



REDE NORDESTE DE BIOTECNOLGIA
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM
BIOTECNOLOGIA

PAULO CILAS MORAIS LYRA JUNIOR

**ANÁLISE FUNCIONAL DE NOVE SNPs DE
SUSCEPTIBILIDADE AO CÂNCER DE OVÁRIO NO
LOCUS 8q21**

VITÓRIA

2018

PAULO CILAS MORAIS LYRA JUNIOR

**ANÁLISE FUNCIONAL DE NOVE SNPs DE SUSCEPTIBILIDADE AO CÂNCER
DE OVÁRIO NO LOCUS 8q21**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Espírito Santo/RENORBIO, como requisito parcial para obtenção do título de Doutor em Biotecnologia, na área de concentração Saúde.

Orientadora: Prof^a. Dr^a. Leticia Batista Azevedo Rangel

Orientador estrangeiro: Dr. Álvaro Monteiro

VITÓRIA

2018

PAULO CILAS MORAIS LYRA JUNIOR

**ANÁLISE FUNCIONAL DE NOVE SNPs DE SUSCEPTIBILIDADE AO CÂNCER
DE OVÁRIO NO LOCUS 8q21**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia da Universidade Federal do Espírito Santo/RENORBIO, como requisito parcial para obtenção do título de Doutor em Biotecnologia, na área de concentração Saúde.

Aprovada em 19 de março de 2017.

COMISSÃO EXAMINADORA

Prof^a. Dr^a. Leticia Batista Azevedo Rangel
Universidade Federal do Espírito Santo
Orientadora

Prof. Dr. Valério Garrone Barauna
Universidade Federal do Espírito Santo

Prof^a. Dr^a. Renata Dalmaschio Daltoé
Universidade Federal do Espírito Santo

Prof^a. Dr^a. Etel Rodrigues Pereira Gimba
Universidade Federal Fluminense/INCa

**Prof. Dr. Marco César Cunegundes
Guimarães**
Universidade Federal do Espírito Santo

AGRADECIMENTOS

À minha filha, Marina, que irá nascer e que me motivou a vencer os obstáculos finais desta caminhada e concluir a contento esta etapa de suma importância em minha almejada carreira acadêmica;

À Luisa, pelo apoio pleno que possibilitou a realização de meu sonho acadêmico, prontificando-se a me acompanhar aos EUA para o fim;

Ao meu pai, Paulo Cilas Morais Lyra, e à minha mãe, Eni Guedes da Silveira Lyra, por me amarem incondicionalmente e me apoiarem em todas as situações e as circunstâncias difíceis que vivenciei ao longo desta jornada;

Aos meus irmãos, Pauline Guedes Lyra e Gustavo Guedes Lyra, pelo incentivo constante, sempre ao meu lado dentro de suas peculiaridades;

À minha família, em especial à minha tia Keila, por ter me ajudado em momentos fundamentais;

À Rosane Itaboraí por todo apoio recebido, como também ao meu primo Everton por seus sábios conselhos;

À professora Leticia Batista Azevedo Rangel que me adotou como filho. Por 10 anos me lapidou e se, hoje chego aqui, é porque ela esteve ao meu lado sempre;

Ao Dr. Álvaro Monteiro por ter acreditado em mim mesmo sem me conhecer. A oportunidade que me ofereceu mudou a minha vida por completo;

Ao Dr. Gustavo Mendoza, profissional incrível que tive a oportunidade de conviver. Ajudou-me, com dedicação e paciência, nuances experimentais durante minha permanência no exterior;

Ao Dr. Guillermo Gonzalez, por me hospedar em sua casa quando precisei e por me dar todo suporte para permanecer nos EUA;

À Xueli Li, à Carly Harro, à Cara Shields, à Olga Golubeva, à Anxhela Gustafson, à Natalia Sumi e à Rebekah Baskin, amigas que me acolheram no laboratório *Cancer Genetics and Genomics* no *Moffitt Cancer Center* e me apoiaram em todos os momentos que precisei de auxílio;

À Nayara Tessarollo, Isabella Guimarães, Taciane Barbosa, Diandra Zipinotti, Marcele de Souza, Ana Maria Rodrigues e Alan Branco, que fazem parte da família LBCMh e que alegraram os meus dias;

Ao *Moffitt Cancer Center* por possibilitar a realização de meu sonho acadêmico nos EUA;

Ao CNPq, por conceder a bolsa de doutorado sanduíche;
À FAPES, por conceder a bolsa de doutorado e a taxa de bancada;
À RENORBIO por me aceitar como aluno do programa;
À UFES pela estrutura disponibilizada;
À Deus, porque sem Ele em minha vida, eu não teria toda gratidão acima descrita.

RESUMO

O câncer de ovário (CaOV) configura como um câncer letal. Fatores genéticos contribuindo para o risco de desenvolvimento do CaOV têm sido investigados através dos estudos de associação ampla do genoma (GWAS), identificando *loci* de risco em diferentes regiões dos cromossomos, dentre eles o *locus* 8q21. Nesse estudo, realizamos uma análise funcional sistemática de nove SNPs candidatos para a causalidade ao CaOV no *locus* proximal ao gene CHMP4C. Após a caracterização da região para prováveis elementos regulatórios e genes associados, testamos os nove SNPs candidatos para atividade alelo específica para regiões com atividade de *enhancer*, como também testes para identificar prováveis fatores de transcrição. O SNP candidato localizado na região codificante do gene CHMP4C foi testado para instabilidade da proteína. Três SNPs foram identificados com funcionalidade alelo específica: rs35094336, rs137960856, rs1116683632. Este estudo elucidou o campo funcional da região 8q21 associado ao CaOV e identificou SNPs funcionais como possíveis mecanismos de associação ao risco de desenvolvimento da doença.

Palavras-chave: Câncer de ovário. GWAS. Fine Mapping. SNPs. CHMP4C.

ABSTRACT

Ovarian cancer (OC) is one of the most lethal cancer because of lack of early diagnosis. Genetic factors contributing to OC have recently been investigated through genome-wide association studies (GWAS), implicating in several independent OC risk loci in different chromosomal regions, among them the locus 8q21. Here, we performed an in-depth functional analysis of 9 SNPs causal candidates in the risk locus on 8q21 proximal to the CHMP4C gene. We first characterized the region for likely regulatory elements and genes associated, then we tested the 9 SNPs candidates for allele-specific activity in two different enhancer scanning assay, as well transcription factors associated. The SNP candidate localized in the coding region of CHMP4C gene was test for protein instability. Three SNPs were identified with allele-specific functionality: rs35094336, rs137960856, rs1116683632. These studies reveal the functional landscape of the 8q21 OC susceptibility locus and identify a network of functional SNPs as a possible mechanism driving OC risk association.

Key words: Ovarian cancer. Genome wide association study. Fine Mapping. SNPs. CHMP4C.

SUMÁRIO

1 INTRODUÇÃO	9
2 REVISÃO DA LITERATURA	11
2.1 EPIDEMIOLOGIA DO CAOVS	11
2.1.1 Classificação do CaOV	11
2.1.2 Diagnóstico do CaOV	13
2.2 ANÁLISE GENÔMICA PARA DETECÇÃO DE DOENÇAS	15
2.2.1 Estudos de associação ampla do genoma (GWAS)	18
2.2.2 Anotações e ensaios para elementos regulatórios	23
2.3 META-ANÁLISE E REPLICAÇÃO DO GWAS IDENTIFICA TRÊS NOVOS LOCI DE SUSCEPTIBILIDADE PARA O CÂNCER DE OVÁRIO	26
2.3.1 Dados moleculares e funcionais preliminares	29
2.4 ANÁLISE DE FINE-MAPPING NA REGIÃO 8q21 EM CASOS DE CÂNCER DE OVÁRIO.....	30
2.4.1 Identificação de SNPs de susceptibilidade candidatos a partir do Fine-Mapping	32
3 OBJETIVO GERAL	34
3.1 OBJETIVOS ESPECÍFICOS	34
4 MATERIAIS E MÉTODOS	35
4.1 ANÁLISE DE FRAGMENTOS GENÔMICOS COM ATIVIDADE DE ENHANCER	35
4.1.1 Análise estatística	36
4.2 ANÁLISE POR STARR-SEQ DE REGIÕES DO GENOMA COM ATIVIDADE DE ENHANCER.....	
4.2.1 Geração da biblioteca e preparação das amostras para col dados	37
4.2.2 Análise de dados in silico para os dados a partir do sequenciamento de nova geração	39
4.3 ANÁLISE DE FATORES DE TRANSCRIÇÃO COM ATIVIDADE ALELO DEPENDENTE POR ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA) ...	39

4.4 TRANSFEÇÕES ESTÁVEIS COM VETORES CONTENDO O GENE CHMP4C PARA AMBOS ALELOS.....	41
4.5 ANÁLISES DA ESTABILIDADE PROTÉICA DE CHMP4C ATRAVÉS DE WESTERN BLOT.....	42
5 RESULTADOS.....	43
5.1 MAPEAMENTO DOS 9 SNPS CANDIDATOS PARA ELEMENTOS FUNCIONAIS RELACIONADOS AO CÂNCER DE OVÁRIO	48
6 DISCUSSÃO	62
7 CONCLUSÃO	67
8 REFERÊNCIAS.....	68

1 INTRODUÇÃO

O câncer de ovário (CaOV) é o sétimo tipo tumoral mais comum entre as mulheres, configurando, paradoxalmente, a principal causa *mortis* por tumores ginecológicos (NEZHAT et al., 2015; PRAT, 2015). O carcinoma ovariano pode ser dividido em três grupos histológicos: epitelial, estromal e de células germinativas, sendo que o epitelial corresponde a 90% dos casos diagnosticados da doença (ROSEN et al., 2010). A agressividade do CAOVS está associada, dentre outros fatores, à ausência de métodos eficazes de diagnóstico, à natureza assintomática de sua fase inicial e à detecção tardia da doença (LI et al., 2012). De fato, as pacientes detectadas com CaOV em estágio inicial e responsivo à quimioterapia apresentam taxa de sobrevivência em torno de 80-90%. Contudo, tais mulheres representam apenas 15-20% de toda a população de mulheres com essa neoplasia (MODUGNO; EDWARDS, 2012). Em contraste, a maioria das mulheres é diagnosticada com CaOV em estágios avançados e já resistentes aos tratamentos quimioterápicos disponíveis na clínica.

Estudos realizados com grupos de familiares ou com gêmeos revelaram fortes indícios de componentes genéticos hereditários envolvidos no risco para o desenvolvimento do CaOV. Contudo, foi constatado que apenas 20% dos componentes hereditários estão relacionados a genes de moderada a alta penetrância funcional em células ovarianas normais, causando o CaOV (JERVIS et al., 2014; PHELAN et al., 2017). Diante desses achados, a ciência tem dedicado esforços na tentativa de elucidar a relação entre elementos hereditários e genes que podem contribuir para o desenvolvimento da doença. Nesse ínterim, estudos de sequenciamento em ampla escala e de associação genômica ampla (*genome-wide association studies* – GWAS) vêm sendo realizados, identificando, com êxito, centenas de *loci* cromossômicos associados ao câncer, incluindo o CaOV.

No que tange ao GWAS, tratam-se de estudos que comparam variantes genéticas comuns dentre uma população de estudo de número expressivo, divididos em casos (pacientes afetados/doentes) e controles (pacientes não afetados/sadios) com fins à determinação de variações genéticas potencialmente associadas à doença de interesse (SALE; MYCHALECKYJ; CHEN, 2009). Os resultados dos estudos supracitados são intrigantes, posto que correlações entre variantes

genéticas observadas foram consideradas, aparentemente, promissoras como agentes para o desenvolvimento da doença. Ainda, análises pós-GWAS são conduzidas para identificar SNPs (*single-nucleotide polymorphism*) que contribuem funcionalmente para o risco do desenvolvimento da doença, denominados SNPs causais.

O processo para identificação de SNP causal, a partir do GWAS, pode ser dividido em duas etapas, a saber: i) a primeira consiste em genotipagem mais detalhada do *locus* previamente identificado, processo denominado *fine-mapping*; ii) a segunda investiga como as variantes identificadas influenciam na funcionalidades celulares para o aumento do risco da doença de interesse. Essa etapa se baseia em testes de anotação funcional e experimentos *in vitro* / *in vivo* (SPAIN; BARRETT, 2015). O objetivo final desses estudos é, portanto, a potencial aplicação clínica dos SNPs estudados como fatores de risco para o desenvolvimento da doença de interesse em indivíduos carreadores do alelo de risco.

Em face ao exposto, o presente trabalho se propôs a realizar análise funcional sistemática, almejando identificar SNPs causais no locus de susceptibilidade 8q21, potencialmente associados ao risco aumentado de desenvolvimento do CaOV. O estudo teve início a partir dos dados gerados por Pharoah e colaboradores (2013), o qual identificou o SNP rs11782652 ($p = 5,5 \times 10^{-9}$), valendo-se de meta-análise de GWAS associado ao desenvolvimento de CaOV epitelial. Em seguida, o mesmo grupo realizou a análise de *fine-mapping* (dados não publicados) a partir do SNP rs11782652, o qual identificou nove SNPs candidatos à causalidade do CaOV. Neste contexto, o presente trabalho se prestou a realizar estudo *in vitro* com fins à caracterização da funcionalidade dos nove SNPs. A realização desse estudo possibilitou, por meio de dados inovadores e promissores, contribuir para o aprofundamento do entendimento dos mecanismos envolvidos no desenvolvimento e na progressão do CaOV, tal a fazer emergir possibilidades inéditas de estratégias diagnósticas e/ou terapêuticas contra a doença que segue desafiando pesquisadores e clínicos.

2 REVISÃO DA LITERATURA

2.1 EPIDEMIOLOGIA DO CAO V

Ao longo de duas décadas, a taxa de mortalidade de CaOV aumentou 1% ao ano em mulheres brancas e 0,4% em mulheres negras, tal que o CaOV configura a quinta causa de óbito por câncer entre as mulheres. Nesse cenário, nos Estados Unidos da América (EUA), somente no ano de 2017, mais de 50% das pacientes portadoras de CaOV foram a óbito em decorrência da doença (AMERICAN CANCER SOCIETY, 2017). No Brasil, dados do Instituto Nacional do Câncer (2016), indicaram que o CaOV ocupou a sétima posição de incidência de câncer entre as mulheres, com exceção do câncer de pele não melanoma. Por outro lado, a taxa de óbito ocupou a segunda posição por câncer ginecológico, atrás apenas do câncer de colo de útero. No Espírito Santo, há uma taxa de incidência de 5,78 por 100 mil habitantes. Quando observado apenas a capital do estado, Vitória, a taxa aumenta para 9,45 por 100 mil habitantes (SANTIAGO; NOVOA, 2016).

Digno de nota, o câncer é considerado uma doença do envelhecimento. De fato, nosso grupo de pesquisa publicou um estudo epidemiológico no Espírito Santo em que foi demonstrado uma prevalência maior de CaOV em mulheres com idade acima de 60 anos (PAES et al., 2011). Ainda em relação à prevalência do câncer em idade avançada, o Instituto Brasileiro de Geografia e Estatística publicou, no ano de 2000, estimativas de crescimento de 9% da população brasileira idosa em 50 anos. Conseqüentemente, à medida que a população envelhece, pode-se antever aumento do número total de casos de CaOV, aspecto com potencial de reflexo econômico negativo tanto sobre o Sistema Único de Saúde (SUS) quanto no sistema privado de saúde do país. Diante do cenário do CaOV, o qual aponta um futuro dramático, o Ministério da Saúde decretou, em 2010, o CaOV como doença prioritária no Brasil, por meio do qual aprova e regulamenta diretrizes diagnósticas e terapêuticas objetivando o manejo desta neoplasia (MAGALHÃES, 2013).

2.1.1 Classificação CaOV

Análises histopatológicas, imunohistoquímicas e genéticas permitem a distinção de cinco tipos principais de câncer de ovário epitelial (EOC), os quais configuram 98% dos carcinomas ovarianos, a saber: i) carcinoma seroso de alto

grau (HGSOC) (70% dos casos de CaOV diagnosticados); ii) endometrióide (ENOC) (10% dos casos de CaOV); iii) células claras (CCOC) (10% dos casos de CaOV); iv) mucinoso (MOC) (3% dos casos de CaOV); e carcinoma seroso de baixo grau (LGSOC) (5% dos casos de CaOV) (LEE *et al.*, 2003) (Figura 1). Com fins meramente informativos, posto que o presente trabalho tem como foco os cânceres de ovário epiteliais, há, ainda, os tumores de menor incidência de células germinativas malignas, disgerminomas, tumores do saco vitelínico e teratomas imaturos (PRAT, 2014).

Sabe-se que, dentre os tumores ovarianos, aproximadamente 10% dos casos de CaOV estão associados com alterações genéticas autossomais de baixa frequência hereditária que levam à predisposição à doença (BOYD, 2003; JELOVAC; ARMSTRONG, 2011). As mutações germinativas nos oncogenes do câncer de mama 1/2 (*BRCA1/2*) estão associadas a 5-15% do total dos casos de CaOV, enquanto que outros fatores genéticos conhecidos somados estão relacionados a, aproximadamente, 2% do total de casos (KIGAWA, 2013), mostrando um campo desconhecido, envolvendo riscos genéticos associados à doença, que necessita de um estudo mais aprofundado.

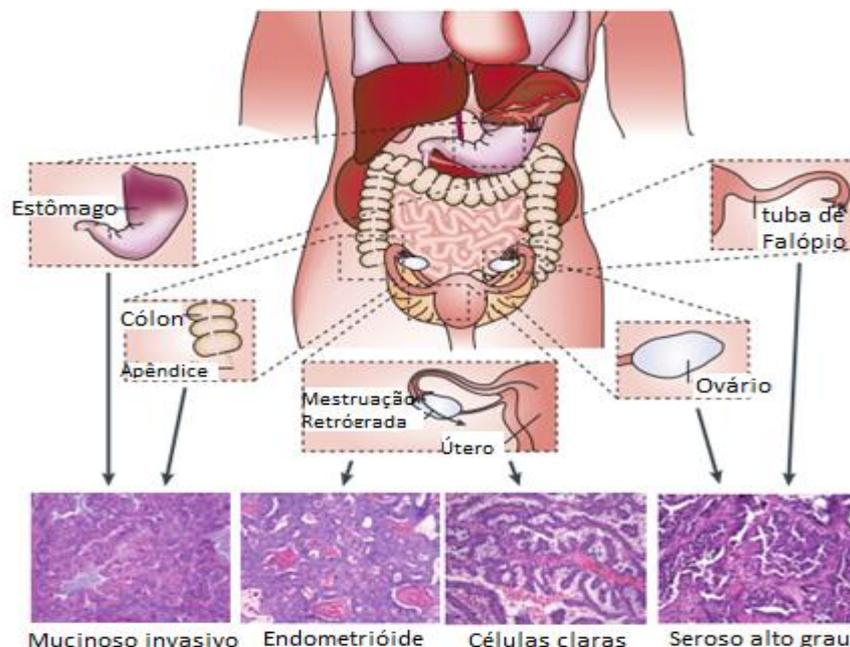


Figura 1: Desenho esquemático para os possíveis tecidos que dão origem aos subtipos de câncer de ovário epiteliais – Mucinoso invasivo, Endometrióide, Células claras e Seroso - com as respectivas imagens histológicas do tecido. Imagens histológicas cortesia R. Drapkin, Dana-Farber, Cancer Institute, EUA, e C. Crum, Brigham e Women Hospital, EUA. Adaptado VAUGHAN *et al.*, 2011.

2.1.2 Diagnóstico do CaOV

É consenso entre clínicos e pesquisadores que o diagnóstico precoce do câncer é fator decisivo para o curso da doença. Todavia, a inexistência de sinais e sintomas patognomônicos do CaOV associado aos métodos diagnósticos disponíveis na clínica, os quais são inespecíficos para a doença, a mesma é frequentemente diagnosticada em estadios avançados comumente correlacionados à ocorrência de metástases e de quimiorresistência. (MCCLUGGAGE, 2011; SOSLOW, 2008; SHIH & KURMAN, 2004). Esse é o caso de cerca de 85% das pacientes, sendo que apenas 15% das mulheres são diagnosticadas em estadios iniciais da doença, quando esta é restrita aos ovários, o que é condição de bom prognóstico. (GOFF et al., 2000).

Atualmente, o estadiamento do CaOV segue diretrizes estabelecidas pela FIGO (Federação Internacional de Ginecologia e Obstetrícia), revisada em 2012, dada a heterogeneidade da doença (PRAT, 2014) (Quadro 1). Apesar dos métodos de rastreio de mulheres assintomáticas não serem eficazes, o conhecimento dos sintomas pode ajudar a identificar a doença em fase precoce. Assim, alguns sintomas sugestivos incluem dor pélvica ou abdominal, urgência ou frequência urinária, inchaço e saciedade da fome precoce (GOFF et al., 2007). As mulheres que apresentam muitos desses sintomas, mais de doze vezes por mês, em um período menor que um ano, apresentam sensibilidade de detecção de CaOV de 56,7% para a doença em estadio inicial e 79,5% em avançado. A especificidade da detecção é de 90% para pacientes com mais de 50 anos e 86,7% para pacientes com menos de 50 anos. Dessa forma, a presença desses sintomas deve ser considerada no diagnóstico diferencial de CaOV (JELOVAC; ARMSTRONG, 2011).

Quadro 1: Estadiamento FIGO, porcentagem de casos e taxa de sobrevida em 10 anos segundo o estadiamento FIGO do CaOV. Adaptado JELOVAC; ARMSTRONG, 2011.

Estadiamento FIGO	Características	Distribuição (%)	Taxa de sobrevida em 10 anos (%)
I	Doença confinada no ovário ou tuba de falópio	20	73
IA	Tumor limitado em 1 ovário (cápsula intacta) ou tuba de falópio, sem ascites		
IB	Tumor em ambos os ovários (cápsula intacta) ou tuba de falópio, sem ascites		
IC	Estádio IA ou IB adicionado de ascites, cápsula rompida, tumor na superfície do ovário		
II	Tumor com metástase na pélvis	5	45
III	Doença confirmada na cavidade abdominal incluindo superfície do fígado, pélvis, para-aórtica, gânglios linfáticos e intestino.	58	21
IIIA	Nódulos linfáticos negativos adicionado de espalhamento microscópico na superfície peritoneal		
IIIB	Nódulos linfáticos negativos, implante peritoneal <2cm		
IIIC	Nódulos linfáticos positivos e/ou implante abdominal >2cm		
IV	Espalhamento para o parênquima do fígado, pulmão, pleura ou outros locais extra-abdominais.	17	<5

O diagnóstico de CaOV se inicia com a detecção de massa anexial durante exame pélvico, sendo a ultrassonografia, o teste de rastreio não invasivo mais utilizado para o fim. A recomendação para a cirurgia depende do grau de malignidade da massa pélvica associado a outros fatores, que incluem a idade da paciente, o início da menopausa, o histórico familiar, o nível sérico de CA-125 e a uni- ou bilateralidade do tumor (CANNISTRA, 2004).

