

Títol del treball:

Identificació de nous gens implicats en predisposició germinal al càncer colorectal

Estudiant: Anna Gratacós Mulleras

Grau en Biologia


Correu electrònic: anna.gratacosm@gmail.com

Tutor: Antoni Benito Mundet

Cotutor: Antoni Castells

Empresa / institució: Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)

Vistiplau tutor i cotutor:



Nom del tutor: Antoni Benito Mundet
Nom del cotutor: Antoni Castells
Empresa / institució: IDIBAPS
Correus electrònics:
antoni.benito@udg.edu
castells@clinic.ub.es

Data de dipòsit de la memòria a secretaria de coordinació: 22/07/2016

Agraïments

M'agradaria donar les gràcies a totes les persones que m'han ajudat a fer possible aquest treball. Primer de tot als membres del grup d'Oncologia Gastrointestinal i Pancreàtica de l'Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), especialment al doctor Antoni Castells i al doctor Sergi Castellví-Bel per acollir-me en el seu grup, ajudar-me i orientar-me durant la meva estada i en l'elaboració del treball. Gràcies també al doctor Antoni Benito per la seva ajuda en l'orientació del treball. També vull agrair a la ja doctora Clara Esteban-Jurado per la seva ajuda i orientació en aquest treball, i a la resta de membres del grup pel bon ambient de treball i els bons moments viscuts. Per altra banda, també agrair als meus amics i companys d'universitat pel seu suport al llarg de tota l'elaboració del treball. A la meva família i parella, perquè sense el seu suport res d'això hagués set possible, i especialment a la meva tieta Maria Àngels Ginés per facilitar-me el contacte a l'IDIBAPS i a la meva germana, Mercè Gratacós Mulleras per la seva ajuda i interès constant, i per les nostres estones de descans.

Moltes gràcies a tothom

Resum

El càncer colorectal (CCR) és la neoplàsia amb major incidència a Espanya tenint en compte els dos gèneres. Es distingeix entre l'espòrdic que correspon a un 65% dels casos i el familiar amb un 35%. En el CCR familiar, la majoria de casos la causa roman desconeguda, però un 5% és causat per mutacions germinals d'alta penetrància que donen lloc a les síndromes mendelianes hereditàries com la síndrome de Lynch, la poliposi adenomatosa familiar i les poliposi associades al gens *MUTHY* i *POLE/POLD1*. Actualment a partir de les tècniques de seqüenciació massiva, s'han identificat nous gens associats a la predisposició germinal al CCR com *NTHL1* o *FAN1*.

El principal objectiu d'aquest estudi és la determinació de nous gens implicats en la predisposició germinal al CCR a partir de l'anàlisi de famílies que presentin una forta agregació per aquesta malaltia i la selecció de variants compartides en els individus de cada família. S'han utilitzat mostres de sang perifèrica dels pacients amb CCR de 16 famílies no emparentades i seqüenciat l'exoma pel mètode Illumina. Posteriorment s'ha realitzat un anàlisi bioinformàtic per reduir el nombre de variants trobades i seleccionar aquelles amb una elevada probabilitat de causar un canvi en la funció del gen. S'han seleccionat els gens candidats per a cada família tenint en compte diversos paràmetres funcionals i seguidament s'ha realitzat una validació de les variants (Integrative Genomics Viewer i seqüenciació Sanger).

S'han seleccionat un total de 27 variants i 24 han estat validades per Sanger. Tanmateix s'han realitzat estudis de segregació en familiars addicionals que presentaven adenomes avançats o altres neoplàsies i estudis somàtics a partir d'ADN tumoral per determinar la pèrdua d'heterozigositat (LOH). Els resultats més destacables s'han trobat en 9 gens: *LOXLA*, *GEMIN5*, *UTP11L*, *MUC6*, *HIST1H4*, *GSK3A*, *SPLIT2*, *CIDEB* i *MCM3*, entre els quals segons la seva funció o vinculació anterior amb el càncer s'han considerat més rellevants *LOXLA*, *HIST1H4*, *GSK3A* i *MCM3*.

S'han identificat noves possibles variants associades a la predisposició germinal al càncer colorectal en gens que estan relacionats amb el cicle cel·lular, l'apoptosi i el dany de l'ADN. Tanmateix *LOXLA* vinculat a càncer de coll i cap, s'ha associat també a CCR. De la mateixa manera, les famílies gèniques MCM i SLIT podrien participar en la predisposició genètica al CCR.

Resumen

El cáncer colorrectal (CCR) es la neoplasia con mayor incidencia en España considerando los dos géneros. Se distingue en esporádico que corresponde a un 65% de los casos y el familiar con un 35%. En el CCR familiar, la causa permanece desconocida en la mayoría de los casos, pero un 5% es causado por mutaciones germinales de alta penetrancia que dan lugar a los síndromes mendelianos hereditarios como el síndrome de Lynch, la poliposis adenomatosa familiar y las poliposis asociadas al gen *MUTYH* i *POLE/POLD1*. Actualmente a partir de las técnicas de secuenciación masiva, se han identificado nuevos genes asociados a la predisposición germinal al CCR como *NTHL1* o *FANL*.

El principal objetivo de este estudio es la determinación de nuevos genes implicados en la predisposición germinal al CCR a partir del estudio de familias que presentan fuerte agregación para esta enfermedad y la selección de variantes compartidas en los individuos de cada familia. Se han utilizado muestras de sangre periférica de pacientes con CCR de 16 familias no emparentadas y secuenciado el exoma a partir del método Illumina. Posteriormente se ha realizado un análisis bioinformático para reducir el nombre de variantes encontradas y seleccionar aquellas con una elevada probabilidad de causar un cambio en la función del gen. Se han seleccionado los genes candidatos por cada familia considerando diversos parámetros funcionales y seguidamente se ha realizado una validación de las variantes (Integrative Genomics Viewer y secuenciación Sanger).

Se han seleccionado un total de 27 variantes y 24 han sido validadas por Sanger. Se han realizado estudios de segregación en familiares adicionales que presentaban adenomas avanzados u otras neoplasias y estudios somáticos a partir de ADN tumoral para determinar la pérdida de heterocigosidad (LOH). Los resultados más relevantes se han encontrado en 9 genes: *LOXLA*, *GEMIN5*, *UTP11L*, *MUC6*, *HIST1H4*, *GSK3A*, *SPLIT2*, *CIDEB* i *MCM3*, entre los cuales a partir de su función o vinculación anterior con el cáncer se han considerado más relevantes *LOXLA*, *HIST1H4*, *GSK3A* y *MCM3*.

Se han identificado nuevas posibles variantes asociadas a la predisposición germinal al cáncer colorrectal en genes que están relacionados con el ciclo celular, la apoptosis y el daño en el ADN. También *LOXLA* vinculado a cáncer de cuello y cabeza se ha asociado a CCR. De la misma manera, las familias génicas MCM y SLIT podrían participar en la predisposición genética al CCR.

Abstract

Colorectal cancer (CRC) is the most frequent neoplasia in both sexes in Spain. It may be sporadic (65%) as well as familial (35%). In the majority of cases the cause of familial CCR remains unknown. However, a 5% is due to high penetrance germline mutations which are characterised as hereditary Mendelian syndromes, such as Lynch syndrome, familial adenomatous polyposis and the poliposis associated with *MUTYH* and *POLE/POLD1*. Nowadays using next generation sequencing techniques, new genes associated with germline predisposition to CRC such as *NTHL1* or *FANL1*, have been identified.

The main objective of this study is the identification of new genes related to germline predisposition to CRC. Families with a strong aggregation for the disease and the selection of variants shared by different individuals in each family were studied. Germline DNA samples used for the whole exome sequencing by Illumina method were obtained from peripheral blood of patients with CCR belonging to 16 unrelated families. A bioinformatic analysis was performed for reducing the number of encountered variants and selecting those with a high probability to cause a change in the gene function. The candidate genes in each family have been selected considering several functional parameters. Finally, a validation of the variants was carried out (Integrative Genomics Viewer and Sanger sequencing).

Twenty-seven variants were selected and 24 of them validated by Sanger sequencing. Segregation studies from additional family members who presented advanced adenomas or other neoplasms as well as somatic studies with tumour DNA to determine the loss of heterozygosity (LOH), were also performed. The most remarkable results were found in 9 genes: *LOXL4*, *GEMIN5*, *UTP11L*, *MUC6*, *HIST1H4*, *GSK3A*, *SPLIT2*, *CIDEB* and *MCM3*. According their function or previous relation with cancer *LOXL4*, *HIST1H4*, *GSK3A* and *MCM3* have been considered the most relevant.

We identified new possible germinal predisposition variants associated to CCR in genes that play a role in cell cycle, apoptosis and DNA repair., Furthermore, *LOXL4*, which is associated with head and neck cancer, is also related to CCR. Similarly, *MCM* and *SLIT* gene families could participate in genetic CCR predisposition.

Índex

1. Introducció	
1.1 Definició del càncer	1
1.2 Càncer colorectal (CCR).....	2
1.2.1 Epidemiologia del càncer colorectal	2
1.2.2 Carcinogènesi en la mucosa colònica	3
1.2.3 Etiologia.....	4
1.2.4 Tipus CCR.....	5
1.3 Tècniques moleculars per descobrir gens de predisposició al CCR.....	8
2. Objectives	9
3. Metodologia	
3.1 Pacients i tractament de les mostres	9
3.2 Anàlisi de dades.....	10
3.3 Priorització de les variants	12
3.4 Prevalidació de les variants	13
3.5 Validació de les variants, estudis de segregació i LOH en ADN tumoral.....	13
4. Resultats	15
5. Discussió	
5.1 Discussió dels resultats	20
5.2 Ètica i sostenibilitat	23
6. Conclusions	24
7. Bibliografia	25

1.Introducció

1.1 Definició de càncer

Es considera càncer al conjunt de malalties causades per canvis en gens que controlen la divisió i mort cel·lular, provocant un desequilibri entre aquests processos. Aquest és un procés dinàmic que s'inicia amb una mutació en una cèl·lula que progressivament va adquirint noves mutacions que potencien una proliferació cel·lular descontrolada o inactiven l'apoptosi, ocasionant que aquelles cèl·lules anòmales sobrevisquin i presentin diverses mutacions en el seu genoma. Consegüentment a aquesta desregulació, es formen acumulacions de cèl·lules en un punt concret del cos causant el que s'anomenen tumors o neoplasmes (Nussbaum *et al.*, 2015). El procés es finalitza quan aquests tumors adquireixen capacitat invasiva en el mateix punt d'origen sent capaços de migrar a través dels vasos sanguinis o limfàtics dipositant-se en altres regions del cos, produint el que s'anomena metastasis. Aquests passen a anomenar-se tumors metastàtics.

Tal com s'ha dit, el càncer és una malaltia causada per canvis genètics. La majoria de mutacions que el causen són somàtiques i es produeixen en el teixit afectat durant la carcinogènesi. Tot i això molts tipus de càncer presenten un component hereditari, aquest presenta mutacions a nivell germinal que augmenten el risc o la predisposició a la carcinogènesi (de la Chapelle, 2004).

El càncer pot ser esporàdic, causat per factors ambientals, o hereditari, associat a predisposició germinal. Aquest dos tipus es poden diferenciar a partir de la hipòtesi dels dos hits proposada per Alfred G. Knudson. Aquest va descriure el càncer com dos esdeveniments que comporten la pèrdua de funció primer d'un al·lel d'un gen i posteriorment del segon produint una inactivació de gens supressors tumorals (GST). En el càncer esporàdic els progenitors no pateixen càncer i un dels seus fills en presenta com a conseqüència d'un esdeveniment puntual a nivell somàtic, que posteriorment s'acompanya d'una segona mutació. Per altra banda, en el càncer hereditari el primer hit o esdeveniment s'hereta i el segon s'adquireix posteriorment per una mutació a nivell somàtic. Primer de tot hi ha una mutació a nivell germinal, en què totes les cèl·lules del cos de la persona estan danyades i per aquest motiu només és necessari un canvi perquè es perdi completament la funció del gen (Knudson, 1996).

1.2 Càncer colorectal

1.2.1 Epidemiologia del càncer colorectal

El càncer colorectal (CCR) representa una de les neoplàsies més freqüents a nivell mundial i és una important causa de mortalitat en el món desenvolupat. Presenta una elevada incidència a Amèrica del Nord, Austràlia, Nova Zelanda i Europa i per altra banda una baixa incidència a l'Àfrica i el Sud d'Àsia tal com s'observa a la *Figura 1* (Ferlay *et al.*, 2010; GLOBOCAN 2012).

L'elevada incidència es deu a l'augment de l'esperança de vida en aquests països i la progressiva occidentalització dels estils de vida (Tenesa & Dunlop, 2009). Aquesta occidentalització consisteix en l'adquisició dels costums i comportaments d'occident com són la dieta, i el sedentarisme que contribueix en un augment dels factors de risc del CCR (Center *et al.*, 2009).

