Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Títol del treball:

CHARACTERIZATION OF NON-CODING GENETIC VARIANTS ASSOCIATED WITH BRUGADA SYNDROME: RESEARCH AND COMMUNICATION

Estudiant:	Anna Torrecillas Sanllehí
Grau:	Doble Grau en Biologia i Biotecnologia
Correu electrònic:	torrecillas.anna@gmail.com

Tutor:	Dr. Miquel Duran
Cotutor:	Dra. Sara Pagans
Empresa/Institució:	Universitat de Girona - IDIBGI

Vistiplau tutor (i cotutor):

Nom del tutor:	Dr. Miquel Duran
Nom del cotutor:	Dra. Sara Pagans
Empresa/Institució:	Universitat de Girona - IDIBGI
Correus electrònics:	miquel.duran@udg.edu
	sara.pagans@udg.edu

Data de dipòsit de la memòria a secretaria de coordinació: **24 de maig de 2018**

CHARACTERIZATION OF NON-CODING GENETIC VARIANTS ASSOCIATED WITH BRUGADA SYNDROME:

RESEARCH AND COMMUNICATION

ANNA TORRECILLAS SANLLEHÍ

UNIVERSITAT DE GIRONA - IDIBGI

DUAL BACHELOR'S DEGREE BIOLOGY/BIOTECHNOLOGY

Tutor/Cotutor: MIQUEL DURAN and SARA PAGANS

ABSTRACT

Brugada syndrome (BrS) is defined as a cardiac disorder associated with high risk of sudden cardiac death and ventricular fibrillation. It is responsible of 4% of all sudden deaths and 20% of sudden cardiac deaths in patients without structural abnormalities. BrS is caused by dysfunction of ion channels involved in generation of the cardiac action potential. Approximately 20-25% of BrS patients have mutations in the coding region of the *SCN5A* gene, which encodes the α-subunit of the cardiac voltage-gated sodium channel.

This work is included in the "Regulome-Seq" project, which aims to study non-coding DNA variants within transcription factor (TF) binding sites nearby BrS-associated genes. To characterize genetic variation in these regions, a set of pre-selected 1,293 *loci* was sequenced in a cohort of 89 diagnosed-BrS cases. To evaluate the consequences of the variants identified that overlap TF CTCF binding sites, the DeepBind algorithm was used to predict the effect of each variant in CTCF binding. 34 of these variants overlapping CTCF binding sites were cloned into a luciferase reporter vector to evaluate their effect in cell-based assays. After co-transfection in H9c2 cells, the luciferase activity obtained was measured to examine the CTCF binding level for each sequence.

The results obtained indicate that DeepBind algorithm and luciferase assays results have a similar trend to evaluate the effect of the variants on CTCF binding. However, the results did not show a significant correlation between both methods. This suggests that the use of other DeepBind algorithms based on Machine Learning to predict BrS effects should be further explored.

The experimental part of this work was complemented by its communication to the public and by innovative steps in general Science dissemination, with the conception of a Molecular Biology workshop on possible mutations in DNA. Communicating genome and DNA-related concepts is a great opportunity to improve scientific culture of society and to increase Science awareness among young students. Most people have a small idea about what BrS is, because they think is "something" related to mutations in their body hearts.

This study seeks to explain the effect and influence of mutations in genetic diseases, to contribute to the open scientific attitude in the improvement of the educational process of Science in secondary education students and society.

RESUM

La síndrome de Brugada (SBr) es defineix com un trastorn cardíac associat a un alt risc de mort sobtada cardíaca i fibril·lació ventricular. És la responsable del 4% de totes les morts sobtades i del 20% de les freqüències cardíaques sobtades en pacients sense anomalies estructurals. La SBr està causada per la disfunció dels canals iònics implicats en la generació del potencial d'acció cardíac. Aproximadament el 20-25% dels pacients amb la SBr tenen mutacions a la regió codificant del gen *SCN5A*.

Aquest treball està inclòs dins el projecte "Regulome-Seq", el qual pretén estudiar variants d'ADN no codificants dins dels llocs d'unió del factor de transcripció (FT) prop dels gens associats a la SBr. Per dur-lo a terme, es va seqüenciar un total de 1.293 *loci* pre-seleccionats en una cohort de 89 pacients diagnosticats amb SBr. Per tal d'avaluar les conseqüències de les variants identificades que cauen en llocs d'unió del FT CTCF, es va utilitzar l'algoritme DeepBind per predir l'efecte de cada variant en la unió de CTCF. 34 d'aquestes variants es van clonar en un vector reporter de luciferasa. Després d'una co-transfecció en cèl·lules H9c2, es va mesurar l'activitat de luciferasa obtinguda per tal d'examinar el nivell d'unió de CTCF per a cada seqüència.

Els resultats obtinguts indiquen que els algoritmes DeepBind i els resultats de l'assaig de luciferasa tenen una tendència similar a l'avaluar l'efecte de les variants en la unió al motiu CTCF. Tanmateix, la correlació entre ambdós mètodes no és tan bona, la qual cosa suggereix que s'ha d'explorar encara més l'ús d'altres algoritmes basats en l'aprenentatge automàtic per predir els efectes de la BrS.

La part experimental d'aquest treball es complementa amb la divulgació científica i la comunicació del projecte mitjançant un taller de Biologia Molecular relacionat amb possibles mutacions de l'ADN. Comunicar el que és genoma i conceptes relacionats amb l'ADN és una bona oportunitat per millorar la cultura científica de la societat i augmentar la consciència dels joves estudiants. La majoria de les persones tenen una petita idea sobre el que és la SBr, creuen que és "alguna cosa" relacionada amb mutacions en el nostre cor.

Aquest estudi pretén així explicar l'efecte de les mutacions en malalties genètiques, per contribuir a l'actitud científica oberta en la millora del procés educatiu de la ciència en els estudiants d'educació secundària i la societat.

RESUMEN

El síndrome de Brugada (SBr) se define como un trastorno cardíaco asociado a un alto riesgo de muerte súbita cardíaca y fibrilación ventricular. Es el responsable del 4% de todas las muertes súbitas y del 20% de las frecuencias cardíacas repentinas en pacientes sin anomalías estructurales. El SBr está causado por la disfunción de los canales iónicos implicados en la generación del potencial de acción cardíaco. Aproximadamente, el 20-25% de los pacientes con la SBr tienen mutaciones en la región codificante del gen *SCN5A*.

Este trabajo está incluido en el proyecto "Regulome-Seq", el cual pretende estudiar variantes de ADN no codificantes en de los sitios de unión de factores de transcripción (FT) cerca de los genes asociados al SBr. Se secuenciaron un total de 1.293 *loci* preseleccionados en una cohorte de 89 pacientes diagnosticados con SBr. Para evaluar las consecuencias de las variantes identificadas que se superponen en sitios de unión del FT del CTCF, se utilizó el algoritmo DeepBind para predecir el efecto de cada variante en la unión de CTCF. 34 de estas variantes se clonaron en un vector reportero luciferasa. Después de una co-transfección en células H9c2, se midió la actividad de luciferasa obtenida para examinar el nivel de unión de CTCF por cada secuencia.

Los resultados obtenidos indican que los algoritmos DeepBind y los resultados del ensayo de luciferasa tienen una tendencia similar al evaluar el efecto de las variantes en la unión al motivo CTCF. Sin embargo, la correlación entre ambos métodos no es buena, lo que sugiere que se tiene que explorar aún más el uso de otros algoritmos basados en el aprendizaje automático para predecir los efectos del SBr.

La parte experimental de este trabajo se complementa con la divulgación científica y la comunicación del proyecto mediante un taller de Biología Molecular relacionado con posibles mutaciones del ADN. Comunicar lo que es genoma y conceptos relacionados con el ADN es una buena oportunidad para mejorar la cultura científica de la sociedad y aumentar la conciencia de los jóvenes estudiantes. La mayoría de las personas tienen una pequeña idea sobre lo que es el SBr, creen que es "algo" relacionado con mutaciones en nuestro corazón.

Este estudio pretende así explicar el efecto de las mutaciones en enfermedades genéticas, para contribuir a la actitud científica abierta en la mejora del proceso educativo de la ciencia en los estudiantes de educación secundaria y la sociedad.

I would like to express my deep gratitude to Dr. Miquel Duran and Dr. Sara Pagans for their patient guidance, enthusiastic encouragement, useful critiques and assistance in keeping my progress on schedule of this research work.

I would also like to thank to Ms. Mel·lina Pinsach, Mr. Adrià Pérez and to all the researchers and technicians in GenCardio (IDIBGI) for their help in offering me the resources in the laboratory. I would also like to extend my thanks to C4D (UdG) for giving me the opportunity to communicate Science.

Finally, I wish to thank my parents, family, University classmates and friends for their support and encouragement throughout my work and studies.

May 24th, 2018

CONTENT INDEX

1. INTRODUCTION

There are two different parts in this Bachelor Thesis, arising from a collaboration between two different research teams of the Universitat de Girona (UdG) that of Dra. Sara Pagans (GenCardio, IDIBGI) and Dr. Miquel Duran (Institut de Química Computacional i Catàlisi). The experimental work has been done at the IDIBGI laboratory in the Santa Caterina Hospital (Salt), while dissemination research of different activities based on the innovation in Science communication within the projects of C4D (UdG) in different exhibits and congresses in Girona, Hospitalet de Llobregat, and Madrid.

1.1. PROJECT DESCRIPTION

The impact of next generation sequencing in society and medicine is difficult to envision, but the world will have soon millions of human genomes sequenced. The size of the human genome is about 3,234 million bases, composed by a combination of four bases. Two people are expected to have more than three million differences (or variants) in their genomes [Campuzano *et al*., 2015]. This means that there is a small variation in our genomes (0,10%) that account for our differences on how we look, susceptibility to disease or response to pharmacological treatments. The information of total genomic DNA varies widely between people, and the proportion of coding (about 1-2% of the genome) and noncoding DNA (nearly 98% of the genome) within these genomes also changes greatly as well. The effects on coding regions are easy to predict, while the effects on noncoding regions are more difficult to determine [Mademont-Soler *et al.*, 2016]. Multiple single nucleotide polymorphisms (SNP) have been found to be associated with cardiovascular diseases in genetic association studies, but their genetic contributions to cardiovascular diseases are poorly understood.

This project aims to understand the role of noncoding DNA studying the effect of variants in transcription factor binding sites in the context of Brugada syndrome (BrS) [Alipanahi *et al*., 2015]. It is a clinical condition in which the heart stops beating suddenly and unexpectedly without any previous symptoms and it causes up to 800,000 deaths per year in the Western World [Zipes & Wellens, 1998]. And to communicate this project we performed a new Science communication workshop [McNutt, 2017], based on mutations, for High school students.

1.2. EXPERIMENTAL AND RESEARCH DESCRIPTION

1.2.1 THE HEART

The heart is a muscular organ localized in the thoracic cavity and specialized in pumping the blood.

The human heart [Colledge *et al.*, 2010] is divided into **4 different chambers** (*Figure 1*): two upper chambers (left atrium and right atrium) and two lower chambers (ventricles). There is a muscular wall, called **interventricular septum**, which separates the left and the right side of the heart preventing blood mixing between the two sites. Mitral and tricuspid valves are localized between the atria and the ventricle of left and right sides [Colledge *et al.*, 2010], respectively, ensuring the unidirectional flow of blood. The heart's wall is composed of 3 layers, from outside to inside: the **epicardium** (the serous layer), the **myocardium** (the muscular wall of heart containing contractile cardiomyocytes) and the **endocardium** (the innermost layer made up of endothelial tissue).

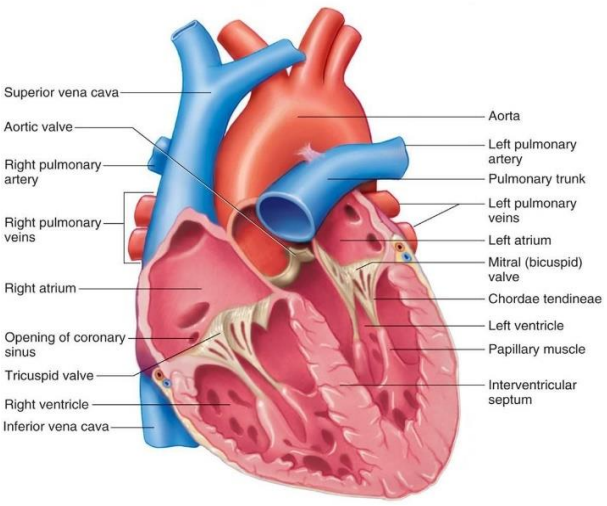


***Figure 1**. Structure of human heart. From The Mc Graw-Hill Companies.*

An electrical system controls the heart and uses electrical signals to contract the heart's walls [Baumgartner *et al.*, 2010]. When the walls contract, blood is pumped into the circulatory system. The cardiac cycle comprises the events of the human heart from the beginning of one heartbeat to the beginning of the next [Chen *et al.*, 2010]. It consists of two coordinated periods: diastole (the heart muscle relaxes and refills with blood) and systole (a period of robust contraction and pumping of blood). The blood carries the oxygen and nutrients that the organs need to work well and it also carries carbon dioxide to the lungs.

