

**INSTITUTO POLITÉCNICO DE LISBOA
ESCOLA SUPERIOR DE TECNOLOGIA DA SAÚDE DE LISBOA**

Deep intronic mutations in colorretal cancer

Mariana Jesus Pimpão Luís

Orientadores

Professora Doutora Ana Marques Ramos

Mestrado em Tecnologias Clínico Laboratoriais

Lisboa, 2021

**INSTITUTO POLITÉCNICO DE LISBOA
ESCOLA SUPERIOR DE TECNOLOGIA DA SAÚDE DE LISBOA**

Deep intronic mutations in colorretal cancer

Mariana Jesus Pimpão Luís

Orientadores

Professora Doutora Ana Marques Ramos

Mestrado em Tecnologias Clínico Laboratoriais

Lisboa, 2021

Dedicatória

À minha família.

Agradecimentos

Em primeiro lugar quero agradecer à Escola Superior de Tecnologias da Saúde de Lisboa, por ser a instituição que me forneceu todos os conhecimentos por mim adquiridos ao longo do meu percurso académico onde, cresci enquanto pessoa bem como futura cientista.

Quero agradecer a todas as pessoas que de alguma forma me ajudaram na realização da minha tese em especial, um agradecimento á professora doutora Ana Ramos por ter aceitado este meu desafio, por ser a minha orientadora de tese, por toda a disponibilidade, por todo o auxílio prestado, por toda a paciência bem como todos os conhecimentos que me transmitiu.

Agradeço também ao Pedro Barbosa, doutorando do Instituto de Medicina Molecular por todo o apoio prestado ao longo da tese, por tudo o que me transmitiu, por toda a disponibilidade que teve para me ensinar e para me ajudar na realização do meu trabalho.

Os últimos são sempre os primeiros por isso agradeço especialmente à minha mãe, que é tudo na minha vida, ao meu pai, à minha irmã, à minha tia, aos meus avós e ao meu namorado, por todo o apoio que me deram durante estes anos de licenciatura, pelo apoio que me deram nos momentos menos bons, mostrando sempre que nunca devia de desistir daquilo que acreditava e de aquilo que me fazia feliz, tenho plena consciência que este percurso só é possível devido a eles.

Resumo

A sequência genómica esteve, durante muitos anos, apenas associada à sua capacidade de codificar proteínas fazendo com que as sequências intrónicas fossem consideradas não funcionais. Os genes são compostos por exões interrompidos por intrões. Os intrões são sequências nucleotídicas não codificantes não sendo, portanto, traduzidos, uma vez que são removidos num processo anterior, denominado *splicing*. Foram identificadas diversas variantes de nucleotídeos localizadas profundamente nos intrões com associação significativa a doenças. Variantes intrónicas profundas são mutações que ocorrem a mais de 100 pares de bases dos limites exão-intrão.

O cancro representa a primeira causa de morte nos países desenvolvidos e a segunda causa de morte nos países em desenvolvimento. Em 2019 o cancro colorretal foi a sexta causa de morte em Portugal. Esta neoplasia é caracterizada por ser uma doença do genoma que evolui e progride por um acumular de mutações somáticas, e alterações epigenéticas com ou sem hereditariedade.

Assim, o objetivo deste trabalho é realizar uma análise exaustiva de mutações intrónicas profundas já identificadas e prever o efeito funcional dessas mutações através de uma análise *in silico*. Através do programa bioinformático, foi possível obter dados e dividi-los em duas tabelas. A primeira tabela consiste nas variantes e genes já reportados clinicamente e a segunda tabela contém variantes que ainda não foram reportadas clinicamente. Existem várias variantes envolvidas no CRR e seria importante começar a estudar essas mutações, especificamente para este tipo de cancro, de modo a perceber se poderão, por exemplo, ser um biomarcador deste cancro.

Palavras-Chave: Genética; Variantes; Cancro Colorretal; Intrões profundos.

Abstract

The genomic sequence was, for many years, only associated with its ability to encode proteins causing the intronic sequences to be considered non-functional. Genes are made of exons interrupted by introns. Introns are non-coding nucleotide sequences and are therefore not translated, as they are removed in an earlier process called *splicing*. Several nucleotide variants located deep in introns with significant disease association have been identified. Deep intronic variants are mutations that occur more than 100 base pairs from the exon-intron boundaries.

Cancer represents the leading cause of death in developed countries and the second leading cause of death in developing countries. In 2019 colorectal cancer was the sixth leading cause of death in Portugal. This neoplasm is characterized for being a disease of the genome that evolves and progresses through accumulation of somatic mutations, and epigenetic alterations with or without heredity.

Thus, the purpose of this study is to carry out an exhaustive analysis of already identified deep intronic mutations and to predict the functional effect of these mutations through an *in silico* analysis. Through a bioinformatics program, it was possible to obtain data and divide them into two tables. The first table consists of the variants and genes that have already been clinically reported and the second table contains variants that have not yet been clinically reported. There are several variants involved in CRR and it would be important to start studying these mutations, specifically for this type of cancer, in order to understand if they could, for example, be a biomarker of this cancer

keywords: Genetics; Variants; Colorectal Cancer; Deep Introns;

Índice

1. Introdução.....	1
1.1. Contextualização	1
1.2. Objetivo Geral e Objetivos Específicos	2
1.2.1. Objetivo geral	2
1.2.2. Objetivos específicos.....	2
2.Enquadramento Teórico – O Carcinoma do cólon e do reto.....	3
2.1. Cancro	3
2.2. Cancro colorretal	5
2.2.1. Epidemiologia.....	5
2.2.2. Anatomia e Fisiopatologia.....	7
2.2.3. Estadio da Doença	10
2.2.4. Sintomas	12
2.2.5. Fatores de risco.....	12
2.2.6. Rastreio e Diagnóstico.....	14
2.3. Genética vs. Cancro Colorretal.....	17
3. Metodologia	21
3.1. Desenho do Estudo	21
3.2. Questão de Partida e Hipóteses de Investigação	21
3.2.1. Questão de partida	21
3.2.2. Hipóteses de investigação.....	21
3.3. Recolha dos dados.....	21
3.4. Análise <i>in sílico</i>	22
4. Resultados.....	23
5. Discussão de resultados.....	25
6. Conclusões e Perspetivas Futuras.....	35
6.1. Conclusões do presente estudo.....	35



6.2. Perspetivas futuras	35
7. Referências bibliográficas	36

Índice de Figuras

Figura 2.1- Esquema proposto por Hanahan e Weinberg atualizado com as quatro novas referências, duas de emergência e outras duas de referência, apresentadas posteriormente. (6).	5
Figura 2.2 - Número estimado de novos casos em 2020 em todo o mundo de homens (A) e Mulheres (B) em todas as idades.	6
Figura 2.3- Número estimado de novos casos em 2020, em todo o mundo, ambos os sexos em todas as idades	6
Figura 2.4- Número estimado de incidências em homens (A) e Mulheres (B) e o número de mortes em homens(C) e mulheres (D), em 2020, em Portugal; todas as idades.....	7
Figura 2.5- Representação do final do trato digestivo, o cólon e reto	8
Figura 2.6- Processo de formação do tumor, invasão dos tecidos e metastização.....	10
Figura 3.1- Página do Nacional Cancer Institute com as condições de pesquisa selecionadas.....	22
Figura 5.1- Resultados obtidos pela ferramenta bioinformática utilizada para a previsão de variantes e possíveis efeitos.	25
Figura 5.2 Codificação da proteína normal á esquerda e codificação da proteína mutada á direita.....	34

Índice de Tabelas

Tabela 2.1- Estádios do cancro colorretal e correspondente taxa de sobrevivência. ..	11
Tabela 4.1- Resultados obtidos pela ferramenta bioinformática utilizada para a previsão de variantes e possíveis efeitos.	23
Tabela 4.2- Resultados obtidos através da ferramenta bioinformática. São apresentadas abaixo apenas as variantes sem estudos realizados sobre as mesmas e o efeito previsto pelo programa bioinformático.....	24
Tabela 4.3- Resultados obtidos através da ferramenta bioinformática. São apresentadas abaixo apenas as variantes já estudadas, o efeito previsto pelo programa bioinformático e o que se encontra descrito sobre as mesmas na literatura	24



Lista de abreviaturas

CCR- Cancro do cólon e do reto;

ARN- Ácido Ribonucleico;

OMS- Organização Mundial de saúde;

PAC- Polipose adenomatosa coli,

MSH2- *MutS Homolog 2;*

MLH1- *mutL homolog 1;*

ADN- Ácido desoxirribonucleico;

SGI- Sistema gastrointestinal;

ARNm- Ácido ribonucleico mensageiro;

DGS- Direção Geral da Saúde;

TNM- (T) Tumor primário, (N) linfonodos, (M) metástase;

INC- Instabilidade cromossomal;

MSI- Microsatélites;

miARNs- Micro ARNs;

snoARNs- Subprodutos funcionais não codificantes de ARN nucleolar;

CCRHSP- Cancro colorretal hereditário sem polipose;

PAF- Polipose Adenomatosa familiar;

CCHNP- Cancro colorretal hereditário sem polipose;

1. Introdução

1.1. Contextualização

A sequência genómica esteve, durante muitos anos, apenas associada à sua capacidade de codificar proteínas fazendo com que as sequências intrónicas fossem consideradas não funcionais. Os genes são compostos por exões interrompidos por intrões. Os intrões são sequências nucleotídicas não codificantes não sendo, portanto, traduzidos, uma vez que são removidos num processo anterior, denominado *splicing*. No entanto, evidências recentes defendem a funcionalidade dos intrões. Foram identificadas diversas variantes de nucleotídeos localizadas profundamente nos intrões com associação significativa a doenças. Variantes intrónicas profundas são mutações que ocorrem a mais de 100 pares de bases dos limites exão-intrão. Estas descobertas estão a promover uma nova Era de pesquisa focada na compreensão de como a variação na sequência intrónica profunda afeta e contribui para os fenótipos da doença, neste caso no CCR. Atualmente existem genes na qual é conhecido que a sua mutação está associada ao desenvolvimento do cancro colorretal tais como o polipose adenomatosa coli (PAC), *MutS Homolog 2 (MSH2)* e *mutL homolog 1 (MLH1)*, mas desconhece-se qual a frequência de mutações intrónicas profundas nesses e noutros genes envolvidos na carcinogénese do CCR.

O cancro representa a primeira causa de morte nos países desenvolvidos e a segunda causa de morte nos países em desenvolvimento. O cancro do cólon e reto (CCR) é considerado atualmente um problema de saúde pública a nível mundial. Em 2019 este tipo de cancro foi a sexta causa de morte em Portugal. Esta neoplasia é caracterizada por ser uma doença do genoma que evolui e progride por um acumular de mutações somáticas, e alterações epigenéticas com ou sem hereditariedade.

Assim, o objetivo deste trabalho é realizar uma análise exaustiva de mutações intrónicas profundas já identificadas e prever o efeito funcional dessas mutações através de uma análise in sílico.

Este estudo é importante porque existem evidências experimentais onde sugerem que alterações em regiões não codificantes em ARNs codificantes reguladores de proteínas são relevantes na tumorigenese e esta descoberta pode ajudar em potenciais diagnósticos, prognósticos e eficácia de tratamentos quimioterápicos, ou mesmo como agentes terapêuticos.

1.2. Objetivo Geral e Objetivos Específicos

1.2.1. Objetivo geral

O objetivo geral deste estudo será a realização de uma análise exaustiva de mutações intrónicas profundas já identificadas e prever o efeito funcional dessas mutações *in silico* em cancro colorretal.

1.2.2. Objetivos específicos

Os objetivos específicos são realizar uma pesquisa e análise exaustiva de mutações intrónicas profundas em cancro colorretal numa base de dados que será a *Cancer Nacional Institute*; prever o efeito funcional das mutações através de um programa bioinformático; criar uma lista de mutações intrónicas profundas reportadas e estudadas até á atualidade e criar uma outra lista com mutações que ainda não tenham sido estudadas e por fim realizar uma análise das mutações, e dos seus efeitos relativamente ao CCR.

2. Enquadramento Teórico – O Carcinoma do cólon e do reto

2.1. Cancro

A palavra cancro tem origem no latim “*Karinoa*” e tem como significado caranguejo(1). Foi usada por Hipócrates na antiga Grécia, para indicar o aspeto de uma massa central de onde partem prolongamentos que invadem tecidos adjacentes como se fossem as patas de um caranguejo (1). Atualmente a palavra cancro é o termo utilizado para designar um acontecimento maligno de células decorrentes da degeneração patológica das próprias células do corpo(2).

Segundo a Organização Mundial de Saúde (OMS), o cancro é definido como um crescimento descontrolado de células com posterior disseminação das mesmas, podendo afetar quase todo o organismo(3). O desenvolvimento do cancro pode levar à invasão dos tecidos adjacentes, assim como ocorrer metastização em outras partes do corpo mais distantes (3), podendo ocorrer em qualquer lugar do corpo (4).

Os cancros com predominância nos homens são o do pulmão, próstata, colorretal, estômago e fígado enquanto nas mulheres são o cancro da mama, colorretal, pulmão, cervical e tireoide(5).

O cancro é uma consequência de alterações em genes que controlam a função celular, em particular o crescimento e a divisão da célula. Este requer estratégias diferentes de tratamento dependendo da sua localização e disseminação(5). Por outras palavras, os tumores são caracterizados por proliferação desregulada e por fenómenos de invasão e metastização(5). Assim é possível diferenciar um tumor maligno de um benigno, uma vez que os tumores benignos não metastizam, enquanto os tumores malignos, por outro lado, não permanecem no seu local de origem e, apresentam características de invasão e metastização, ameaçando assim o bem-estar e a própria vida do hospedeiro (5). Estas células tem a capacidade de ignorar sinais para pararem de se dividir ou para iniciar um processo conhecido como morte celular ou apoptose. As células cancerígenas conseguem induzir as células normais próximas a produzir vasos sanguíneos, este fenómeno é denominado de angiogénese (3,4).

As alterações que ocorrem no ADN celular podem ocorrer devido a transmissão hereditária ou a exposição ambiental a agentes que danificam o ADN, tais como o fumo do tabaco ou o stress. Estas mutações podem contribuir para o cancro quando ocorrem em quatro tipos de genes: proto-oncogenes, genes supressores de tumores, genes reguladores da apoptose ou genes responsáveis pela reparação do ADN (6).

Em 2000, Douglas Hanahan e Robert A. Weinberg, definiram seis capacidades biológicas – “*hallmarks of cancer*” - que são adquiridas pelas células cancerígenas em diferentes pontos do desenvolvimento do tumor e incluem: autonomia da sinalização proliferativa, evasão de supressores de crescimento, invasão e metástase, potencial replicativo ilimitado, angiogénese e resistência à morte celular(6). Alguns anos mais tarde, os mesmos autores revisitaram este conceito e novas características das células tumorais foram acrescentados à lista inicial(6). Assim, como novos “*hallmarks of cancer*”, surge os processos inflamatórios como fonte promotora de tumores, a capacidade das células tumorais evitarem a destruição por parte do sistema imunitário do hospedeiro, reprogramação do metabolismo energético e a instabilidade e mutabilidade do genoma(6).

Relativamente à instabilidade e mutabilidade do genoma, estes ocorrem devido a uma sucessão de alterações no genoma das células neoplásicas que, conferem vantagens seletivas nas células que permitem o seu crescimento e eventual domínio no ambiente em que se encontra(6). A capacidade dos sistemas de manutenção do genoma para detetar e resolver defeitos do ADN, garante que as taxas de mutação espontâneas são geralmente muito baixas. Existe uma variedade de efeitos que afetam vários componentes do sistema de manutenção de ADN(6).

Os avanços na análise molecular-genética dos genomas das células neoplásicas demonstraram mutações que alteram a função e a instabilidade genómica continua durante a progressão do tumor(6). Embora as especificidades da alteração do genoma variem entre os vários tipos de cancro, podemos concluir que os defeitos de manutenção e reparação do genoma que já foram documentados são seletivamente vantajosos e levam à progressão do tumor. Estes defeitos de reparação e manutenção do genoma são inerentes à grande maioria das células neoplásicas(6).