Els països desenvolupats tot i presentar una major incidència, tenen un menor índex de mortalitat a causa del screening poblacional que es realitza per detectar el CCR (a partir del test de sang oculta en femta i colonoscòpies que permeten detectar pòlips o adenomes en estadis primerencs i CCR), també per la millora dels tractaments i per les campanyes de conscienciació (Center *et al.*, 2009). Per altra banda, tot i que els països en vies de desenvolupament tenen una menor incidència, presenten un 64% de les morts per càncer a nivell mundial. Aquests percentatges es deuen a l'elevat nombre d'infeccions, a la manca de recursos i tractaments, la dificultat de detecció i a la baixa esperança de vida (Torre *et al.*, 2015).

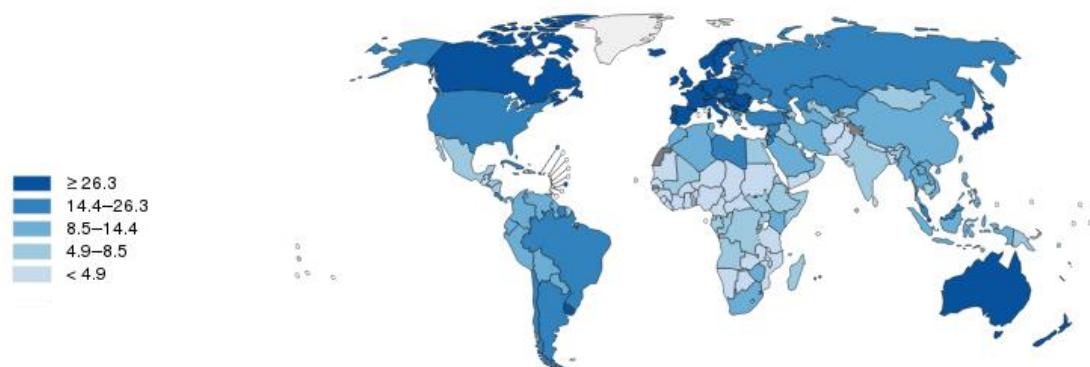


Figura 1. Mapa mundial de la incidència del càncer colorectal per país estandarditzada per l'edat i en els dos sexes. La llegenda mostra el nombre de nous casos diagnosticats per cada 100.000 habitants durant l'any 2012. Extret i modificat de GLOBOCAN 2012, *International Agency for Research on Cancer*.

El CCR és el tercer més comú en homes (746.000 casos, 10% del total de casos de càncer) després del de pròstata i pulmó, i el segon més comú en dones (614.000 casos, 9,2% del total) després del de pit, tal com s'observa a la *Figura 2* (Ferlay *et al.*, 2015). L'any 2012 a Espanya sense diferenciació per

sexes va ser el càncer que presentava una major incidència, en què es van diagnosticar 32.240 nous casos de CCR i 14.700 persones van morir per aquesta malaltia (GLOBOCAN 2012).

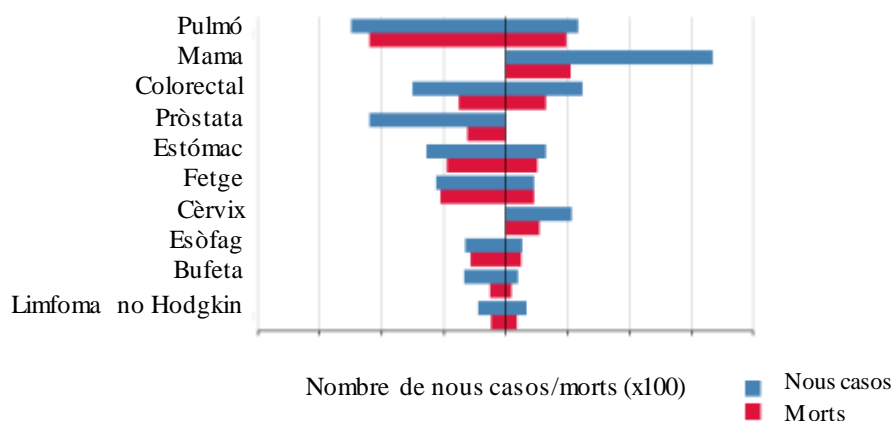


Figura 2. Incidència i mortalitat dels 10 càncers més freqüents a nivell mundial segons el sexe estandarditzada per l'edat. Les dades són per cada 100.000 habitants durant l'any 2012. Extret i modificat de GLOBOCAN 2012, *International Agency for Research on Cancer*.

1.2.2 Carcinògenesi en la mucosa colònica

La majoria de neoplasmes malignes que sorgeixen a l'intestí gros són adenocarcinomes que provenen de cèl·lules epitelials columnars secretores de la mucosa del còlon i recte i produeixen secrecions o moc (Chu, 2011).

El CCR sorgeix a partir de lesions precursoras anomenades pòlips, masses cel·lulars que sobresurten de la paret intestinal que pateixen alteracions anòmales en l'organització intracel·lular i intercel·lular fins a la formació de l'adenocarcinoma. La majoria de lesions formen pòlips adenomatosos o adenomes però fa pocs anys es va determinar l'existència d'una altra classe de pòlips anomenats serrats que constitueixen el 30% dels casos de CCR (Carballal *et al.*, 2013).

La carcinògenesi s'inicia a les criptes colòniques. En condicions normals, aquestes contenen cèl·lules mare que permeten la regeneració de la mucosa necessària pel constant desgast que pateix aquesta zona. Aquestes cèl·lules mare substitueixen les cèl·lules epitelials velles aproximadament cada 5 dies. La proliferació d'aquestes cèl·lules de la cripta dona lloc a cèl·lules normals i també d'aberrants que provenen de la mateixa cèl·lula mare i que són resistents a l'apoptosi. D'aquesta manera es provoquen alteracions en la proliferació i també dinàmiques anormals. Aquesta alteració és la lesió neoplàsica més primerenca que es coneix i forma el que s'anomenen criptes o focus aberrants (Alrawi *et al.*, 2006). Aquests formen acumulacions de cèl·lules epitelials anomenats microadenomes. Seguidament es forma el pòlip, el qual difereix en el seu potencial maligne en funció de la mida i l'estructura d'aquest. Quan aquest pateix una alteració anòmala en la seva organització definida com a displàsia es genera el que

s'anomena pòlip adenomatós o adenoma. Aquest pot anar patint mutacions i altres canvis fins a la formació de l'adenocarcinoma o adenoma maligne (Stewart & Wike, 2006).

La patogènia del CCR és heterogènia i és un procés que compren múltiples passos. La lesió precursora pre-maligna que predomina en aquest tipus de càncer és la formació d'un adenoma displàstic. Aquest procés s'inicia per mutacions a *APC* en un 70% de les vegades. Seguit per mutacions activadores de *KRAS* (oncogen) i inactivadores de GST com *TP53* per acabar formant l'adenocarcinoma (Figura 3). A més, aquestes mutacions estan sovint acompanyades per inestabilitat cromosòmica, canvis en el nombre o l'estructura dels cromosomes (Brenner *et al.*, 2014).

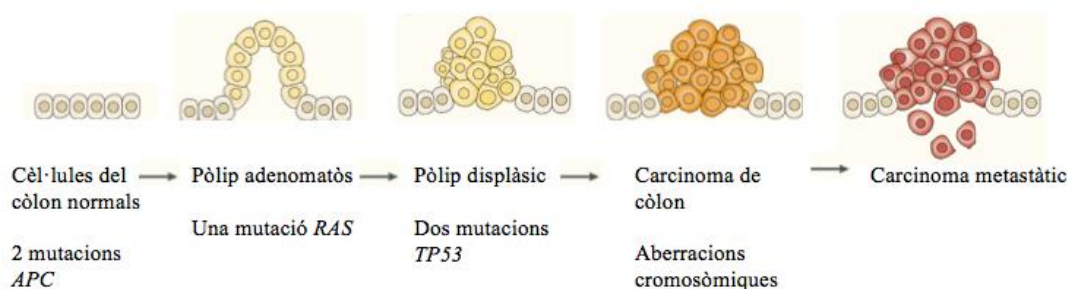


Figura 3. Model de progressió cel·lular de l'epiteli de la mucosa intestinal durant la formació d'un tumor que condueix al càncer colorectal. Esquema de la successió de principals mutacions en la mucosa colònica fins a la formació d'un adenocarcinoma. Extret i modificat de Fearon, E. R. & Vogelstein, B. (1990).

També cal esmentar que els tumors no són idèntics entre pacients, per aquest motiu no hi ha un únic tractament i aquest és difícil i diferent per cada pacient (Aran *et al.*, 2015). Per aquest motiu és important arribar a la medicina personalitzada per tractar cada pacient d'acord amb l'historial i les característiques tumorals que presenta.

1.2.3 Etiologia

El CCR és una malaltia multifactorial que no presenta una causa específica i en la que hi intervé l'edat, factors ambientals i genètics incrementant el risc de patir aquesta malaltia (Aran *et al.*, 2015).

Dins dels factors de risc ambientals es poden distingir dos grups. El primer comprèn els modificables que incrementen el risc de CCR, com el consum d'alcohol, tabac, l'alt consum de carn vermella o processada, el sedentarisme i l'obesitat (Johnson *et al.*, 2013; Baena & Salinas, 2015; Chan *et al.*, 2011). Per altra banda, dins del segon grup estarien els factors ambientals modificables que s'ha verificat que disminueixen el risc de CCR com són l'activitat física, els antiinflamatoris no esteroides, els probiòtics i una dieta rica en fruita i vegetals (Baena & Salinas, 2015; Ouakrim *et al.*, 2015; Wollowski *et al.*,

2001, Tantamango-Bartley *et al.*, 2013). A més, també s'ha vist que el calci, la vitamina D i l'àcid fòlic poder actuar com a agents quimiopreventius pel CCR (Lamprecht & Lipkin, 2003).

També s'ha associat l'edat com a factor que incrementa el risc de CCR, fent que als 80 anys hi hagi un 50% de probabilitats de patir algun càncer. Es creu que el dany cel·lular està relacionat amb l'edat proporcionant mutacions o alteracions en certes cèl·lules i així facilitant la formació del càncer (López-Otín *et al.*, 2013). Tanmateix, s'ha associat un increment del risc de patir CCR en persones que presenten malalties inflamatòries cròniques de còlon com la colitis o la malaltia de Crohn (Aran *et al.*, 2015).

Per últim, els factors genètics explicarien al voltant d'un 30% dels casos de CCR (Short *et al.*, 2015). Es tractaria de mutacions germinals que es transmeten a la descendència i poden provocar agregació familiar, és a dir un augment del nombre de casos CCR en aquella família. La presència d'un familiar de primer grau amb CCR incrementa la probabilitat de que un membre d'aquella família presenti en algun moment de la seva vida CCR, fent que es dobli respecte al risc de la població general sense antecedents familiars. Per altra banda, l'increment del risc en presentar dos familiars amb CCR és al voltant de 3 vegades més (Fuchs *et al.*, 1994).

Aquestes variants genètiques heretables poden ser d'alta o de baixa penetrància. Aproximadament un 5% dels casos són causats per variants d'alta penetrància poc freqüents o rares en la població que estan associades a síndromes de CCR mendelians (Jasperson *et al.*, 2013). Per altra banda, les variants de baixa penetrància són comunes a la població. S'ha comprovat que l'acumulació de variants de baixa penetrància amb factors ambientals tindria un efecte additiu en el risc de patir CCR familiar. Tanmateix, algunes d'aquestes variants podrien també modificar el risc de càncer en pacients que presenten variants d'alta penetrància (Valle, 2014).

1.2.4 Tipus CCR

El càncer colorectal es divideix en esporàdic (65-70% dels casos) i familiar (35% dels casos), on un 5% fa referència a síndromes hereditàries, tal com s'observa a la *Figura 4* (Burt, 2007).

Es poden classificar segons l'agregació familiar. Els casos esporàdics són causats per factors ambientals o per l'edat però no estan associats a una herència i per tant no s'observa agregació familiar. Per altra banda, en els casos familiars, es presenten dos o més casos de CCR en la mateixa família. També es pot considerar del tipus familiar si es tenen parents que pateixen altres càncers però que aquests estan relacionats amb mutacions en gens que predisposen també al CCR com és el cas de *BRCA1* (associat a càncer de mama i colorectal) (Phelan *et al.*, 2013). Del total de casos familiars, avui en dia únicament un 5% es poden explicar com a mutacions en gens d'alta penetrància de predisposició al CCR i són els que s'anomenen hereditaris. Dins d'aquest grup presenten una forta agregació familiar i un patró

d'herència mendeliana. Cal tenir present que els pacients que presentin aquestes mutacions germinals en gens associats al CCR hereditari presenten un augment del risc a patir-lo però no necessàriament l'han de desenvolupar pel fet de ser portadors d'aquella variant encara que tenen una probabilitat alta de fer-ho (Jasperson et al., 2013). Excepte en els casos hereditaris, els casos familiars són aquells on no es coneix encara la causa de l'agregació.

La barrera entre el càncer esporàdic i el familiar no està ben definida pel fet d'haver-hi moltes variants que aporten increments variables de risc de susceptibilitat al càncer colorectal i, tanmateix en els dos tipus hi pot haver involucrats factors ambientals. Altrament, la barrera entre càncer familiar i hereditari tampoc està ben definida, ja que avui es coneixen majoritàriament les variants d'alta penetrància associades als casos hereditaris, però encara falta per identificar variants de baixa i moderada penetrància. Per tots aquests motius esmentats, la barrera entre CCR hereditari, familiar i esporàdic no està estrictament definida i podria existir cert nivell de solapament entre les 3 entitats (Burt, 2007).

