

**UNIVERSIDADE NOVE DE JULHO
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA - CIÊNCIAS DA SAÚDE**

GIULIANA YUKARI SEO

**O PAPEL DO POLIMORFISMO GENÉTICO DO
EGF+61 A/G NO RISCO DE DESENVOLVIMENTO
E PROGRESSÃO DO ADENOCARCINOMA
COLORRETAL**

SÃO PAULO

2022

GIULIANA YUKARI SEO

**PAPEL DO POLIMORFISMO GENÉTICO DO EGF+61 A/G NO
RISCO DE DESENVOLVIMENTO E PROGRESSÃO DO
ADENOCARCINOMA COLORRETAL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em
Medicina - Ciências da Saúde da Universidade Nove de Julho, como requisito
para obtenção do grau de mestre em ciências da saúde

Orientador: Prof. Dr. Ramon Andrade Bezerra de Mello

**SÃO PAULO
2022**

Seo, Giuliana Yukari.

O papel do polimorfismo genético do EGF+61 A/G no risco de desenvolvimento e progressão do adenocarcinoma colorretal. / Giuliana Yukari Seo. 2022.

49 f.

Dissertação (Mestrado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2022.

Orientador (a): Prof. Dr. Ramon Andrade Bezerra de Mello.

1. Neoplasia do cólon. 1. EGFR. 2. EGF+61. 3. Polimorfismo genético. 4. Mutação. 5. Câncer.

I. Mello, Ramon Andrade Bezerra de.

II. Título

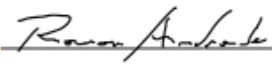
CDU 616

São Paulo, 13 de dezembro de 2022.

TERMO DE APROVAÇÃO

Aluna: **GIULIANA YUKARI SEO**

Título da dissertação: **O PAPEL DO POLIMORFISMO GENÉTICO DO EGF+61 A/G NO RISCO E PROGNÓSTICO NO ADENOCARCINOMA DO CÓLON.**

Presidente: PROF. DR. RAMON ANDRADE BEZERRA DE MELLO 

Membro: PROF. DR. JOSÉ ANTÔNIO SILVA JUNIOR 

Membro: PROF. DR. DANIEL HUMBERTO POZZA 

Dedico este trabalho ao meu noivo Ícaro.

AGRADECIMENTOS

Agradeço a Deus por guiar e iluminar meu caminho. Aos meus pais por serem minha base.

A meu orientador, Prof. Dr. Ramon Andrade Bezerra de Mello, pela a orientação no trabalho de pesquisa e pelos ensinamentos em oncologia clínica. Aos professores da Universidade Nove de Julho pela excelência no ensino e qualidade técnica de todos.

Aos alunos de medicina Rafaela Cremonese, Joao Vittor Luciano e Bruno Antunes pelo apoio e desenvolvimento de projetos e artigos na área de oncologia. Aos biomédicos Gildo Junior e Estella Gudín pelo suporte na área laboratorial.

Agradeço a Universidade Nove de Julho pela bolsa de estudos, e a possibilidade de me tornar um pouco mais sábia do que quando comecei esta trajetória.

LISTA DE TABELAS E FIGURAS

Figura 1: Sequencia adenoma-carcinoma do cólon normal ao carcinoma

Figura 2: Exemplificação do mecanismo de ativação e funcionamento do EGFR

Tabela 1: Classificação TNM pelo American Joint Comitee on Cancer 8 edicao

Tabela 2: Agrupamento prognostico pelo AJCC

Tabela 3: Classificação de funcionalidade pelo *Eastern Cooperative Oncology Group*

Tabela 4: Sequências dos *primers* e sondas para a detecção do polimorfismo no gene EGF+61 (A/G)

Tabela 5: Sequência dos primers e sondas para a detecção do polimorfismo do gene EGF+61 (A/G)

Figura 3: Curva de amplificação do PCR em tempo real evidenciando o genótipo AA

Figura 4: Curva de amplificação do PCR em tempo real evidenciando o genótipo AG

Figura 5: Curva de amplificação do PCR em tempo real evidenciando o genótipo GG

Tabela 6: Fluxograma do processo de seleção dos pacientes e resultados de participantes finais

Tabela 7: características clínicas dos participantes

Tabela 8: frequência dos genótipos dos polimorfismos do EGF+61 A/G nos pacientes com câncer colorretal e nos pacientes saudáveis

Figura 6: Código para criação das variáveis *dummy*

Figura 7: Colunas PCR

Figura 8: Linha de código para grupo de treino e grupo de teste

Figura 9: Modelo de treino com acurácia de 100%

Figura 10: Código para predição do *target* com os 25% dos dados restantes

Figura 11: Matriz de confusão

Figura 12: Resultado probabilidade participantes portadores do genótipo A/G de desenvolver câncer

Figura 13: Possibilidade dos pacientes com câncer colorretal portadores do genótipo AG de apresentarem progressão de doença durante o acompanhamento

LISTA DE ABREVIATURAS E SIGLAS

Akt: Serina-treonina cinase

APC : polipose adenomatosa familiar

CCR: câncer colorretal

DNA: ácido desoxirribonucleico

DNAg: ácido desoxirribonucleico genômico

EDTA: ácido etilenodiamino tetra-acético

EGF: fator de crescimento endotelial

EGF+61 A/G: polimorfismo do fator de crescimento epidérmico

EGFR: receptor do fator de crescimento endotelial

INCA: Instituto Nacional do Câncer

JNK: c-Jun N-terminal cinase

MAPK: Proteína-quinases ativadas por mitógenos

Ng: nanogramas

nM: nanômetros

°C : graus celsius

PCR: reação de cadeia polimerase

Pmolar: picomolar

SH2: Src homologo 2

SNPs: polimorfismos genéticos

TGF α : fator de crescimento transformador α

TP53: gene supressor tumoral

RESUMO

Introdução: A neoplasia do cólon é bastante frequente nos homens e nas mulheres a nível mundial, no Brasil e especificamente em São Paulo, alguns polimorfismos genéticos (SNPs) podem estar associado ao desenvolvimento de neoplasias de origem epitelial, como é o caso do polimorfismo do fator de crescimento epidérmico (*EGF*+61 A/G) **Objetivo:** Analisar o papel de polimorfismos genéticos do *EGF*+61 no risco e prognóstico no adenocarcinoma do cólon, o papel desta variante polimórfica na susceptibilidade de desenvolver adenocarcinoma do cólon e reto **Materiais e métodos:** estudo transversal caso-controle análise foi realizada por linguagem de programação Python e algoritmos de *machine learning* para realizar a modelagem estatística. **Resultados:** participaram do estudo 187 indivíduos, destes 112 do grupo caso e 72 do grupo controle, 48,21% do sexo masculino e 50,89% do sexo feminino. A média de idade do grupo caso foi de 58 anos e do grupo controle de 59 anos. Os pacientes do grupo caso 6% (7 pacientes) apresentam-se com estadio clínico I, 31% (35 pacientes) em estadio clínico II, 38% (43 pacientes) no estadio clínico III e 23% (26 pacientes) em estadio clínico IV. É possível observar que probabilisticamente 27 de 47 dos participantes sem câncer portadores do polimorfismo A/G poderão não apresentar a condição oncológica e 20 de 47 poderão desenvolver a condição, ou seja 57,45% e 42,55% respectivamente e dos pacientes com câncer e polimorfismo, 55,56% não terão progressão de doença e 44,44% poderão ter progressão. **Conclusão:** Nosso estudo revelou correlação dos pacientes portadores do polimorfismo *EGF* +61 A/G com a possibilidade de 42% de desenvolvimento de câncer colorretal ao longo da vida. Os pacientes portadores do polimorfismo e também de neoplasia possuem 44% de chance de progressão de doença.

Palavras-chave: neoplasia do cólon, EGFR, *EGF*+61, polimorfismo genético, mutação, câncer.

ABSTRACT

Introduction: Colon cancer is quite common in men and women worldwide, in Brazil and specifically in São Paulo, some genetic polymorphisms (SNPs) may be associated with the development of neoplasms of epithelial origin, such as colon polymorphism epidermal growth factor (EGF+61 A/G) **Objective:** analyze the role of genetic polymorphisms of EGF+61 in the risk and prognosis of colon adenocarcinoma, the role of this polymorphic variant in the susceptibility of developing colon and rectum adenocarcinoma **Materials and methods:** cross-sectional case-control study analysis was performed using Python programming language and machine learning algorithms to perform statistical modeling. **Results:** 187 individuals participated in the study, of which 112 were in the case group and 72 in the control group, 48.21% male and 50.89% female. The mean age of the case group was 58 years and the control group 59 years. Of the patients in the case group, 6% (7 patients) present with clinical stage I, 31% (35 patients) with clinical stage II, 38% (43 patients) with clinical stage III and 23% (26 patients) with clinical stage IV. It is possible to observe that probabilistically 27 out of 47 of the participants without cancer with the A/G polymorphism may not have the oncological condition and 20 out of 47 may develop the condition, that is 57.45% and 42.55%, respectively, of patients with cancer and polymorphism, 55.56% will not have disease progression and 44.44% may have progression. **Conclusion:** Our study revealed a correlation of patients with the EGF +61 A/G polymorphism with a 42% possibility of developing colorectal cancer throughout life. Patients with the polymorphism and also with neoplasia have a 44% chance of disease progression.

Keywords: colon cancer, EGFR, EGF+61, genetic polymorphism, mutation, cancer

SUMÁRIO

1.INTRODUÇÃO	12
2. OBJETIVOS	20
2.1. Primários	20
2.2. Secundário	20
3. MATERIAIS E MÉTODOS.....	21
3.1. Tipo de estudo, população e amostra	21
3.2. Infraestrutura e recursos humanos disponíveis.....	21
3.3. Variáveis consideradas	22
3.4. Extração de DNA	22
3.5. Reação em cadeia da polimerase em tempo real	23
3.6. Análise estatística	27
4. RESULTADOS	28
5. DISCUSSÃO	36
6. CONCLUSÃO.....	38
7.REFERENCIAS.....	39

1.INTRODUÇÃO

A redefinição dos padrões de vida, alterações demográficas, do padrão de nutrição, redução da taxa de mortalidade e de natalidade, indicam prolongamento da expectativa de vida e envelhecimento da população. Aumenta também as doenças crônico-degenerativas, dentre elas as cardiovasculares e o câncer ⁽¹⁾.