1.2.1.1. Electrical System of the heart and Cardiac Action Potential

The membrane potential is the difference in electric potential between the inside and the outside of the cell. Membrane potentials are determined by: ion concentrations inside and outside the cell; permeability of the membrane to these ions through ion channels [Balser, 2001], and active transport via ion pumps.

The electrical system of the heart is also called the cardiac conduction system [Balser, 2001], and is composed by: the **sinoatrial** (SA) node, located in the right atrium; the **atrioventricular** (AV) node, located on the interatrial septum; and the His-Purkinje system, located along the walls of the ventricles.

Each heartbeat begins with a signal from the SA node [Chen *et al.*, 2010], thus the SA node is sometimes called the heart natural pacemaker [Christoffels *et al.*, 2010].

The cardiac action potential [Peñaranda *et al.*, 2012] results from the balance between depolarizing and repolarizing ionic currents in cardiomyocytes. The resting membrane potential in cardiomyocytes is from -85 mV to -90 mV. These currents are mediated by the coordinated opening and closing of cardiac ion channels. The cardiac action potential (*Figure 2*) is initiated by a voltage stimulus [Balser, 2001], which triggers activation of voltage-gated sodium channels. This potential can be divided into four or five different phases: **phase 0** or depolarization, **phase 1** or early repolarization, **phase 2** or plateau phase, **phase 3** or final repolarization and **phase 4** (when the cell reaches its resting potential). Mutations in genes encoding ion channels involved in these different phases [Morita *et al.*, 2002] can alter the cardiac action potential and cause cardiac dysfunctions.

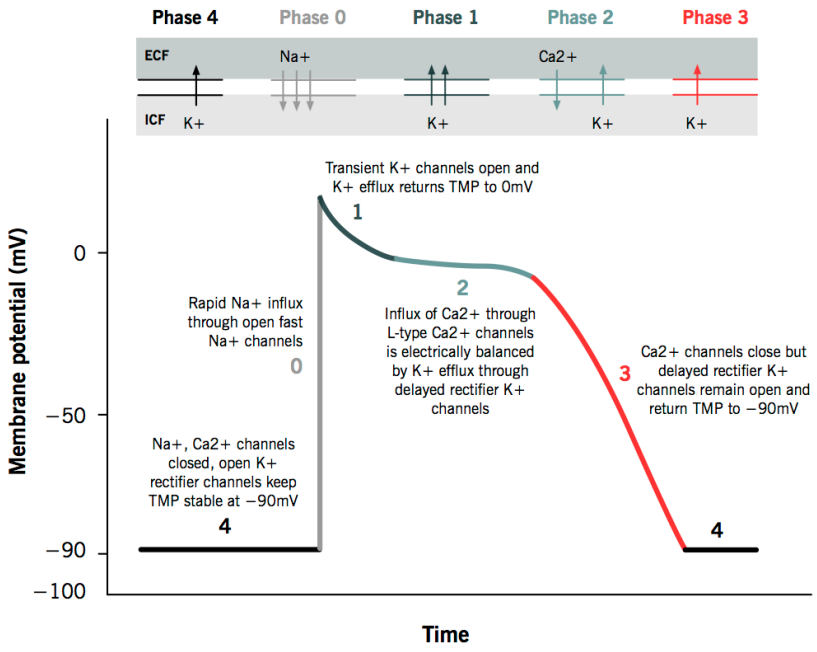


***Figure 2**. The cardiac action potential. The action potential in cardiomyocytes is composed of 5 phases (0-4), beginning and ending with phase 4. Ions involved in the potential are also shown in the different phases. From McMaster Pathophysiology Review*.

1.2.1.2. Electrocardiogram

The electrocardiogram (ECG) is a non-invasive device used to record the electrical activity of the heart from different angles to identify and locate a cardiac pathology (*Figure 3*). The ECG [Allen *et al.*, 2011] produces a distinctive waveform in response to the electrical changes taking place within the heart.

The first part of the wave [Allen *et al.*, 2011], called the **P wave**, is a small increase in voltage of about 0,10 mV and corresponds to the depolarization of the atria during atrial systole. The **QRS complex** which features a small drop in voltage (Q), a large voltage peak (R) and another small drop in voltage (S), and corresponds to the depolarization of the ventricles during ventricular systole.

***Figure 3**. Action potentials from different parts of the human heart are recorded as P wave, QRS complex and T wave in the ECG. From Clinical ECG Interpretation.*

The **T wave** is a small peak that follows the QRS complex and represents the ventricular repolarization during the relaxation phase of the cardiac cycle. Variations in the waveform and distance between the waves of the ECG can be used clinically to diagnose cardiac disorders [Hatcher & Basson, 2009].

1.2.2. SUDDEN CARDIAC DEATH

Cardiovascular diseases (CVD) are a class of pathological disorders that involve the heart or blood vessels [Fernández *et al*., 2017]. CVD are the most common cause of death globally as of 2008, accounting for 30% of deaths. The risk factors linked to these diseases are inherited and environmental, and include genetic factors, high blood pressure, excessive levels of cholesterol, diabetes, tobacco consumption and obesity.

Sudden Cardiac Death (SCD) is defined as natural death due to cardiac causes occurring suddenly and fast on apparently healthy individuals [Katritsis *et al*., 2016]. SCD is responsible for more than 60% of all deaths caused by CVD [Brugada & Brugada, 1992]. In the young, death can be the first and only clinical manifestation of SCD. Globally, this SCD causes nearly around one million deaths per year in the Western World, which represent a higher mortality than AIDS, lung and breast cancer together.

The most important cause of SCD is coronary artery disease, corresponding to 75-80% of the cases. The other 15-20% have a genetic component, are inherited, and have a higher prevalence in the young population. Genetic pathologies associated with SCD are classified in two different categories [Ackerman *et al*., 2013]:

I. **CARDIOMYOPATHIES**: They are associated with structural abnormalities in the heart which are the responsible of arrhythmia [Antzelevitch *et al*., 1998]. These diseases are caused by mutations in genes which are encoding sarcomeric proteins, components of the cytoskeleton and desmosomal proteins [Lim *et al*., 2012].

Cardiomyopathies [Lim *et al.*, 2012] represent about 5-10% of SCDs and they include: Hypertrophic Cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM) and Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC).

II. **CHANNELOPATHIES (**ELECTRICAL DISORDERS**)**: They are disorders of the electrical activity of the heart due to ion channel dysfunction [Fernández *et al.*, 2017], without presenting structural heart defects. The same genetic variant may be associated with different channelopathies [Cowan *et al.*, 2008]. In addition, there is an overlap of the syndrome's causes, thus the clinical phenotype of more than one disease may coexist in a single patient (*Figure 4*). They include [Lim *et al.*, 2012] the following diseases: Long QT Syndrome (LQTS), Brugada Syndrome (BrS), Short QT Syndrome (SQTS) and Catecholaminergic Ventricular Polymorphic Tachycardia (CVPT).

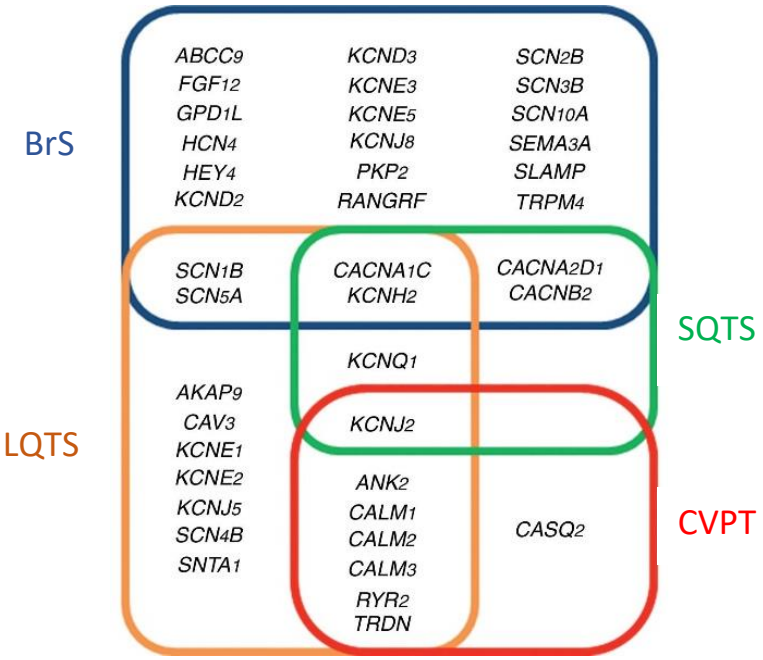


***Figure 4**. Overlapping diagram between genes associated with BrS, Short QT Syndrome, Long QT Syndrome and Catecholaminergic Ventricular Polymorphic Tachycardia. From Revista Española de Medicina Legal.*

1.2.3. BRUGADA SYNDROME

BrS was described [Antzelevitch *et al.*, 2008] as a new cardiac disorder in 1992 consisting on syncopal episodes and/or sudden death in patients with a structurally normal heart [Antzelevitch *et al.*, 1998]. BrS exhibits a characteristic ECG pattern, with an ST segment elevation in the right precordial leads (V1 to V3) and a negative T wave, which gives to the ECG a shark's fin shape [Brugada & Brugada, 1992]. This ST segment elevation in the ECG of BrS patients is often dynamic.

Three types of ST segment elevation (*Figure 5*) are generally recognized: ***type I*** is a diagnostic criteria of BrS and is characterized by a coved ST segment elevation exceeding or at 2 mm (0,2 mV) followed by a negative T wave [Brugada & Brugada, 1992]; ***type 2*** has a saddleback appearance with an ST segment elevation of more than 2 mm followed by a through displaying higher than 1 mm ST elevation and a positive or biphasic T wave; and ***type 3*** has either a saddleback or coved appearance with an ST segment elevation of less than 1 mm. *Type 2* and *type 3* EGG are not diagnostic of BrS.

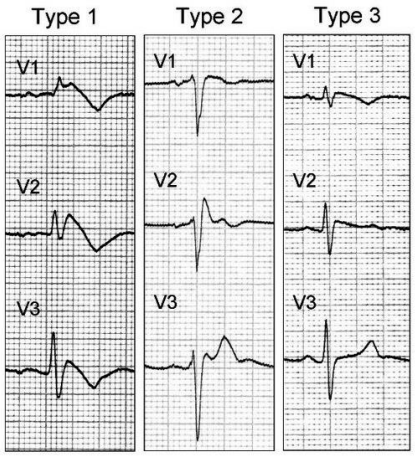


Figure 5*. Representation of the three types of ECGs associated with BrS. From Brugada Syndrome Wikipedia.*

BrS is defined as a genetic disease with an autosomal dominant pattern of transmission [Antzelevitch & Fish, 2006]. It is associated with high risk of SCD and ventricular fibrillation (VF) and is responsible of 4% of all sudden death and 20% of SCD in patients without structural abnormalities. In men, the incidence is higher than in women [Konstandina *et al*., 2016], probably due to gender differences in ion channels expression. Because individuals with BrS exhibit a characteristic ECG pattern [Peng-Sheng & Priori, 2008], BrS is also described as an electrophysiological disorder caused by dysfunction of ion channels involved in generation of the action potentials, especially sodium channels.

The first genetic basis of BrS was described after the identification [Berne & Brugada, 2012], in patients with BrS, of genetic variants in the *SCN5A* gene, which encodes the α-subunit of the voltage-gated cardiac sodium channel ($Na_V1.5$). $Na_V1.5$ channel is the responsible of the cellular intake of sodium and plays a key role in phase 0 of the cardiac action potential [Kapplinger *et al*., 2010].

SCN5A [Bezzina *et al*., 2013], located on locus 3p21, is the main gene associated to BrS [Mademont-Soler *et al*., 2016]. *SCN5A* variants account for nearly 18-30% of BrS cases. To date, over 300 mutations in *SCN5A* have been linked to BrS. Approximately 30 of these mutations have been studied in heterologous expression systems and have shown to result in loss-of-function of the channel or failure to express correctly, resulting in reduced sodium current [Kapplinger *et al*., 2010].

In addition to *SCN5A*, other 17 genes [Mademont-Soler *et al*., 2016] have also been described to be associated with BrS. Those genes encode for calcium channels or sodium channel beta subunits [Selga *et al*., 2015], among others. Nevertheless, in approximately 70% of BrS cases, the genetic cause is still unknown [Antzelevitch & Eyal, 2008]. Our working hypothesis is that BrS cases without mutations in *SCN5A* or associated genes could be explained by genetic variants in non-coding regulatory regions that result in an alteration of ion channels genes expression [Tarradas *et al.*, 2017].

However, the study of genetic variants in non-coding regulatory is quite challenging. The functional effect of genetic variants identified in coding regions, which represent 2% of the genome, is easy to predict because we know the genetic code. In contrast, the functional effect of genetic variants found in non-coding regions [Bezzina *et al.*, 2013], which represent the remaining 98%, is difficult to predict because the code used by these regions is unknown. Transcription factors (TF) play a very important role in the interpretation of this non-coding code [Tarradas *et al.*, 2017].

1.2.4. THE CTCF REGULATOR AND THE CARDIAC TRANSCRIPTION FACTORS

Eukaryotic transcription is a complex process which involves many different components: sequence-specific DNA binding factors, chromatin regulators, transcription machinery and cofactors that work together modulating gene expression based on the current needs of the cell. Sequence-specific regulators or TF control multiple aspects of transcription [Chen *et al*., 2017].