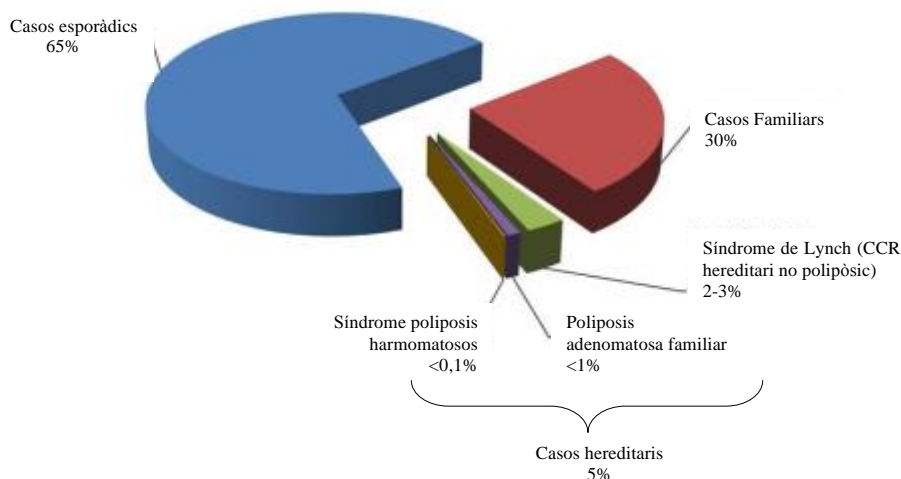


Figura 4. Representació percentatges tipus càncer colorectal en funció del total de casos. Extret i modificat de Randall W. Burt (2000).

Aquesta neoplàsia no presenta en tots els casos una penetrància completa. La majoria de síndromes ja conegudes associades a càncer colorectal hereditari presenten un patró d'herència autosòmic dominant, tanmateix, hi ha síndromes com és el cas de la poliposi associada a *MUTYH* (PAM) que requereixen una mutació bial·lèlica en el gen *MUTYH*, per tal d'expressar el fenotip i iniciar la carcinogènesi i per tant presenten un patró autosòmic recessiu (Buecher *et al.*, 2014).

Fins al moment, s'han associat 13 gens d'alta penetrància a síndromes Mendelians lligats al CCR en edats juvenils o primerenques. Aquests són els gens reparadors de l'ADN (*mismatch repair genes*) que inclouen *MLH1*, *MSH2*, *MSH6* i *PMS2* associats a la síndrome de Lynch, el gen *APC*, relacionat amb la síndrome de poliposi adenomatosa familiar (PAF) i l'atenuada, *MUTYH*, que com s'ha dit està associat

a PAM, *NTHLI* per la formació d'adenomes convencionals, *SMAD4* i *BMPRIA* associats a poliposi juvenils, *LKBI/STK11* per la síndrome Peutz-Jeghers, *GREM1* per pòlips mixtos serrats i tipus juvenils, i *POLD1* i *POLE* descoberts recentment que donen lloc a la síndrome de correcció associat a la polimerasa (Tomlinson, 2015).

La síndrome hereditària més freqüent és la de Lynch i els pacients que presenten aquesta síndrome tenen un 50-80% de risc de desenvolupar CCR al llarg de la seva vida. Es caracteritza per l'absència de pòlips durant la carcinogènesi. El diagnòstic d'aquest síndrome inclou el compliment dels criteris d'Amsterdam. Aquests criteris indiquen que com a mínim hi ha d'haver tres membres amb CCR, un d'ells ha de ser de primer grau dels altres dos, ha d'afectar dues generacions i ha de presentar-se un o més casos abans dels 50 anys (Half & Robert, 2004).

Per altra banda es troben les síndromes polipòsiques adenomatoses. En aquest grup s'hi troba el PAF, que és la segona síndrome més comuna dins del CCR hereditari. Es caracteritza pel desenvolupament de centenars o milers d'adenomes a l'adolescència i presenta una penetrància del 100%. Aquesta síndrome és causada per mutacions germinals en el gen *APC*, que codifica per un supressor tumoral que participa en la via de senyalització Wnt (de la Chapelle, 2004). Aquesta via activa la proliferació responsable de la renovació constant de les cèl·lules del còlon. Una altra síndrome polipòsica adenomatosa és PAM, on existeixen mutacions a *MUTYH*. Aquest gen participa en la via de reparació de l'ADN per escissió de bases (BER) (Buecher *et al.*, 2014). A més, s'han identificat mutacions en homozigosis en el gen *NTHLI* associat a la mateixa via com a altre gen causant de poliposi adenomatosa (Weren *et al.*, 2015).

Per últim, un altre tipus de poliposi és l'hamartomatosa. Aquesta, compren síndromes genètiques rares que es caracteritzen pel desenvolupament en el tracte gastrointestinal d'aquest tipus de pòlips poc freqüents comparats amb els pòlips neoplàsics o hiperplàsics. Aquest tipus de pòlip és més comú en nens. Comprèn tres síndromes, la poliposi juvenil, la poliposi Peutz-Jeghers i les síndromes hamartomatoses associats a *PTEN* (Jelsig *et al.*, 2014).

Per altra banda, pel que fa al CCR en inicis primerencs, s'han associat dues síndromes autosòmics recessius que predisposen al càncer. Aquests són la síndrome de Bloom (mutacions en *BLM*) i l'anèmia de Fanconi (mutacions en *FAN1*) Per aquestes malalties, s'han determinat variants d'aquests gens associades amb CCR (de Voer *et al.*, 2015; Seguí *et al.*, 2015).

A part de les síndromes hereditàries, hi ha un 40-50% de pacients que compleixen els criteris clínics d'Amsterdam per la síndrome de Lynch i no presenten mutacions en els gens associats amb aquesta

síndrome. Aquests es classifiquen com a casos de CCR familiar de tipus X. Un gen que s'ha confirmat que dóna susceptibilitat és *RPS20* (Nieminen *et al.*, 2014).

1.3 Tècniques moleculars per descobrir gens de predisposició al CCR

Actualment hi ha diferents estratègies a nivell molecular que permeten la identificació de nous gens implicats amb el CCR. Inicialment, es realitzaven estudis de desequilibri de lligament per estudiar la segregació en els individus d'una família que presenten aquesta malaltia a partir de marcadors informatius com són els microsatèl·lits o SNPs (*single nucleotide polymorphism*) que es troben repartits per tot el genoma. Se seleccionaven les regions que es transmetien en bloc entre els membres afectats i es calculava la probabilitat que es transmetés juntament. Si aquesta probabilitat era més alta que a l'atzar, es considerava que la regió estava associada a la malaltia. Aquests estudis van permetre la identificació de gens d'alta penetrància responsables de les síndromes hereditàries de CCR (Burt & Neklason, 2005).

Des que es van conèixer les variants d'alta penetrància associades amb aquestes síndromes, s'han seguit estudiant gens de baixa penetrància que poden estar associats a la predisposició al CCR. La identificació d'aquestes variants es realitza a partir d'estudis d'associació del genoma complet (*Genome-wide association studies*, GWAS). Aquests estudis es basen en el genotipatge de milers de SNPs comuns distribuïts per tot el genoma i es fa una comparació entre un grup control de persones sanes i un altre d'individus amb CCR per tal de determinar quines variants es troben significativament associades al CCR (Purcell *et al.*, 2007). Avui en dia es coneixen 62 variants de baixa penetrància identificades a partir d'aquest mètode, de les quals algunes d'elles estan localitzades en els gens *GREM1*, *BMP4*, *CDH1*, *SMAD7*, *BMP2*, els quals formen part de la via de senyalització TGF β . Aquesta via participa en la proliferació, diferenciació i migració cel·lular i a partir d'aquests estudis GWAS es va poder associar per primera vegada a la susceptibilitat genètica al CCR (Tenesa & Dunlop, 2009).

A partir de les variants d'alta o baixa penetrància que s'han associat al CCR, es pot realitzar seqüenciació dirigida per Sanger per tal d'estudiar una regió concreta del genoma d'individus amb CCR. Tanmateix, la seqüenciació massiva de nova generació (*Next generation sequencing*, NGS) permet la seqüenciació de milions de fragments d'ADN genòmic de manera massiva i simultània, i permet la identificació de totes les variants del genoma sense distingir aquelles patogèniques de les no patogèniques. És possible seqüenciar tot el genoma (*Whole genome sequencing*, WGS) o també es pot seqüenciar les regions codificants i les seves regions flanquejants intròniques (*Whole exome sequencing*, WES). La WES té un cost més baix i per aquest motiu és el mètode més utilitzat per l'estudi de malalties mendelianes, ja que se sap que aproximadament un 85% de les mutacions causants d'aquestes malalties es troben a les regions codificants (Botstein & Richs, 2003).

2. Objectives

The main objective of this study is the identification of new genes responsible for familial colorectal cancer predisposition by whole-exome sequencing of selected cases with familial aggregation but no alteration of genes involved in the known colorectal cancer hereditary forms (polyposis forms and Lynch syndrome).

1. Identification of new genes responsible for familial CRC by whole-exome sequencing in patients with familial aggregation and with no alteration of the known hereditary CRC genes.
2. Selection of a reduced number of candidate variants for each family with familial aggregation of CRC using programmed analysis pipeline in R language and manual curation.
3. Validation of the selected variants for each family by Sanger sequencing.
4. Segregation studies of the identified variants in the carriers' family, in blood peripheral samples from additional cancer and also with advanced adenomas or polyps.
5. Identification of the second hit in tumour samples analysing loss of heterozygosis (LOH) by Sanger sequencing.

3. Material i mètodes

3.1 Pacients i tractament de les mostres

En aquest estudi s'ha treballat amb un total de 46 pacients amb càncer colorectal de 16 famílies amb forta agregació familiar. Les dades s'han obtingut de 4 clíniques d'alt risc en CCR; Hospital Clínico San Carlos de Madrid, Hospital Clínic de Barcelona, Hospital Donostia de San Sebastián i Hospital Complejo Hospitalario Universitario de Ourense. Les clíniques d'alt risc per CCR consisteixen en unitats funcionals hospitalàries on s'atenen els pacients amb un risc elevat a partir CCR de forma integral i multidisciplinària implicant a serveis com els de Gastroenterologia, Endoscòpia, Oncologia, Infermeria, Genètica o Psicologia. S'han exclòs les famílies que presentaven mutacions en els gens *APC*, als *missmatch repair genes* (*MLH1*, *MLH2*, *MSH6*, *PMS2*) i també les famílies amb mutacions en homozigosis o heterozigosis en els gens *APC* i *MUTYH*. Les famílies s'han escollit seguint el següent criteri: presència de dos o més familiars que presentin CCR, dos o més generacions afectades i com a mínim tenir un familiar diagnosticat abans dels 60 anys, seguint el pedigrí que s'observa a la *Figura 5*.

Aquest estudi ha estat aprovat pel comitè d'ètica i investigació clínica de cada hospital participant i cada pacient ha consentit en la utilització de la seva mostra per a aquest projecte d'investigació.

Les mostres d'ADN germinal s'han obtingut de limfòcits de sang perifèrica del pacient per la seqüenciació d'exoma. Tanmateix en alguns casos s'ha obtingut mostra del tumor del pacient per estudiar la pèrdua d'heterozigositat (LOH) a nivell somàtic. Per al processament de les mostres de sang s'ha utilitzat el kit QIAamp DNA Blood kit i per les mostres de teixit tumoral el QIAamp Tissue Kit (Qiagen, Redwood City, CA) tal com s'ha descrit a l'estudi de Esteban-Jurado *et al.* (2014).

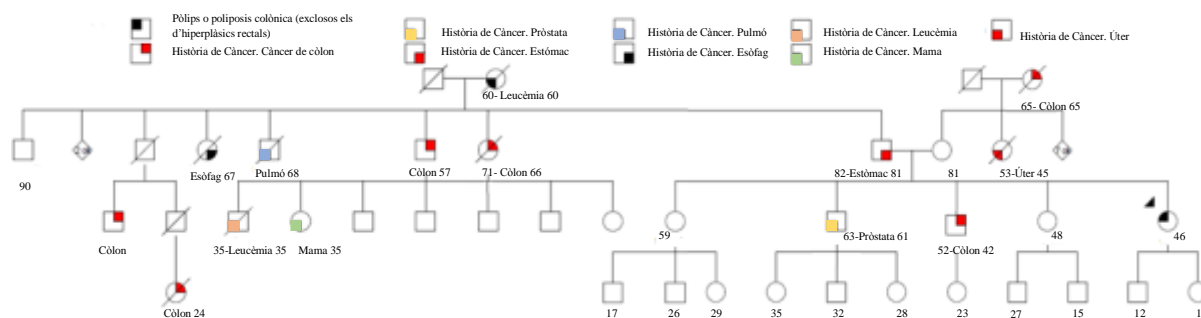


Figura 5: Genealogia família 21, extret d'una de les clíniques d'alt risc que han participat en aquest estudi.

Dins d'aquestes famílies també s'han considerat familiars que presentaven adenomes avançats i altres neoplàsies extracolòniques com a primer signe d'inici de la neoplàsia i d'aquests s'ha obtingut mostres d'ADN a partir de sang perifèrica per realitzar estudis de segregació.