O Brasil encontra-se em transição epidemiológica, com aumento de idosos e conseqüentemente de tumores, dentre eles o câncer colorretal (CCR)⁽²⁾. De acordo com dados coletados pelo Instituto Nacional do Câncer (INCA) ⁽³⁾, estima-se cerca de 40mil casos novos no Brasil para o triênio de 2020 a 2022.

Essa incidência não é homogênea no país, predominando nas regiões Sul e Sudeste e principalmente nos Estados de São Paulo, Rio de Janeiro e Rio Grande do Sul, onde também ocorrem as maiores taxas de mortalidade por este tipo de neoplasia ^(4,5). Isso se deve a desigualdades sociodemográficas, culturais e acesso a serviços de saúde ⁽³⁾.

A mortalidade por câncer em 2013 por todo o país foi estimada em 15.415 casos; sendo 7.387 homens e 8.024 mulheres. Além dos casos já documentados, o aumento da incidência e da mortalidade já é esperado para os próximos anos, visto que entre 1979 e 2000 houve aumento médio anual de 3,5% ocupando atualmente a quinta causa de óbito por neoplasias. ^(6,7).

O câncer colorretal é consequência da prevalência de seus múltiplos fatores de risco, os quais incluem a obesidade, a alimentação rica em carne vermelha e alimentos processados, alcoolismo, tabagismo, além do envelhecimento da população nacional nas últimas décadas ^(2,8,9). Esses componentes induzem mutações cumulativas, que são peças-chave no processo de várias etapas da carcinogênese colorretal⁽¹⁰⁾.

Todas as células tumorais apresentam instabilidade genômica, seja por mutações hereditárias ou adquiridas. Estas alterações podem ocorrer em vários

níveis, seja em um único nucleotídeo, um segmento de DNA, também chamado microssatélite, genes inteiros ou até mesmo cromossomos inteiros ⁽¹¹⁻¹⁴⁾.

O acúmulo de mutações e alterações genéticas que favorecem a progressão tumoral ⁽¹⁵⁾ como a progressão adenoma-carcinoma, exemplificado na figura 1, a qual pode ocorrer através de duas importantes vias que afetam a expressão tanto genes supressores de tumores como da polipose adenomatosa familiar (APC) e o gene supressor tumoral (TP53), como oncogenes (genes KRAS, PI3KCA, BRAF e NRAS). No primeiro, alterações mutagênicas podem ocorrer através de instabilidade cromossômica, originada a partir da duplicação ou deleção de trechos cromossômicos, correspondendo a 85% dos casos registrados de CCR. No segundo, as mutações podem ocorrer por instabilidade microssatélite, corresponde a 15% dos casos, não sendo incomum ambos os eventos ocorrerem de forma concomitante como mostra a figura abaixo ⁽¹⁶⁾. Neles, além de alterações nos genes responsáveis pela supressão e progressão de tumores, podem estar presentes alterações nos mecanismos de reparo do DNA, e modificações de ordem epigenética ⁽¹⁷⁾.

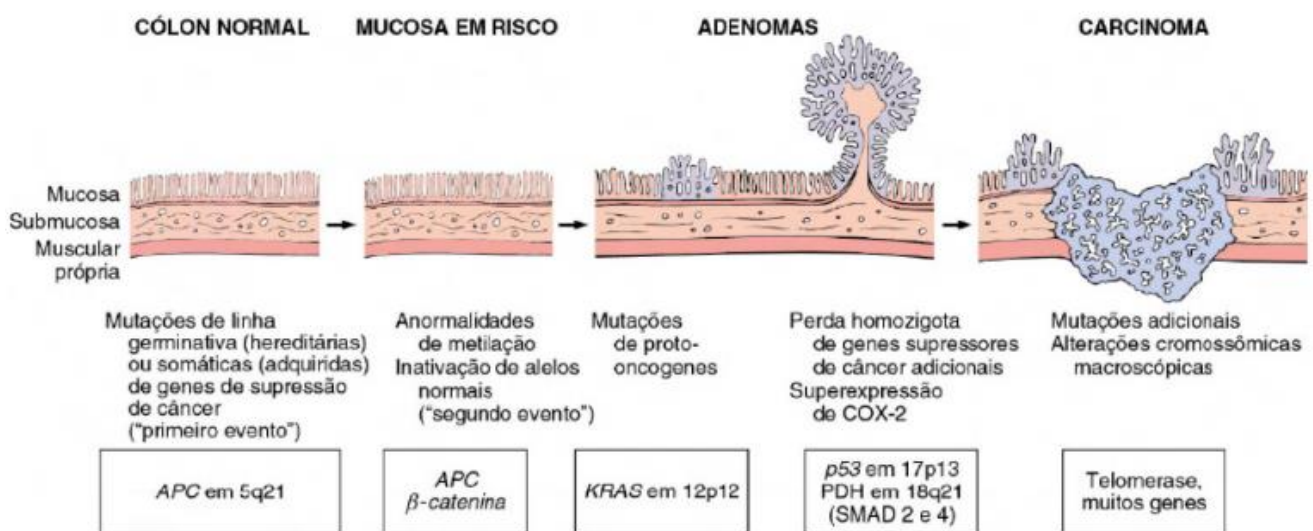


Figura 1: Sequencia adenoma-carcinoma do cólon normal ao carcinoma.

Desta forma, sendo uma doença passível de prevenção e diagnóstico precoce, torna-se importante a implantação de programas de rastreamento como a colonoscopia e pesquisa de sangue oculto nas fezes além de mudança

de estilo de vida, os quais impactam na incidência e mortalidade da patologia (18,19).

O câncer colorretal é mais comum em torno dos 60 anos de idade e apresenta uma longa fase pré-clínica, levando anos até iniciar sintomas ⁽²⁰⁾. Os principais sinais e sintomas são dor, mudança do hábito intestinal, perda de peso, anemia, massa palpável, vômitos, tenesmo, sangramento retal ⁽²¹⁾.

Para o estadiamento do câncer colorretal é utilizado o *American Joint Comitee on Cancer* (AJCC), que atualmente encontra-se na oitava edição, e classifica o tumor de acordo com o tamanho (T), invasão linfonodal (N) e metástase a distancia (M), assim gerando o agrupamento prognóstico, como segue na tabela 1 e 2 ⁽²²⁾.

T	Tumor primário
T0	Sem evidencia de tumor
Tis	Carcinoma <i>in situ</i>
T1	Tumor invade submucosa
T2	Tumor invade muscular própria
T3	Tumor invade através da muscular própria
T4	Tumor invade peritoneo visceral ou adere a órgão/ estrutura adjacente
T4a	Invade peritoneo visceral
T4b	Invade órgão adjacente
N	Linfonodos regionais
N0	Sem acometimento
N1	1 a 3 linfonodos positivos
N1a	1 linfonodo
N1b	2 ou 3 linfonodos
N1c	Sem linfonodos mas há depósitos tumorais
N2	4 ou mais linfonodos regionais positivos
N2a	4 a 6 linfonodos regionais

N2b	7 ou mais linfonodos regionais
M	Metástase a distancia
M0	Sem metástase a distancia
M1	Metástase a distancia presente
M1a	Metástase em um órgãos exceto peritônio
M1b	Metástase em 2 ou mais órgãos exceto peritônio
M1c	Metástase peritoneal com ou sem acometimento de outros órgãos

Tabela 1: Classificação TNM pelo *American Joint Comitee on Cancer* 8 edição

	T	N	M
Estadio 0	Tis	N0	M0
Estadio I	T1, T2	N0	M0
Estadio IIA	T3	N0	M0
Estadio IIB	T4a	N0	M0
Estadio IIC	T4b	N0	M0
Estadio IIIA	T1, T2 T1	N1/N1c N2a	M0
Estadio IIIB	T3-T4a T2-T3 T1-T2	N1/N1c N2a N2b	M0
Estadio IIIC	T4a T3-T4a T4b	N2a N2b N1/N2	M0
Estadio IVA	Qualquer T	Qualquer N	M1a
Estadio IVB	Qualquer T	Qualquer N	M1b
Estadio IVC	Qualquer T	Qualquer N	M1c

Tabela 2: Agrupamento prognostico pelo *American Joint Comitee on Cancer* 8 edição

Outro fator relacionado ao desenvolvimento tumoral a nível celular são os fatores de crescimento e seus receptores ⁽²³⁾, as células tumorais podem

expressar maior quantidade de receptores de superfície, assim, aumentando sua sensibilidade a estes fatores ⁽²⁴⁾. Um exemplo é o fator de crescimento endotelial (EGF) e seu receptor (EGFR), muito estudados pois são elementos envolvidos em uma cascata de sinalização responsável por proliferação, diferenciação e migração celular ^(25,26).

O fator de crescimento endotelial (EGF) foi descoberto em 1962 por Cohen ⁽²⁷⁾ e, desde a caracterização inicial da sua interação com receptor de superfície celular (EGFR), em 1975 por Carpenter et al ⁽²⁸⁾, este tem sido extensivamente estudado ⁽²⁹⁾. Sendo considerado um fator de crescimento proteico pertencente à família EGF, a qual consiste de, no mínimo, oito membros ⁽³⁰⁾

O receptor do fator de crescimento epidérmico (EGFR) é uma proteína transmembranar ativada pela ligação de seus ligantes específicos, incluindo o fator de crescimento epidérmico e o fator de crescimento transformador α (TGF α) ⁽³¹⁾. Após ativado passa por uma transição de uma forma monomérica inativa para um homodímero ativo ⁽³¹⁾. Além de formar homodímeros após a ligação do ligante, o EGFR pode parear com outro membro da família de receptores ErbB, para criar um heterodímero ativado ⁽³²⁾.

O EGFR dimerizado então, estimula a atividade intracelular da proteína tirosina quinase. Ocorre a autofosforilação de vários resíduos de tirosina (Y) do EGFR ⁽³³⁾, o que provoca ativação e sinalização a jusante por várias outras proteínas que se associam às tirosinas fosforiladas por meio de seus próprios domínios Src homolog 2 (SH2) de ligação à fosfotirosina. Essas proteínas de sinalização a jusante iniciam várias cascatas de transdução de sinal, principalmente as vias Proteína-quinases ativadas por mitógenos (MAPK), Serina-treonina cinase (Akt) e c-Jun N-terminal cinase (JNK), levando à síntese de DNA e proliferação celular ⁽³⁴⁾. E modulam fenótipos como migração celular, adesão e proliferação, como demonstrado na figura 2.

