proteína adaptadora citosólica (denominada “proteína de respuesta primaria de diferenciación mielóide 88”) que se expresa de manera ubicua en el organismo y desempeña un papel esencial como transductor de la señal en las vías de señalización de la interleucina-1 y del receptor tipo Toll<sup>437</sup>. Además de su importancia para la respuesta inmunitaria innata, *MYD88* también se ha relacionado con el mantenimiento de la función fisiológica en el corazón adulto<sup>438-440</sup>. Un estudio realizado por Chen *et al.* generó una cepa de ratón con sobreexpresión de la proteína *MYD88* en el corazón<sup>438</sup>. Estos ratones transgénicos presentaban corazones estructuralmente normales, pero mostraban una disminución de la contractilidad del ventrículo izquierdo, un fenotipo que también se observa en algunas canalopatías cardíacas como SBr y LQTS<sup>441,442</sup>. Estos resultados sugieren que una desregulación de la expresión de *MYD88* podría tener un papel importante en la aparición de arritmias cardíacas. Por este motivo, la observación que *MYD88* interacciona con una región CTCF (*EXOG-CTCF*) que modula la expresión de *SCN5A* en presencia de dCas9-KRAB es interesante y permite establecer una posible conexión entre ambas regiones. Una posible hipótesis sería que la región *EXOG-CTCF* también estuviera implicada en la regulación del gen *MYD88*.

Otro dato relevante a destacar es la identificación de una inserción de 6 pb (rs545602132) en la región promotora de *MYD88*. Esta inserción se identificó en un proyecto de tesis de nuestro grupo a partir de los datos de secuenciación de las regiones reguladoras en la cohorte de 89 pacientes<sup>8</sup>. La inserción se identificó en cuatro individuos con SBr y es específica para SBr. Una hipótesis propuesta en la tesis de M. Pinsach<sup>8</sup> es que esta inserción podría estar afectando la actividad del promotor *MYD88*, aumentando la expresión del gen *MYD88* y causando efectos similares a los observados por Chen *et al.*<sup>438</sup>. Otra posible explicación podría ser que la inserción estuviese creando un nuevo TFBS involucrado en su regulación transcripcional.

Los resultados de esta tesis, junto con los estudios anteriores, sugieren que la inserción rs545602132 podría afectar la interacción del promotor de *MYD88* con el de *EXOG*, alterando las estructuras topológicas que involucran la región promotora de *EXOG* y a su vez, la expresión de *SCN5A*. Por lo tanto, de manera global y en base al ejemplo de la inserción, podemos hipotetizar que variantes que se superpongan a lugares de unión de FT involucrados en la organización tridimensional del genoma podrían obstaculizar la regulación del TAD-*SCN5A*. Este acontecimiento daría como resultado una expresión alterada del gen *SCN5A*, que a su vez puede estar asociada con arritmias/canalopatías cardíacas asociadas con los canales de sodio, como

el SBr. En el caso del ensayo CRISPRi con *EXO-CTCF*, podemos hipotetizar que dCas9-KRAB estaría interfiriendo con la correcta unión de CTCF a la región promotora, alterando la correcta regulación espacial de la zona y, en consecuencia, la expresión de *SCN5A*.

▪ Discusión en base a los datos de interacción de Jin *et al.*<sup>444</sup>

El análisis exhaustivo del estudio de Jin *et al.* mediante Hi-C en fibroblastos humanos permitió encontrar nuevas interacciones entre las regiones de estudio de esta tesis. Por un lado, identificamos una interacción entre la región a la que se dirige la gRNA de ***EXO-CTCF*** con una región del intrón 1 de *SCN5A*. Las gRNAs de las líneas Intrón 1-CTCF e Intrón 1-NKX2-5 se dirigen a esta región de interacción del intrón 1. En el caso de la línea **Intrón 1-NKX2-5**, en los experimentos CRISPRi se observó una disminución significativa en la expresión de *SCN5A*, mientras que en la línea **Intrón 1-CTCF** no se observó ninguna alteración significativa.

La región del intrón 1 de *SCN5A* se ha descrito previamente como un elemento regulador en *cis* localizado en el promotor del gen<sup>333</sup>. En cambio, NKX2-5 es un factor de transcripción cardíaco imprescindible para el buen funcionamiento de los cardiomiocitos<sup>443</sup>. Una posible explicación de la disminución de la expresión de *SCN5A* en la línea Intrón 1-NKX2-5 podría ser que la presencia de dCas9-KRAB alterase la unión de este FT y diera lugar a una desregulación en la expresión de *SCN5A*. Cabe destacar también que el estudio de Tarradas *et al.* demostró que el intrón 1 de *SCN5A* es una región clave para modular la expresión de *SCN5A* mediante la unión de GATA4. Esta región contiene TFBS claves para la unión de GATA4, por lo que dCas9-KRAB pudo haber alterado las interacciones de GATA4 en la región.

Para las regiones intrónicas en *SCN10A* seleccionadas para el ensayo CRISPRi, solamente observamos una disminución significativa de la expresión de *SCN5A* para las líneas ***SCN10A-GATA4*** y ***SCN10A-rs6801957***. Es interesante destacar que, según los datos de Jin *et al.*, la región donde se dirige la gRNA de la línea *SCN10A-rs6801957* interacciona con la región promotora del intrón 1 estudiada en los experimentos CRISPRi. Estos resultados están de acuerdo con la función de la región intrónica de *SCN10A* como una región *enhancer* que interactúa con el promotor de *SCN5A*<sup>51</sup>. La variante rs6801957, situada en el intrón 16 de *SCN10A*, afecta la unión de los FT TBX3/TBX5, dando lugar a una disminución de la expresión de *SCN5A*, y se ha asociado con defectos en la conducción eléctrica cardíaca<sup>51,445</sup>. El estudio realizado en esta tesis confirma la función de esta región como *enhancer* de *SCN5A* utilizando un modelo de hiPSC-CMs y valida que el sistema CRISPRi es válido para detectar alteraciones en la expresión génica incluso en regiones muy distales. En base a nuestros resultados, podemos sugerir que dCas9-KRAB estaría imitando el efecto de la variante rs6801957 al interferir con la unión de TBX3/TBX5.

En nuestros experimentos de CRISPRi, las líneas ***SCN10A-CTCF*** y ***5' SCN5A-CTCF*** no dieron lugar a una alteración significativa de la expresión de *SCN5A*. No obstante, los datos de Jin *et*

*al.*, muestran que existe una interacción entre las dos regiones a las que se dirigen las gRNAs de estas dos líneas (**Tabla 18**; apartado 2.6. de Resultados). La observación de que ninguna de las líneas con dianas para CTCF, salvo **EXO-CTCF**, haya presentado una disminución significativa de la expresión de *SCN5A* podría estar relacionado con una función redundante de estas regiones en la regulación de la expresión génica. Diversos estudios han descrito mecanismos evolutivos, mediante los cuales, la alteración de una región CTCF se puede compensar con la presencia de otras regiones con función redundante en la misma región regulatoria<sup>387,460</sup>. Esta hipótesis cobra aún más sentido analizando las conexiones entre las regiones CTCF estudiadas en nuestro estudio. de, En este sentido, se podría hipotetizar que la región para la gRNA de **EXO-CTCF** tiene una mayor relevancia en la coordinación transcripcional a partir del efecto significativo sobre la expresión de *SCN5A* y la observación que interacciona con un mayor número de regiones del locus *SCN5A-SCN10A*.

## 2.2. Limitaciones, fortalezas y factores de variabilidad en el estudio CRISPRi

- Valoración del significado clínico de los resultados

A nivel de posible relevancia clínica, es interesante discutir el resultado de la línea **SCN10A-CTCF**, dado que el intervalo de confianza obtenido en los resultados de qPCR en relación a la disminución de la expresión de *SCN5A* (condición +DOX) podría considerarse clínicamente significativo. Es importante hacer una distinción clara entre significancia estadística y relevancia clínica (o significación clínica)<sup>446,447</sup>. Además de la significación, los valores p incorporan los números de casos y la variabilidad de los datos de la muestra. Incluso si el límite de la significación estadística se establece de antemano, la persona encargada de la investigación aún debe juzgar por sí misma la relevancia clínica de las diferencias estadísticamente significativas. El mismo valor numérico para la diferencia, puede ser "estadísticamente significativo" si el tamaño muestral es grande y "no significativo" si el tamaño muestral es más pequeño (por ejemplo, solamente 3 réplicas). Por otro lado, los resultados de alta relevancia clínica no pierden automáticamente su importancia si no hay significación estadística. La causa puede ser que el tamaño muestral sea demasiado pequeño o que la dispersión en las muestras sea demasiado grande. Por esta razón, una decisión sobre la relevancia de los resultados basada únicamente en el valor p puede ser simplista. Sin embargo, desafortunadamente, existe la falsa convicción entre buena parte de la comunidad científica que la significación estadística es equivalente a la relevancia clínica<sup>446-448</sup>.

En este sentido, resultados como los obtenidos en la línea **SCN10A-CTCF**, con un p-valor muy ajustado y un intervalo de confianza mayoritariamente dentro de la significación, deben ser estudiados con más profundidad y no descartarlos automáticamente por haber obtenido un

primer resultado no significativo. Para ello, se deberían repetir los experimentos con el objetivo de obtener resultados con una menor desviación estándar.

- Uso de un modelo *in vitro* de hiPSC-CMs

Un factor a tener en cuenta como fuente de variabilidad experimental es el uso de un modelo celular basado en un protocolo de diferenciación cardíaca. Aunque distintos estudios han demostrado que los hiPSC-CMs son relativamente inmaduros y no reflejan completamente los diversos aspectos estructurales y funcionales de las células cardíacas adultas, la dificultad para obtener subpoblaciones puras de células cardíacas primarias con alta integridad requiere el uso de un sistema *in vitro*<sup>449,450</sup>.

A pesar de que los protocolos actuales para la generación de hiPSC-CMs son muy eficientes, existen limitaciones al usar este modelo, derivadas principalmente del estado de inmadurez de las células<sup>451,452</sup>. Características estructurales y funcionales, como el acoplamiento excitación-contracción (que requiere de túbulos transversales), la gestión adecuada del calcio, un metabolismo eficiente (oxidativo), y la baja expresión de ciertos genes, son rasgos en gran medida ausentes en hiPSC-CMs que asemejan las propiedades electrofisiológicas de este modelo a un estadio fetal en lugar de uno adulto<sup>451-454</sup>.

A pesar de todo ello, a día de hoy sigue siendo uno de los mejores modelos de los que se dispone para estudiar las bases genéticas de las enfermedades cardíacas. Existen otros modelos como sistemas de expresión heterólogos, peces cebra, modelos murinos y otros animales. Aunque estos modelos presentan sus pros y contras, la principal ventaja de los modelos de hiPSC-CMs es que pueden contener el *background* genético específico de un paciente. Además, estas células reproducen un entorno similar al de los cardiomiocitos, dando la posibilidad, por ejemplo, de detectar cardiotoxicidad en fármacos probando sus efectos tóxicos en los cardiomiocitos<sup>300</sup>. Con los protocolos de diferenciación optimizados y un trabajo exhaustivo para conseguir rondas de diferenciación lo más puras posibles, sumado al correcto tratamiento estadístico de las muestras y el uso de replicados biológicos (rondas de diferenciación independientes), podemos considerar que los resultados obtenidos en esta tesis son fiables y válidos. Con todo, se debe repetir y aumentar la muestra en todos aquellos experimentos dudosos o con problemas celulares asociados al protocolo de diferenciación.

- Estudios funcionales con el sistema Operetta

En este trabajo, se midió la concentración de sodio intracelular de 15 rondas de diferenciación de cinco líneas CRISPRi distintas mediante el sistema Operetta. Estos experimentos fueron realizados en el laboratorio de Genómica y Bioinformática del CiMUS, Santiago de Compostela, ya que allí pudimos tener acceso al sistema Operetta, necesario para obtener las mediciones. Esta circunstancia añadió una serie de dificultades que se deben tener en cuenta para interpretar

los resultados. En primer lugar, el protocolo seguido para el análisis funcional de los hiPSC-CMs no pudo ser optimizado ni probado con anterioridad, por lo que todas las modificaciones o imprevistos se tuvieron que resolver durante el desarrollo de los experimentos. Cabe recordar que el procedimiento completo que implica descongelar y preparar las hiPSCs, diferenciar a cardiomiocitos y realizar los experimentos puede conllevar dos meses de trabajo si no hay errores, contaminaciones ni ningún imprevisto que altere el correcto estado de las células. Por lo tanto, el tiempo fue un recurso limitante que permitió poco margen de error y la necesidad de obtener los resultados de la mejor manera posible, aunque ello implicase eliminar alguna muestra. En segundo lugar, a pesar de que tenemos una amplia experiencia en el protocolo de diferenciación, el experimento requería un paso de escalado de placas de 12 pocillos a placas de 384 pocillos en el replaqueo. Para ello se hicieron experimentos de prueba para encontrar que número de células era el más adecuado para sembrar en este tipo de pocillos (conseguir la confluencia). No obstante, no teníamos experiencia sobre el comportamiento de los cardiomiocitos a pequeña escala. A pesar de las limitaciones, se obtuvieron resultados fiables y de calidad para las 15 rondas de diferenciación que superaron con éxito todas estas problemáticas.

La línea EXOG-CTCF fue la única que presentó unos resultados sólidos como para confirmar el efecto funcional de los ensayos CRISPRi, obteniendo una disminución del sodio intracelular de entre el 15%-20% en las dos rondas con tratamiento completo con DOX. A pesar de ello, los resultados globales obtenidos sugieren una buena reproducibilidad y resultan coherentes, observándose siempre una tendencia a la baja en las mediciones del sodio intracelular en las muestras +DOX, y nunca al revés. También debe valorarse la detección de diferencias significativas en un modelo en el cual las mediciones de sodio se hacen de manera menos precisa que en técnicas como el *patch-clamp*. Esta técnica utiliza los principios del *voltage clamp*, permitiendo el registro de corriente a través de una pequeña unión con la membrana celular de una sola célula aislada. Además, permite atribuir las alteraciones observadas a canales iónicos específicos<sup>455</sup>. En nuestro caso las mediciones se realizan de manera menos específica, midiendo la concentración de sodio intracelular de un conjunto de células en un momento dado. A pesar de ello, la técnica usada está avalada por diversos estudios donde evalúan satisfactoriamente corrientes de sodio y calcio, además del potencial de acción cardíaco<sup>456-458</sup>.

En resumen, el objetivo de estos experimentos fue el de realizar un estudio que pudiese complementar las técnicas de electrofisiología más complejas, para así poder seleccionar aquellas líneas con resultados significativos y, a posteriori, poder estudiar más detalladamente su electrofisiología con los otros métodos.