These factors recognize and bind specific sequences of DNA, modulating the expression levels of certain genes.

Cardiac development and function are closely regulated by a set of TF and chromatin remodeling proteins that form evolutionarily conserved regulatory networks [Tarradas *et al.*, 2017]. Cardiac TF such as NKX2-5, GATA4 or TBX5 play a key role during embryonic and adult expression. They specifically interact with non-coding regulatory sequences of cardiac genes to activate or suppress their expression. Genome-wide association studies (GWAS) have identified genetic variants in binding sites of these cardiac TF that are associated to the risk of cardiac arrhythmias [Poulos *et al*., 2016].

CCCTC-binding factor (CTCF) is a highly conserved TF containing an eleven-zinc finger DNA-binding domain capable of interacting with both DNA and other proteins [Somi *et al*., 2015]. CTCF [Ong & Corces, 2014] can function as a transcriptional activator, as a repressor or as an insulator protein [Somi *et al*., 2015], blocking the communication between enhancers and promoters. Unlike most sequence-specific regulators, which need cooperative interactions with other TF [Chen *et al*., 2017], CTCF can access DNA in a relatively autonomous way through its rich interface junction.

Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) studies performed in 20 human cell types showed that CTCF recognizes about 50,000 different DNA sequences which are abnormally long. Computational analysis of these sequences identified a CTCF-binding consensus motif of around 20 bp [Ong & Corces, 2014].

In summary, these observations suggest that the variants in regulatory elements (TF binding sites) of cardiac ion channel genes may have functional effects in the development and homeostasis of the electrophysiological properties of the heart.

1.2.5. MACHINE LEARNING AND DEEPBIND TECHNOLOGY

This work is part of a project called Regulome-Seq, which aims to characterize genetic variation at regulatory regions nearby genes associated with BrS. To carry out this project, 1,293 genomic loci that potentially host TF-binding sites nearby BrS-associated genes were pre-selected after computational analysis of public databases [Konstadina *et al.*, 2016].

To characterize genetic variation in these regions, this set of 1,293 loci was selectively captured and sequenced in a cohort of 89 diagnosed-BrS cases (BrS cohort) that did not carry variants in coding regions. After sequencing, a total of 5,508 variants were identified.

Once identified, the next step was to examine the functional relevance of these variants. However, to experimentally validate all the variants identified in this project is unrealistic because it would require a high amount of time and money which makes it impossible to carry on.

For this reason, we are using a computational tool named DeepBind [Min *et al*., 2017], which is based on machine learning, to predict the effect of these variants on TF binding [Alipanahi *et al.*, 2015]. This program uses artificial neural networks that are able to capture information that is not explicit and give you a final prediction. DeepBind uses deep learning to analyze how proteins bind to DNA and RNA [Ngiam *et al.*, 2011], allowing it to detect mutations that could disrupt cellular processes and cause disease. DeepBind can be applied to both microarray and sequencing data and it can learn from millions of sequences through parallel implementation on graphics processing units (GPU).

The DeepBind algorithm is trained using real ChIP-seq data as input and the program creates a model that best fits the actual data (*Figure 6*) [DeepBind, 2014]. Once the model is created, it is possible to analyze any sequence.

The basis of the program consists in a screening of the positions of the sequence being treated and gives a prediction of the effect of a variant on the binding of a given transcriptional factors.

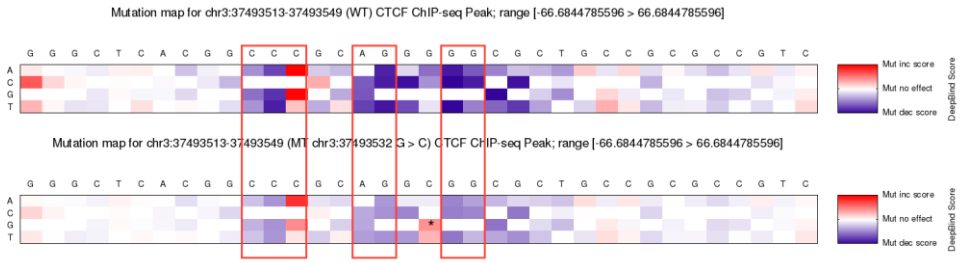


***Figure 6**. Example of mutation maps for reference (above) and alternative (below) variants using DeepBind [Alipanahi et al., 2015]. Red score indicates that the mutation is predicted to increase binding, while blue score indicates reduced binding. The color intensity shows the effect that the mutation makes in the sequence. From IDIBGI.*

The research team will be able to link these algorithm predictions to BrS data. The resulting DeepBind models can be used to identify binding sites in problem sequences and to quantify the effects of new mutations [Min *et al.*, 2017].

1.3. SCIENCE COMMUNICATION

The ability to communicate Science is based on innovation [Binder *et al.*, 2016]. Scientists apply scientific thinking to determine how to better communicate their Science, arts thinking and performing arts are also used as a novel tool in this field. Science progresses through experimentation, evidence, and being able to communicate the relevance and impact of new scientific ideas and discoveries can enhance society culture in other fields [McNutt, 2017]: a scientific experiment is rather unfinished until its results communicated. Communication should not only take place between experts because this process does not end until the results of the research reach the society [Brossard *et al.*, 2005].

However, communication of Science is not for everyone and not every scientist is able to do it correctly and properly (even though in an ideal situation, all researchers should be able to communicate well enough). It is quite hard to expect that all scientists use Twitter [Binder *et al.*, 2016] and other social networks, participate in different workshops in their local schools, or run a blog to communicate Science to High school students and society in general. At the same time, scientists who are willing to take care of this communication [Gauchat, 2011], are fighting to get more attention and air time for Science and, in general, for culture in newspapers and mass media. Actually, one may think that a society becomes more democratic when there is a normalization of Science on public and private TV.

There is an Albert Einstein's quote which can state perfectly this idea: "If you can't explain it simply, you don't understand it well enough". As experts, scientists have a deep knowledge of particular subjects but they would have to be able to adapt this knowledge to other levels of understanding, educational level, and environments [National Academy of Sciences, 2017].

Although public communication may seem very different from scholarly communication of Science [Vieta, 2016], there are several points which are quite similar. Moreover, no one can argue that giving a talk to peers at a scientific conference is the same as standing in front of a group of middle schoolers to teach them about molecular biology [National Academy of Sciences, 2015], or that writing a peer-reviewed research article is the same as writing in a Science blog for High school students.

The main objective of scientific communication is thus to accurately and clearly communicate scientific knowledge; hence it is intimately linked with the scientific method. What is really important is to know who the audience is and what kind of language and examples the scientist should use [Gauchat, 2011]. In this way, it is better to use preferably a local language (i.e., Catalan) if a scientist wants to communicate new results in a local area, but English, as a *lingua franca*, can also be used. Moreover, it is interesting to know that some studies [National Academy of Sciences, 2017] report that the combination of research's explanations with the use of visually spectacular experiments with homemade material or props is the best way to explain difficult concepts [Vieta, 2016]. Usually, we use a wide diversity of methodologies to communicate Science in a more interesting point of view [McNutt, 2017].

Other objectives for which a dissemination activity is carried out can be diverse [Vieta, 2016]: to literate society in Science and technology, to promote scientific vocations, to increase the scientific culture between society by making it a more democratic society, to obtain public and private funds for scientific research, and to fight against the growth of pseudo-sciences.

Scientific communication plays a very important role at the Universitat de Girona because since 2008 many dissemination activities promoted by the **Càtedra de Cultura Científica i Comunicació Digital** (C4D) and **Institut de Química Computacional i Catàlisi** (IQCC) were developed. C4D tries to connect educational methodologies with Science, Internet technologies, magic, and recreational Science.

This current project has different financial ways to approach Science to non-scientific people using low-cost communication where low-cost must be considered in terms of time, energy, attention and money. In particular, the interplay between magic and Science has always been a key issue for C4D. The project takes advantage of their interface to promote scientific vocations and to increase the scientific culture of society.

2. OBJECTIVES

This Bachelor Thesis has two main objectives as it consists of two related, different parts. First, to identify and characterize genetic variants in regulatory regions of genes associated with BrS to understand the code and meaning of non-coding genome regions. Second, to communicate what BrS is and why is so important to understand the genome as a human society.

As previously explained in the Introduction, sequencing of pre-selected regions in 89 BrS patients could identify more than 5,000 non-coding variants in these regions. The hypothesis of this project was that those variants that overlap with TF binding sites would be more likely to affect the binding of the other factors and therefore be more relevant to BrS. Then, in order to evaluate whether the presence of a variant is affecting the CTCF factor, DeepBind is used to predict the effect of each variant on the CTCF region.

Taking all these into consideration, the specific objectives of this work are the following:

1. To clone reference variants and alternative variants identified in CTCF motifs in a pGL4 reporter plasmid.

2. To perform luciferase assays with vectors pGL4 and CTCF-VP64 in H9c2 cardiac cells.

3. To characterize these reporter luciferase assays in order to correlate and compare the experimental results with the DeepBind predictions to study the effect of variants in the CTCF region.

4. To establish a method to communicate what mutations and BrS are and why are being studied in this research team.

5. To design new Science communication workshops and improve the existing ones with the use of magic, illusion and visual objects as a communicative resource.

6. To contribute to the open scientific attitude in the improvement of the educational process of Science in secondary education students and society with the idea that Science is another field of culture.

3. MATERIALS AND METHODS

This study includes the analysis of the 34 variants overlapping CTCF-binding sites identified in BrS patients [Chen *et al*., 2017]. For each variant, two different plasmid constructs were created: a version corresponding to the sequence with the CTCF intact motif (*wild-type* or reference sequence) and the same sequence with the mutant CTCF motif identified (alternative sequence).

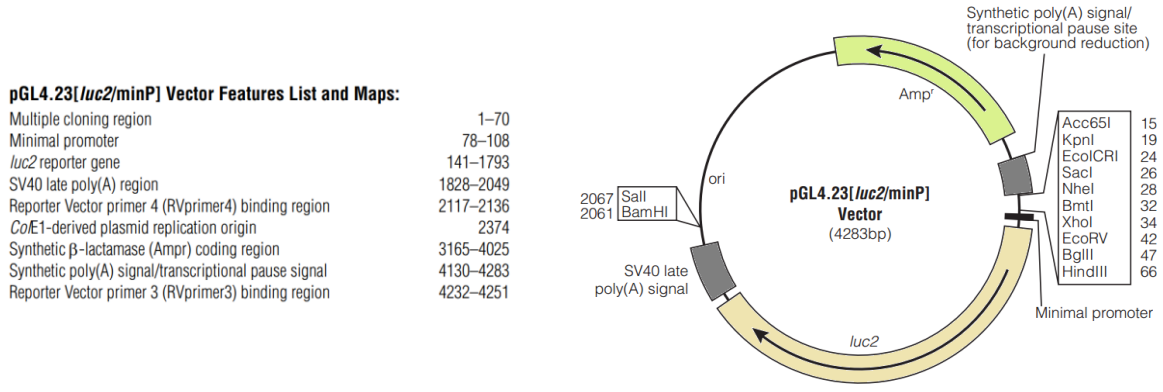
3.1. HUMAN PATIENTS' SAMPLES

The variants with I have been working corresponded to 89 patients of BrS who did not have any mutation in the coding regions of the genes that encode for ion channels. Patients diagnosed with BrS were collected over some years in previous studies. The clinical diagnosis was accepted as positive when the patients had a BrS (*type 1*) ECG spontaneously diagnostic.

All patients included in the genetic study had signed a written informed consent. This study complied with the requirements of the 1975 Declaration of Helsinki and was approved by the ethical committee of the institution (Hospital Josep Trueta - IDIBGI, Girona).

3.2. PLASMID CONSTRUCTS AND CLONING

The 68 sequences corresponding to the CTCF motifs (including reference and alternative) were cloned into a pGL4.23 vector (*Figure 7*) [Cheng *et al*., 2017]. This vector has a multiple cloning region (MCR) located upstream of a minimal promoter, a gene of resistance to Ampicillin that allows its selection in *E. coli*, and a reporter *luc2* gene [Wheeler, 2016]. The *luc2* gene is a synthetically-derived luciferase sequence with humanized codon optimization, and it is a part of pGL4 backbone vector.



pGL4.23[*luc2*/minP] Vector Features List and Maps:

Multiple cloning region	1–70
Minimal promoter	78–108
luc2 reporter gene	141–1793
SV40 late poly(A) region	1828–2049
Reporter Vector primer 4 (RVprimer4) binding region	2117–2136
*Col*E1-derived plasmid replication origin	2374
Synthetic β-lactamase (Ampr) coding region	3165–4025
Synthetic poly(A) signal/transcriptional pause signal	4130–4283
Reporter Vector primer 3 (RVprimer3) binding region	4232–4251

***Figure 7**. pGL4.23 Plasmid DNA Restriction Map [Wheeler, 2016]. The luc2 gene is a synthetically-derived luciferase sequence with humanized codon optimization. From Promega.*

This pGL4.23 displays increased reporter gene expression with codon optimization of synthetic genes for mammalian expression as well as reduced background and risk of expression artifacts with removal of cryptic DNA regulatory elements [Promega Corporation, 2015].