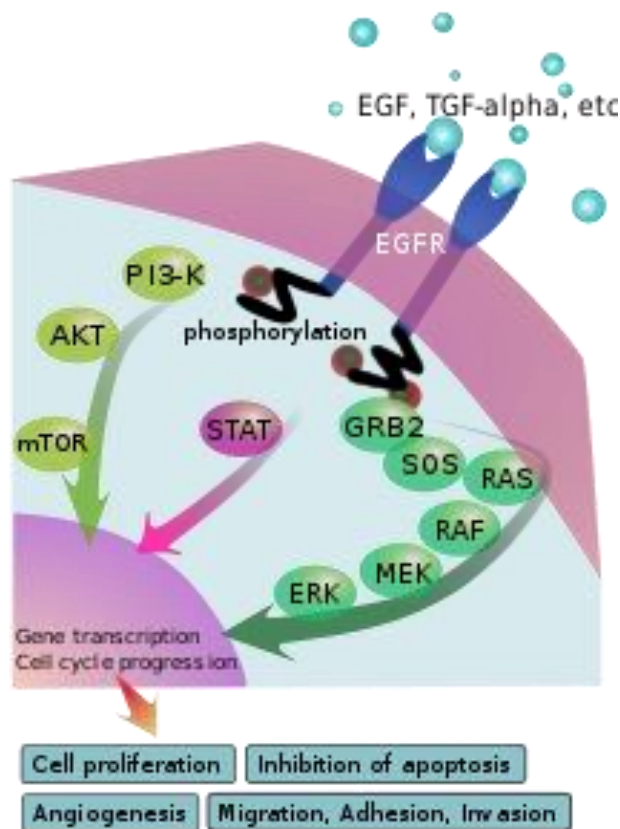


Figura 2: Exemplificação do mecanismo de ativação e funcionamento do EGFR

Cerca de 50% dos CCR metastáticos tem mutações no gene RAS ⁽³⁵⁾. Portanto, é frequente a desregulação do EGFR no câncer colorretal seja por aumento de expressão, amplificação gênica e mutação, assim ocorre o aumento do crescimento e desenvolvimento tumoral, da angiogenese, além da inibição da apoptose, sendo esta uma mutação pesquisada nos pacientes, com implicação terapêutica ⁽³⁶⁻³⁸⁾. Acredita-se que até 60% dos tumores de cólon possam ter o receptor EGF super-expresso ^(38.39)

Atualmente foram desenvolvidas medicações específicas para bloquear este receptor e sua via de sinalização, são os anticorpos monoclonais e inibidores de tirosino-kinase ⁽⁴⁰⁾. Fazem parte do acervo para o tratamento de câncer colorretal, o Cetuximab, um anticorpo monoclonal quimérico que compete para ligação ao domínio extracelular do EGFR, assim previne sua ativação, já aprovado em associação à quimioterapia em pacientes com câncer colorretal metastático ⁽⁴¹⁻⁴³⁾

O Panitumumab, um anticorpo monoclonal recombinante que se liga ao receptor do EGFR humano inibindo a sua auto-fosforilação também aprovado em para pacientes metastáticos refratários a quimioterapia ⁽⁴⁴⁾ Este medicamento em monoterapia de segunda linha ou em associação a quimioterapia em primeira linha demonstrou ganho principalmente de sobrevida livre de progressão mas não houve ganho significativo de sobrevida global ⁽⁴⁵⁾.

O estudo BOND evidenciou a eficácia do cetuximab, em pacientes com neoplasia colorretal metastáticos em combinação com irinotecano, além do benefício do panitumumab no tratamento e sobrevida livre de progressão destes pacientes. ^(46.47)

O gene que codifica o fator de crescimento endotelial foi inicialmente identificado no cromossomo 4q por Morton ⁽⁴⁸⁾. Logo após, Shahbazi, em 2002, com ênfase na substituição de adenina por guanina na posição +61 (EGF+61), supôs uma maior expressão sérica de EGF devido à proximidade do locus +61G e a região envolvida na regulação do gene EGF, assim, associou este polimorfismo ao risco e desenvolvimento de melanoma ^(49.52).

Su et al, em 2014 demonstrou associação do polimorfismo com tumor de laringe e faringe ⁽⁵³⁾, também foi positiva a associação com câncer de pulmão não pequenas células no estudo na população de Portugal de De Mello, 2012, sendo indicado a detecção de EGFR e mutação KRAS nos pacientes metastáticos portadores desta neoplasia ⁽⁵⁴⁾.

Em 2012 Araujo associou a variante G do polimorfismo com risco de câncer de colo de útero ⁽⁵⁵⁾. Uma metanálise em hepatocarcinoma também correlacionou o EGF+61 A/G com o risco da neoplasia, especialmente na população asiática ⁽⁵⁶⁾. Também houve relação com o polimorfismo e risco de desenvolvimento de gliomas ^(57.58).

Já na população brasileira não houve associação do polimorfismo e risco de câncer de pulmão, de acordo com Laus, 2019 ⁽⁵⁹⁾. No estudo em câncer de mama triplo negativo, não houve evidencia de amplificação do gene EGFR no estudo de Nakajima, 2014 ⁽⁶⁰⁾. Acredita-se que indivíduos com o genótipo A/A produzam menos EGF do que os indivíduos que tem a variante com G ⁽⁶¹⁾. Assim, observamos que há resultados tanto positivos quanto negativos com relação a este polimorfismo.

Apesar do desenvolvimento destas novas medicações para o tratamento do câncer colorretal, continua sendo mais benéfico em termos de saúde coletiva, focar na prevenção ao invés do tratamento, até pelo fato de o tratamento nem sempre atingir níveis esperados ⁽⁶²⁾.

Novos métodos para colaborar na detecção e rastreamento de neoplasias têm sido desenvolvidos, dentre eles o *machine learning* ⁽⁶³⁻⁶⁵⁾, um método que cria e automatiza modelos analíticos através de algoritmos baseados em exemplos e observações ⁽⁶⁶⁻⁶⁸⁾. Neste método, o modelo de inteligência artificial, que une matemática com computação, aprende a partir dos dados fornecidos pelos pesquisadores, por exemplos fornecidos, ao invés de regras, desta forma o método torna-se capaz de prever prognóstico, diagnóstico e até mesmo tratamento ⁽⁶⁹⁾.

Em câncer de mama por exemplo, utiliza-se como rastreamento a mamografia, ressonância magnética, ultrassom de mama e biopsia, pelo método de *machine learning* pode-se prever o risco do paciente de portar a neoplasia ou não, evitando assim gastos desnecessários ⁽⁷⁰⁾.

De forma semelhante foi estudado o antígeno prostático específico (PSA), o toque retal de próstata, ultrassom de próstata trans-retal e correlação com biopsias positivas para câncer de próstata, o método de *machine learning* foi capaz de evitar 50% de biopsias desnecessárias ⁽⁷¹⁾.

Sendo o Hospital São Paulo - UNIFESP, um centro científico de referência em Oncologia no Sudeste e no Brasil, com um considerável número de pacientes portadores de neoplasias de cólon e reto acompanhados em seus serviços clínicos. Os pacientes com adenocarcinoma colorretal atendidos nesse hospital ainda não foram submetidos a uma pesquisa por polimorfismos genéticos de *EGF+61 A/G* e não há estudos conhecidos do gênero na população paulista.

Frente a isso, este trabalho justifica-se na importância de trazer um estudo que possa revelar o perfil genético desse polimorfismo na população, identificar sua relação com o adenocarcinoma de cólon e levantar possibilidades de prevenção e terapêutica aliadas à oncogenética em estudos futuros.

2. OBJETIVOS

2.1. Primários

Analisar o papel do polimorfismo genético do *EGF*+61 A/G no risco de de desenvolvimento do adenocarcinoma do cólon e reto e o papel desta variante polimórfica com a probabilidade de progressão de doença nos pacientes portadores da condição oncológica.

2.2. Secundário

Aplicar o método de *machine learning* no risco prognóstico dos pacientes com adenocarcinoma de cólon e reto

3. MATERIAIS E MÉTODOS

3.1. Tipo de estudo, população e amostra

Trata-se de um estudo transversal caso-controle, o estudo foi aprovado pelo comitê de ética do hospital São Paulo número do parecer 4.199.733. Todos os participantes foram maiores de 18 anos de idade acompanhados no Hospital São Paulo e que concordaram em participar após assinatura do Termo de Consentimento Livre e Esclarecido.

O primeiro grupo experimental constou de pacientes com diagnóstico de adenocarcinoma de cólon ou reto comprovado por análise histopatológica e o segundo grupo de pacientes saudáveis, sem neoplasia conhecida que aguardavam na fila do Laboratório Central, para realizar coleta de exame de sangue.

Outros dados sobre a clínica desses pacientes foram adquiridos pela revisão de prontuário. Para um nível de confiança de 95% foi estimado um cálculo amostral de 139 pacientes no grupo com neoplasia e 139 pacientes no grupo de risco.

3.2. Infraestrutura e recursos humanos disponíveis

A análise pré-analítica dos pacientes que seriam posteriormente selecionados para o estudo foi realizada a partir dos recursos do setor de arquivos do Hospital São Paulo, garantindo assim as respectivas informações necessárias dos pacientes.

Para a realização das etapas de extração e amplificação das amostras o Laboratório de Biologia Molecular da UNIFESP disponibilizou os recursos como pipetas, ponteiros e tubos necessários além do nano espectrofotômetro e termociclador 7500 Real Time para realização do sequenciamento do polimorfismo estudado. Materiais para extração do DNA foram adquiridos pelos pesquisadores.

3.3. Variáveis consideradas

As variáveis consideradas para os pacientes com neoplasia foram: Idade à admissão, diagnóstico histológico, estadiamento inicial de acordo com o *American Joint Comitee on Cancer* oitava edição, localização do tumor primário (cólon descendente ou ascendente, reto), classificação funcional pelo *Eastern Cooperative Oncology Group* (ECOG) tabela 3, datas das progressões, data do óbito.