Embora recomendada clinicamente, a dosagem de CA-125 ainda é questionável, visto que esse marcador pode estar elevado tanto em condições não-malignas, como endometriose e em casos de inflamação pélvica, bem como em doenças malignas, incluindo tumores do endométrio e do pâncreas (GUPTA; LIS, 2009). Logo, abordagens específicas com fins ao diagnóstico precoce e preciso do CaOV são de extrema urgência clínica, vislumbrando melhorias na qualidade de vida das portadoras da doença, bem como no aumento da taxa de sobrevivência das mesmas. Introduce-se, portanto, a possibilidade de aplicar ferramentas de análise genômica na pesquisa de marcadores ou padrões de expressão de genes potencialmente correlacionáveis ao desenvolvimento e/ou a progressão do CaOV.

2.2 ANÁLISE GENÔMICA PARA DETECÇÃO DE DOENÇAS

O genoma é de caráter único do indivíduo, de modo que a comparação entre genomas de humanos descreve diferenças inúmeras dentre os mesmos. Essas diferenças incluem desde pequenas variações individuais de nucleotídeos, os SNPs, bem como modificações mais amplas, como deleções, inserções e variações do número de cópias do gene no DNA. Seja qual for o evento genômico, há a possibilidade de se observar diferenças fenotípicas no indivíduo carreador de alterações quando comparado ao seu par com genoma selvagem.

Dentre as características fenotípicas abordadas, pode-se incluir o aumento dos riscos para o desenvolvimento de certas doenças (STRACHAN; READ, 2011). Até o ano 2000, antes da introdução dos GWAS, o método primário para investigação de variantes no genoma humano era por meio de estudos hereditários em casos de famílias portadoras de doenças hereditárias, que se mostrou eficaz para desordens de um único gene (ROPER, 2010). No entanto, para doenças comuns e complexas, os resultados foram de reprodutibilidade insuficiente

(STRACHAN; READ, 2011). Por outro lado, estudos baseados na metodologia de GWAS procuram, dentro de uma população, variantes genéticas com o objetivo correlacioná-los ao traço da doença de interesse. A diferença entre os estudos clássicos de hereditariedade genética e o novo método de GWAS está compilada na Figura 2.

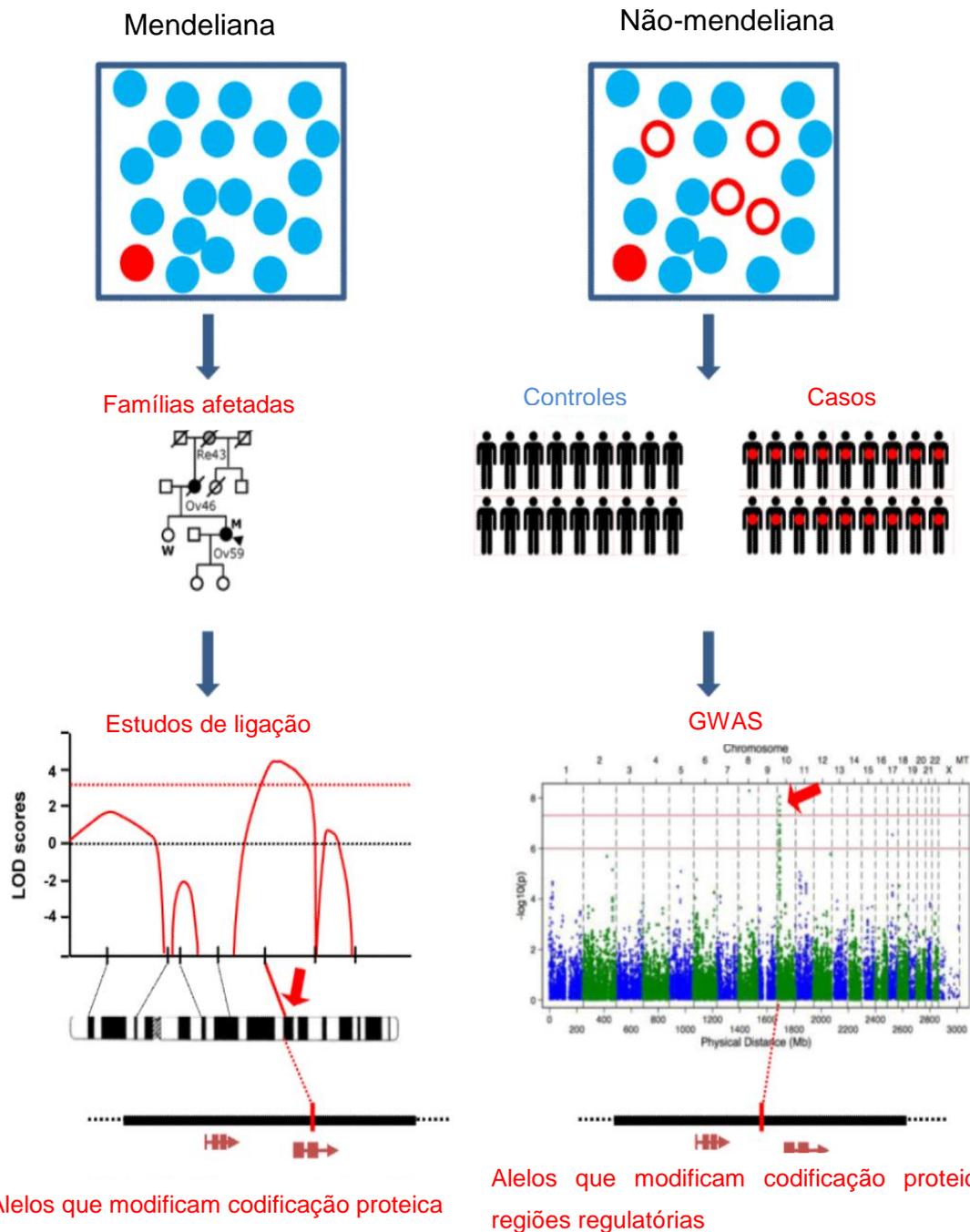


Figura 2 - Diferentes estratégias para identificação de loci de predisposição à doença. Em doença mendeliana, indivíduos que apresentam variantes de alta penetrância que conferem risco (círculo vermelho) tendem a expressar o fenótipo da doença (círculo com o centro vermelho), contudo são raros na população. Em doença não-mendeliana, os indivíduos que apresentam a variante comum de risco (círculo vermelho) tendem a ter baixa penetrância e, por isso, muitos desses não irão manifestar o fenótipo da doença (círculo com o centro branco). Doença mendeliana pode ser estudada por meio de ligação baseada em hereditariedade e o *locus* de risco é posicionalmente identificado (seta vermelha) (tipicamente em logaritmo da probabilidade (LOD) com *score* > 3 (linha pontilhada vermelha)). Doença não-mendeliana pode ser investigada com GWAS (sequenciamento de casos e controles para comparação) e o *locus* de risco é posicionalmente identificado (seta vermelha) (tipicamente com $p \leq 5 \times 10^{-8}$ (linha vermelha pontilhada)). Adaptado de MONTEIRO; FREEDMAN, 2014.

2.2.1 Estudos de associação ampla do genoma (GWAS)

No passado, os estudos dos SNPs estavam limitados à tecnologias disponíveis que resultavam na seleção de um único *locus* ou poucos *loci* por ciclo experimental, sendo necessário selecionar regiões específicas previamente. Para tanto empregavam-se critérios, tais quais: i) a relação descrita previamente com potencial para desencadear carcinogênese; ii) funções no metabolismo; iii) controle do ciclo celular; iv) síntese hormonal, dentre outros. Sendo assim, os estudos referidos eram limitados aos achados sobre SNPs associados a funções fundamentais para a homeostasia ou doenças previamente documentadas. Por outro lado, GWAS ampliou a possibilidade de descoberta de alvos gênicos inéditos ou vias possivelmente envolvidas com a patologia em questão (MONTEIRO; FREEDMAN, 2014). Desse modo, com o avanço tecnológico, o número de SNPs estudados em conjunto foi ampliado, tornando possível a inclusão de variantes comuns presentes no genoma. Assim, os estudos de associação com teste direto de polimorfismos genéticos em ampla escala têm se tornado um dos principais métodos da genética por ter o foco em alelos de baixa penetrância, diferentemente dos estudos anteriores de ligação genética (Figura 3).

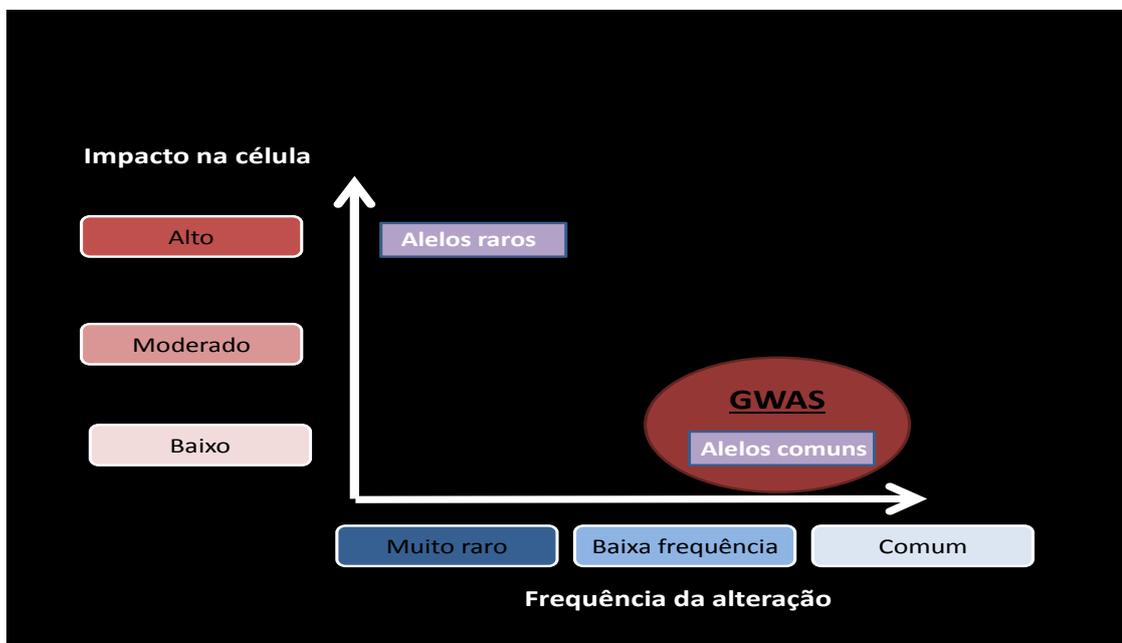


Figura 3 - Características das variantes de baixa penetrância (comuns) e de alta penetrância (raros). Os estudos clássicos de genética têm por objetivo estudar alelos raros – alteração com alto impacto na célula e com frequência rara na população. O foco do GWAS é fazer associações observando, principalmente, alelos comuns – alteração de baixo impacto na célula, mas encontrado com grande frequência na população (adaptado de LYRA-JUNIOR et al., 2017).

O GWAS é uma tecnologia emergente com capacidade de analisar o genoma humano por completo e identificar variantes comuns em uma determinada população associada a fatores genéticos de uma doença específica. O primeiro sucesso do GWAS foi a identificação de variantes no gene do Fator H do complemento como fatores de risco para a degeneração macular relacionado à idade (EDWARDS, 2005; HAINES, 2005; KLEIN, 2005). Dessa maneira, há evidências de que o GWAS possa prever indivíduos com maior probabilidade para o desenvolvimento de doenças e direcionar novas estratégias para prevenção e tratamento das mesmas (BUSH; MOORE, 2012).

A tecnologia *chip-based microarray technology*, que permite os estudos de GWAS, é executada em plataformas de genotipagem que podem avaliar de centenas a milhares de SNPs simultaneamente. As duas plataformas mais utilizadas para a maioria dos GWAS foram desenvolvidas pela Illumina (San Diego, CA) e Affymetrix (Santa Clara, CA). A Affymetrix imprime pequenas sequências de DNA, de determinado alelo, em um chip. Em seguida, o DNA das amostras é adicionado ao chip e nesse momento pode ocorrer a formação de dupla fita entre as sequências fixas e os fragmentos de DNA. Caso ocorra o pareamento, sondas marcadas, específicas para determinado alelo, completam a dupla fita formada anteriormente. Por fim, o sinal para detecção é amplificado e captado por um sensor. Nesse método, as variantes são detectadas pela diferença de hibridização das sondas às amostras. Todos os passos, com mais detalhes, da técnica estão ilustrados na Figura 4.

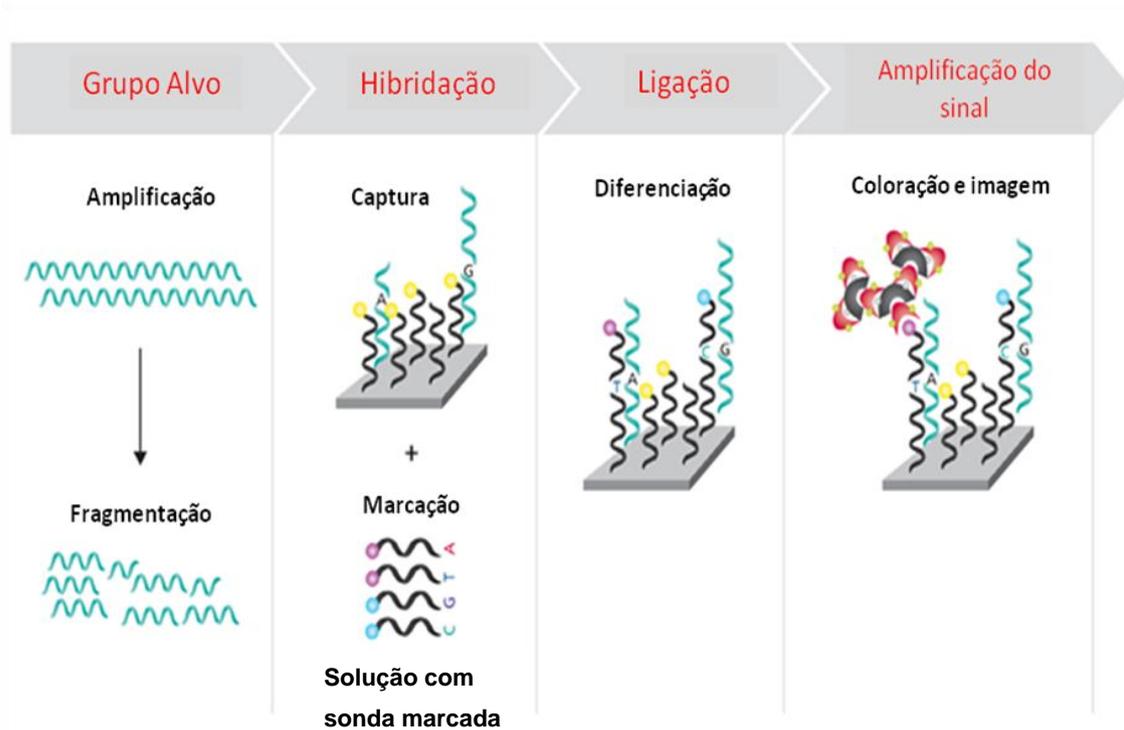


Figura 4 – Detalhamento metodológico com fins à detecção da variante presente na amostra a partir da tecnologia desenvolvida pela Affymetrix para GWAS. O material é inicialmente amplificado e fragmentado (fitas azuis), em seguida o material é colocado em uma plataforma (retângulo cinza) contendo determinada sequência de DNA (fita preta com círculo amarelo) juntamente com sondas marcadas (fitas pretas com círculos em azul ou roxo). Em seguida, realiza-se a ligação das sondas ao fragmento fixo na plataforma seguido da coloração (complexo com arcos vermelhos e cinza) e leitura do experimento com o intuito de determinação da variante presente. Adaptado de https://www.affymetrix.com/products_services/arrays/specific/axiom_mydesign.affx

A plataforma Illumina, por sua vez, utiliza uma tecnologia baseada em *bead*, na qual uma sequência longa de DNA é inserida a uma matriz sólida para detectar a variante. Após adicionar fragmentos da amostra à plataforma, ocorre a ligação, do fragmento derivado da amostra com a sequência complementar fixada no *bead* e há uma interrupção da dupla fita uma base antes do SNP alvo. A próxima etapa consiste em alongar a dupla fita interrompida, tendo como molde o DNA da amostra. Nucleotídeos marcados são incorporados e detectados através da coloração emitida por excitação de elétrons (Figura 5). Embora mais dispendiosa, a Illumina confere maior especificidade. Portanto, é possível conduzir estudos de associação usando conjuntos de SNPs para as variantes mais conhecidas no genoma, e assim, investigar, por associações, eliminando a necessidade de conhecimento prévio sobre a função das variantes selecionadas (BURTON et al., 2007; BUSH; MOORE, 2012).

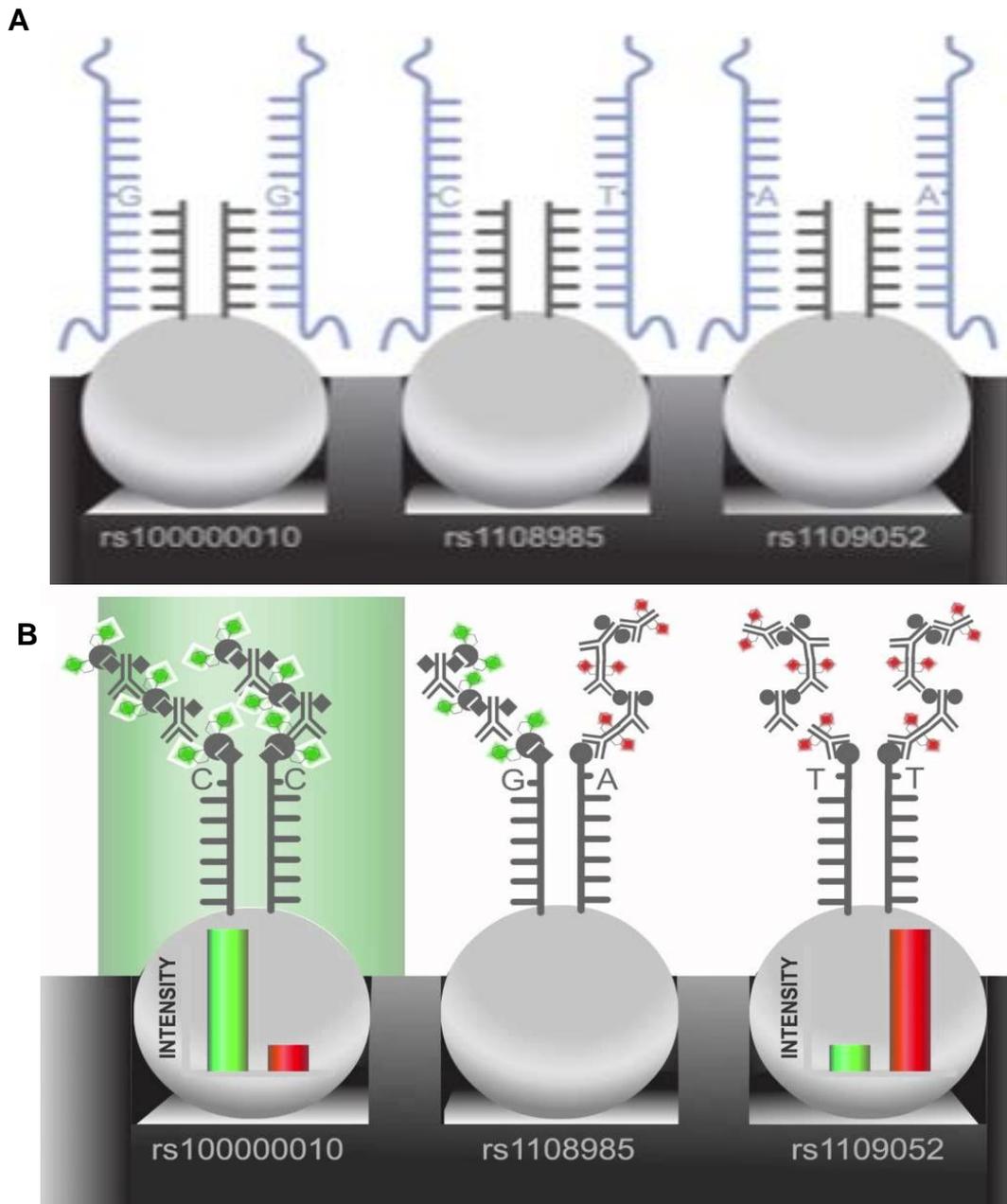


Figura 5 – Detalhamento experimental para a detecção da variante presente na amostra a partir da tecnologia desenvolvida pela Illumina. A) Na primeira etapa, um fragmento da sequência de interesse é fixado na plataforma por um *bead* (Fita em cinza). Em seguida, fragmentos da amostra (fitas azuis) são adicionados e há o pareamento específico formando uma dupla fita entre as sequências amostra/beads, interrompida no nucleotídeo correspondendo ao SNP de interesse. B) O nucleotídeo marcado incorporado à fita (quadrados verdes ou vermelhos), correspondente à variante de interesse, possibilitando a emissão de luminescência, tal a indicar o nucleotídeo inserido e determinando a variante presente na amostra. A figura mostra a detecção de três SNPs diferentes, sendo o primeiro identificado como homocigoto GG – marcação em verde para implementação do nucleotídeo citosina (C) na fita complementar, o segundo heterocigoto CT – marcação em verde para implementação do nucleotídeo guanina (G) e em vermelho para implementação do nucleotídeo adenina (A) na fita complementar e o terceiro homocigoto AA – marcação em vermelho para implementação do nucleotídeo timina (T) na fita complementar, para os SNPs rs10000010, rs1108985, rs1109052, respectivamente. (<https://www.illumina.com/science/technology/beadarray-technology/infinium-assay.html>).

Embora o GWAS se proponha a identificar variantes espalhados por todo o genoma relacionados à doença, há limitações no sequenciamento de todos os SNPs em todas as amostras. Para sobrepor o problema, os estudos utilizam de um fenômeno da genética denominado desequilíbrio de ligação (*linkage disequilibrium – LD*), que se refere a uma associação não ao acaso quando comparados alelos em dois *loci* (SLATKIN, 2008). A mensuração do grau de LD são comparações que observam a frequência de haplótipos em relação a uma frequência pré-determinada, essa se baseia na frequência de alelos que compõem os vários haplótipos (VANLIERE; ROSENBERG, 2008). Dessa maneira, ao selecionar um SNP chave, é possível determinar os alelos presentes no haplótipo pela correlação em LD das regiões. Os artigos de GWAS trazem o nome de imputação para os SNPs que são determinados sem que ocorra o sequenciamento direto, para tanto utiliza-se um painel referência de diferentes populações feita pelo projeto 1000 Genomes Project, representativo dos haplótipos mais frequentes (AUTON et al., 2015; MCVEAN et al., 2012).