Single-stranded DNA oligonucleotides sequences selected for cloning were ordered at *Conda Laboratories*. *Figure 8* shows an example of oligonucleotides designed to generate the reference and alternative CTCF-binding motifs corresponding to variant 31 and insert them into pGL4.23. Forward oligonucleotides hybridize with reverse ones, although reverse oligonucleotides contain 4 extra nucleotides at both ends that do not hybridize with the forward chain. Hybridization of forward and reverse oligonucleotides generated cohesive ends that were complementary to the ones generated after digestion of the pGL4.23 vector with KpnI and NheI restriction enzymes. In addition, an extra nucleotide (in bold in *figure 8*) was added at both ends of all oligonucleotides forward and reverse to create completely different restriction sites after ligation.

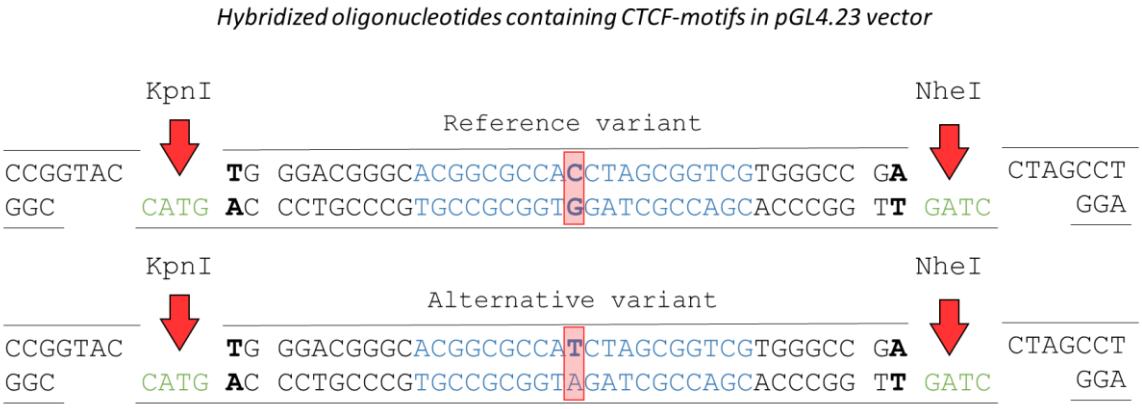


***Figure 8**. Cloning strategy of hybridized oligonucleotides containing CTCF-motifs in pGL4.23 vector. Example of the variant 31, in reference (top) or alternative (bottom) form. In this case, the position of the variant is altered with a change of a T, replacing a C, in the forward sequence. The blue sequence corresponds to the CTCF motif and the based is highlighted in orange at the position of the variant. From IDIBGI.*

3.3. ANNEALING

The oligonucleotides of the 34 variants of study (reference and alternative) were first hydrated with Tris-EDTA (TE) buffer to have them at a concentration of 100 µM. To generate each variant, we performed annealings with the forward and reverse oligonucleotides. For each annealing reaction, 6,5 µL nuclease free ddH2O, 1 µL forward oligo (100 µM), 1 µL reverse oligo (100 µM), 1 µL T4 PNK Buffer (with ATP) and 0.5 µL PNK enzyme were mixed and incubated for 5 minutes at 95°C (*Table 1*).

The annealing procedure in the thermocycler is divided into different steps which are gradually warmed up and cooled down to room temperature in a PCR thermocycler.

***Table 1**. Sketch of temperatures.*

50 minutes	30 minutes at 37°C
	5 minutes at 95°C
	Gradually cooled 1°C/12 seconds from 95°C to 25°C
	25°C for 5-10 minutes
	∞ at 4°C

3.4. LIGATION

Following annealing, each pair of hybridized oligonucleotides was ligated to the multiple cloning region (MCR) of pGL4.23 previously digested with KpnI and NheI. Below, a diagram of the pGL4.23 MCR highlighting the restriction sites where the hybridized oligonucleotides of interest were introduced.

…C T G G C C G G T A C C T G A G C T C G C T A G C C T C G A G…

The ligation reaction included 6 µL nuclease free ddH2O, 1 µL oligos (1/250 diluted), 1 µL pGL4.23 plasmid (50 ng/µL), 1 µL ligase buffer (with ATP) and 1 µL ligase. Reactions were performed at room temperature for 1 hour.

3.5. TRANSFORMATION

Ligation products were transformed into **DH5α™ MAX Efficiency®** cell competent bacteria [Thermo Fisher, 2017]. With this strain, transformation efficiency is up to 1×10^9 transformants/µg of DNA plasmid, and gives a high production of plasmids and a greater stability of the inserts thanks to the *endA1* and *recA1* mutations. Since transformation of these bacteria with an excess of DNA diminishes the efficiency of transformation, it is recommended to transform 1 ng to 10 ng of DNA.

In this experiment, the transformation protocol was performed with 16 µL of bacteria and 1 µL of the ligation product. Competent cells and DNA were first incubated 30 minutes on ice, following a heat shock step at 42°C for 45 seconds. After incubating on ice for 2 minutes, 250 µL of SOC medium (Thermo Fisher) was added, and bacteria were recovered in the incubator shaker at 220 rpm for 1 hour at 37°C. Bacteria cultures were plated in LB-agar plates containing 100 µg/mL of ampicillin. Bacteria plates were incubated O/N at 37°C.

3.6. SEQUENCING OF CLONED PLASMIDS

To verify a correct insertion into the plasmid named pGL4.23 (*Figure 9*), the colonies obtained after the transformation were sequenced following the next steps:

3.6.1. MINIPREPS

A plasmid preparation is a method of DNA extraction and purification for plasmid DNA. Many methods have been developed to purify plasmid DNA from bacteria. MiniPrep is used to check the preparations and to decide to make bacteria grow or not in MidiPrep or MaxiPrep.

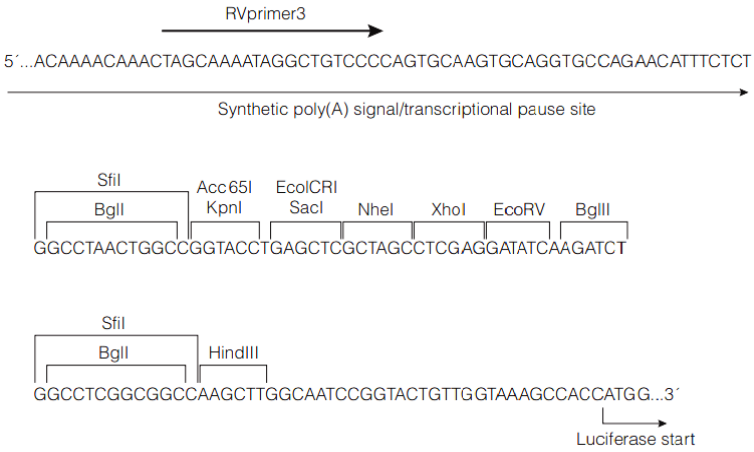


Figure 9*. Multiple cloning region for the pGL4.23 [luc2/minP] vector. From Promega.*

In this study, two colony forming units of each cloned variant were inoculated in separate tubes with 4 mL of LB medium, a a nutrient-rich media, (with 100 µg/mL of ampicillin), and incubated O/N at 37°C on a shaker at 220 rpm. Extraction of plasmid DNA was performed following the MiniPrep protocol (QIAprep Spin Miniprep Kit, QIAGEN), with some modifications.

Briefly, 2 mL of bacterial suspension was centrifuged for 1 minute at 12,000 rpm, and the supernatant was discarded. Bacteria pellet was resuspended in 250 µL of suspension buffer followed by 250 µL of lysis buffer. After 5 minutes, lysis reaction was stopped with neutralization buffer. Bacterial lysate was then centrifuged for 5 minutes at 13,200 rpm. The supernatant was applied into a QIAprep spin column, washed thoroughly and eluted in 50 µL of EB buffer.

Finally, plasmid DNA was precipitated with isopropanol and washed with 70% ethanol, in order to obtain a sediment with the purified DNA plasmid. It is important to wait an hour at room temperature to get the sediment dried. Later on, this purified DNA fragment was resuspended with an amount from 20 µL to 40 µL of TE buffer depending on the quantification.

3.6.2. QUANTIFICATION OF DNA PLASMIDS

Plasmid DNA was quantified using the NanoDrop™ spectrophotometer (Thermo Scientific). If the final concentration of DNA was higher than 3,000 ng/µL, the plasmid was pre-diluted to measure the concentration. The ratio of absorbance at 260 nm and 280 nm is used to assess the purity of DNA and RNA. A ratio of approximately 1.8 is generally accepted as "pure" for DNA. If the ratio is lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm. All DNA plasmid samples analysed had correct ratios of absorbance.

3.6.3. BIGDYE REACTION

The sequencing reaction is based on the enzymatic Sanger method [Promega, 2007]. The classical chain-termination method requires: single-stranded DNA template, DNA primer, DNA polymerase, deoxynucleosidetriphosphates (dNTPs) and some modified di-deoxynucleotidetriphosphates (ddNTPs) radioactively or fluorescent labelled. These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides when a modified ddNTP is incorporated.

To carry out the BigDye reaction [Applied Biosystems, 2011], we used 200 ng of plasmid DNA (it is recommended to use between 150 ng to 300 ng) and the in-house designed primer CTCF-cloning test (5'-TTCCTATGCCTACTGCCTCG-3'), which is oriented in reverse direction. This BigDye reaction uses BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Scientific), which provides pre-mixed reagents for Sanger sequencing reactions. The kit reagents are suitable for performing fluorescence-based cycle sequencing reactions on single-stranded or double-stranded DNA [Applied Biosystems, 2011].

With this kit is possible to carry out a reaction in the same well where the four terminators marked are added. The BigDye reactions were carried out in 96-well plate sealed with thermic lid and placed in the thermocycler with the following amounts of reagents: 2 µL nuclease free ddH2O, 0.5 µL CTCF-cloning test primer (10 µM), 1 µL BigDye buffer, 0.5 µL BigDye reaction mix, 1 µL DNA (200 ng/µL). Big Dye reaction consisted in:

96°C	1 min	
96°C	10 sec	×25 cycles
50°C	5 sec	
60°C	4 min	
4°C	∝	

The denaturalization, hybridization and extension steps are repeated 25 times in this thermocycler cycle to get a better result.

At the end of the BigDye reaction, fragments of different sizes were obtained according to the incorporation of the ddNTP at the 3' ends. Furthermore, the fragments are marked with different fluorochromes according to the ddNTP that they have incorporated.

3.6.4. PRECIPITATION

The product of the BigDye reaction was precipitated to purify DNA for further sequencing. First, 80 µL of ethanol-acetate were added to each well and incubated for 10 minutes at room temperature, and the plate was centrifuged for 45 minutes at 2,000 g. The supernatant was then removed and the plate was left to dry out for 10 minutes. Next, 150 µL of 70% ethanol was added and the plate was centrifuged again left to dry out. Finally, 10 µL of formamide were added to the wells to be sequenced and the plate was incubated 7 minutes at 95°C in the thermocycler.

3.6.5. SEQUENCING ANALYSIS

The product of the BigDye reaction was sequenced using the Genetic Analyzer 3130XL (Applied Biosystems) sequencer. This sequencer allows to identify the ddNTP of each fragment with a resolution by the difference in length as it incorporates a laser fluorescence detector. The complete sequence of the amplified variant fragment was analyzed with the *Sequence analysis* software, which allowed the validation of the correct insertion of the hybridized oligonucleotides into the pGL4.23 vector.

3.7. MIDIPREPS

The clones corresponding to the variants that had been correctly inserted were grown in larger quantities to prepare MidiPreps. The MidiPrep purification leads to higher amounts and higher purity of DNA plasmid than MidiPreps. In this study, we followed the protocol provided by the manufacturer (Plasmid Plus Midi Kit, QIAGEN).

The clones corresponding to the variants that had been cloned correctly became to grow from 1mL that was conserved at 4°C from MiniPrep. LB medium is a nutrient-rich media commonly used to culture bacteria in the lab. LB agar plates are frequently used to isolate individual (clonal) colonies of bacteria carrying a specific plasmid.

The MidiPrep (Plasmid Plus Midi Kit, QIAGEN) can extract and purify DNA plasmid. In this study is enough to use MidiPreps instead of MaxiPreps because for luciferase assays it was not necessary to transfect high amounts of DNA.

MidiPreps have a greater degree of purity than MiniPreps and more amounts of DNA plasmid are obtained. The MidiPrep protocol (Plasmid Plus Midi Kit, QIAGEN) provided by the manufacturer was followed in the laboratory. It was quite similar to MiniPrep steps.

First, 1 mL of bacterial culture kept at 4°C from the MiniPrep was added to 100 mL LB medium with 100 µg/mL of ampicillin and incubated at 37°C O/N on the shaker (220 rpm). Similar to the MiniPrep protocol, the bacterial pellet obtained after centrifugation was lysed in lysis buffer and then neutralized. Then, the bacterial lysate was applied into the QIAfilters, which retained the remaining cells and other products. A treatment with Endotoxin Removal (ER) buffer was performed during 30 minutes on ice to eliminate possible endotoxins.

Next, the purified lysate was applied to the QIAGEN-tips (columns that retain DNA). After washing with different specific buffers from the kit, impurities were removed. DNA was eluted from the columns with 15 mL of elution buffer. Plasmid DNA was precipitated with isopropanol and washed with 70% ethanol. The DNA sediment was dried out at room temperature (for about 5-10 minutes) and resuspended in 50 µL Tris-EDTA (TE) buffer.