Na Figura 2.1, podemos observar as seis características de referência das células tumorais inicialmente propostas por Hanahan e Weinberg, em 2000, e as restantes quatro “*hallmarks of cancer*” apresentadas posteriormente (6).

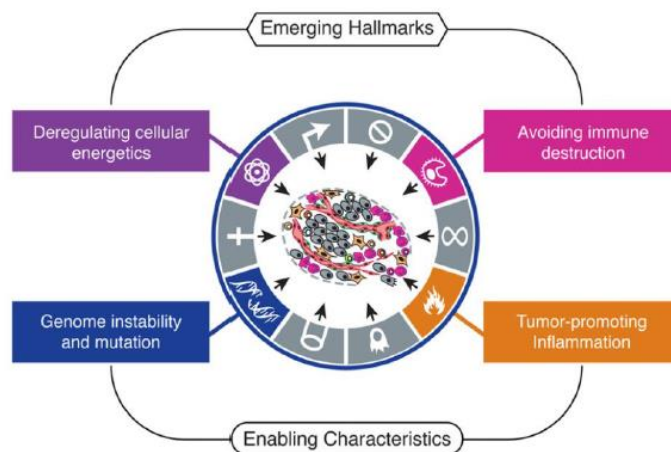


Figura 2.1- Esquema proposto por Hanahan e Weinberg atualizado com as quatro novas referências, duas de emergência e outras duas de referência, apresentadas posteriormente. (6).

Existem mais de 100 tipos de cancros que são, normalmente, denominados de acordo com o órgão ou tecido em que se formaram. Assim, o cancro que tem início no cólon, chama-se “cancro do cólon” e o cancro que tem início no reto, chama-se “cancro retal”. Contudo, ambos os cancros mencionados anteriormente podem ser denominados de cancro colorretal (7).

2.2. Cancro colorretal

O cancro, como referido anteriormente é uma doença caracterizada pela divisão e disseminação não controlada de células anormais. O cólon e reto (*colorectum*), juntamente com o ânus, compõem o intestino grosso, o segmento final do sistema gastrointestinal (SGI)(8,9).

2.2.1. Epidemiologia

No mundo

O cancro colorretal CCR é o terceiro cancro com maior incidência no sexo masculino correspondendo a uma taxa de 10,6% da totalidade de casos de cancro (Figura 2.2A), já no sexo feminino, é a segunda forma de cancro em termos de incidência, apresentando uma taxa de 9,4% (Figura 2.2.B). O CCR é terceiro cancro mais comuns em todo o mundo com 19 292 789 casos estimados em 2020 (Figura 2.3) (10,11).

O CCR era bastante raro, mas tornou-se um cancro predominante nos países desenvolvidos em comparação com os países em vias de desenvolvimento. Este aumento está relacionado com o aumento do índice de envelhecimento da população, os maus hábitos alimentares, tabagismo, baixa atividade física e obesidade(12). No entanto, é de realçar que a mortalidade têm uma taxa muito maior nos países em vias de desenvolvimento em comparação com os países desenvolvidos, refletindo a baixa sobrevivência em regiões menos desenvolvidas e menos acesso a cuidados de saúde(8). Na maioria das populações, o risco médio do desenvolvimento de CCR, ao longo da vida, está num intervalo de 3–5%. O CCR apresenta uma incidência e mortalidade que aumentam com a idade do indivíduo e são maiores para o sexo masculino, o que parece refletir as diferenças na exposição a hormonas e a fatores de risco(8,12).

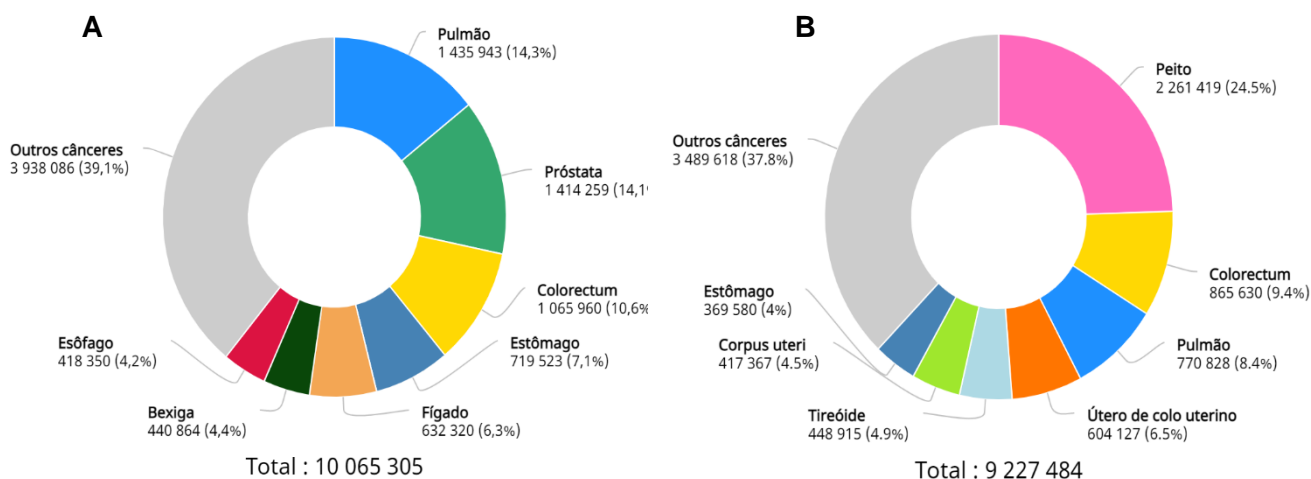


Figura.2.2 - Número estimado de novos casos em 2020 em todo o mundo de homens (A) e Mulheres (B) em todas as idades.

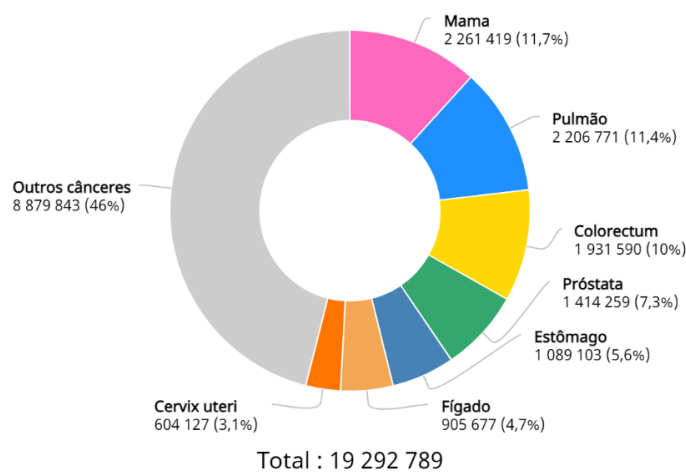


Figura 2.3- Número estimado de novos casos em 2020, em todo o mundo, ambos os sexos em todas as idades

Em Portugal

O CCR é o terceiro cancro com maior incidência no sexo masculino correspondendo a uma taxa de 19% da totalidade de casos de cancro em 2020 (Figura 2.4 A)(8). No sexo feminino, é a segunda forma de cancro em termos de incidência, apresentando uma taxa de 15.3% (Figura 2.4 B). Relativamente á taxa de mortalidade, nos homens em 2020 houve cerca de 14,1% (Figura 2.4C) de mortalidade enquanto nas mulheres foi de 9,5% (Figura 2.4D)(8).

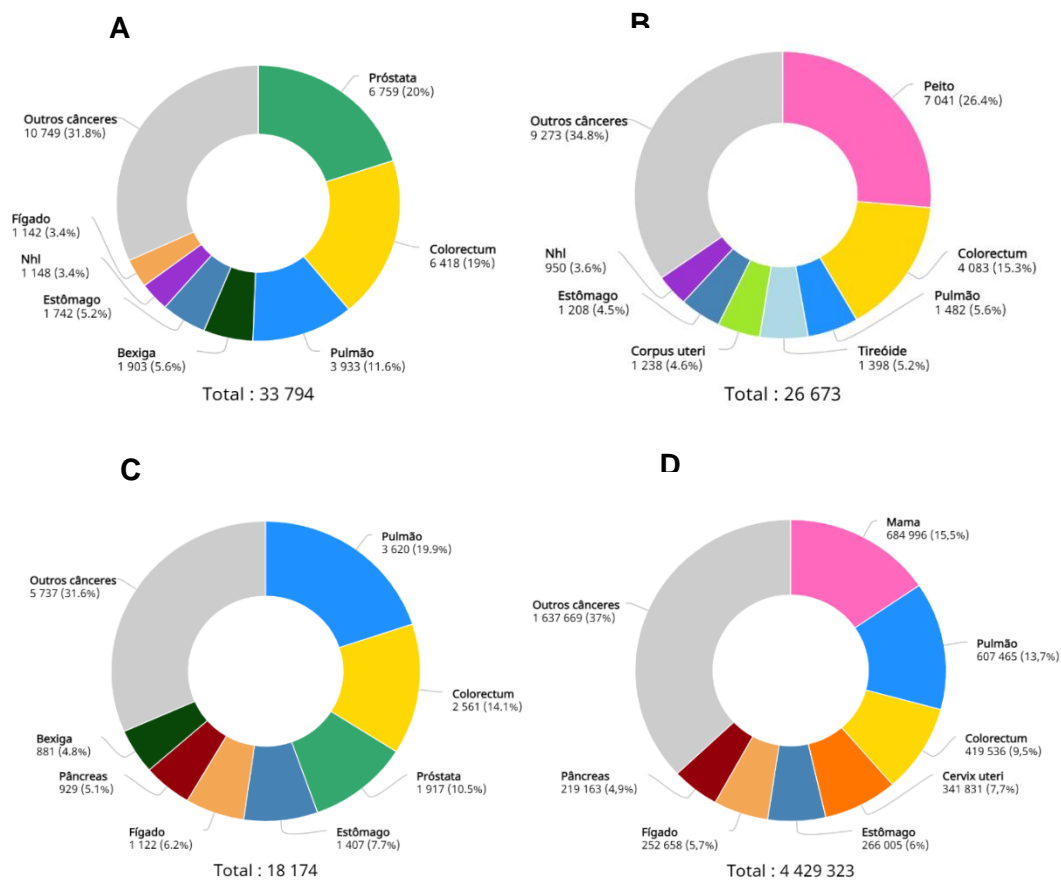


Figura 2.4- Número estimado de incidências em homens (A) e Mulheres (B) e o número de mortes em homens(C) e mulheres (D), em 2020, em Portugal; todas as idades.

2.2.2. Anatomia e Fisiopatologia

Anatomia e Fisiologia

Conhecer a fisiologia de uma doença é essencial para a compreender e poder atuar no tratamento da mesma(9).

O cólon e o reto são elementos do intestino grosso, e fazem parte do aparelho digestivo, participando na absorção de nutrientes dos alimentos digeridos provenientes do intestino delgado e expulsão do desperdício do organismo(9).

O processo digestivo inicia quando o alimento é colocado na boca. Os nutrientes são absorvidos no intestino delgado e no duodeno. No cólon ou intestino grosso, existem microrganismos que constituem a chamada flora intestinal que decompõem certas substâncias por fermentação, tornando-as mais fáceis de processar(1,2). As substâncias que não são decompostas por enzimas digestivas no intestino delgado e, não estão sujeitas à fermentação por microrganismos no cólon, são excretadas inalteradas pelo reto. O cólon e o reto não são órgãos digestivos essenciais para um processo de digestão(1,2).

O cólon, e o reto estão localizados no final do trato digestivo. O cólon tem cerca de 1,5 metros de comprimento e está dividido em quatro partes. A primeira é o cólon ascendente começa na parte superior do ceco e sobe pelo lado direito até à superfície inferior do fígado, onde curva para a esquerda(1,2). De seguida encontra-se o cólon transversal que tem cerca de 40 a 45cm de comprimento, sendo o segmento mais largo e de maior mobilidade do intestino grosso forma um arco de comprimento variável da direita para a esquerda terminando na zona inferior ao baço designada de região hipogástrica. É nesta zona que se inicia o cólon descendente, com cerca de 20 a 25cm de comprimento e é considerada a região mais estreita do cólon. Esta desce ao longo do lado esquerdo até à zona pélvica. Nesta zona existe uma união ao cólon sigmoide, que na sua parte inferior se une ao reto(1,2). O reto tem cerca de 15 a 16cm e é dividido em três segmentos sendo eles o terço superior, o terço médio e o terço inferior. Esta divisão em terços é de relevância clínica uma vez que o tratamento dos tumores difere de acordo com o terço do reto em que eles estão localizados e é a zona que se une ao ânus através do canal anal (1,2).

Além disso, é constituído por Cinco camadas: do interior para o exterior, a mucosa, a muscular da mucosa, a submucosa, a muscular, a sub-serosa e a serosa (Figura 2.5)

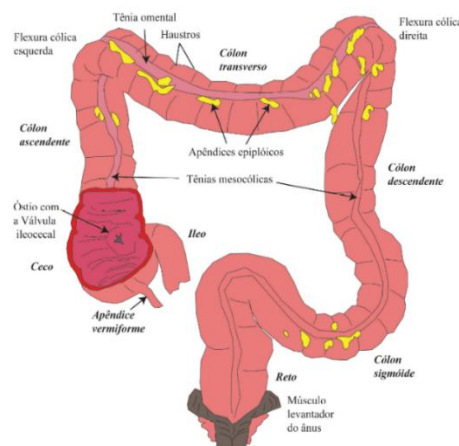


Figura 2.5- Representação do final do trato digestivo, o cólon e reto

Patologia

O CCR é uma neoplasia localizada no cólon ou no reto. A perda da estabilidade genómica foi observada na maioria das lesões neoplásicas iniciais no cólon e é provavelmente um evento molecular e fisiopatológico central na iniciação e formação de CCR. A perda de estabilidade acelera a acumulação de mutações e alterações epigenéticas em proto oncogenes, genes que regulam a apoptose ou a regeneração do ADN ou genes supressores de tumor, que levam á formação de células malignas do cólon selecionando as células mais agressivas e malignas (4,12,13).

Os protos oncogenes estão envolvidos no crescimento e divisão celular e quando estes sofrem mutações, podem alterar o seu funcionamento, tornando-os mais ativos que o normal. Desta forma, os protos oncogenes são convertidos em oncogenes, estimulando o crescimento e a sobrevivência das células, mesmo quando estas deviam sofrer apoptose e ser eliminadas pelo sistema imunitário(4).

Os genes supressores de tumores também estão envolvidos no controlo do crescimento e divisão das células. Quando ocorrem certas mutações nestes genes, pode ocorrer uma inativação dos mesmos, levando a uma divisão descontrolada de células(4).

Os genes de reparação de ADN estão envolvidos, como o nome indica, na reparação dos danos que ocorrem no genoma. Caso adquira mutação nestes genes, isto pode levar a que os outros tenham uma menor probabilidade de reparar lesões no material genético, aumentando a probabilidade do surgimento de células cancerígenas(4).

Os genes reguladores da apoptose estão envolvidos na morte celular programada. Quando ocorre uma mutação num destes genes, a morte celular não é bem conduzida, levando a que células anormais continuem a sua divisão. Tal vai originar uma anomalia exponencial com possibilidade de formação de neoplasia(4).

A carcinogénese que é o processo onde uma célula mutada pode multiplicar-se e originar um crescimento de células com autonomia e que não respondem á sinalização de crescimento e a mecanismos de apoptose, adquirindo a capacidade de proliferar ilimitadamente e de invadir outros órgãos e tecidos, ou seja, metastizar-se(14). A metastização ocorre quando as células cancerígenas circulam para fora do tumor inicial, neste caso, do cólon ou do reto, através do sistema linfático ou circulatório (9).

As células cancerígenas atravessam as cinco camadas do cólon para depois invadir outros tecidos através do sistema circulatório. Quando ocorre metastização, as células são usualmente encontradas nos gânglios linfáticos. Se as células cancerígenas já tiverem atingido estes gânglios, é provável que também se tenham disseminado para outros gânglios linfáticos, ou mesmo para outros órgãos(9).