### 2.3. Conclusiones generales y futuros estudios

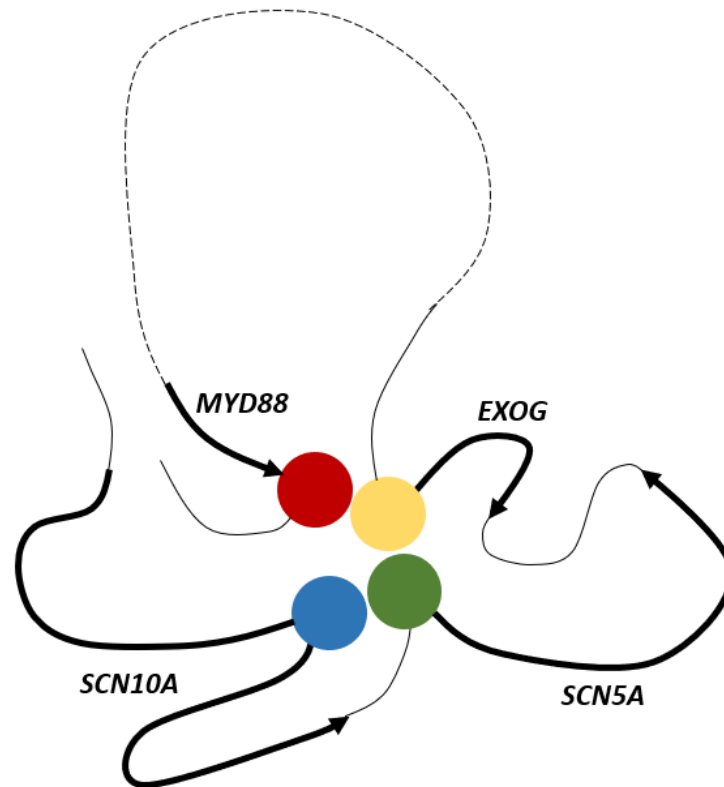
La identificación de una región reguladora del gen *SCN5A* en el promotor de *EXOG* pone de manifiesto la complejidad para entender el conjunto de interacciones existentes dentro del locus de *SCN5A*. De todas las regiones estudiadas, el promotor de *EXOG* se postula como la más relevante de todas por diversos factores:

1. En los ensayos CRISPRi, la línea celular *EXOG-CTCF* fue la que presentó una mayor disminución en los niveles de mRNA de *SCN5A* (exceptuando las regiones control en el TSS), además de ser la disminución con mayor significancia estadística.
2. En los estudios funcionales vinculados a CRISPRi, *EXOG-CTCF* fue la línea en la que se observó una mayor disminución del sodio intracelular y la que mostró mayor consistencia en los resultados.
3. La región diana para la gRNA de *EXOG-CTCF* se localiza en una región con múltiples interacciones con otras regiones reguladoras, según las bases de datos de interacciones de cromatina estudiadas. Las bases de datos le atribuyen interacciones con todas las regiones estudiadas y otras como el gen *MYD88*, excepto con las regiones intrónicas del *SCN10A*.

Nuestros datos sugieren que el complejo dCas9-KRAB, al interferir con esa región involucrada en múltiples contactos (y la unión de FT), estaría afectando en mayor medida la desregulación de la estructura tridimensional del locus (**Figura 123**). Esta propuesta contradice el modelo presentado por Boogaard *et al.*<sup>51</sup>, el cual no le atribuyó una función de interacción al gen *EXOG*.

El resto de regiones estudiadas también aportan información valiosa. Las dos regiones intrónicas en *SCN10A* (*SCN10A*-rs6801957 y *SCN10A*-GATA4) refuerzan el papel del *enhancer* de *SCN10A* como elemento regulador de *SCN5A*. Las líneas Intrón 1-NKX2-5 y 5'*SCN5A*-GATA4 refuerzan la importancia de los FT NKX2-5 y GATA4, además de la función reguladora de estas dos regiones cercanas al *SCN5A*.

En conclusión, el sistema CRISPRi desarrollado para la identificación de nuevas regiones reguladoras del *SCN5A* funcionó satisfactoriamente para regiones tanto cercanas como lejanas al gen. El uso de esta técnica puede ser una herramienta útil para caracterizar estas regiones y crear una traslación hacia el diagnóstico clínico de nuevas variantes de riesgo. Sin duda, las regiones con potencial clínico deberán pasar un proceso de validación funcional más exhaustivo para poderse vincular con la enfermedad de manera más robusta. Por ejemplo, las regiones específicas identificadas mediante CRISPRi y asociadas a una alteración de la expresión génica que contengan variantes asociadas a una enfermedad, serían candidatas para su estudio clínico.



**Figura 123. Modelo de interacciones del promotor *EXOG* en el locus *SCN5A-SCN10A*.** Cada círculo representa un elemento regulador: en azul, el *enhancer* de *SCN10A*; en verde, el promotor de *SCN5A* (que incluye las regiones de interacción del intrón 1 de *SCN5A* y 5' *SCN5A*); en amarillo, el promotor de *EXOG*; y en rojo, el promotor de *MYD88*. Las regiones en negrita representan cada uno de los genes indicados, siendo el resto regiones intergénicas. El *enhancer* de *SCN5A* interacciona con el intrón 1 y 5' *SCN5A*, mientras que el promotor de *EXOG* interacciona simultáneamente con el promotor de *MYD88*, el intrón 1 y 5' *SCN5A*. La región interlineada entre *EXOG* y *MYD88* indica que hay una amplia región entre ellos (y otros genes en medio de ambos no representados). Modelo realizado a partir de los datos de Boogaard *et al.*<sup>51</sup>, Man *et al.*<sup>49</sup> y los resultados de esta tesis.

Los resultados obtenidos en esta tesis son prometedores para continuar investigando estas regiones. Para confirmar la relevancia de estas regiones en el SBr, se propone, como siguiente paso, la realización de experimentos ChIP-seq para dilucidar en detalle los FT involucrados específicamente en las regiones diana de las gRNAs utilizadas y sus alrededores. Asimismo, experimentos de Hi-C aplicando los ensayos CRISPRi podrían ser clave para poder capturar el mapa de interacciones derivado del uso de dCas9-KRAB y evaluar si hay cambios en la estructura tridimensional. La realización de todos estos experimentos a gran escala permitiría obtener un mapa global de las regiones y FT implicados en la regulación del locus *SCN5A-SCN10A*. Por lo tanto, debido a la implicación del SBr con el gen *SCN5A*, disponer de esa información supondría entender de mejor manera nuevos mecanismos asociados a la fisiopatología de la enfermedad, especialmente en aquellos casos de SBr sin causa conocida.



### 3. Determinación y estudio del haplotipo asociado al síndrome de Brugada. Avances y futuras investigaciones.

#### 3.1. Comprobación experimental del haplotipo mediante Oxford Nanopore sequencing y comparativa con los grupos control

El poder y las ventajas de las tecnologías NGS se basan en su capacidad de analizar simultáneamente millones de pares de bases de DNA. Esto ha permitido a los investigadores e investigadoras comprender mejor el papel de la variación genética en términos de evolución, enfermedad o resistencia a los medicamentos, entre otros. En el campo del diagnóstico genético, las tecnologías de secuenciación de última generación se han convertido en métodos ampliamente utilizados para una gran variedad de trastornos hereditarios, facilitando nuevos avances en la predicción de enfermedades y la toma de decisiones terapéuticas para pacientes en riesgo<sup>390</sup>. Sin embargo, la precisión de estos test genéticos puede variar en función de una serie de variables, como la tecnología de secuenciación utilizada, la cobertura, y el tratamiento bioinformático utilizado para el descubrimiento de variantes. Algunas de las variantes detectadas mediante NGS pueden tener implicaciones médicas para los pacientes y sus familiares; por lo tanto, los criterios de calidad para aceptar las variantes diagnosticadas en el ámbito clínico son bastante conservadores (las variantes deben cubrirse  $\geq 30X$ ) y en ocasiones deben complementarse con secuenciación de Sanger para la validación de variantes de baja calidad. Sin embargo, la validación con secuenciación Sanger no se utiliza de manera rutinaria en los hospitales<sup>390</sup>.

Uno de los objetivos de esta tesis fue comprobar experimentalmente los haplotipos asociados al SBr, identificados por la Dr. Pinsach<sup>351</sup> a partir de algoritmos de predicción utilizando datos de secuenciación con *short-reads* de Illumina. El análisis de *short-reads* no permite la determinación empírica de los haplotipos debido a la amplia extensión genómica en la que se encuentran las 7 SNVs de interés que componen el haplotipo (12.4 kb aproximadamente). Para solventar esta limitación, en esta tesis se secuenciaron los haplotipos de los 86 pacientes de SBr incluidos en la cohorte de estudio mediante ONT, obteniendo como resultado la resolución experimental de 83 de ellos. Es decir, en el 96.5% de los pacientes, los dos métodos coincidieron. La tasa de éxito de la secuenciación con lecturas largas es notable, y la tasa de error de ONT (6.5%)<sup>294,295</sup> no fue un impedimento para esclarecer el *phasing* del haplotipo en la gran mayoría de casos. Es importante subrayar que los 86 pacientes fueron secuenciados con la técnica de ONT conjuntamente en un solo *pool*, identificando posteriormente cada paciente en el análisis. De esta manera, ONT aparece como una aproximación prometedora para aplicarse a la clínica de manera masiva una vez esta tecnología haya evolucionado lo suficiente y las tasas de error sigan bajando, como ha pasado durante los últimos años<sup>295</sup>. Cabe destacar que estos errores afectan principalmente a las secuencias en las cuales se repite muchas veces el mismo nucleótido. Esto es debido a la naturaleza intrínseca del método de secuenciación. Los nucleótidos se identifican

en función del cambio de voltaje generado en el nanoporo al pasar de un nucleótido a otro. Por ejemplo, si pasan seis timinas seguidas, el voltaje del nanoporo se mantendrá estable durante ese tiempo. La forma de determinar cuántos nucleótidos pasan reside, pues, en la velocidad de paso de la secuencia por el nanoporo. En este sentido es donde también se encuentran las mejoras a nivel de *software*, para detectar con mayor precisión la velocidad de paso y poder inferir con menor error las repeticiones nucleotídicas en las secuencias<sup>283,284,292</sup>.

Una vez determinados los haplotipos en la cohorte de SBr con ambas tecnologías, Pinsach-Abuin *et al.* los comparó con tres poblaciones control de individuos sanos (1KG, Welllderly y GTEx.; apartado 3.1. de Resultados). Se observó que Hap1 es significativamente más abundante en los individuos con SBr que en los controles, estando presente en casi el 75% de los pacientes con SBr y en menos del 40% en la población sana. Por el contrario, Hap2 y Hap3 son significativamente más frecuentes en los controles que en los casos de SBr<sup>351</sup>. Además, tal y como se ha comentado previamente en la sección de resultados, análisis posteriores de Pinsach-Abuin *et al.* determinaron que existe un haplotipo de riesgo asociado a SBr (Hap<sup>1/1</sup>) y otro asociado a protección (Hap<sup>2/3</sup>)<sup>351</sup>.

En relación a la genética del SBr, los resultados de Pinsach-Abuin permitieron hacer algunas observaciones importantes. En primer lugar, el análisis de estimaciones de *odds ratio* usando modelos de transmisión hereditaria determinó que el haplotipo de riesgo, Hap1, encajaría con un modelo de herencia recesiva. Este hallazgo es novedoso y significa un cambio importante en la genética del SBr si tenemos en cuenta que, en la actualidad, el modelo de transmisión más ampliamente aceptado para el SBr es autosómico dominante. En línea con nuestros resultados, dos estudios recientes también han sugerido un modelo de herencia recesiva para el SBr: una variante homocigota en *TRPM4*<sup>392</sup> y variantes en el gen *KCNE5* ligado al cromosoma X<sup>393</sup>. El modelo recesivo no es inusual para variantes patogénicas no codificantes<sup>394</sup>. Por ejemplo, el estudio de Bauwens *et al.* determinó que las variantes patogénicas no codificantes explican la mayoría de los casos no resueltos en la enfermedad asociada a *ABCA4* (distrofia de retina), con un modelo de heredabilidad autosómico recesivo<sup>461</sup>.

En el artículo de Pinsach-Abuin *et al.* también se observó que el análisis de los haplotipos en el *enhancer* *SCN5A-SCN10A* entre las superpoblaciones humanas presenta diferencias significativas. Es relevante que el genotipo Hap<sup>1/1</sup> es poco común en la población del este de Asia, lo que es consistente con estudios epidemiológicos que muestran que los casos de SBr en las poblaciones de China, Japón, Taiwán y Corea del Sur tienen menos variantes deletéreas que los caucásicos en el locus *SCN5A*<sup>395</sup>. Otra observación relevante fue la identificación del genotipo protector al SBr, Hap<sup>2/3</sup>. En este sentido, cabe destacar también estudios previos que describieron variantes protectoras contra arritmias en las regiones codificantes de los genes

*KCNQ1* y *SCN5A*<sup>396,397</sup> y variantes protectoras contra el SBr en una región codificante del gen *SCN10A* y en un lugar no codificante *downstream* del gen *HEY2* en la población japonesa<sup>398,399</sup>.

Es importante reconocer la creciente evidencia que sugiere que el SBr puede no estar causado por una única variante patogénica, sino por la presencia de múltiples variantes de susceptibilidad que actúen sinérgicamente a través de diferentes vías<sup>351,352</sup>. Estos estudios sugieren que el *background* genético de cada individuo (es decir, la presencia de distintas variantes genéticas) pueden conferir distintas susceptibilidades al SBr. En base a estas observaciones, no se puede descartar la posibilidad de que las combinaciones de variantes no codificantes interrumpen la expresión de genes asociados al SBr como otro mecanismo subyacente para explicar algunos casos de SBr. En general, las variantes comunes en regiones reguladoras podrían agravar o atenuar la manifestación de la enfermedad junto a un defecto genético primario, que suele estar descrito en una región codificante, pero estas variantes también pueden ser responsables del fenotipo de la enfermedad en ausencia de variantes codificantes<sup>247</sup>. A medida que aparecen nuevos estudios, cada vez es más evidente que tanto las regiones codificantes como las no codificantes tienen un papel fundamental en la fisiopatología de BrS<sup>327,400</sup>.

### 3.2. Análisis del perfil de expresión génica en las líneas SBr y GPG2

Con el fin de estudiar el papel de los haplotipos en el locus *SCN5A-SCN10A* a nivel funcional, en esta tesis se analizaron las diferencias en los perfiles de expresión génica entre un individuo sano (Hap<sup>1/3</sup>) y un individuo con SBr (Hap<sup>1/1</sup>). Este estudio es relevante por la particularidad de estudiar un paciente con SBr sin variantes en genes asociados a SBr, por lo que las alteraciones en la expresión génica asociadas a la enfermedad solamente se podrían explicar por el resto de su *background* genético, incluido Hap<sup>1/1</sup>. Los experimentos de RNA-seq se realizaron a partir de las líneas hiPSC-CMs de ambos individuos (líneas SBr y GPG2), concretamente a partir de 4 rondas de diferenciación independientes para cada línea.

El análisis de los resultados de RNA-seq mostró cambios en el perfil de expresión génica entre las líneas de hiPSC-CMs de SBr y GPG2, lo que sugiere cambios específicos a causa de la enfermedad. De entre todos los DEGs obtenidos en el RNA-seq, los genes *SCN5A*, *CACNA1C*, *ACTC1*, *MYH6*, *SCNN1A*, *PRKAG2*, *TTN*, *SEMA3A*, *KCNQ1* y *NKX2-5* se consideraron los que presentan mayor relevancia en relación con el SBr o el haplotipo Hap<sup>1/1</sup>.

En primer lugar, se observó una reducción en la expresión del gen *SCN5A* en las hiPSC-CMs de SBr en comparación con GPG2. Esta observación también ha sido documentada en estudios de expresión génica tanto de tejido humano como de ratones<sup>402,403</sup>, aunque hay que tener en cuenta que son modelos distintos al nuestro. Uno de los pocos estudios RNA-seq en la bibliografía realizado a partir de hiPSC-CMs de pacientes con SBr describió también una reducción de la expresión de *SCN5A*, así como también de *KCND3* y *KCNJ2*<sup>404</sup>, que codifican

para el miembro 3 de la subfamilia D del canal de potasio dependientes de voltaje y el miembro 2 de la subfamilia J del canal de potasio dependiente de voltaje, respectivamente. El papel de *KCNJ2* no se había descrito previamente en SBr. En nuestros resultados de RNA-seq, uno de los DEGs destacados fue *KCNQ1*, un gen que tampoco había sido asociado con SBr hasta el momento y presentó una reducción de su expresión respecto al control.