Grau	ECOG performance
0	Ativo, sem restrição a atividades
1	Restricao a atividades físicas extremas
2	Capaz de auto cuidado, independente por mais de 50% do tempo
3	Dependente para cuidados, confinado a cama mais do que 50% do tempo
4	Incapacitado, totalmente dependente para cuidados, confinado a cama
5	Morte

Tabela 3: Classificação de funcionalidade pelo *Eastern Cooperative Oncology Group*

Foram excluídos do estudo pacientes com diagnostico histopatológico que não seja de adenocarcinoma, pacientes cuja localização do tumor inicial não esteja no cólon ou reto. Além dos casos que não obtiveram sucesso na extração e reação em cadeia polimerase (PCR) em tempo real.

3.4. Extração de DNA

Para a extração de Ácido Desoxirribonucleico (DNA) de leucócitos foram coletados 5 mL de sangue periférico, contendo, cada um deles ácido etilenodiamino tetra-acético (EDTA) a 15% para impedir a coagulação

sanguínea. As amostras foram manipuladas a partir das etapas de extração e purificação de DNA utilizando o PROTOCOLO 2 - *30 minutes DNA extraction from blood sample* do Kit de Extração de DNA Easy-DNA™ gDNA Purification Kit; Thermo Fisher Scientific, o qual consiste nas etapas de isolamento e precipitação do DNA para que se chegasse em um valor mínimo de 30 nanogramas (ng) de DNA, quantificado pelo nano espectrofotômetro DS-11 Denovix.

3.5. Reação em cadeia da polimerase em tempo real

As amostras de DNA genômico (DNAg) foram submetidas a uma reação de reação de cadeia polimerase (PCR) em tempo real utilizando a metodologia TaqMan® (Applied Biosystem). Foram empregados sondas iniciadores e sondas específicas (Applied Biosystem, Foster City. CA) correspondentes à região próxima do gene EGF+61, de acordo com a metodologia previamente descrita.

Para as reações foram utilizados um sistema de PCR em tempo real. As condições suficientes empregadas para uma reação foram: 1 picomolar (pmolar) dos iniciadores direto e reverso; 100 nanometros (nM) de sondas marcadas com VIC® e FAM; 30ng de DNAg e água livre de nucleases (Sigma-Aldrich) e quantidade suficiente para volume final de 50microlitros. As condições de ciclagem foram: 50 graus celsius (°C) por 2 minutos, 95°C por 10 minutos, seguidos de 40 ciclos de 95°C por 15 segundos, 60°C por 1 minuto. Em todas as reações foram utilizados controles negativos e três controles positivos, sendo uma amostra de cada genótipo. Cada amostra foi analisada em duplicata.

A tabela 4 mostra as sequências dos *primers* e sondas empregados para a detecção do polimorfismo no gene EGF +61 (A/G)

Primer/Sonda	Seqüência (5' - 3')
Forward primer	CTT CTT TCA GCC CCA ATC CA
Reverse primer	GAA AAA GAA AGG AAG TGA TGG A
A. Sonda específica (reporter dye FAM)	FAM AGG GTT GTA GCT GGA A
G – sonda específica (reporter dye VIC)	VIC AGG GTT GTG GCT GGA

Tabela 4: Sequências dos *primers* e sondas para a detecção do polimorfismo no gene EGF+61 (A/G)

As amostras de DNA genômico primeiramente extraídos e quantificados foram submetidas a uma reação de PCR em tempo real utilizando a metodologia TaqMan, a qual consiste na emissão de fluorescência gerada proporcionalmente à cada amplificação do alvo estudado.

A metodologia se baseia na utilização de uma sonda específica para um alvo estudado, tendo em uma das suas extremidades um fluoróforo (VIC/FAM) e na outra extremidade uma molécula, denominada *Quencher*, na qual ao haver a ligação entre a sonda e a fita simples do DNA a sonda se degrada devido à atividade da taq polimerase, gerando conseqüentemente a separação do *quencher* do fluoróforo liberando a devida luminosidade.

Para a reação de amplificação foi necessário 1,0ul de amostra contendo um valor mínimo de 30ng de DNA, 0,5ul das sondas e *primers* específicos para a região de interesse, neste caso o EGF+61 e 5,0 ul do Taqman *genotyping master mix*.

Primer/ Sonda	Sequência 5' 3'
Foward Primer	CTT CTT TCA GCC CCA ATC CA
Reverse Primer	GAA AAA GAA AGG AAG TGA TGG A

Sonda Específica – reporter dye FAM	FAM AGG GTT GTA GCT GGA A
Sonda Específica – repórter dye VIC	VIC AGG GTT GTG GCT GGA

Tabela 5: Sequência dos Primers e Sondas para a detecção do polimorfismo do gene EGF+61 (A/G)

Foi utilizado também os controles a partir de amostras que expressam as variantes dos genótipos: AG, AA e GG.

Após a adição das amostras e dos reagentes a placa foi levada para uma rápida centrifugação, em seguida ao termociclador, o qual realizou uma ciclagem de duração média de 120 minutos, com a ocorrência das etapas de desnaturação, a qual consiste em degradar a dupla fita de DNA em uma temperatura de 95°C; anelamento, que consiste na implementação da sonda e do primer específico na nossa região alvo em uma temperatura de 60°C; extensão, a qual consiste na atividade da Taq polimerase, realizando a implementação dos nucleotídeos necessários para moldar uma nova região alvo.

As amostras foram denominadas através de sequência numérica de 1 até 187 e seus resultados foram analisados através da expressão das curvas que cada amostra apresentavam, com base no método de discriminação alélica. Observamos os exemplos de curva de amplificação do PCR em tempo real nas figuras abaixo, na figura 3 há o genótipo AA, na figura 4 observamos o genótipo AG e na figura 5 a curva de PCR do genótipo GG.