Tal com s'ha dit, l'objectiu de l'estudi és la identificació de nous gens implicats en la predisposició germinal al CCR, per aquest motiu s'han estudiat les regions codificants del genoma germinal de cada pacient per determinar aquells canvis amb un potencial fenotípic més evident. S'ha dut a terme una seqüenciació selectiva de la regió codificant o exoma germinal utilitzant la plataforma HiSeq2000 (Illumina, San Diego, CA) i SureSelectXT Human All Exon V4 per l'enriquiment dels exons (Agilent, Santa Clara, CA). S'ha escollit aquest tipus de seqüenciació mitjançant tècniques de NGS perquè permet el processament d'un gran nombre de mostres simultàniament en una mateixa ronda de forma ràpida i relativament econòmica en comparació amb la seqüenciació convencional Sanger.

3.2 Anàlisi de dades

Una vegada seqüenciades les mostres, s'han mapejat els productes a partir de l'anotació del genoma humà de referència hg19/GRCh37 i s'han anotat les variants seguint l'estudi de Esteban-Jurado *et al.* (2014). Al Centre Nacional d'Anàlisi Genòmica (<http://www.cnag.cat>), es van anotar les conseqüències funcionals de cada variant (*missense*, *frame shift*, regió de *splicing*, mutació sinònima,

codó stop) i la posició (exons, introns, regions no transcrites). S'han considerat interessants aquelles variants que assegurin un canvi important (*high effect impact*) en l'estructura de la proteïna, com és el cas les mutacions de canvi en la pauta de lectura (*frameshift*), un prematur codó stop (*premature stop codon*) o en el lloc de splicing (*canonical splicing site*). En aquests casos l'efecte que tindrà truncant la seqüència de la proteïna o afegint un codó stop abans que s'acabi la seqüència, repercutirà en el funcionament d'aquesta. També s'han seleccionat les mutacions puntuals de canvi d'un aminoàcid per un altre (*missense*). Pel que fa a aquests últims tipus de canvi, s'han utilitzat sis eines bioinformàtiques per la predicció de la patogenicitat que permeten seleccionar aquells canvis d'aminoàcid que són més probables d'afectar el funcionament de la proteïna. Aquests són PhyloP, SIFT, PolyPhen, MutationTaster, CADD i LRT (Esteban-Jurado *et al.*, 2014).

A partir de la *pipeline* automàtica en llenguatge R, s'ha realitzat un primer filtratge en el qual s'han seleccionat les variants en heterozigosis ja que s'ha buscat un patró d'herència dominant. Així mateix s'han seleccionat únicament aquelles variants compartides en els membres de la mateixa família i que presentessin una cobertura de >10x, aquelles que es predeïa un canvi important en la funció del gen (variants que alteren la pauta de lectura, de codó d'aturada, regions canòniques d'*splicing* i *missenses*) i que presentaven <25% del total de pacients seqüenciats. Per altra banda es van eliminar les que presentaven una freqüència al·lèlica $\geq 0,5\%$ en les bases de dades del Projecte 1000 Genomes, de l'Exome Variant Server, Geuvadis European Exome Variants Server, del Centre Nacional d'Anàlisi Genòmica i de l'ExAC, ja que s'han estudiat variants rares, és a dir que no són comunes en la població i que per tant no es consideren polimorfismes genètics.

Referent a les mutacions *missense*, s'han posat filtres per tal que presentessin ≥ 3 prediccions patogèniques de les eines bioinformàtiques. També s'han afegit informació bibliogràfica sobre les funcions dels gens on es localitzaven les variants extretes del NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene>) i del Gene Ontology (GO) (<http://geneontology.org>) i també s'ha filtrat segons si podia relacionar-se amb càncer per la seva funció biològica o bé per la bibliografia existent.

Les dades obtingudes s'han comparat amb la seqüenciació d'exoma d'una línia germinal externa de pacients que presenten leucèmia limfocítica crònica obtinguda del *International Cancer Genome Consortium* (<http://icgc.org>). Les variants trobades en ambdues neoplàsies no s'han considerat implicades en la neoplàsia d'estudi i s'han seleccionat només aquelles exclusivament presents en la cohort CCR.

Finalment s'han verificat totes les anotacions de les variants (posició, domini, canvis d'aminoàcids...), s'han anotat les interaccions de cada gen a partir de NCBI Gene i la conservació de l'aminoàcid del genoma de referència en 100 vertebrats mitjançant la base de dades del UCSC (<https://genome.ucsc.edu>).

3.3 Priorització de les variants

Una vegada s'ha obtingut les dades amb tota la informació necessària, s'han seleccionat aquelles variants que eren més probables d'estar relacionades amb la predisposició germinal al CCR. En aquest punt s'han seguit uns criteris més restrictius (*Figura 6*) seleccionant aquelles variants que presentessin una freqüència al·lèlica en la població d'entre 0-0,001 i que les *missenses* presentessin més prediccions deletèries o patogèniques per eines bioinformàtiques. També que el gen estigués altament conservat i que la variant no es trobés en el genoma d'una altra espècie, a més, que el canvi es trobés en un domini de la proteïna i que aquest tingués una funció indispensable per l'activitat d'aquesta i que presentés interaccions amb gens que s'han relacionat anteriorment amb càncer o gens que participen en vies associades a càncer com són el cicle cel·lular o l'apoptosi. Per últim s'ha descartat aquells gens anotats a l'OMIM que s'han relacionat amb altres malalties i aquells associats amb càncer o amb una funció o terme GO no relacionat amb el procés de carcinogènesi.

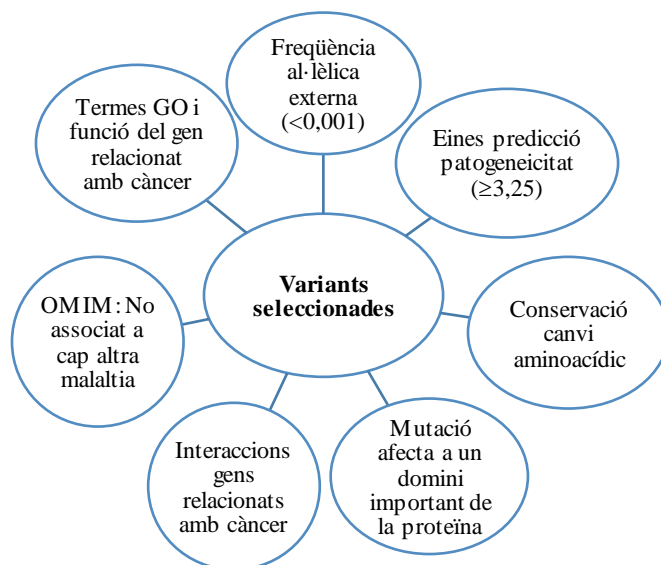


Figura 6. Esquema dels criteris emprats per a la prioritització de les variants candidates a la predisposició a CCR.

La dificultat d'aquesta selecció de variants és l'existència de lindars en alguns paràmetres com són els filtres de freqüència, les eines bioinformàtiques, però la inexistència d'un llindar en la resta de filtres. Aquests filtres més subjectius han proporcionat informació addicional que ha permès seleccionar

aquelles variants més interessants tenint en compte tota la informació relacionada amb els gens on estan localitzades.

3.4 Prevalidació de les variants

S'ha dut a terme un estudi de prevalidació de les variants a partir del programa IGV (Integrative Genomics Viewer), una eina per l'exploració visual de diverses dades genòmiques (Thorvaldsdóttir *et al.*, 2013). Aquest programa permet filtrar per reduir els falsos positius generats durant la NGS (Guo *et al.*, 2012).

S'ha comprovat si cada variant seleccionada es trobava en heterozigosis en l'individu seqüenciat (presenta aproximadament la meitat dels *reads* idèntics a la seqüència de referència i l'altra meitat presenten el canvi de la variant). També s'ha comprovat si la variant es trobava en els extrems del gen, ja que presentarà una cobertura més baixa i per últim s'ha comprovat el *strand bias*, és a dir el biaix per direcció. Aquest últim paràmetre es refereix a la distribució desigual dels *reads* de seqüenciació alineats de les direccions *forward* i *reverse* en una posició concreta (Park *et al.*, 2014).

Per tots aquests motius s'ha prevalidat les variants per comprovar que segueixin aquests criteris que s'han passat per alt durant la selecció.

3.5 Validació de les variants, estudis de segregació i de pèrdua d'heterozigositat en mostres tumorals

Les variants seleccionades s'han validat per reacció en cadena de la polimerasa (PCR) utilitzant primers específics (*Taula 1*) dissenyats pel programa Primer3 (<http://primer3.ut.ee>). Primer de tot s'han optimitzat les condicions de la PCR amplificant a diferents temperatures: 56, 57, 58, 59 i 60 °C en 35 cicles (temps de desnaturalització, unió i extensió dels primers 10'', 10'' i 15'') i a 40 cicles (temps de desnaturalització, unió i extensió de les primeres 15'', 15'' i 30'') en d'altres a partir de mostres control de pacients sans utilitzant el termociclador model Veriti de Life Technologies. Una vegada optimitzats, s'han processat les mostres d'ADN germinal dels pacients amb CCR de cada família amb els seus primers específics i amb un control negatiu per cada parell de primers seguint les proporcions de la mix especificades a la *Taula 2*. Posteriorment s'ha realitzat una electroforesi en gel d'agarosa D1 low eeo al 3% a 100 volts durant 40 minuts per verificar la correcta amplificació de les bandes entre 100 i 200 pb carregant 3 µl del marcador de pes molecular i 6 µl de la mostra amplificada o la mostra control. Els resultats s'ha visualitzat mitjançant ImageQuant LAS 4000 (GE Healthcare Life Sciences) i s'han seqüenciat per Sanger (GATC Biotech, Cologne, Germany) juntament amb una mostra control. També s'ha realitzat un estudi de segregació a partir de mostres d'ADN germinal disponibles de membres addicionals de les famílies seleccionades els quals presentaven adenomes avançats.

Per últim, en les famílies on hi havia mostres d'algun ADN tumoral s'ha estudiat la pèrdua d'heterozigositat a nivell somàtic comparant per seqüenciació pel mètode de Sanger els resultats de l'ADN germinal i el tumoral del mateix individu. Tots els resultats obtinguts per Sanger s'han visualitzat usant el programa Chromas Lite (Versió 2.1.1).

Taula 1. Descripció dels primers per a PCR dissenyats utilitzats per seqüenciació Sanger, estudis de segregació i de LOH.

GEN	Primer forward 5'-3'	Primer reverse 5'-3'	EXIENSIO PCR (pb)
<i>EPHB6</i>	TCTCCGATCACTGACCTCT	CTTGACACCAAGTGGCTAT	188
<i>CENPE</i>	AACCCTTAACTGTGGAGGTG	CCGTAAAGCAT AATATCCGTATCA	187
<i>SIK3</i>	AACAGCAGTCCCTGTGTGAT	GT TTGATGCTGCTGTTGTG	213
<i>RREB1</i>	CTGAGTGGCACAGCCTTG	GTGGTGTACTGCCAGTGCT	191
<i>ARHGAP12</i>	GAATTTTCAGCAAGGTGATTGA	GATGGCAAAAACAAGAAGCTG	208
<i>CIDEB</i>	GCCCCTTCTTGACTTTTCTT	ATAGGAACCGAGGAACCTGG	206
<i>GEMIN5</i>	TCGATTAACCTAGCGTCTGG	CTGATGGAGTGCATGTTCTG	193
<i>RGS3</i>	GAATGAGAGGCTGTGGAG	GCTCACAGCTAGGCAAAAAGA	173
<i>SLIT1</i>	GACTTTCACCAGGTCCAGAG	GGCAAAAAGAGGTCCCATC	162
<i>BCL9L</i>	AGTTTCTCTTTGGGGCTTG	CACTCCACACACACATCC	174
<i>ASCC3</i>	AAGAGGCTCCTTTGTTGACC	AATGGGTTCGTTGAGTGA	157
<i>LOXL4</i>	GACGCTGGCTGTCTTCTTAG	CTCCGTACTCACTGTCCAT	196
<i>NSMCE1</i>	TTGGCTCTTCCACAAACATA	GCTCCTCAGACATCAGTTGG	214
<i>BAP1</i>	CATGTGGCTTCTCTCTGG	ATCCTCCTCGTCATCCTCAT	193
<i>HERC</i>	ACCCAAGTCCCTTTTCTCTG	GGGAAAGAATGGGAAATACC	204
<i>UTP11L (FAM_NOV_3)</i>	CCAGTCTGTGTCATTTGGTG	TCCAGTCTGATTGGTAACTCCT	186
<i>CHD3</i>	GGGAGGTGATCTGTATAGACG	GAAGAGAGGAGCTGCAGAGA	180
<i>BLM</i>	AGCTAGAGGCGATCAATGC	AGGAGGTTTAAAGAGGTCCCTAA	184
<i>MUC6</i>	GAGGAGCCGTAGCTGGAAT	CACCTTGTAAGAAGGGTGCAG	157
<i>MCM3</i>	TGTTTCCTGAGAGCATTTGG	ACCTGATTGTTCATCCTGCAT	226
<i>CLPTMIL</i>	TCTTTGATTTCCCTGGCCTTT	TTCTCAAGAGCCAAGCACAC	221
<i>HISTH4</i>	TGGGTAAAGGAGGAGCTAAG	AATCACGTTCTCCAGGAACA	170
<i>PSMB10</i>	TGGTGTCCAAGTGAAGCAG	ATCTTCTCGCAGCTCTTGTC	170
<i>SLIT2</i>	TACAAAGGGGAAGCAAAGG	CTGCACTCTAATCGGGACAC	207
<i>GSK3A</i>	AGAGAGGTCTGTGGCTGAAG	TCAAGTGAGGAGGGATGAGA	175
<i>UTP11L (FAM23)</i>	TTTCTTCTCCAGCGAATAGC	GGGGTTGATTTACACACAGC	226
<i>MCM4</i>	CAGGACGAAGCCTATGACAG	AGGGTTACCTCGATGAGAGC	185
<i>HIST3H2A</i>	TTAGTCTCGTTTTTCGGTTG	AGGATCTCGGCAGTCAAGTA	238

Taula 2. Volum de cada reactiu de la mix de la PCR per a una mostra.