Alguns estudos mostram que 70% dos CCR desenvolvem-se de pólipos adenomatosos esporádicos(14). Os pólipos são formações que se desenvolvem na mucosa do intestino, mais frequentemente no cólon. Existem vários tipos de pólipos, a maioria dos quais é benigna(15). No entanto, o pólipo adenomatoso, que corresponde a cerca de dois terços de todos os pólipos, encontra-se associado a mutações do ADN das células que o constituem, formando uma situação de displasia, o que pode facilmente evoluir para carcinoma. Os adenomas planos, que correspondem a cerca de 10% de todos os pólipos, são os mais difíceis de detetar na endoscopia, e podem ter uma elevada taxa de se tornarem malignos ou podem predispor para um fenótipo cancerígeno mais agressivo (15).

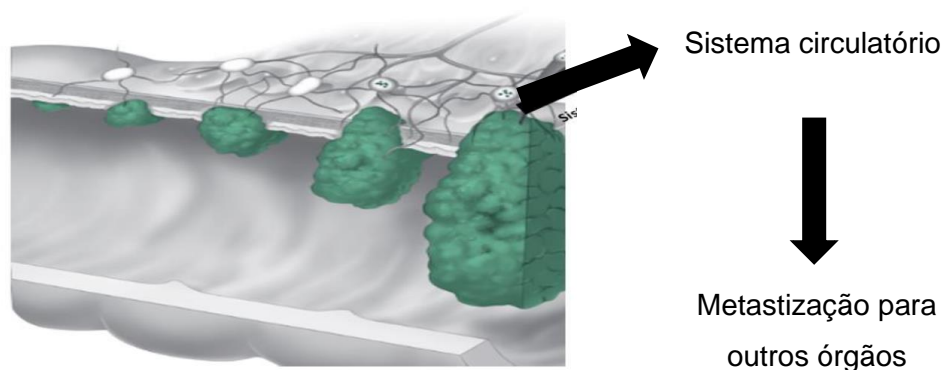


Figura 2.6- Processo de formação do tumor, invasão dos tecidos e metastização

2.2.3. Estadio da Doença

O estadio da doença é o processo pelo qual são avaliadas se as células disseminaram para outros locais do corpo ou tecidos vizinhos. É a inclusão de um cancro numa de várias fases de evolução previamente definidas. O cancro é diagnosticado e é necessário saber em que estadio se encontra o cancro para que seja feito um tratamento adequado(16). Para determinar o estadio da doença devem ser realizados alguns procedimentos tais como:

- ❖ Colonoscopia: caso a mesma não tenha sido efetuada na fase de diagnóstico; é executada para detetar sinais de outro(s) cancro(s) em todo o órgão;
- ❖ Tomografia computadorizada do tórax e abdómen: permite verificar se o cancro metastizou para o fígado, os pulmões ou outros órgãos.
- ❖ Ressonância Magnética do reto: permite avaliar a profundidade do cancro nas diferentes camadas da parede do reto, e se os gânglios linfáticos na vizinhança do reto já foram, ou não, invadidos pelo cancro.

- ❖ ecografia endorectal: consiste na introdução de uma sonda pelo ânus e permite, se necessário, complementar a informação da RM do reto: avaliar a profundidade do cancro nas diferentes camadas da parede do reto, e o estado dos gânglios linfáticos loco-regionais(7,17).

Apesar dos testes disponíveis, muitas vezes o estadiamento só é determinado de uma forma inequívoca após a cirurgia para remoção do tumor(17).

De acordo com os exames de diagnóstico referidos, o estadiamento pode classificar o cancro CCR em vários estágios, sendo eles:

- ❖ Estádio 0: o tumor encontra-se, apenas, no revestimento interior do cólon e do reto. Pode ser também denominado carcinoma *in situ*. (7,17)
- ❖ Estádio I: o tumor desenvolveu-se para dentro da parede do cólon e do reto, ou seja, o cancro invadiu a submucosa ou a muscular. No entanto, ainda não se espalhou para o exterior do cólon. (7,17)
- ❖ Estádio II: o tumor desenvolveu-se mais profundamente na parede do cólon e do reto. Pode ter invadido tecidos vizinhos, no entanto, as células cancerígenas não chegaram aos gânglios linfáticos. (7,17)
- ❖ Estádio III: o tumor já atingiu os gânglios linfáticos vizinhos, mas não chegou a outras partes do corpo. (7,17)
- ❖ Estádio IV: o tumor já apresenta metástases noutras partes do corpo(7,17).

O CCR pode ser classificado com um método baseado no sistema TNM, proposto pela *American Joint Committee on Cancer* (18).

A taxa de sobrevivência do CCR depende, do estadio em que o tumor se encontra. Assim, a probabilidade de sobrevivência num indivíduo com CCR no estágio IV vai ser muito menor que a probabilidade de um indivíduo no estágio I (Tabela) (18).

Tabela 2.1- Estádios do cancro colorretal e correspondente taxa de sobrevivência.

Estádio	Taxa de sobrevivência
Estádio 0	100%
Estádio I	90-100%
Estádio II	75-85%
Estádio III	30-40%
Estádio IV	<5%

2.2.4. Sintomas

Existem vários sintomas associados ao CCR que não são específicos deste tipo de cancro e que quando presentes, podem ser indicadores de outras doenças como tumores benignos, por isso, o médico representa um papel essencial no diagnóstico da doença. O CCR precoce muitas vezes não tem sintomas, o que é uma das razões para o rastreio ser tão importante. À medida que um tumor cresce, pode sangrar ou bloquear o intestino (8,16,19).

Em estádios mais avançados os sintomas apresentam-se geralmente como: Alteração de hábitos intestinais, como obstipação, diarreia ou, mais frequentemente, uma alternância das duas (8,16,19);

- ❖ Sensação de que o intestino não esvazia completamente;
- ❖ Perda de sangue pelo ânus, misturado ou não com as fezes e de cor vermelho vivo, castanho-escuro ou mesmo preto, que pode levar a anemia;
- ❖ Fezes menores que o habitual;
- ❖ Dor abdominal, constante ou em cólica;
- ❖ Emissão de muco pelo ânus misturado ou não com as fezes
- ❖ Perda de peso inexplicada;
- ❖ Cansaço constante;
- ❖ Náuseas e vômitos(8,16,19).

A avaliação oportuna dos sintomas consistentes com CCR é essencial para todos os indivíduos, independentemente da idade, dada a incidência crescente em adultos jovens(8).

2.2.5. Fatores de risco

Os fatores de risco é qualquer comportamento ou condição que aumente o risco de ter uma doença(7).

O aumento da incidência de CCR é considerado um marcador de mudança económica. Vários estudos mostraram que pessoas com hábitos de vida saudáveis apresentam risco 27% a 52% menor de CCR comparando com aquelas que não têm esses comportamentos. Os fatores de risco hereditários ou historial médico pessoal ou familiar de CCR ou doenças benignas por exemplo pólipos, têm um risco aumentado e devem começar um rastreio de CCR antes dos 45 anos(8,14).

Conhecer os fatores de risco pode ser uma forma de prevenir a doença, uma vez que os fatores comportamentais podem ser evitados. Alguns dos riscos associados ao CCR são:

- ❖ **Idade:** o risco de CCR aumenta com a idade. Aliás, é conhecido que 90% dos casos e 93% das mortes ocorrem em pessoas com mais de 50 anos. A idade média de aparecimento desta neoplasia é entre os 65 e os 72 anos (7,8).
- ❖ **Etnia:** as taxas da presença deste cancro são maiores em pessoas de etnia africanas e mais baixas em pessoas de etnia asiática. A incidência em indivíduos de etnia africana foi 25% maior que nos de etnia caucasiana e 50% maior que a asiática (8).
- ❖ **Sedentarismo:** um dos principais fatores de risco comportamentais é a inatividade física. As pessoas ativas têm 25% menos risco de sofrer desta patologia e a taxa de sobrevivência é maior (8).
- ❖ **Dieta:** estudos sugerem que uma dieta pobre em frutas e vegetais, cálcio, fibra, cálcio e folatos mas rica em gorduras, especialmente gordura animal, carboidratos, açúcar processado, pode aumentar o risco de CCR. Sugerem ainda que pessoas com uma dieta muito pobre em fruta e vegetais, podem ter risco aumentado para esta doença (7,8,20).
- ❖ **Medicação:** existe a evidência de que os anti-inflamatórios não esteroides reduzem o risco do aparecimento do CCR, no entanto, não é aconselhado a toma como prevenção, uma vez que estes induzem hemorragias. Além disso, também foi demonstrado que a medicação hormonal para a pós-menopausa e que os contraceptivos diminuem o risco do aparecimento da doença(8).
- ❖ **Tabagismo:** uma pessoa fumadora, pode ter risco aumentado de desenvolver pólipos e CCR. É atribuído ao tabagismo ativo ou anterior, um risco de CCR cerca de 50% maior do que não fumadores. O tabagismo também está associado a uma menor sobrevivência ao CCR, particularmente para os fumantes ativos(8,20).
- ❖ **Álcool:** O consumo de álcool leve a moderado (até duas bebidas por dia) foi associado a um risco ligeiramente menor (8%) do que nenhum consumo / consumo ocasional, enquanto beber em excesso foi associado a um risco 25% maior. O homens têm o risco mais elevado do que as mulheres devido aos hábitos de alcoolismo ou às diferenças hormonais no metabolismo do álcool (8).
- ❖ **Doença de Crohn ou colite ulcerosa:** são doenças crónicas que provocam a inflamação do cólon, o que leva a um risco aumentado de desenvolver CCR. É estimado que 18% dos doentes com colite ulcerosa durante 30 anos desenvolvem esta neoplasia (8,20).
- ❖ **História Familiar e hereditária:** Doentes com um familiar de primeiro grau com história de CCR têm uma probabilidade duas a três vezes maior de também desenvolver a doença, especialmente se o familiar teve a doença ainda jovem

(8,20). Caso um indivíduo tenha mais do que um familiar direto com a patologia, o risco aumenta três a seis vezes quando comparado com o resto da população. No entanto, a taxa de sobrevivência nestes casos é maior, uma vez que a atenção com a doença é maior e o diagnóstico precoce é maior (8,20). Um estudo recente descobriu que 5% dos pacientes com CCR têm uma mutação da linha germinativa herdada associada a uma condição hereditária de alto risco conhecida, e outros 5% têm mutações associadas a risco moderadamente elevado. Cerca de 30% dos pacientes com CCR têm história familiar da doença, tornando este um dos fatores de risco mais importantes(8,21). As síndromes hereditárias que predispõem para o CCR são a Polipose Adenomatosa Familiar (PAF) onde pessoas com esta condição têm uma mutação ou perda do gene PAC, que leva ao aparecimento de centenas ou milhares de pólipos no intestino grosso em idade jovem(8,21). O A síndrome de Lynch, ou CCR Hereditário sem Polipose (HNPCC) em que pessoas com esta condição têm certas mutações nos genes que provocam uma falha nos mecanismos de reparação do ADN. A consequência é que um tumor colorretal benigno pode evoluir para cancro a um ritmo mais rápido (em média, 2 a 3 anos) do que nos indivíduos que não tenham a síndrome de Lynch (8,21).

- ❖ **Alterações genéticas:** certas mutações em genes como os proto oncogenes ou genes supressores de tumores aumentam o risco do aparecimento da patologia (20).
- ❖ **Pólipos colorretais:** os pólipos são comuns em pessoas com idade superior a 50 anos. A maioria dos pólipos é benigna; no entanto, alguns pólipos, como os adenomas, podem tornar-se cancerígenos; detetar e remover os pólipos é um modo de prevenção do CCR (20).

Existem fatores que podem ajudar contra o desenvolvimento do CCR como uma dieta rica em vegetais, frutas e produtos integrais. O aumento da atividade física pode ajudar também a reduzir este risco de CCR. Foi comprovado que a aspirina reduz o risco de CCR em pessoas com a síndrome de Lynch (21).

2.2.6. Rastreio e Diagnóstico

Os rastreios são essenciais para a deteção da doença mesmo antes desta apresentar quaisquer sinais ou sintomas porque normalmente os sintomas surgem apenas num estadio mais avançado. Existem vários exames para a deteção das

doenças sendo eles a pesquisa de sangue oculto nas fezes, a sigmoidoscopia, colonoscopia e a deteção de marcadores tumorais (7,21,22).

Se apresentar quaisquer sinais ou sintomas de um possível CCR, o médico deverá confirmar se são provocados por um tumor ou por qualquer outra causa ou patologia. Num doente com história clínica e exame clínico com suspeita de cancro do cólon e reto, devem ser pedidos alguns exames para confirmar ou despistar essa suspeita e, caso estes exames detetem algo suspeito deve proceder-se á biopsia onde normalmente o tecido anómalo é removido (7,21,22).

O diagnóstico efetivo do cancro do cólon e reto só é possível pela análise ao microscópio, pela Anatomia Patológica, de pólipos ou amostras de lesões resultantes da colheita dos mesmos através de uma colonoscopia(7,21,22).

Os pólipos são um dos principais fatores de risco do CCR. O seu lento crescimento de, aproximadamente, 10 anos fornece a oportunidade de prevenção ou mesmo a deteção do cancro em estágios iniciais, diminuindo a taxa de mortalidade, tanto pela diminuição da incidência da doença como por proporcionar um tratamento mais eficaz.

Estima-se que 20% da população desenvolve pólipos adenomatosos durante a vida, mas apenas 5 a 6% da população desenvolve CCR(14). No entanto, os rastreios por colonoscopia geralmente têm tendência a encontrar CCR prevalente em 0,5–0,9% das pessoas na mesma faixa etária.

Os rastreios também são importante nesta neoplasia uma vez que os adenomas e os cancros iniciais são detetáveis e tratáveis (7,21,22).

Pesquisa de sangue oculto nas fezes

O CCR pode produzir sangue e outros componentes que podem ser detetados nas fezes, mesmo antes de aparecerem sintomas da doença. Os exames de pesquisa de sangue nas fezes são um método de rastreio mais usado para CCR. É utilizado preferencialmente o método imunológico em comparação com o método pela resina de guaiaco, porque produz menos falsos positivos. Caso este método dê positivo passa-se para outro método de rastreio para uma confirmação, como as colonoscopias e sigmoidoscopia. A sensibilidade desta análise para neoplasia colorretal é de apenas 50-60%, embora sua especificidade seja alta (8,14,23,24).

Em Portugal, a norma da DGS 003/2014 recomenda a utilização anual deste método, como forma de rastreio. Este é utilizado em pacientes assintomáticos com idades compreendidas entre 50 e 74 anos, exceto em casos de antecedentes pessoais de adenomas ou CCR, doença inflamatória intestinal ou antecedentes familiares de

adenomas de primeiro grau ou CCR ou síndromes hereditários de cancro do cólon e reto(8,14,23,24).

Colonoscopia

A colonoscopia é um exame realizado ao cólon e reto, que tem o objetivo de detetar pólipos, regiões anormais ou cancro.

Segundo a norma da DGS 003/2014, a colonoscopia é utilizada quando o teste imunológico para pesquisa de sangue nas fezes dá positivo ou para as exceções à pesquisa de sangue oculto nas fezes. Caso a colonoscopia total se apresente normal, este procedimento apenas deve ser repetido passado 10 anos. (8,14,24)

Estudos observacionais sugerem que a colonoscopia pode ajudar a reduzir a incidência de CCR em cerca de 40% e a mortalidade em cerca de 60%.

Este método tem algumas limitações tais como um maior risco de complicações em comparação com outros exames de rastreio, como lacerações intestinais e sangramento, especialmente quando um pólipos é removido ou os pacientes são de faixas etárias mais avançadas”. A qualidade da colonoscopia, é variável e está associada a lesões não encontradas, que às vezes progridem para CCR antes do próximo exame agendado. A colonoscopia de baixa qualidade está associada a uma maior probabilidade de morte de CCR(8,14,24).