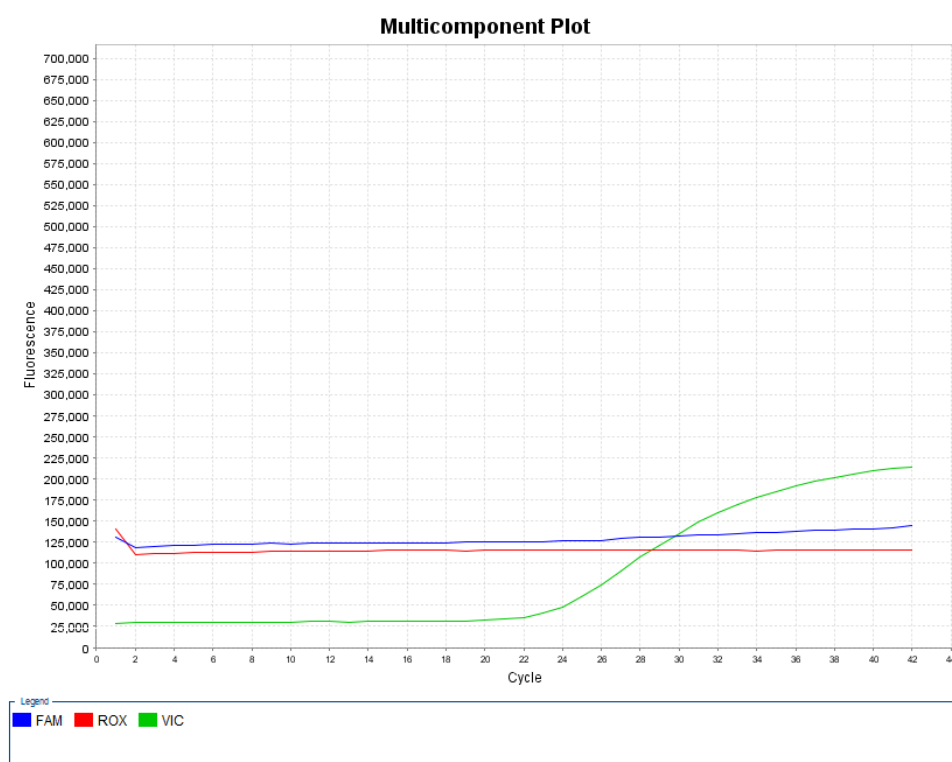


Figura 3: Curva de amplificação do PCR em tempo real evidenciando o genótipo AA

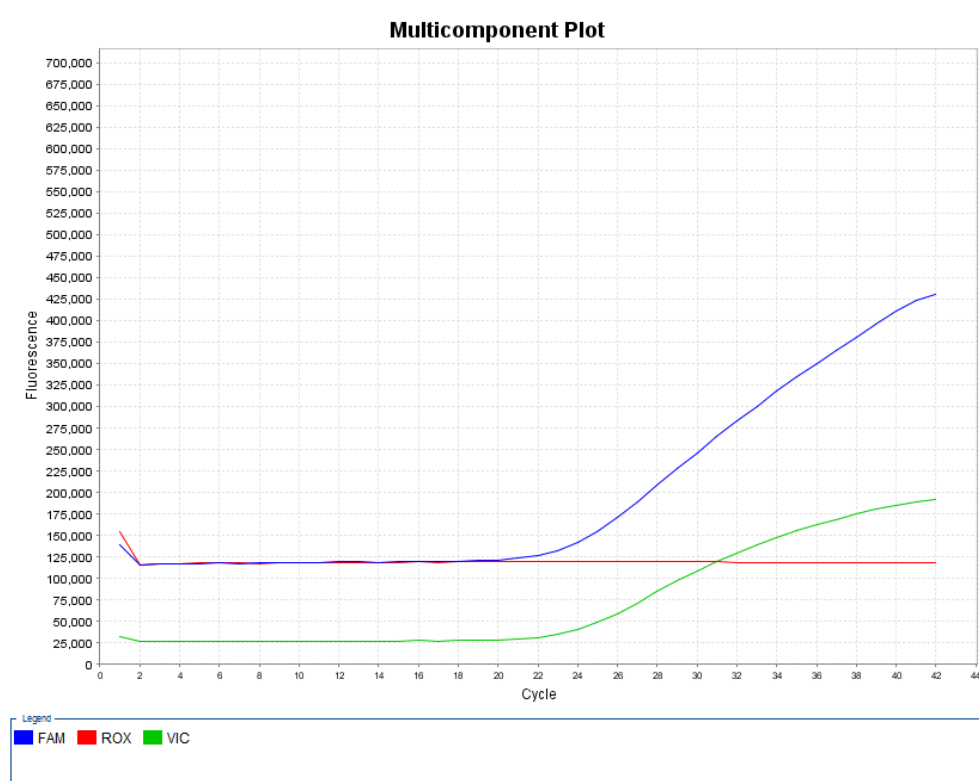


Figura 4: Curva de amplificação do PCR em tempo real evidenciando o genótipo AG

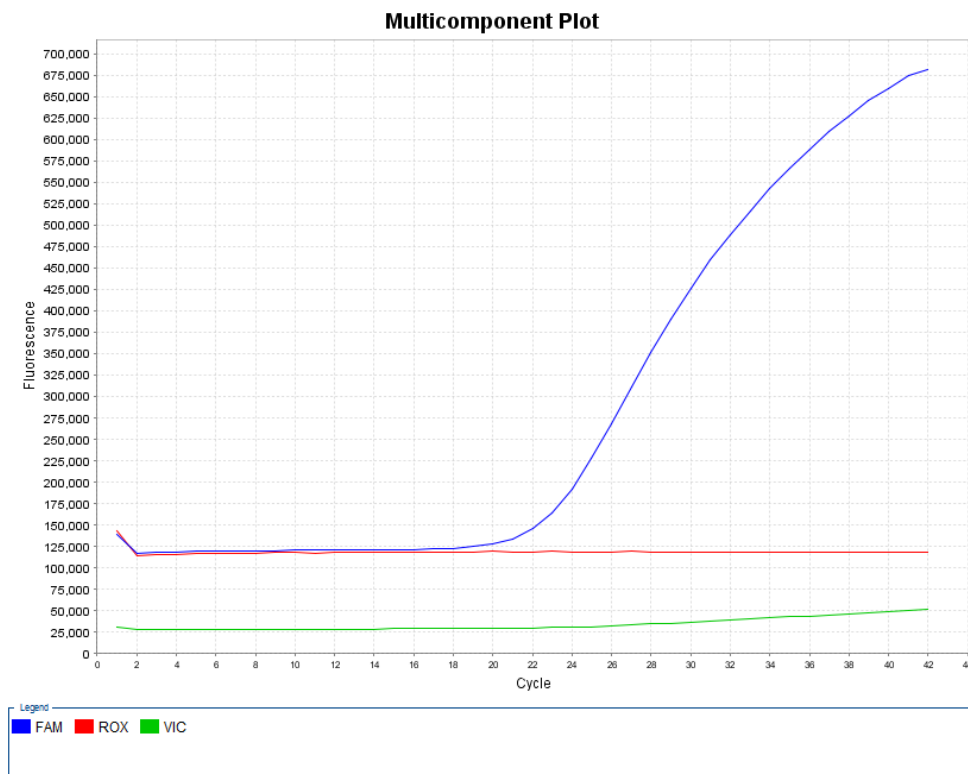


Figura 5: Curva de amplificação do PCR em tempo real evidenciando o genótipo GG

3.6. Análise estatística

A análise estatística foi realizada através de linguagem de programação Python e algoritmos de *machine learning* para realizar a modelagem estatística. Realizado análise preditiva e o modelo de previsão final pelo Random Forest Classifier. Através deste classificador verificado as relações entre a variável resposta, AG, e as variáveis explicativas ou independentes, AA e GG, e obtido a previsão de possibilidade de desenvolver câncer ou não, assim como o prognóstico destes pacientes com nucleotídeo AG, por meio de uma análise de progressão. O modelo de previsão final foi feito por *random forest classifier*. Foram utilizadas as bibliotecas *scikit-learn*, *pandas* e *numpy*.

4. RESULTADOS

Participaram do estudo 187 indivíduos, sendo 112 do grupo com neoplasia e 72 do grupo de risco. O estudo ocorreu em meio a pandemia de covid-19, na qual muitos pacientes foram remarcados, ou atendidos por telemedicina, assim não foi atingido o número de participantes estimado. Dessa forma foi optado por uma coleta 2 pacientes com câncer para 1 paciente de risco sem câncer para aumentar o poder estatístico. (Tabela 3)

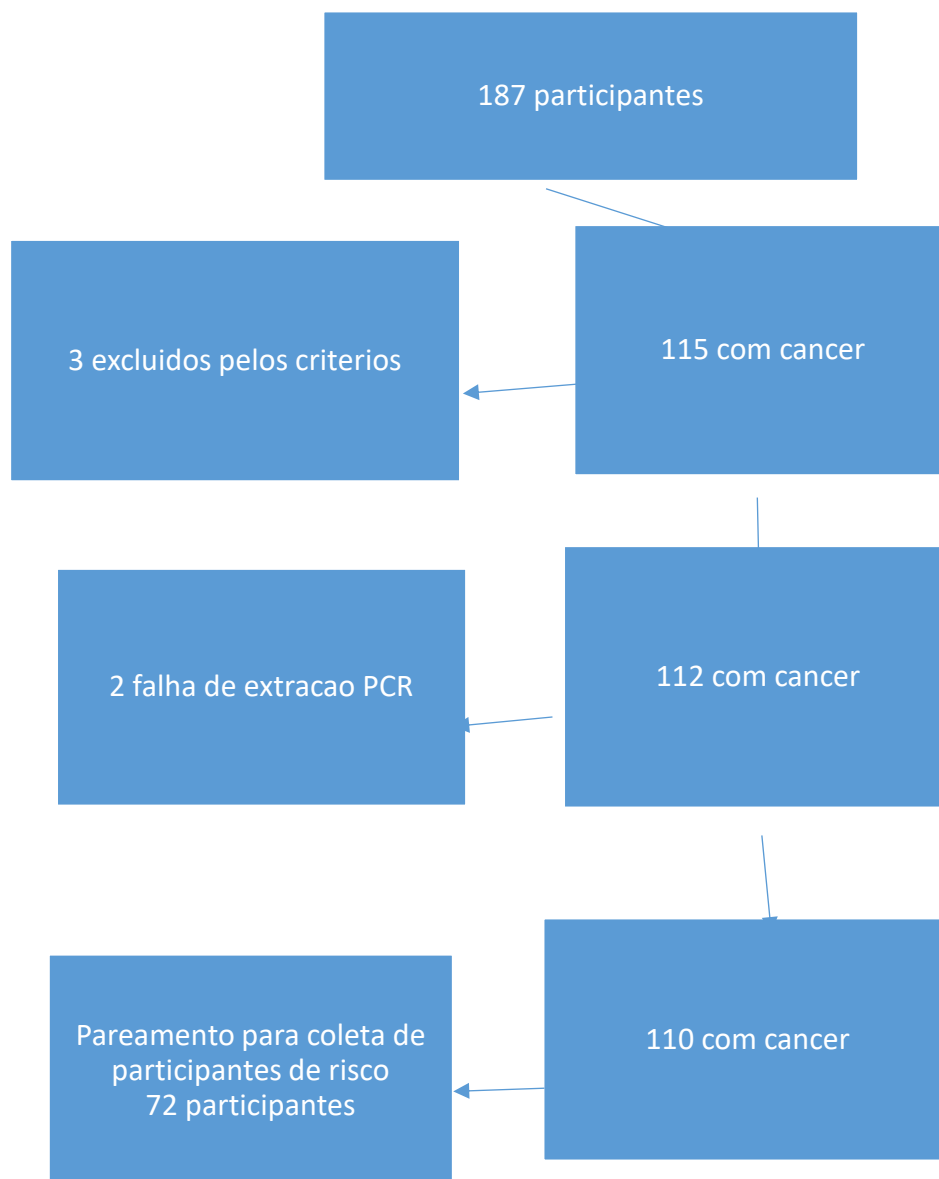


Tabela 6: Fluxograma do processo de seleção dos pacientes e resultados de participantes finais

A média de idade do grupo caso foi de 58 anos e do grupo controle de 59 anos. Os pacientes do grupo caso 6% (7 pacientes) apresentam-se com estadio clinico I, 31% (35 pacientes) em estadio clinico II, 38% (43 pacientes) no estadio clinico III e 23% (26 pacientes) em estadio clinico IV. Conforme demonstrado na tabela 7.

	Casos (112) (%)	Controles (72) (%)
Gênero		
Masculino	54 (48%)	33 (45%)
Feminino	58 (52%)	39 (55%)
Idade	33-83 anos	31-87 anos
Etnia		
Branco	70 (62%)	44 (61%)
Pardo	25 (22%)	19 (26%)
Negro	8 (7%)	7 (9%)
Amarelo	5 (4%)	2 (2%)
Sem informação	4 (3%)	0
Localização		
Cólon	85 (75%)	-
Reto	27 (25%)	-
Estadiamento		
I	7 (6%)	-
II	35 (31%)	-
III	43 (38%)	-
IV	26 (23%)	-

Tabela 7: Características clínicas dos participantes

A distribuição dos genótipos dos pacientes com câncer e sem câncer estão demonstradas na tabela 8. Dois participantes do grupo dos portadores de neoplasia não obtiveram sucesso na extração de DNA e foram excluídos do estudo.

Genótipo	Casos (%)	Controles (%)
AA	25 (22%)	20 (27%)
AG	55 (50%)	27 (36%)
GG	30 (28%)	27 (36%)

Tabela 8: Frequência dos genótipos dos polimorfismos do EGF+61 A/G nos pacientes com câncer colorretal e nos pacientes saudáveis

Os dados obtidos são qualitativos ou categóricos nominais e o nucleotídeo AG representa a variável resposta. O objetivo classificar os pacientes AG quanto a probabilidade de desenvolver câncer. Através da modelagem estatística foi construído relações entre a variável resposta, AG, e as variáveis explicativas ou independentes, AA e GG para obter as previsões quanto à possibilidade de ter câncer ou não, como também o prognóstico destes pacientes, através de uma análise preditiva quanto a sua progressão.

A característica de interesse é o PCR, que possui três categorias: AA, AG e GG. O modelo aprendeu de forma supervisionada a fazer previsões baseada nestas categorias, para isso foram criadas a partir desta característica o seu *one-hot-encoding*, variáveis *dummy*. As variáveis *dummy* foram criadas a partir desta linha de código (figura 6)

```
[37]: df2= pd.get_dummies (data = df1, columns = ['pcr'] )
```

Figura 6: Código para criação das variáveis *dummy*

Este algoritmo é utilizado em variáveis acima de duas categorias. A variável *dummy* assume apenas o valor 0 ou 1 para indicar a presença ou ausência das condições específicas, estas categorias foram classificadas de forma mutuamente excludentes. Assim foram criadas três colunas: pcr_AA, pcr_AG e pcr_GG, conforme figura 7.

```
[61]: display(df2)
```

	pcr_AA	pcr_AG	pcr_GG
0	0	1	0
1	0	1	0
2	0	1	0
3	0	0	1
4	1	0	0
...	...	...	...