Producte	Quantitat x1
GoTaq G2 Colorless Master Mix	12,5 µl
2X: ADN polimerasa+ dNTPs+ MgCl ₂ .	
Primer forward	2 µl (20 pmol)
Primer reverse	2 µl (20 pmol)
Nuclease-free H ₂ O	8 µl
ADN	0,5 µl (10-50 ng)
Volum total	25µl

4. Resultats

En aquest estudi s'ha treballat amb un total de 46 pacients amb càncer colorectal de 16 famílies amb forta agregació familiar i amb les dades obtingudes de la seqüenciació de l'exoma complet en aquests pacients. Després dels procediments descrits anteriorment incloent l'anàlisi de les dades de seqüenciació, la prioritització, prevalidació i validació, es va arribar a unes variants finals tal com s'indica en la *Figura 7*.

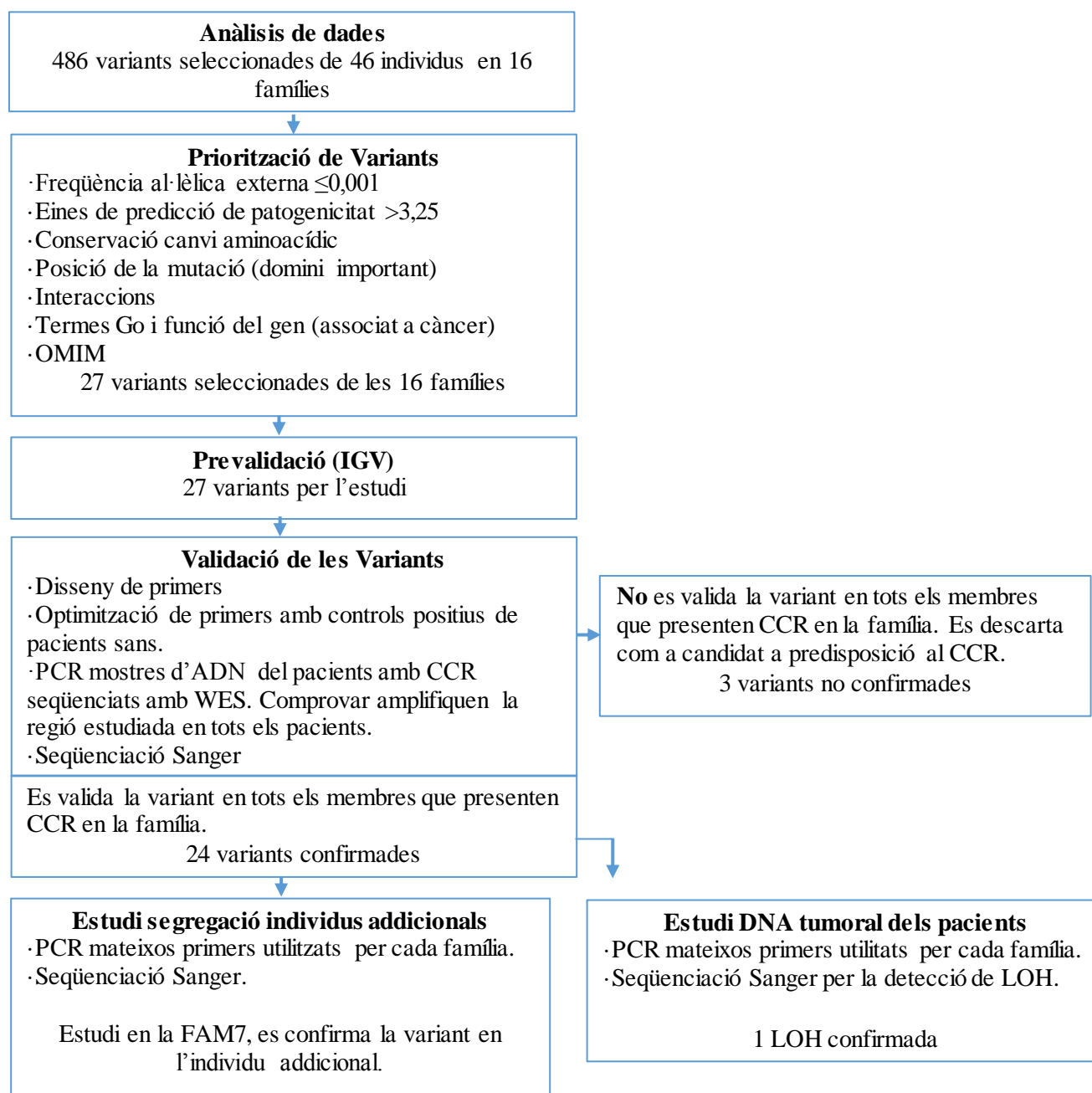


Figura 7. Esquema de la metodologia realitzada en l'estudi amb les variants resultants del tractament de dades, la prioritització, la prevalidació, la validació de les variants seleccionades i dels estudis de segregació i LOH.

En la prioritització de variants s'ha tingut en compte tant aquelles que s'han associat a CCR, com aquelles que no s'han relacionat anteriorment amb aquesta malaltia. Tanmateix, s'han eliminat les freqüents en la població i les que es troben presents en diverses famílies però no en tots els membres d'aquestes. Com s'ha especificat anteriorment en la *Figura 6* s'han seleccionat les variants que complissin uns límits en la freqüència al·lèlica, tinguessin prediccions patogèniques en les eines bioinformàtiques en el cas de les variants *missense* i una cobertura de seqüenciació adequada. D'altra banda, la resta de paràmetres estudiats com la conservació de l'aminoàcid, si el canvi es produeix en un domini important de la proteïna i la funció d'aquesta proteïna, són eines que han servit per filtrar però que no permeten saber si el canvi té un efecte important en l'estructura tridimensional de la proteïna, en la seva funció o en les seves interaccions i per tant no presenten uns límits, ja que no donen una mesura quantitativa sinó qualitativa. Per a la prioritització, s'han seleccionat les variants que compleixen els límits establerts i addicionalment presenten una funció interessant, interaccions associades amb gens que s'han associat amb càncer i que el canvi aminoacídic es trobi molt conservat i no present en cap altre animal evolutivament proper a l'*Homo sapiens* per tal de verificar que no pertany a una regió variable del genoma.

En aquesta prioritització, s'han obtingut un total de 27 variants, en les que algunes famílies han presentat 4 candidates i d'altres només 1 (*Taula 3*). A partir de les variants seleccionades s'ha aconseguit en tots els casos amplificar el fragment dels gens que contenia la variant d'interès per l'estudi, tal com es pot observar per algunes d'elles a la *Figura 8*.

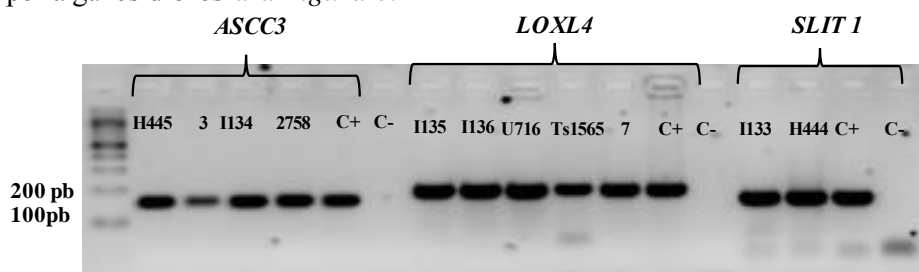


Figura 8. Resultats de l'amplificació del fragment d'interès de cada gen en *SCC3* de la FAM10, *LOXL4* de la FAM12 i *SLIT1* de la FAM9, a 58°C, en 40 cicles (10'',10'',15'') excepte *SLIT1* que s'ha optimitzat a 58°C, a 40 cicles (15'',15'',30''). Les mostres H445, I134, I135, I136, U716, I133 i H444 són mostres de sang perifèrica de pacients amb CCR. Per altra banda les mostres 3, Ts1565 i 7 són d'ADN tumoral i la mostra 2758 és de sang perifèrica que correspon a un familiar addicional amb 12 pòlips benignes. C⁺, fa referència al control positiu i C⁻ al control negatiu.

Tal com es pot observar a la *Figura 8*, s'ha realitzat una correcta amplificació de les mostres en tots els casos observant-se una banda intensa entre els 100-200 pb. Tanmateix es pot apreciar en alguns casos una segona banda poc definida per sota els 100 pb, aquesta correspon a dímers de primer que han hibridat i s'han quedat retinguts al final del gel d'agarosa. No s'ha observat en cap cas una banda d'amplificació en la regió de les mostres en els controls negatius.

Taula 3. Descripció de les variants seleccionades de les 16 famílies, segons la freqüència externa, les eines bioinformàtiques, si es troba en un domini, les interaccions i funcions i referències en l'OMIM.

Gen	Mutació	Freqüència	In silico	Domini	Conservació	Interaccions (selecció)	Funció	OMIM
<i>CENPE</i>	c.7678A>C (p.Lys2560Gln)	0	3,5	Globular autoinhibitori	59/73	<i>BUB1B,CENFE,FANCA4</i>	Segregació dels cromosomes pel fus mitòtic. Gen candidat a CCR (DeRycke <i>et al.</i> , 2013)	Microencefàlia autosòmica recessiva
<i>EPHB6</i>	c.2384G>A (p.Arg795His)	5/120126	5,25	Unió a ATP	99/99	<i>BCL6,EPHB1,EPHB4</i>	Supressió de l'invasió del càncer. Variants associades a CCR familiar	602757
<i>SIK3</i>	c.1871G>A (p.Arg624Gln)	3/121204	5,25	-	100/100	<i>HDAC5,MAP3K7 SMAD2, SMAD3</i>	Progressió de la fase G1a S del cicle cel·lular	614776
<i>RREB1</i>	c.3206C>T (p.Ala1069Val)	20/116936	5,25	-	92/98	<i>AR,BLM,HDAC1,HDAC2, HDAC3</i>	Regulació de la transcripció. Unió a elements de resposta de <i>RAS</i>	602209
<i>ARHGAP12</i>	c.1970A>T (p.Asn657Ile)	3/121220	4,25	Rho	80/99	<i>GRB2,MAPK11,SRPK1, XPO</i>	Supressió tumoral per regulació de la invasió i adhesió cel·lular	610577
<i>CIDEB</i>	c.116G>A (Arg39Pro)	5/121300	5,25	Lloc heterodímer	76/79	<i>CIDEA</i>	Apoptosis com a resposta del dany a l'ADN	604441
<i>GEMIN5</i>	c.194A>G (p.Glu65Gly)	0	4,25	WD40	72/96	<i>BARD1,LGR4,MAP2K4, MAP3K5,MAPK3</i>	Splicing	607005
<i>RGS3</i>	c.1059A>T (p.Lys353Asn)	0	5,25	PDZ, DegQ	93/96	<i>EPHRINB1,ESR1</i>	Senyal proteïnes G. Sobre-expressat en tumors amb p53 mutat. Inhibició de <i>MAPK</i>	602189
<i>SLIT1</i>	c.4504C>T (p.Arg1502Trp)	11/119992	5,25	-	90/98	<i>GPC1,ROBO1</i>	Regulador negatiu de <i>Sdf1</i> i <i>Cxcr4</i> implicats en càncer de mama	603742
<i>BCL9L</i>	c.16A>T (p. Asn6Tyr)	6/121274	4,25	-	56/80	<i>CTNNB1,EWSR1,YAP1</i>	Progressió de tumors intestinals primerencs. Via canònica Wnt	609004
<i>ASCC3</i>	c. 2527A>C (p. Ile843Leu)	0	4,25	HELICc	98/99	<i>CDK1,EIF3F,SOX2</i>	Proliferació cel·lular	614217
<i>LOXL4</i>	c.1531G>A (p.Gly511Lys)	34/120418	5,25	Scavenger receptor Cys-rich	86/100	<i>MAST1,PRMPT6,TRIP13</i>	Progressió del càncer de cap i coll	607318
<i>NSMCE1</i>	c.476T>C (p.Leu159Pro)	1/120694	5,25	SMC_Nse1	73/73	<i>PMF1</i>	Regulació positiva estimulada per dany a l'ADN	-
<i>BAP1</i>	c.1133C>T (p.Ala378Val)	0	4,25	-	85/98	<i>ANAPC7,BAP1,BRCA1 CASP8,ESR1,IGF1R, PTEN,PTPRJ, XRCC3</i>	Regulació del cicle cel·lular. Supressor tumoral per desubiquització. Via de control de <i>BRCA1</i>	603089: Síndrome de predisposició a tumors