TE buffer solubilizes DNA while protecting it from its degradation, since EDTA (a chelator of divalent cations such as Mg^{2+}, which are required for nuclease enzymatic activity). Finally, DNA was quantified in the Nanodrop.

3.8. H9c2 CELL CULTURE

Cardiac cells derived from embryonic rat ventricle (H9c2 cells) were obtained from the American Type Culture Collection (ATCC). H9c2 cells were used to perform dual luciferase reporter assays to test the effects of the variants on CTCF binding. These cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich) supplemented with 10% of FBS or Fetal Bovine Serum (Sigma-Aldrich), 1% of Glutamax (Gibco, Life Technologies) and 1% of penicillin-streptomycin (Gibco, Life Technologies) in a biological incubator at 37°C with 5% CO_2.

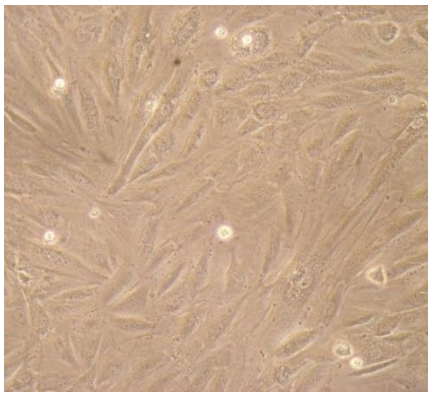


***Figure 10**. H9c2 cells morphology. From IDIBGI.*

The luciferase reporter assays were performed in H9c2 cells (*Figure 10*). This cell line comes from embryonic cardiac ventricular rat cells. H9c2 cells were grown at 37°C with 5% CO2 in Dulbecco's Modified Eagle Medium (Sigma-Aldrich) supplemented with 10% Fetal Bovine Serum (FBS, Sigma-Aldrich), 1% penicillin/streptomycin (Sigma-Aldrich) and 1% glutamine (Invitrogen).

H9c2 cells were kept in 75cm2 culture flasks, where they grew forming a monolayer. Subcultures of H9c2 cells were splitted when they reached approximately an 80-90% of confluence.

First, the medium of the culture flask was aspirated, and the cells were washed with phosphate-buffered saline (PBS, Gibco, Thermo Scientific). Then, cells were trypsinized in Trypsin-EDTA 0,05% (Gibco, Thermo Scientific), with the minimum volume to let the product cover the surface of the flask, by incubating them 2 minutes at 37°C in the incubator.

Trypsinized cells were resuspended in 10 mL of DMEM supplemented with antibiotic. This suspension of cells was subcultivated in a new culture flask, at a determined ratio to control the confluence between the cell suspension on the medium.

3.9. LUCIFERASE REPORTER ASSAY

The luciferase reporter assay is used to study transcriptional activity of a promoter or enhancer. This assay is widely used because it is relatively inexpensive and gives quantitative measurements instantaneously. Light emission has been used for almost 100 years to detect experimental changes in biological assays. Luminescence is light emission as a result of a chemical reaction without production of heat. A photon-emitting enzymatic oxidation reaction is called bioluminescence and it occurs in a diverse range of organisms by a class of luciferases.

Firefly (*Photinus pyralis*) and **renilla** (*Renilla reniformis*) luciferases, 61 kDa and 36 kDa, respectively, have been the primary molecules in luciferase-based applications because of their stability. The reaction catalyzed by luciferase enzymes (*Figure 11*) consists on an oxidation of luciferin to oxiluciferin, leading to the emission of luminescence that can be measured in a luminometer. Coelenterazine is oxidized by renilla luciferase to produce coelenteramide and light [Promega Corporation, 2008].

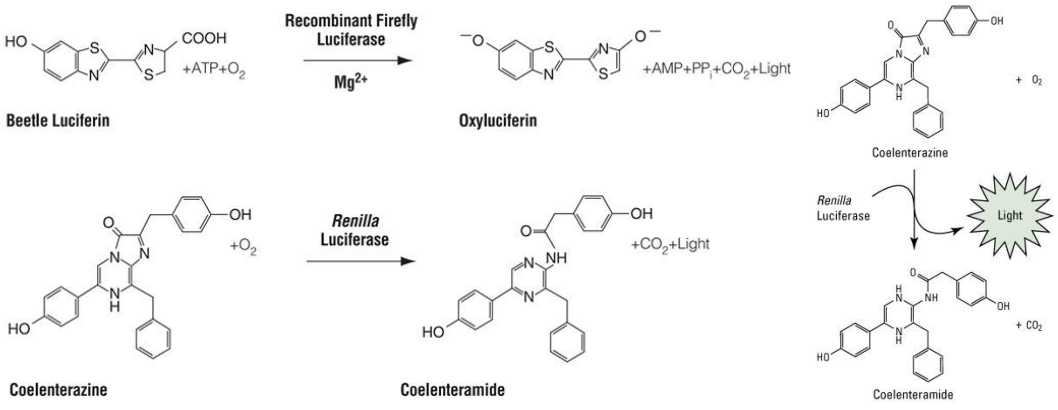


***Figure 11**. Enzymatic reactions catalyzed by firefly or renilla luciferase. From Promega.*

In this study, we used a **dual reporter assay system** (*Figure 12*), since we measured two different enzyme activities to improve experimental accuracy and to allow a more reliable interpretation of the experimental data. The experimental reporter used is the firefly luciferase, which is cloned in the pGL4.23 in our plasmids of interest, and is designed to analyze the effect of the variations on the CTCF motif when it is bound to DNA. The control reporter used in this dual system is renilla (and it is photosensitive). Constitutive expression of renilla luciferase under the EF1α promoter was used in combination with the pGL4.23 vector to normalize firefly luciferase measurements. After the first firefly luciferase signal is read, a second reagent is added to the sample initiating simultaneously the renilla luciferase reaction.

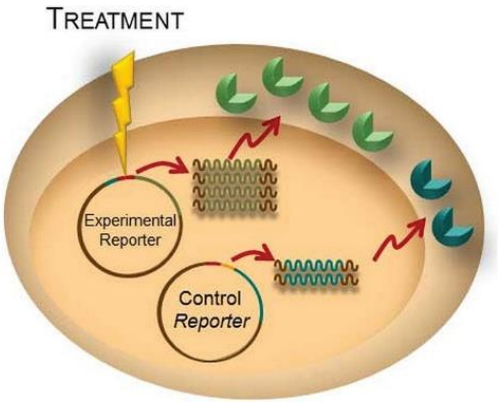
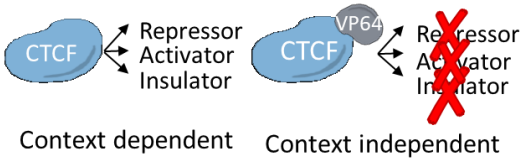


***Figure 12**. Dual Reporter Assay system. From Promega Connections.*

CTCF has a context-dependent activity and can take different functions such as repressor, activator or insulator. However, in this study, we are interested to evaluate its DNA binding abilities, which can be achieved when it is fused to the VP64 protein. VP64 is a transcriptional activator formed by 4 tandem copies of the Herpes Simplex Viral Protein 16 (VP16). When VP64 is fused to another protein domain that can be linked to the gene promoter, it acts as a strong transcriptional activator (*Figure 13*). Therefore, co-transfection of the CTCF-VP64 vector together with the pGL4.23 and renilla vectors allowed CTCF to bind to the CTCF motifs analyzed, activating transcription of the luciferase gene, which then catalyzed the bioluminescent reaction.

***Figure 13**. Context-dependent activity of CTCF (left). When fused to VP64, CTCF becomes context independent and acts as a potent activator when binds to its motif. From IDIBGI.*



The CTCF-VP64 expression vector (200 ng), the EF1α renilla reporter vector (5 ng) and the CTCF motif-luciferase (pGL4.23, 100 ng) were co-transfected in H9c2 cells using Lipofectamine 2000 from Life Technologies and following the manufacturer's specifications (Promega) [Invitrogen, 2013].

The volume of lipofectamine used in these transfections was determined by applying a 1:2 ratio (µg DNA: µL lipofectamine). Transfections were performed in 12-well plates with 60,000 cells per well that were plated 24 hours' prior transfection. For each variant, three biological replicas were performed. Cells were incubated for 6 hours with the DNA-lipofectamine mix in DMEM without antibiotic. Then, medium was replaced by complete DMEM medium with antibiotic, and incubated at 37°C during 2 days.

48 hours after transfection, cells were washed twice in PBS and lysed in 100 µL of Passive lysis buffer (Promega). Cell lysates were recovered and transferred into 1.5 mL tubes. 10 µL of cell lysates were transferred into a 96-well plate. Firefly and renilla luciferase activity were measured using a Glomax 96 microplate luminometer (Promega), which automatically applies dual luciferase reagents (LARII and Stop & Glo Reagent).

3.10. COMPUTATIONAL ANALYSES

DeepBind predictions are bioinformatic tools used to identify the correlation and the efficiency of experimental conditions [DeepBind, 2014]. These CTCF binding to the different study sequences predictions are expressed by *scores*.

DeepBind software was downloaded from *http://tools.genes.toronto.edu/deepbind* webpage based in a Toronto University's source, and the predictions were performed with 36 bp sequences and their specific primers for the correct *in silico* binding.

4. RESULTS AND DISCUSSION

4.1. EXPERIMENTAL RESULTS

4.1.1. CLONING CTCF VARIANTS

The analysis of the non-coding variants identified in BrS patients showed that a total of 66 variants overlap a CTCF motif. From those variants, 80% correspond to SNVs (single-nucleotide variants) and the remaining 20% to indels (insertions/deletions). These indels located in non-coding regions can create new binding sites for TF or disrupt them completely.

In this study, 34 variants of the CTCF motif (with their reference and alternative version) were cloned into the vector pGL4.23. Each variant come from different chromosome locations obtained from preliminary analysis of the Regulome-Seq project. The reference variants correspond to the common variants that can be find in the population (using 1,000 genomes database), whereas the alternative ones correspond to variants identified in patients with BrS. There is an example of one sequence below in *figure 14*:

Reference variant (1,000 Genomes)	GATGGTTCCATGGC**G**CCACCCACTGTCCAGGAATGC
Alternative variant (BrS patient)	GATGGTTCCATGGC**A**CCACCCACTGTCCAGGAATGC

***Figure 14**. The blue region corresponds to the CTCF motif of this variant. The reference nucleotide variant is shown in green, and the alternative version in red. From IDIBGI.*

After cloning, all vectors were sequenced to check that the variants had been correctly introduced. Sequences were examined using Sequencing Analysis Software (Applied Biosystems), a multipurpose software used to view, display, analyze, edit, save, and print sequencing files. With this software, the electropherograms, four-color picture of the sequence analyzed with peaks that are representing the bases, can be analyzed. The interpretation of electropherograms is made by this software (although it should be revised by the user), and in the upper part the corresponding base is assigned.

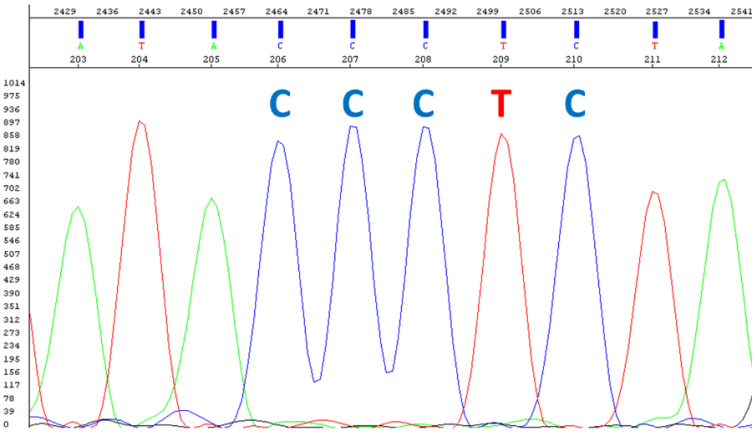


***Figure 15**. Electropherogram corresponding to pGL3.24 plasmid containing variant 35. This variant is the alternative one. It is represented the CCCTC-binding factor (CTCF). From Sequencing Analysis Software (Applied Biosystems).*

The electropherogram represents the DNA sequence of the variant subjected to the analysis [Gilder *et al.*, 2004]. The analysis of all the sequences cloned showed that variants 31 to 66 (except variants 57 and 60) had high-quality electropherograms, since all had well-defined peaks that corresponded to each base in the CTCF region (*Figure 15*). Importantly, all clones had the expected sequence of the variants, confirming that the cloning strategy and procedure had been correctly designed and accomplished. After confirming the correct sequence of the clones, a MidiPrep was prepared for each clone for further transfection experiments.

In general, the CTCF motif is in a central position of the sequence, having several adjacent nucleotides on each side. However, not all cloned sequences contain the CTCF motif in the same position, in some cases it is more shifted to 5' or 3' depending on the binding motif position. Depending on the position of the variants in the CTCF motif, the variants were classified into four different groups (*Figure 16*): **core nucleotides** are those which are very important in the CTCF motif; **non-core nucleotides** are inside CTCF motif but they are not the most important ones in the motif; **nucleotides close to motif** correspond to the 4-side-nucleotides left and the right of the motif; and nucleotides **far from motif** correspond to those that are further away that these 4-side-nucleotides.