Sigmoidoscopia

A sigmoidoscopia é um exame idêntico á colonoscopia, a diferença é que analisa apenas o cólon sigmoide e reto com o intuito de detetar regiões anormais ou cancro. Dados de ensaios clínicos realizados mostram uma redução de cerca de 20% a 25% na incidência de CCR e uma redução de 25% a 30% na mortalidade por CCR, com maiores reduções em homens de que em mulheres (8,14).

Marcadores Tumorais

Para que haja um diagnóstico precoce e consecutivamente um melhor tratamento é necessário os realizar análises aos marcadores tumorais (25,26).

Os marcadores tumorais são macromoléculas presentes no tumor, no sangue ou em outros líquidos biológicos, cujo aparecimento e ou alterações em suas concentrações estão relacionados com a génese e o crescimento de células neoplásicas que secretadas pelo tecido como resposta ao tumor. Estes são normalmente utilizados para o rastreio, diagnóstico e classificação dos tumores, como também para o

prognóstico e a monitorização de recorrência do tumor ou de metástases nos casos de cancro (25,26). Estes podem ser hormonas, ácidos nucleicos, poliaminas, antigénios, enzimas, proteínas e lípidos específicos da membrana celular. Estes limitam-se a determinar o estadiamento, o prognóstico e uma possível resposta terapêutica. O antigénio carcino-embrionário (CEA) e o antigénio carbohidrato 19-9 (CA 19-9) são os marcadores tumorais mais estudados que têm sido utilizados para o controlo dos cancros gastrointestinais (25,26)

2.3. Genética vs. Cancro Colorretal

Em média em cada gene tem entre 8 a 10 pequenos fragmentos de codificação denominados exões interrompidos por sequências não codificantes extensas, na forma de intrões aproximadamente 20 vezes maiores. Os intrões são uma marca da evolução eucariótica. Durante a transcrição, toda a informação da sequência de um gene é copiada para ARN mensageiro (ARNm) precursor (pré-ARNm), que inclui exões e intrões (27).

Os genes são constituídos por uma região reguladora que é composta pelo promotor (onde se liga a ARN-polimerase) a montante da região codificadora, e uma região terminadora. As regiões intrónicas são transcritas para o pré-ARNm, e são removidas do ARNm através do processo denominado *splicing* de ARN (28). A remoção eficiente e precisa de intrões é catalisada por uma ribonucleoproteína altamente sofisticada chamada spliceossoma (27). O spliceossoma é composto por cinco pequenas ribonucleoproteínas, sendo elas o U1, U2, U4, U5, e U6 snRNAs snRNP, e por um conjunto de fatores proteicos associados. Dentro do spliceossoma, as sequências de *splicing* no pré-ARNm são forçadas a rearranjos tridimensionais que permitem a ativação de um centro catalítico de ARN e desencadeiam a reação de *splicing* (27). A grande maioria dos intrões são processados pelo complexo U1, U2, U4, U5, e U6 snRNP, conhecido como o spliceossoma maior. No entanto, o *splicing* de uma minoria de intrões é realizada por um complexo snRNP chamado spliceossoma menor. Em geral, os *splicing* maiores e menores partilham muitas características comuns e o mecanismo é quase idêntico. O spliceossoma menor é composto por quatro snRNAs distintos denominados como U11, U12, U4atac e U6atac (27).

Depois do *splicing*, os intrões são excisados e na maioria dos casos degradados. No entanto, alguns dos intrões não são degradados e dão origem a subprodutos funcionais não codificantes de ARN nucleolar (snoRNAs). Alguns genes não têm capacidade de codificação de proteínas e a sua função primária pode ser a de gerar

snoARNs a partir dos seus intrões. Para além dos snoARNs, uma classe de micro ARNs não convencionais (miARNs) é também produzida a partir de intrões (27).

O *splicing* alternativo aumenta a variedade do transcriptoma e do proteoma ao gerar múltiplas isoformas de ARNm a partir de um único gene (27). A retenção do intrão, em ARNms maduros era considerada uma consequência de *splicing* errada porque, os ARNms que contêm intrões são frequentemente alvo de degradação pelo exossoma no núcleo. Um estudo recente mostrou que, muitos intrões são ativamente retidos em transcrições poliadeniladas e contribuem para a desregulação da expressão genética. O termo intrões retidos foi proposto para descrever a classe de intrões que são transitoriamente retidos em transcrições nucleares, e que podem sofrer *splicing* e ainda afetar a expressão genética por outros mecanismos (27).

Os limites intrão-exão são delimitados por curtas sequências nos locais de *splicing* (ss) 5'UTR (dador) e 3'UTR (aceitador) que medeiam o reconhecimento pelo spliceossoma.

As sequências intrónicas foram inicialmente consideradas não funcionais. Os elementos conservados nos intrões incluem as sequências de aceitação do local de *splicing*, os locais de ligação para proteínas reguladoras, as sequências de genes de ARN não codificantes, bem como regiões adicionais. Vários estudos revelaram que os primeiros intrões (ou seja, na extremidade 5' dos genes) são normalmente os mais longos e mais conservados. A conservação do primeiro intrão está provavelmente relacionada com a presença de elementos reguladores e a um padrão específico de organização da cromatina (27).

Em humanos 94% dos genes são interrompidos, em média, por sete intrões(29). As sequências intrónicas podem melhorar ou inibir o *splicing* e podem ser associadas a mutações causadoras de doenças. Entre outras doenças, as mutações intrónicas têm sido bastante estudadas em cancro estando associadas ao cancro da bexiga, faringe e pulmão (30) Existem mutações que afetam intensificadores de *splicing* intrónico ou silenciadores que têm o potencial de resultar em *splicing* mal regulado (27). A recente introdução de abordagens de sequenciamento do genoma completo resultou na identificação de um número crescente de variantes patogénicas localizadas profundamente nos intrões. Variantes intrónicas profundas são mutações que ocorrem a mais de 100 pares de bases dos limites do exão-intrão (27). Estas descobertas estão a promover uma pesquisa focada na compreensão em como a variação na sequência intrónica profunda afeta o *splicing* pré-ARNm e contribui para os fenótipos da doença. Existem pacientes portadores de alterações de ADN consideradas variantes genéticas

de significado desconhecido. Estes casos costumam ser o caso de variantes sinónimas, missense e intrónicas (27). Quanto mais longo for um gene, mais provável é que seja afetado por mutações patogénicas. Não é, portanto, surpreendente que numerosas mutações intrónicas profundas tenham sido descritas em genes particularmente longos como os associados à neurofibromatose. Mutações intrónicas profundas que promovem a inclusão de um pseudo-exão têm sido descritas em várias síndromes tumorais hereditárias (27). Estas incluem neurofibromatose dos tipos 1 e 2, melanoma, ataxia-telangiectasia, retinoblastoma, síndrome de Lynch, cancro da mama, e polipose adenomatosa familiar. Possivelmente, a ligação independente de *splicing* de U1 snRNP a regiões intrónicas suprime tanto a inclusão de pseudo-exões como a clivagem prematura e a poliadenilação a partir de sinais de poliadenilação localizados em intrões. Em alguns casos, a mutação intrónica profunda leva ao aparecimento de mais do que uma isoforma de ARNm pelo *splicing* aberrante (31).

A grande parte das mutações intrónicas profundas não tem qualquer efeito nos sítios de *splicing* canónicos. No entanto, existem mutações que criam um novo sítio de *splicing*, que por consequência, interferem com o reconhecimento de sítios de *splicing* naturais (31). O enfraquecimento dos sítios de *splicing* canónico é frequentemente observado quando as mutações intrónicas profundas estão a menos de 150 pb das junções exão- intrão naturais. Outras formas de mutação intrónica profunda são as que perturbam as zonas reguladoras de transcrição. Foram relatadas variantes genéticas que causam doenças através da inativação de genes ARN codificados por intrão (31).

Os rearranjos cromossómicos que resultam em fusões genéticas são frequentemente detetados em células cancerígenas. Os rearranjos que envolvem as regiões intrónicas profundas foram descritos, em associação com doenças monogénicas. Existem evidências a indicar que a perda de expressão destes contribui para o fenótipo da doença (27). Outros estudos sugerem um papel dos snoARNs no cancro. Em particular, os SNORD50A e SNORD50B ARNs, que são codificados por genes localizados dentro dos intrões 4 e 5 do longo gene não-proteico codificador SNHG5, são recorrentes perdas em muitos tipos de cancro (27). Os SNORD50A e SNORD50B ARNs ligam-se diretamente ao K-Ras e a sua perda desencadeia a hiperativação da sinalização de Ras-ERK1/ERK2 contribuindo para a oncogénese (27).

Os carcinomas colorretais podem ser classificados como esporádicos, onde ocorrem mutações pontuais que afetam células individuais e as descendentes. Este tipo de cancro é heterogéneo, onde existem vários genes mutados. Normalmente este tipo de cancro tem uma sequência de acontecimentos tendo início na formação de um adenoma e termina num estado de carcinoma (10). A primeira mutação ocorre na

polipose adenomatosa do colo (PAC), um gene supressor de tumor, desencadeando a formação de adenomas não malignos, também chamados de pólipos. Estima-se que aproximadamente 15% desses adenomas têm uma evolução para o estado de carcinoma em dez anos (10). No CCR hereditário um dos alelos está mutado, fazendo com que uma mutação pontual no outro alelo desencadeie o aparecimento da célula tumoral e, posteriormente, o carcinoma (10). Dentro das formas hereditárias, o CCR hereditário sem polipose (CCHSP) que está relacionado a mutações nos mecanismos de reparo do ADN, ou Síndrome de Lynch, o mais prevalente, que é causada por mutações herdadas num dos alelos que codificam para proteínas de reparo de ADN, como MSH2, MLH1, MLH6, PMS1 e PMS2 (32). A mutação nos genes referidos anteriormente vão fazer com que se acumule mutações de ADN, que ocorrem em fragmentos de ADN de microssatélites (MSI) com sequências de nucleótidos repetidas. Este tipo de cancro também é causado por mutações hereditárias, embora não sejam classificados como cancro hereditário (32). No entanto, outras formas hereditárias podem conduzir à neoplasia como a polipose adenomatosa familiar (PAF) que apresentam instabilidade cromossomal (INC) e que tem como característica formação de múltiplos pólipos potencialmente malignos no cólon e que aparece associada à mutação no gene responsável pela reparação do ADN, designadamente, o *Adenomatous Poliposis Coli* (PAC) que tem a função de controlar a via de sinalização do WNT (32,33). A mutação no gene PAC é seguida por mutações em KRAS, TP53. A maioria dos pacientes com esta síndrome tem idades mais jovens e desenvolve um grande número de adenomas colorretais (32,33).

O CCR pode também ser classificado por familiar. Deve ser feito um aconselhamento genético para famílias com histórico de cancro. É importante realçar que para as famílias, por mais extensa que seja a história familiar de cancro (a menos que esteja presente em ambos os lados), uma pessoa terá mais de 50% de hipótese de não desenvolver aquele tumor específico (32,33).

3. Metodologia

3.1. Desenho do Estudo

Este é um estudo científico, tipo descritivo-correlacional, onde é feita uma questão de partida e apresentadas algumas hipóteses de investigação, no qual foram coletadas informações e descrições de pesquisa, com vista à identificação de relações entre variáveis. A recolha dos dados teve início em fevereiro de 2021 e término em maio do mesmo ano.

3.2. Questão de Partida e Hipóteses de Investigação

3.2.1. Questão de partida

Existiram mutações intrónicas profundas em genes associados ao cancro colorretal?

3.2.2. Hipóteses de investigação

Para responder a esta questão de partida, foram formuladas 2 hipóteses de investigação:

- Há mutações intrónicas profundas associadas ao CCR;
- Há mutações intrónicas profundas estudadas consideráveis *Driver mutations* no CCR;

3.3. Recolha dos dados

As variantes intrónicas profundas foram recolhidas da base de dados *Nacional Cancer Institute*.

Foi selecionado o cancro primário em cancro do cólon e do reto e mutações que ocorreram apenas em intrões e zonas de *splicing*. Seguidamente recolhi em 91 mil mutações que me foram apresentadas apenas 10mil mutações intrónicas profundas. As mutações recolhidas foram então utilizadas num programa bioinformático tal como descrito abaixo.

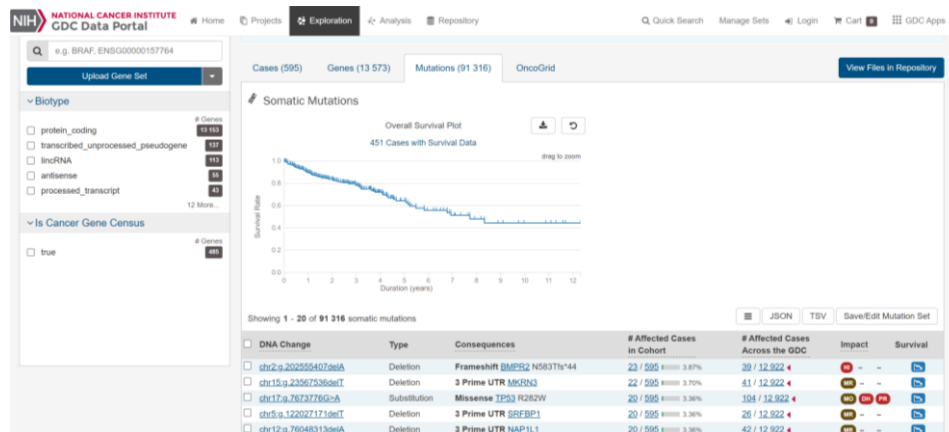


Figura 3.1- Página do Nacional Cancer Institute com as condições de pesquisa selecionadas.

3.4. Análise in silico

Para prever os efeitos das variantes in Silico foi utilizada uma ferramenta bioinformática chamada VETA que faz uma previsão precisa de variantes e possíveis consequências das mesmas a partir de uma sequência primária usando um spliceAI, um bloco residual, que consiste em camadas de lotes, unidades lineares retificadas (ReLU), e unidades organizadas de uma forma específica que prevê se cada posição numa transcrição pré-ARNm leva a criação ou destruição de um dador de *splicing*, aceitador de *splicing*, ou nenhum, usando apenas a sequência genómica da transcrição pré-ARNm (34).

O score é obtido numa escala de 0 a 100 sendo que apenas os valores que são maiores que 0,2 têm um efeito relevante. Quanto maior for o valor mais significativo é o resultado, ou seja, que a variante tem uma grande probabilidade de realizar o efeito obtido bio informaticamente.

Através do programa podemos obter quatro efeitos possíveis sendo eles: Variante que leva á criação de um *splice acceptor* (DS_AG), variante que leva á criação destruição de um *splice acceptor* (DS_AL), variante que leva á criação de um *splice donor* (DS_DG) e variante que leva á criação de um *splice acceptor* (DS_DL).

Na tabela 4.1 podemos observar os dados obtidos da ferramenta bioinformática utilizada para a previsão de variantes e possíveis consequências. Podemos também observar nesta tabela quais as variantes já foram estudadas e as que nunca foram estudadas ou que não existe informações a falar delas.

4. Resultados

Com os dados obtidos na tabela 4.1, foi possível dividir informação em duas tabelas. A tabela 4.2 contém variantes que não se encontram estudadas ou que não tem quaisquer informações online, e o respetivo efeito previsto pelo programa bioinformático. A tabela 4.3 apresenta variantes, os genes á qual correspondem, significado clínico previsto e por fim o que se encontra descrito na literatura.

Como referido anteriormente, o *score* obtido é mostrado numa escala de 0 a 100 sendo que apenas os valores que são maiores que 0,2 têm um efeito relevante. Quanto maior o valor mais significativo é o resultado, ou seja, a variante tem uma grande probabilidade de realizar o efeito obtido bio informaticamente.

Com esta previsão pode observar-se que a maioria das variantes provoca uma destruição de um aceitador de *splicing* sendo que o valor máximo obtido nestas variantes foi de 0.86.