<u>HERC2</u>	c.12490G>A (p. Asp4164Asn)	0	4,25	Regió 3-5 RCC1 repetida	99/99	<i>BRCA1,CDK8,EPCAM, MAPK6,TP53</i>	Reparació del ADN. Mutacions somàtiques en càncer gàstric i CCR amb inestabilitat de microsatèl·lits	615516: Retard mental
<u>UTP11L</u>	c.518C>A (p. Thr173Asn)	2/121384	4,25	Utp11	88/97	<i>CRNN,SPRR3</i>	Regulació positiva del procés apoptòtic	609440
<u>BLM</u>	c.2069C>T (p. Pro690Leu)	1/121140	5,25	Lloc unió ATP	96/96	<i>BRCA1,BRIP1,ATR,ATM, FANCD2,FANCD,FANCM, FAN1,MLH1,MRE11A,MSH2, MSH6,RAD51,TP53,XRCC2</i>	Cicle cel·lular. Recombinació meiótica. Augment del risc de patir CCR	210900: Síndrome de Bloom
<u>MUC6</u>	c.1778delC (p.M594Cfs*3)	0	FS	-	-	<i>TFF1</i>	Inhibició de l'invasió en càncer	158370 158374: associat a susceptibilitat a càncer gàstric
<u>MCM3</u>	c.2401C>T (p. Arg801Trp)	3/121370	3,25	-	55/99	<i>CCND1,CDK2,EP300, MCM4,MAPK6,MGMT, NOTCH1,POLD1,RAD52, SMARCA5</i>	Cicle cel·lular, replicació del ADN Elevada expressió associada a neoplasmes	602693
<u>CLPTM1L</u>	c.976+1G>A	0	SP	-	-	<i>NCSTN,NTRK1,PIK3C3, SPTAN1,TUBG1,ZBTB33</i>	Apoptosi. Localitzat a regió de susceptibilitat amb <i>TERT</i> per molts tipus de càncers	612585
<u>HIST1H4H</u>	c.82C>T (p. Gln28*)	3/121380	ST	-	-	<i>CSN5,H2AFZ,ITGA4, MCM2,RAD21</i>	Mitosi. Replicació de l'ADN. Assemblatge nucleosoma	602828
<u>SLIT2</u>	c.4400G>T (p. Gly1467Val)	32/121258	4,25	Homòleg <i>SLIT2</i>	100/100	<i>ANKRD12, MYC, PTCH1, ROBO1, ROBO2, SLIT2</i>	Regulació apoptosi. Candidat supressor tumoral per CCR	603742
<u>PSMB10</u>	c.172C>G (p. Arg58Gly)	0	5,25	Subunitat proteosoma β	81/94	<i>ANXA7,PSME1,PSME2,PSMB3, PSMD6,PTN</i>	Resposta dany de l'ADN. Regulació apoptosi. Transducció senyal <i>p53</i>	176847
<u>GSK3A</u>	c.1295T>C (p. Ile432Thr)	2/106326	3,25	-	78/100	<i>BCL3,MYC,TOR</i>	Senyalització apoptosi. Diana terapèutica per càncer de pàncrees.	606784
<u>UTP11L</u>	c.665G>A (p. Arg222His)	3/117600	5,25	-	92/100	<i>CRNN,SPRR3</i>	Regulació apoptosi	609440
<u>MCM4</u>	c.2039A>C (p. Glu680Ala)	0	4,25	MCM	98/100	<i>CDK1,CDK2,EP300,HDAC5, MCM3,POLK,RAD21</i>	Replicació de l'ADN. La seva fosforilació la inhibeix	609981: deficiència cèl·lules NK amb un dany en la reparació del ADN
<u>HIST3H2A</u>	c.106C>A (p. Arg36Ser)	1/114798	5,25	Lloc unió ADN	86/93	<i>RPS6KA5,CHD1L,MCM5, RNF2</i>	Desassemblatge del nucleosoma. Reparació de dany a l'ADN per rajos UV	615015

CCR, càncer colorectal. FS, frameshift; NS, nonsense; SP, canonical splice site.

Freqüència al·lèlica, ExAC: Exome Aggregation Consortium (versió 0.3.1). El valor 0 indica l'absència de la variant a la base de dades. Aquestes dades són proporcionades per varis projectes de seqüenciació a gran escala que proporcionen informació de 60706 individus no emparentats. In silico: nombre de eines de predicció de patogenicitat bioinformàtiques. Domini: regió de la proteïna on es troba el canvi. Conservació: Dades del UCSC genome browser que proporcionen informació de la conservació de 100 espècies (11 primats, 52 mamífers i 37 vertebrats) en la posició concreta de la mutació. Algunes espècies, sobretot vertebrats s'observen gaps en la posició de la mutació i en aquests casos pel contacte no s'ha tingut en compte aquella espècie. Interaccions: aquelles que s'han associat a proteïnes associades amb càncer. OMIM: Introducció del número, també s'ha anotat si estan associats a malalties hereditàries.

Les variants seleccionades s'han confirmat per seqüenciació pel mètode de Sanger en tots els casos, excepte pels gens *PSMB10* i *HIST3H4* que no presentaven la variant i que correspondrien a errors de la NGS. Per altra banda, en el cas d'*ARHGAP1* no es va poder determinar la presència o no de la variant. Per tant finalment s'han pogut confirmar 24 de les 27 variants (Taula 4). També en les famílies 7, 8 que es tenia mostra de membres que presentaven adenomes avançats i limfoma, s'han realitzat estudis de segregació. A més, s'ha determinat la pèrdua d'heterozigositat (LOH) en les famílies que presentaven una bona segregació en tots els membres seqüenciats si es tenien mostres tumorals disponibles.

Taula 4. Resultats validació de les variants prioritzades, de l'estudi de segregació i pèrdua d'heterozigositat en les mostres tumorals.

Gen	Mutació	Mostres	Validació	Segregació	LOH
<i>CENPE</i>	c.7678A>C (p.Lys2560Gln)	FAM1	2/2	-	NA
<i>EPHB6</i>	c.2384G>A (p. Arg795His)	FAM1	2/2	-	NA
<i>SIK3</i>	c.1871G>A (p.Arg624Gln)	FAM1	2/2	-	NA
<i>RREB1</i>	c.3206C>T (p.Ala1069Val)	FAM2	2/2	-	-
<i>ARHGAP12</i>	c.1970A>T (p.Asn657Ile)	FAM4	-	-	NA
<i>CIDEB</i>	c.116G>A (p.Arg39Pro)	FAM7	2/2	1/1	N
<i>GEMIN5</i>	c.194A>G (p.Glu65Gly)	FAM8	3/3	NA	NA
<i>RGS3</i>	c.1059A>T (p.Lys353Asn)	FAM9	2/2	-	-
<i>SLIT1</i>	c.4504C>T (p.Arg1502Trp)	FAM9	2/2	-	-
<i>BCL9L</i>	c.16A>T (p. Ala5Tyr)	FAM9	2/2	-	-
<i>ASCC3</i>	c.2527A>C(p. Ile843Leu)	FAM10	2/2	-	N
<i>LOXL4</i>	c.1531G>A (p.Gly511Lys)	FAM12	3/3	-	N
<i>NSMCE1</i>	c.476T>C (p.Leu159Pro)	FAM_NOVA_1	2/2	-	N
<i>BAP1</i>	c.1133C>T (p.Ala378Val)	FAM_NOVA_2	2/2	-	NA
<i>HERC2</i>	c.12490G>A(p.Asp4164Asn)	FAM_NOVA_2	2/2	-	NA
<i>UTP11L</i>	c.518C>A (p.Thr173Asn)	FAM_NOVA_3	2/2	-	NA
<i>UTP11L</i>	c.665G>A (p.Arg222His)	FAM23	4/4	-	-
<i>BLM</i>	c.2069C>T(p. Pro690Leu)	FAM_NOVA_4	2/2	-	N
<i>MUC6</i>	c.1778delC (p.M594Cfs*3)	FAM19	3/3	-	N
<i>MCM3</i>	c.2401C>T(p.Arg801Trp)	FAM19	3/3	-	S
<i>CLPTMIL</i>	c.976+1G>A	FAM19	2/2	-	NA
<i>HIST1H4</i>	c.82C>T (p. Gln28*)	FAM19	3/3	-	N
<i>SLIT2</i>	c.4400G>T(p.Gly1467Val)	FAM21	3/3	-	-
<i>PSMB10</i>	c.172C>G(p.Arg58Gly)	FAM21	0/3	-	-
<i>GSK3A</i>	c.1295T>C (p.Ile432Thr)	FAM22	4/4	-	-
<i>MCM4</i>	c.2039A>C (p. Glu680Ala)	FAM39	2/2	-	-
<i>HIST3H2A</i>	c.106C>A (p.Arg36Ser)	FAM39	1/2	-	-

CCR, càncer colorectal; Validació, segregació dels pacients amb CCR; Segregació CCR, segregació en familiars addicionals que presentin adenomes avançats, pòlips, presentin algun altre tipus de càncer; LOH, pèrdua d'heterozigositat; N, no ; NA, error en l'amplificació; S, sí.

Mostra: identificador de cada variant detectada per seqüenciació completa de l'exoma la variant en més d'un individu de la família. Validació: Comprovació de la variant en tots els individus seqüenciats per exoma a partir del mètode Sanger. Segregació: Proporció de casos de familiars amb adenomes avançats, pòlips o altres tipus de càncer que s'han seqüenciat pel mètode Sanger i presenten la mateixa variant en heterozigosis. LOH: delació de l'al·lel normal en la posició de la variant en ADN tumoral comparant amb l'ADN germinal del mateix pacient.

A partir dels resultats obtinguts, els millors candidats per a la predisposició germinal de CCR són *LOXLA*, *GEMIN5*, *UTP11L*, *MUC6*, *HIST1H4*, *GSK3A* i *SLIT2*, ja que presenten tres o més individus a la família amb CCR presentant forta agregació. També cal remarcar en els estudis de segregació que s'ha identificat com a bon candidat al CCR, *CIDEB* en la família 7 on s'ha comprovat la segregació en els familiars amb CCR i confirmat la presència d'aquesta variant en una dona de 26 anys amb limfoma. Per últim també cal destacar un altre gen bon candidat, *MCM3*, en aquest es confirma la variant en els familiars afectats i en la mostra tumoral d'un dels afectats s'ha observat una lleugera pèrdua d'heterozigositat tal com es pot observar a la *Figura 9*.

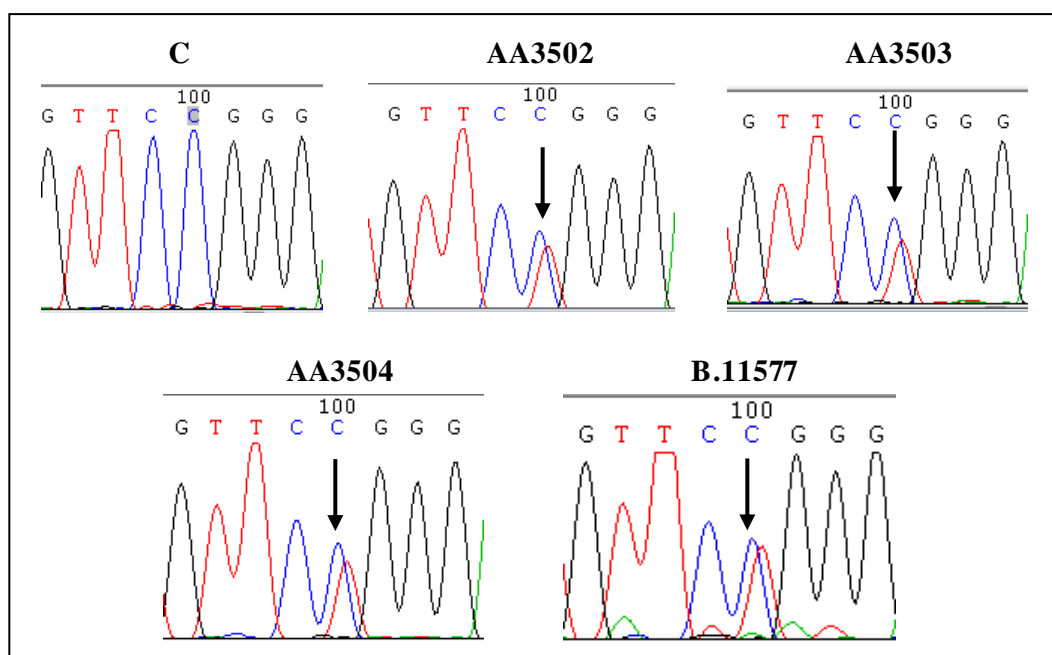


Figura 9: Resultats seqüenciació Sanger visualitzats pel programa Chromas de la variant en *MCM3* de la família 19 en la mostra control (C), els tres familiars amb CCR (AA3502, AA3593, AA3504) i la mostra d'ADN tumoral (B.11577) que correspon al germinal AA3504.