Uma tendência dos SNPs com associação ao risco da doença é a localização em regiões não codificantes do genoma, igualmente distribuídas em regiões intrônicas e regiões intergênicas (BLATTLER et al., 2014; FREEDMAN et al., 2011). Dessa forma, a possibilidade de identificar SNPs em regiões de pouco impacto na funcionalidade celular, que não tenha uma relação direta no desenvolvimento da doença, torna-se maior. Em detrimento dos resultados, análises mais detalhadas de outros SNPs associados ao primeiro podem ser necessárias a fim de identificar SNPs causais, ou seja, SNPs que expliquem funcionalmente o aumento do risco à doença.

Existem dois métodos adotados para determinação dos SNPs causais. O primeiro é parte da estatística derivada da propriedade do LD, discutido anteriormente, em que o grau de associação (r^2) é calculado para obtenção dos SNPs mais relacionados ao SNP principal – ou tag-SNP. Sua utilização é controversa, uma vez que não há padrões definidos para exclusão/inclusão de SNPs, tornando o método demasiadamente empírico. O método mais apurado para detecção de SNPs associados ao tag-SNP é conhecido como *fine-mapping*. O *fine-mapping* emprega um ensaio de genotipagem denso que contém todos os SNPs comuns dentro do *loci* de risco identificados na primeira etapa do GWAS, o qual,

juntamente com imputações, permite que investigadores façam uma análise completa das regiões de risco (BROWNING; BROWNING, 2007; HOWIE; DONNELLY; MARCHINI, 2009; LI et al., 2010). A maior parte das análises de fine-mapping têm sido feitas por consórcios internacionais com interesse mútuo sobre uma doença específica; os exemplos são: the immunochip (TRYNKA et al., 2011), the metabochip (VOIGHT et al., 2012) e the iCOGs array (MICHAILIDOU et al., 2013).

2.2.2 Anotações e ensaios para elementos regulatórios

A caracterização das regiões regulatórias contendo os alelos de risco são fundamentais para o entendimento das possíveis modificações em funções que podem ocorrer na célula. Por essa razão, são rastreados elementos que sabidamente estão envolvidos de alguma forma no controle da expressão gênica. As sequências mais abundantes, dentre os elementos regulatórios, são de *enhancers*. Esses estão na fita de DNA, contendo entre 50 a 1500 pares de base, e que, quando ativados, aumentam a probabilidade da transcrição de um gene em particular ocorrer. Além disso, outras regiões regulatórias também inseridas no DNA, como os promotores, os *insulators* e os silenciadores, apesar de não serem tão abundantes, podem ser alvos de estudo por influenciar a funcionalidade celular (MASTON; EVANS; GREEN, 2006). Destaca-se o fato de que, com exceção de regiões promotoras específicas, os elementos regulatórios são, por muitas vezes, célula específica (HEINTZMAN et al., 2010). Isso explicaria, ao menos parcialmente, a obtenção de listas distintas de variantes associados à doença para cada GWAS realizado.

O estudo de modificações em histonas ou sensibilidade à DNase (ou hipersensibilidade) mostrou-se ferramenta eficaz para anotações de elementos regulatórios sendo tecido específicos (VISEL; ET AL., 2009) e mais informativo do que estudar sequências conservadas (BIRNEY et al., 2007; BLOW et al., 2010). Utilizar anotações da cromatina para identificar SNPs funcionais dentro de uma sequência regulatória do locus de susceptibilidade, tem sido bastante proposto pela área científica (COETZEE et al., 2010). Além das anotações de regiões reguladoras, é possível obter mais informações do locus de interesse através de associações dos fatores de transcrição com os elementos responsivos. Tanto modificações de histona

como de regiões de ligação dos fatores de transcrição, são atualmente identificados utilizando a metodologia de sequenciamento de cromatina imunoprecipitada (chromatin immunoprecipitation sequencing – ChIP-seq). Os sinais captados, normalmente, correspondem a sequências de DNA inferiores a 1kb para análises detalhadas. Para atividade de *enhancers* em regiões regulatórias pode-se utilizar genes-reporter tanto *in vitro* (JIA et al., 2009b) quanto *in vivo* (VISEL; ET AL., 2009).

Integrar o conhecimento das sequências regulatórias que estão nos *loci* de risco com o catálogo de SNPs associados ao risco, talvez seja a maneira mais eficiente de priorizar tanto os locais de regulação quanto as variantes funcionais mais prováveis. Em um estudo, dois SNPs no cromossomo 8q24 foram associados ao risco de câncer de próstata e ao câncer colorretal, sendo que um SNP está localizado no sítio de ligação FoxA1 e, curiosamente, muitas regiões de *enhancers* transcricionais foram identificadas no *locus* 8q24, sendo dois deles associados aos receptores de andrógenos (JIA et al., 2009a). Nesse caso, é postulado que o alelo de risco para o câncer de próstata alterou a funcionalidade de FoxA1 o que pode ter resultado na forte responsividade ao tratamento com andrógeno.

Em um estudo em paralelo, um SNP no *locus* 8q24 para câncer colorretal, encontrado também em região com anotação para *enhancer* transcricional, foi testado quanto às atividades alelo-dependente. Como resultado, foi mostrado que a região contendo o SNP interagia fisicamente com o proto-oncogene MYC com ligação alelo dependente do fator de transcrição 7-like 2 (TCF7L2) (POMERANTZ et al., 2009). Apesar de encontrar essa forte relação, não houve aprofundamento a partir dos dados iniciais. O trabalho prossegue na tentativa de elucidar com mais detalhes a funcionalidade dos SNPs. Para tanto, poderia ser empregado métodos bioquímicos que avaliam a diferença na ligação de fatores de transcrição e atividade, como por exemplo ChIP-seq ou *electrophoretic mobility shift assay* (EMSA). Assim, as sequências regulatórias contendo SNPs funcionais podem ser associadas à genes alvos.

Uma vez que há evidências suficientes que apontem genes candidatos de susceptibilidade, mais estudos detalhados de funcionalidade são necessários para caracterizar o papel dos genes na patogênese da doença que está sendo considerada. Obter um melhor entendimento dos mecanismos biológicos do

desenvolvimento do câncer geralmente recai em modelos de análise que refletem a doença e na aplicação de tecnologias que facilitem a análise desses modelos. Os estudos de funcionalidade se baseiam principalmente em: 1) identificação do SNP principal (ou tag-SNP) e os possíveis SNPs causais associados, 2) genes candidatos relacionados aos *loci* de susceptibilidade, 3) variações funcionais alelo dependente. A Figura 6 resume os passos básicos para o estudo funcional no contexto do GWAS. Dessa forma, acredita-se que é possível alcançar aplicação clínica a partir dos dados obtidos.

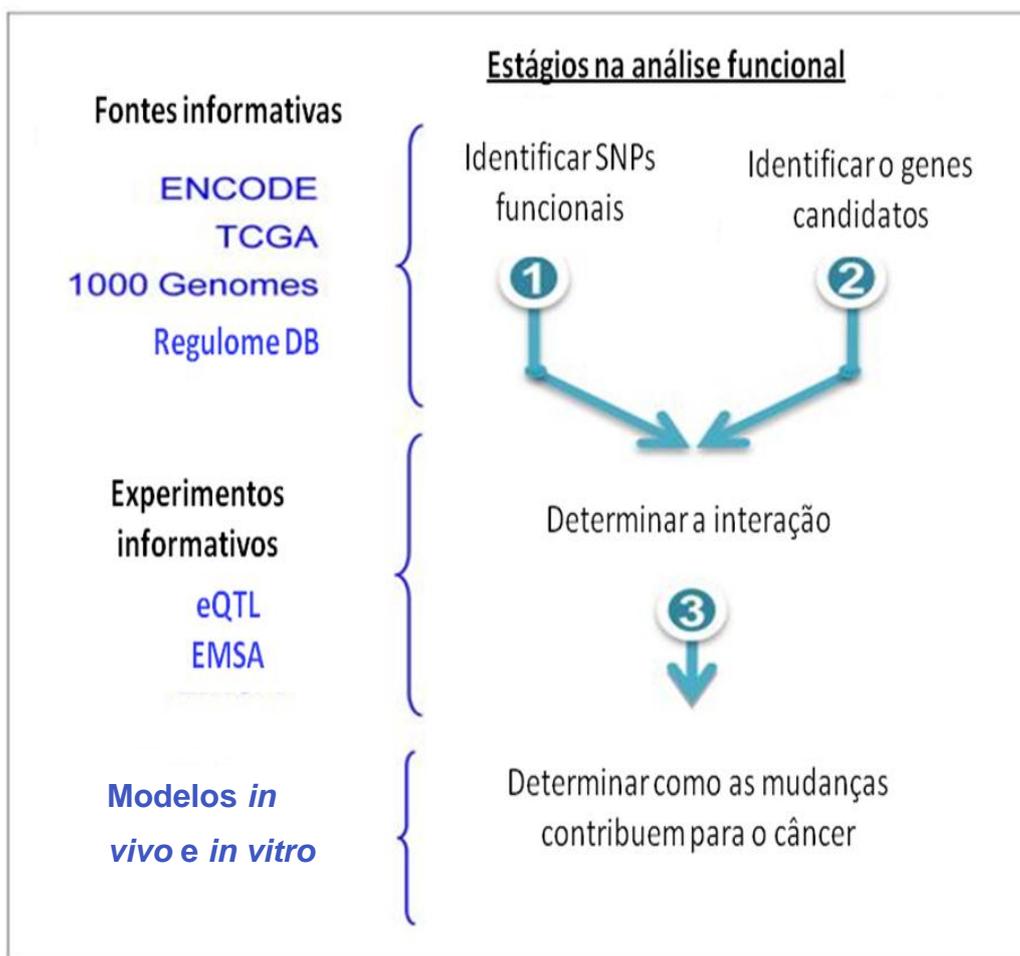


Figura 6 – Os três estágios da análise funcional. O primeiro estágio é a identificação de SNPs funcionais (a partir do r^2 proveniente do LD ou fine-mapping, quando disponível para a região). No mesmo íterim, são identificados os genes candidatos no *locus*. Utiliza-se banco de dados para obter informações publicamente disponíveis a respeito da região - por exemplo TCGA, 1000 Genomes, Regulome DB. No segundo estágio são realizados experimentos que acrescentam mais informações a respeito das atividades funcionais ou regulatórias do *locus* de susceptibilidade – por exemplo eQTL e EMSA. Por último, são realizados modelos *in vitro* e *in vivo* para definir as mudanças que de fato contribuem para o desenvolvimento da doença. No caso, o foco do presente estudo é o CaOV. Adaptado de MONTEIRO; FREEDMAN, 2014.

2.3 META-ANÁLISE E REPLICAÇÃO DO GWAS IDENTIFICA TRÊS NOVOS *LOCI* DE SUSCEPTIBILIDADE PARA O CÂNCER DE OVÁRIO

Pharoah e colaboradores (2013) realizaram um estudo em que foram retirados dados de dois GWAS, um dos Estados Unidos e outro do Reino Unido. O estudo nos Estados Unidos foi composto por quatro estudos independentes de caso-controle, incluindo 1.952 casos de EOC e 2.052 controles. O outro estudo, consiste em um GWAS multi-center de duas fases que incluiu 1.817 casos de EOC e 2.354 controles na primeira fase e 4.162 casos de EOC e 4.810 controles na segunda fase. Com base nesses estudos, foram selecionados 24.551 SNPs para compor uma nova plataforma de genotipagem, a iCOGS. Foram analisadas as associações para, aproximadamente, 2.5 milhões de SNPs incluindo os diretamente genotipados e imputados. Ademais, foram selecionados 24.551 SNPs associados com o risco para todas as histologias (11.647 SNPs) ou câncer de ovário seroso (12.904 SNPs) baseados na classificação do valor de p , em que p era menor que $5,0 \times 10^{-8}$.

Após aplicar todos os filtros para o controle de qualidade, foram testados 22.252 SNPs para associação com o risco de EOC invasivo e para EOC seroso invasivo em 18.174 casos de EOC (10.316 casos de seroso) e 26.134 controles. As análises primárias foram baseadas em indivíduos de ancestrais europeus (16.283 casos e 23.491 controles). Associações previamente reportadas foram confirmadas (SNPs em 2q31, 3q25, 8q24, 9p22, 17q21 e 19p13). Foram identificados três novos *loci* com significância genômica ($p < 5 \times 10^{-8}$) (Gráfico 1), sendo dois deles significantes para todos EOC invasivos (8q21 e 10p12) e o outro apenas para EOC seroso invasivo (17q12)(PHAROAH et al., 2013).

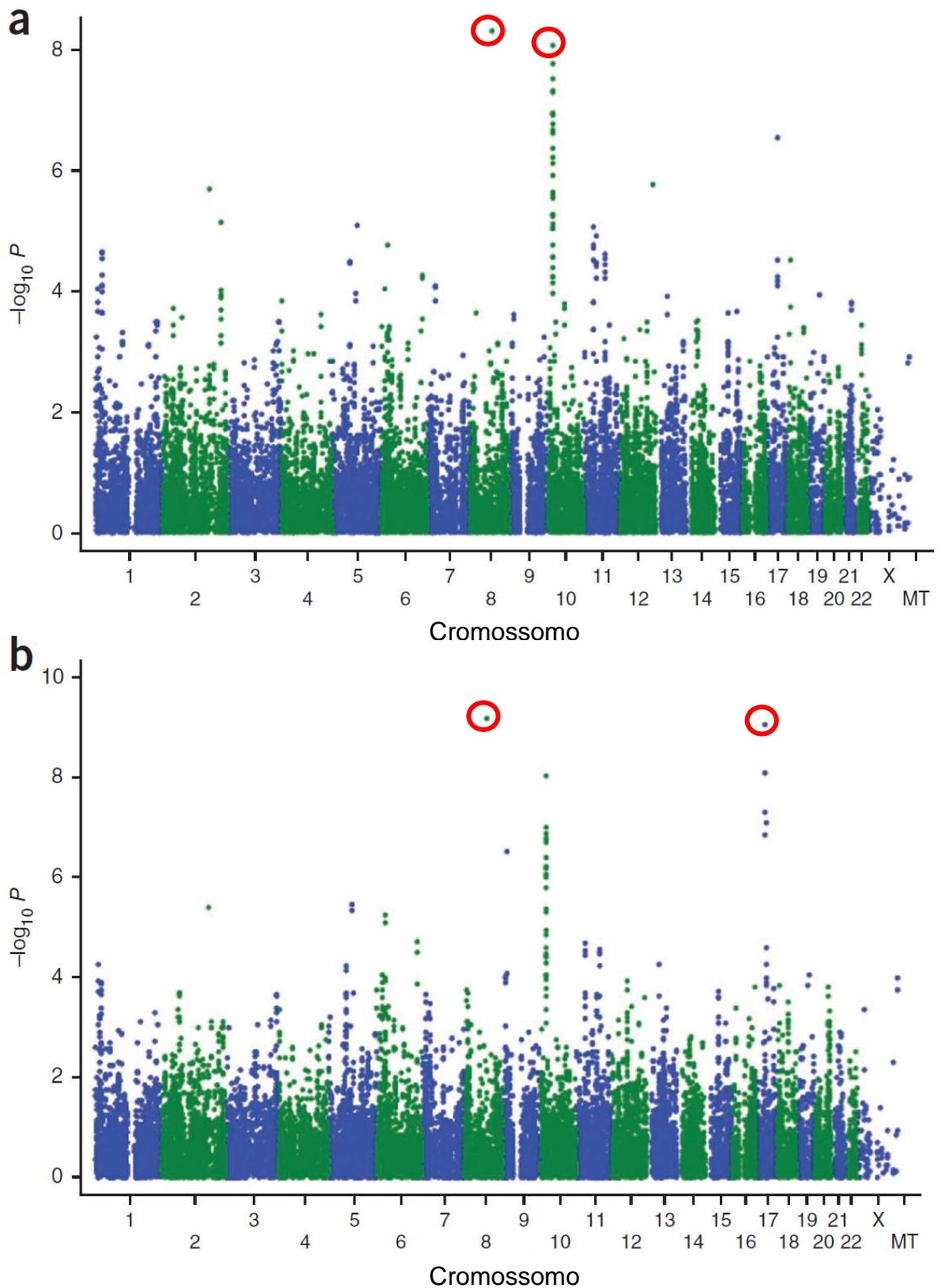


Gráfico 1 – A associação de SNPs genotipados no risco de desenvolvimento do câncer de ovário. Gráfico de Manhattan plot mostrando a associação entre genótipos de 22.076 SNPs e o risco com todos os subtipos de EOC invasivo (a), identificado 2 *loci* (8q21 e 10p12), em que os SNPs obtiveram valor de $p < 5 \times 10^{-8}$ e subtipo seroso invasivo de EOC separadamente (b) identificando um *locus* 17q12 em que o SNP obteve valor de $p < 5 \times 10^{-8}$ (PHAROAH et al., 2013).

A associação mais relevante para todos os EOC invasivos foi do SNP rs11782652 no *locus* 8q21 (Razão de possibilidades (OR) = 1,19, Intervalo de confiança (CI) de 95%: 1,12-1,26, $p=5,5 \times 10^{-9}$). Ademais, o SNP apresentou uma variação pequena na associação por subtipo histológico ($p=0,0002$) com maior grau de associação no subtipo seroso (Tabela 1) (PHAROAH et al., 2013). No intuito de verificar possíveis SNPs causais associados ao SNP rs11782652, foi realizada análise do haplótipo através dos dados disponíveis no HapMap. Não foram identificados SNPs com alto grau de associação ao SNP rs11782652 na região de, aproximadamente, 70 kb (chr8:82,598,658-82,599,158-82,669,813) (Figura 7).

Tabela 1- Odds Ratio e teste de associação por subtipo histológico na população mais associada ao SNP da região 8q21. Adaptado de PHAROAH et al., 2013.

SNP	Locus	Alelo Ref/Alt	Frequência do alelo de risco	Subtipo	Ancestral	OR	CI 95%	Valor de p
rs11782652	8q21	A > G	0.07	Todos invasivo	Europeu	1,19	1,12-1,26	$5,5 \times 10^{-9}$
				Seroso	Europeu	1,24	1,16-1,33	$7,0 \times 10^{-10}$
				Endometrióide	Europeu	1,04	0,92-1,19	0,50
				Células claras	Europeu	1,12	0,95-1,33	0,18
				Mucinoso	Europeu	0,95	0,78-1,15	0,58
				Seroso	Asiático	-	-	-
				Seroso	Africano	1,73	0,92-3,28	0,91
				Seroso	Misturado	1,15	0,81-1,64	0,43

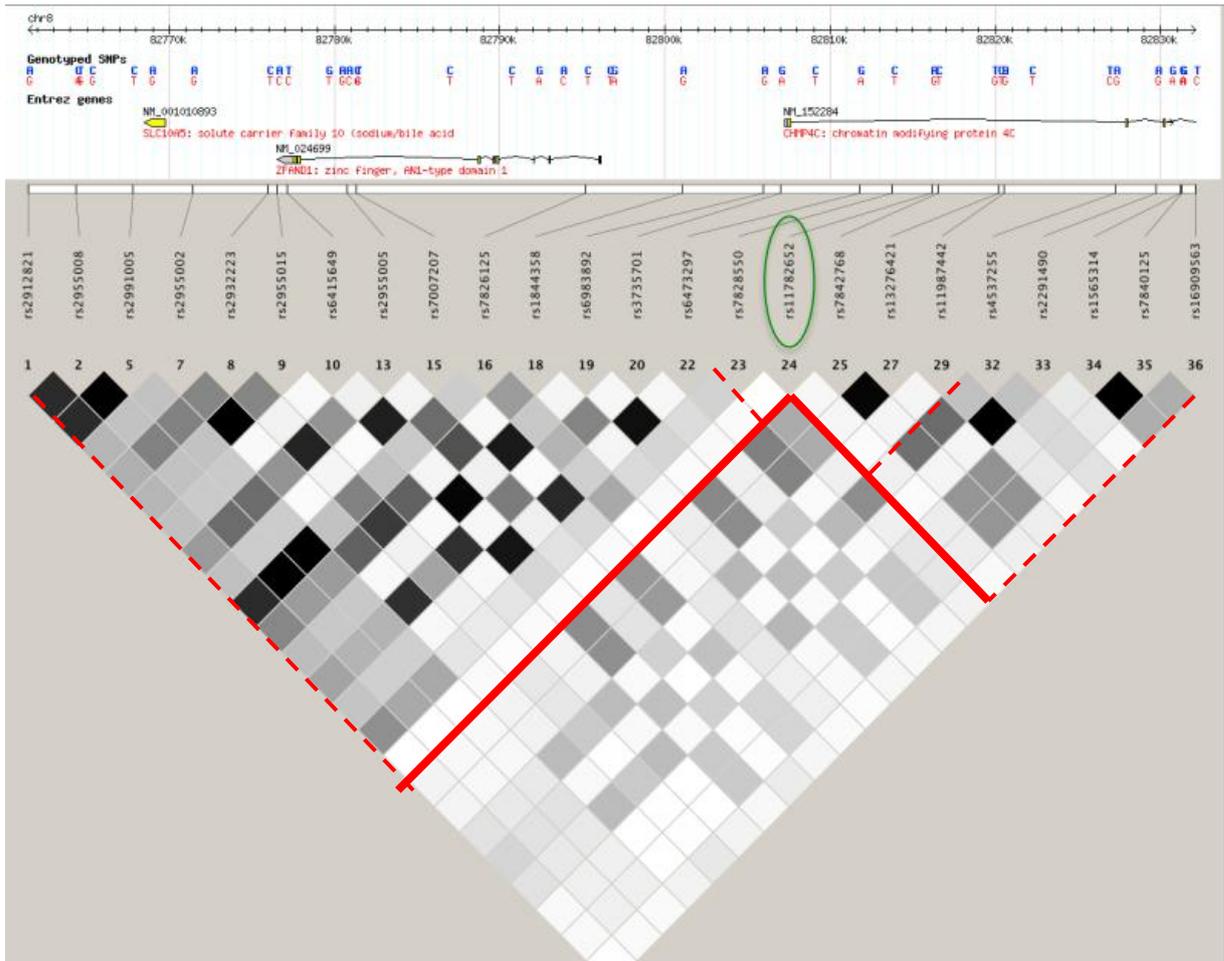


Figura 7- Linkage Disequilibrium – r^2 plot para grau de associação dos SNPs no haplótipo da região 8q21 (chr8:82,598,658-82,599,158-82,669,813) a partir do banco de dados do HapMap. A associação do SNP rs11782652 aos SNPs adjacentes é observada ao traçar uma linha reta a partir do SNP principal (linha contínua vermelha) e outra linha reta a partir do SNP secundário (linha tracejada vermelha). O ponto de intercessão das linhas forma um quadrado, a cor do quadrado quanto mais escuro (preto) mais forte a associação e quanto mais claro (branco) menor a associação. O tag-SNP rs11782652 apresenta baixa associação com os SNPs adjacentes (PHAROAH et al., 2013).