Tabela 4.1- Resultados obtidos pela ferramenta bioinformática utilizada para a previsão de variantes e possíveis efeitos.

#CHROM	POS	REF ALT	HGVSc	NM	Gene	Protein	HGVSp	gnomADg_AF	Existing_variation	DS	AGDS	ALDS	DGDS_DL
1	36444025	T C	ENST00000356837.9:c.121A>G	NM_001330493.2	OSCP1	NP_659484.4	1:g.36444025T>C		None	0.00	0.27	0.00	0.07
1	160801902	C T	ENST00000263285.11:c.454+1820C>NM_002348.4	LY9	NP_002339.2	1:g.160801902C>T	6.977300017751988e-06	rs72038465&COSV54431226	0.00	0.00	0.50	0.00	
1	17255332	G T	ENST00000367723.8:c.782+1706G>T	014283.5	SUCO	NP_055098.1	1:g.17255332G>T		COSV55263092	0.00	0.00	0.03	0.63
1	174809474	C T	ENST00000251507.9	NM_001386446.1	RABGAP1L	NP_001353375.1	1:g.174809474C>T		rs781982432&COSV52293653	0.00	0.00	0.01	0.23
10	17152074	C A	ENST00000377799.8:c.1075+1433G>NM_004412.7	TRDMT1	NP_004403.1	10:g.17152074C>A		None		0.00	0.35	0.00	0.00
10	32907089	C T	ENST00000396033.8:c.2331+1279G>NM_002211.4	ITGB1	NP_002202.2	10:g.32907089C>T	0.00013281000428833081rs200899694			0.00	0.01	0.00	0.21
10	113160663	C T	ENST000000627217.2	NM_001146274.1	TCF7L2	NP_001139746	10:g.113160663C>T		COSV63336375&COSV533395	0.00	0.14	0.00	0.24
10	133272063	T C	ENST00000445355.8:c.958-109A>G	NM_001109.5	ADAM8	NP_001100.3	10:g.133272063T>C		None	0.00	0.28	0.05	0.00
11	68179470	C A	ENST00000304363.9:c.377+662G>T	NM_017635.5	KMT5B	NP_060105.3	11:g.68179470C>A	6.981800197536359e-06	rs1242113446	0.00	0.00	0.14	0.37
11	78121191	G T	ENST00000299626.10:c.369-17C>A	NM_024079.5	ALGB	NP_076984.2	11:g.78121191G>T		None	0.00	0.28	0.00	0.00
11	83212903	GT G	ENSG00000137494	NM_001300972.1	ANKRD42	NP_001287901.1	11:g.83212914del	8.018800144782229e-05	rs537626867	0.47	0.48	0.00	0.00
11	84316823	A G	ENST00000531895.5	NM_001142699.3	DLG2	NP_001674.1	11:g.84316823A>G		COSV54675309	0.00	0.00	0.00	0.57
11	125086382	G T	ENST00000308074.4:c.29+122G>T	NM_001145290.2	SLC37A2	NP_001138762.1	11:g.125086382G>T		None	0.04	0.24	0.00	0.00
13	49265956	G A	ENSG0000102543	NM_030911.4	CDADC1	NP_112173.1	13:g.49265956G>A	6.9780999183421955e-06	rs1456271345	0.00	0.00	0.00	0.23
14	35829815	G T	None			14:g.35829815G>T		None		0.00	0.27	0.00	0.00
14	67592310	C T	ENST00000216452.9:c.474+325G>A	NM_004569.5	PIGH	NP_004560.1	14:g.67592310C>T		None	0.01	0.00	0.00	0.30
16	66609796	G T	ENST00000424011.6:c.400-87G>T	NM_181553.4	CMTM3	NP_853531.1	16:g.66609796G>T		None	0.00	0.24	0.00	0.00
16	88859593	TG T	None			16:g.88859594del		None		0.81	0.19	0.00	0.01
17	44317515	C T	ENST00000426726.8:c.1199-581C>T	NM_001144825.2	RUNDC3A	NP_001138297.1	17:g.44317515C>T	2.7915000828215852e-05	rs764379763	0.00	0.26	0.00	0.00
17	50088228	G A	None	NM_002204.4	ITA3	NP_002195.1	17:g.50088228G>A	2.7923999368795194e-05	rs1276437993	0.90	0.00	0.00	0.00
19	54219875	G A	ENST00000448962.5:c.221-11866G>T	NM_001013.4	RPS9	NP_001004.2	19:g.54219875G>A		None	0.00	0.24	0.00	0.15
19	57814533	C T	None			19:g.57814533C>T		None		0.00	0.00	0.01	0.54
2	46958547	G T	ENST00000394850.6:c.517+1540G>T	NM_020458.4	TTC7A	NP_085191.2	2:g.46958547G>T		rs543740098	0.00	0.00	0.22	0.00
2	113756048	T A	None			2:g.113756048T>A		None		0.00	0.01	0.00	0.58
2	160282306	T G	None			2:g.160282306T>G		None		0.01	0.86	0.00	0.20
2	197420192	G A	ENST00000335508.11:c.415+236C>T	NM_012433.4	SF3B1	NP_036565.2	2:g.197420192G>A		None	0.00	0.21	0.00	0.18
20	63941521	C T	ENST00000399908.9	NM_017859.4	UCKL1	NP_060329.2	20:g.63941521C>T	0.0013744999887421727	rs1564676949	0.41	0.00	0.00	0.00
3	57864842	G T	None			3:g.57864842G>T		None		0.00	0.33	0.00	0.34
3	138188400	GT ENST00000538260.5	NM_001363941.2	ARMC8	NP_001350870.1	3:g.138188400dup	7.083499804139137e-05	rs1446739040	0.37	0.57	0.00	0.00	
4	157361112	A G	ENST00000296528.12:c.2292-426A>C	NM_001083619.3	GRIA2	NP_001077088.2	4:g.157361112A>G		None	0.00	0.00	0.02	0.39
6	97270208	A G	ENST00000275053.8:c.603-216T>C	NM_001350599.2	MMS22L	NP_001337528.1	6:g.97270208A>G		None	0.00	0.25	0.00	0.02
6	102064049	G C	ENST00000421544.6:c.2563-4298G>C	NM_021956.5	PIRK2	NP_068775.1	6:g.102064049G>C		None	0.00	0.00	0.00	0.43
6	128008069	T C	ENST00000532331.5:c.2333+1061A>NM_002844.4	PTPRK	NP_002835.2	6:g.128008069T>C	2.1112999093020335e-05	rs1450222112	0.00	0.16	0.00	0.21	
6	167893861	G A	ENSG00000130396	NM_001386888.1	AFDN	NP_001373817.1	6:g.167893861G>A		COSV60043856	0.32	0.50	0.00	0.00
7	27803133	G T	ENST000000265393.10	NM_006024.7	TAX1BP1	NP_006015.4	7:g.27803133G>T	6.980400030442879e-06	rs996499746&COSV55291745	0.00	0.27	0.00	0.26
9	4650808	A G	ENST00000475086.5:c.177+581T>C	NM_001353486.2	SPATA6L	NP_001340415.1	9:g.4650808A>G		None	0.00	0.32	0.00	0.23
9	125335222	G T	ENST00000394105.6:c.2479G>T	NM_015635.4	GAPVD1	NP_001269609.1	9:g.125335222G>T		None	0.00	0.05	0.00	0.25
X	53279565	T C	None			X:g.53279565T>C		None		0.00	0.05	0.00	0.25
X	72274437	T C	None			X:g.72274437T>C		None		0.04	0.24	0.00	0.00

Tabela 4.2- Resultados obtidos através da ferramenta bioinformática. São apresentadas abaixo apenas as variantes sem estudos realizados sobre as mesmas e o efeito previsto pelo programa bioinformático

#CHROM	POS	REF ALT	HGVSc	NM	Gene	Protein	HGVsg	gnomADg_AF	Existing variation	DS	AG	DS	ALDS	DGDS	DL
1	36444025	T C	ENST00000356637 9.c.121A>G	NM_01330493.2	OSCP1	NP_659484.4	1.g.36444025T>C		None	0.00	0.27	0.00	0.07		
10	17152074	C A	ENST00000377799 8.c.1075+1433G>T	NM_004412.7	TRDMT1	NP_004403.1	10.g.17152074C>A		None	0.00	0.35	0.00	0.00		
10	133272063	T C	ENST00000445355 8.c.958-109A>G	NM_001109.5	ADAM8	NP_001100.3	10.g.133272063T>C		None	0.00	0.28	0.05	0.00		
11	78121191	G T	ENST00000299626 10.c.369-17C>A	NM_024079.5	ALG8	NP_076984.2	11.g.78121191G>T		None	0.04	0.24	0.00	0.00		
11	125086382	G T	ENST00000308074 4.c.'29+122G>T	NM_001145290.2	SLC37A2	NP_001138762.1	11.g.125086382G>T		None	0.00	0.27	0.00	0.00		
14	35829815	G T	None				14.g.35829815G>T		None	0.00	0.27	0.00	0.00		
14	67592310	C T	ENST00000218452 9.c.474+325G>A	NM_004569.5	PIGH	NP_004560.1	14.g.67592310C>T		None	0.01	0.00	0.00	0.30		
16	66609796	G T	ENST00000424011 6.c.400-87G>T	NM_181553.4	CMTM3	NP_853531.1	16.g.66609796G>T		None	0.00	0.24	0.00	0.00		
16	88859593	TG T	None				16.g.88859594del		None	0.81	0.19	0.00	0.01		
19	54219975	G A	ENST00000448962 5.c.221-1186G>A	NM_001013.4	RPS9	NP_001004.2	19.g.54219975G>A		None	0.00	0.24	0.00	0.15		
19	57814533	C T	None				19.g.57814533C>T		None	0.00	0.00	0.01	0.54		
2	113756048	T A	None				2.g.113756048T>A		None	0.00	0.01	0.00	0.58		
2	160282306	T G	None				2.g.160282306T>G		None	0.01	0.86	0.00	0.20		
2	197420192	G A	ENST00000335508 11.c.415+236C>T	NM_012433.4	SF3B1	NP_036565.2	2.g.197420192G>A		None	0.00	0.21	0.00	0.18		
3	57864842	G T	None				3.g.57864842G>T		None	0.00	0.33	0.00	0.34		
4	157361112	A G	ENST00000296526 12.c.2292-426A>G	NM_01083619.3	GRIA2	NP_001077088.2	4.g.157361112A>G		None	0.00	0.00	0.02	0.39		
6	97270208	A G	ENST00000275053 8.c.607-216T>C	NM_001350599.2	MMS22L	NP_001337528.1	6.g.97270208A>G		None	0.00	0.25	0.00	0.02		
6	102064049	G C	ENST00000421544 6.c.2563-4298G>C	NM_021956.5	GRIK2	NP_068775.1	6.g.102064049G>C		None	0.00	0.00	0.00	0.43		
9	4656088	A G	ENST00000475086 5.c.177+581T>C	NM_001353486.2	SPATA6L	NP_001340415.1	9.g.4656088A>G		None	0.00	0.32	0.00	0.23		
9	125335222	G T	ENST00000394105 6.c.2479G>T	NM_015635.4	GAPVD1	NP_001269609.1	9.g.125335222G>T		None	0.00	0.05	0.00	0.25		
X	53279565	T C	None				X.g.53279565T>C		None	0.00	0.05	0.00	0.25		
X	72274437	T C	None				X.g.72274437T>C		None	0.04	0.24	0.00	0.00		

Das variantes referidas acima no quadro podemos observar que apenas existem três variantes com elevada probabilidade de realizar o efeito obtido. com valores significativos acima dos 0.5.

Abaixo encontra-se na tabela 4.3 pode-se visualizar todos os genes já estudados juntamente com o resultado obtido bioinformaticamente e o que se encontra descrito na literatura.

Tabela 4.3- Resultados obtidos através da ferramenta bioinformática. São apresentadas abaixo apenas as variantes já estudadas, o efeito previsto pelo programa bioinformático e o que se encontra descrito sobre as mesmas na literatura

#CHROM	POS	REF ALT	HGVSc	NM	Gene	Protein	HGVsg	gnomADg_AF	Existing variation	DS	AG	DS	ALDS	DGDS	DL
1	172553332	G T	ENST00000367723 8.c.762+1706G>T	NM_014283.5	SUCO	NP_055098.1	1.g.172553332G>T		COSV55263092	0.00	0.00	0.03	0.63		
1	174800474	C T	ENST00000251507 8.c.	NM_001368446.1	RABGAP1L	NP_001353375.1	1.g.174800474C>T		rs761962432&COSV52293653	0.00	0.00	0.01	0.23		
10	32907089	C T	ENST00000396033 8.c.2331+1279G>	NM_002211.4	ITGB1	NP_002202.2	10.g.32907089C>T	0.00013281000428833067	rs200899894	0.00	0.01	0.00	0.21		
10	113160663	C T	ENST00000306272 7.2	NM_001146274.1	TCF7L2	NP_001139746	10.g.113160663C>T		COSV53336375&COSV5333965	0.00	0.14	0.00	0.24		
11	86179470	C A	ENST00000304363 9.c.377+662G>T	NM_017635.5	KMT5B	NP_060106.3	11.g.86179470C>A	6.981800197536356e-06	rs1242113446	0.00	0.00	0.14	0.37		
11	83212903	GT G	ENSG00000137494	NM_001300972.1	ANKRD42	NP_001287901.1	11.g.83212914del	8.01880014478228e-05	rs537626867	0.47	0.48	0.00	0.00		
11	84316823	A G	ENST00000531895.5	NM_001142699.3	DLG2	NP_001674.1	11.g.84316823A>G		COSV54675309	0.00	0.00	0.00	0.57		
13	49265956	G A	ENSG0000102543	NM_030911.4	CDADC1	NP_112173.1	13.g.49265956G>A	6.9780999183421955e-06	rs1456271345	0.00	0.00	0.00	0.23		
17	44317515	C T	ENST00000426726 8.c.1199-581C>T	NM_001144825.2	RUNDC3A	NP_001138297.1	17.g.44317515C>T	2.7915000828215852e-05	rs1276437983	0.00	0.26	0.00	0.00		
17	50088228	G A	None	NM_002204.4	IT3A3	NP_002195.1	17.g.50088228G>A	2.7923999368795194e-05	rs1276437993	0.90	0.00	0.00	0.00		
2	46958547	G T	ENST00000394850 8.c.517+1540G>T	NM_020458.4	TTC7A	NP_065191.2	2.g.46958547G>T		rs543740098	0.00	0.00	0.22	0.00		
20	63941521	C T	ENST00000369908.9	NM_017859.4	UCKL1	NP_060329.2	20.g.63941521C>T	0.0013744999887421727	rs564676949	0.41	0.00	0.00	0.00		
3	138188400	G T	ENST00000532826.0	NM_001363941.2	ARMC8	NP_001350870.1	3.g.138188407dup	7.083499804139137e-05	rs1446739040	0.37	0.57	0.00	0.00		
6	128008069	T C	ENST00000532331 5.c.2333+1061A>	NM_002844.4	PTPRK	NP_002835.2	6.g.128008069T>C	2.1112999093020335e-05	rs1450222112	0.00	0.16	0.00	0.21		
6	167893861	G A	ENSG00000130396	NM_001386888.1	AFDN	NP_001373817.1	6.g.167893861G>A		COSV60043856	0.32	0.50	0.00	0.00		
7	27803133	G T	ENST00000265393 10	NM_006024.7	TAX1BP1	NP_006015.4	7.g.27803133G>T	6.980400030442979e-06	rs99649746&COSV55291745	0.00	0.27	0.00	0.00		

5. Discussão de resultados

O objetivo geral deste estudo é a realização de uma análise exaustiva de mutações intrónicas profundas já identificadas e prever o efeito funcional dessas mutações *in silico* no CCR.

Após serem obtidos dados do programa bioinformático, foi possível dividir informação em duas tabelas. A primeira tabela consiste nas variantes e genes já reportados clinicamente e a segunda tabela contém variantes que ainda não foram reportadas clinicamente. Todos os genes referidos anteriormente têm, de alguma forma, ligação ao CCR.