5. Discussió

5.1 Discussió dels resultats

La seqüenciació de l'exoma de 46 pacients amb CCR de 16 famílies amb forta agregació familiar han permès la identificació de les variants en els gens *LOXLA*, *GEMIN5*, *UTP11L*, *MUC6*, *HIST1H4*, *GSK3A*, *SLIT2*, *CIDEB* i *MCM3* per a la predisposició al CCR.

LOXLA pertany a la família gènica LOX, aquesta codifica per enzims extracel·lulars que s'encarreguen de catalitzar la desaminació oxidativa de molècules de col·lagen i elastina per tal d'insolubilitzar-les en la matriu extracel·lular. D'aquesta manera participa en el manteniment de l'estabilitat del teixit connectiu (Lucero & Kagan, 2006). Tanmateix, *LOXLA* pot bloquejar el factor de transcripció NF- κ B a

través de la inactivació de la via PI3K/Akt i Raf/MEK i actuar com a supressor tumoral. Pel que fa a nivell genòmic, el silenciament del gen per una metilació en el promotor s'ha associat a càncer de bufeta. A més d'aquesta funció, s'ha determinat una regulació positiva en l'expressió d'aquest gen associat a la proliferació i invasió cel·lular i metàstasis en càncer de coll i cap i també en càncer gàstric a partir de la via FAK/Src (Jeay *et al.*, 2003; Wu *et al.*, 2007; Li *et al.*, 2015; Weise *et al.*, 2008). La mutació c.1531G>A (p.Gly511Lys) es localitza en el domini Scavenger receptor Cys-rich, aquest intervé en les interaccions proteïna-proteïna i proteïna-lligand de l'enzim Lisil oxidasa (Hohenester *et al.*, 1999). Per aquest motiu la proteïna pot presentar canvis en l'afinitat, estabilitat i en la transducció del senyal amb algunes de les seves interaccions. Per altra banda, aquesta família presenta dos familiars de 14 i 16 anys amb càncer de cap, en els que es podria comprovar la presència d'aquesta variant.

GEMIN5 codifica per una proteïna clau en el complex SMN i és l'encarregada d'unir-se als snRNAs. Mutacions en aquest gen afecten la unió dels snRNAs al complex de splicing i poden influir en la tumorogènesis i en la progressió del tumor (Battle *et al.*, 2006; Lee *et al.*, 2006). La mutació c.194A>G (p.Glu65Gly) es localitza en el domini WD40, aquest domini conté còpies repetides en tàndem de 40 aminoàcids que s'uneixen per formar un domini de forma circular. Aquest domini està implicat en la transducció del senyal, en la regulació de la transcripció i en el control cel·lular (Xu & Min, 2011). Mutacions en aquest domini poden interferir en l'estructura de la proteïna i ocasionar problemes amb algunes de les seves interaccions com és *MAPK3* que pertany a la cascada *MAPK/ERK* associada a la iniciació o regulació de la mitosi.

UTP11L codifica per una ribonucleoproteïna requerida per la biogènesi del 18S rRNA. Tanmateix aquesta proteïna s'ha associat a una regulació positiva de l'apoptosi al cervell (Phipps *et al.*, 2011; Heese *et al.*, 2003). La mutació c.665G>A (p.Arg222His) pot causar defectes en l'estructura de la proteïna que interfereixen en la biogènesi dels ribosomes que poden causar efectes importants en el metabolisme i la viabilitat cel·lular. En aquest estudi s'han considerat dues variants d'aquest gen com a candidates, de les quals les dues s'han validat en les seves respectives famílies, per aquest motiu pren un major interès aquest gen per a la predisposició al CCR.

MUC6 codifica per una mucina gàstrica que s'encarrega de la protecció del tracte gastrointestinal de proteases, àcids, microorganismes patògens i danys mecànics (Toribara *et al.*, 1993). S'ha relacionat aquest gen com a inhibidor de la invasió de cèl·lules tumorals en càncer de pàncrees però també es troba expressat de forma aberrant en la progressió d'alguns casos de CCR (Leir & Harris, 2011). La mutació c.1778delC (p.M594Cfs*3) produeix que es trunqui la proteïna ocasionant un efecte important en la funció d'aquesta.

HIST1H4 codifica per un membre de la família d'histones H4, que participa en l'assemblatge del nucleosoma. A més participa en la regulació del cicle cel·lular en la replicació d'ADN quan es troba acetilada (Miotto & Struhl, 2010). La mutació c.82C>T (p.Gln28*) provoca la formació d'un codó d'aturada a l'inici de la proteïna causant la pèrdua de la seva activitat, ja que conté els dominis d'interacció amb les altres histones i punts d'acetilació que es troben upstream de la mutació. Errors en l'empaquetament del nucleosoma poden ocasionar problemes durant el cicle, la divisió cel·lular i l'expressió gènica per formar d'aquesta manera cèl·lules aberrants.

GSK3A, codifica per un glicogen 3 α sintasa quinasa, és una serin treonin quinasa que està implicada en el control de diverses proteïnes com el glicogen sintasa i factors de transcripció com *JUN* (Ali *et al.*, 2001). També participa en la via de Wnt i la PI3K/Akt, ambdues relacionades amb la proliferació i creixement cel·lular i associades a càncer colorectal (Wu & Pan, 2010). La mutació c.1295T>C (p.Ile432Thr) no es troba en cap domini important de la proteïna, però pot afectar a l'estructura, estabilitat i interaccions d'aquesta.

SLIT2, codifica per una de les glicoproteïnes de la família SLIT. Té un paper important en la guia axònica i migració neuronal on *SLIT2* és el lligand per la família de receptors ROBO. Estudis recents proposen *SLIT2* com a supressor tumoral ja que s'ha trobat inactivat en càncer de mama, de pulmó i colorectal (Dallol *et al.*, 2002; Dallol *et al.*, 2003). La metilació de *SLIT2* s'ha associat amb alguns subtipus de càncer gàstric (Kim *et al.*, 2016). A més, la mutació c.4400G>T (p.Gly1467Val) es troba a una posició de l'homòleg *SLIT2* C-terminal. Aquesta regió permet la interacció amb els receptors ROBO i és una regió difusible. Canvis en aquesta regió de la proteïna provocaran canvis en la funció i possiblement disminuint la seva activitat i interaccions. A més, s'ha seleccionat un altre gen que pertany a la família SLIT, la variant en *SLIT1* que també s'ha validat. Per aquest motiu aquesta família gènica podria estar associada a la predisposició al CCR.

CIDEB, codifica per un efector B de tipus DFFA inductor de mort cel·lular (Inohara *et al.*, 1998). La mutació c.116G>A (p.Arg39Pro) afecta a un domini heterodimèric d'interacció amb altres lligands i proteïnes, que pot provocar una desregulació del procés d'apoptosi.

Per últim hi ha *MCM3*, pertany a la família de proteïnes de manteniment de mini-cromosoma (MCM) altament conservades que participen en la iniciació de la replicació del genoma eucariota. El complex MCM té un paper determinant en la replicació d'ADN (Bailis & Forsburg, 2004). Per altra banda, s'ha observat una sobreexpressió de *MCM3* en diferents càncers humans proposant la seva implicació en el creixement i la proliferació de les cèl·lules cancerígenes (Ha *et al.*, 2004). La mutació c.2401C>T (p.Arg801Trp) no està situada a cap domini de la proteïna però canvis en l'estructura d'aquesta podrien influir la seva afinitat, expressió i funció. En aquest estudi, s'ha seleccionat també

MCM4 com a candidat en la família 39 que també s'ha validat en els dos membres afectats. Tanmateix, la mutació c.2039A>C (p.Glu680Ala) es localitza en un domini important de la proteïna, el MCM. Per tant, es podria proposar que canvis en els gens que constitueixen aquesta família gènica podrien ser possibles responsables de la predisposició de CCR.

Entre tots els candidats, els que per les seves funcions, característiques i el tipus de canvi en el gen més probablement causa un canvi important en la funció són *LOXLA*, *HIST1H4*, *GSK3A* i *MCM3*.

També cal destacar els resultats de la família 19 en què s'han verificat quatre bons candidats per a la predisposició al CCR (*MUC6*, *CLPTMIL*, *HIST1H4* i *MCM3*), en aquesta família es troben 6 individus amb càncer de còlon dels quals només 3 s'han pogut utilitzar per a l'estudi. Mitjançant estudis de segregació en els altres membres afectats es podria acabar de determinar quina variant afecta significativament al CCR.

5.2 Ètica i sostenibilitat

L'objectiu d'aquest estudi és la determinació de noves variants genètiques implicades amb el CCR, per aquest motiu per dur a terme l'estudi es requereixen mostres de pacients amb CCR que donin el consentiment per tal de ser inclosos al projecte d'investigació. Cada pacient és informat a partir d'un document on es descriu l'objectiu del projecte, els procediments que es realitzaran amb la mostra de sang que subministra per l'estudi i asseguren que els resultats obtinguts a partir de la mostra es dedicaran exclusivament per l'estudi a la predisposició al CCR segons els termes previstos en la legislació vigent (Llei 14/2007 i RD1716/2011). Tanmateix s'explica on es guardarà la mostra, el benefici que pot causar de cara a un cert temps en la seva família si es dona el cas que es troba una nova variant associada i també els riscos, i deixant clar la participació voluntària del pacient i la confidencialitat de l'estudi. Una vegada el pacient ha donat el seu consentiment, les mostres passen a disposició dels investigadors amb un codi determinat per cada hospital i no es té coneixement del nom ni la localització del pacient. A més, el pacient pot anul·lar el conveni en tot moment fent que la mostra d'aquest es deixi d'utilitzar.

Pel que fa a la sostenibilitat, s'ha seguit de manera rigorosa els protocols de separació de residus per un correcte tractament dipositant tot el material que ha estat en contacte amb material biològic en els seus contenidors corresponents que es recullen externament i s'incineren. Tanmateix el material utilitzat al laboratori és autoclavat, les puntes de pipeta contenen un filtre per minimitzar contaminacions i el gel d'agarosa es llença al contenidor de residus citotòxics pel seu tractament posterior. De la mateixa manera, s'ha complert amb les normes de seguretat de treball en un laboratori. En l'elaboració de la PCR s'han seguit uns controls molt exigents d'esterilitat. En la preparació de les mix s'ha treballat en una sala estèril amb material exclusiu per aquesta i s'ha realitzat sota una cabina de flux laminar

prèviament netejada amb DNazap i posteriorment irradiada amb tot el material que s'utilitza excepte la polimerasa i els primers durant 10-15 minuts amb llum ultraviolada. Posteriorment s'ha de fer un canvi de sala per posar les mostres dels pacients sota una altra campana de flux laminar esterilitzada amb les mateixes condicions i per realitzar la PCR, d'aquesta manera s'eviten possibles contaminacions.

Per millorar l'eficàcia del mètode, s'hauria de fer un estudi a nivell de proteïna. Caldria fer estudis funcionals per determinar l'efecte d'aquestes mutacions a nivell d'expressió, d'estabilitat i d'interacció amb altres proteïnes. Aquests estudis a partir de totes les variants seleccionades i validades es realitzaran a partir d'una *Massive Parallel pipeline* pel clonatge de variants d'ADN i que permet examinar el fenotip d'aquestes mutacions (Wei *et al.*, 2014).

6. Conclusions

- Whole exome sequencing, validation, segregation and tumour studies help us to filter all the variants in each family in order to keep only some of them. Positive results do not indicate a certain relation to colorectal cancer (CRC) susceptibility, it is also needed to carry out with functional and in vitro studies.
- Most of the variants selected were confirmed by Sanger sequencing, indicating the sequencing method's efficiency.
- *CIDEB*, *ASCC3*, *LOXLA*, *NSMCE1*, *BLM*, *MUC6* and *HIST1H4* do not have LOH. However, they are still good candidates because it can be another genetic change along the whole sequence not founded by this technique.
- *LOXLA*, *HIST1H4*, *GSK3A* and *MCM3* are probably better candidates to CRC susceptibility due to their mutations, functions, pathways and associations with cancer.
- Genes from the same family can be involved in the same disease which possibly happens with MCM family and SLIT family.
- A gene can be involved in different cancer types as it probably happens with *LOXLA*.