2.3.1 Dados moleculares e funcionais preliminares

O SNP rs11782652 associado ao risco de CaOV está localizado em região não codificante do DNA, íntron 1 do gene CHMP4C. Como dito anteriormente, o tag-SNP pode atuar como um marcador para o verdadeiro SNP causal, o qual pode interferir em diversas funções celulares dependendo da localização do SNP no genoma. Para identificar o possível SNP causal e gene alvo no *locus* 8q21, foi avaliado o papel funcional de SNPs e genes em uma região de 1Mb centralizando o SNP rs11782652 na região (PHAROAH et al., 2013).

Inicialmente foram imputados genótipos da região no banco de dados do projeto *1000 Genomes* e testados para associação todas as variantes de menor

frequência que apresentaram *Minor allele frequency* (MAF) > 0,02. Foi comparado o logaritmo de probabilidade dos modelos de regressão, em que oito SNPs (rs11782652, rs74544416, rs787400051, rs78724141, rs74758321, rs137960856, rs76837345, rs35094336) foram considerados fortemente associados (equivalente à probabilidade de 1000:1) como possíveis candidatos para SNP causal (PHAROAH et al., 2013).

Não foram encontradas evidências de correlação entre o genótipo rs11782652 e expressão gênica em células normais de ovário e células epiteliais da tuba de falópio para os nove genes na região (FABP5, PMP2, FABP4, FABP12, IMPA1, SLC10A5, ZFAND1, CHMP4C e SNX16), mas houve uma associação estatisticamente significativa entre o rs11782652 e a expressão de CHMP4C em tecidos primários de EOC ($p = 3,9 \times 10^{-4}$) e linfonodos transformados ($p = 0,012$). Também foram encontradas evidências de associação para o rs11782652 e a metilação da região (*methylation quantitative trait locus* (mQTL)) para três genes em tecidos tumorais: ZFAND1 ($p = 0,003$), CHMP4C ($p = 0,001$) e SNX16 ($p = 0,001$). No entanto, a metilação do gene foi correlacionada apenas com a expressão do gene CHMP4C (PHAROAH et al., 2013).

No geral, os dados do artigo sugerem que oito SNPs podem estar envolvidos no desenvolvimento do CaOV e vários genes no *locus* 8q21 podem atuar no desenvolvimento somático do EOC. No entanto, as evidências indicam fortemente que CHMP4C (*encoding chromatin-modifying protein 4C*) é o provável gene candidato para a susceptibilidade ao CaOV.

2.4 ANÁLISE DE FINE-MAPPING NA REGIÃO 8q21 EM CASOS DE CÂNCER DE OVÁRIO

O estudo de *Fine-mapping* da região 8q21, desenvolvido por Pharoah e colaboradores (dados não publicados), foi feito como parte do COGS (Collaborative Oncologic Gene-Environment Study) com o objetivo de identificar os SNPs com maior probabilidade de causar o câncer de ovário no *locus*. Todos os SNPs dentro do intervalo de 1Mb no *locus* 8q21 foram identificados a partir do projeto 1000 genomas (DURBIN et al., 2010) e HapMap III (The International HapMap Consortium, 2003). Apenas as variantes com alelo de menor frequência encontrado

pelo menos duas vezes no projeto 1000 genomas e com pontuação > 0.8 em Illumina Design foram incluídos. Além disso, somente as variantes correlacionadas com o SNP de risco ($r^2 > 0.1$) foram selecionadas para a genotipagem, novamente a plataforma utilizada foi a iCOGS.

O ensaio iCOGS genotipou os casos e os controles de 43 estudos que constituem o Ovarian Cancer Association Consortium (OCAC) previamente descrito por Pharoah, (2013). Em conjunto com os dados OCAC iCOGS, também foram utilizados dados de outros estudos de GWAS para câncer de ovário a fim de aumentar o poder estatístico. As amostras foram excluídas caso: i) indivíduos sem ancestral europeu; ii) valores de genotipagem inferiores a 95%; iii) baixa ou alta heterogeneidade ambígua; iv) se encontrado marcadores para o cromossomo Y ou sexo ambíguo. Ainda, foram excluídos indivíduos para cada par de amostras por serem parentes de primeiro grau ou amostras duplicadas entre iCOGS e qualquer outro GWAS previamente realizado. Ademais, os SNPs foram excluídos caso fossem monomórficos, apresentassem valores de identificação menores que 95% e mostrassem desvios do equilíbrio de Hardy-Weinberg ($p < 1 \times 10^{-7}$ ou $p < 1 \times 10^{-12}$ em controles) ou baixa concordância entre as duplicatas (<98%). Para o estudo, também foram excluídos SNPs raros (Frequência do alelo recessivo (MAF) < 1% ou contagem alélica < 5).

Assim, o grupo final genotipado consistiu em 11.069 casos e 21.722 controles do COGS (OCAC-iCOGS), 2.165 casos e 2.564 controles de GWAS norte americano, 1.762 casos e 6.118 controles de GWAS do reino unido e 441 casos e 441 controles da clínica Mayo. Todas as pessoas analisadas tiveram o termo de consentimento assinado, com informações e amostra de sangue, sob protocolos aprovados pela instituição executora. Em geral, os estudos de 11 países somaram 15.437 casos de mulheres diagnosticadas com EOC invasivo, sendo 9.627 casos com subtipo seroso e 30.845 controles.

Para haver uma maior cobertura da genotipagem, foram imputados os genótipos de SNPs incluídos em GWAS anteriores. Foi feita a imputação separadamente das amostras do iCOGS e dos GWAS. Os dados das variantes imputadas foram provenientes do *1000 Genome Project* usando a versão 3 de abril

2012 (CLARKE et al., 2012). Ao final, o bloco analisado compreendeu 8.068 SNPs, dos quais 2.350 foram genotipados pelas amostras da COGS.

2.4.1 Identificação de SNPs de susceptibilidade candidatos a partir do Fine-Mapping

A seleção dos SNPs genotipados e imputados foram testados para o potencial risco de associação com câncer de ovário. A associação mais significativa encontrada foi do SNP rs11782652 ($p = 1 \times 10^{-11}$), corroborando os achados de Pharoah de 2013. Ademais, baseado na estatística do log da probabilidade, a partir da análise de associação, foram encontrados mais oito SNPs com probabilidade de 100:1 de ser o variante funcional, a saber: rs74544416, rs787224141, rs78740005, rs111683632, rs137960856, rs74758321, rs76837345 e rs35094336, sendo que sete desses SNPs foram identificados por Pharoah em 2013, diferenciando apenas a probabilidade estatística que antes era de 1000:1. Esses SNPs restringem o sinal de associação para uma região de aproximadamente 17 kb (chr8: 82.653.590 – 82.670.992, Human Genome Browser hg19 assembly) (Figura 8)(dados não publicados).

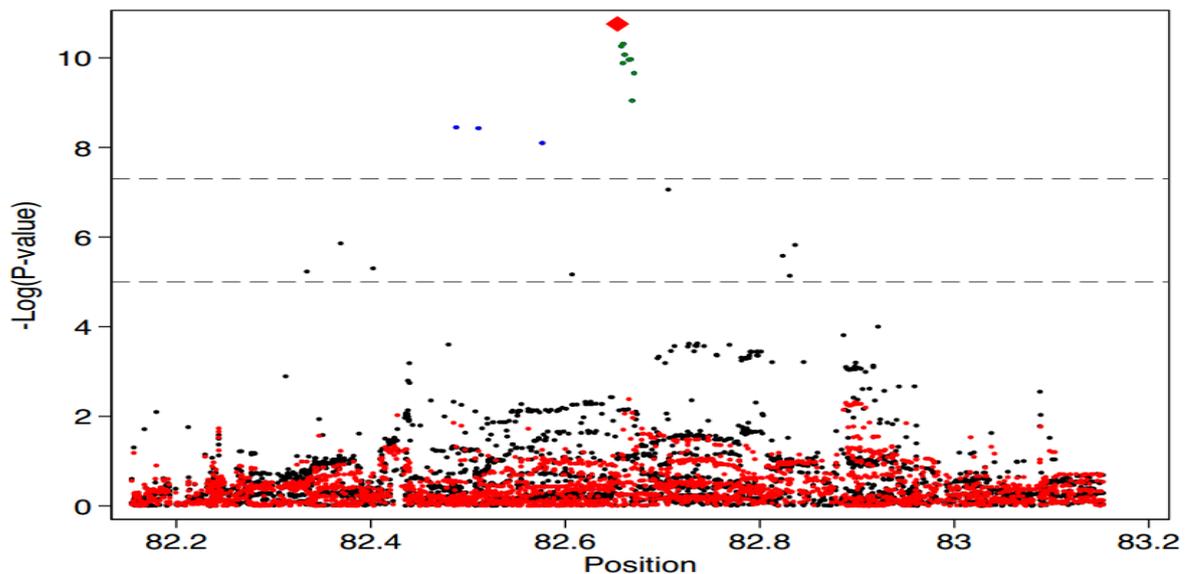


Figura 8 - Análise de Fine mapping determina a associação entre câncer de ovário com a região no locus 8q21. SNPs genotipados (vermelho) e imputados (preto) representados em suas respectivas localizações no cromossomo 8 (eixo x) e $-\log_{10}$ (valor de p) na associação com o câncer de ovário. O traço horizontal cinza representa o ponto de corte para a probabilidade de grande associação. Os pontos azuis representam os SNPs com probabilidade de 1000:1, os pontos verdes representam os SNPs com probabilidade de 100:1 e o losango vermelho representa o SNP com maior associação (Pharoah *et al*, Dados não publicados).

Os dados reportados, por ambos os estudos apresentados até o momento, norteiam os próximos passos da pesquisa com o objetivo de alcançar uma aplicação clínica. É irrefutável o fato de que o gene CHMP4C apresenta dados robustos que o relaciona diretamente com a susceptibilidade ao CaOV e, também, que há ao menos nove SNPs com características de atuarem modificando funções celulares que sujeitam o indivíduo a desenvolver a malignidade. Dessa forma, é imperativo a necessidade de análises funcionais que elucidem variações na presença do alelo de risco associado à doença e o possível impacto celular para aumentar a susceptibilidade ao desenvolvimento da doença.

3 OBJETIVO GERAL

O presente trabalho teve como alvo investigar as funcionalidades celulares que podem sofrer alterações quando presente o alelo de risco dos SNPs candidatos e possíveis implicações na susceptibilidade ao desenvolvimento do CaOV.

3.1 OBJETIVOS ESPECÍFICOS

- Analisar os nove prováveis SNPs causais *in silico* para obtenção de dados que suportem a hipótese de implicação funcional celular;
- Avaliar sistematicamente os nove prováveis SNPs causais *in vitro* a fim de identificar alterações na funcionalidade que sejam alelo dependentes.

4 MATERIAIS E MÉTODOS

4.1 ANÁLISE DE FRAGMENTOS GENÔMICOS COM ATIVIDADE DE *ENHANCER*

A análise de uma região para atividade de *enhancer* foi realizada segundo o protocolo proposto por Buckley e colaboradores (2015). Brevemente, segmentos genômicos (aqui denominados como “tiles”), de até 2kb, foram amplificados por PCR tendo como fita molde o DNA genômico ou o cromossomo artificial bacteriano (BAC) RP11-464L17 (ThermoFisher), utilizando primers contendo locais de recombinação att B (Senso - 5' CCAACTTTGTACAAAAAGCAGGCTACGCGT, Anti-senso – 5' CCAACTTTGTACAAGAAAGCTGGGTCTCGAG). Para a amplificação utilizou-se KAPA HiFi HotStart PCR kit (Kapabiosystems) seguido de purificação do gel (Qiagen). A tecnologia de clonagem do tipo gateway (Gateway BP clonase – Life Tech) foi utilizada para clonar os amplicons no vetor pDONR 221. As reações de recombinação (Gateway LR Clonase II – Life Tech) foram utilizadas para clonar os *tiles* genômicos nas orientações senso e antisenso (Figura 9) *upstream* ao promotor SV40 que conduz a expressão da luciferase *firefly* no vetor pGL3-Promoter-LR (Promega) (Figura10), o qual foi obtido através do Gateway Vector Conversion System (Life Tech).

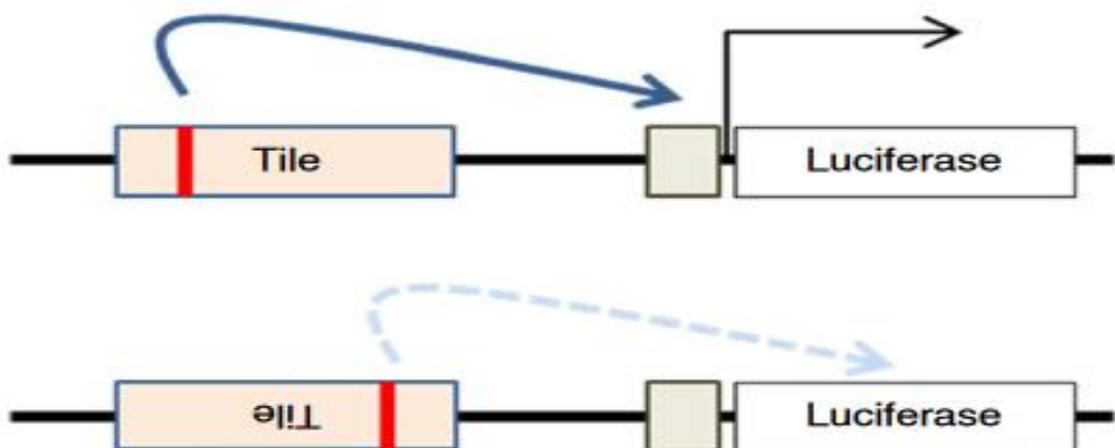


Figura 9 - Esquema dos *tiles* nas duas orientações em relação ao promotor SV40 e a luciferase. Em vermelho a posição do SNP de interesse (BUCKLEY et al., 2015).

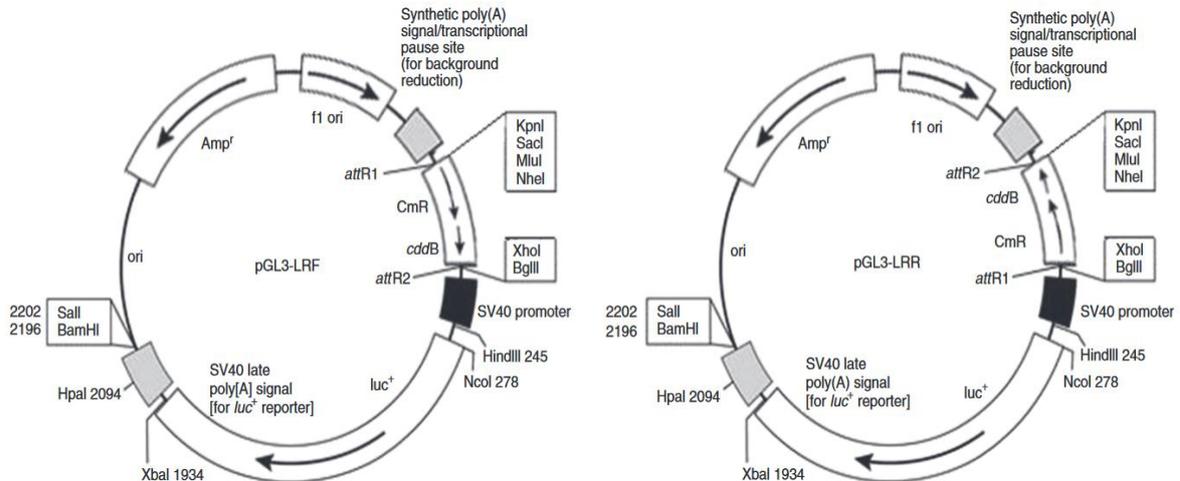


Figura 10 - Mapa dos vetores. Mapa do vetor pGL3-LR – F (senso) e R (anti-senso) são adaptados do mapa original do pGL3-Promoter da Promega. *Luc*⁺, gene da luciferase; ori, origem da replicação; *amp*^r, gene de resistência à ampicilina; f1 ori, oriem da replicação f1 phage; *attR1* (5'-ACAAGTTTGTACAAAAAGCTGAAC-3') e *attR2* (5'-GACCACTTTGTACAAGAAAGCTGAAC-3'); CmR, gene de resistência ao cloranfenicol; *cddB*, gene codificante uma *DNA gyrase lethal* para *E.coli* (BUCKLEY et al., 2015).

Para o ensaio repórter, usamos a linhagem iOSE^{CMYC} (célula epitelial da superfície do ovário – Tert-imortalizado e parcialmente transformado com o *Myc*) alocada em placas de 96 poços com 5×10^3 células/poço. Os vetores pGL3 contendo os *tiles* individualmente foram co-transfectados com o vetor pRL-CMV, que é o controle interno, usando Fugene HD (Promega) na razão de 3 de Fugene HD (volume em μ L): 1 DNA (ng). As leituras de DNA foram obtidas após 24 horas da transfecção usando Dual-Glo Luciferase Assay System (Promega). A atividade da luciferase foi calculada a partir da normalização pelo controle interno e o nível de significância foi determinado pela comparação da atividade de luciferase entre o alelo de menor frequência com o de maior frequência (teste t não pareado; valor de $p \leq 0,05$). Os alelos foram gerados utilizando *QuickChange PCR mutagenesis* (BRAMAN; PAPWORTH; GREENER, 1996). Para o sequenciamento, utilizamos os primers RVprimer3 e GLprimer2 (Promega) para verificar o variante após a mutagênese.

4.1.1 Análise estatística

Os resultados foram expressos como média \pm desvio padrão de experimentos realizados em triplicatas biológicas e avaliados através de Análise de Variância (ANOVA) de uma via. Um valor $p < 0,05$ foi considerado estatisticamente significativo. Análises estatísticas foram realizadas no software GraphPadPrism v.5

4.2 ANÁLISE POR STARR-SEQ DE REGIÕES DO GENOMA COM ATIVIDADE DE *ENHANCER*

4.2.1 Geração da biblioteca e preparação das amostras para coleta de dados

Para a análise de STARR-seq (self-transcribing active regulatory region sequencing) (ARNOLD et al., 2013), que também identifica regiões do genoma com atividade de *enhancer*, foram realizadas etapas sucessivas descritas sucintamente a seguir. Brevemente, o primeiro passo foi a geração da biblioteca da região. Para tanto o BAC DNA RP11-464L17 (ThermoFisher) foi fragmentado usando o sonicador (Covaris M220) para obtenção de fragmentos, em média, de 500 pares de base. Em seguida, foi realizada uma reação para reparar as extremidades e a cauda A. Posteriormente, foi realizada a ligação de adaptadores e a limpeza dos fragmentos através de *beads* magnéticos que se ligam ao DNA. A seguir, foi realizada a amplificação da biblioteca através da PCR, utilizando primers pré-estabelecidos para os adaptadores e a limpeza novamente dos fragmentos. Na sequência, foi feita a recombinação dos fragmentos da biblioteca gerada em vetores de varredura para STARR-seq. Essa etapa foi realizada utilizando enzimas de digestão Agel e Sall (NEB) e In-fusion HD (Clontech). De posse dos vetores, foi feita a transformação de bactérias utilizando MegaX DH10B electrocompetent (Thermo Fisher) para replicação dos vetores. Por fim, realizou-se a transfecção de células iOSE4^{CMYC} com 20 µg da biblioteca e após 24 horas foi coletado o RNA, utilizando os kits Qiagen RNeasy mini kit e Qiagen Shredder (Qiagen). Também foram coletados os plasmídeos remanescentes, através do kit de mini-prep (Qiagen), para utilizá-los como controle interno da biblioteca que foi utilizada. Após todas essas etapas, o RNA foi convertido em cDNA, indexado e feito o sequenciamento de nova geração – *Next-Generation Sequence* - (Illumina) (Figura 11).

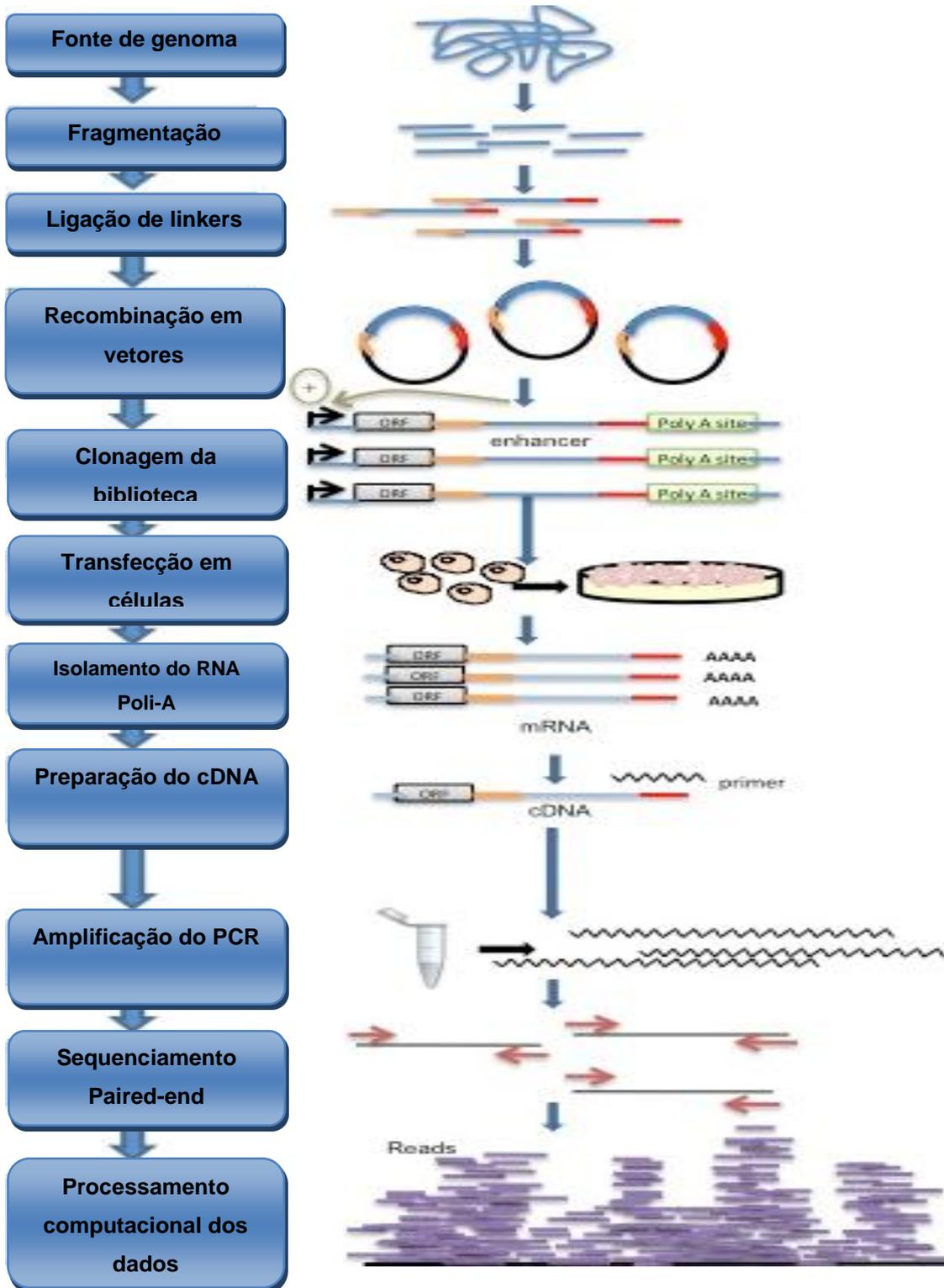


Figura 11 - Visão geral da sequência experimental do STARR-seq para identificação de regiões com atividade de *enhancer*. (MUERDTER; BORYŃ; ARNOLD, 2015).