Figura 7: Colunas PCR

Com o *encoder* já criado, em machine learning não conseguimos trabalhar nas linhas de códigos com dados qualitativos, mas sim numéricos, a partir deste ponto foi especificado qual o nosso *target*. Em *y* armazenamos o *pcr_AG*, sendo esta a nossa variável resposta/alvo. Em *x* armazenamos a coluna *pcr_GG* e *pcr_AA*, nossa variável preditora. Assim, com a variável preditora e a variável alvo já criadas, dividimos os dados em dois grupos: o grupo de treino e o grupo de teste, conforme figura 8

```
[40]: x_treino, x_teste, y_treino, y_teste = train_test_split(x, y, test_size = 0.25, random_state = 1)
```

Figura 8: Linha de código para grupo de treino e grupo de teste

É importante primeiramente treinar o modelo para melhorar as condições do aprendizado e assim ir ajustar a melhor acurácia. Desta forma, o modelo foi treinado nas variáveis *x_treino* e *y_treino* utilizando 75% dos dados.

Após o treinamento concluído sem presença ou a menor quantidade neste modelo de falsos positivos e negativos e uma boa acurácia, foi aplicado o aprendizado obtidos nos dados de *x_treino* e *y_treino* no *x_teste* e *y_teste*.

Assim o modelo de treino aprendeu de tal forma com os dados, que foi possível observar, conforme figura 9, que a acurácia foi de 100% e não foi identificado a presença de falso positivo ou falso negativo, assim o algoritmo de *random forest classifier* aprendeu de forma satisfatória.

```
[47]: rf = RandomForestClassifier(n_estimators = 1000, random_state = 42)
      rf.fit(x_treino, y_treino)
```

```
[47]: RandomForestClassifier(n_estimators=1000, random_state=42)
```

Figura 9: Modelo de treino com acurácia de 100%

Com o modelo já treinado, foi utilizado os 25% dos dados restantes na variável `x_teste` e `y_teste` e feito a inferência probabilística classificatória de predição em relação ao nosso target, `pcr_AG`, conforme figura 10

```
[48]: resultado_rf = rf.predict(x_teste)
      print(pd.crosstab(y_teste, resultado_rf, rownames = ['Real'], colnames = ['Predito'], margins = True))
```

Predito	0	1	All
Real			
0	27	0	27
1	0	20	20
All	27	20	47

Figura 10: Código para predição do *target* com os 25% dos dados restantes

O resultado de todo o aprendizado e a probabilidade esperada destes pacientes que possuem a sequência de nucleotídeos AG de ter ou não câncer foi visualizada através da matriz de confusão da figura 11:

```

model = RandomForestClassifier(n_estimators = 1000, random_state = 42)
model.fit(x_treino, y_treino)

# Estimativa da acurácia no conjunto de teste
predictions = model.predict(x_teste)
print('Accuracy score =', accuracy_score(y_teste, predictions))

# Matriz de confusão
cm = confusion_matrix(y_teste, predictions)
labels = ['sem câncer', 'com câncer']
cmd = ConfusionMatrixDisplay(cm, display_labels=labels)
cmd.plot(values_format='d')
plt.show()
print(classification_report(y_teste, predictions, target_names=labels))

```

Figura 11: Matriz de confusão

É possível observar na figura 12 que probabilisticamente 27 de 47 participantes poderão não apresentar a câncer colorretal e 20 de 47 poderão desenvolver a condição, ou seja 57% e 42% respectivamente. Atraves deste método, qualquer dado adicionado a este modelo já treinado, será classificado adequadamente.

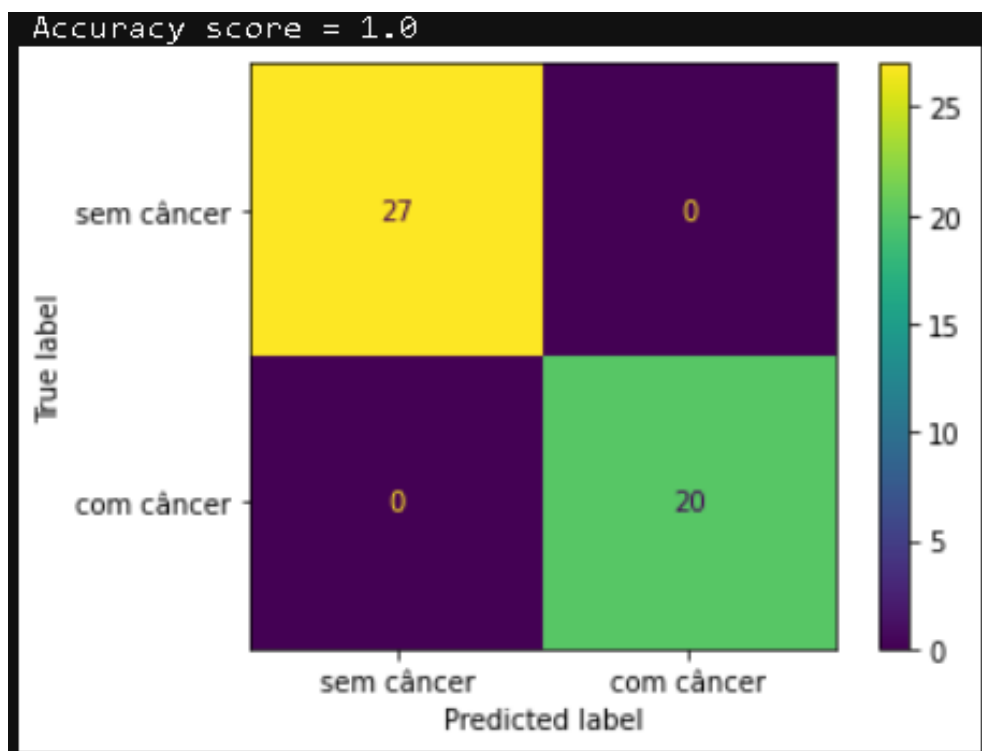


Figura 12: Resultado probabilidade participantes portadores do genótipo A/G de desenvolver câncer

O modelo já classificado quanto a possibilidade de ter câncer ou não, foram categorizados esses pacientes com câncer quanto ao seu prognóstico, ou seja, utilizado o mesmo *dataframe*, adicionando as colunas 'estadiamento' e 'progressão' contidas no mesmo *dataset* original e que fizeram parte da predição estatística.

Os dados de treino contaram com 75% dos dados contidos no *dataset* e 25% para o teste final. O *target* para o modelo foi o nucleotídeo AG e o modelo nos deu o output quanto a probabilidade destes pacientes com câncer desenvolverem uma progressão ou não. O estadiamento no conjunto de dados tem as informações: I, II, III e IV e caso ocorra progressão. O modelo treinado atingiu uma acurácia de 100%, não foi observado a presença de falsos resultados o que levou a uma precisão de 100%.

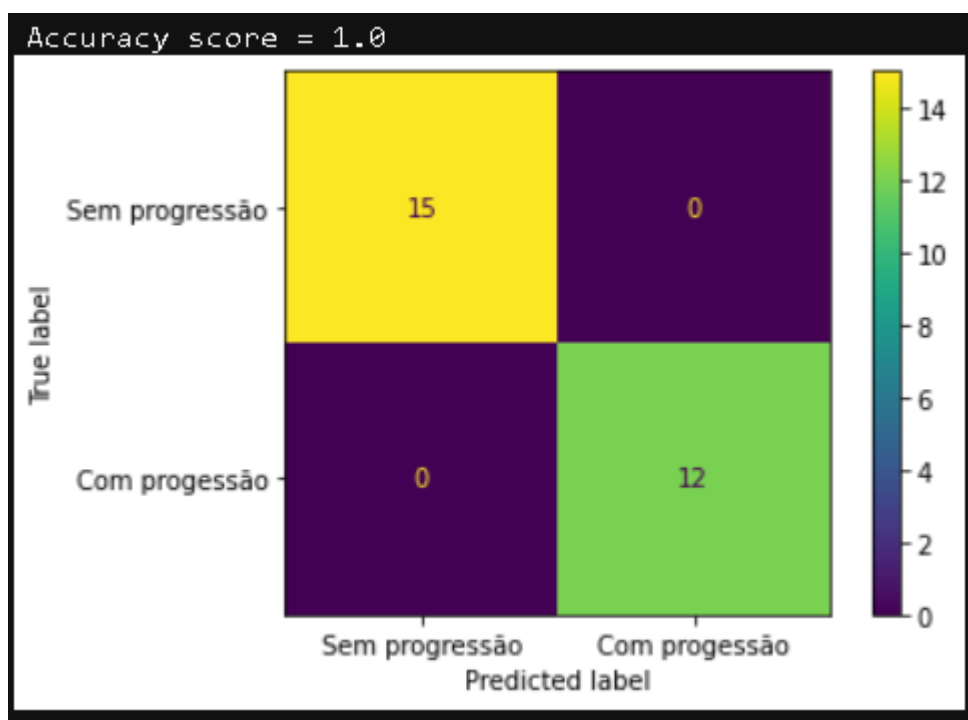


Figura 13: Possibilidade dos pacientes com câncer colorretal portadores do genótipo AG de apresentarem progressão de doença durante o acompanhamento

A matriz de confusão acima nos aponta que a previsão para o modelo, após o treinamento com os dados, a partir dos dados históricos, 55% dos pacientes não terão progressão e 44% poderão ter progressão da doença, conforme figura 13.

5. DISCUSSÃO

O fator de crescimento epitelial tem papel importante em células normais para sobrevivência e proliferação celular, mas também em células tumorais agindo na angiogenese, metastatizacão e inibição de apoptose ⁽⁷²⁾. Este aparentemente simples polimorfismo na posição +61A/G do gene EGF, já foi estudado na susceptibilidade a tumores porem, como evidenciamos houveram diferentes resultados nos tumores estudados, também devemos avaliar a etnia da população estudada e tamanho da amostra.

Cheung et al associou o polimorfismo com aumento de chance de desenvolvimento de doença do refluxo esofágico e neoplasia de esôfago na população canadense ⁽⁷³⁾. Também buscando correlacionar fatores de risco, Browmick evidenciou aumento dos níveis de EGFR nos pacientes portadores de glioblastoma com esta mutação ⁽⁷⁴⁾. Nosso estudo evidencia que 27 de 47 participantes poderão não apresentar a câncer colorretal e 20 de 47 poderão desenvolver a condição, ou seja 57% e 42% respectivamente.