Cabe destacar que para las líneas celulares usadas para el RNA-seq (SBr y GPG2), los niveles de sodio intracelular fueron evaluados a nivel funcional con el sistema de medición Operetta. La reducción del sodio intracelular fue del 36% en la línea de SBr respecto al control. Este resultado estaría de acuerdo con disminución de la expresión de *SCN5A*, que posteriormente se traduciría en un menor número de canales de sodio en la membrana, observada en los experimentos de RNA-seq. La disminución de los niveles de sodio intracelular en la línea SBr en comparación con GPG2 es estadísticamente significativa y mayor que cualquiera de las alteraciones observadas en los ensayos CRISPRi, donde el rango fue de un 4.5% a un 27.8% de disminución de sodio intracelular en los casos estadísticamente significativos. Ambos experimentos no son comparables, ya que en el caso de CRISPRi se están inhibiendo regiones específicas del genoma en un *background* genético sano. Sin embargo, estos resultados indican que la comparación de hiPSC-CMs con un *background* genético asociado a una enfermedad cardíaca (SBr) con hiPSC-CMs control presenta diferencias más contundentes que con el método CRISPRi.

Nuestros resultados también muestran que los niveles de mRNA de otros genes previamente asociados a SBr como *CACNA1C*, *TTN*, *SEMA3A* y *SCNN1A* también presentan alteraciones importantes entre GPG2 y SBr. En el caso de ***SCNN1A***, se observó un aumento de la expresión en los hiPSC-CMs SBr respecto a GPG2. Este gen codifica para la subunidad  $\alpha$  del canal de sodio epitelial. En este sentido, el efecto observado podría ser debido a un mecanismo compensatorio causado por la alteración de otros canales de sodio, como *SCN5A*. Los cambios en el perfil de expresión génica en respuesta a los cambios en las corrientes de canales iónicos se han descrito previamente para otras canalopatías y se considera que es una respuesta celular compensatoria<sup>401,462</sup>. En línea con esta hipótesis, nuestros resultados también mostraron alteraciones relevantes en los niveles de mRNA, como la disminución en la expresión de ***NKX2-5*** en las células SBr respecto a GPG2. *NKX2-5* codifica para la proteína homeobox 5 y actúa como activador transcripcional. Aunque previamente no se ha asociado con el SBr, el FT *NKX2-5* es clave para muchos procesos cardíacos y regula los canales iónicos codificados por los genes *SCN5A*, *CACNA1C* y *KCNH2*<sup>249,471</sup>. En la misma línea, los resultados de CRISPRi de esta tesis también sugieren que *NKX2-5* participaría en la regulación transcripcional de *SCN5A* (gRNA **Intrón 1-NKX2-5**). En base a estas observaciones, podemos hipotetizar que las alteraciones en la expresión de los canales iónicos y *NKX2-5* podrían estar correlacionadas. De hecho, *NKX2-5* es esencial para establecer el programa transcripcional necesario para el desarrollo muscular cardíaco. Modelos de hiPSC-CMs *knock-out* para *NKX2-5* han demostrado

una diferenciación cardíaca defectuosa, asociada con una disminución en la expresión de genes de canales iónicos y contractilidad reducida<sup>467</sup>.

Finalmente, es importante destacar la observación de diferencias en la expresión de genes estructurales en las células SBr respecto a GPG2. Tres de estos genes se incluyen entre los principales DEGs del RNA-seq (*ACT1*, *MYH6* o *TTN*). La visión de la genética del SBr ha evolucionado recientemente para reconocer que, en la mayoría de los casos, el SBr es un síndrome que puede resultar del efecto combinado de múltiples variantes, incluso en múltiples genes, presentes en una misma persona (síndrome oligogénico)<sup>405,406</sup>, y se están desarrollando un número importante de estudios para determinar todos los genes asociados al SBr. Los genes candidatos comúnmente incluyen aquellos que codifican para canales de sodio, potasio y calcio<sup>249,407</sup>, así como, menos comúnmente, proteínas sarcoméricas y estructurales<sup>408,409</sup> y genes mitocondriales<sup>410</sup>. Por lo tanto, la aparición de DEGs con una función más estructural y que difiere de la definición clásica de canalopatía, podría ser plausible teniendo en cuenta la heterogeneidad genética que se puede dar en el SBr. En línea con estas observaciones, la revisión de Moncayo-Arlandi & Brugada discute el vínculo molecular del SBr con la miocardiopatía arritmogénica<sup>466</sup>. Los discos intercalares que conectan los cardiomiocitos controlan la adhesión y comunicación de célula a célula. Varias estructuras macromoleculares (desmosomas, uniones adherentes, *GAP junctions* y complejos de canales de sodio) coexisten en el disco intercalar y le confieren sus propiedades mecánicas y eléctricas. Tradicionalmente, se sospechaba/se trabajaba en la premisa que cada estructura tenía una función única en el disco intercalar. Sin embargo, existen evidencias que sugieren que estos complejos actúan de manera conjunta en la comunicación y adhesión intercelular, formando una única entidad estructural y funcional: el conexoma. Estos conceptos han dado apoyo conceptual a la superposición de SBr y la miocardiopatía arritmogénica, ambas basadas en la alteración del disco intercalar<sup>466</sup>.

El experimento de RNA-seq realizado en esta tesis tuvo ciertas limitaciones debido a una serie de factores experimentales que pueden haber afectado el número total de DEGs identificado:

1. En primer lugar, el estudio sólo incluyó la secuenciación de hiPSC-CMs de un individuo para cada grupo de estudio (SBr vs control), debido a que en el momento del estudio no se disponía de hiPSCs de más pacientes con SBr y Hap<sup>1/1</sup>. Es importante tener en cuenta que a diferencias en la expresión de algunos genes identificadas podrían ser debidas a la variabilidad interindividual no relacionada con la enfermedad. En la última década, muchos estudios han establecido que los patrones de expresión génica varían entre individuos, entre tipos de tejido y entre células isogénicas en un entorno homogéneo<sup>463,464</sup>. Estas diferencias de la expresión conducen a la variabilidad fenotípica en la población. También se ha observado que los genes con alta variabilidad de expresión en la población probablemente

tengan una expresión dependiente de la edad, pero no del sexo<sup>465</sup>. En nuestro caso, las líneas celulares SBr y GPG2 provienen de dos individuos de sexo masculino de 68 y 55 años, respectivamente. Wang *et al.*, en un estudio similar al nuestro, realizó un análisis RNA-seq a partir de hiPSC-CMs de un solo individuo con SBr y un solo individuo control. En este caso, también obtuvo una cifra elevada de DEGs (>2000)<sup>411</sup>.

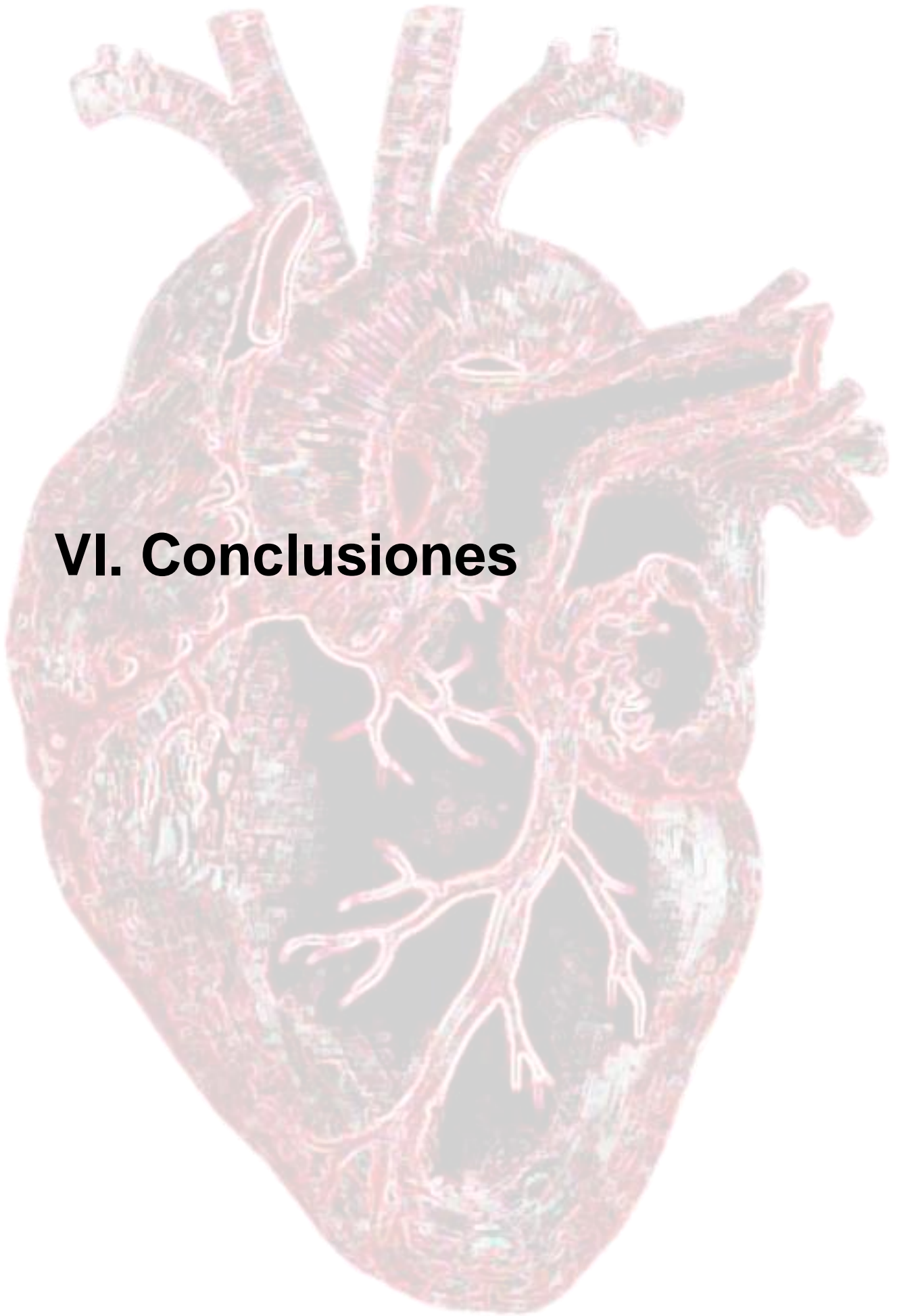
Para solventar esta limitación se necesitarían una *n* elevada para cada grupo de estudio (SBr y control). Por contra, conseguir hiPSC de muchos individuos y específicos para la condición requerida es un proceso complejo y asociado a un elevado coste.

2. Variabilidad inherente al proceso experimental de diferenciación a hiPSC-CMs. A pesar de que el protocolo de diferenciación a hiPSC-CMs está bien establecido, los resultados obtenidos al realizar las diferenciaciones suelen ser heterogéneos. La calidad de una ronda de diferenciación puede depender de muchos factores difíciles de controlar y que actúan simultáneamente, como el número de pases de la línea celular o la propia manipulación de las hiPSCs. Además, puede existir también variabilidad entre pocillos de una misma ronda de diferenciación, por lo que probablemente haya un factor de aleatoriedad en el éxito o el fracaso de este proceso. A lo largo de esta tesis solamente se usaron rondas de diferenciación que llegasen al final del proceso con la calidad fenotípica que marca el protocolo.
3. Clasificación de los DEGs obtenidos. A partir de nuestro análisis con cuatro paquetes de detección de expresión diferencial se obtuvieron >3000 DEGs, una cifra difícil de manejar y valorar cuales de esos genes estaban o no involucrados con el SBr.

Además de estos factores, tampoco se pudo discernir si los DEGs más relevantes del estudio podrían estar asociados al Hap<sup>1/1</sup> o simplemente al SBr. Para determinarlo, sería necesario añadir un nuevo grupo de estudio para tener pacientes con SBr, pero sin el haplotipo de riesgo. Una posible solución tanto a los problemas de variabilidad intraindividual como a la relación entre los resultados y el haplotipo es la creación de una línea celular isogénica. En este sentido, el último objetivo de esta tesis consistió en realizar una edición genómica a la línea celular de SBr Hap<sup>1/1</sup> en una línea Hap<sup>3/3</sup>. Esto permitiría obtener una línea con un *background* idéntico al del paciente con SBr, pero sin el haplotipo de riesgo. La realización de un RNA-seq mediante la línea SBr y la línea modificada isogénica esclarecería el perfil transcripcional asociado al haplotipo de riesgo. El estudio de haplotipos intermedios también resulta interesante, con el fin de determinar el efecto individual de cada variante del haplotipo. Finalizada esta tesis, se dispone de una línea de hiPSCs con el haplotipo modificado en la posición rs1 (rs6801957), lo que dio lugar a un haplotipo intermedio, concretamente Hap<sup>9/9</sup>. Este haplotipo no existe en la cohorte de SBr, por lo que su estudio podría determinar si un solo cambio en el haplotipo de riesgo (en este caso la posición rs1), es suficiente para obtener una mejora del fenotipo celular asociado al SBr.



## **VI. Conclusiones**







1. Los experimentos de IP y WB demuestran que K318, K322 y K326, situadas en la región de unión al DNA, son dianas de acetilación de GATA4 por p300.
2. Ensayos luciferasa en células H9c2 expresando vectores de expresión de GATA4 indican que la acetilación del residuo K318 de GATA4 es importante por su actividad transcripcional sobre los promotores *SCN5A* y *PDX1*. Los resultados también sugieren que la acetilación de los residuos K322 y K326 no estaría relacionada con un incremento de la actividad transcripcional de GATA4.
3. El sistema CRISPRi desarrollado en esta tesis permite evaluar la implicación de regiones no codificantes en la regulación de la expresión génica. Mediante este sistema, es posible evaluar la función reguladora de secuencias proximales o distales al gen de interés. También permite evaluar alteraciones funcionales en los niveles de sodio intracelular en hiPSC-CMs.
4. Utilizando el sistema CRISPRi en hiPSC-CMs se han identificado siete regiones en el locus *SCN5A-SCN10A* asociadas a la regulación transcripcional del gen *SCN5A*.
5. Nuestro estudio describe, por primera vez, que la región del promotor de *EXOG* actúa como región reguladora de la expresión del gen *SCN5A*. La inhibición de esta región con Cas9, además, altera los niveles de sodio intracelular en el modelo de hiPSC-CMs.
6. Mediante secuenciación de lecturas largas de Oxford nanopore se han validado experimentalmente los haplotipos del locus *SCN5A-SCN10A* identificados previamente con algoritmos de predicción en 83 de los 86 pacientes de la cohorte de SBr.
7. El análisis de RNA-seq ha identificado 3176 genes diferencialmente expresados entre hiPSC-CMs de un paciente con SBr y Hap<sup>1/1</sup> respecto a un control sano con el haplotipo Hap<sup>1/3</sup>. De ellos, los diez genes más relevantes están mayoritariamente relacionados con canales iónicos de sodio, calcio o potasio, factores de transcripción cardíacos y genes estructurales, siendo el más relevante el gen *SCN5A*.
8. La línea de hiPSC-CMs SBr (Hap<sup>1/1</sup>) presentó una disminución del 36.5% en sus niveles de sodio intracelular respecto a la línea de hiPSC-CMs GPG2 (Hap<sup>1/3</sup>).
9. Se ha desarrollado un protocolo de edición genómica CRISPR/Cas9 con el que se han generado homocigotos *knock-in* en hiPSCs con eficiencias de hasta el 60.4% de actividad HDR.

10. Se ha realizado de forma satisfactoria una edición genómica de la variante rs1 del haplotipo Hap<sup>1/1</sup> en la línea hiPSC de SBr obteniendo una nueva línea hiPSC con el haplotipo Hap<sup>9/9</sup>.