4.2.2 Análise *in silico* para os dados a partir do sequenciamento de nova geração

Os arquivos gerados a partir do sequenciamento (BAM files) foram alinhados com o genoma humano (Human Genome version 19) utilizando o programa computacional Bowtie2 (LANGMEAD; SALZBERG, 2012). Em seguida, utilizamos o programa Integrative Genomics Viewer (ROBINSON et al., 2011) para visualização das regiões que foram mapeadas com alto número de cópias e também para observar se a biblioteca gerada estava cobrindo a região de interesse. O programa *Model-based Analysis of ChIP-Seq data* (MACS)2 identifica regiões genômicas enriquecidas em dados de ChIP- e DNase-seq que são estatisticamente significantes usando um algoritmo. O programa seleciona os picos ao analisar o genoma inteiro, calculando o valor de p para cada pico utilizando uma dinâmica distribuição de Poisson para capturar tendências (ZHANG et al., 2008). Por fim, as regiões de *enhancer* foram caracterizadas por apresentar alto número de cópias transcritas.

4.3 ANÁLISE DE FATORES DE TRANSCRIÇÃO COM ATIVIDADE ALELO DEPENDENTE POR ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

O protocolo para extrato de proteínas nucleares de células iOSE4^{CMYC} foi baseado no modelo proposto por Dignam e colaboradores (1983) para extratos nucleares. Brevemente, o extrato nuclear foi preparado usando uma solução hipotônica de lise (10mM HEPES pH 7,9, 1,5mM de MgCl₂, 10mM KCl, suplementado por Ditioneitol (DTT) e inibidores de protease) seguido por ruptura celular usando uma seringa e isolando a fração nuclear com uma solução de extração (20mM HEPES pH 7,9, 1,5mM MgCl₂, 0,42M NaCl, 0,2mM EDTA, 25% v/v glicerol, suplementado com DTT e inibidor de protease). Para cada SNP, uma sonda de DNA foi gerada para conter o alelo mais frequente ou o alelo menos frequente com 20 pares de base de cada lado (Tabela 2). As sondas senso e anti-senso foram aneladas (10µM de cada) utilizando um bloco de aquecimento a 90°C por 5 minutos (resfriamento gradativo), seguido de marcação com ATP [γ -32P] (Perkin Elmer) usando T4 polinucleotideo quinase (NEB) e retirando o ATP [γ -32P] não ligado através do kit QIAquick Nucleotide Removal (Qiagen). As sondas marcadas foram incubadas com o extrato de proteínas nucleares isoladas previamente usando

LightShift Poly(dI-dC) (Thermo) em solução tampão de ligação (10mM Tris, 50mM KCl, 1mM DTT, pH 7,4). A reação foi então transferida para o gel 6% poliacrilamida de eletroforese, nas condições: 83 Volts, *overnight* a 4°C. O gel foi seco e filmes foram expostos ao gel por 5-8 horas. Os ensaios de EMSA foram realizados em pelo menos dois experimentos independentes e cada sonda foi adicionada em duplicata para cada gel. Mudanças no padrão de migração foram analisadas por inspeção visual e apenas variações reprodutíveis foram considerados positivas.

Tabela 2 – Tabela com as sondas utilizadas para a técnica de EMSA e as respectivas sequência, alelo de maior frequência - Major - (senso e anti-senso) e alelo de menor frequência – Minor - (senso e anti-senso).

Nome das sondas	Sequência
rs11782652 Major A	ATAGGATTTACACAGTTACAAGGAATTATGGCAAGAGAGTA
rs11782652 Major T	TACTCTCTTGCCATAATTCCTTGTAAGTGTGTAAATCCTAT
rs11782652 Minor G	ATAGGATTTACACAGTTACAGGGAATTATGGCAAGAGAGTA
rs11782652 Minor C	TACTCTCTTGCCATAATTCCTTGTAAGTGTGTAAATCCTAT
rs78740005 Major T	CCAGATTTGAAGGTAAACAATCCTAGAGCATGATGTTCTAT
rs78740005 Major A	ATAGAACATCATGCTCTAGGATTGTTTACCTTCAAATCTGG
rs78740005 Minor C	CCAGATTTGAAGGTAAACAACCCTAGAGCATGATGTTCTAT
rs78740005 Minor G	ATAGAACATCATGCTCTAGGGTTGTTTACCTTCAAATCTGG
rs78724141 Major G	AGGCACTGAACCAGCCTGACGAGATAGGGAAGACATGCACT
rs78724141 Major C	AGTGCATGTCTTCCCTATCTCGTCAGGCTGGTTCAGTGCCT
rs78724141 Minor T	AGGCACTGAACCAGCCTGACTAGATAGGGAAGACATGCACT
rs78724141 Minor A	AGTGCATGTCTTCCCTATCTAGTCAGGCTGGTTCAGTGCCT
rs111683632 Major C	TGCAGTGAGCCAAGATCGCGCCACTGCACTCCAGCCTGGTG
rs111683632 Major G	CACCAGGCTGGAGTGCAGTGGCGCGATCTTGGCTCACTGCA
rs111683632 Minor G	TGCAGTGAGCCAAGATCGCGGCACTGCACTCCAGCCTGGTG
rs111683632 Minor C	CACCAGGCTGGAGTGCAGTGCCGCGATCTTGGCTCACTGCA
rs74758321 Major A	TTTGTCATATGCCCTATCAGACAAAATTTCAAATGGGTTT
rs74758321 Major T	AAACCCATTTTCAAATTTTGTCTGATAGGGCATATGACAAA
rs74758321 Minor G	TTTGTCATATGCCCTATCAGGCAAATTTCAAATGGGTTT
rs74758321 Minor C	AAACCCATTTTCAAATTTTGCCTGATAGGGCATATGACAAA
rs76837345 Major A	GTGATAATGAAGATTTTTGGATCCCTGGATTATCTCCAAAA
rs76837345 Major T	TTTTGGAGATAATCCAGGGATCCAAAAATCTTCATTATCAC
rs76837345 Minor G	GTGATAATGAAGATTTTTGGGTCCCTGGATTATCTCCAAAA
rs76837345 Minor C	TTTTGGAGATAATCCAGGGACCCAAAAATCTTCATTATCAC
rs35094336 Major G	TCAAACAATTGGCAGCTTGGGCTACCTAAACTAAAACACAT
rs35094336 Major C	ATGTGTTTTAGTTTAGGTAGCCCAAGCTGCCAATTGTTTGA
rs35094336 Minor A	TCAAACAATTGGCAGCTTGGACTACCTAAACTAAAACACAT
rs35094336 Minor T	ATGTGTTTTAGTTTAGGTAGTCCAAGCTGCCAATTGTTTGA

rs74544416 Major GTGA	ATCTGTTTCATGAAAACATGTGAGTGACTCTGGTCTCCCTCT
rs74544416 Major GTGA	AGAGGGAGACCAGAGTCACTCACATGTTTTTCATGAACAGAT
rs74544416 Minor Deleção	ATCTGTTTCATGAAAACATGTGACTCTGGTCTCCCTCT
rs74544416 Minor Deleção	AGAGGGAGACCAGAGTCACTCACATGTTTTTCATGAACAGAT
rs137960856 Major A	GTCCTATACATTACATGAGCAATGAAATTGAAACTCTTTTCTG
rs137960856 Major A	CAGAAAAGAGTTTCAATTTTCATTGCTCATGTAATGTATAGGAC
rs137960856 Minor G	GTCCTATACATTACATGAGCAGTGAAATTGAAACTCTTTTCTG
rs137960856 Minor G	CAGAAAAGAGTTTCAATTTTCACTGCTCATGTAATGTATAGGAC

4.4 TRANSFECCÕES ESTÁVEIS COM VETORES CONTENDO O GENE CHMP4C PARA AMBOS ALELOS

Transfecções estáveis para as condições: a) CHMP4C – rs35094336 alelo G b) CHMP4C – rs335094336 alelo A c) pCMV6-entry vetor vazio (figura 12), foram realizadas utilizando FuGENE HD (Promega) de acordo com as instruções do fabricante. Brevemente, células iOSE4^{CMYC} foram cultivadas em placas de 6 poços por 24 horas de forma que estivessem entre 70 a 90% de confluência na hora da transfecção. No dia da transfecção, 19 µg do plasmídeo para cada condição, contendo o gene de resistência à Neomicina e 56 µL de FuGENE HD em Opti-MEM (Invitrogen) foram adicionados individualmente em tubos eppendorf. Após 10 minutos de incubação em temperatura ambiente, o complexo foi adicionado individualmente aos poços. Após 48 horas, 200µg/ml de Geneticina (Sigma) foi adicionada ao meio completo. O meio completo com Geneticina foi trocado a cada quatro dias gerando colônias resistentes à Geneticina.

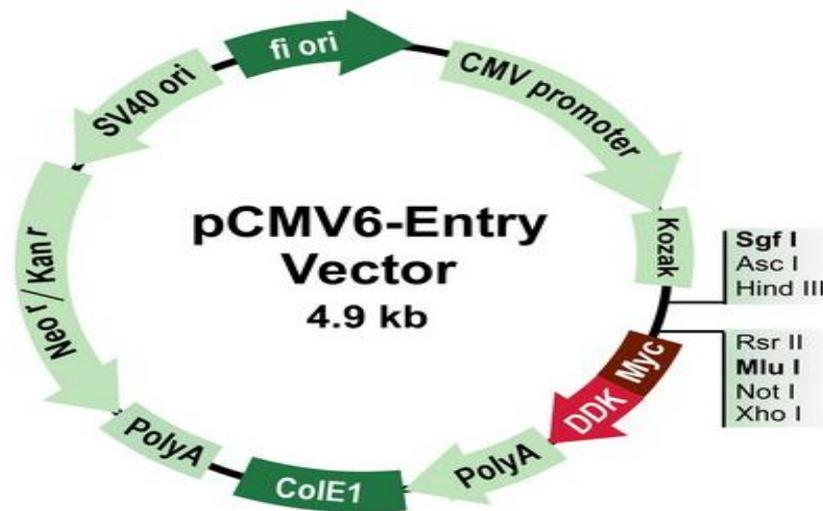


Figura 12 – Desenho esquemático do vetor vazio pCMV6-Entry.

4.5 ANÁLISES DA ESTABILIDADE PROTÉICA DE CHMP4C ATRAVÉS DE WESTERN BLOT

As células contendo as condições: a) CHMP4C – rs35094336 alelo G b) CHMP4C – rs335094336 alelo A c) pCMV6-entry vetor vazio, após cultivo para estabilização e seleção de colônias transfectadas, foram tratadas com cicloheximida (Sigma), inibidor clássico da síntese proteica, por 0h, 6h, 12h, 24h, 36h e 48h. As células foram coletadas em PBS e iniciado o protocolo para obtenção do extrato proteico celular. Brevemente, o lisado das células foi preparado utilizando solução de lise NETN (100mM NaCl, 20mM Tris-Cl (pH 8,0) e 0,5 mM EDTA). As proteínas foram resolvidas por gel de eletroforese desnaturante SDS-PAGE e transferidas para uma membrana. As proteínas fundidas ao epítipo denominado DDK (flag) foram detectadas por anticorpo policlonal anti-DDK (Flag) e o anticorpo IgG conjugado de camundongo anti-coelho e visualizado com substrato Pierce ECL Western blotting (Thermo Scientific) de acordo com o protocolo fornecido.

5 RESULTADOS

Foram testados todos os nove SNPs encontrados pelo *fine-mapping* com probabilidade de 100:1, associados ao tag-SNP rs11782652. Todos os SNPs estão em uma área de aproximadamente 18kb (chr8:82.653.396-82.671.115) no gene CHMP4C, sendo oito SNPs localizados em íntrons e um localizado no éxon 5 do gene CHMP4C. Os detalhes dos nove SNPs candidatos como a alteração, a classe, a função e a frequência alélica estão na Tabela 3.

Tabela 3 - Lista dos SNPs com probabilidade de 100:1 e suas características: Alteração, classe, função e frequência alélica.

SNP candidato	Alteração	Classe	Localização genômica	Frequência alélica
rs11782652	A/G	Único	Íntron	A:94,549%(4735/5008);G:5,451%(273/5008)
rs74544416	A/G	Único	Íntron	A:95,947%(4805/5008);G:4,053%(203/5008)
rs78724141	G/T	Único	Íntron	G:95,527%(4784/5008);T:4,473%(224/5008)
rs78740005	T/C	Único	Íntron	T:95,887%(4802/5008);C:4,113%(206/5008)
rs111683632	C/G	Único	Íntron	C:95,527%(4784/5008);G:4,473%(273/5008)
rs137960856	GTGA/*	Deleção	Variante de splice	GTGA:93,890%(105445/112307); -:6,110%(6862/112307)
rs74758321	A/G	Único	Íntron	A:95,407%(4778/5008);G:4,593%(230/5008)
rs76837345	A/G	Único	Íntron	A:96,506%(4833/5008);G:3,494%(175/5008)
rs35094336	G/A	Único	Alteração não sinônima	G:94,490%(119089/126034);A:5,510%(6945/126034)

O SNP rs35094336 está localizado no éxon 5 do gene CHMP4C, sendo uma alteração não sinônima (quando a alteração de uma das bases do DNA resulta em alteração do aminoácido codificado na proteína), o que o coloca como principal SNP causal. Para caracterizar os possíveis impactos da alteração, utilizamos a ferramenta *in silico* PolyPhen-2 (Polymorphism Phenotyping v2). Essa ferramenta prediz o possível impacto da substituição de um aminoácido na estabilidade e função de proteínas humanas usando considerações estruturais e comparativos

evolucionários (ADZHUBEI; JORDAN; SUNYAEV, 2013). O resultado obtido através da ferramenta é mostrado na Tabela 4. Uma representação 3D da proteína CHMP4C com o local onde ocorre a variação (apontada pela seta) está representada na figura 13.

Tabela 4 – Dados obtidos após análise utilizando a ferramenta *in silico* PolyPhen-2. A predição para o SNP rs35094336 é de que há um provável dano na proteína quando alterado o alelo G (mais frequente) para o alelo A (menos frequente), ocasionando em troca do aminoácido Alalina por Treonina.

SNP	Gene	Alelo	Mudança	Detalhe	Predição	Score
rs35094336	CHMP4C	G/A	Alalina > Treonina	Não sinônimo	Provável dano	0,997

Ala232Thr

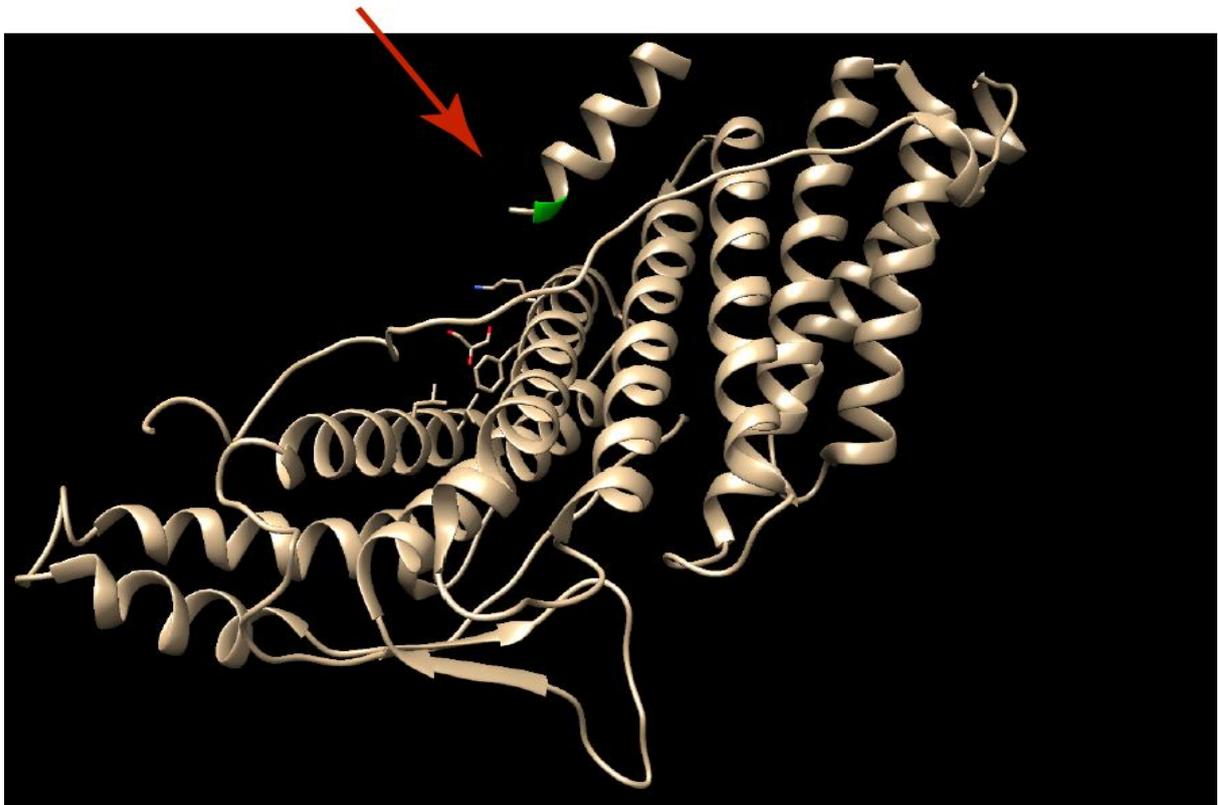


Figura 13 – Representação 3D da proteína CHMP4C. A seta representa a alteração do aminoácido Alalina para Treonina na posição 232 da sequência proteica de CHMP4C. Esse resíduo é localizado na Alfa-hélice anfipática da porção C-terminal que é conservada em todas as proteínas da família CHMP4.

A fim de identificar a presença de outros genes, além de CHMP4C, potencialmente associado aos SNPs candidatos do *locus* 8q21, foi realizada outra análise *in silico*. O *Topologically Association Domain* (TAD) é uma região de interação do cromossomo a partir da conformação espacial, ou seja, a sequência de DNA dentro de um TAD interage fisicamente muito mais entre elas do que com uma sequência fora do TAD. A função dos TADs não é completamente entendida, contudo, em alguns casos, a disrupção da TAD pode resultar em doenças porque a mudança organizacional em 3D do cromossomo interrompe a regulação gênica (POMBO; DILLON, 2015). Na Figura 14, podemos analisar a região de 2Mb ao redor dos SNPs candidatos (bloco roxo). Constatamos que não há forte interação entre as regiões vizinhas e os SNPs candidatos, indicando que a funcionalidade está, muito provavelmente, contida apenas em relação ao gene CHMP4C.

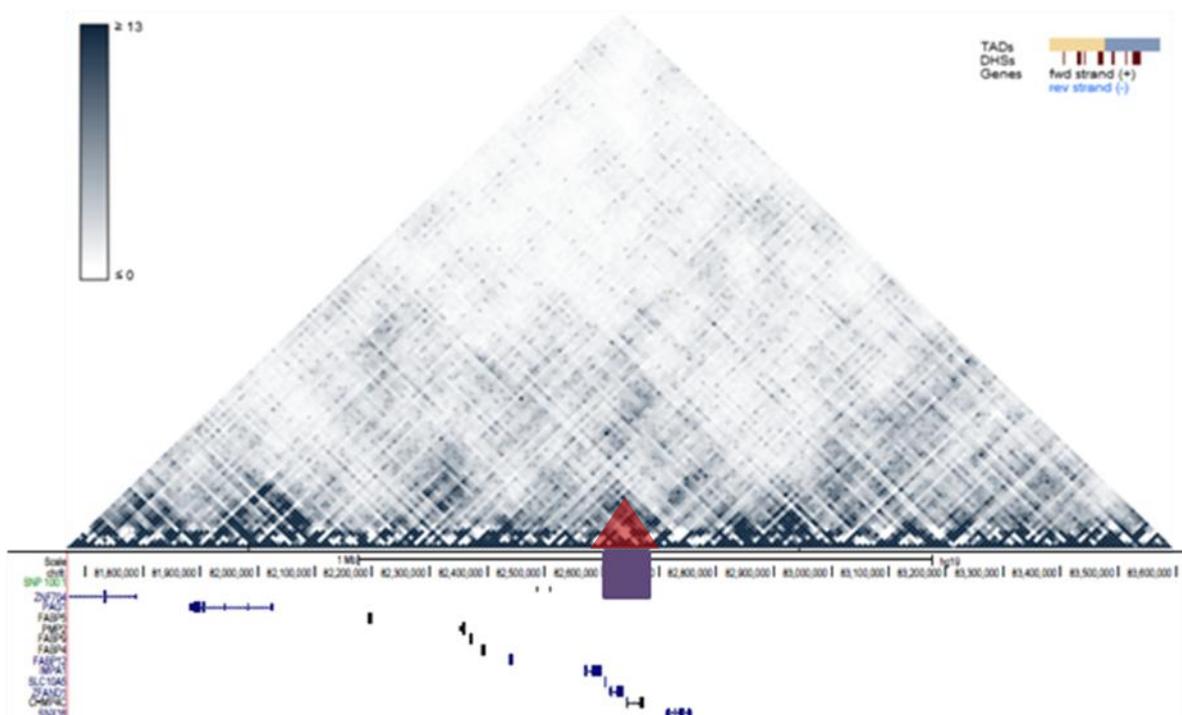


Figura 14 - Bloco obtido para análise de domínios com associação topológica. O bloco em lilás mostra a região em que os nove SNPs candidatos estão localizados. O triângulo em vermelho mostra que há pouca interação da região que contém os nove SNPs candidatos com regiões em que outros genes estão presentes, confirmando que CHMP4C deve ser o principal gene a ser investigado. Acesso: <http://promoter.bx.psu.edu/hi-c/view.php>

A localização dos SNPs candidatos é importante para determinar as possíveis funções. Nós avaliamos a sobreposição entre os nove SNPs e regiões que apresentam características de elementos regulatórios ativos, tendo como referência: *Formaldehyde Assisted Isolation of Regulatory Elements* seguido de

sequenciamento (FAIRE-seq), que é indicativo de cromatina ativa; Histona H3 Lisina 27 acetilada (H3K27Ac), indicativo de *enhancer* e região promotora; Histona H3 Lisina 4 Mono-metilada (H3K4Me1), indicativo de *Enhancer*. Os dados para esses elementos regulatórios foram obtidos através de um painel de estudos realizados com linhagens de células normais e tumorais (COETZEE et al., 2015). Essa análise identificou apenas um SNP, rs11782652, sobrepondo as marcações de características de cromatina em células iOSE11, UWB1.289, CaOV3 e iFTSEC33 (COETZEE et al., 2015; LAWRENSON et al., 2009) (FIGURA 15A e B).