De acordo com Wu et al⁽⁷⁵⁾, o genótipo GG está associado a maior susceptibilidade ao desenvolvimento de câncer colorretal em caucasianos, resultado também encontrado por Lin et al ⁽¹⁰⁾ e Zhu et al ⁽⁷⁶⁾ ambos na população chinesa. Por outro lado, Lau et al⁽⁷⁷⁾ não encontrou a mesma relação na população da Malásia. Observa-se que grande parte dos estudos realizados foram em população oriental, com características diferentes da população brasileira.

A meta-análise realizada em 2012 por Li et al⁽⁷⁸⁾ avaliou 41 estudos caso-controle e evidenciou que as variantes dos genótipos +61 possuem maior risco de desenvolvimento de câncer. Piao et al⁽⁴⁸⁾ em sua meta-análise também evidenciou a correlação entre o polimorfismo e o risco de câncer gastrointestinal, principalmente colorretal e esofágico mas não em tumores gástricos. Estudos estes que corroboram nossos resultados.

Li et al também evidenciou relação do alelo +61G com maior risco de câncer de colorretal, esôfago, estomago e hepatocelular, estes dados foram

mais significativos em pacientes de etnia mista e europeia do que asiáticos ⁽⁴⁷⁾. Dado o pequeno número de pacientes de etnia asiática não obtivemos relevância estatística, entretanto a população brasileira trata-se de uma etnia predominantemente mista.

O método de análise *machine learning* ainda é pouco utilizado na área medica, Kennion faz uma análise sistemática dos estudos que utilizaram o método e identifica que dentre as razões do pouco entusiasmo em sua implementação é o não entendimento de como funciona este método de análise, assim como a dificuldade em identificar o melhor algoritmo⁽⁷⁹⁾.

Este método de análise, que envolve estatística e computação tem apresentado grande efeito em medicina preventiva, como no nosso caso, em que, através da coleta de biopsia líquida, podemos prever risco de desenvolvimento e progressão de neoplasia colorretal. Importante ressaltar que este método não substitui a investigação e o diagnóstico, mas colabora para identificar casos de alto risco, para que estes possam ser diagnosticados e tratados precocemente ⁽⁸⁰⁾.

Este estudo apresenta como limitações o tamanho da amostra, além disso fatores dietéticos e ambientais podem diferir dos estudos prévios em populações asiáticas e o fato de nenhum destes estudos ter utilizado o método de *machine learning* como avaliação. Apesar disso foi possível evidenciar relação entre a presença deste polimorfismo e o risco de desenvolvimento e progressão do câncer colorretal.

6. CONCLUSÃO

Nosso estudo revelou correlação dos pacientes portadores do polimorfismo EGF +61 A/G com a possibilidade de 42% de desenvolvimento de câncer colorretal. Os pacientes portadores do polimorfismo e da neoplasia em estudo possuem 44% de chance de progressão de doença. Confirmamos a aplicabilidade do método de *machine learning* na avaliação de risco prognóstico destes pacientes.

7.REFERENCIAS

1. MS / INCA / Coordenação de Prevenção e Vigilância / Divisão de Vigilância e Análise de Situação, 2021.
2. Jemal, A. et al. Global cancer statistics. CA: A Cancer Journal for Clinicians, v. 61, n. 2, p. 69–90, 2011.
3. Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2020 : incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva. – Rio de Janeiro : INCA, 2019.
4. Habr-Gama, A. Câncer colorretal: a importância de sua prevenção. Arquivos de Gastroenterologia, v. 42, n. 1, p. 2–3, mar. 2005.
5. Mendes, JDV.; Cecilio, MAM. Tendências Regionais da Mortalidade por Câncer no Estado de São Paulo 2000 a 2010. BEPA. Boletim Epidemiológico Paulista. v. 9, n. 104, p. 24–45, 2012.
6. Facina, T. Incidência de Câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Revista Brasileira de Oncologia, 2014.
7. Habr-Gama A. Câncer colorretal – A importância de sua prevenção. Arq Gastroenterol, v. 42 – no.1 – jan./mar. 2005.
8. Zhu Y, Chen Z, Jiang H, Lu B. The genetic association between EGF A61G polymorphism (rs4444903) and risk of colorectal cancer: An update meta-analysis and trial sequential analysis. Medicine (Baltimore). 2019 Jan;98(2):e14007.
9. Zandonai AP, Sonobe HM, Sawada NO. Os fatores de risco alimentares para câncer colorretal relacionado ao consumo de carnes. Ver Esc Enferm USP, 2012; 46(1): 234-9.
10. Lin, L. et al. Association of epidermal growth factor +61 A/G polymorphism in Chinese patients with colon cancer. Genetic testing and molecular biomarkers, v. 16, n. 9, p. 1142–5, set. 2012.
11. Balmain, A., Gray, J.; Ponder, B. The genetics and genomics of cancer. *Nat Genet* 33 (Suppl 3), 238–244 (2003).
12. Maser, R.S.; De Pinho, R.A. Connecting chromosomes, crisis, and cancer. *Science* 297, 565–569 (2002).

13. Kolodner, R.D., Putnam, C.D.; Myung, K. Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* 297, 552–557 (2002).
14. Rouse, J. & Jackson, S.P. Interfaces between the detection, signaling, and repair of DNA damage. *Science* 297, 547–551 (2002).
15. Brown, K., Strathdee, D., Bryson, S., Lambie, W.; Balmain, A. The malignant capacity of skin tumours induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr. Biol.* 8, 516–524 (1998).
16. Strum, WB. Colorectal Adenomas. *New England Journal of Medicine*, 2016; v. 374, p. 1065 – 1075.
17. Markowitz, SD; Bertagnolli, MM. Molecular Basis of Colorectal Cancer. *New England Journal of Medicine*, 2009; v. 361, p. 2441 – 2460.
18. De Jonge V, et al. Quality evaluation of colonoscopy reporting and colonoscopy performance in daily clinical practice. *Gastrointest Endosc.* 2012; 75:98–106.
19. Valori R, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition--Quality assurance in endoscopy in colorectal cancer screening and diagnosis. *Endoscopy.* 2012; 44(Suppl 3):SE88–105. Result of extensive, international consensus on quality measures for colorectal cancer screening.
20. Kuipers EJ et al. Colorectal cancer. *Nat Rev Dis Primers.* 2015 Nov 5;1:15065.
21. Capelhuchnick P, et al. Sinais e sintomas do câncer colorretal e diagnóstico precoce. *Ver. Bras. Colo-Proct*, 1991; 11(4):125-127.
22. Amin MB, et al. *AJCC Cancer Staging Manual* (8th edition). Springer International Publishing: American Joint Commission on Cancer; 2017
23. Sporn, MB, Roberts, AB. *Ann Intern Med.* 117, 408-414 (1992).
24. Normanno N, Bianco C, Strizzi L, Mancino M, Maiello MR, De Luca A, Caponigro F, Salomon DS. The ErbB receptors and their ligands in cancer: an overview. *Curr Drug Targets.* 2005 May;6(3):243-57
25. Bublil, EM.; Yarden Y. The EGF receptor family: spearheading a merger of signaling and therapeutics. *Current opinion in cell biology*, v. 19, n. 2, p. 124–134, 2007.

26. Krasinskas AM. EGFR Signaling in Colorectal Carcinoma. *Patholog Res Int*. 2011 Feb 14;2011:932932
27. Cohen S. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the newborn animal. *J Biol Chem*, v. 237, p.1555-1562, 1962
28. Carpenter G, Lembach KJ, Morrison MM, Cohen S. Characterization of the binding of 125-I-labeled epidermal growth factor to human fibroblasts. *J Biol Chem*, v.250, p.4297-4304, 1975.
29. Endres NF, Engel K, Das R, Kovacs E, Kuriyan J. Regulation of the catalytic activity of the EGF receptor. *Curr Opin Struct Biol*, v.21, p.777-784, 2011.
30. Riese DJ et al. The Epidermal Growth Factor Receptor Couples Transforming Growth Factor- α , Heparin-binding Epidermal Growth Factor-like Factor, and Amphiregulin to Neu, ErbB-3, and ErbB-4. *J Biol Chem*, v.271, p.20047-20052, 1996.
31. Yarden Y, Schlessinger J. Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. *Biochemistry*. 26 (5): 1443–51, 1987.
32. Maruyama IN. Mechanisms of activation of receptor tyrosine kinases: monomers or dimers. *Cells*. 3 (2): 304–30, 2014.
33. Abulrob, Abedelnasser; Lu, Zhengfang; Baumann, Ewa; Vobornik, Dusan; Taylor, Rod; Stanimirovic, Danica; Johnston, Linda J. Nanoscale Imaging of Epidermal Growth Factor Receptor Clustering: EFFECTS OF INHIBITORS. *Journal of Biological Chemistry*. 285 (5): 3145–3156, 2010.
34. Downward J, Parker P, Waterfield MD. Autophosphorylation sites on the epidermal growth factor receptor. *Nature*. 311 (5985): 483–5, 1984.
35. Jean-Yves Douillard, M.D., Ph.D., Kelly S. Oliner, Ph.D., Salvatore Siena, M.D., et al. Panitumumab–FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer. *N Engl J Med* 2013; 369:1023-34.
36. Araujo, A. P. et al. Association between EGF+ 61A/G polymorphism and gastric cancer in Caucasians. 2011.
37. Yarden Y, Sliwkowski MX. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol*. 2001 Feb;2(2):127-37

38. Lockhart C, Berlin J. D. The epidermal growth factor receptor as a target for colorectal cancer therapy. *Semin Oncol*, 2005 Feb 32: 52-60.
39. Castro-Carpeño, J.; et al. EGFR and colon cancer: a clinical view. *Clin Transl Oncol* (2008) 10:6-13
40. Baselga J. Why the epidermal growth factor receptor? The rationale for cancer therapy. *Oncologist*. 2002;7 Suppl 4:2-8
41. Cutsem, EV, et al. Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. *N Engl J Med* 2009;360:1408-17.
42. Bokemeyer C, et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J Clin Oncol*. 2009 Feb 10;27(5):663-71.
43. Heinemann, V, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *Lancet Oncol* 2014; 15: 1065–75.
44. Kim TW, et al. A phase 3 trial evaluating panitumumab plus best supportive care vs best supportive care in chemorefractory wild-type KRAS or RAS metastatic colorectal cancer. *Br J Cancer*. 2016 Nov 8;115(10):1206-1214.
45. Brazilian Medical Association et al. Metastatic colorectal cancer: treatment with panitumumab. *Revista da Associação Médica Brasileira*. 2018, v. 64, n. 7
46. Cunningham D, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med*. 2004 Jul 22;351(4):337-45.
47. Van Cutsem E, et al. Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J Clin Oncol*. 2007 May 1;25(13):1658-64.
48. Morton CC; Byers MG; Nakai H et al. Human genes for insulin-like growth factors I and II and epidermal growth factor are located on 12q22–q24.1, 11p15, and 4q25–q27, respectively. *Cytogenet Cell Genet* 41:245-249.