## **VII. Bibliografía**



1. Marban, E. Cardiac channelopathies. *Nature* **415**, 213-218 (2002).
2. OpenStax CNX. *Anatomy & Physiology*. Disponible en: [https://cnx.org/contents/FPtK1z mh@9.1:Y5T\\_wVSC@4/Heart-Anatomy](https://cnx.org/contents/FPtK1z mh@9.1:Y5T_wVSC@4/Heart-Anatomy).
3. Betts, J. G. et al. *Anatomy & Physiology*. (OpenStax, 2013).
4. Zimmer, T. et al. The beta1 subunit but not the beta2 subunit colocalizes with the human heart Na<sup>+</sup> channel (hH1) already within the endoplasmic reticulum. *J. Membr. Biol.* **186**, 13-21 (2002).
5. Garg, V. et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443-447 (2003).
6. Sands, Z. et al. Voltage-gated ion channels. *Curr. Biol.* **15**, R44-R47 (2005).
7. Bezanilla, F. Voltage-gated ion channels. *IEEE Trans. Nanobioscience* **4**, 34-48 (2005).
8. Pinsach-Abuin, M. Regulome-seq: a Novel Approach for the Identification of Non-coding Variants Associated with Human Disease. Assessment of its Applicability in 89 Brugada syndrome Individuals. Tesis doctoral, *Universitat de Girona* (2019). Repositorio de la tesis: <http://hdl.handle.net/10803/666922>.
9. Wood, J. N. & Iseppon, F. Sodium channels. *Brain Neurosci. Adv.* **2** (2018).
10. Catterall, W. A. et al. International Union of Pharmacology. XLVII. Nomenclature and structure function relationships of voltage-gated sodium channels. *Pharmacol. Rev.* **57**, 397-409 (2005).
11. Cusdin, F. S. et al. Trafficking and cellular distribution of voltage-gated sodium channels. *Traffic* **9**, 17-26 (2008).
12. Brackenbury, W. J. & Isom, L. L. Na Channel  $\beta$  Subunits: Overachievers of the Ion Channel Family. *Front. Pharmacol.* **2**, 53 (2011).
13. Bouza, A. A. & Isom, L. L. Voltage-gated sodium channel  $\beta$  subunits and their related diseases. *Handb. Exp. Pharmacol.* **246**, 423-450 (2018).
14. Veerman, C. C. et al. The cardiac sodium channel gene SCN5A and its gene product NaV1.5: Role in physiology and pathophysiology. *Gene* **573**, 177-187 (2015).
15. Hull, J. M. & Isom, L. L. Voltage-gated sodium channel  $\beta$  subunits: The power outside the pore in brain development and disease. *Neuropharmacology* **132**, 43-57 (2018).
16. Jiang, D. et al. Structure of the Cardiac Sodium Channel. *Cell* **180**, 122-134 (2020).
17. Salvage, S. C. et al. Cell-Adhesion Properties of  $\beta$ -Subunits in the Regulation of Cardiomyocyte Sodium Channels. *Biomolecules* **10**, 989 (2020).
18. Tarradas, A. Els factors GATA4 i GATA5 en la regulació transcripcional del gen que codifica pel canal de sodi cardíac (SCN5A). Tesis doctoral, *Universitat de Girona* (2016). Repositorio de la tesis: <http://hdl.handle.net/10803/405732>.
19. Catterall, W. A. Voltage-gated calcium channels. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
20. Hille, B. *Ion Channels of Excitable Membranes*, Third Edition. (Sinauer Associates, 2001).
21. Pongs, O. Ins and outs of cardiac voltage-gated potassium channels. *Curr. Opin. Pharmacol.* **9**, 311-315 (2009).
22. Vacher, H. et al. Localization and targeting of voltage-dependent ion channels in mammalian central neurons. *Physiol. Rev.* **88**, 1407-1447 (2008).
23. González, C. et al. K(+) channels: function-structural overview. *Compr. Physiol.* **2**, 2087-2149 (2012).
24. Mangoni, M. E. & Nargeot, J. Genesis and Regulation of the Heart Automaticity. *Physiol. Rev.* **88**, 919-982 (2008).
25. Amin, A. S. et al. Cardiac ion channels in health and disease. *Heart Rhythm* **7**, 117-126 (2010).
26. Nerbonne, J. M. & Kass, R. S. Molecular physiology of cardiac repolarization. *Physiol. Rev.* **85**, 1205-1253 (2005).
27. Keating, M. T. & Sanguinetti, M. C. Molecular and cellular mechanisms of cardiac arrhythmias. *Cell* **104**, 569-580 (2001).



28. Wickenden, A. D., *et al.* The role of action potential prolongation and altered intracellular calcium handling in the pathogenesis of heart failure. *Cardiovasc. Res.* **37**, 312–323 (1998).
29. Campuzano, O. *et al.* Genetics and cardiac channelopathies. *Genet. Med.* **12**, 260–267 (2010).
30. Adabag, A. S. *et al.* Sudden cardiac death: Epidemiology and risk factors. *Nat. Rev. Cardiol.* **7**, 216–225 (2010).
31. Wong, C. X. *et al.* Epidemiology of Sudden Cardiac Death: Global and Regional Perspectives. *Hear. Lung Circ.* **28**, 6–14 (2019).
32. Paratz, E. D. *et al.* Cardiac arrest and sudden cardiac death registries: a systematic review of global coverage. *Open Hear.* **7**, 1–10 (2020).
33. Khan, H. M. & Leslie, S. J. Risk factors for sudden cardiac death to determine high risk patients in specific patient populations that may benefit from a wearable defibrillator. *World J. Cardiol.* **11**, 103–119 (2019).
34. Hayashi, M. *et al.* The Spectrum of Epidemiology Underlying Sudden Cardiac Death. *Circ. Res.* **116**, 1887–1906 (2015).
35. Chugh, S. S. *et al.* Epidemiology of Sudden Cardiac Death: Clinical and Research Implications. *Prog. Cardiovasc. Dis.* **51**, 213–228 (2008).
36. Priori, S. G. *et al.* Task force on sudden cardiac death, European Society of Cardiology: Summary of recommendations. *Europace* **4**, 3–18 (2002).
37. Deo, R. & Albert, C. M. Epidemiology and genetics of sudden cardiac death. *Circulation* **125**, 620–637 (2012).
38. Friedlander, Y. *et al.* Family history as a risk factor for primary cardiac arrest. *Circulation* **97**, 155–160 (1998).
39. Albert, C. M. *et al.* Prospective study of sudden cardiac death among women in the United States. *Circulation* **107**, 2096–2101 (2003).
40. Burke, M. A. *et al.* Clinical and mechanistic insights into the genetics of cardiomyopathy. *J. Am. Coll. Cardiol.* **68**, 2871–2886 (2016).
41. Coronary Heart Disease Research. *National heart, lung and blood institute*. Disponible en: <https://www.nhlbi.nih.gov/health/coronary-heart-disease>.
42. Wexler, R. K. *et al.* Cardiomyopathy: an overview. *Am. Fam. Physician* **79**, 778–784 (2009).
43. Magi, S. *et al.* Sudden cardiac death: focus on the genetics of channelopathies and cardiomyopathies. *J. Biomed. Sci.* **24**, 56 (2017).
44. Fernández-Falgeras, A. *et al.* Cardiac Channelopathies and Sudden Death: Recent Clinical and Genetic Advances. *Biology (Basel)* **6**, 7 (2017).
45. Campuzano, O. *et al.* Genetics and cardiac channelopathies. *Genet. Med.* **12**, 260–7 (2010).
46. Kim, J. B. Channelopathies. *Korean J. Pediatr.* **57**, 1–18 (2014).
47. Ackerman, M. J. *et al.* HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies. *Europace* **13**, 1077–1109 (2011).
48. Zipes, D. P. *et al.* ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: A report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Com. *Circulation* **114** (2006).
49. Gollob, M. H. *et al.* Recommendations for the use of genetic testing in the clinical evaluation of inherited cardiac arrhythmias associated with sudden cardiac death: Canadian cardiovascular society/Canadian heart rhythm society joint position paper. *Can. J. Cardiol.* **27**, 232–245 (2011).
50. van den Boogaard, M. *et al.* Genetic variation in T-box binding element functionally affects *SCN5A/SCN10A* enhancer. *J. Clin. Invest.* **122**, 2519–2530 (2012).
51. van den Boogaard, M. *et al.* A common genetic variant within *SCN10A* modulates cardiac *SCN5A* expression. *J. Clin. Invest.* **124**, 1844–1852 (2014).

52. Schwartz P. J. *et al.* Prevalence of the Congenital Long-QT Syndrome. *Circulation* **120**, 1761–1767 (2009).
53. Roden D. M. Long-QT Syndrome. *N. Engl. J. Med.* **358**, 169–176 (2008).
54. Zipes, D.P. The Long QT Interval Syndrome. A Rosetta Stone for Sympathetic related ventricular tachyarrhythmias. *Circulation* **84**, 1414–1419 (1991).
55. Mizusawa, Y. & Horie, M. Characteristics C. Genetic and Clinical Advances in Congenital Long QT Syndrome. *Circulation* **78**, 2827–2833 (2014).
56. Ceyhan-Birsoy, O. *et al.* Next generation sequencing-based copy number analysis reveals low prevalence of deletions and duplications in 46 genes associated with genetic cardiomyopathies. *Mol. Genet. Genomic Med.* **4**, 143–151 (2016).
57. Morita, H. *et al.* The QT syndromes: long and short. *Lancet.* **372**, 750–763 (2008).
58. Guerrier, K. *et al.* Short QT Interval Prevalence and Clinical Outcomes in a Pediatric Population. *Circ. Arrhythm. Electrophysiol.* **8**, 1460–1464 (2015).
59. Crotti, L. *et al.* Congenital Short QT Syndrome. *Indian Pacing. Electrophysiol. J.* **10**, 86–95 (2010).
60. Patel, C. *et al.* Basic Science for the Clinical Electrophysiologist Short QT Syndrome: From Bench to Bedside. *Circ. Arrhythm. Electrophysiol.* **3**, 401–408 (2010).
61. Rudic, B. *et al.* Short QT Syndrome – Review of Diagnosis and Treatment. *Arrhythm. Mech.* **3**, 76–79 (2014).
62. Priori, S. G. *et al.* Clinical and Molecular Characterization of Patients With Catecholaminergic Polymorphic Ventricular Tachycardia. *Circulation* **106**, 69–74 (2002).
63. Katz, G. *et al.* Catecholaminergic Polymorphic Ventricular Tachycardia from Bedside to Bench and Beyond. *Curr. Probl. Cardiol.* **34**, 9–43 (2009).
64. Leenhardt, A. *et al.* Catecholaminergic Polymorphic Ventricular Tachycardia. *Circ. Arrhythm. Electrophysiol.* **5**, 1044–1052 (2011).
65. Priori, S. G. *et al.* Mutations in the Cardiac Ryanodine Receptor Gene (hRyR2) Underlie Catecholaminergic Polymorphic Ventricular Tachycardia. *Circulation* **103**, 196–200 (2001).
66. Berne, P. & Brugada, J. Brugada Syndrome. *Circulation* **76**, 1563–1571 (2012).
67. Remme C. A. *et al.* Overlap Syndrome of Cardiac Sodium Channel Disease in Mice Carrying the Equivalent Mutation of Human. *Circulation* **114**, 2584–2595 (2006).
68. Brugada, R. *et al.* Sodium Channel Blockers Identify Risk for Sudden Death in Patients With ST-Segment Elevation and Right Bundle Branch Block but Structurally Normal Hearts. *Circulation* **101**, 510–515 (2000).
69. Miescher, F. Die Histochemischen und physiologischen. (Vogel, 1897).
70. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribosenucleic Acid. *Nature* **171**, 737–738 (1953).
71. Lander E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
72. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
73. Ezkurdia, I. *et al.* Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866–5878 (2014).
74. Leslie, A. *et al.* Eukaryotic Genome Complexity. *Nat. educat.* **1**, 96 (2008).
75. Alberts, B. *et al.* Molecular biology of the cell. (Garland Science, 2008).
76. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–R29 (2006).
77. Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
78. Ala, U. Competing Endogenous RNAs, Non-Coding RNAs and Diseases: An Intertwined Story. *Cells* **9**, 1574 (2020).
79. Jo, B-S. & Choi, S. S. Introns: The Functional Benefits of Introns in Genomes. *Genomics Inform.* **13**, 112–118 (2015).

80. Chi K. The dark side of the human genome. *Nature* **538**, 275–277 (2016).
81. Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).
82. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
83. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2012).
84. Sun, M. *et al.* Systematic functional interrogation of human pseudogenes using CRISPRi. *Genome Biol.* **22**, 240 (2021).
85. Xiao-Jie, L. *et al.* Pseudogene in cancer: real functions and promising signature. *Med. Genet.* **0**, 1–8 (2014).
86. Vanin, E. F. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272 (1985).
87. Treangen, T. & Salzberg, S. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev.* **13**, 36–46 (2012).
88. Biscotti, M. A. *et al.* Repetitive DNA in eukaryotic genomes. *Chromosome Res.* **23**, 415–420 (2015).
89. Zhang, F. *et al.* Copy Number Variation in Human Health, Disease, and Evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–481 (2009).
90. Patrushev, L. & Minkevich, I. The Problem of the Eukaryotic Genome Size. *Biochemistry.* **73**, 1519–1552 (2008).
91. Pace, J. & Feschotte, C. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**, 422–432 (2007).
92. Han J. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob. DNA* **1**, 1–12 (2010).
93. Ponicsan, S. L. *et al.* Genomic gems- SINE RNAs regulate mRNA production. *Curr. Opin. Genet. Dev.* **20**, 149–155 (2011).
94. Ran, F. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
95. Gasperskaja, E. & Kučinskas, V. The most common technologies and tools for functional genome analysis. *Acta Med. Litu.* **24**, 1-11 (2017).
96. Cookson, W. *et al.* Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
97. Lee, T. I. & Young, R. A. Transcriptional Regulation and Its Misregulation in Disease. *Cell* **152**, 1237–1251 (2013).
98. Keung, A. J. *et al.* Chromatin regulation at the frontier of synthetic biology. *Nat. Rev. Genet.* **16**, 159–171 (2015).
99. Dixon, J. R. *et al.* Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62**, 668–680 (2016).
100. O'Connor, T. *et al.* CisMapper: predicting regulatory interactions from transcription factor ChIP-seq data. *Nucleic Acids Res.* **45**, e19–e19 (2017).
101. Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479 (2003).
102. Gershenzon, N. I. & Ioshikhes, I. P. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* **21**, 1295–1300 (2005).
103. Morris, J. R. *et al.* Enhancer choice in cis and in trans in *Drosophila melanogaster*: role of the promoter. *Genetics* **167**, 1739–1747 (2004).
104. Maston, G. A. *et al.* Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).

105. Yang, C. *et al.* Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52–65 (2007).
106. Bulger, M. & Groudine, M. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev. Biol.* **339**, 250–257 (2010).
107. Atchison, M. L. Enhancers: mechanisms of action and cell specificity. *Annu. Rev. Cell Biol.* **4**, 127–153 (1988).
108. Mora, A. *et al.* In the loop: promoter–enhancer interactions and bioinformatics. *Brief. Bioinform.* **17**, 980–995 (2016).
109. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
110. Sagai, T. *et al.* Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).
111. Shim, S. *et al.* Cis-regulatory control of corticospinal system development and evolution. *Nature* **486**, 74–79 (2012).
112. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).
113. Perry, M. W. *et al.* Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
114. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Publ. Gr.* **47**, 8–12 (2015).
115. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
116. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
117. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
118. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2014).
119. Zhang, J. *et al.* Super enhancers—Functional cores under the 3D genome. *Cell Prolif.* **54**, e12970 (2021).
120. Kohli, S. *et al.* Transcription factors in heart: promising therapeutic targets in cardiac hypertrophy. *Curr. Cardiol. Rev.* **7**, 262–271 (2011).
121. Luse, D. S. The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? *Transcription* **5** (2013).
122. Knuesel, M. T. *et al.* The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes Dev.* **23**, 439–451 (2009).
123. Reynolds, N. *et al.* Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140**, 505–512 (2013).
124. Deckert, J. & Struhl, K. Histone acetylation at promoters is differentially affected by specific activators and repressors. *Mol. Cell. Biol.* **21**, 2726–2735 (2001).
125. Stefanovic, S. *et al.* GATA-dependent regulatory switches establish atrioventricular canal specificity during heart development. *Nat. Commun.* **5**, 3680 (2014).
126. Zhou, P. *et al.* Regulation of GATA4 transcriptional activity in cardiovascular development and disease. *Develop. Biol.* **100**, 143–169 (2012).
127. Bruneau, B. G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
128. Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
129. Schlesinger, J. *et al.* The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS genetics* **7**, e1001313 (2011).