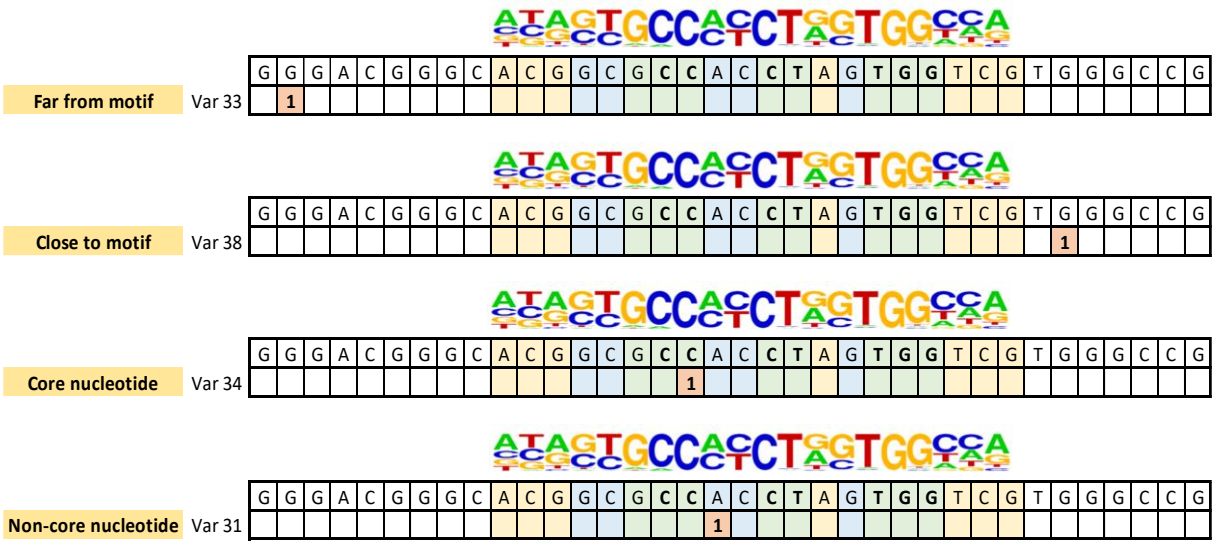


***Figure 16**. Motif detector logo is represented in each example and depending on the position where the variant is it is possible to classify the variants in four different groups: far from motif (white), close to motif (orange), core nucleotide (green) and non-core nucleotide (blue). From all the variants studied, 10 of them correspond to the far from motif group, 5 variants are close to motif, 10 variants are classified in core nucleotides and 9 variants in non-core nucleotides. The CTCF motif detector logo was found in SPANR (Splicing-based Analysis of Variants) and it is based in D00328.018 ID number. From IDIBGI.*

From all the variants studied, 10 of them correspond to the far from motif group, 5 variants are close to motif, 10 variants are classified in core nucleotides and 9 variants in non-core nucleotides.

4.1.2. LUCIFERASE ASSAYS

To evaluate the effect of the variants studied in CTCF binding, luciferase assays were performed for each variant. To eliminate the variability between samples, the activity of the reporter luciferase of each sample was normalized by the activity of the reporter Renilla.

The mean of 3 biological replicas was made for each variant, obtaining a final value of Luciferase/Renilla and represented in a plot. Luciferase assays were repeated in those cases where the results between the replicas showed a high variability in standard deviation.

The effectiveness and correct design of the luciferase protocol had been previously validated. Luciferase reporter experiments were performed after transfection of H9c2 cells with increasing amounts of the CTCF-VP64 expression vector together with the pGL4.23 vector (either containing a random sequence or the consensus CTCF motif). CTCF present in H9c2 cells would be able to bind to the consensus region, but not to a random sequence. As expected, the results obtained showed that when the CTCF-VP64 was transfected with the pGL4.23 containing and a random sequence, luciferase activity values were almost non-existent. In contrast, co-transfection with the consensus CTCF motif lead to a dose-dependent increase in luciferase activity.

This result demonstrate that the CTCF-VP64 protein can bind to the CTCF cloned consensus motif and activate the expression of the luciferase reporter gene. The luciferase enzyme can catalyze then the bioluminescent reaction in the presence of the substrate. On the other hand, the presence of random sequences without any CTCF motif did not allow the binding of CTCF-VP64 and this kept the reporter gene silenced.

Additionally, in these luciferase experiments with the CTCF motif, the higher luciferase activity was observed when the higher dose of CTCF-VP64 transfected was used. Therefore, at higher values of Luciferase obtained, better levels of binding with the CTCF to the sequence analyzed happened. At low concentrations of CTCF-VP64 the effect of this activity is not enough to activate the expression of the luciferase gene.

To examine the effect of each alternative variant versus its reference version in CTCF binding, H9c2 cells were transfected with CTCF-VP64 expression vectors together with each of the pGL4.23 cloned variants. 48 hours after transfection, cells were lysed and luciferase assays performed.

The normalized luciferase activity obtained for each variant indicated the CTCF-VP64 binding level in the CTCF motif in that sequence. In the presence of the CTCF consensus motif, CTCF-VP64 will bind and activate the expression of the luciferase gene. The presence of a variant within the motif can affect CTCF binding, and therefore the luciferase expression (*Figure 17*).

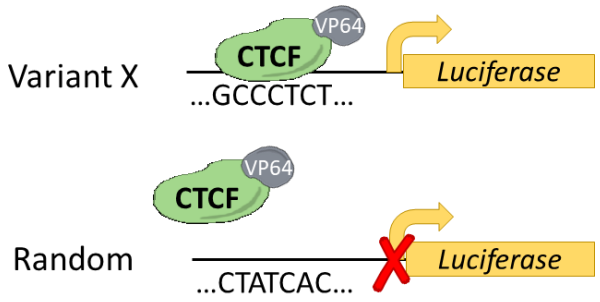


***Figure 17**. The presence of a variant within the motif can affect CTCF binding, and therefore the luciferase expression. There is an example of a variant that contains a CTCF motif (top) and a random sequence (bottom). From IDIBGI.*

To predict the effects of CTCF motif binding with each sequence is really difficult, because it is not known exactly which gene is regulating these genes. So, software programs were developed to try to predict these bindings. CTCF present in H9c2 cells would be able to bind to the consensus region, but not to a random sequence. The CTCF binding with the VP64 transactivator in low amounts does not achieve a transcriptional effect significantly higher. The global results obtained in this luciferase assays demonstrated that the designed reporter system works correctly and it was able to analyze the functional effect of our variants.

The luciferase assays performed showed that the designed reporter system works correctly and it was able to analyze the effect functional of our variants on CTCF binding. The results of the luciferase assays obtained for all the variants are shown in *figure 18*. From the 34 variants analyzed (reference and alternative), 13 variants showed a statistically significant difference between the reference and the alternative version.

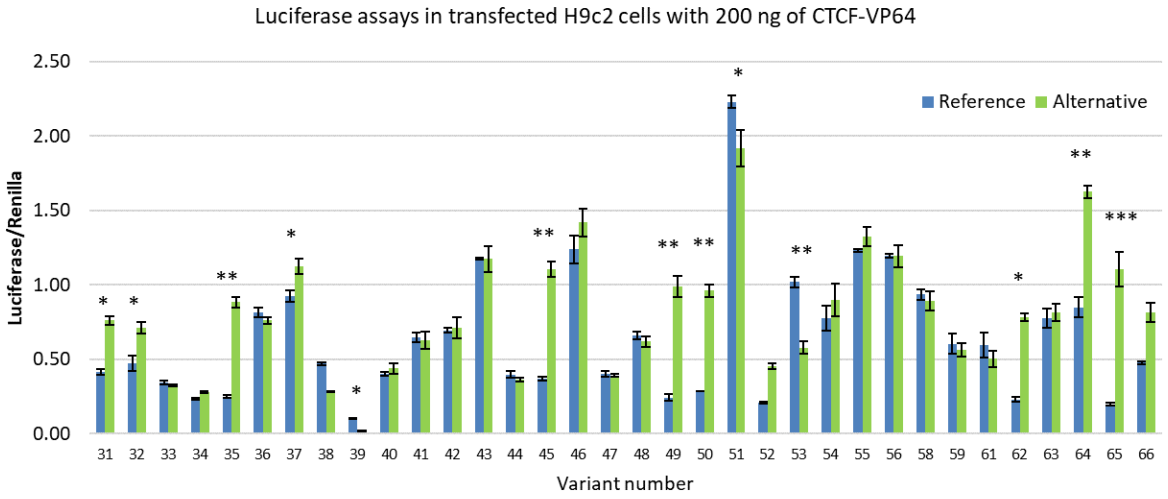


***Figure 18**. Luciferase experiments in H9c2 cells transfected with 200 ng of CTCF-VP64, 200 ng of reference and alternative variants and 10 ng variants of EF1α. Firefly luciferase activity was normalized by renilla. Variants 57 and 60 were eliminated for its value of standard deviation between replicas. The statistical significance was analyzed using a t-test to compare the variants' means and p-values are represented in the figure 21 as: * ($p \leq 0.05$), ** ($p \leq 0.01$) and *** ($p \leq 0.001$). From IDIBGI.*

This means that 38.23% of the variants studied would affect the binding of the CTCF. From those variants that affect the binding, we observed different effects. 10 variants with statistical differences increase CTCF binding in their alternative version compared to the reference one, while 3 of them decrease CTCF binding. Alternative variants 52 and 66 have not statistical differences versus its reference, but they show to increase the binding as well. The main tendency in all the variants (those with statistical significance differences and those without this statistical differences) is to increase the value of the binding in the alternative sequences (21 out to our 34 variants).

As expected, it was observed that all variants in the reference version have a different luciferase activity value, since they correspond to different sequences located at different points in the genome. In addition to that, although they are common in the human population, they differ in their sequence and the binding of CTCF is affected.

4.1.3. Luciferase assays versus DeepBind predictions

To compare the results obtained in the luciferase assays and DeepBind predictions, the effect of the variants on CTCF binding observed in luciferase assays was compared to the predictions obtained with the DeepBind with the same variants. To perform these comparisons, the difference in binding was calculated between reference and alternative values per each variant. The values were represented in boxplots (*Figures 19 & 20*) to evaluate the tendencies in both methods, and the ANOVA statistical test was used to determine differences between positions relative to the four CTCF groups.

The graphs obtained show that the tendency between luciferase assays as a measure of CTCF binding and DeepBind predictions are very similar. There are no significant differences for the luciferase assays results (*Figure 19*) that are far from CTCF motif, close to CTCF and non-core nucleotides positions in the CTCF motif. But when the variant is in the core nucleotides, differences among other position groups were observed. The tendency is different in the DeepBind plot, where no differences were observed between far and close positions to CTCF and between inside positions. In both cases, the higher result of absolute difference |Alternative-Reference| was found in **core nucleotides** positions.

Therefore, the results show, for both methods, that the effect of a variant on CTCF binding is greater when the variant is located within the CTCF motif than that it is outside the motif.

Additionally, within the CTCF motif, the effect of the variant is bigger when it affects the core nucleotides. However, this difference between variants inside and outside the CTCF motif is higher in experimental results. This suggest, that the variants identified in the CTCF motif will likely affect CTCF binding in these regions. This could have effects on genome and alter gene expression.