A variante C.121A>G provoca uma alteração do transcrito na posição 50 da proteína OSCP1 (35). Na figura 5.1 podemos observar a sequência de aminoácidos da proteína e abaixo a codificação com a respetiva mutação. De acordo com o resultado bioinformático obtido, a variante provoca uma destruição de um aceitador de *splicing*, o que pode levar a uma transcrição do ADN incorreta e por consequência a proteína deixar de ter a sua função de transporte transmembranar(35). A proteína OSCP1 tem como função biológica a atividade de transporte transmembranar de xenobióticos (35). Esta proteína pode também estar envolvida no desenvolvimento e/ou progressão do carcinoma nasofaríngeo. Pode estar altamente expresso na placenta e em linhas celulares derivadas de tumores como as do útero, testículo e segmentos olfativos e mucosa nasal e em menor quantidade o timo e intestino delgado (35). Esta variante não tem significado clínico reportado no clinvar.(36)

Proteína de referência

```

1 MSVRTLPLLF LNLGGEMLYI LDQRLRAQNI PGDKARKDEW TEVDRKRVLN DIISTMFNRK
61 FMEELFKPQE LYSKKALRTV YERLAHASIM KLNQASMDKL YDLMTMAFKY QVLLCPRPKD
121 VLLVTFNHL D TIKGFIRDSP TILQQVDETL RQLTEIYGGL SAGEFQLIRQ TLLIFFQDLH
181 IRVSMFLKDK VQNNNGRFVL PVSQPVVWGT EVPGLIRMFN NKGEEVKRIE FKHGGNYVPA
241 PKEGSFELYG DRVLKLGTMN YSVNQPVETH VSGSSKNLAS WTQESTAPNP LAKEELNFLA
301 RLMGGMEIKK PSGPEPGFRL NLFTTDEEEE QAALTRPEEL SYEVINIQAT QDQQRSEELA
361 RIMGFEITE QPRLSTSKGD DLLAMMDEL*

```

Proteína prevista a partir de sequência de codificação variante

```

1 MSVRTLPLLF LNLGGEMLYI LDQRLRAQNI PGDKARKDEW AEVDRKRVLN DIISTMFNRK
61 FMEELFKPQE LYSKKALRTV YERLAHASIM KLNQASMDKL YDLMTMAFKY QVLLCPRPKD
121 VLLVTFNHL D TIKGFIRDSP TILQQVDETL RQLTEIYGGL SAGEFQLIRQ TLLIFFQDLH
181 IRVSMFLKDK VQNNNGRFVL PVSQPVVWGT EVPGLIRMFN NKGEEVKRIE FKHGGNYVPA
241 PKEGSFELYG DRVLKLGTMN YSVNQPVETH VSGSSKNLAS WTQESTAPNP LAKEELNFLA
301 RLMGGMEIKK PSGPEPGFRL NLFTTDEEEE QAALTRPEEL SYEVINIQAT QDQQRSEELA
361 RIMGFEITE QPRLSTSKGD DLLAMMDEL*

```

Figura 5.1- Resultados obtidos pela ferramenta bioinformática utilizada para a previsão de variantes e possíveis efeitos.

A variante c.454+1820C>T de acordo com o programa bioinformático provoca a criação de um dador de *splicing* no gene LY9. Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína. A proteína LY9 é um recetor SLAM, ou seja, é uma molécula sinalização de ativação linfocítica (37,38) Estes recetores são desencadeados por interações célula-célula provocam a ativação e diferenciação de uma grande variedade de células imunes. Encontra-se envolvida na regulação e interconexão da resposta imune inata e adaptativa (37,38). Participa em reações de adesão entre linfócitos T e células secundárias por interações homofílicas. Promove a diferenciação de células T para células T específicas que levam a secreção do IL-17 (37,38) Esta proteína pode também ser um regulador negativo da resposta imune e desativar respostas a autoanticorpos e inibir a secreção de IFN-gama por células CD4⁺T (37,38).

A variante c.762+1706G>T de acordo com a análise *in silico* provoca uma criação de um aceitador de *splicing* no gene SUCO. Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína. O gene SUCO codifica uma proteína de desenvolvimento necessária na modelagem óssea durante a embriogénese tardia e regula a síntese de colagénio tipo I em osteoblastos durante a maturação pós-natal, participa na síntese proteica, promove a proliferação do osteoblasto, está envolvido no ciclo celular e metabolismo celular. (39). É uma proteína altamente expressa no pâncreas e testículos e em menor grau na próstata, ovários, coração, timo, intestino delgado e baço (40,41). Um estudo publicado sobre *SUCO* mostra uma superexpressão nos tecidos do carcinoma hepatocelular (HCC), e que a alta expressão do gene estava significativamente relacionada à baixa taxa de sobrevivência dos pacientes com HCC e pode ser um potencial biomarcador diagnóstico para este cancro.(39) Outros estudos demonstraram que a falta deste gene levou ao desenvolvimento neuronal anormal relacionado com a epilepsia e que este gene foi encontrado expresso em cancro retal(39). Sendo um gene ligado á codificação do fator de ossificação, presumo que não seja uma mutação *driver* para o CCR(39).

A proteína ativadora rab GTPase 1-like (RABGAP1L) faz parte da família Ras. As Rab têm como função mediar a formação de vesículas e o seu transporte, juntamente com actina e tubulina, e fusão de membrana(42). O sistema de transporte de membrana intracelular rege a secreção proteica durante a exocitose e a absorção de nutrientes extracelulares durante a endocitose em células eucarióticas. (42). Quando recrutada pela ANK2 e levada para PI3P, inativa a RAB22A e promove o transporte para a extremidade principal das células migratórias. Esta proteína é encontrada na medula óssea em células que produzem neutrófilos, basófilos, eosinófilos e monócitos

desencadeando leucemias mieloides agudas. Esta mutação ainda não foi reportada no clinvar (43,44). As Rab GTPases também têm um papel importante na progressão do cancro, sendo que já foi estudado que existe mutações nas vias de sinalização deste gene. A expressão do rab25 é diminuída no cancro do cólon(42). Outro estudo mostrou que o rab25 pode funcionar como um supressor de tumor em células intestinais, levando ao adenocarcinoma colorretal humano. A perda de Rab25 leva ao aumento da tumorigenese por alteração na regulação do tráfico de proteínas para a superfície celular(42). A deficiência de rab25 promove significativamente o desenvolvimento da neoplasia. Outros estudos também já associaram as GTPases á migração e invasão do cancro (45). São anormalmente expressos em diferentes tecidos cancerígenos. Quase 70 proteínas rabs e rab-like são codificadas pelo genoma humano(46). Sendo um gene supressor de tumor podemos afirmar tratar-se de uma mutação que pode desencadear o processo de tumorigenese. A variante c.1075+1433G>T de acordo com a análise *in silico* provoca uma destruição de um aceitador de *splicing* no gene Rab GTPase no gene TRDMT, o que pode provocar a retenção de um intrão e, por consequência a proteína fica maior.

A proteína tRNA (citosina(38)-C(5))-metiltransferase tem uma atividade catalítica especificamente citosinas metiladas citosina no anti códon de tRNA (Asp). Tem uma maior expressão nos testículos, ovário e timo e em níveis mais baixos no baço, próstata, cólon, intestino delgado e leucócitos sanguíneos periféricos. Esta proteína ainda não foi reportada na base de dados.

A variante c.2331+1279G>A de acordo com a análise *in silico* provoca uma criação de um aceitador de *splicing* no gene ITGB1(47). Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína(47). A beta integrina 1 (ITGB1) é uma proteína heterométrica que contém subunidades alfa e beta(47). São recetores de membrana envolvidos em vários processos como adesão celular e reconhecimento em vários processos como a embriogénese, hemostasia, reparação de tecidos, resposta imune e difusão metastática de células tumorais (47). Dependendo do tipo das subunidades alfa e beta, esta proteína pode ser um recetor para o colágeno, para fibronetina, reconhece um ou mais domínios dentro das regiões de CS-1 e CS-5, é um receptor para fibrinogénio, para a VCAM1 e para a lamimina, está presente nos oócitos e está envolvido na fusão espermatozoides-óvulos(48). Estimula a migração de células endoteliais e está envolvido na promoção da motilidade celular endotelial e angiogénese(48). O ITGA5:ITGB1 atua como receptor de fibrillina-1 (FBN1) e media a adesão celular dependente de R-G-D ao FBN1 já o

ITGA5:ITGB1 é um receptor para IL1B e a vinculação é essencial para a sinalização IL1B e Reduz ligeiramente a interação com o ACAP1(48).

O fator de transcrição *7 like 2* (TCFL2), participa na homeostase da glicose no sangue e do tecido intestinal (49)(50). Codifica um fator de transcrição de alta mobilidade (HMG) que desempenha um papel fundamental na via de sinalização Wnt. Atua como repressor na ausência de CTNNB1 e como ativador na presença do mesmo(49). É uma proteína necessária para a manutenção do compartimento epitelial de células-tronco do intestino delgado. Mutações neste gene estão associadas a um risco aumentado de desenvolver diabetes tipo 2. Existe expressão deste gene por exemplo em gordura e no cólon(49). É um gene com uma grande frequência de mutações em CCR e normalmente encontra-se subexpresso no cancro retal primário resistente á quimioterapia(50). A grande maioria dos tumores colorretais têm mutações na via Wnt/ β -Catenin que por consequência afetam a proliferação, migração, invasão e a capacidade de iniciação tumoral das células do CCR (50). Com esta afirmação podemos dizer que esta mutação será uma mutação *driver*. Estudos demonstram que a necessidade vital deste gene em células intestinais saudáveis é amplamente perdida no curso da carcinogénese colorretal. As células do CCR deficientes de TCF7L2 aumentam a motilidade e a invasão do cancro(50).

A proteína Disintegrina e metaloprotease que contém o domínio 8 (ADAM8) é uma proteína de membrana estruturalmente relacionada a desintegrinas de veneno de cobra(51). Está envolvida em vários processos tais como interações célula- célula e matrizes celulares como fertilização, desenvolvimento muscular e neurogénese(51). O *splicing* alternativo resulta em várias variantes de transcrição(52). Através do programa bioinformático, a variante c.958-109A>G vai provocar uma destruição e um aceitador de *splicing* no gene ADAM8. Por consequência da variante pode provocar a retenção de um intrão e, por consequência a proteína fica maior(52).

A Histona-lysina N-metiltransferase (KMT5B) codifica uma proteína que contém o domínio SET que são domínios de interação proteína-proteína que medeiam interações com proteínas semelhantes às fosfatases(53). Várias variantes de transcrição que codificam diferentes isoformas foram encontradas para este gene(53). É uma proteína que regula a transcrição e manutenção da integridade do genoma(54). Também tem a função na miogénese regulando a expressão de genes-alvo, como o EID3.Facilita a formação de focos após danos no ADN e a reparação do ADN não homólogo(54). Quando ocorre mutações neste gene pode haver formação de um retardo mental autossómico dominante 51(54). A variante c.377+662G>T de acordo

com o programa bioinformático leva à criação de um aceitador de *splicing* no gene KMT5B(54). Esta variante pode provocar alterações na função da proteína por exemplo no reparo do ADN. A variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína(54).

A ALG8 alfa-1,3-glucosyltransferase (ALG8) catalisa a adição do segundo resíduo de glicose ao precursor de oligossacarídeo ligado ao lípido(55). Esta proteína também é necessária para a maturação do *pkd1/ policistina 1* e localização para a membrana plasmática(55). Mutações neste gene têm sido associadas à desordem congénita da *glycosylation* tipo 1h (CDG-1h) e doença hepática policística 3(56). A mutação c.369-17C>A de acordo com o resultado obtido *in silico*, leva à destruição de um aceitador de *splicing* no gene ALG8(56). Esta variante pode provocar a retenção de um intrão e, por consequência a proteína fica maior(56).

A Discos grandes homólogos 2 (DLG2) é uma proteína utilizada na percepção da dor crónica através de um recetor NMDA(57). Regula a expressão de recetores NMDA em neurónios, interage com a cauda do citoplasma das subunidades recetoras NMD(58). Interage com a cauda citoplasmática das subunidades recetoras NMDA, bem como com canais de potássio. Regula também a estabilidade sináptica em sinapses colinérgicas. Existe uma alta expressão desta proteína no cérebro e glândula adrenal(57,58).

A proteína da família transportadora *solute 37* membro 2 (SLC37A2) é um anti transportador da glicose 6-fosfato(59). Transporta a glicose citoplasmática no lúmen do retículo endoplasmático e transportar o fosfato inorgânico na direção oposta(59). Pode não desempenhar um papel na regulação homeostática dos níveis de glicose no sangue(59). É expressa em vários locais tais como na glândula adrenal e glândula salivar. Expressão ampla na glândula adrenal (RPKM 13.0), glândula salivar(60). A variante c.*29+122G>T de acordo com a análise *in silico* vai criar um aceitador de *splicing*. Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, à diminuição do tamanho da proteína. Esta variante pode também fazer alterações na estrutura ou função da proteína. SLC37A2.(60)

O gene CDADC1 tem a função de catalisar a deaminação da citidina e desoxycitidina em uridina e desoxiuridina, respectivamente e desempenha um papel importante no desenvolvimento testicular e espermatogênese(61).

Fosfatidylinositol glicano âncora biossíntese classe H (PIGH) é uma proteína que codifica uma proteína associada ao retículo endoplasmático que está envolvida na biossíntese de glicosylphosphatidylinol (GPI)(62). Esta proteína é encontrada em células sanguíneas à superfície celular(63). A proteína codificada por este gene é uma subunidade do GPI N-acetylglucosaminy (GlcNAc) transferase que transfere GlcNAc para fosfatidylinositol (PI) no retículo endoplasmático(63). A variante.474+325G>A no gene acima referido levou à criação de um aceitador de *splicing* de acordo com o resultado bioinformático. Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína

A proteína CKLF como domínio transmembranar marvel 3 pertence a uma família de quimiocina sendo proteínas de sinalização. Estes encontram-se localizados no cromossoma 16(64). A variante encontrada neste gene, c.400-87G>T, de acordo com o programa bioinformático, leva à destruição de um aceitador de *splicing*(64). Com esta variante a proteína pode provocar a retenção de um intrão e, por consequência a proteína fica maior(64). Já foram descobertas variantes neste gene com diferentes 5' UTRs mas que acabavam por codificar a mesma proteína(64).

Domínio RUN com 3A é uma proteína que tem efeito de RAP2A em células neuronais com função reguladora de GTPases, regulação positiva na sinalização mediada pelo CGMP e transdução de sinal mediado pela GTPase(65,66). Esta é encontrada em vários tecidos sendo em maior percentagem no cérebro e na glândula adrenal. A variante encontrada neste gene foi c.1199-581C>T, e tem como efeito a destruição de um aceitador de *splicing*(65,66). Esta variante pode provocar a retenção de um intrão e, por consequência a proteína fica maior(65,66).

A Integrina alfa-3/beta-1 (ITA3) é uma proteína recetora para fibronectina, laminina, colagénio, epiligrina, trombospondina e CSPG4(67,68). Esta proteína participa em processos de adesão, formação de processos de degradação e matriz, promovendo assim a invasão celular. Não tem ainda significância clínica comprovada no clinvar(67,68).

O gene ribossómico S9 (RPS9) codifica um componente da subunidade 40S(69). Os ribossomas contêm uma subunidade 40S e outra 60S que, juntas contêm 4 espécies de ARN e aproximadamente 80 proteínas estruturalmente distintas(69). Esta proteína pertence à família S4P e esta localizada no citoplasma(69). Foi observado diferentes expressões deste gene nos cancros colorretais quando comparadas com os tecidos normais adjacentes, mas ainda não foi demonstrado que a elevada expressão deste gene induza a uma maior gravidade da doença(69). A variante encontrada neste gene

foi c.221-11866G>A, e tem como efeito a destruição de um aceitador de *splicing*. Esta variante pode provocar a retenção de um intrão e, por consequência a proteína fica maior.