130. Srivastava, D & Olson, E. N. A genetic blueprint for cardiac development. *Nature* **407**, 221–226 (2000).
131. He, A. *et al.* Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl. Acad. Sci. USA* **108**, 5632–5637 (2011).
132. Lyons, I. *et al.* Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene *Nkx2-5*. *Genes Dev.* **9**, 1654–1666 (1995).
133. Tanaka, M. *et al.* The cardiac homeobox gene *Csx/Nkx2.5* lies genetically upstream of multiple genes essential for heart development. *Development* **126**, 1269–1280 (1999).
134. Schott, J. J. *et al.* Congenital heart disease caused by mutations in the transcription factor *NKX2-5*. *Science* **281**, 108–111 (1998).
135. Furtado, M. B. *et al.* A novel conditional mouse model for *Nkx2-5* reveals transcriptional regulation of cardiac ion channels. *Differentiation* **91**, 29–41 (2016).
136. Briggs, L. E., *et al.* Perinatal loss of *Nkx2-5* results in rapid conduction and contraction defects. *Circ. Res.* **103**, 580–590 (2008).
137. Lin, Q. *et al.* Control of mouse cardiac morphogenesis and myogenesis by transcription factor *MEF2C*. *Science* **276**, 1404–1407 (1997).
138. Edmondson, D. G. *et al.* *Mef2* gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* **120**, 1251–1263 (1994).
139. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, *dHAND*. *Nat. Genet.* **16**, 154–160 (1997).
140. George, R. M. *et al.* Neonatal Deletion of *Hand1* and *Hand2* within Murine Cardiac Conduction System Reveals a Novel Role for *HAND2* in Rhythm Homeostasis. *J. Cardiovasc. Dev. Dis.* **9**, 214 (2022).
141. Ostadal, B. & Dhalla, N. S. Cardiac Adaptations: Molecular Mechanisms. *Advances in Biochemistry in Health and Disease.* (Springer-Verlag, 2013).
142. Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor *Tbx5* in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).
143. Arnolds, D.E. *et al.* *TBX5* drives *Scn5a* expression to regulate cardiac conduction system function. *J. Clin. Invest.* **122**, 2509–2518 (2012).
144. Pikkarainen, S. *et al.* GATA transcription factors in the developing and adult heart. *Cardio. Res.* **63**, 196–207 (2004).
145. Lowry, J. A. & Atchley, W. R. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *J. Mol. Evol.* **50**, 103–115 (2000).
146. Aronson BE, Stapleton KA, Krasinski SD. Role of GATA factors in development, differentiation, and homeostasis of the small intestinal epithelium. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306**, G474–490 (2014).
147. Morrisey, E. E. *et al.* GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *J. Biol. Chem.* **272**, 8515–8524 (1997).
148. Gao, J. *et al.* GATA family transcriptional factors: emerging suspects in hematologic disorders. *Exp. Hematol. Oncol.* **4**, 28 (2015).
149. Molkentin, J. D. *et al.* Direct activation of a GATA6 cardiac enhancer by *Nkx2.5*: evidence for a reinforcing regulatory network of *Nkx2.5* and GATA transcription factors in the developing heart. *Develop. Biol.* **217**, 301–309 (2000).
150. Ferreira, R. *et al.* GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell. Biol.* **25**, 1215–1227 (2005).
151. Charron, F. & Nemer, M. GATA transcription factors and cardiac development. *Semin. Cell Dev. Biol.* **10**, 85–91 (1999).
152. Chen, Y. *et al.* DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep.* **2**, 1197–1206 (2012).

153. Peterkin, T. *et al.* Redundancy and evolution of GATA factor requirements in development of the myocardium. *Develop. Biol.* **311**, 623-635 (2007).
154. Al Sayed, Z. R. *et al.* Human model of IRX5 mutations reveals key role for this transcription factor in ventricular conduction. *Cardiovasc. Res.* **117**, cvaa259 (2021).
155. Canac, R. *et al.* Deciphering Transcriptional Networks during Human Cardiac Development. *Cells.* **11**, 3915 (2022).
156. Luna-Zurita, L. *et al.* Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell* **164**, 999–1014 (2016).
157. Tarradas, A. *et al.* Transcriptional regulation of the sodium channel gene (SCN5A) by GATA4 in human heart. *J. Mol. Cell. Cardiol.* **102**, 74-82 (2017).
158. Kouzarides T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
159. Allis, C. D. *et al.* New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636 (2007).
160. Parthun, M. R. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**, 5319-5328 (2007).
161. Yanazume, T. *et al.* Cardiac p300 is involved in myocyte growth with decompensated heart failure. *Mol. Cell. Biol.* **23**, 3593-3606 (2003).
162. Kakita, T. *et al.* p300 protein as a coactivator of GATA-5 in the transcription of cardiac-restricted atrial natriuretic factor gene. *J. Biol. Chem.* **274**, 34096-34102 (1999).
163. Takaya, T. *et al.* Identification of p300-targeted acetylated residues in GATA4 during hypertrophic responses in cardiac myocytes. *J. Biol. Chem.* **283**, 9828-983 (2008).
164. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907 (2014).
165. Yilbas, A. E. *et al.* Activation of GATA4 gene expression at the early stage of cardiac specification. *Front. Chem.* **2**, 12 (2014).
166. Trivedi, C. M. *et al.* Hopx and Hdac2 interact to modulate Gata4 acetylation and embryonic cardiac myocyte proliferation. *Dev. Cell* **19**, 450-459 (2010).
167. Kawamura T. *et al.* Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes. *J. Biol. Chem.* **280**, 19682-19688 (2005).
168. Morales, V. *et al.* Chromatin structure and dynamics: functional implications. *Biochimie* **83**, 1029–1039 (2001).
169. Cheung, P. *et al.* Signaling to chromatin through histone modifications. *Cell* **103**, 263-271 (2000).
170. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).
171. Orlandi, K. & McKnight, J. Bulky Histone Modifications May Have an Oversized Role in Nucleosome Dynamics. *Bioessays* **42**, e1900217 (2020).
172. Hayes, J. J. & Wolffe, A. P. The interaction of transcription factors with nucleosomal DNA. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **14**, 597–603 (1992).
173. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
174. Black, J. C. & Whetstine J. R. Tipping the lysine methylation balance in disease. *Biopolymers* **99**, 127-135 (2013).
175. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* **13**, 37-50 (2013).
176. Margueron, R. *et al.* The key to development: Interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
177. Shi, Y. & Whetstine, J.R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell.* **25**, 1–14 (2007).
178. Shi, Y. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2014).

179. Wissmann, M. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell. Biol.* **9**, 347–353 (2007).
180. Cloos, P. A. *et al.* Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140 (2008).
181. D'Oto, A. *et al.* Histone demethylases and their roles in cancer epigenetics. *J. Med. Oncol. Ther.* **1**, 34–40 (2016).
182. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
183. Petty, E. & Pillus, L. Balancing chromatin remodeling and histone modifications in transcription. *Trends Genet.* **29**, 621–629 (2013).
184. Oftedal, G. The Metaphorical Role of the Histone Code. *Metaph. Anal. Sci. Hum.* **453**, 253–267 (2022).
185. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
186. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
187. Geahigan, K. B. *et al.* The dynamic impact of CpG methylation in DNA. *Biochemistry* **39**, 4939–4946 (2000).
188. Derreumaux, S. *et al.* Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element. *Nucleic Acids Res.* **29**, 2314–2326 (2001).
189. Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).
190. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–30 (2015).
191. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
192. Ji, X. *et al.* 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
193. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
194. Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678–1689 (2013).
195. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).
196. Peters, J-M., *et al.* The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22**, 3089–3114 (2008).
197. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456–E6465 (2015).
198. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
199. Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep* **17**, 2042–2059 (2016).
200. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
201. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* **453**, 948–51 (2008).
202. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299–312 (2007).
203. Zentner, G. E. *et al.* Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21**, 1273–83 (2011).

204. Hon, G. C. *et al.* Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **18**, R195-201 (2009).
205. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
206. Natarajan, A. *et al.* Predicting cell-type-specific gene expression from regions of open chromatin. *Genome Res.* **22**, 1711–1722 (2012).
207. Qiu, Z. *et al.* Identification of regulatory DNA elements using genome-wide mapping of DNase I hypersensitive sites during tomato fruit development. *Mol. Plant* **9**, 1168–1182 (2016).
208. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
209. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* **121**, 1876–1886 (2010).
210. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cisregulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
211. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87-D92 (2020).
212. Buenrostro, J.D. *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213-1218 (2013).
213. Gilmour, D. S. & Lis, J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4275–4279 (1984).
214. Ecker, J. R. *et al.* Genomics: ENCODE explained. *Nature* **489**, 52–55 (2012).
215. Schmitt, A. D. *et al.* Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755 (2016).
216. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
217. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–597 (2015).
218. Mifsud, B. *et al.* High-resolution capture Hi-C to map long-range promoter contacts in human cells. *Nat. Genet.* **47**, 598–606 (2015).
219. Li, G. *et al.* ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**, R22 (2010).
220. Mandegar, M. A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541-553 (2016).
221. Spencer, D. H. *et al.* in *Clinical Genomics* (eds. Kulkarni, S. & Pfeifer, J.) 109–127 (Academic Press, 2015).
222. The 1000 Genomes projec Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
223. Lin, M. *et al.* Effects of short indels on protein structure and function in human genomes. *Sci. Rep.* **7**, 9313 (2017).
224. Girirajan, S. *et al.* Human copy number variation and complex genetic disease. *Annu. Rev. Genet.* **45**, 203–226 (2011).
225. Escaramís, G. *et al.* A decade of structural variants: description, history and methods to detect structural variation. *Brief. Funct. Genomics* **14**, 305–314 (2015).
226. Griffiths, A. J. F. *et al.* An introduction to genetic analysis - Induced mutations (W. H. Freeman, 2000).
227. Genetics. *Neupsy Key*. Disponible en: <https://neupsykey.com/genetics-3>



228. Sehn, J. K. in *Clinical Genomics* (eds. Kulkarni, S. & Pfeifer, J.) 129–150 (Academic Press, 2015).
229. Khan Academy Courses. Available at:  
<https://www.khanacademy.org/science/biology/dnaas-the-genetic-material/>.
230. Cooper, G. M. in *The Cell - DNA Repair* (Sinauer Associates, 2000).
231. Molla, K. A. *et al.* Predictable NHEJ Insertion and Assessment of HDR Editing Strategies in Plants. *Front. Genome. Ed.* **4**, 825236 (2022).
232. Lupski, J. R. *et al.* Clan genomics and the complex architecture of human disease. *Cell* **147**, 32–43 (2011).
233. Schork, N. J. *et al.* Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).
234. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primers* **1**, 59 (2021).
235. Zuk, O., *et al.* The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1193–1198 (2012).
236. The International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
237. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
238. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
239. Ahn, S-M. *et al.* The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.* **19**, 1622–1629 (2009).
240. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
241. Kim, J-I. *et al.* A highly annotated whole-genome sequence of a Korean individual. *Nature* **460**, 1011–1015 (2009).
242. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
243. Coventry, A. *et al.* Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* **1**, 131 (2010).
244. Adzhubei, I. *et al.* Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **7**, 7.20 (2013).
245. Telenti, A. *et al.* Deep sequencing of 10,000 human genomes. *Proc. Natl. Acad. Sci. USA* **113**, 11901–11906 (2016).
246. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
247. Cerrone, M. *et al.* Beyond the one gene-one disease paradigm complex genetics and pleiotropy in inheritable cardiac disorders. *Circulation* **140**, 595–610 (2019).
248. Walsh, R. & Wilde, A. M. SCN5A variants in Brugada syndrome: True, true false, or false true. *J. Cardiovasc. Electrophysiol.* **30**, 128–131 (2019).
249. Monasky, M. M. *et al.* Brugada syndrome: Oligogenic or mendelian disease? *Int. J. Mol. Sci.* **21**, 1687 (2020).
250. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
251. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).
252. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
253. Sakabe, N. J. *et al.* Transcriptional enhancers in development and disease. *Genome Biol.* **13**, 1–11 (2012).
254. Spielmann, M. & Mundlos, S. Looking beyond the genes: The role of non-coding variants in human disease. *Hum. Mol. Genet.* **25**, R157–R165 (2016).

255. Khurana, E. *et al.* Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).
256. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
257. Waszak, S. M. *et al.* Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* **2015**, 162, 1039–1050.
258. De Gobbi, M. *et al.* A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* **312**, 1215–1217 (2006).
259. Arzate-Mejía, R. G. *et al.* In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nat. Commun.* **11**, 1–6 (2020).
260. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
261. Crotti, L. *et al.* NOS1AP is a genetic modifier of the long-QT syndrome. *Circulation* **120**, 1657–1663 (2009).
262. Kousi, M. & Katsanis, N. Genetic modifiers and oligogenic inheritance. *Cold Spring Harb. Perspect. Med.* **5**, a017145 (2015).
263. Machanick, P. & Bailey, T. L. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
264. Telenti, A. & di Iulio, J. Regulatory genome variants in human susceptibility to infection. *Hum. Genet.* **139**, 759–768 (2019).
265. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).
266. Farley, E. K. *et al.* Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc. Natl. Acad. Sci. USA* **113**, 6508–6513 (2016).
267. Lodder, E. M. & Bezzina, C. R. Arrhythmogenic Right Ventricular Cardiomyopathy: Growing Evidence for Complex Inheritance. *Circ. Cardiovasc. Genet.* **6**, 525–528 (2013).
268. Gourraud, J-B. *et al.* The Brugada Syndrome: A Rare Arrhythmia Disorder with Complex inheritance. *Front. Cardiovasc. Med.* **3**, 1–11 (2016).
269. Bezzina, C. *et al.* A single Na<sup>+</sup> channel mutation causing both long-QT and Brugada syndromes. *Circ. Res.* **85**, 1206–1213 (1999).
270. Awad, M. M. *et al.* Mechanisms of Disease: molecular genetics of arrhythmogenic right ventricular dysplasia / cardiomyopathy. *Nat. Clin. Pract.* **5**, 258–267 (2008).
271. Cobo-Marcos, M. *et al.* Usefulness of Genetic Testing for Hypertrophic Cardiomyopathy in Real-world Practice. *Rev. Esp. Cardiol.* **66**, 746–747 (2013).
272. Alfares, A. A. *et al.* Results of clinical genetic testing of 2,912 probands with hypertrophic cardiomyopathy: expanded panels offer limited additional sensitivity. *Genet Med.* **17**, 880–888 (2015).
273. Lopes, L.R. *et al.* Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. *J. Med. Genet.* **50**, 228–239 (2013).
274. Simpson, S. *et al.* Genetics of Human and Canine Dilated Cardiomyopathy. *Int. J. Genomics* **2015**, 1–13 (2015).
275. Duijvenboden, K. V. *et al.* Gene regulatory elements of the cardiac conduction system. *Brief Funct. Genomics* **13**, 28–38 (2013).
276. Bokil, N. J. *et al.* Molecular genetics of long QT syndrome. *Mol. Genet. Metab.* **101**, 1–8 (2010).
277. Daley, J. M. *et al.* DNA joint dependence of Pol X family polymerase action in nonhomologous end joining. *J. Biol. Chem.* **280**, 29030–29037 (2005).
278. McVey, M. & Lee, S. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529–538 (2008).
279. Burgess, D. J. Genomics: next generation sequencing for reference genomes. *Nat. Rev. Genet.* **19**, 125 (2018).