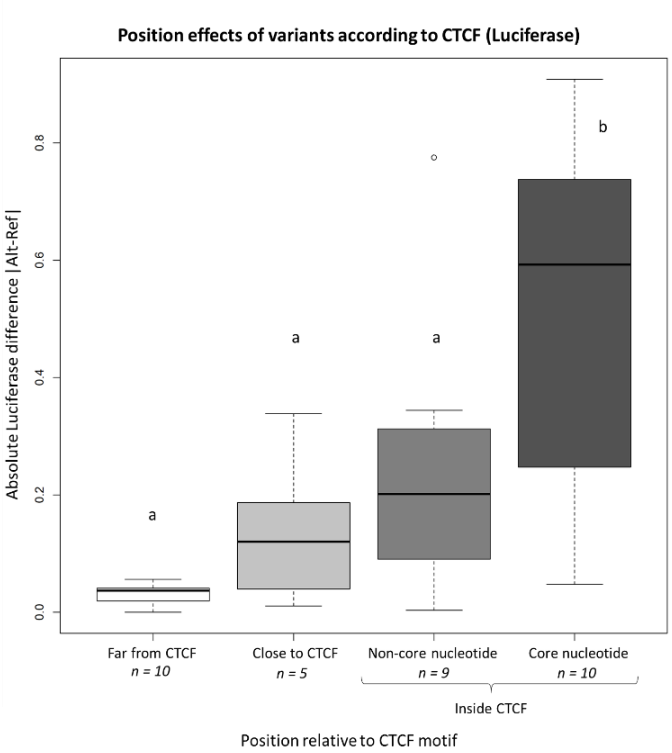
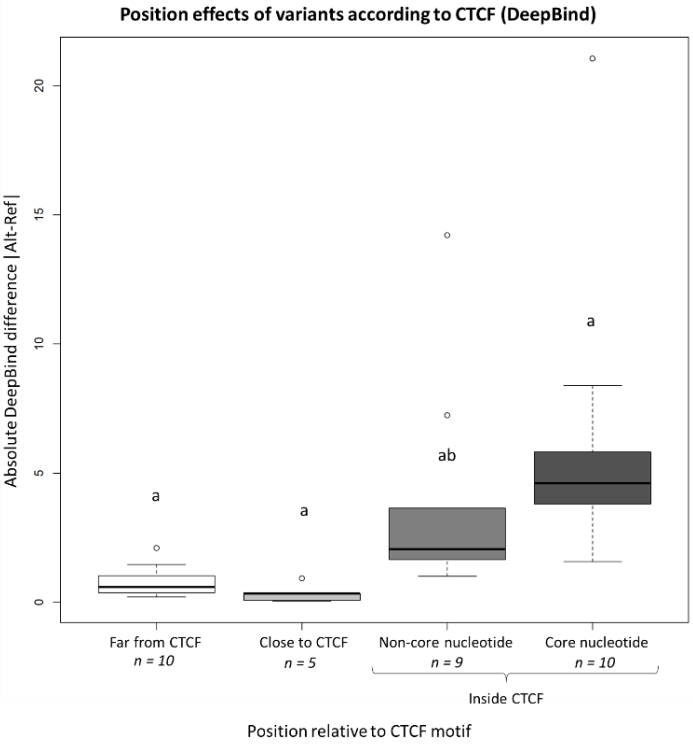


***Figure 19**. Binding differences distribution according to whether the variant is found far, close or inside (core or non-core nucleotides) the CTCF motif in the Luciferase assays for the 34 analyzed variants. The statistical significance was analyzed by an ANOVA and a multiple comparisons of means (Tukey Contrasts). A p-value (α is 0,05) = < 0.001 for mean's comparison between core nucleotide position and far position. A p-value = 0.0088 for the mean's comparison between core nucleotide position and close position. And a p-value = 0.0240 between positions inside the CTCF motif. On the comparison between all the groups, those with an "a" have not any statistically significant differences. And parameter "b" indicates that core nucleotides in CTCF binding are different compared to the other groups. From IDIBGI.*

***Figure 20**. Distribution of binding differences according to whether the variant is found far, close or inside (core or non-core nucleotides) the CTCF motif in the DeepBind for the 34 analyzed variants. The statistical significance was analyzed by an ANOVA and a multiple comparisons of means (Tukey Contrasts).*

A p-value (where α is defined by a value of 0,05) = 0.0128 for the mean's comparison between core nucleotide position and far position, and a p-value = 0.0335 for mean's comparison between core nucleotide position and close position.

On the comparison between all the groups, those with an "a" have not any statistically significant differences. And parameter "ab" indicates that non-core nucleotides in CTCF binding is a group a little bit different from the others, but it cannot be considered really different. From IDIBGI.



The binding ability to the CTCF motif for our variants (*Figures 19 & 20*) differs significantly according to the statistical tests. However, the luciferase assays have been validated correctly, because there is an almost identical tendency between both results (Luciferase assays and DeepBind) data.

The main limitation of the luciferase assays is that they measure the binding of CTCF to a sequence of DNA cloned into a non-chromatinized or circular plasmid. Therefore, the effects on the chromatin structure are not taken into account in this project. Bioinformatics are in silico predictions which are still being developed to improve their reliability by training these systems with real data. These systems present a number of advantages and disadvantages that can be contrasted with the experimental validation method.

To improve the results and get a better correlation between luciferase assays and DeepBind predictions it will be necessary to develop more experiments because our "n" is not big enough.

4.2. SCIENCE COMMUNICATION WORKSHOP PERCEPTION

4.2.1. Miscellaneous Science communication activities

The first part of this Bachelor Thesis' results is based on experimental processes in IDIBGI laboratories. Working in the GenCardio (IDIBGI) laboratory in Santa Caterina's Hospital (Salt) was a such great opportunity to be involved in a really research project. I have contributed with what I could in the laboratory and I have learned a lot from that. However, this current project is not finished and the results in this study are from a small part.

The second part of the project is based on communication of some difficult experimental results related with BrS. It has been seen that there is a lack in Science culture, so there is a great interest to try to improve this problem. A more curious and democratic society may have more arguments. Although I had had some communication opportunities and formation in the past, my skills in this field have been increased with this work. I have understood the actual high importance of making Science accessible to everyone, for example. There is always a way to adapt some scientific knowledge and have it reached to the general public.

And this is what I have learned from C4D. For example, this 2018, in *Temps de Flors* festival in Girona there was a Rubik's cube made of flowers (*Figure 21*).

***Figure 21**. Some pictures of #rubikdeflor in Temps de Flors, 2018.*

The main objective to show this Rubik's cube in *Temps de Flors* (#rubikdeflor) was to communicate one of the C4D's projects about Magic and Science, in particular the III Summer Camp #cemc18. In addition, on weekends and weekdays, some people have been teaching how to solve Rubik's cube. I also had the opportunity to participate in II Encuentro de Magia, Ciencia y Educación (ECME, #ecme17) last December in MUNCyT museum in Alcobendas (Madrid).

***Figure 22**. Some pictures of conferences and workshops in ECME, 2018.*

The main idea of this meeting was to explain new methodologies to incorporate more fun Science at different levels from Universities and High Schools. 20 researchers, educators, and magicians were communicating to more than 100 attendants their roles and what they did in their labs to contribute to the project "From the Magic of Science to the Science of Magic". This Bachelor Thesis showcases the buildup of a scientific and creative citizenship by means of open and online Science, scientific volunteering, and flexible associationism using the Internet (Science 2.0) and Magic. In #ECME17 (*Figure 22*) I was recording videos and taking pictures of speakers, listening at them, posting some Twitter posts in @magsci profile and learning lots of new communications techniques.

We created a folder[1] in our cloud storage to share some useful files that were used and recorded in this congress and other activities that have already been explained before or are going to be explain later. It was a huge opportunity to establish a first contact with relevant people in the world of communication, magic and Science, such as: Artur Antúnez (@arturantunez), Jorge Luengo (@jorge_luengo), Paul Nugent and Caroline Ainslie.

[1] Torrecillas, A. [Treball Final de Grau]. (2018). Characterization of non-coding genetic variants associated with Brugada syndrome: Research and communication: Support Material. [Google Drive folder]. Retrieved from https://drive.google.com/open?id=1-KSQn_CsWgESIGY4NHzxOUB-Q0lBI5Wl

4.2.2. Workshop: Why do Superheroes have powers?

Most people have a remote idea about what DNA is, they think is “something” that we have inside our cells. The concept of mutations is a little bit more difficult for them. What is really sure is that Science is culture and there are some festivals who give this idea. YoMo (Hospitalet de Llobregat, Barcelona) is a nice example to discover this. For four days during this February 2018, more than 15,000 school children and teenagers from across Catalonia attended the showcase of Science and technology in Hospitalet de Llobregat. The workshop (*Figure 23*) was part of the low-cost communication project in C4D. This means that scientists have to be able to communicate difficult concepts using as less objects and funds as possible. The objective in YoMo was to attract teachers and schoolers to our workshop to explain what BrS was.

Divulging biomedical science, at the level of molecular biology, may be seem a difficult task. The workshop "***Why superheroes have powers?***" aims to facilitate the understanding of concepts of molecular biology and processes linked to DNA, RNA, amino acids and proteins that are key in the processes of life.

Specifically, it aims to understand the medical biochemistry research that is being carried out at the IDIBGI lab (GenCardio Research group), working in proteins associated with the *SCN5A* gene. This Molecular Biology workshop could be incorporated at class because some teachers ask about that. For this situation, we can be so proud of our new workshop that we presented in YoMo.

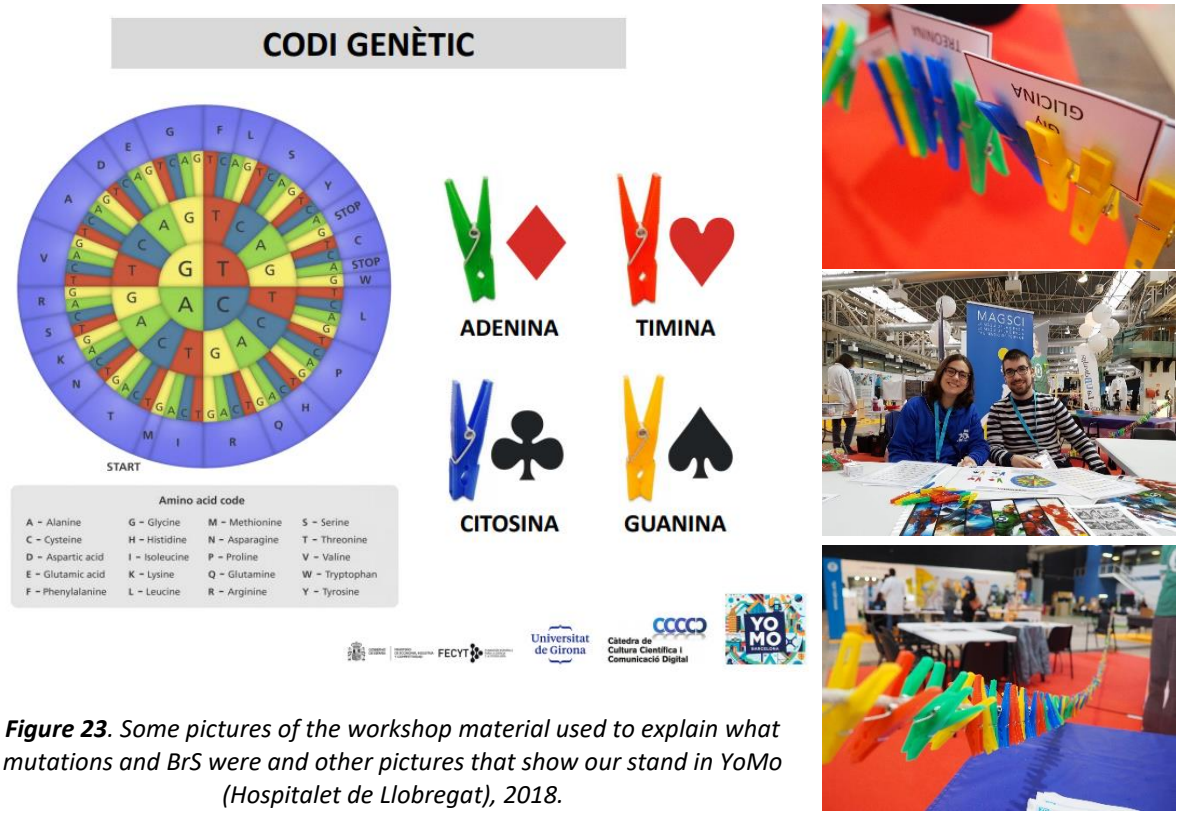


Figure 23*. Some pictures of the workshop material used to explain what mutations and BrS were and other pictures that show our stand in YoMo (Hospitalet de Llobregat), 2018.*

The workshop consists of several parts:

- Explanation of the basic concepts of genetics, DNA, RNA and proteins/enzymes.
- Explain the consequences of some mutations in the human genome: which can be positive (give superpowers) or negative (cause illnesses, such as heart disease).
- Delivery of the comic "*How I became the amazing Spiderman*".
- Ask for answers to simple test-type questions.
- Contribution to the construction of a protein (coded by the *SCN5A*) chain gene of triplets with different colored clothes peg as nitrogenous bases).

First of all, we had to explain what Molecular Biology is and why it is related to Medical Science. There were two panels with some superheroes in our stand to attract people to come in and take a seat. Then, with the question: "*Do you want to know why superheroes have powers?*" teenagers and school children from 12 to 16 years old were coming to know the answer. Lots of teachers came there too to ask us some questions about the workshop. The explanation took about 10 to 15 minutes per each group of people (*Figure 24*) who attended the stand. It was the right time to explain carefully our topic. Our reflexion could be what one may wonder about would it happen if instead of using superheroes we considered a DNA molecule. We used four-different-coloured clothes peg to represent each nucleotide. Adenine (A) was represented in green colour, thymine (T) was red, guanine (G) was yellow and cytosine (C), blue. The four different nitrogenous bases are often represented by the first letter: A, T, G and C. Then, we explained how can you find this letters forming special molecules called the proteins.

***Figure 24**. Explaining our project to some High school students in YoMo (Barcelona).*

If one of these clothes peg was not in its correct place in our genome, a mutation would occur. One must note that in other approach at C4D cards were used to represent the same concept. As there are black and red cards it is easy to connect two suits per each colour and to distinct the pairings according to their number of hydrogen bonds. In that case, Magic is useful to explain the DNA formation and diamonds, clubs, hearts and spades are playing another role.

"*Why superheroes have powers?*" workshop allowed us to value the proposal very positively. We value very highly the use of cheap material to create Science and be able to communicate Science. My role in this workshop was to design an idea knowledge, to discover a the BrS, to share information with society and to appear in TV, among others.

4.2.3. Assess on Science communication

With live theatre shows, interactive workshops, and dozens of hands-on activities, YoMo (The Youth Mobile Festival) is designed to bring classroom learning to life by showing the exciting and rewarding career opportunities. What I value the most from the experiences, festivals and congresses with C4D in the field of scientific communication has been all I could learn that has not been taught in my University classes.