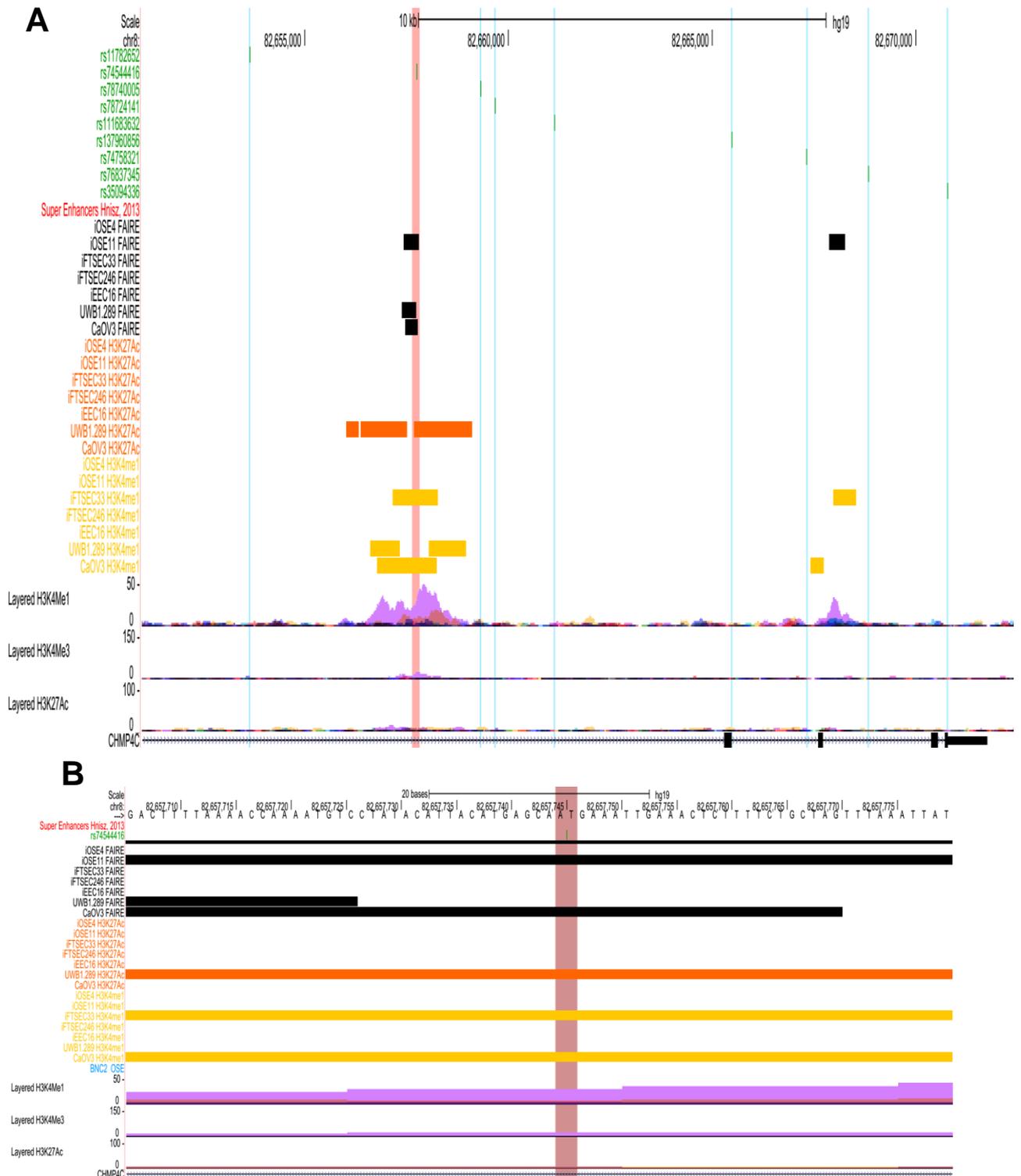


Figura 15 – Visualização da área em que os nove SNPs candidatos estão no *locus* 8q21 com os dados obtidos através do UCSC. À esquerda estão os nove SNPs candidatos (verde) e os dados organizados por experimento e linhagem celular : Super enhancer = Vermelho; FAIRE-seq = preto; H3K27AC = laranja; H3K4me1 = amarelo; Layered H3K4me1 = preto; Gene CHMP4C = azul). A) As linhas azuis identificam o posicionamento dos oito SNPs e a relação negativa de biomarcadores para a região; a linha vermelha identifica o SNP rs11782652 e a relação positiva com biomarcadores. B) Ampliação da região do SNP rs11782652, confirmando a relação positiva com biomarcadores para atividade de cromatina.

5.1 MAPEAMENTO DOS NOVE SNPS CANDIDATOS PARA ELEMENTOS FUNCIONAIS RELACIONADOS AO CÂNCER DE OVÁRIO

Para validar a funcionalidade das regiões de *enhancer* encontradas nas análises *in silico* em células ovarianas, utilizamos um teste repórter. Primeiramente, geramos sete *tiles* genômicos (Tabela 5) que foram clonados no sentido senso e anti-senso logo depois do promotor SV40 da expressão da luciferase, visto que é conhecido que o loop formado para a transcrição pode afastar fisicamente a porção do *enhancer* impedindo sua atividade (Figura 16). Cada constructo foi transfectado em células ovarianas iOSE4^{CMYC} (LAWRENSON et al., 2010) e os níveis de luciferase foram mensurados após 24 horas da transfecção. Interessantemente, nenhuma região apresentou alteração significativa quando comparado os alelos (normal vs. risco) em ambas as direções (senso e anti-senso) (Gráfico 2).

Tabela 5 – Lista de SNPs e seus respectivos tiles e a posição da sequência genômica que o tile mimetiza.

SNP	Nome do tile	Posição genômica
rs11782652	D12	chr8:82,652,644-82,654,974
rs74544416	E2	chr8:82,656,266-82,658,539
rs78724141	E3	chr8:82,658,365-82,660,469
rs78740005	E3	chr8:82,658,365-82,660,469
rs111683632	E4	chr8:82,659,948-82,662,106
rs137960856	E7	chr8:82,665,245-82,667,150
rs74758321	E8	chr8:82,667,181-82,669,040
rs76837345	E8	chr8:82,667,181-82,669,040
rs35094336	E9	chr8:82,668,879-82,671,154

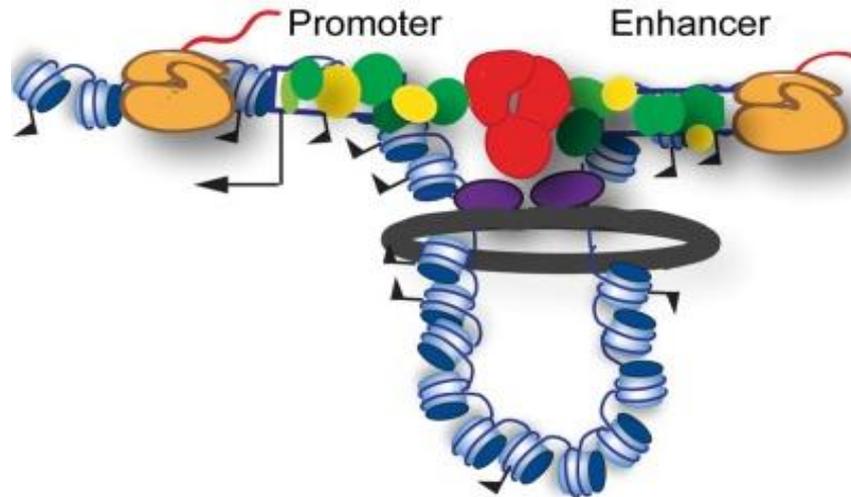


Figura 16 – Modelo de organização espacial da cromatina. *Enhancers* podem ativar a expressão gênica à distância específica. O looping promotor-*enhancer* é mediado por coesina (arco cinza), CTCF (roxo) e pelo mediador complexo (vermelho) que trazem o *enhancer* para próximo do seu promotor alvo. A seta marca o ponto de início da transcrição do gene. Fatores de transcrição (verde) e co-fatores (amarelo) se ligam ao *enhancer* e são trazidos também para a maquinaria transcricional no promotor (adaptado de MORA et al., 2015).

Como o resultado do teste de luciferase não validou as evidências *in silico* para a funcionalidade de regiões com atividade de *enhancer*, utilizamos outro teste para mitigar quaisquer dúvidas remanescentes. Sendo assim, realizamos o teste de STARR-seq para identificar novamente regiões do genoma associadas à funcionalidade de *enhancer* em células de origem epitelial ovariana. A abordagem em relação ao primeiro teste repórter se diferencia ao utilizar-se de uma biblioteca genômica construída utilizando BAC DNA, representativa de uma região de, aproximadamente, 282 Kb. Sucessivas etapas de controle de qualidade (CQ) foram realizadas através de sequenciamento por Next Generation Sequence (NGS) confirmando que a biblioteca era representativa da região de interesse. Em seguida, procedemos com a transfecção de células iOSE^{CMYC} com os constructos da biblioteca genômica por 24 horas, seguido de sequenciamento de NGS para identificação das regiões com função de *enhancer* (Figura 17).

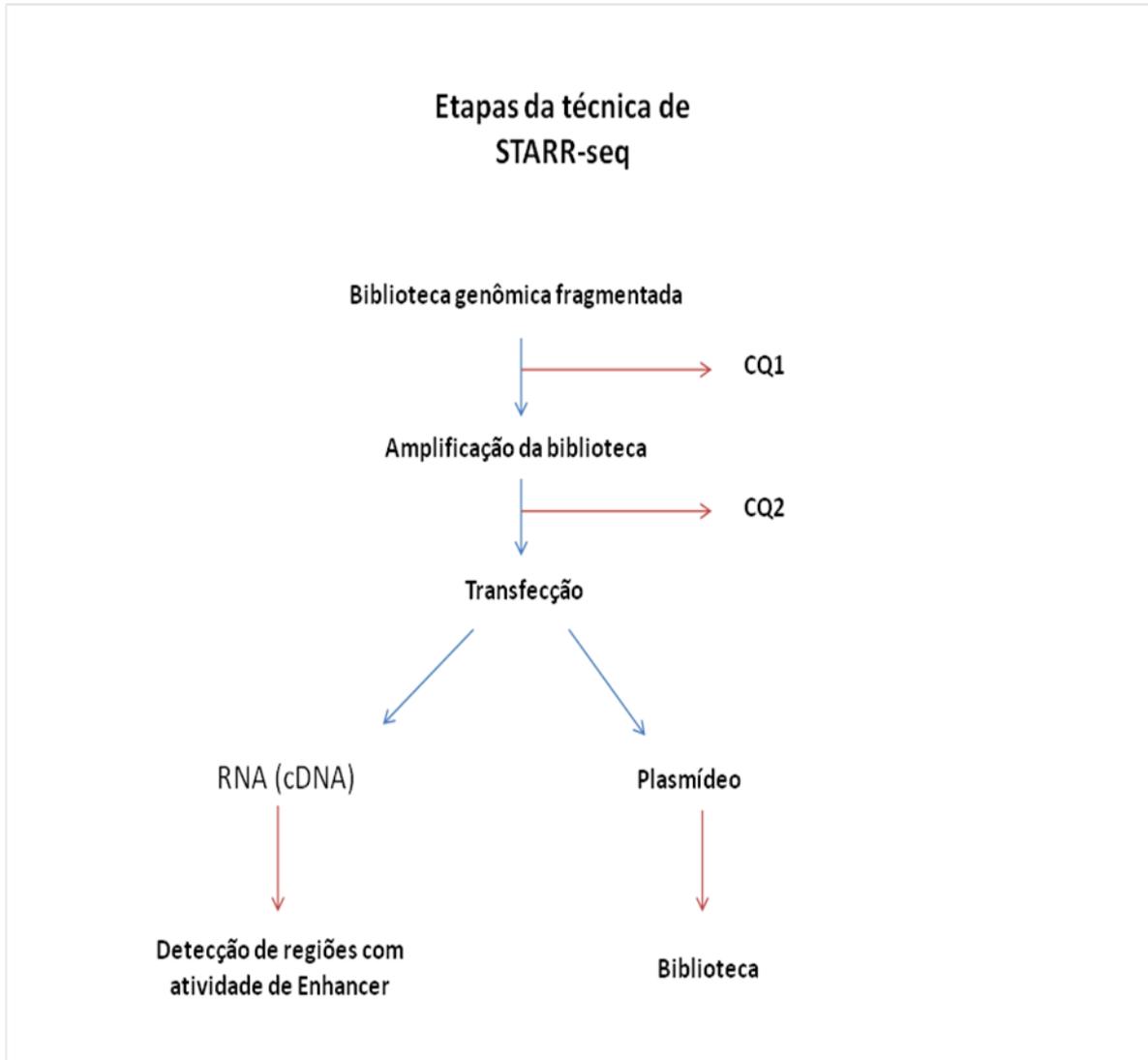


Figura 17 – Etapas da técnica de STARR-seq. Inicialmente realizamos a fragmentação do BAC (500 pares de base) confirmado através de bioanalyser, em seguida o primeiro sequenciamento (CQ1 – NGS) para confirmação de que a sequência da região de interesse permanecia completa. A segunda etapa foi amplificação da biblioteca, através de transformação de bactérias, seguido de novo sequenciamento (CQ2 – NGS) para confirmação de que a sequência da região de interesse permanecia intacta. Após as etapas de preparação da biblioteca, células iOSE^{CMYC} foram transfectadas com vetores contendo porções randômicas da região de interesse. Após 24 horas da transfecção, foi coletado o RNA (transformado em cDNA) – utilizado para identificar as regiões com atividade de *enhancers* – e os plasmídeos remanescentes para reconstituição da biblioteca – garantindo que não houveram falhas na sequência de interesse.

A fragmentação da BAC DNA RP11-464L17 em sequências randômicas de, em média, 500 pares de base, foi confirmada através da análise de gel de agarose por eletroforese (Figura18).

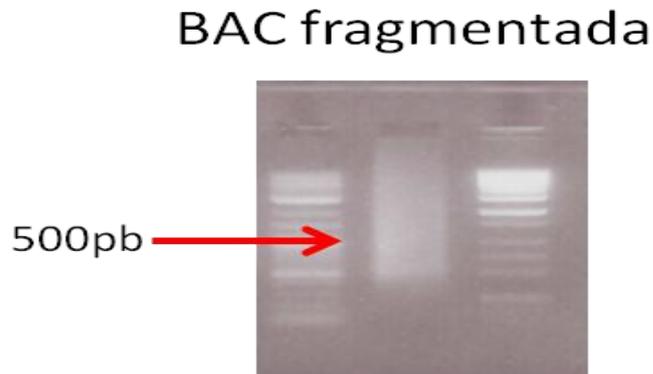


Figura 18 – Eletroforese em gel de agarose 10% para separação dos fragmentos após sonificação da BAC DNA RP11-464L17, confirmando que há fragmentos de 500 pares de base na amostra.

A cobertura genômica que foi capturada após os sucessivos sequenciamentos por NGS (Figura 19). Na primeira etapa de controle de qualidade, houveram falhas de alinhamento com região de interesse, como também ruídos em regiões que não estavam contidos no BAC DNA RP11-464L17. Contudo, no CQ2 houve um maior alinhamento das sequências obtidas pós amplificação da biblioteca e diminuição do ruído em regiões não específicas. Sendo assim, prosseguimos com os testes de transfecção celular e coleta de dados. A cobertura genômica com a coleta dos plasmídeos mostrou-se satisfatória por não apresentar ruídos fora da região específica. Eram esperadas lacunas no alinhamento da amostra contendo os plasmídeos, uma vez que parte da biblioteca foi introduzida nas células. O alinhamento do cDNA também mostrou-se satisfatório por não apresentar ruídos além das margens da região de interesse.

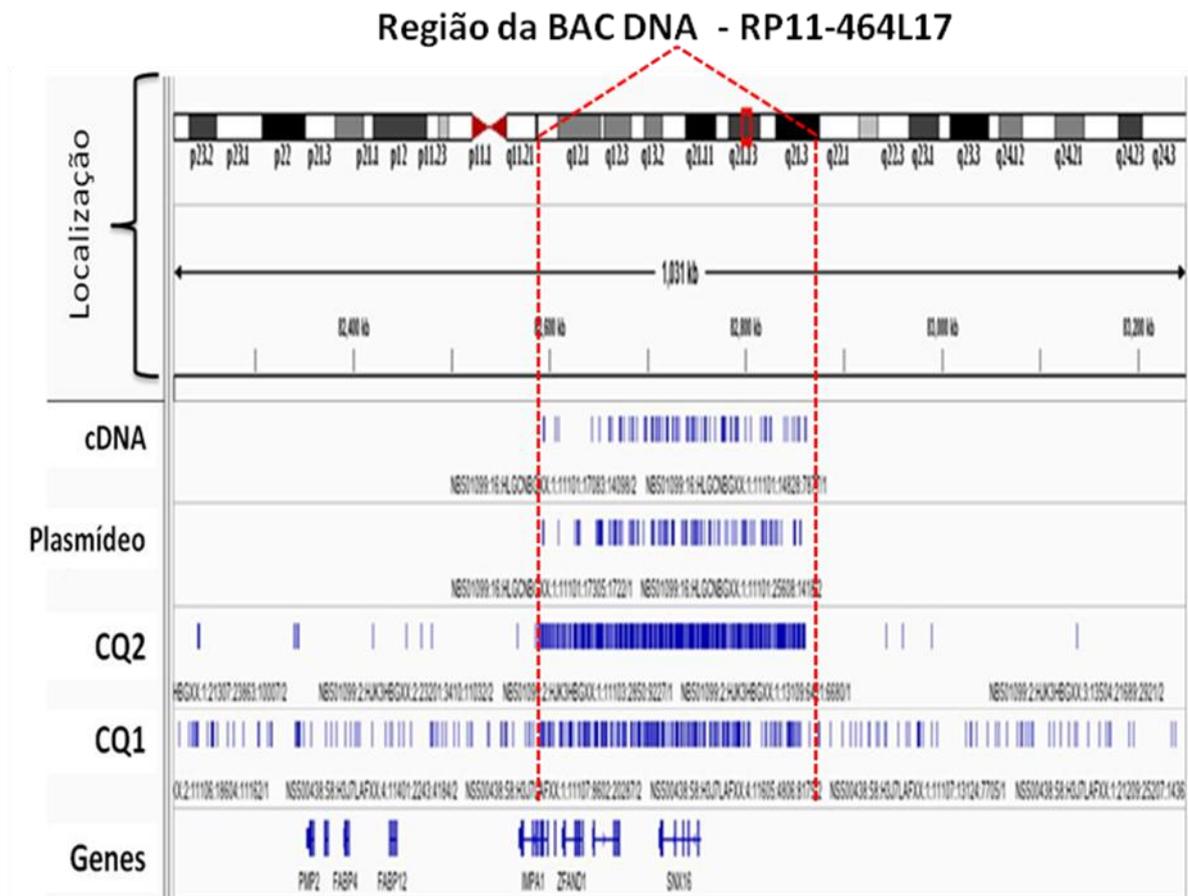


Figura 19 – Visualização, no programa IGV, do alinhamento com o genoma dos dados obtidos após sequenciamento (NGS) das amostras Controle de qualidade 1 (CQ1), Controle de qualidade 2 (CQ2), Plasmídeo e cDNA. Quando ocorre o alinhamento da sequência obtida pelo sequenciamento (NGS) da amostra com a sequência do genoma é feito um traço azul correspondente à região. Em vermelho, a delimitação da sequência representativa presente na BAC DNA RP11-464L17 (chr8:82,583,319-82,800,877).

Uma vez confirmado o sucesso do sequenciamento e do alinhamento das sequências, realizamos a análise que determina as regiões com atividade de *enhancer* na amostra. Para tanto, foi utilizada a ferramenta MACSII – CallPeaks que identifica regiões enriquecidas. Assim, foram detectadas dezenove regiões enriquecidas, traduzindo-se em regiões de atividade de *enhancer* (Tabela 6). Contudo, nenhuma região identificada é representativa do local em que os nove SNPs candidatos estão presentes (Figura 20), corroborando a análise inicial do ensaio de luciferase, em que nenhuma das regiões de interesse contendo esses SNPs possuem atividade de *enhancer in vitro*.

Tabela 6 – Resultado obtido após análise por MACSII-CallPeaks. Dezenove regiões foram consideradas positivas para enriquecimento (maior número de cópias), o que indica uma região com atividade de *enhancer*.

Cromossomo	Posição de início	Posição final	Altura do pico com as sequências empilhadas
chr8	56636499	56636500	10.59264
chr8	82594793	82594794	1.75333
chr8	82643807	82643808	3.35223
chr8	82672655	82672656	2.3267
chr8	82676245	82676246	2.75885
chr8	82698498	82698499	10.59264
Chr8	82704749	82704750	2.93522
chr8	82730363	82730364	3.10525
Chr8	82742185	82742186	2.51475
chr8	82776161	82776162	2.75885
Chr8	82778342	82778343	2.91432
chr8	82780250	82780251	4.98013
Chr8	82789877	82789878	2.75885
chr8	82792083	82792084	2.75885
Chr8	82800044	82800045	2.3267
chr8	82816277	82816278	2.91432
Chr8	82839607	82839608	3.70151
chr8	82847747	82847748	4.45838
chr8	82854413	82854414	2.3267

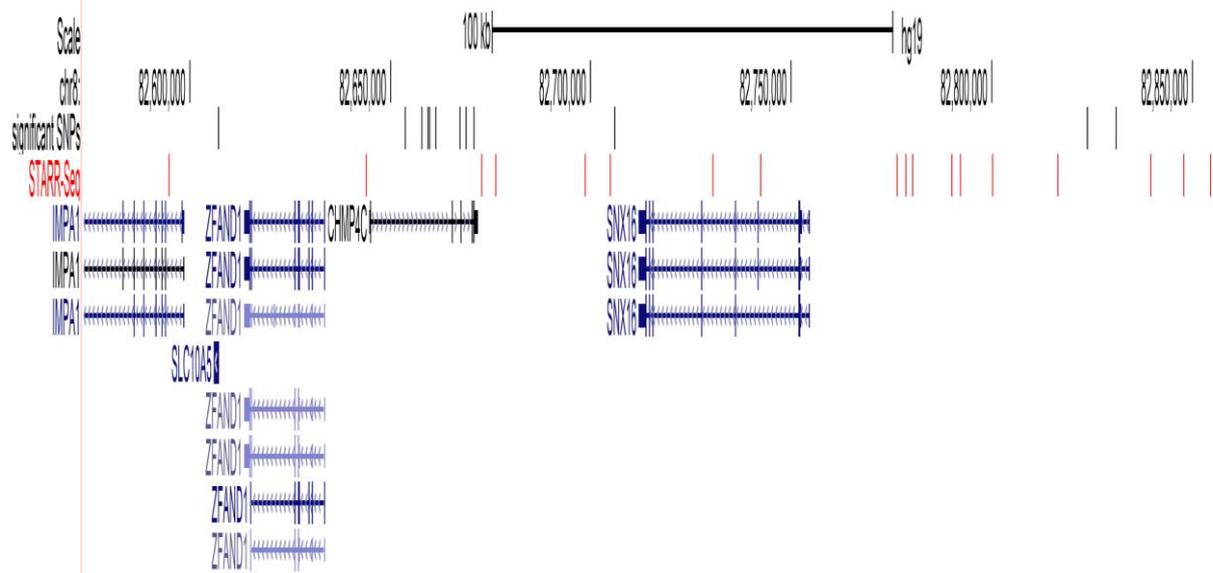


Figura 20 – Visualização no programa IGV das dezenove regiões positivas para enriquecimento de cópias (traços vermelhos). É possível notar que não há alinhamento das regiões enriquecidas com os nove SNPs candidatos (traços pretos). Em azul, na parte inferior da figura, os genes que estão na região da biblioteca.