280. Harrington, C. T. *et al.* Fundamentals of pyrosequencing. *Arch. Pathol. Lab. Med.* **137**, 1296-1303 (2013).
281. Bao, S. *et al.* Evaluation of next-generation sequencing software in mapping and assembly. *J. Hum. Genet.* **56**, 406-414 (2011).
282. Emiyu, K. & Lelisa, K. Review on Illumina Sequencing Technology. *Austin J. Vet. Sci. Anim. Husb.* **9**, 1088 (2022).
283. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36**, 338-345 (2018).
284. Payne, A. *et al.* Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files. *Bioinformatics* **35**, 2193-2198 (2019).
285. Amarasinghe, S. L. *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
286. Quail, M. *et al.* A tale of three next generation sequencing platforms: comparison of Ion torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
287. Asimaki, A. *et al.* A New Diagnostic Test for Arrhythmogenic Right Ventricular Cardiomyopathy. *N. Engl. J. Med.* **360**, 1075-1084 (2009).
288. Lin, Y. L. *et al.* Comparison of GATK and DeepVariant by trio sequencing. *Sci Rep.* **12**, 1809 (2022).
289. Rodríguez, P. Edición de Genes de Tomate que Codifican Potenciales Factores Provirales para el Virus del Mosaico del Pepino Dulce. Tesis doctoral, *Universidad de Murcia* (2020). Repositorio de la tesis: <https://digitum.um.es/digitum/bitstream/10201/99544/1/TD-PRS-22092020.pdf>
290. Udaondo, Z. *et al.* Comparative Analysis of PacBio and Oxford Nanopore Sequencing Technologies for Transcriptomic Landscape Identification of *Penaeus monodon*. *Life (Basel)* **11**, 862 (2021).
291. Ljungström, V. Exploring next-generation sequencing in chronic lymphocytic leukemia. Tesis doctoral, *Uppsala Universitet* (2016).
292. Wick, R. R. *et al.* Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
293. Beaulaurier, J. *et al.* Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat. Commun.* **6**, 7438 (2015).
294. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One.* **16**, e0257521 (2021).
295. Watson, M. & Warr, A. Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* **37**, 124-126 (2019).
296. Bowden, R. *et al.* Sequencing of human genomes with nanopore technology. *Nat. Commun.* **10**, 1869 (2019).
297. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228-232 (2016).
298. Papadatos, G. A. *et al.* Slowed conduction and ventricular tachycardia after targeted disruption of the cardiac sodium channel gene *Scn5a*. *Proc. Natl. Acad. Sci. USA* **99**, 6210-6215 (2002).
299. Mantegazza, M. & Broccoli, V. SCN1A/Nav1.1 channelopathies: Mechanisms in expression systems, animal models, and human iPSC models. *Epilepsia* **60**, S25-S38 (2019).
300. Sendfeld, F. *et al.* Experimental models of Brugada syndrome. *Int. J. Mol. Sci.* **20**, 2123 (2019).
301. Nishida, K. *et al.* Canine model of Brugada syndrome using regional epicardial cooling of the right ventricular outflow tract. *J. Cardiovasc. Electrophysiol.* **15**, 936-941 (2004).
302. Morita, H. *et al.* Repolarization heterogeneity in the right ventricular outflow tract: Correlation with ventricular arrhythmias in Brugada patients and in an in vitro canine Brugada model. *Hear. Rhythm* **5**, 725-733 (2008).
303. Griffin, A. *et al.* Preclinical animal models for Dravet syndrome: Seizure phenotypes, comorbidities and drug screening. *Front. Pharmacol.* **9**, 1-15 (2018).

304. Schutte, S. S. *et al.* Model systems for studying cellular mechanisms of SCN1A-related epilepsy. *J. Neurophysiol.* **115**, 1755–1766 (2016).
305. Gurdon, J. B. *et al.* Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**, 64–65 (1958).
306. Kaufman, M. H. & Evans, M. J. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
307. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA* **78**, 7634–7638 (1981).
308. Cowan, C. A. *et al.* Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).
309. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
310. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 5858 (2007).
311. Mertens, J. *et al.* Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat. Rev. Neurosci.* **17**, 424–437 (2016).
312. Veerman, C. C. *et al.* Immaturity of Human Stem-Cell-Derived Cardiomyocytes in Culture: Fatal Flaw or Soluble Problem? *Stem Cells Dev.* **24**, 1–57 (2015).
313. Tohyama, S. *et al.* Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12**, 127–137 (2013).
314. Sharma, A. *et al.* Derivation of highly purified cardiomyocytes from human induced pluripotent stem cells using small molecule-modulated differentiation and subsequent glucose starvation. *J. Vis. Exp.* **2015**, 3–10 (2015).
315. Parikh, S. S. *et al.* Thyroid and Glucocorticoid Hormones Promote Functional T-Tubule Development in Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *Circ Res.* **121**, 1323–1330 (2017).
316. Liang, W. *et al.* Disease modeling of cardiac arrhythmias using human induced pluripotent stem cells. *Expert Opin. Biol. Ther.* **19**, 313–333 (2019).
317. Reilly, L. *et al.* Challenges and innovation: Disease modeling using human-induced pluripotent stem cell-derived cardiomyocytes. *Front. Cardiovasc. Med.* **9**, 966094 (2022).
318. Nasrallah, A. *et al.* CRISPR-Cas9 Technology for the Creation of Biological Avatars Capable of Modeling and Treating Pathologies: From Discovery to the Latest Improvements. *Cells.* **11**, 3615 (2022).
319. Gilbert, L.A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2013).
320. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
321. Brugada, P. & Brugada, J. Right bundle branch block, persistent ST segment elevation and sudden cardiac death: A distinct clinical and electrocardiographic syndrome. A multicenter report. *J. Am. Coll. Cardiol.* **1992**, 20, 1391–1396.
322. Morita, H. *et al.* Atrial fibrillation and atrial vulnerability in patients with Brugada syndrome. *J. Am. Coll. Cardiol.* **40**, 1437–1444 (2002).
323. Kapplinger, J. D. *et al.* An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart. Rhythm.* **7**, 33–46 (2010).
324. Brugada, R. *et al.* Brugada Syndrome. *GeneReviews [Internet]* (2005; updated: 2022).
325. Hosseini, S. M. *et al.* Reappraisal of reported genes for sudden arrhythmic death: Evidence-based evaluation of gene validity for brugada syndrome. *Circulation* **138**, 1195–1205 (2018).

326. Priori, S. G. *et al.* Clinical and genetic heterogeneity of right bundle branch block and ST-segment elevation syndrome: A prospective evaluation of 52 families. *Circulation* **102**, 2509–2515 (2000).
327. Probst, V. *et al.* SCN5A Mutations and the role of genetic background in the pathophysiology of brugada syndrome. *Circ. Cardiovasc. Genet.* **2**, 552–557 (2009).
328. Saber, S. *et al.* Complex genetic background in a large family with Brugada syndrome. *Physiol. Rep.* **3**, 1–7 (2015).
329. Daimi, H. *et al.* Role of SCN5A coding and non-coding sequences in Brugada syndrome onset: What's behind the scenes? *Biomed. J.* **42**, 252–260 (2019).
330. Olivotto, I. & Coppini, R. Channelopathies, cardiac hypertrophy, and the theory of light. *Eur. Heart J.* **39**, 2908–2910 (2018).
331. Leoni, A. L. *et al.* Variable Nav1.5 Protein Expression from the Wild-Type Allele Correlates with the Penetrance of Cardiac Conduction Disease in the Scn5a+/- Mouse Model. *PLoS ONE.* **5**, e9298 (2010).
332. Bezzina, C. R. *et al.* Common sodium channel promoter haplotype in Asian subjects underlies variability in cardiac conduction. *Circulation* **113**, 338–344 (2006).
333. Yang, P. *et al.* Polymorphisms in the cardiac sodium channel promoter displaying variant in vitro expression activity. *Eur. J. Hum. Genet.* **16**, 350–357 (2008).
334. Park, J. K. Genetic variants in SCN5A promoter are associated with arrhythmia phenotype severity in patients with heterozygous loss-of-function mutation. *Heart. Rhythm.* **9**, 1090–1096 (2012).
335. Yagihara, N. *et al.* Variants in the SCN5A Promoter Associated With Various Arrhythmia Phenotypes. *J. Am. Heart Assoc.* **5**, 1–10 (2016).
336. Matsumura, H. *et al.* H558R, a common SCN5A polymorphism, modifies the clinical phenotype of Brugada syndrome by modulating DNA methylation of SCN5A promoters. *J. Biomed. Sci.* **24**, 1–10 (2017).
337. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat. Genet.* **42**, 1068–1076 (2010).
338. Chambers, J. C. *et al.* Genetic variation in SCN10A influences cardiac conduction. *Nat. Genet.* **42**, 149–152 (2010).
339. Akopian, A. N. *et al.* A tetrodotoxin-resistant voltage-gated sodium channel expressed by sensory neurons. *Nature* **379**, 257–262 (1996).
340. Benn, S. C. *et al.* Developmental expression of the TTX-resistant voltage-gated sodium channels Nav1.8 (SNS) and Nav1.9 (SNS2) in primary sensory neurons. *J. Neurosci.* **21**, 6077–6085 (2001).
341. Fukuoka, T. & Noguchi, K. Comparative study of voltage-gated sodium channel-subunits in non-overlapping four neuronal populations in the rat dorsal root ganglion. *Neurosci. Res.* **70**, 164–171 (2011).
342. Verkerk, A. O. *et al.* Functional Nav1.8 channels in intracardiac neurons: The link between SCN10A and cardiac electrophysiology. *Circ. Res.* **111**, 333–343 (2012).
343. Stroud, D. M. *et al.* Contrasting Nav1.8 activity in Scn10a-/- ventricular myocytes and the intact heart. *J. Am. Heart Assoc.* **5**, e002946 (2016).
344. Yang, T. *et al.* Blocking Scn10a channels in heart reduces late sodium current and is antiarrhythmic. *Circ. Res.* **111**, 322–332 (2012).
345. Dybkova, N. *et al.* Differential regulation of sodium channels as a novel proarrhythmic mechanism in the human failing heart. *Cardiovasc. Res.* **114**, 1728–1737 (2018).
346. Bezzina, C. R. *et al.* Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat. Genet.* **45**, 1044–1049 (2013).

347. Veerman, C. C. *et al.* The brugada syndrome susceptibility gene HEY2 modulates cardiac transmural ion channel patterning and electrical heterogeneity. *Circ. Res.* **121**, 537–548 (2017).
348. Pfeufer, A. *et al.* Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153–159 (2010).
349. Man, J. C. K. *et al.* An enhancer cluster controls gene activity and topology of the SCN5A-SCN10A locus in vivo. *Nat. Commun.* **10**, 1–15 (2019).
350. London, B. Whither Art Thou, SCN10A, and What Art Thou Doing? *Circ. Res.* **111**, 268–270 (2012).
351. Pinsach-Abuin M. *et al.* Analysis of Brugada syndrome loci reveals that fine-mapping clustered GWAS hits enhances the annotation of disease-relevant variants. *Cell Rep. Med.* **2**, 100250 (2021).
352. Pérez-Agustín, A. *et al.* Role of Non-Coding Variants in Brugada Syndrome. *Int. J. Mol. Sci.* **21**, 8556 (2020).
353. Yang, B. *et al.* The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat. Med.* **13**, 486–491 (2007).
354. Wahlquist, C. *et al.* Inhibition of miR-25 improves cardiac contractility in the failing heart. *Nature* **508**, 531–535 (2014).
355. Luo, X. *et al.* MicroRNA-26 governs profibrillatory inward-rectifier potassium current changes in atrial fibrillation. *J. Clin. Investig.* **123**, 1939–1951 (2013).
356. Van Rooij, E. *et al.* Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl. Acad. Sci. USA* **105**, 13027–13032 (2008).
357. Lu, Y. *et al.* MicroRNA-328 contributes to adverse electrical remodeling in atrial fibrillation. *Circulation* **122**, 2378–2387 (2010).
358. Daimi, H. *et al.* Regulation of SCN5A by microRNAs: MiR-219 modulates SCN5A transcript expression and the effects of flecainide intoxication in mice. *Heart Rhythm.* **12**, 1333–1342 (2015).
359. Zhou, R. *et al.* Whole Genome Network Analysis of Ion Channels and Connexins in Myocardial Infarction. *Cell. Physiol. Biochem.* **27**, 299–304 (2011).
360. Zhang, X. *et al.* A common variant alters SCN5A-miR-24 interaction and associates with heart failure mortality. *J. Clin. Investig.* **128**, 1154–1163 (2018).
361. Labun, K. *et al.* CRISPR Genome Editing Made Easy Through the CHOPCHOP Website. *Curr. Protoc.* **1**, e46 (2021).
362. Brinkman, E. K. & van Steensel, B. Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER. *Methods Mol. Biol.* **1961**, 29–44 (2019).
363. Lin, Y. *et al.* Heparin Promotes Cardiac Differentiation of Human Pluripotent Stem Cells in Chemically Defined Albumin-Free Medium, Enabling Consistent Manufacture of Cardiomyocytes. *Stem Cells Transl. Med.* **6**, 527–538 (2017).
364. Sharma, A. *et al.* Derivation of highly purified cardiomyocytes from human induced pluripotent stem cells using small molecule-modulated differentiation and subsequent glucose starvation. *J. Vis. Exp.* **18**, 52628 (2015).
365. Montague, T. G. *et al.* CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407 (2014).
366. Distler, J. H. *et al.* Nucleofection: a new, highly efficient transfection method for primary human keratinocytes. *Exp. Dermatol.* **14**, 315–320 (2005).
367. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell. Biol.* **20**, 490–507 (2019).
368. Taylor, S.C. *et al.* The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends Biotechnol.* **37**, 761–774 (2019).