Innovation ways to explain a funnier Chemistry and Mathematics topics are easy to find on the Internet, but they are less common in other fields like Molecular Biology. Tools in Molecular Biology field are not popular and there is no information on how to correctly communicate these field concepts while adapting the language to the public. For this reason, we tried to design the previous "*Why superheroes have powers?*" workshop.

4.3. ETHICS & SUSTAINABILITY STATEMENT

The main problem that scientists are facing from the results obtained in genome sequencing studies is that they still do not know how to interpret all the information contained in the genome, especially when analyzing DNA variants in non-coding regions.

When studying inherited diseases such as BrS, the identification of potentially pathogenic variants in a patient raises the question of whether the relatives should also be genetically tested. In these circumstances, genetic counselors usually recommend the closest relatives to get analyzed and find out whether they also have this pathogenic mutation. Sharing genomic data can provide us with enormous tools to understand inherited diseases. However, scientists should be extremely cautious in terms of genomic data privacy protection, avoiding to make public potential damaging information.

This study, that uses genetic data from BrS patients, was approved by the ethical committee of the Hospital Josep Trueta, Girona and conforms the principles of the 1975 Declaration of Helsinki.

5. CONCLUSIONS

Attending to the objectives planned at the beginning of this Bachelor Thesis (in section 2), the following conclusions can be drawn:

1. The luciferase assays are a valid method to analyze the CTCF binding level. The results obtained show that 13 of the 34 alternative variants evaluated with this reporter system present a statistically significant difference in CTCF binding versus its reference variant.

2. The effect of variants observed in luciferase assays is, in general, similar than the effect predicted using DeepBind algorithm, although the degree of correlation between DeepBind and luciferase is not robust.

3. In both methods (Luciferase and DeepBind), the binding differences between the reference and alternative sequences are greater when the variant is located within the CTCF motif. However, this difference in binding between variants inside the motif (core nucleotides and non-core nucleotides in the motif) and outside the motif (far or close) is only significant for DeepBind predictions.

4. This study validates the effectiveness of the DeepBind software to use it as a reliable predictor of the effects of the variants on TF binding. Although in this project DeepBind is used to predict the effects of noncoding variants identified in BrS patients, we expect that in the future it could be applied to any type of disease.

5. The use of low-cost objects was effective to communicate Science, especially in certain social and age sectors. The design of a new workshop using low-cost communication was a great experience. Working with new methodologies allowed interesting contributions in teenagers as innovative proposals for the 21st century school.

6. The established strategy to communicate the concepts of mutations and BrS in YoMo was successful and people who attend to our stands left with an idea about the relevance of this project. The open scientific attitude contributes to the improvement of the training process and to the idea that Science is another field of culture.

REFERENCES

Ackerman, M., Marcou, C., & Tester, D. (2013). Medicina personalizada: Diagnóstico genético de cardiopatías/canalopatías hereditarias. *Revista Española de Cardiología, 66*(4), 298-307. doi: 10.1016/j.recesp.2012.12.010

Alipanahi, B., Delong, A., Weirauch, M., & Frey, B. (2015). Predicting the sequence specificities of DNA-binding and RNA-binding proteins by deep learning. *Nature Biotechnology, 33*(8), 831-839. doi: 10.1038/nbt.3300

Allen, D., *et al*. (2011). *ECG Interpretation* (5th ed.). Amber: Lippincott Williams & Wilkins.

Antzelevitch, C., & Eyal, N. (2008). Brugada syndrome: Recent advances and controversies. *Current Cardiology Reports, 10*(5), 376-383. doi: 10.1007/s11886-008-0060-y

Antzelevitch, C., & Fish, J. (2006). Therapy for the Brugada syndrome. *Handbook of Experimental Pharmacology, 171*, 305-330.

Antzelevitch, C., Brugada, P., Brugada, J., Brugada, R., Nademanee, K., & Towbin, J. (1998). *Clinical approaches to tachyarrhythmias: The Brugada syndrome* (1st ed.). New York: Futura Publishing Company, Inc.

Applied Biosystems. (2011). *BigDye™ Direct Cycle Sequencing Kit*. Waltham, MA: Thermo Fisher.

Balser, J. (2001). The cardiac sodium channel: gating function and molecular pharmacology. *Journal of Molecular and Cellular Cardiology, 33*, 599-613. doi: 10.1006/jmcc.2000.1346

Baumgartner, H., Bonhoeffer, P., De Groot, N., de Haan, F., Deanfield, J., Galie, N., *et al.* (2010). Guidelines for the management of grown-up congenital heart disease. *European Heart Journal, 31*(23), 2915-2957. doi: 10.1093/eurheartj/ehq249

Berne, P., & Brugada, J. (2012). Brugada Syndrome. *Circulation Journal, 76*(7), 1563-1571. doi: 10.1253/circj.cj-12-0717

Bezzina, C., Barc, J., Mizusawa, Y., *et al*. (2013). Common variants at SCN5A, SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nature Genetics, 45*(9), 1044-1049. doi: 10.1038/ng.2712

Binder, A., Hillback, E., & Brossard, D. (2016). Conflict or caveats? How media portrayals of scientific uncertainty influence perceptions of new technologies. *Risk Analysis, 36*(4), 831-846. doi: 10.1111/risa.12462

Brugada, P., & Brugada, J. (1992). Right bundle branch block, persistent ST segment elevation and sudden cardiac death: A distinct clinical and electrocardiographic syndrome: A multicenter report. *Journal of the American College of Cardiology, 20*(6), 1391-1396. doi: 10.1016/0735-1097(92)90253-J

Brossard, D., Lewenstein, B., & Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen Science project. *International Journal of Science Education, 27*(9), 1099-1121. doi: 10.1080/09500690500069483

Campuzano, O., Sarquella-Brugada, G., Brugada, R., & Brugada, J. (2015). Review article: Genetics of channelopathies associated with Sudden Cardiac Death. *Global Cardiology Science and Practice, 39.* doi: 10.5339/gcsp.2015.39

Chen, P., Joung, B., Shinohara, T., Das, M., Chen, Z., & Lin, S. (2010). The initiation of the heartbeat. *Circulation Journal, 74*(2), 221-225. doi: 10.1253/circj.cj-09-0712

Chen, Y., Chang, Y., Wu, C., Ko, Y., Tsao, S., Chen, J., & Lin, S. (2017). Epstein-Barr virus Rta-mediated accumulation of DNA methylation interferes with CTCF binding in both host and viral genomes. *Journal of Virology, 91*, 12-15. doi: 10.1128/jvi.00736-17

Christoffels, V., Smits, G., & Moorman, A. (2010). Development of the pacemaker tissues of the heart. *Circulation Research, 106*(2), 240-254. doi: 10.1161/circresaha.1 09.205419

Colledge, N., Walker, B., & Ralston, S. (2010). *Davidson's principles and practice of medicine* (21st ed.). Edinburgh: Churchill Livingstone.

Cowan, J., Morales, A., Dagua, J., & Hershberger, R. (2008). Genetic testing and genetic counseling in cardiovascular genetic medicine: Overview and preliminary recommendations. *Congestive Heart Failure, 14*(2), 97-105. doi: 10.1111/j.1751

DeepBind. (2014). *University of Toronto*. Retrieved from http://tools.genes.toronto.edu/deepbind

Fernández, A., Sarquella-Brugada, G., Brugada, J., & Brugada, R. (2017). Cardiac Channelopathies and Sudden Death: Recent genetic advances. *Biology, 6*(1), 7. doi: 10.3390/biology6010007

Gauchat, G. (2011). The cultural authority of science: Public trust and acceptance of organized science. *Public Understanding of Science, 20*(6), 751-770. doi: 10.1177/0963662510365246

Gilder, J., Ford, S., Doom, T., & Krane, D. (2004). Systematic differences in electropherogram peak heights reported by different versions of the GeneScan® Software. *Journal of Forensic Sciences, 49*(1), 1-4.

Hatcher, C., & Basson, C. (2009). Specification of the cardiac conduction system by transcription factors. *Circulation Research, 105*(7), 620-630. doi: 10.1161/circresaha.109.204123

Invitrogen. (2013). *Lipofectamine® 2000 DNA Transfection Reagent Protocol*. Carlsbad, CA: Life Technologies.

Kapplinger, J., Tester, D., Alders, M., Benito, B., Berthet, M., & Brugada, J. (2010). An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm Journal, 7*(1), 33–46. doi: 10.1016/j.hrthm.2009.09.069

Katritsis, D., Gersh, B., & Camm, J. (2016). A clinical perspective on sudden cardiac death. *Arrhythmia & Electrophysiology Review, 5*(3), 177-182. doi: 10.15420/aer.2016.11.2

Konstandina, K., Rozana, K., Beti, G., & Antonio, G. (2016). Brugada syndrome: A case report. *Sanamed, 11*(1), 53-56. doi: 10.5937/sanamed1601053k

Lim, S., Vos, T., Flaxman, A., Danaei, G., Shibuya, K., Adair-Rohani, H., *et al*. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2224-2260. doi: 10.1016/S0140-6736(12)61766

Mademont-Soler, I., Pinsach, M., Riuró, H., *et al*. (2016). Large Genomic Imbalances in Brugada Syndrome. *PLoS ONE, 11*(9). doi: 10.1371/journal.pone.0163514

McNutt, M. (2017). Editorial: Improving Scientific Communication. *Science, 342*(6154), 13. doi: 10.1126/science.1246449

Min, S., Lee, B., & Yoon, S. (2017). Deep Learning in Bioinformatics. *Briefings in Bioinformatics Journal, 18*(5), 851-869. doi: 10.1093/bib/bbw068

Morita, H., Kusano, K., Nagase, S., et al. (2002). Atrial Fibrillation and atrial vulnerability in patients with Brugada Syndrome. *Journal of the American College of Cardiology, 40*, 1437-1444. doi: 10.1016/j.jacc.2007.10.060

National Academy of Sciences. (2017). *Communicating Science Effectively: A Research Agenda*. Washington, DC: The National Academies Press.

National Academy of Sciences. (2015). *The Science of Science Communication*. Washington, DC: The National Academies Press.

Ngiam, J., Coates, A., Lahiri, A., *et al*. (2011). On optimization methods for deep learning. *Proceedings of the 28th International Conference on Machine Learning*, 265-72.

Ong, C., & Corces, V. (2014). Modes of transcriptional regulation. CTCF: An architectural protein bridging genome topology and function. *Nature Reviews Genetics, 15*(4), 234-246. doi: 10.1038/nrg3663

Peng-Sheng, C., & Priori, S. (2008). The Brugada Syndrome. *Journal of the American College of Cardiology, 51*(12), 1176-1180. doi: 10.1016/j.jacc.2007.12.006

Peñaranda, A., Cantalapiedra, I., Bragard, J., & Echebarria, B. (2012). Cardiac dynamics: A simplified model for action potential propagation. *Theoretical Biology & Medical Modelling, 9*, 50. doi: 10.1186/1742-4682-9-50

Poulos, R., Thoms, J., & Guan, Y. (2016). Functional mutations form at CTCF-cohesin binding sites in melanoma due to uneven nucleotide excision repair across the motif. Cell Reports Journal, 17(11), 2865-2872. doi: 10.1016/j.celrep.2016.11.055

Promega Corporation. (2007). *Normalizing genetic reporter assays: Approaches and considerations for increasing consistency and statistical significance*. Madison, WI: Promega Corporation.

Promega Corporation. (2008). *Instructions for use of* pRL renilla luciferase *reporter vectors: Technical manual*. Madison, WI: Promega Corporation.

Promega Corporation. (2015). *Instructions for use of Dual-Luciferase® reporter assay: Technical manual*. Madison, WI: Promega Corporation.

Promega Corporation. (2015). *Instructions for use of pGL4 luciferase reporter vectors: Technical manual*. Madison, WI: Promega Corporation.

Selga, E., Campuzano, O., Pinsach, M., *et al*. (2015). Comprehensive genetic characterization of a Spanish Brugada syndrome cohort. *PLoS ONE, 10*(7). doi: 10.1371/journal.pone.0132888

Somi, K., Nam-Kyung, Y., & Bong-Kiun, K. (2015). CTCF as a multifunctional protein in genome regulation and gene expression. Gene regulation: A master controller of the genome. *Experimental & Molecular Medicine, 47*, 166. doi: 10.1038/emm.2015.33

Tarradas, A., Pinsach, M., Mackintosh, C., Llorà, O., *et al*. (2017). Transcriptional regulation of the sodium channel gene (SCN5A) by GATA4 in human heart. *Journal of Molecular and Cellular Cardiology, 102*, 74-82. doi: 10.1016/j.yjmcc.2016.10.013

Thermo Fisher (2017). *MAX Efficiency™ DH5α® competent cells*. Waltham, MA: Thermo Fisher.

Vieta, J. (2016). *Nous formats en la comunicació de la química. Explorant el seu paper en el foment de la cultura i les vocacions científiques* (Tesi doctoral). Universitat de Girona.

Wheeler, R. (2016). Certificate of analysis: pGL4.23[luc2/minP] vector. U.S. *Patent No. 5,670,356*. Fitchburg, WI: U.S. Patent and Trademark Office.

Zipes, D., & Wellens, H. (1998). Sudden Cardiac Death. *American Heart Association, 98*, 2334-2351. doi: 10.1161/01.cir.98.21.2334