Para investigar outros efeitos alelo específicos, realizamos o ensaio eletroforético de motilidade alterada (EMSA) para identificar a diferença na ligação de proteínas nucleares no alelo de referência *versus* o alelo de risco para cada SNP candidato. Usamos extratos nucleares de células iOSE4^{CMYC} incubados com as sondas de cada alelo a ser analisado. Foi possível identificar dois SNPs, rs111683632 e rs137960856, com constante diferença na ligação de proteínas quando comparado o alelo mais frequente com o menos frequente (Figura 21).

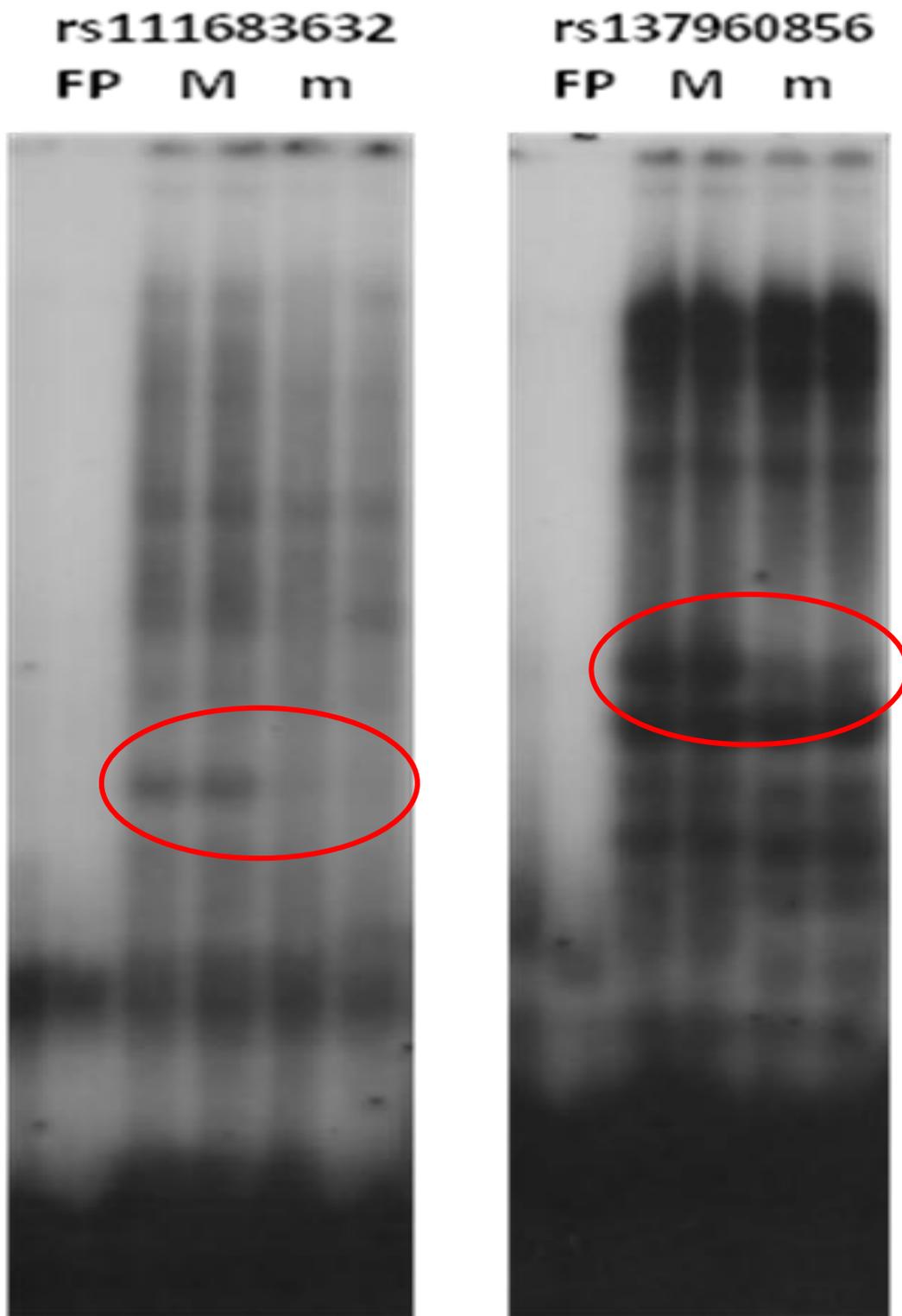
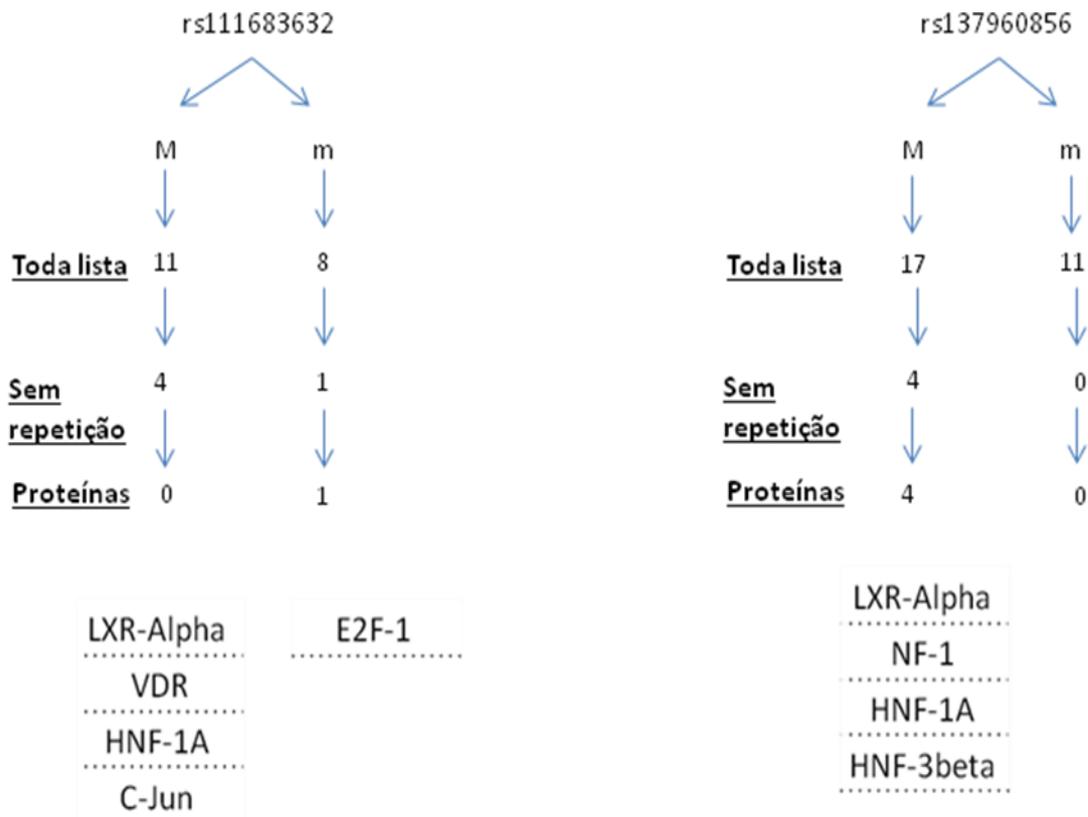


Figura 21 - Ensaio de EMSA para detecção de fatores de transcrição que se ligam em cada sonda contendo um diferente alelo para os nove SNPs candidatos. FP: *Free Probe* – Controle negativo. M: Alelo com maior frequência na população. m: Alelo com menor frequência na população. Para cada condição foram feitos dois poços.

Tendo os SNPs rs111683632 e rs137960856 apresentado variação na ligação de proteínas alelo específico, propusemos relacionar quais as possíveis proteínas estariam envolvidas nessa atividade. Para tanto, realizamos análise *in silico* com duas ferramentas de predição: PROMO e JASPAR. Ambas as ferramentas calculam quais proteínas podem ter afinidade por liga-se a uma dada sequência de DNA.

Realizamos os testes para os SNPs rs111683632 e rs137960856 utilizando as respectivas sequências das sondas geradas para o EMSA (Tabela 2) e para cada análise obtivemos uma lista com as prováveis proteínas associadas. Em seguida, delineamos um método de exclusão que consistiu em eliminar proteínas com predição para ambos os alelos, uma vez que o interesse é identificar atividade alelo específico de ligação. Com isso, foram obtidas listas de proteínas com ligação alelo específica (Figura 22).

PROMO



JASPAR

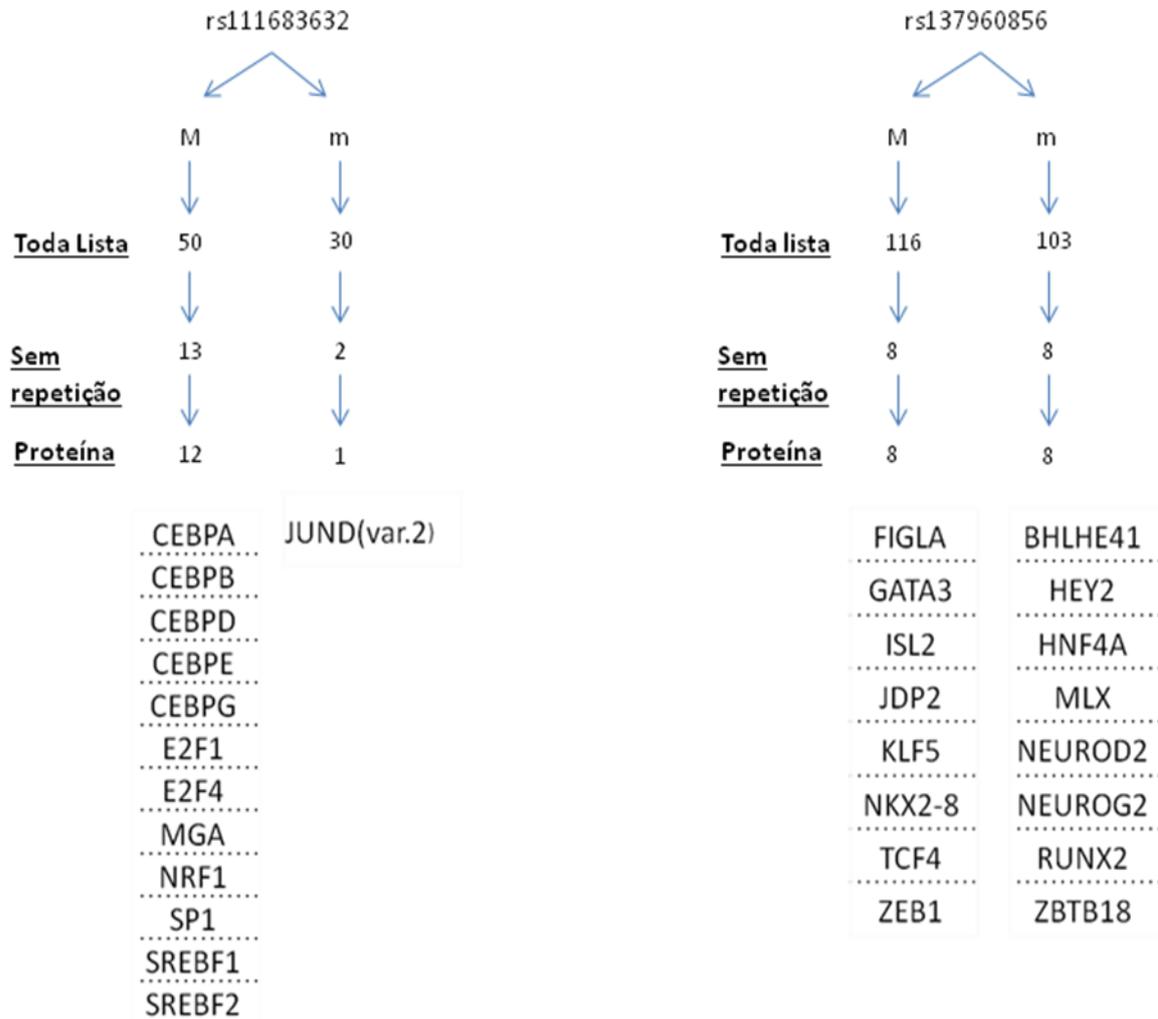


Figura 22 - Lista de prováveis fatores de transcrição que apresentaram potencial de ligação nas sequências de sondas desenhadas para cada alelo, sendo M o de maior frequência e m o de menor frequência, dos SNPs rs111683632 (C/G) e rs137960856 (GTGA/----). A lista foi obtida através de dois diferentes testes *in silico* (PROMO e JASPAR). Para cada análise seguiu-se o protocolo: selecionar toda a lista e comparar proteínas que eram particulares de um alelo e, por fim, eliminar proteínas repetidas gerando a lista final apresentada.

Como um dos SNPs está localizado na porção do éxon 5 do gene *charged multivesicular body protein 4C* (CHMP4C), rs35094336, tentamos avaliar se houve alteração da função celular na presença dos diferentes alelos (A/G). Como mostrado anteriormente, a troca do nucleotídeo na sequência gênica representa uma alteração com provável dano na proteína. É conhecido que CHMP4C é uma subunidade do complexo ESCRT-III que está envolvido no tempo de divisão celular. Essa função está relacionada com a distribuição espaço-temporal durante estágios tardios da citocinese. De acordo com a literatura, CHMP4C funcionaria em um complexo juntamente com a proteína de parada celular Aurora-B que previne a resolução prematura das pontes intracelulares do cromossomo e o acúmulo de danos no DNA (Figura 23) (CARLTON et al., 2012b).

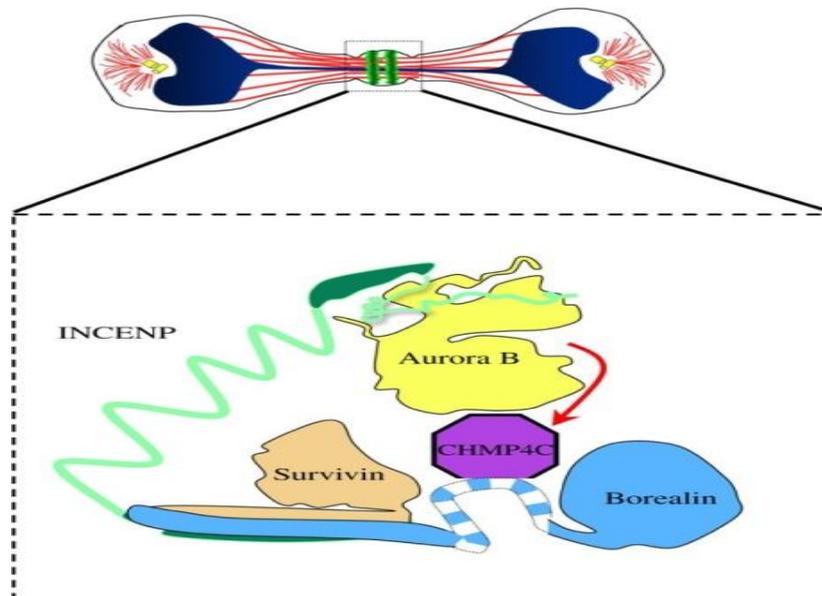


Figura 23 – Ilustração do envolvimento da proteína CHMP4C no complexo Aurora B checkpoint para o *timing* da divisão celular (CARMENA, 2012).

Baseado nisso, iniciamos nossos testes transfetando as células com o vetor contendo a sequência do gene CHMP4C com ambos os alelos para serem expressos em células iOSE4^{CMYC}. A fim de obter um resultado mais próximo da realidade, realizamos os testes em células com expressão estável após a transfecção por meio de seleção utilizando Geneticina (G418). No teste realizado, observamos variações na estabilidade (meia vida) da proteína CHMP4C comparando os alelos transfetados, referência G e de risco A (Figura 24).

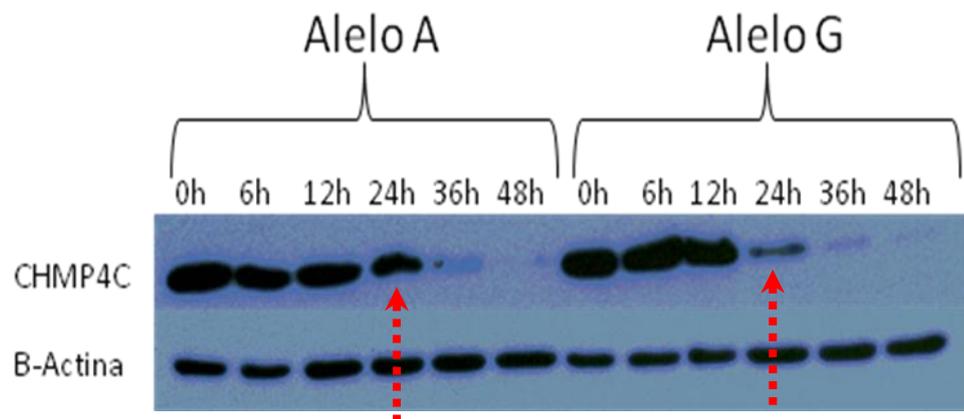


Figura 24 – Imunoblot para detecção da variação da estabilidade da proteína CHMP4C alelo dependente com proteínas da célula iOSE4^{CMYC} com transfecção estável. O gene CHMP4C foi clonado no vetor pCVM6-entry em duas versões para conter os dois alelos (A e G) do SNP rs35094336, fundida ao epítipo DDK(Flag). A extração e marcação com o anticorpo anti-DDK(flag) foi realizada após 0h, 6h, 12h, 24h, 36h, 48h da adição de ciclohexamida. B-actina foi utilizada como normalizador.

O resultado do experimento mostra que há uma alteração significativa na estabilidade da proteína entre os alelos, sendo o alelo A mais estável em relação ao alelo G (Figura 24).

6 DISCUSSÃO

Mais de 30 *loci* foram identificados através do GWAS para a susceptibilidade ao câncer de ovário, gerando a necessidade de análise funcional pós-GWAS. Sendo assim, estudos são realizados para que identifiquem e relacionem os *loci* detectados com a sua funcionalidade. Nosso objetivo foi identificar alelos de risco, a partir de uma lista de SNPs candidatos gerada previamente, que possam ser caracterizados funcionalmente como causais e elucidar os possíveis mecanismos envolvidos no *locus* de risco 8q21 para o câncer de ovário.

Apesar de o GWAS detectar o SNP associado ao risco da doença, não necessariamente ele é o agente funcional para o desenvolvimento da mesma. Dessa forma, é fundamental um novo sequenciamento genômico do *locus* encontrado previamente, denominado de *fine-mapping*. O *fine-mapping* é uma estratégia minuciosa para identificar o SNP funcional associado ao risco da doença. Pharoah e colaboradores (dados não publicados) combinaram estudos anteriores de GWAS com consórcios mundiais para aumentar a significância estatística, bem como ampliar o número de SNPs analisados. Isso foi possível visto que os SNPs estão concentrados em uma região menor, o que facilita o sequenciamento de um maior número de SNPs. Ainda, utilizaram dados de imputação de SNPs que estão sabidamente associados ao haplótipo. Por fim, nove SNPs causais candidatos foram encontrados com probabilidade de 100:1 (rs74544416, rs787224141, rs78740005, rs111683632, rs137960856, rs74758321, rs76837345, rs35094336 e rs11782652).

Uma vez identificados os SNPs causais candidatos, foi realizada uma análise *in silico* com o objetivo de obter informações para direcionar os experimentos funcionais para cada SNP de acordo com as suas características. Nesse estudo encontramos oito SNPs localizados em regiões intrônicas do gene *CHMP4C* e um SNP no éxon 5 do gene *CHMP4C*. É sabido que regiões de íntrons, apesar de não terem participação na formação da proteína final, estão envolvidas na regulação de genes proximais e distais. De fato, ao coletar informações em bancos de dados para regiões com atividade de promotores, de *enhancer*, de ligação de fatores de transcrição e de ação da DNase, quase todos os SNPs estão correlacionados a essas regiões específicas em células de ovário. Isso indica que há razão para que

esses SNPs estivessem entre os SNPs funcionais de risco associado ao câncer de ovário, o que nos deu base para prosseguir com experimentos funcionais.

Apesar de ser relatado que a maioria dos SNPs encontrados em estudos de GWAS e fine-mapping estão em regiões intrônicas ou intergênicas, para esse estudo, curiosamente um SNP – rs35094336 – está em uma região codificante do gene *CHMP4C* (éxon 5). As alterações em regiões codificantes têm maior apelo por estarem diretamente ligadas à produção de proteínas e com isso modificações na fita de DNA que podem levar a alterações fenotípicas. Nesse intuito, identificamos que o SNP rs35094336 tem predição para provável dano na proteína por conter uma alteração *missense*, que se caracteriza pela troca do aminoácido alalina por treonina. Stehr e colaboradores (2011) mostraram que a inativação de supressores de tumores correlaciona-se frequentemente com alterações na proteína, enquanto o aumento da atividade de oncogenes está relacionado a alterações específicas em local funcional, conseqüentemente, padrões distintos de alterações podem potencialmente serem usados para pré-classificar genes que são associados ao câncer, mas que ainda não possuem função conhecida (STEHR et al., 2011).

A organização espacial do genoma é intimamente ligada à sua função biológica. No núcleo de células eucarióticas, cromossomos em interfase ocupam territórios distintos e inúmeros modelos têm tentado propor como os cromossomos se dobram dentro desses territórios (CREMER; CREMER, 2010). Avanços recentes em tecnologia genômica têm impulsionado estudos da organização genômica em três dimensões. Em particular, Hi-C tem sido introduzido como método identificador de interação de cromatina no genoma, elucidando que regiões genômicas mesmos distantes linearmente podem estar próximas devido à conformação espacial e com isso haver regulação gênica (LIEBERMAN-AIDEN et al., 2009). Nesse intuito, analisamos se a região em que os SNPs se encontram está relacionada com outros genes, além do *CHMP4C*. No entanto, não há evidências, baseado na TAD (*topological associated domain*), de que há regiões interagindo com os SNPs. Dessa forma, delineamos os experimentos apenas para a região dos SNPs e o gene *CHMP4C*.