369. Jabato, F.M. *et al.* Gene expression analysis method integration and co-expression module detection applied to rare glucide metabolism disorders using ExpHunterSuite. *Sci. Rep.* **11**, 15062 (2021).
370. Bruntraeger, M. *et al.* Editing the Genome of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9 Ribonucleoprotein Complexes. *Methods Mol. Biol.* **1961**, 153-183 (2019).
371. Yang, P. *et al.* Cloning and initial characterization of the human cardiac sodium channel (SCN5A) promoter. *Cardiovasc. Res.* **61**, 56-65 (2004).
372. He, A. *et al.* PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes develop.* **26**, 37-42 (2012).
373. Chen, M. *et al.* Inhibition of cardiomyocyte hypertrophy by protein arginine methyltransferase 5. *J. Biol. Chem.* **289**, 24325-24335 (2014).
374. Wang, J. *et al.* SUMO-1 modification activated GATA4-dependent cardiogenic gene activity. *J. Biol. Chem.* **279**, 49091-49098 (2004).
375. Lee, S. & Lee, D. K. What is the proper way to apply the multiple comparison test? *Korean. J. Anesthesiol.* **71**, 353-360 (2018).
376. Carrasco, M. *et al.* GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Invest.* **122**, 3504-3515 (2012).
377. Yang, X. J. & Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol. Cell* **31**, 449-461 (2008).
378. Ang, Y.S. *et al.* Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167**, 1734-1749 (2016).
379. Peng, L. *et al.* Deacetylase-independent function of SIRT6 couples GATA4 transcription factor and epigenetic activation against cardiomyocyte apoptosis. *Nucleic Acids Res.* **48**, 4992-5005 (2020).
380. Moskowitz, I. P. *et al.* The T-Box transcription factor Tbx5 is required for the patterning and maturation of the murine cardiac conduction system. *Development* **131**, 4107-4116 (2004).
381. Kapoor, A. *et al.* Multiple SCN5A variant enhancers modulate its cardiac gene expression and the QT interval. *Proc. Natl. Acad. Sci. USA* **116**, 10636-10645 (2019).
382. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431-440 (2006).
383. Ban, K. *et al.* Purification of cardiomyocytes from differentiating pluripotent stem cells using molecular beacons that target cardiomyocyte-specific mRNA. *Circulation* **128**, 1897-1909 (2013).
384. Molina, C. E. *et al.* Identification of optimal reference genes for transcriptomic analyses in normal and diseased human heart. *Cardiovasc. Res.* **114**, 247-258 (2018).
385. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611-622 (2009).
386. Barutcu, A. R. *et al.* A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, 1444 (2018).
387. Kentepozidou, E. *et al.* Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **21**, 5 (2020).
388. Greenland, S. *et al.* Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, 337-350 (2016).
389. Erikson, G. A. *et al.* Whole-genome sequencing of a healthy aging cohort. *Cell* **165**, 1002-1011 (2016).
390. Rehm, H. L. *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* **15**, 733-747 (2013).
391. Szumilas, M. Explaining odds ratios. *J. Can. Acad. Child. Adolesc. Psychiatry* **19**, 227-229 (2010).

392. Janin, A. *et al.* TRPM4 mutations to cause autosomal recessive and not autosomal dominant Brugada type 1 syndrome. *Eur. J. Med. Genet.* **62**, 103527 (2019).
393. David, J. P. *et al.* Deletion in mice of X-linked, Brugada syndrome- and atrial fibrillation-associated Kcne5 augments ventricular KV currents and predisposes to ventricular arrhythmia. *FASEB J.* **33**, 2537-2552 (2019).
394. Wells, A. *et al.* Ranking of non-coding pathogenic variants and putative essential regions of the human genome. *Nat Commun.* **10**, 5241 (2019).
395. Milman, A. *et al.* Ethnic differences in patients with Brugada syndrome and arrhythmic events: New insights from Survey on Arrhythmic Events in Brugada Syndrome. *Heart Rhythm* **16**, 1468-1474 (2019).
396. Maekawa, K. *et al.* Genetic polymorphisms and haplotypes of the human cardiac sodium channel alpha subunit gene (SCN5A) in Japanese and their association with arrhythmia. *Ann. Hum. Genet.* **69**, 413-428 (2005).
397. Zhang, Y. *et al.* Single nucleotide polymorphisms and haplotype of four genes encoding cardiac ion channels in Chinese and their association with arrhythmia. *Ann. Noninvasive Electrocardiol.* **13**, 180-190 (2008).
398. Behr, E. R. *et al.* Role of common and rare variants in SCN10A: results from the Brugada syndrome QRS locus gene discovery collaborative study. *Cardiovasc. Res.* **106**, 520-529 (2015).
399. Nakano, Y. *et al.* Common Variant Near HEY2 Has a Protective Effect on Ventricular Fibrillation Occurrence in Brugada Syndrome by Regulating the Repolarization Current. *Circ. Arrhythm. Electrophysiol.* **9**, e003436 (2016).
400. Martínez-Campelo, L. *et al.* Searching for genetic modulators of the phenotypic heterogeneity in Brugada syndrome. *PLoS One* **17**, e0263469 (2022).
401. Roden, D. M. Repolarization reserve: a moving target. *Circulation* **118**, 981-2 (2008).
402. Gaborit, N. *et al.* Transcriptional profiling of ion channel genes in Brugada syndrome and other right ventricular arrhythmogenic diseases. *Eur. Heart. J.* **30**, 487-96 (2009).
403. Martin, C. A. *et al.* Reduced Na(+) and higher K(+) channel expression and function contribute to right ventricular origin of arrhythmias in Scn5a+/- mice. *Open. Biol.* **2**, 120072 (2012).
404. Liang, P. *et al.* Patient-Specific and Genome-Edited Induced Pluripotent Stem Cell-Derived Cardiomyocytes Elucidate Single-Cell Phenotype of Brugada Syndrome. *J. Am. Coll. Cardiol.* **68**, 2086-2096 (2016).
405. Campuzano, O. *et al.* Update on genetic basis of Brugada syndrome: monogenic, polygenic or oligogenic? *Int. J. Mol. Sci.* **21**, 7155 (2020).
406. Monasky, M. M. *et al.* Evaluating the use of genetics in Brugada syndrome risk stratification. *Front. Cardiovasc. Med.* **8**, 652027 (2021).
407. Pappone, C. *et al.* The omics of channelopathies and cardiomyopathies: what we know and how they are useful. *Eur. Heart. J. Suppl.* **22**, 105-109 (2020).
408. Monasky, M. M. *et al.* Commentary: next generation sequencing and linkage analysis for the molecular diagnosis of a novel overlapping syndrome characterized by hypertrophic cardiomyopathy and typical electrical instability of Brugada syndrome. *Front. Physiol.* **8**, 1056 (2017).
409. Pappone, C. *et al.* Epicardial ablation in genetic cardiomyopathies: a new frontier. *Eur. Heart. J. Suppl.* **1**, 61-66 (2019).
410. Tafti, M.F. *et al.* Novel and heteroplasmic mutations in mitochondrial tRNA genes in Brugada syndrome. *Cardiol. J.* **25**, 113-119 (2018).
411. Wang, H. *et al.* Hyperactivation of platelet-derived growth factor signalling contributes to arrhythmogenesis in Brugada syndrome. *Clin. Transl. Med.* **12**, e715 (2022).
412. Perkins, J. *et al.* ExpHunterSuite: Package For The Comprehensive Analysis Of Transcriptomic Data. *R package version 1.5.0* (2022).



413. Browning, B. L. *et al.* A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338-348 (2018).
414. Delaneau, O. *et al.* Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
415. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
416. Edward, M. *et al.* Effect of Highly Fragmented DNA on PCR. *Nucleic Acids Res.* **24**, 5026–5033 (1996).
417. Brionne, A, *et al.* ViSEAGO: a Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Min.* **12**, 16 (2019).
418. Morota, G. *et al.* MeSH-Informed Enrichment Analysis and MeSH-Guided Semantic Similarity Among Functional Terms and Gene Products in Chicken. *G3 (Bethesda)* **6**, 2447-2453 (2016).
419. Campuzano, O. *et al.* Genetic interpretation and clinical translation of minor genes related to Brugada syndrome. *Hum Mutat.* **40**, 749-764 (2019).
420. Boczek, N. J. *et al.* Characterization of SEMA3A-encoded semaphorin as a naturally occurring Kv4.3 protein inhibitor and its contribution to Brugada syndrome. *Circ. Res.* **115**, 460-469 (2014).
421. Gee, P. *et al.* Extracellular nanovesicles for packaging of CRISPR-Cas9 protein and sgRNA to induce therapeutic exon skipping. *Nat. Commun.* **11**, 1334 (2020).
422. Kim, S. *et al.* Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012-1019 (2014).
423. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
424. Richardson, C. *et al.* Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
425. Zuris, J. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* **33**, 73–80 (2015).
426. Kagita, A. *et al.* Efficient ssODN-Mediated Targeting by Avoiding Cellular Inhibitory RNAs through Precomplexed CRISPR-Cas9/sgRNA Ribonucleoprotein. *Stem Cell Reports* **16**, 985-996 (2021).
427. Petazzi, P. *et al.* CRISPR/Cas9-Mediated Gene Knockout and Knockin Human iPSCs. *Methods Mol. Biol.* **2454**, 559-574 (2022).
428. Wen, W. *et al.* High-Level Precise Knockin of iPSCs by Simultaneous Reprogramming and Genome Editing of Human Peripheral Blood Mononuclear Cells. *Stem Cell Reports.* **10**, 1821-1834 (2018).
429. Tigchelaar, W. *et al.* In EXOG-depleted cardiomyocytes cell death is marked by a decreased mitochondrial reserve capacity of the electron transport chain. *Bioessays* **38**, S136-S145 (2016).
430. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627 (2015).
431. Montefiori, L. E. *et al.* A promoter interaction map for cardiovascular disease genetics. *eLife* **7**, e35788 (2018).
432. Javierre, B. M. *et al.* Lineage-Specific genome architecture links enhancers and Non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 (2016).
433. Freire-Pritchett, P. *et al.* Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* **6**, e21926 (2017).
434. Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).

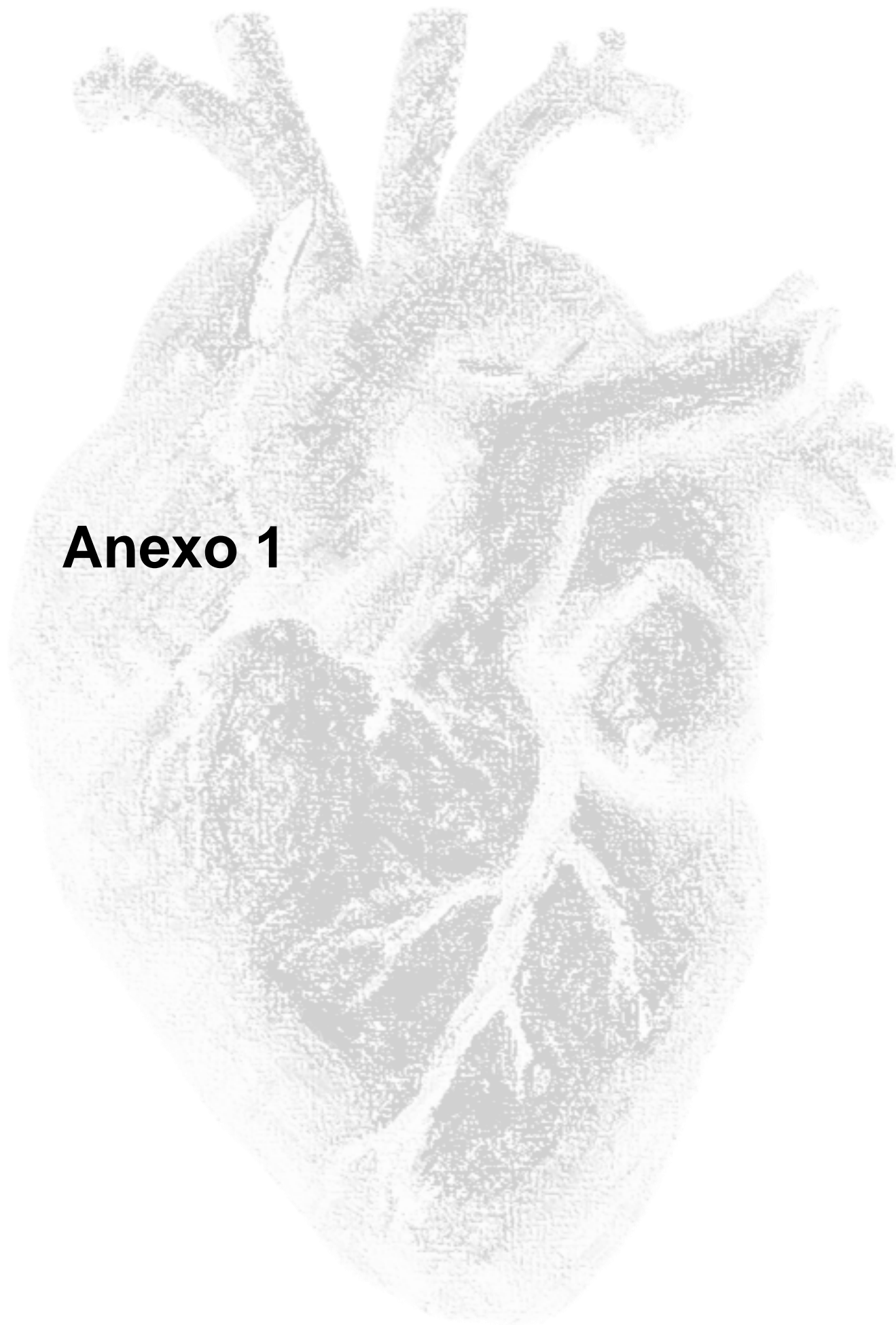
435. Siersbæk, R. *et al.* Dynamic rewiring of Promoter-Anchored chromatin loops during adipocyte differentiation. *Mol. Cell* **66**, 420–435 (2017).
436. Nykamp, K. *et al.* Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. *Genet. Med.* **19**, 1105–1117 (2017).
437. Janssens, S. & Beyaert, R. A universal role for MyD88 in TLR/IL-1R-mediated signaling. *Trends Biochem. Sci.* **27**, 474–482 (2002).
438. Chen, W. *et al.* Overexpression of myeloid differentiation protein 88 in mice induces mild cardiac dysfunction, but no deficit in heart morphology. *Braz. J. Med. Biol. Res.* **49**, e4794 (2016).
439. Feng, Y. *et al.* MyD88 and Trif signaling play distinct roles in cardiac dysfunction and mortality during endotoxin shock and polymicrobial sepsis. *Anesthesiology* **115**, 555–567 (2011).
440. Singh, M. V. *et al.* MyD88 mediated inflammatory signaling leads to CaMKII oxidation, cardiac hypertrophy and death after myocardial infarction. *J. Mol. Cell. Cardiol.* **52**, 1135–1144 (2012).
441. van Hoorn, F. *et al.* SCN5A mutations in Brugada syndrome are associated with increased cardiac dimensions and reduced contractility. *PLoS One* **7**, e42037 (2012).
442. Hummel, Y. M. *et al.* Ventricular dysfunction in a family with long QT syndrome type 3. *EP Eur.* **15**, 1516–1521 (2013).
443. Akazawa, H. & Komuro, I. Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. *Pharmacol Ther.* **107**, 252–268 (2005).
444. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
445. Hu, D. *et al.* Mutations in SCN10A are responsible for a large fraction of cases of Brugada syndrome. *J. Am. Coll. Cardiol.* **64**, 6–79 (2014).
446. Andrade, C. The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian J. Psychol. Med.* **41**, 210–215 (2019).
447. Ioannidis, P. A. Publishing research with P-values: Prescribe more stringent statistical significance or proscribe statistical significance?. *Eur. Heart J.* **40**, 2553–2554 (2019).
448. du Prel, J.B. *et al.* Confidence interval or p-value?: part 4 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.* **106**, 335–339 (2009).
449. Gherghiceanu, M. *et al.* Cardiomyocytes derived from human embryonic and induced pluripotent stem cells: comparative ultrastructure. *J. Cell. Mol. Med.* **15**, 2539–2551 (2011).
450. Karakikes, I. *et al.* Human induced pluripotent stem cell-derived cardiomyocytes: insights into molecular, cellular, and functional phenotypes. *Circ. Res.* **117**, 80–88 (2015).
451. Machiraju, P. & Greenway, S. C. Current methods for the maturation of induced pluripotent stem cell-derived cardiomyocytes. *World J. Stem Cells* **11**, 33–43 (2019).
452. Ronaldson-Bouchard, K. *et al.* Advanced maturation of human cardiac tissue grown from pluripotent stem cells. *Nature* **556**, 239–243 (2019).
453. Huang, C. Y. *et al.* Enhancement of human iPSC-derived cardiomyocyte maturation by chemical conditioning in a 3D environment. *J. Mol. Cell Cardiol.* **138**, 1–11 (2020).
454. Tang, B. L. Maturing iPSC-Derived Cardiomyocytes. *Cells* **9**, 213 (2020).
455. Hill, C. L. & Stephens, G. J. An Introduction to Patch Clamp Recording. *Methods Mol. Biol.* **2188**, 1–19 (2021).
456. Martínez, A. L. *et al.* Identification of Sodium Transients Through NaV1.5 Channels as Regulators of Differentiation in Immortalized Dorsal Root Ganglia Neurons. *Front. Cell Neurosci.* **16**, 816325 (2022).
457. Martínez, A. L. *et al.* A New Model of Sensorial Neuron-Like Cells for HTS of Novel Analgesics for Neuropathic Pain. *SLAS Discov.* **24**, 158–168 (2019).
458. Martínez, A. L. *et al.* Development of a novel in vitro assay to screen for neuroprotective drugs against iatrogenic neurite shortening. *PLoS One* **16**, e0248139 (2021).
459. Kaushal, A. *et al.* CTCF loss has limited effects on global genome architecture in *Drosophila* despite critical regulatory functions. *Nat. Commun.* **12**, 1011 (2021).