7. Bibliografia

- Ali, A., Hoeflich, K. P., & Woodgett, J. R. (2001). Glycogen synthase kinase-3: properties, functions, and regulation. *Chemical reviews*, *101* (8), 2527-2540.
- Alrawi, S. J., Schiff, M., Carroll, R. E., Dayton, M., Gibbs, J. F., Kulavlat, M., ... Anderson, G. R. (2006). Aberrant Crypt Foci. *Anticancer research*, *26* (1A), 107-119.
- Aran, V., Victorino, A. P., Thuler, L. C., & Ferreira, C. G. (2015). Colorectal Cancer: Epidemiology, Disease Mechanisms and Interventions to Reduce Onset and Mortality. *Clinical Colorectal Cancer*, *22* (4), 1-9.
- Baena, R., & Salinas, P. (2015). Diet and colorectal cancer. *Maturitas*, *80* (3), 258-264.
- Bailis, J. M., & Forsburg, S. L. (2004). MCM proteins: DNA damage, mutagenesis and repair. *Current Opinion in Genetics & Development*, *14* (1), 17-21.
- Battle, D. J., Lau, C. K., Wan, L., Deng, H., Lotti, F., & Dreyfuss, G. (2006). The *GEMIN5* protein of the SMN complex identifies snRNAs. *Molecular Cell*, *23* (2), 273-279.
- Buecher, B., Mariani, P., Audollent, R., De Singly, B., Lièvre, A., & Cacheux, W. (2015). *MUTYH* Status and Colorectal Cancer Risk: Implication for Surveillance. *Current Colorectal Cancer Reports*, *11* (1).
- Brenner, H., Kloor, M., & Pox, C. P. (2014). Colorectal cancer. *The Lancet*, *383* (9927), 1490-1502.
- Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics*, *33* (1.3s), 228-237.
- Burt, R., & Neklason, D. W. (2005). Genetic testing for inherited colon cancer. *Gastroenterology*, *128* (6), 1696-1716.
- Burt, R. (2007). Inheritance of colorectal cancer. *Drug Discovery Today: Disease Mechanisms*, *4* (4), 293-300.
- Carballal, S., Moreira, L., & Balaguer, F. (2013). Pólipos serrados y síndrome de poliposis serrada. *Cirugía Española*, *91* (3), 141-148.
- Center, M. M., Jemal, A., Smith, R. A., & Ward, E. (2009). Worldwide Variations in Colorectal Cancer. *CA: A Cancer Journal for Clinicians*, *59* (6), 366-378.
- Chan, D. S. M., Lau, R., Aune, D., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2011). Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. *PLoS ONE*, *6* (6). e20456.
- Chu, K. M. (2011). *Epidemiology and Risk Factors of Colorectal Cancer. Early diagnosis and treatment of cancer: Colorectal Cancer*. Elsevier Inc.
- Dalol, A., Da Silva, N. F., Viacava, P., Minna, J. D., Bieche, I., Maher, E. R., & Latif, F. (2002). *SLIT2*, a human homologue of the *Drosophila Slit2* gene, has tumor suppressor activity and is frequently inactivated in lung and breast cancers. *Cancer research*, *62* (20), 5874-5880.

- Dallol, A., Morton, D., Maher, E. R., & Latif, F. (2003). *SLIT2* axon guidance molecule is frequently inactivated in colorectal cancer and suppresses growth of colorectal carcinoma cells. *Cancer research*, *63* (5), 1054-1058.
- De la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. *Nature Reviews Cancer*, *4* (10), 769-780.
- DeRycke, M. S., Gunawardena, S. R., Middha, S., Asmann, Y. W., Schaid, D. J., McDonnell, S. K., ... & Serie, D. J. (2013). Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiology Biomarkers & Prevention*, *22* (7), 1239-1251.
- de Voer, R. M., Hahn, M.-M., Mensenkamp, A. R., Hoischen, A., Gilissen, C., Henkes, A., ... Kuiper, R. P. (2015). Deleterious Germline *BLM* Mutations and the Risk for Early-onset Colorectal Cancer. *Scientific Reports*, *5* (14060), 1-7.
- Esteban-Jurado, C., Garre, P., Vila, M., Lozano, J. J., Pristoupilova, A., Beltrán, S., ... Castellvi-Bel, S. (2014). New genes emerging for colorectal cancer predisposition. *World Journal of Gastroenterology*, *20* (8), 1961-71.
- Fearon, E. R. & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61* (5), 759-767.
- Ferlay, J., Shin, H.R., Bray, F., Foman, D., Mathers, C., & Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, *127* (12), 2893-2917.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, Maxwell, P., Forman, D., & Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, *136* (5), E359-386.
- Fuchs, S. C., Giovannucci, L. E., Graham, A. C., Hunter, J. D., Frank, E. S., & Willett, C. W. (1994). A prospective study of family history and the risk of colorectal cancer. *The New England Journal of Medicine*, *331* (25), 1669-1674.
- Guo, Y., Li, J., Li, C.-I., Long, J., Samuels, D. C., & Shyr, Y. (2012). The effect of strand bias in Illumina short-read sequencing data. *BMC Genomics*, *13* (1), 666.
- Ha, S. A., Shin, S. M., Namkoong, H., Lee, H., Cho, G. W., Hur, S. Y., ... & Kim, J. W. (2004). Cancer-associated expression of minichromosome maintenance 3 gene in several human cancers and its involvement in tumorigenesis. *Clinical Cancer Research*, *10* (24), 8386-8395.
- Half, Elizabeth E., and Robert S. Bresalier. (2004). Clinical management of hereditary colorectal cancer syndromes. *Current Opinion in Gastroenterology*, *20* (1), 32-42.
- Heese, K., Nagai, Y., & Sawada, T. (2003). Comparative gene identification 94 a pivotal regulator of apoptosis. *Neuroscience*, *116* (2), 321-324.
- Hohenester, E., Sasaki, T., & Timpl, R. (1999). Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. *Nature Structural & Molecular Biology*, *6* (3), 228-232.
- Inohara, N., Koseki, T., Chen, S., Wu, X., & Núñez, G. (1998). *CIDE*, a novel family of cell death activators with homology to the 45 kDa subunit of the DNA fragmentation factor. *The EMBO journal*, *17* (9), 2526-2533.

- International Agency for Research on Cancer. (2012). GLOBOCAN. Recuperat de <http://globocan.iarc.fr/Default.aspx>
- Jasperson, W. K., Tuohy, M. T., Neklason, W. D., & Randall, W. B. (2013). Hereditary and Familial Colon Cancer. *Gastroenterology*, *10* (2), 249–253.
- Jeay, S., Pianetti, S., Kagan, H. M., & Sonenshein, G. E. (2003). Lysyl oxidase inhibits ras-mediated transformation by preventing activation of *NF-κB*. *Molecular and Cellular Biology*, *23* (7), 2251–2263.
- Jelsig, A. M., Qvist, N., Brusgaard, K., Nielsen, C. B., Hansen, T. P., & Ousager, L. B. (2014). Hamartomatous polyposis syndromes: a review. *Orphanet Journal of Rare Diseases*, *9* (1), 1–10.
- Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B., & Berry, D. A. (2013). Meta-analyses of colorectal cancer risk factors. *Cancer Causes & Control*, *24* (6), 1207–22.
- Kim, M., Kim, J. H., Baek, S. J., Kim, S. Y., & Kim, Y. S. (2016). Specific expression and methylation of *SLIT1*, *SLIT2*, *SLIT3*, and miR-218 in gastric cancer subtypes. *International journal of Oncology*, *48* (6), 2497–2507.
- Knudson, A. G. (1996). Hereditary cancer: two hits revisited. *Journal of Cancer Research and Clinical Oncology*, *122* (3), 135–140.
- Lamprecht, S. A., & Lipkin, M. (2003). Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms. *Nature Reviews Cancer*, *3* (8), 601–14.
- Lee, J. H., Horak, C. E., Khanna, C., Meng, Z., Yu, L. R., Veenstra, T. D., & Steeg, P. S. (2008). Alterations in Gemin5 expression contribute to alternative mRNA splicing patterns and tumor cell motility. *Cancer Research*, *68* (3), 639–644.
- Leir, S. H., & Harris, A. (2011). *MUC6* mucin expression inhibits tumor cell invasion. *Experimental Cell Research*, *317* (17), 2408–2419.
- Li, R. K., Zhao, W. Y., Fang, F., Zhuang, C., Zhang, X. X., Yang, X. M., ... & Yang, S. L. (2015). Lysyl oxidase-like 4 (LOXL4) promotes proliferation and metastasis of gastric cancer via *FAK/Src* pathway. *Journal of Cancer Research and Clinical Oncology*, *141* (2), 269–281.
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153* (6), 1194–1217.
- Lucero, H. A., & Kagan, H. M. (2006). Lysyl oxidase: an oxidative enzyme and effector of cell function. *Cellular and Molecular Life Sciences*, *63* (19–20), 2304–2316.
- Miotto, B., & Struhl, K. (2010). HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Molecular Cell*, *37* (1), 57–66.
- Nieminen TT, O'Donohue MF, Wu Y, Lohi H, Scherer SW, Paterson AD, Ellonen P, Abdel-Rahman WM, Valo S, Mecklin JP, Järvinen HJ, Gleizes PE, Peltomäki P. (2014). Germline mutation of *RPS20*, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology*, *147* (3), 595–598.
- Nussbaum, R. L., McInnes, R. R., & Willard, H. F. (2015). *Thompson & Thompson genetics in medicine (8th edition)*. Elsevier Health Sciences.

- Ouakrim, D. A., Dashti, S. G., Chau, R., Buchanan, D. D., Clendenning, M., Rosty, C., ... Win, A. K. (2015). Aspirin, Ibuprofen, and the Risk for Colorectal Cancer in Lynch Syndrome. *Journal of the National Cancer Institute*, *107* (9), 1–11.
- Park, M. H., Rhee, H., Park, J. H., Woo, H. M., Choi, B. O., Kim, B. Y., ... Koo, S. K. (2014). Comprehensive analysis to improve the validation rate for single nucleotide variants detected by next-generation sequencing. *PLoS ONE*, *9* (1). e86664.
- Phelan, C. M., Iqbal, J., Lynch, H. T., Lubinski, J., Gronwald, J., Moller, P., ... Narod, S. a. (2013). Incidence of colorectal cancer in *BRCA1* and *BRCA2* mutation carriers: results from a follow-up study. *British Journal of Cancer*, *110* (2), 530–534.
- Phipps, K. R., Charette, J. M., & Baserga, S. J. (2011). The small subunit processome in ribosome biogenesis progress and prospects. *Wiley Interdisciplinary Reviews: RNA*, *2* (1), 1-21.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81* (3), 559-575.
- Randall W. Burt. (2000). Colon Cancer Screening. *Gastroenterology*, *119* (3), 837-853.
- Seguí, N., Mina, L. B., Lázaro, C., Sanz-Pamplona, R., Pons, T., Navarro, M., ... Valle, L. (2015). Germline Mutations in *FANL* Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology*, *149* (3), 563–566.
- Short, E., Thomas, L. E., Hurley, J., Jose, S., & Sampson, J. R. (2015). Inherited predisposition to colorectal cancer: towards a more complete picture. *Journal of Medical Genetics*, *52* (12), 791-796.
- Stewart, S.L., & Wike, J.M. (2006). A Population-Based Study of Colorectal Cancer. *Cancer*, *107* (5), 1128-1141.
- Tantamango-Bartley, Y., Jaceldo-Siegl, K., Fan, J., & Fraser, G. (2013). Vegetarian diets and the incidence of cancer in a low-risk population. *Cancer Epidemiology Biomarkers and Prevention*, *22* (2), 286–294.
- Tenesa, A., & Dunlop, M. G. (2009). New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nature Review Genetics*, *10* (6), 353–358.
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, *14* (2), 178-192.
- Tomlinson, I. (2015). The Mendelian colorectal cancer syndromes. *Annals of Clinical Biochemistry: An International Journal of Biochemistry and Laboratory Medicine*, *52* (6), 690-692.
- Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W. L., Gum, E., Hicks, J. W., ... & Kim, Y. S. (1993). Human gastric mucin. Identification of a unique species by expression cloning. *Journal of Biological Chemistry*, *268* (8), 5879-5885.
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-tieulent, J., & Jemal, A. (2015). Global Cancer Statistics, 2012. *CA: A Cancer Journal of Clinician*, *65* (2), 87–108.
- Valle, L. (2014). Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World Journal of Gastroenterology*, *20* (29), 9828–9849.

- Wei, X., Das, J., Fragoza, R., Liang, J., de Oliveira, F. M. B., Lee, H. R., ... & Lipkin, S. M. (2014). A massively parallel pipeline to clone DNA variants and examine molecular phenotypes of human disease mutations. *PLoS Genet*, *10* (12), e1004819.
- Weise, J. B., Rudolph, P., Heiser, A., Kruse, M. L., Hedderich, J., Cordes, C., ... & Görögh, T. (2008). *LOXLA* is a selectively expressed candidate diagnostic antigen in head and neck cancer. *European Journal of Cancer*, *44* (9), 1323-1331.
- Weren, R. D., Ligtenberg, M. J., Kets, C. M., De Voer, R. M., Verwiel, E. T., Spruijt, L., ... & Hoischen, A. (2015). A germline homozygous mutation in the base-excision repair gene *NTHL1* causes adenomatous polyposis and colorectal cancer. *Nature genetics*, *47* (6), 668-671.
- Wollowski, I., Rechkemmer, G., & Pool-Zobel, B. L. (2001). Protective role of probiotics and prebiotics in colon cancer. *The American Journal of Clinical Nutrition*, *73* (2), 451s-455s.
- Wu, G., Guo, Z., Chang, X., Kim, M. S., Nagpal, J. K., Liu, J., ... & Sidransky, D. (2007). *LOXL1* and *LOXLA* are epigenetically silenced and can inhibit ras/extracellular signal-regulated kinase signaling pathway in human bladder cancer. *Cancer Research*, *67* (9), 4123-4129.
- Wu, D., & Pan, W. (2010). *GSK3*: a multifaceted kinase in Wnt signaling. *Trends in Biochemical Sciences*, *35* (3), 161-168.
- Xu, C., & Min, J. (2011). Structure and function of WD40 domain proteins. *Protein & Cell*, *2* (3), 202-214.