Na tentativa de comprovar a funcionalidade real dos SNPs iniciamos os testes *in vitro*. Utilizando o protocolo de varredura para *enhancer* foi possível detectar a

presença de regiões regulatórias transcricionais que são ativadas de maneira célula-específicas (BUCKLEY et al., 2015). Nenhuma região foi identificada com atividade. Com a finalidade de confirmar o resultado obtido pelo método de luciferase, lançamos mão de um segundo método, STARR-seq. Esta ferramenta identifica regiões com atividades regulatórias através de uma tecnologia mais robusta (MUERDTER; BORYŃ; ARNOLD, 2015). Entretanto, as regiões contendo os SNPs também não mostraram nenhum tipo de atividade nos testes realizados. Dessa forma, podemos afirmar que, para as células de epitélio superficial de ovário normal, os SNPs candidatos não possuem efeito regulatório.

Os locais de ligação no DNA, do inglês “*DNA binding sites*”, são regiões em que outras moléculas podem se ligar. Essas regiões são definidas como sequências curtas de DNA (normalmente entre 4 a 30 pares de base), podendo compreender até 200 pares de base para regiões recombinantes). Esses locais são normalmente associados com proteínas ou complexos especializados, conhecidos como fatores de transcrição e que estão associados à regulação transcricional (HALFORD; MARKO, 2004). O teste de EMSA identificou dois SNPs com ligação de proteínas nucleares dependente do alelo presente, rs137960856 e rs111683632. O SNP rs137960856 é uma deleção de 4 pares de base (GTGA), a predição usando PROMO e JASPAR juntos indicam uma lista de onze candidatos que se ligam na sonda com o alelo mais frequente (LRX-Alpha; NF-1; HNF-1A; HNF-3beta; FIGLA; GATA3; ISL2; JDP2; KLF5; NKX2-8; TCF4; ZEB1). Para o SNP rs111683632 a predição indica que os fatores de transcrição ligados à sonda contendo o alelo mais frequente são: LXR-alpha; VDR; HNF-1A; C-jun; CEBP; E2F1 e 4, MGA. NRF1, SP1, SREBF1 e 2.

A proteína *Liver X receptor alpha* (LXR-alpha) está presente em ambas as sondas. LXR-alpha é uma proteína receptora nuclear codificada pelo gene *NR1H3*. LXR-alpha regula sua própria supressão via indução de SREBP1 (presente entre os fatores de transcrição que se ligam na sonda do SNP rs111683632). Sua atividade está envolvida com a regulação da função de macrófagos, controle transcricional envolvendo homeostase lipídica, inflamação e ativação de hormônios tireoidianos (CHRISTOFFOLETE et al., 2010). Além desta, a proteína *Hepatocyte nuclear factor 1 homeoboxA* (HNF-1A) também é encontrada como fator de transcrição que se liga nas sondas contendo o alelo de maior frequência. HNF-1A é expresso em órgãos de

origem endodérmica, o que inclui fígado, rins, pâncreas, intestinos, estômago, timo, queratinócitos e melanócitos na pele humana (REAL HERNANDEZ et al., 2015). Estudos têm mostrado que esta proteína afeta o crescimento de células epiteliais do intestino e a diferenciação celular em geral (LUSSIER et al., 2010). Ademais, foram encontrados níveis significativamente mais baixos de HNF-1A em tumores pancreáticos e adenomas hepatocelular do que em tecidos normais adjacentes, sugerindo que HNF-1A pode ter um papel importante como um supressor tumoral (BLUTEAU et al., 2002; LUO et al., 2015).

Outra proteína importante a ser ressaltada é a GATA-3. GATA-3 pertence à família GATA de fatores de transcrição. Essa proteína regula a diferenciação de células luminal do epitélio em glândulas mamárias (KOUROS-MEHR et al., 2006). Intrigantemente, a GATA-3 é um dos três genes com mutação em mais de 10% dos casos de câncer de mama (CANCER GENOME ATLAS NETWORK, 2012). Apresenta-se envolvida no desenvolvimento do câncer de mama do tipo luminal A, nas vias com receptor de estrogênio alfa (DYDENSBORG et al., 2009; WILSON; GIGUÈRE, 2008), como também sinalização de receptores de androgênio em tumores receptor de estrogênio negativo/receptor de androgênio positivo (SANGA et al., 2009). Embora o câncer de mama e ovário sejam patologias clinicamente distintas, as recentes pesquisas de larga escala genômica têm descoberto muitas similaridades, particularmente no que se refere às alterações genéticas e epigenéticas (LONGACRE et al., 2016). Sendo assim, podemos estar diante de mais uma dessas similaridades entre essas neoplasias.

O SNP rs35094336 está localizado na região codificante (éxon 5) do gene CHMP4C, e assim as chances de ser um SNP funcional e estar envolvido no risco de câncer se tornam elevadas. Um estudo realizado em 2012 conclui que CHMP4C age como um fator regulador essencial para os pontos de verificação na seção dependente da Aurora B, mostrando que na ausência de CHMP4C as células completam a seção de forma mais rápida refletindo na resolução das pontes de cromatina de forma mais abrupta. Dessa maneira, os autores concordam que CHMP4C é um fator protetor para que não ocorra acúmulo de danos ao DNA (CARLTON et al., 2012a).

Nossos achados de instabilidade da proteína de CHMP4C confirmam que há alteração entre os alelos [G/A]. Sendo o alelo A mais estável do que o alelo G, isso nos leva a acreditar que o fator de risco associado ao câncer esteja na presença deste alelo no gene do CHMP4C. Não foi possível detectar com clareza o impacto que ocorre na célula quando a proteína está mais estável. Mais estudos são necessários para elucidar melhor mecanismos que podem ser afetados a partir desse fenômeno.

Podemos afirmar que o CHMP4C ainda é um gene/proteína em processo de descoberta, visto que, nos últimos cinco anos, apenas treze artigos foram publicados envolvendo esse gene, sendo apenas um correlacionado com o câncer (LI et al., 2015). Nesse estudo é encontrado que a expressão de Aurora B induzida por radiação aumenta a fosforilação de CHMP4C em câncer de pulmão de células não-pequenas o que mantém o ciclo celular de verificação e viabilidade celular, bem como a resistência à apoptose. Logo, a depleção de CHMP4C aumenta a sensibilidade celular para radiação caracterizando uma nova função para este gene na resistência à radiação (LI et al., 2015). Diante do exposto, é de fundamental importância que mais estudos para CHMP4C sejam realizados para elucidar melhor seu papel no mecanismo de risco/desenvolvimento no câncer de ovário.

7 CONCLUSÃO

Realizamos uma análise abrangente do panorama regulatório do *locus* de susceptibilidade 8q21 ao CaOV, bem como identificamos regiões regulatórias e gene alvo que provavelmente contribuem para o risco. Nossos dados revelaram os SNPs rs35094336, rs137960856 e rs111683632 como SNPs funcionais no *locus* 8q21. O SNP rs35094336 como modulador da estabilidade da proteína de CHMP4C. Os SNPs rs137960856 e rs111683632 apresentaram atividade alelo dependente para a ligação de fatores de transcrição. As consequências *in vivo* desses SNPs continuam incertas e serão abordadas em trabalhos futuros.

8 REFERÊNCIAS

- , - et al. The International HapMap Project. **Nature**, v. 426, n. 6968, p. 789–796, 18 dez. 2003.
- ADZHUBEI, I.; JORDAN, D. M.; SUNYAEV, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. In: **Current Protocols in Human Genetics**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013. p. 7.20.1-7.20.41.
- AMERICAN CANCER SOCIETY. Cancer Facts & Figures. **Facts & Figure 2017**, p. 1–76, 2017.
- ARNOLD, C. D. et al. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. **Science**, v. 339, n. 6123, p. 1074–1077, 1 mar. 2013.
- AUTON, A. et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 30 set. 2015.
- BIRNEY, E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature**, v. 447, n. 7146, p. 799–816, 14 jun. 2007.
- BLATTLER, A. et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. **Genome Biology**, v. 15, n. 9, p. 469, set. 2014.
- BLOW, M. J. et al. ChIP-Seq identification of weakly conserved heart enhancers. **Nature Genetics**, v. 42, n. 9, p. 806–810, 22 set. 2010.
- BLUTEAU, O. et al. Bi-allelic inactivation of TCF1 in hepatic adenomas. **Nature genetics**, v. 32, n. 2, p. 312–5, out. 2002.
- BOYD, J. Specific keynote: hereditary ovarian cancer: what we know. **Gynecol Oncol.**, v. 88, p. S8–S10, 2003.
- BRAMAN, J.; PAPWORTH, C.; GREENER, A. Site-directed mutagenesis using double-stranded plasmid DNA templates. **Methods in molecular biology (Clifton, N.J.)**, v. 57, p. 31–44, 1996.
- BROWNING, S. R.; BROWNING, B. L. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. **The American Journal of Human Genetics**, v. 81, n. 5, p. 1084–1097, nov. 2007.
- BUCKLEY, M. et al. Enhancer scanning to locate regulatory regions in genomic loci. **Nature Protocols**, v. 11, n. 1, p. 46–60, 10 dez. 2015.
- BURTON, P. R. et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. **Nature**, v. 447, n. 7145, p. 661–678, jun. 2007.
- BUSH, W. S.; MOORE, J. H. Chapter 11: Genome-Wide Association Studies. **PLoS Computational Biology**, v. 8, n. 12, p. e1002822, dez. 2012.
- CANCER GENOME ATLAS NETWORK. Comprehensive molecular portraits of

- human breast tumours. **Nature**, v. 490, n. 7418, p. 61–70, 4 out. 2012.
- CANNISTRA, S. Cancer of the ovary. **J Med**, v. 351, p. 2519–2529, 2004.
- CARLTON, J. G. et al. ESCRT-III governs the Aurora B-mediated abscission checkpoint through CHMP4C. **Science (New York, N.Y.)**, v. 336, n. 6078, p. 220–5, abr. 2012a.
- CARLTON, J. G. et al. ESCRT-III governs the Aurora B-mediated abscission checkpoint through CHMP4C. **Science (New York, N.Y.)**, v. 336, n. 6078, p. 220–5, 13 abr. 2012b.
- CARMENA, M. Abscission checkpoint control: stuck in the middle with Aurora B. **Open Biology**, v. 2, n. 7, p. 120095–120095, 18 jul. 2012.
- CHRISTOFFOLETE, M. A. et al. Regulation of thyroid hormone activation via the liver X-receptor/retinoid X-receptor pathway. **The Journal of endocrinology**, v. 205, n. 2, p. 179–86, maio 2010.
- CLARKE, L. et al. The 1000 Genomes Project: data management and community access. **Nature Methods**, v. 9, n. 5, p. 459–462, 27 abr. 2012.
- COETZEE, G. A. et al. A systematic approach to understand the functional consequences of non-protein coding risk regions. **Cell Cycle**, v. 9, n. 2, p. 256–259, 15 jan. 2010.
- COETZEE, S. G. et al. Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. **Human Molecular Genetics**, v. 24, n. 13, p. 3595–3607, 2015.
- CREMER, T.; CREMER, M. Chromosome territories. **Cold Spring Harbor perspectives in biology**, v. 2, n. 3, p. a003889, mar. 2010.
- DIGNAM, J.; LBOVITZ, R.; ROEDER, R. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. **Nucleic Acids Research**, v. 11, n. 5, p. 1475–1489, 1983.
- DURBIN, R. M. et al. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, n. 7319, p. 1061–1073, 28 out. 2010.
- DYDENSBORG, A. B. et al. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. **Oncogene**, v. 28, n. 29, p. 2634–42, 23 jul. 2009.
- EDWARDS, A. O. Complement Factor H Polymorphism and Age-Related Macular Degeneration. **Science**, v. 308, n. 5720, p. 421–424, abr. 2005.
- FREEDMAN, M. L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. **Nature Genetics**, v. 43, n. 6, p. 513–518, jun. 2011.
- GOFF, B. et al. Ovarian carcinoma diagnosis. **Cancer**, v. 89, p. 2068–2075, 2000.
- GOFF, B. et al. Development of an ovarian cancer symptom index: possibilities for earlier detection. **Cancer**, v. 109, p. 221–227, 2007.
- GUPTA, D.; LIS, C. G. Role of CA125 in predicting ovarian cancer survival - a review

- of the epidemiological literature. **Journal of Ovarian Research**, v. 2, n. 1, p. 13, 2009.
- HAINES, J. L. Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. **Science**, v. 308, n. 5720, p. 419–421, abr. 2005.
- HALFORD, S. E.; MARKO, J. F. How do site-specific DNA-binding proteins find their targets? **Nucleic acids research**, v. 32, n. 10, p. 3040–52, 2004.
- HEINTZMAN, N. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. **Nature**, v. 459, n. 7243, p. 108–112, 2010.
- HOWIE, B. N.; DONNELLY, P.; MARCHINI, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. **PLoS Genetics**, v. 5, n. 6, p. e1000529, jun. 2009.
- JELOVAC, D.; ARMSTRONG, D. K. Recent Progress in the Diagnosis and Treatment of Ovarian Cancer. v. 61, n. 3, p. 183–203, 2011.
- JERVIS, S. et al. Ovarian cancer familial relative risks by tumour subtypes and by known ovarian cancer genetic susceptibility variants. **Journal of Medical Genetics**, v. 51, n. 2, p. 108–113, 2014.
- JIA, L. et al. Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus. **PLoS Genetics**, v. 5, n. 8, p. e1000597, ago. 2009a.
- JIA, L. et al. Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus. **PLoS Genetics**, v. 5, n. 8, p. e1000597, 14 ago. 2009b.
- KIGAWA, J. New strategy for overcoming resistance to chemotherapy of ovarian cancer. **Yonago Acta Medica**, v. 56, n. 2, p. 43–50, 2013.
- KLEIN, R. J. Complement Factor H Polymorphism in Age-Related Macular Degeneration. **Science**, v. 308, n. 5720, p. 385–389, abr. 2005.
- KOUROS-MEHR, H. et al. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. **Cell**, v. 127, n. 5, p. 1041–55, 1 dez. 2006.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 4 mar. 2012.
- LAWRENSON, K. et al. In vitro three-dimensional modelling of human ovarian surface epithelial cells. **Cell proliferation**, v. 42, n. 3, p. 385–93, jun. 2009.
- LAWRENSON, K. et al. Senescent fibroblasts promote neoplastic transformation of partially transformed ovarian epithelial cells in a three-dimensional model of early stage ovarian cancer. **Neoplasia (New York, N.Y.)**, v. 12, n. 4, p. 317–25, abr. 2010.
- LEE, K. et al. CHAPTER 2 WHO histological classification of tumours of the ovary. In: **World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Breast and Female Genital Organs**. [s.l.: s.n.]. p. 117–45.
- LI, J. et al. Ovarian serous carcinoma: recent concepts on its origin and carcinogenesis. **Journal of hematology & oncology**, v. 5, n. 1, p. 8, 2012.

- LI, K. et al. CHMP4C Disruption Sensitizes the Human Lung Cancer Cells to Irradiation. **International Journal of Molecular Sciences**, v. 17, n. 1, p. 18, 24 dez. 2015.
- LI, Y. et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. **Genetic Epidemiology**, v. 34, n. 8, p. 816–834, dez. 2010.
- LIEBERMAN-AIDEN, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. **Science (New York, N.Y.)**, v. 326, n. 5950, p. 289–93, 9 out. 2009.
- LONGACRE, M. et al. A Comparative Analysis of Genetic and Epigenetic Events of Breast and Ovarian Cancer Related to Tumorigenesis. **International journal of molecular sciences**, v. 17, n. 5, 18 maio 2016.
- LUO, Z. et al. Hepatocyte nuclear factor 1A (HNF1A) as a possible tumor suppressor in pancreatic cancer. **PloS one**, v. 10, n. 3, p. e0121082, 2015.
- LUSSIER, C. R. et al. Loss of hepatocyte-nuclear-factor-1alpha impacts on adult mouse intestinal epithelial cell growth and cell lineages differentiation. **PloS one**, v. 5, n. 8, p. e12378, 24 ago. 2010.
- LYRA-JUNIOR, P. C. M. et al. GWAS in Breast Cancer. In: PHAM, P. VAN (Ed.). . **Breast Cancer - From Biology to Medicine**. 1st. ed. [s.l.] InTech, 2017.
- MAGALHÃES, J. P. How ageing processes influence cancer. **Nature Reviews Cancer**, v. 13, p. 357–365, 2013.
- MASTON, G. A.; EVANS, S. K.; GREEN, M. R. Transcriptional Regulatory Elements in the Human Genome. **Annual Review of Genomics and Human Genetics**, v. 7, n. 1, p. 29–59, set. 2006.
- MCCLUGGAGE, W. G. Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis. **Pathology**, v. 43, n. 5, p. 420–32, ago. 2011.
- MCVEAN, G. A. et al. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v. 491, n. 7422, p. 56–65, 31 out. 2012.
- MICHAILIDOU, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. **Nature Genetics**, v. 45, n. 4, p. 353–361, 27 mar. 2013.
- MODUGNO, F.; EDWARDS, R. P. Ovarian cancer: prevention, detection, and treatment of the disease and its recurrence. Molecular mechanisms and personalized medicine meeting report. **International journal of gynecological cancer : official journal of the International Gynecological Cancer Society**, v. 22, n. 8, p. S45-57, out. 2012.
- MONTEIRO, A. N. A.; FREEDMAN, M. L. Lessons from post-genome-wide association studies: Functional analysis of cancer predisposition loci. **Journal International medicine**, v. 274, n. 5, p. 414–424, 2014.
- MORA, A. et al. In the loop: promoter–enhancer interactions and bioinformatics. **Briefings in Bioinformatics**, p. bbv097, 19 nov. 2015.

MUERDTER, F.; BORYŃ, Ł. M.; ARNOLD, C. D. STARR-seq - principles and applications. **Genomics**, v. 106, n. 3, p. 145–50, set. 2015.

NEZHAT, F. R. et al. Expert Reviews. n. September, p. 262–267, 2015.

PAES, M. F. et al. A retrospective analysis of clinicopathological and prognostic characteristics of ovarian tumors in the State of Espírito Santo, Brazil. **Journal of ovarian research**, v. 4, n. 1, p. 14, 2011.

PHAROAH, P. ET AL. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. **Nature genetics**, v. 45, n. 4, p. 362–370, 2013.

PHAROAH PD, TSAI YY, RAMUS SJ, PHELAN CM, GOODE EL, LAWRENSON K, BUCKLEY M, FRIDLEY BL, TYRER JP, SHEN H, WEBER R, KAREVAN R, LARSON MC, SONG H, TESSIER DC, BACOT F, VINCENT D, CUNNINGHAM JM, DENNIS J, DICKS E; AUSTRALIAN CANCER STUDY; AUSTRALIAN OVARIAN, S. T. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. **Nature genetics**, v. 45, n. 4, p. 362–370, 2013.

PHELAN, C. M. et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. **Nature Genetics**, v. 49, n. 5, p. 680–691, 2017.

POMBO, A.; DILLON, N. Three-dimensional genome architecture: players and mechanisms. **Nature reviews. Molecular cell biology**, v. 16, n. 4, p. 245–57, abr. 2015.

POMERANTZ, M. M. et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. **Nature Genetics**, v. 41, n. 8, p. 882–884, 28 ago. 2009.

PRAT, J. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. **International Journal of Gynecology and Obstetrics**, v. 124, n. 1, p. 1–5, 2014.

PRAT, J. FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum: abridged republication. **Journal of Gynecologic Oncology**, v. 26, n. 2, p. 87, 2015.

REAL HERNANDEZ, L. M. et al. Berry Phenolic Compounds Increase Expression of Hepatocyte Nuclear Factor-1 α (HNF-1 α) in Caco-2 and Normal Colon Cells Due to High Affinities with Transcription and Dimerization Domains of HNF-1 α . **PLoS one**, v. 10, n. 9, p. e0138768, 2015.

ROBINSON, J. T. et al. Integrative genomics viewer. **Nature Biotechnology**, v. 29, n. 1, p. 24–26, jan. 2011.

ROPERS, H. H. Single gene disorders come into focus--again. **Dialogues in clinical neuroscience**, v. 12, n. 1, p. 95–102, 2010.

ROSEN, D. et al. NIH Public Access. **Ratio**, v. 36, n. 3, p. 490–499, 2010.

SALE, M. M.; MYCHALECKYJ, J. C.; CHEN, W.-M. Planning and Executing a Genome Wide Association Study (GWAS). In: [s.l: s.n.]. p. 403–418.

- SANGA, S. et al. Gene expression meta-analysis supports existence of molecular apocrine breast cancer with a role for androgen receptor and implies interactions with ErbB family. **BMC medical genomics**, v. 2, p. 59, 11 set. 2009.
- SANTIAGO, I. N. G.; NOVOA, A. C. **Estimativa 2016 Incidência de Câncer no Brasil**. [s.l: s.n.].
- SHIH, I.-M.; KURMAN, R. J. A proposed Model Based on Morphological and Molecular Genetic Analysis. **Ovarian Tumorigenesis**, v. 164, n. 5, p. 1511–1518, 2004.
- SLATKIN, M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. **Nature Reviews Genetics**, v. 9, n. 6, p. 477–485, 1 jun. 2008.
- SOSLOW, R. A. Histologic subtypes of ovarian carcinoma: an overview. **International journal of gynecological pathology: official journal of the International Society of Gynecological Pathologists**, v. 27, n. 2, p. 161–74, abr. 2008.
- SPAIN, S. L.; BARRETT, J. C. Strategies for fine-mapping complex traits. **Human Molecular Genetics**, v. 24, n. R1, p. R111–R119, 2015.
- STEHR, H. et al. The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. **Molecular Cancer**, v. 10, n. 1, p. 54, 2011.
- STRACHAN, T.; READ, A. P. **Human Molecular Genetics**. 4th. ed. New York: Garland Science/Taylor & Francis Group, 2011.
- TRYNKA, G. et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. **Nature Genetics**, v. 43, n. 12, p. 1193–1201, 6 nov. 2011.
- VANLIERE, J. M.; ROSENBERG, N. A. Mathematical properties of the measure of linkage disequilibrium. **Theoretical Population Biology**, v. 74, n. 1, p. 130–137, ago. 2008.
- VISEL, A.; ET AL. CHIP-seq accurately predicts tissue-specific activity of enhancers. **Nature**, v. 457, n. 7231, p. 854–8, 2009.
- VOIGHT, B. F. et al. The Metabochip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. **PLoS Genetics**, v. 8, n. 8, p. e1002793, 2 ago. 2012.
- WILSON, B. J.; GIGUÈRE, V. Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. **Molecular cancer**, v. 7, p. 49, 4 jun. 2008.
- ZHANG, Y. et al. Model-based Analysis of CHIP-Seq (MACS). **Genome Biology**, v. 9, n. 9, p. R137, 2008.