O gene TTC7A é Tetratricopeptida com repetição na proteína 7. Faz parte de vários processos biológicos, tais como ciclo celular, transporte de proteínas, secreção e controlo de proteínas (70). É um componente necessário para a localização do fosfatidylinositol 4-quinase (PI4K) na membrana plasmática. Esta proteína atua como um regulador da síntese do fosfatilinositol 4-fosfato(71). Existe variantes neste gene que interrompem o desenvolvimento intestinal causando uma doença inflamatória intestinal precoce e atresia intestinal(72). A variante encontrada neste gene, c.517+1540G>T, de acordo com o programa bioinformático, irá levar á criação de um dador de *splicing*. Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína(72). O *splicing* alternativo resulta em várias variantes de transcrição(72).A característica marcante da deficiência TTC7A a variação de mutações intestinais(73). Assim sendo poderá ser uma mutação *driver* em CCR.

A proteína do fator *splicing* com subunidade 3B é um componente do complexo de *splicing* no pré-mARN do gene SF3B. É uma proteína necessária para montar o complexo(74). A ligação estável do U2 snRNP à sequência do ponto de ramificação (BPS) no pré-mRNA. Poderá também estar envolvido na ligação do complexo E e no *splicing* de intrões nucleares pré-mARN(74). A variante encontrada, c.415+236C>T, e após análise bioinformática vai provocar uma destruição de um aceitador de *splicing*. pode provocar a retenção de um intrão e, por consequência a proteína fica maior. Existe formação de várias variantes de transcrição a codificar diferentes isoformas através do *splicing* alternativo. Poderá ser uma das consequências da variantes encontrada(75). Este gene é encontrado no linfonodo, medula óssea, entre outros tecidos(75).

A proteína Uridine-cytidine quinase-like 1 codificada pelo gene UCKL1 é uma quinase uridina(76). Tem uma atividade catalítica como a uridina- citidina quinase(76). A quinase uridina está envolvida numa síntese aprimorada do ARN, ou seja, é regulado em células cancerígenas(77). Tem a função de ligante ao antigénio nuclear *Epstein-Barr* 3 e á NK(78). Está envolvida no primeiro passo da subcamada que sintetiza o CTP a partir da citidina. Este processo faz parte do metabolismo da pirimidina(76,78). Esta também pode ajudar no acumular de UTP para a transformação e proliferação. As moléculas NK estão associadas á degradação da proteína UCKL1, quando em contato com elas(78). O *splicing* alternativo resulta em várias variantes de transcrição. Esta proteína pode estar expressa em vários tecidos tais como baço, ovários e cólon(78). A

atividade catalítica das quinases uridinas é aumentada de 5 a 13 vezes em vários tumores como no do cólon, por exemplo. A baixa regulação do UCKL1 faz com que haja uma proliferação celular mais lenta e acabam por sofrer apoptose(77). As células com baixa quantidade de UCKL1 são mais propensas ao reconhecimento de células NK. (77). Um estudo procurou averiguar se UCKL-1 desempenhava um papel importante na sobrevivência das células tumorais. No entanto, o estudo demonstrou que havia uma maior expressão deste gene em pacientes de HCC mas que não aumenta significativamente a probabilidade de desenvolver cancro(77).

ARMC8 é uma proteína que contém repetições de tatu 8 e tem como função a regulação da migração celular, proliferação, manutenção tecidual, transdução de sinais e tumorigenese(79,80). Esta proteína está presente em vários tecidos mas tem maior expressão na medula óssea e tireoide. Também é expressa no cólon, mas em menor percentagem(81). Em estudos recentes mostram que *armc8* está envolvido na degradação da e-Catherine e cateninas associadas a tumores malignos(79). Já foi mostrado em estudos anteriores que o nível de expressão do *armc8* é maior no cancro do cólon quando comparado com os tecidos normais do cólon e por consequência está associado a um pior estágio, e prognóstico(79,82).

O gene *Gria 2* codifica o receptor inotrópico de glutamato AMPA subunidade tipo 2 que funciona como um canal de iões com ligação ao sistema nervoso central(83). É uma proteína importante na transmissão sináptica excitatória L- glutamato e atua como neurotransmissor excitatório em sinapses no sistema nervoso central, predominante no cérebro dos mamíferos e são ativados em uma variedade de processos neurofisiológicos normais(83). Quando entra em contato com o L-glutamato o canal de cálcio abre e converte o sinal químico em impulso elétrico(83). Vários estudos sugerem que a edição do pré- mARN é essencial para a função cerebral e a edição defeituosa do ARN pode ser relevante para etiologia lateral amiotrófica(84). A variante c.2292-426A>G encontrada neste gene, de acordo com o programa bioinformático leva á criação de um aceitador de *splicing*(84). Esta variante pode levar ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína. O *Splicing* alternativo leva a variantes de transcrição codificando diferentes isoformas(84).

O gene *MMS22* codifica a proteína de reparação de ADN. É uma proteína necessária para manter a integridade do genoma durante a replicação do ADN, reconhecendo e reparando o ADN(85). Esta proteína também se pode ligar á proteína associada á histona NFKBIL2 e ajuda a regular o estado da cromatina. Este gene está sobre expresso no cancro do pulmão e esófago(86). A variante c.607-216T>C no gene

MMS22, de acordo com o programa bioinformático, leva á destruição de um aceitador de *splicing*. Esta variante pode provocar uma disfunção da proteína ou a nível de estrutura, pode provocar a retenção de um intrão e, por consequência a proteína fica maior.

O gene GRIK2 codifica a proteína que é um receptor ionotrópico de L-glutamato pertencente á família kainate da subunidade 2(87). Estes são recetores neurotransmissores excitatórios em muitas sinapses do sistema nervoso central que existem predominantemente em mamíferos e são ativados em processos neurofisiológicos(87,88). A subunidade codificada por este gene está sujeita a alterações de ARN para alterar a estrutura e função do complexo do recetor. Variantes de *splicing* na transcrição já foram descobertas e estão associadas á incapacidade cognitiva autossómica recessiva. Este gene é expresso no cérebro, coração e outros 5 tecidos(88).

O gene PTPRK codifica o recetor tyrosina phosphatase tipo K. É uma proteína pertencente a uma subfamília de fosfatos de proteína ligada á membrana(89,90). Os PTP são moléculas de sinalização que regulam vários processos celulares tais como o crescimento celular, diferenciação e transformação oncogénica(91). Possui uma região extracelular, uma região transmembranar e dois domínios catalíticos(91). Estudos demonstraram que o PTPRK tem desfosforilado o receptor de fator de crescimento epidérmico (EGFR) e, posteriormente, proíbe a proliferação de células epidérmicas (EGF)-mediadas de células epiteliais e queratinócitos(89). Este gene encontra-se em vários tecidos como por exemplo no cólon(91). Outros estudos mostram que este gene age como um supressor de tumor em tumores endócrinos esporádicos e carcinoma intestinal juvenil(89). A desregulação do PTPRK aumenta a suscetibilidade às lesões intestinais, incluindo neoplasia intraepitenéial, adenoma e adenocarcinoma(92). A variante c.2333+1061A>G encontrada no gene acima referido, de acordo com o programa bioinformático leva á criação de um aceitador de *splicing* ou seja, esta variante levou ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho da proteína.

O gene AFDN codifica a proteína afadin que está envolvida na sinalização e organização de junções celulares durante a embriogénese(93). Também foi descrito como proteína de fusão do gene de leucemia linfoblástica aguda e leucemias mieloides agudas(94). Foram descritas várias variantes de *splicing* de transcrição onde codificaram diferentes isoformas do gene. Este gene encontra-se expresso no esófago, pulmão, cólon, entre outros(94).

O gene TAX1BP1 codifica a proteína vinculante tax1 1 que tem como função inibir apoptose induzida pelo TNF mediando a atividade anti-apoptótica TNFAIP3. Pode desempenhar um papel de sinalização de citocina pró-inflamatória(95). Foram encontradas variantes de *splicing* de transcrição codificando diferentes isoformas para o gene. Este gene encontra-se expresso em vários tecidos tais como na glândula adrenal, tireoide, colon, entre outros(96).

A variante c.2479G>T de acordo com o programa bioinformático leva a criação de um aceitador de *splicing*. Os efeitos desta variante podem ser observados na figura 5.2 onde se verifica uma deleção a partir do nucleótido 827 na proteína GAPVD1, ou seja, esta variante levou ao *splicing* de uma sequência adicional e, por consequência, a diminuição do tamanho do transcrito (97). A proteína GAPVD1 participa em processos de endocitose, internalização do recetor de insulina e no transporte de LC2A4/GLUT4 e ativa as GTPases quando existe troca de nucleótidos de guanina para a proteína rab31 relacionada com a ras(98,99). Na ausência de insulina o rab31 é mantido ativo e promove ciclos entre vesículas de armazenamento. Com a estimulação da insulina, a proteína é translocada para a membrana plasmática libertando LC2A4/GLUT4. Também têm função de transporte e degradação do EGFR(98). Encontra-se localizada no citoplasma, endossomas e membrana plasmática(99).

Proteína de referência

```

1  MVKLDIHTLA HHLKQERLYV NSEKQLIQLR NADVLKTAEK LYRTAWIAKQ QRINLDRLLI
61  TSAEASPAEC CQHAKILEDY QFVDGYKQLG FQETAYGEFL SRLRENPRLI ASSLVAGEKL
121 NQENTQSVIY TVFTSLYGNQ IMQDESEYLL QVLRYLIEFE LKESDNPRRL LRRGTCAFESI
181 LFKLFSEGLF SAKLFLTATL HEPIMQLLVE DEDHLETDPN KLIERFSPSQ QEKLFGKGS
241 DRFRQKVQEM VESNEAKLVA LVNKFIFYLK QNTYCFPHSL RWIVSQMYKT LSCVDRLEVG
301 EVRAMCTDLL LACFICPAVV NPEQYGIISD APINEVARFN LMQVGRLLQQ LAMTGSEEGD
361 PRTKSSLGKF DKSCVA AFLD VVIGGRAVET PPLSSVNLE GLSRTVVYIT YSQLITLVNF
421 MKSVMSGDQL REDRMALDNL LANLPPAKPG KSSLEMPY NTPQLSPATT PANKKNRLLPI
481 ATRSRRTNM LMDLHMDHEG SSQETIQEVQ PEEVLVSLG TGPQLTPGMM SENEVLNMQL
541 SDGGQDVPV DENKLGKPD KTLRFSLCSD NLEGISEGPS NRSNSVSSLD LEGESVSELG
601 AGPSGSGNVE ALQLEHEQA TTQDNLDDKL RKF EIRDMMG LTDDRDISET VSETWSTDVL
661 GSDFDPNIDE DRLQEIAGAA AENMLGSLLC LPGSGSVLLD PCTGSTISET TSEAWSVEVL
721 PSDSEAPDLK QEERLQELLES CSGLGSTSDD TDVREVSSRP STPGLSVVSG ISATS EDIPN
781 KIEDLRSECS SDFGGKDSVT SPDMEITHD FLYILQPKQH FQHIEAEADM RIQLSSSAHQ
841 LTSPPSQSES LLAMFDPLSS HEGASAVVRP KVHYARPSHP PDPPILEGA VGGNEARLNP
901 FGSHVLTPAE MEAFKQRHSY PERLVRSRSS DIVSVRRRPG SDPSWNRPRG NEERELPPAA
961 AIGATSLVAA PHSSSSSPSK DSSRGETEER KDSDBEKSDR NRPWNRKRFV SAMPKAPIPF
1021 RKKEKQEKDK DDLGPDFRST LTDDPSRSLA AQAQVAEDIL DKYRNAIKRT SPSDGAMANY
1081 ESTEVMDGGE SAHDSRDEA LQNISADDLP DSASQAHPQ DSAFSYRDAK KKLRLALCSA
1141 DSVAFVPLTH STRNGLPDHT DPEDNEIVCF LKVQIAEAIN LQDKNLMAQL QETMRCVCRF
1201 DNRTCRKLLA SIAEDYRKRA PYIAYLTRCR QGLQTTQAH LERLLQRVLR KEVANRYFTT
1261 VCVRLLESK EKKIREFIQD FQKLTAAADDK TAQVEDFLQF LYGAMAQDVI WQNASEEQLQ
1321 DAQLA IERSV MNRIFLAFY PNQDGDILRD QVLHEHIQRL SKVVTANHRA LQIPEVYLRE
1381 APHPSAQSEI RTISAYKTPR DKVQCILRMC STIMNLLSLA NEDSVPGADD FVPLVFLVLI
1441 KANPPCLLST VQYISSFYAS CLSGEESYWN MQFTAAVEFI KIIDRRK*

```

Proteína prevista a partir de sequência de codificação variante

```

1  MVKLDIHTLA HHLKQERLYV NSEKQLIQLR NADVLKTAEK LYRTAWIAKQ QRINLDRLLI
61  TSAEASPAEC CQHAKILEDY QFVDGYKQLG FQETAYGEFL SRLRENPRLI ASSLVAGEKL
121 NQENTQSVIY TVFTSLYGNQ IMQDESEYLL QVLRYLIEFE LKESDNPRRL LRRGTCAFESI
181 LFKLFSEGLF SAKLFLTATL HEPIMQLLVE DEDHLETDPN KLIERFSPSQ QEKLFGKGS
241 DRFRQKVQEM VESNEAKLVA LVNKFIFYLK QNTYCFPHSL RWIVSQMYKT LSCVDRLEVG
301 EVRAMCTDLL LACFICPAVV NPEQYGIISD APINEVARFN LMQVGRLLQQ LAMTGSEEGD
361 PRTKSSLGKF DKSCVA AFLD VVIGGRAVET PPLSSVNLE GLSRTVVYIT YSQLITLVNF
421 MKSVMSGDQL REDRMALDNL LANLPPAKPG KSSLEMPY NTPQLSPATT PANKKNRLLPI
481 ATRSRRTNM LMDLHMDHEG SSQETIQEVQ PEEVLVSLG TGPQLTPGMM SENEVLNMQL
541 SDGGQDVPV DENKLGKPD KTLRFSLCSD NLEGISEGPS NRSNSVSSLD LEGESVSELG
601 AGPSGSGNVE ALQLEHEQA TTQDNLDDKL RKF EIRDMMG LTDDRDISET VSETWSTDVL
661 GSDFDPNIDE DRLQEIAGAA AENMLGSLLC LPGSGSVLLD PCTGSTISET TSEAWSVEVL
721 PSDSEAPDLK QEERLQELLES CSGLGSTSDD TDVREVSSRP STPGLSVVSG ISATS EDIPN
781 KIEDLRSECS SDFGGKDSVT SPDMEITHD FLYILQPKQH FQHIEA*

```

Figura 5.2 Codificação da proteína normal à esquerda e codificação da proteína mutada à direita

6. Conclusões e Perspetivas Futuras

6.1. Conclusões do presente estudo

O objetivo deste trabalho era realizar uma análise exaustiva de mutações intrónicas profundas já identificadas e prever o efeito funcional dessas mutações através de uma análise *in silico*. Através do programa bioinformático, foi possível obter dados e dividi-los em dois grupos. O primeiro grupo são as variantes que ainda não foram reportadas clinicamente onde está inserida a variante intrónica c.221-11866G>A que corresponde ao gene RPS9. Este gene já foi observado em CCR, mas não existe ainda confirmação que a sua elevada expressão induz uma maior gravidade da doença.

O segundo grupo é constituído por variantes e genes que já foram reportados clinicamente. Neste grupo está incluída a variante c.2333+1061A>G correspondente ao gene PTPRK que regula vários processos celulares tais como o crescimento celular, diferenciação e transformação oncogénica e já foi encontrado no cólon. Outros estudos mostraram que este gene age como um supressor de tumor em tumores endócrinos esporádicos e carcinoma intestinal juvenil e que a desregulação do PTPRK aumenta a suscetibilidade às lesões intestinais, incluindo adenoma e adenocarcinoma. Outra variante que pertence a este grupo é a c.2331+1279G>A que corresponde ao gene ITGB1, que é um recetor de membrana envolvido em vários processos, como a adesão celular e reconhecimento em vários processos como a embriogénese, hemostasia, reparação de tecidos, resposta imune e difusão metastática de células tumorais. Foi demonstrado que estimula também a migração de células endoteliais e está envolvido na motilidade celular endotelial e angiogénese.