460. Kentepozidou, E. *et al.* Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **21**, 5 (2020).
461. Bauwens, M. *et al.* ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet. Med.* **21**, 1761–1771 (2019).
462. O'Leary, T. *et al.* Correlations in ion channel expression emerge from homeostatic tuning rules. *Proc. Natl. Acad. Sci. USA* **110**, E2645-E2654 (2013).
463. Alemu, E. Y. *et al.* Determinants of expression variability. *Nucleic Acids Res.* **42**, 3503–14 (2014).
464. Roberfroid, S. *et al.* Gene expression variability in clonal populations: causes and consequences. *Crit. Rev. Microbiol.* **42**, 969–84 (2016).
465. Bashkeel, N. *et al.* Human gene expression variability and its dependence on methylation and aging. *BMC Genomics* **20**, 941 (2019).
466. Moncayo-Arlandi, J. & Brugada, R. Unmasking the molecular link between arrhythmogenic cardiomyopathy and Brugada syndrome. *Nat. Rev. Cardiol.* **14**, 744–756 (2017).
467. Anderson, D. J. *et al.* *NKX2-5* regulates human cardiomyogenesis via a *HEY2* dependent transcriptional network. *Nat. Commun.* **9**, 1373 (2018).
468. Spoel, S. H. Orchestrating the proteome with post-translational modifications. *J. Exp. Bot.* **69**, 4499-4503 (2018).
469. Manolio, T. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
470. Shimizu, S. *et al.* Multimerization of the GATA4 transcription factor regulates transcriptional activity and cardiomyocyte hypertrophic response. *Int. J. Biol. Sci.* **18**, 1079-1095 (2022).
471. Huang, L. *et al.* Molecular pathological study on LRRC10 in sudden unexplained nocturnal death syndrome in the Chinese Han population. *Int. J. Legal. Med.* **131**, 621–628 (2017).





# Anexo 1





V1: 31/07/12

**INFORMED CONSENT FOR THE STUDY OF THE GENETIC BASIS OF HEART DISEASES****(GENCARDIO-RESEARCH)****PRINCIPAL INVESTIGATOR: Ramon Brugada, MD, PhD, FACC, FESC****AIMS**

We appreciate your collaboration in the study GENCARDIO-RESEARCH that takes place at the Cardiovascular Genetic Center in Girona. Your participation will help improve our knowledge of heart disease. In recent years significant progress has been made in the investigation of genetic factors in cardiovascular diseases. For such studies it is necessary to collect biological samples (blood, saliva, tissue) and DNA obtained from healthy volunteers and from people with cardiac problems. DNA is the genetic material that is present in all our cells and carries a code in the form of "gene" that determines our personal physical characteristics, like eye color, skin, etc. These genes can also influence the risk that some people to develop certain diseases. The main objective of this project is to ascertain what features of DNA affect the risk of a person to develop heart disease. We are currently conducting a study that aims to collect and store blood samples to isolate DNA. This DNA is stored in the Cardiovascular Genetics Centre and allows scientists to investigate what genes are associated with heart disease, what diseases are influenced by the environment, and which genes influence the response to certain treatments.

All personal information that is collected or generated by this study will be protected in accordance to the law. To this end we use the measures as detailed below:

**PROCES DESCRIPCION**

During your participation in the study

- We will inform you about the objectives of the project and answer any questions that you may have.
- To participate in our study you will not receive any financial reward.
- The data recorded in your file can be treated statistically for the purpose of research.



V1: 31/07/12

- The data may be provided and processed, anonymously, to a third party who may use it only for research purposes.
- A biological sample of blood, saliva, tissue will be collected in order to extract DNA.
- The Cardiovascular Genetic Center agrees that all information received and all the samples will be codified prior to shipment to outside researchers so the donor cannot be identified.
- Products obtained from the samples will be archived and kept for a minimum period of five years at the Cardiovascular Genetic Center.
- Products obtained from the samples may be used in biomedical research studies conducted by other institutions, national or foreign, provided that: 1) have been considered of scientific interest, 2) that meet the requirements established by the Scientific Committees and External expert Advisors in ethical, economic, environmental, legal and social matters.
- You will be able to know the research studies in which your samples and clinical data have been used, and we will provide you with the results of these studies related to your disease. The Cardiovascular Genetics Center will have copy of the information from each study.
- The Cardiovascular Genetics Center is committed not to sell the samples or data obtained from clinical samples.
- You are entitled to request the destruction of your sample and all the related information which has been stored at the Cardiovascular Genetic Center.

Contact information:

Dr. Ramon Brugada Terradellas

Cardiovascular Genetic Center

Parc Hospitalari Martí i Julià Carrer Dr. Castany, s/n

Edifici Mancomunitat 2

17190 -Salt- ramon@brugada.org

V1: 31/07/12

### **DONOR'S STATEMENT**

I have been informed by the health professional mentioned below:

- About the advantages and disadvantages of this procedure.
- About the site of collection, storage and processing of the samples and data.
- About the purpose for using my samples and data (genetic studies, public health or statistical that meet all the law requirements, the Advisory Committee of experts in ethical, economic, environmental, legal and Social, and the Scientific Committee).
- That my samples and data will be provided anonymized to the investigators who work with them.
- That, if due to the genetic analysis I have any injury, the Cardiovascular Genetic Center cannot offer me any economic compensation or take care of any medical bills. I will be offered the appropriate treatment as it is provided to the rest of the community.
- That at any time I can revoke my consent and request the destruction of my data and samples stored at the Cardiovascular Genetic Center.
- That at any time I can request information about genetic studies where my samples have been used.
- I understood all the information and I have been able to ask all the questions that I thought appropriate,

I agree to be contacted by the Center's personnel in order to get additional information YES\_\_\_  
NO\_\_\_

I wish that my DNA sample (or a sample from a deceased family member or minor of whom I am responsible) be destroyed once the study has finalized. YES\_\_\_NO\_\_\_

V1: 31/07/12

Please, sign parts A, B o C in next page as appropriate.

**HEALTH PROFESSIONAL'S STATEMENT REGARDING PROPERLY INFORMING THE DONOR**

Name:

Signature:

**A. ADULT WITH FULL MENTAL CAPACITY**

Patient's name

Patient's signature

**B. MINOR OR PATIENT WITH DIMINISHED MENTAL CAPACITY**

Minor's name / patient with diminished mental capacity

Parent 's name/tutor/legal representative

Parent's signature/tutor/legal representative

**C. DECEASED**

Deceased's name:

Parent's name /tutor / legal representative

Parent's signature /tutor / legal representative







**Anexo 2**



**Tabla A- 1. Términos HPO utilizados para el análisis de similitud semántica.** Se describen los códigos HPO utilizados con su respectivo significado.

Término HPO (ID)	Nombre del término HPO
HP:0001663	Fibrilación ventricular
HP:0001664	<i>Torsade de pointes</i>
HP:0001695	Paro cardíaco
HP:0004751	Taquicardia ventricular paroxística
HP:0001250	Arresto cardíaco
HP:0011715	Bloqueo trifascicular
HP:0001645	Muerte súbita cardíaca
HP:0005184	Intervalo QT prolongado
HP:0011705	Bloqueo auriculoventricular de primer grado
HP:0004308	Arritmia ventricular
HP:0001688	Bradicardia sinusal
HP:0011712	Bloqueo de rama derecha
HP:0001279	Síncope
HP:0005949	Episodios de apnea en la infancia
HP:0005172	Bloqueo fascicular posterior izquierdo
HP:0003198	Miopatía
HP:0011710	Bloqueo de rama del haz
HP:0001678	Bloqueo auriculoventricular
HP:0002900	Hipopotasemia
HP:0006673	Función sistólica reducida
HP:0012722	Bloqueo cardíaco
HP:0001962	Palpitaciones
HP:0003457	ECG anormal
HP:0004757	Fibrilación auricular paroxística
HP:0001635	Insuficiencia cardíaca congestiva
HP:0006671	Taquicardia auricular paroxística
HP:0011713	Bloqueo de rama izquierda
HP:0005110	Fibrilación auricular
HP:0012251	Elevación del segmento ST
HP:0031677	Taquicardia ventricular polimórfica
HP:0004755	Taquicardia supraventricular
HP:0001649	Taquicardia
HP:0004749	<i>Flutter</i> auricular
HP:0003116	Ecocardiograma anormal
HP:0012723	Bloqueo sinoauricular

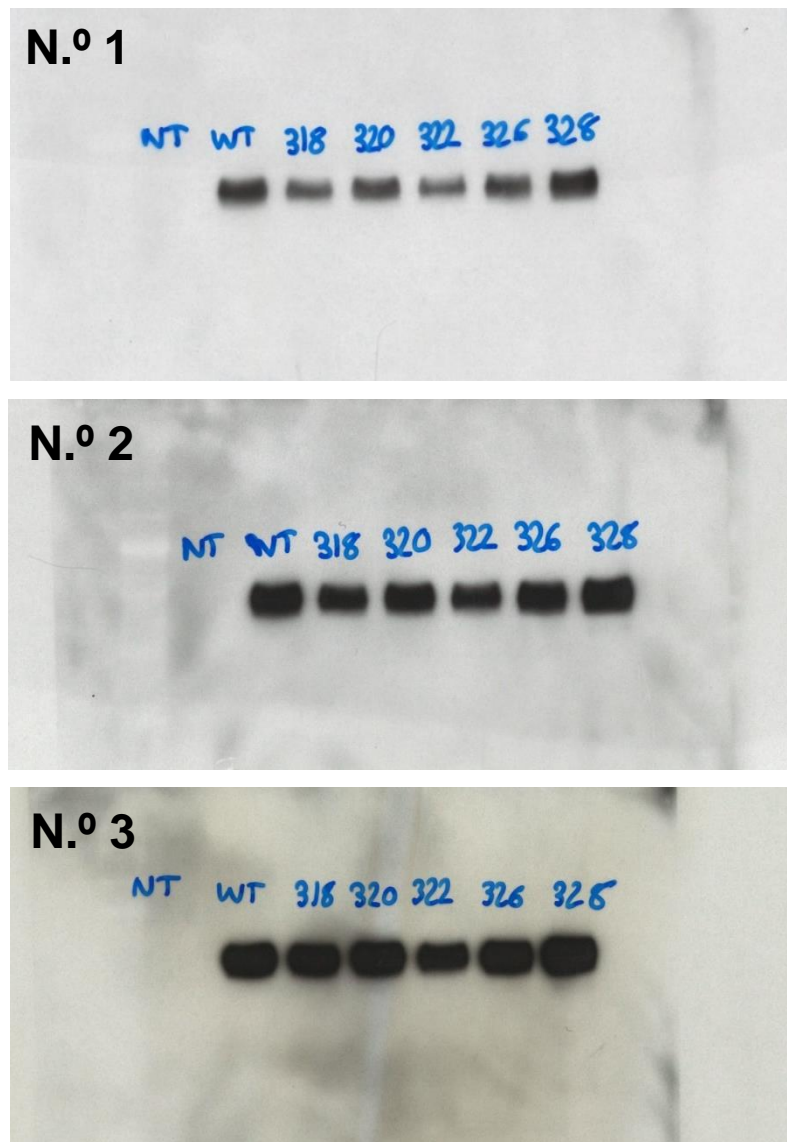




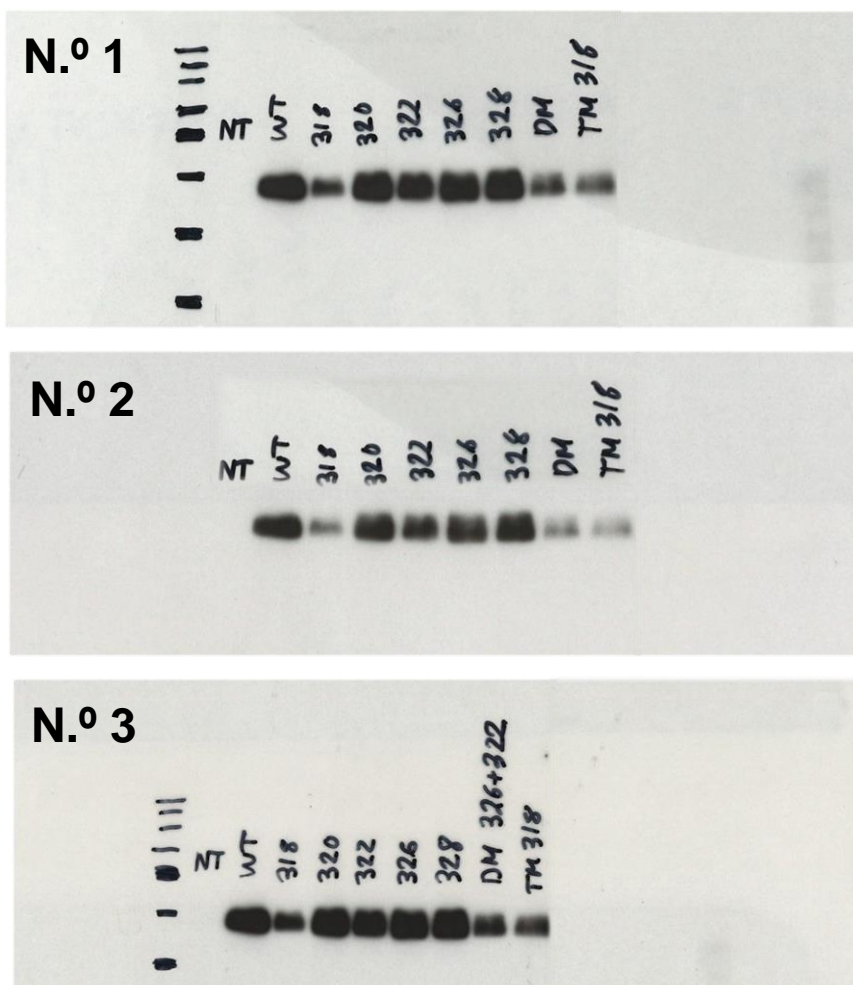


**Anexo 3**





**Figura A- 1. Niveles de acetilación de GATA4 en mutantes simples de GATA4.** WB  $\alpha$ -GATA4 de los lisados de HEK293T transfectados con mutantes simples de GATA4, tratados previamente con inhibidores de HDAC (TSA y NA), e inmunoprecipitados (IP) para lisinas acetiladas; la condición NT corresponde al control sin transfectar. Se muestran 3 WB independientes para el mismo experimento.



**Figura A- 2. Niveles de acetilación de GATA4 en mutantes simples, dobles y triples de GATA4.** WB  $\alpha$ -GATA4 de los lisados de HEK293T transfectados con mutantes simples, doble (DM) y triple (TM) de GATA4, tratados previamente con inhibidores de HDAC (TSA, y NA), e inmunoprecipitados (IP) para lisinas acetiladas; la condición NT corresponde al control sin transfectar. Se muestran 3 WB independientes para el mismo experimento.