Existem diversas variantes envolvidas no CCR mas, para a maioria das variantes ainda não existem informações ou estudos realizados neste tipo de cancro em específico, demonstrando-se assim importante estudar mais aprofundadamente estas mutações e direcionar o estudo para o CCR, de modo a perceber se, por exemplo, poderão ser um biomarcador deste tipo de cancro ou que características e funções poderá ter neste cancro.

6.2. Perspetivas futuras

Realização de estudos a nível laboratorial para confirmação dos dados obtidos *in silico* das variantes. Seria também interessante estudar as variantes que ainda não têm significado clínico reportado de modo a perceber melhor as suas características e o que realmente as relaciona com o cancro colorretal.

7. Referências bibliográficas

1. Oriá RB, Anne G, Brito DEC, Digestório S, Paulo SÃO. Integração Básico-Clínica. 2016. 837 p.
2. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science* (80-). 2002;298(5602):2381–5.
3. OMS | Publicações da OMS. WHO. 2013;
4. What Is Cancer? - National Cancer Institute [Internet]. [cited 2021 Jul 19]. Available from: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
5. Melford john. A guide to cancer. 2018.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* [Internet]. 2011;144(5):646–74. Available from: <http://dx.doi.org/10.1016/j.cell.2011.02.013>
7. Cancro do Cólon e Reto | CUF [Internet]. [cited 2021 Jul 19]. Available from: <https://www.cuf.pt/saude-a-z/cancro-do-colon-e-reto>
8. Paulus J. Colorectal Cancer Facts and Figures 2020-2022. *Amfile*///C/Users/Ali/Downloads/introduction (2).docxerican Cancer Soc. 2020;66(11):1–9.
9. Cancro do Cólon e do Recto : Liga Portuguesa Contra o Cancro [Internet]. [cited 2021 Jul 19]. Available from: <https://www.ligacontracancro.pt/cancro-do-colon-e-do-recto/>
10. Mármol I, Sánchez-de-Diego C, Dieste AP, Cerrada E, Yoldi MJR. Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *Int J Mol Sci*. 2017;18(1).
11. Global Cancer Observatory [Internet]. [cited 2021 Jul 19]. Available from: <https://gco.iarc.fr/>
12. Schrijvers D, Aapro M, Zakotnik B, Audisio R, van Halteren H, Hurria A. ESMO handbook of cancer in the senior patient. *ESMO Handb Cancer Sr Patient*. 2010;iii–193.
13. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-vega F, et

- al. HHS Public Access. 2019;33(1):125–36.
14. Boyle P, Langman JS. Epidemiology. Bmj. 2000;321(7264):805.
 15. Pólipos do cólon | Programa Harvard Medical School Portugal [Internet]. [cited 2021 Jul 19]. Available from: <https://hmsportugal.wordpress.com/2012/03/30/polipos-do-colon/>
 16. Cancro colorretal: mortalidade | Hospital da Luz [Internet]. [cited 2021 Jul 19]. Available from: <https://www.hospitaldaluz.pt/pt/guia-de-saude/dicionario-de-saude/C/49/cancro-colorretal-mais-mortifero>
 17. Estadiamento - Cancro do Cólon e do Recto : Liga Portuguesa Contra o Cancro [Internet]. [cited 2021 Jul 19]. Available from: <https://www.ligacontracancro.pt/cancro-do-colon-e-do-recto-estadiamento/>
 18. Ministério da Saúde. Classificação de Tumores Malignos [Internet]. Uicc. 2004. 254 p. Available from: <http://www1.inca.gov.br/tratamento/tnm/index.asp>
 19. Sintomas - Cancro do Cólon e do Recto : Liga Portuguesa Contra o Cancro [Internet]. [cited 2021 Jul 19]. Available from: <https://www.ligacontracancro.pt/cancro-do-colon-e-do-recto-sintomas/>
 20. Sinais de Alerta - Cancro do Cólon e do Recto : Liga Portuguesa Contra o Cancro [Internet]. [cited 2021 Jul 19]. Available from: <https://www.ligacontracancro.pt/cancro-do-colon-e-do-recto-sinais-de-alerta/>
 21. ESMO. Cancro Colorretal: Um Guia Para O Doente. Anticancer Fund. 2016;1–51.
 22. Diagnóstico - Cancro do Cólon e do Recto : Liga Portuguesa Contra o Cancro [Internet]. [cited 2021 Jul 19]. Available from: <https://www.ligacontracancro.pt/cancro-do-colon-e-do-recto-diagnostico/>
 23. Colorectal Cancer Screening (PDQ®)—Patient Version - National Cancer Institute [Internet]. [cited 2021 Jul 19]. Available from: <https://www.cancer.gov/types/colorectal/patient/colorectal-screening-pdq>
 24. Rastreio Oportunístico do Cancro do Cólon e Reto – Normas de Orientação Clínica [Internet]. [cited 2021 Jul 19]. Available from: <https://nocs.pt/rastreio-oportunistico-cancro-colon-e-retos/>
 25. Markers T, Larynx IN. Marcadores tumorais no câncer da laringe. 2005;34(C):7–14.

26. Sisik A, Kaya M, Bas G, Basak F, Alimoglu O. CEA and CA 19-9 are still valuable markers for the prognosis of colorectal and gastric cancer patients. *Asian Pacific J Cancer Prev.* 2013;14(7):4289–94.
27. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet.* 2017;136(9):1093–111.
28. Lourenço D de A, Choupina AB. Bioinformática aplicada à caracterização de íntrons *Bioinformatics applied to the characterization of introns.* 2019;6(1):62–70.
29. Skoko N, Baralle M, Tisminetzky S, Buratti E. InTRONs in biotech. *Mol Biotechnol.* 2011;48(3):290–7.
30. Zweier. 基因的改变 NIH Public Access. *Bone* [Internet]. 2014;23(1):1–7. Available from:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>
31. Greaves M. Europe PMC Funders Group Evolutionary Determinants of Cancer. 2016;5(8):806–20.
32. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. *Nat Rev Dis Prim* [Internet]. 2015;1(February 2016):1–25. Available from: <http://dx.doi.org/10.1038/nrdp.2015.65>
33. Da Silva FC, Wernhoff P, Dominguez-Barrera C, Dominguez-Valentin M. Update on hereditary colorectal cancer. *Anticancer Res.* 2016;36(9):4399–406.
34. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* [Internet]. 2019;176(3):535-548.e24. Available from: <http://dx.doi.org/10.1016/j.cell.2018.12.015>
35. OSCP1 - Protein OSCP1 - Homo sapiens (Humano) - GENE E proteína OSCP1 [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q8WVF1>
36. Mutalyzer 2.0.34 — Verificador de nomes [Internet]. [cited 2021 Oct 29]. Available from: https://mutalyzer.nl/name-checker?description=NM_001330493.2%3Ac.121A%3EG
37. LY9 - T-lymphocyte surface antigen Ly-9 precursor - Homo sapiens (Human) - LY9 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q9HBG7-4>

38. rs772038455 RefSNP Report - dbSNP - NCBI [Internet]. [cited 2021 Oct 29]. Available from: https://www.ncbi.nlm.nih.gov/snp/rs772038455?horizontal_tab=true
39. Yue C, Liang C, Ge H, Yan L, Xu Y, Li G, et al. SUCO as a promising diagnostic biomarker of hepatocellular carcinoma: Integrated analysis and experimental validation. *Med Sci Monit.* 2019;25:6292–303.
40. SUCO - SUN domain-containing ossification factor precursor - Homo sapiens (Human) - SUCO gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q9UBS9>
41. SUCO SUN domain containing ossification factor [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=51430>
42. Subramani D, Alahari SK. Integrin-mediated function of Rab GTPases in cancer progression. *Mol Cancer* [Internet]. 2010;9(1):312. Available from: <http://www.molecular-cancer.com/content/9/1/312>
43. RABGAP1L - Rab GTPase-activating protein 1-like - Homo sapiens (Human) - RABGAP1L gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q5R372>
44. rs761982432 RefSNP Report - dbSNP - NCBI [Internet]. [cited 2021 Oct 29]. Available from: https://www.ncbi.nlm.nih.gov/snp/rs761982432?horizontal_tab=true#hgvs_tab
45. Tezcan G, Garanina EE, Zhuravleva MN, Hamza S, Rizvanov AA, Khaiboullina SF. Rab GTPase mediating regulation of NALP3 in colorectal cancer. *Molecules.* 2020;25(20):1–20.
46. Li G, Marlin MC. Rab-GTPases. *Encycl Ref Mol Pharmacol.* 2006;799–799.
47. ITGB1 integrin subunit beta 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=3688>
48. ITGB1 - Integrin beta-1 precursor - Homo sapiens (Human) - ITGB1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/P05556>

49. TCF7L2 transcription factor 7 like 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=6934>
50. Wenzel J, Rose K, Haghghi EB, Lamprecht C, Rauen G, Freißen V, et al. Loss of the nuclear Wnt pathway effector TCF7L2 promotes migration and invasion of human colorectal cancer cells. *Oncogene* [Internet]. 2020;39(19):3893–909. Available from: <http://dx.doi.org/10.1038/s41388-020-1259-7>
51. ADAM8 - Disintegrin and metalloproteinase domain-containing protein 8 precursor - Homo sapiens (Human) - ADAM8 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/P78325>
52. ADAM8 ADAM metalloproteinase domain 8 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=101>
53. KMT5B lysine methyltransferase 5B [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=51111>
54. KMT5B - Histone-lysine N-methyltransferase KMT5B - Homo sapiens (Human) - KMT5B gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q4FZB7>
55. ALG8 ALG8 alpha-1,3-glucosyltransferase [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=79053>
56. ALG8 - Probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase - Homo sapiens (Human) - ALG8 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q9BVK2>
57. DLG2 discs large MAGUK scaffold protein 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=1740>
58. DLG2 - Disks large homolog 2 - Homo sapiens (Human) - DLG2 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q15700>

59. SLC37A2 - Glucose-6-phosphate exchanger SLC37A2 - Homo sapiens (Human) - SLC37A2 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q8TED4>
60. SLC37A2 solute carrier family 37 member 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=219855>
61. CDADC1 - Cytidine and dCMP deaminase domain-containing protein 1 - Homo sapiens (Human) - CDADC1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q9BWV3>
62. PIGH - Phosphatidylinositol N-acetylglucosaminyltransferase subunit H - Homo sapiens (Human) - PIGH gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q14442>
63. PIGH phosphatidylinositol glycan anchor biosynthesis class H [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=5283>
64. CMTM3 - CKLF-like MARVEL transmembrane domain-containing protein 3 - Homo sapiens (Human) - CMTM3 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q96MX0>
65. RUNDC3A - RUN domain-containing protein 3A - Homo sapiens (Human) - RUNDC3A gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q59EK9>
66. RUNDC3A RUN domain containing 3A [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=10900>
67. ITGA3 - Integrin alpha-3 precursor - Homo sapiens (Human) - ITGA3 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/P26006>
68. ITGA3 integrin subunit alpha 3 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=3675>
69. RPS9 ribosomal protein S9 [Homo sapiens (human)] - Gene - NCBI [Internet].

- [cited 2021 Oct 29]. Available from:
<https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=6203>
70. Leveau C, Gajardo T, El-Daher MT, Cagnard N, Fischer A, de Saint Basile G, et al. Ttc7a regulates hematopoietic stem cell functions while controlling the stress-induced response. *Haematologica*. 2020;105(1):59–70.
71. TTC7A - Tetratricopeptide repeat protein 7A - Homo sapiens (Human) - TTC7A gene & protein [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.uniprot.org/uniprot/Q9ULT0>
72. TTC7A tetratricopeptide repeat domain 7A [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=57217>
73. 24_blood12510correspondencecombined 1673.
74. SF3B1 - Splicing factor 3B subunit 1 - Homo sapiens (Human) - SF3B1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.uniprot.org/uniprot/O75533>
75. SF3B1 splicing factor 3b subunit 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=23451>
76. UCKL1 - Uridine-cytidine kinase-like 1 - Homo sapiens (Human) - UCKL1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.uniprot.org/uniprot/Q9NWZ5>
77. Buivydiene A, Liakina V, Kashuba E, Norkuniene J, Jokubauskiene S, Gineikiene E, et al. Impact of the uridine–cytidine kinase like-1 protein and il28b rs12979860 and rs8099917 snps on the development of hepatocellular carcinoma in cirrhotic chronic hepatitis c patients—a pilot study. *Med*. 2018;54(5):1–10.
78. UCK2 uridine-cytidine kinase 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from:
<https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=7371>
79. Jiang F, Shi Y, Lu H, Li G. Armadillo repeat-Containing protein 8 (ARMC8) silencing inhibits proliferation and invasion in osteosarcoma cells. *Oncol Res*. 2016;24(5):381–9.

80. ARMC8 - Armadillo repeat-containing protein 8 - Homo sapiens (Human) - ARMC8 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/C9J8P8>
81. ARMC8 armadillo repeat containing 8 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=25852>
82. Gul IS, Hulpiau P, Sanders E, Van Roy F, Van Hengel J. Armc8 is an evolutionarily conserved armadillo protein involved in cell–cell adhesion complexes through multiple molecular interactions. *Biosci Rep.* 2019;39(8):1–12.
83. GRIA2 - Glutamate receptor 2 precursor - Homo sapiens (Human) - GRIA2 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/P42262>
84. GRIA2 glutamate ionotropic receptor AMPA type subunit 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2891>
85. MMS22L - Protein MMS22-like - Homo sapiens (Human) - MMS22L gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q6ZRQ5>
86. MMS22L MMS22 like, DNA repair protein [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=253714>
87. GRIK2 - Glutamate receptor ionotropic, kainate 2 precursor - Homo sapiens (Human) - GRIK2 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q13002>
88. GRIK2 glutamate ionotropic receptor kainate type subunit 2 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=2898>
89. Matsushita M, Mori Y, Uchiumi K, Ogata T, Nakamura M, Yoda H, et al. PTPRK suppresses progression and chemo-resistance of colon cancer cells via direct inhibition of pro-oncogenic CD133. *FEBS Open Bio.* 2019;9(5):935–46.
90. PTPRK - Receptor-type tyrosine-protein phosphatase kappa precursor - Homo

- sapiens (Human) - PTPRK gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q15262>
91. PTPRK protein tyrosine phosphatase receptor type K [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=5796>
 92. Świerczewska M, Sterzyńska K, Wojtowicz K, Kaźmierczak D, Iżycki D, Nowicki M, et al. PTPRK expression is downregulated in drug resistant ovarian cancer cell lines, and especially in ALDH1A1 positive CSCS-like populations. *Int J Mol Sci.* 2019;20(8).
 93. AFDN - Afadin - Homo sapiens (Human) - AFDN gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/P55196>
 94. AFDN afadin, adherens junction formation factor [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=4301>
 95. TAX1BP1 - Tax1-binding protein 1 - Homo sapiens (Human) - TAX1BP1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q86VP1>
 96. TAX1BP1 Tax1 binding protein 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=8887>
 97. Mutalyzer 2.0.34 — Name Checker [Internet]. [cited 2021 Oct 29]. Available from: https://mutalyzer.nl/name-checker?description=NM_015635.4%3Ac.2479G%3ET
 98. GAPVD1 - GTPase-activating protein and VPS9 domain-containing protein 1 - Homo sapiens (Human) - GAPVD1 gene & protein [Internet]. [cited 2021 Oct 29]. Available from: <https://www.uniprot.org/uniprot/Q14C86>
 99. GAPVD1 GTPase activating protein and VPS9 domains 1 [Homo sapiens (human)] - Gene - NCBI [Internet]. [cited 2021 Oct 29]. Available from: <https://www.ncbi.nlm.nih.gov/gene/26130>