49. Cohen S. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. *J Biol Chem.* 1962 May;237:1555-62.
50. Shahbazi M, et al. Association between functional polymorphism in EGF gene and malignant melanoma. *Lancet.* 2002 Feb 2;359(9304):397-401.
51. HAN, C. et al. EGFR mutations, gene amplification, and protein expression and KRAS mutations in primary and metastatic tumors of nonsmall cell lung cancers and their clinical implications: a meta-analysis. *Cancer investigation*, v. 29, n. 9, p. 626–634, nov. 2011.
52. HAN, W.; LO, H.-W. Landscape of EGFR signaling network in human cancers: biology and therapeutic response in relation to receptor subcellular locations. *Cancer letters*, v. 318, n. 2, p. 124–134, maio 2012
53. Su, N.-W. et al. EGF and EGFR genetic polymorphisms predict prognosis in locally advanced pharyngolaryngeal squamous cell carcinoma patients receiving postoperative concurrent chemoradiotherapy. *OncoTargets and therapy*, v. 7, p. 2197–204, jan. 2014.
54. De Mello, R. A. et al. Association between EGF +61 genetic polymorphisms and non-small cell lung cancer increased risk in a Portuguese population: a case-control study. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, v. 33, n. 5, p. 1341–8, out. 2012.
55. Araújo, A P et al. Epidermal Growth Factor Genetic Variation Associated With Advanced Cervical Cancer in Younger Women. *Am J of Clin Oncol*: Jun 2012 - 35(3): 247-250.
56. Jiang, G. et al. Association between epidermal growth factor gene +61A/G polymorphism and the risk of hepatocellular carcinoma: a meta-analysis based on 16 studies. *BMC cancer*, v. 15, p. 314, jan. 2015.
57. XU, X. et al. A functional +61G/A polymorphism in epidermal growth factor is associated with glioma risk among Asians. *PloS one*, v. 7, n. 7, p. e41470, jan. 2012.
58. Wang S, Zhao Y, Ruan Z, Chen H, Fan W, Chen J, Wu Q, Qian J, Zhang T, Huang Y, Lu D. Association between EGF +61 G/A and glioma risk in a Chinese population. *BMC Cancer.* 2010 May 21;10:221.

59. Laus AC et al. EGF+61 A>G polymorphism is not associated with lung cancer risk in the Brazilian population. *Mol Biol Rep.* 2019 Apr;46(2):2417-2425.
60. Nakajima, H. et al. Protein expression, gene amplification, and mutational analysis of EGFR in triple-negative breast cancer. *Breast cancer (Tokyo, Japan)*, v. 21, n. 1, p. 66–74, jan. 2014.
61. McLeish JA, Thursfield VJ, Giles GG. Survival from colorectal cancer in Victoria: 10-year follow up of the 1987 management survey. *ANZ J Surg* 2002;72: 352-356
62. PYTHON SOFTWARE FOUNDATION. Python Language Site: Documentation, 2020. Página de documentação.
63. API design for machine learning software: experiences from the scikit-learn project, Buitinck *et al.*, 2013.
64. Deo, RC. Machine Learning in Medicine. *Am. Heart Ass.* 2015; 132, 1920-1930.
65. Bishop, CM. (2006). *Pattern recognition and machine learning (Information science and statistics)*. Springer-Verlag New York, Inc.
66. Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. *Electron Markets* 31, 685–695 (2021).
67. Rajkomar, A; Dean, J; Kohane, J. Machine Learning in Medicine. *N Engl J Med* 2019; 380:1347-1358.
68. Piao, Y. et al. EGF +61A>G polymorphism and gastrointestinal cancer risk: a HuGE review and meta-analysis. *Gene*, v. 519, n. 1, p. 26–33, 25 abr. 2013.
69. Naji, MA, et al. Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis. *Procedia Computer Science* 191 (2021) 487–492.
70. Shravya, Ch; Pravalika, K; Subhani, S. Prediction of Breast Cancer Using Supervised Machine Learning Techniques. *Int Jour Innov Techn and Exploring Engineering.* 2019; 8: 2278-3075.
71. Chiu, P.KF., Shen, X., Wang, G. *et al.* Enhancement of prostate cancer diagnosis by machine learning techniques: an algorithm development and validation study. *Prostate Cancer Prostatic Dis* 25, 672–676 (2022).

72. Spindler KL, Nielsen JN, Ornskov D, Brandslund I, Jakobsen A. Epidermal growth factor (EGF) A61G polymorphism and EGF gene expression in normal colon tissue from patients with colorectal cancer. *Acta Oncol.* 2007;46(8):1113-7.
73. Cheung, W.Y., et al., 2009. Epidermal growth factor A61G gene polymorphism, gastroesophageal reflux disease and esophageal adenocarcinoma risk. *Carcinogenesis* 30, 1363–1367.
74. Bhowmick, DA; et al. A Functional Polymorphism in the EGF Gene Is Found with Increased Frequency in Glioblastoma Multiforme Patients and Is Associated with More Aggressive Disease. *Cancer Res* (2004) 64 (4): 1220–1223.
75. Wu GY, Hasenberg T, Magdeburg R, Bönninghoff R, Sturm JW, Keese M. Association between EGF, TGF-beta1, VEGF gene polymorphism and colorectal cancer. *World J Surg.* 2009 Jan;33(1):124-9.
76. Zhu X, Shen Y, Xie Q. The association between EGF A61G polymorphism and risk of colorectal cancer in a Chinese population: a case-control study. *Biosci Rep.* 2019 May 15;39(5):BSR20190495.
77. LI, T.-F.; REN, K.-W.; LIU, P.-F. Meta-analysis of epidermal growth factor polymorphisms and cancer risk: involving 9,779 cases and 15,932 controls. *DNA and cell biology*, v. 31, n. 4, p. 568–74, abr. 2012.
78. Kennion, O; Maitland, S; Brady, R. Machine learning as a new horizon for colorectal cancer risk prediction? A systematic review. *Health Sciences Review* 4 (2022) 100041
79. Kopad, S; et al. Machine Learning-Based Identification of Colon Cancer Candidate Diagnostics Genes. *Biology* 2022, 11, 365.
80. Jarrett, D. Applications and limitations of machine learning in radiation oncology *Br J Radiol* 2019; 92: 20190001.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Estudo: O PAPEL DO POLIMORFISMO GENÉTICO DO EGF+61 NO RISCO E PROGNÓSTICO NO ADENOCARCINOMA DO CÓLON (GEN-COLON STUDY)
Iniciais do Participante da Pesquisa:
Número de Inclusão:
Nome do Participante da Pesquisa:

Você está sendo convidado a participar deste estudo clínico. É importante que entenda por que o estudo está sendo feito, como a sua informação será usada, o que está envolvido no estudo e os possíveis benefícios, riscos e desconfortos. Use o tempo que precisar para ler cuidadosamente a informação a seguir. Se preferir, você pode conversar com outras pessoas antes de tomar a sua decisão.

Quais são os objetivos deste estudo?

Este estudo clínico está sendo realizado para analisar o papel de polimorfismos genéticos, ou seja, mudanças nos genes, ou em uma parte neles, que pode aumentar a chance de desenvolver câncer em algumas pessoas e principalmente o câncer de intestino (cólon), se você tiver esta mudança em seus genes, o seu câncer pode ser mais ou menos agressivo.

Aproximadamente 100 participantes de pesquisa como você, participarão deste estudo e outros 100 participantes saudáveis.

Qual será minha participação no estudo?

Você está sendo convidado a coletar 05 ml de sangue para analisar o papel de polimorfismos genéticos no adenocarcinoma do cólon, algumas informações serão coletadas do seu prontuário.

A qualquer momento durante a sua participação neste estudo você é livre para decidir interromper, retirar o seu consentimento e finalizar a sua participação neste estudo. Isto não irá afetar o padrão de cuidados que você tem o direito de receber.

O que eu preciso fazer, se eu decidir participar do estudo?

✓ Coletar 05 ml de sangue e permitir a consulta de informações no seu prontuário.

Quais são os possíveis efeitos adversos, riscos ou desconfortos eu poderei apresentar?

Os inconvenientes que você poderá ter é uma dor leve, no local puncionado para a coleta do sangue, na hora da coleta. E outros efeitos menos esperados são dor, tontura, mal-estar passageiro e hematoma (roxo) no local puncionado.

Todas as informações obtidas a seu respeito neste estudo, serão analisadas em conjunto com de outros voluntários, não sendo divulgado a sua identificação ou de outros pacientes em nenhum momento, o pesquisador responsável dará acesso aos resultados de exames ao seu médico ou a você, sempre que solicitado.

Se ocorrer qualquer problema ou dano pessoal durante ou após o procedimento de coleta de sangue o(a) senhor(a), será garantido o direito a tratamento imediato e gratuito na instituição e o(a) Sr(a) terá direito a indenização determinada por lei.

Quais os possíveis benefícios da minha participação?

Este estudo visa a conhecer o papel de polimorfismos genéticos na história natural do câncer, não traz nenhum benefício direto aos participantes. Porém, poderá ajudar futuramente com o mesmo diagnóstico de câncer de cólon.

Quais são os custos da sua participação?

Não há despesas na sua participação em qualquer fase do estudo, bem como não há compensação financeira relacionada a sua participação.

Após ser esclarecido(a) sobre a pesquisa e sua participação como voluntário(a) e, havendo uma confirmação livre espontânea em aceitar a participar como

voluntário(a), você deverá colocar seu nome, assinar, rubricar e datar o documento em duas vias originais.

Em caso de uma lesão relacionada ao estudo ou sempre que você tiver dúvidas sobre o estudo, por favor, entre em contato com:

Dr. Ramon Andrade B de Mello

E-mail para contato: ramondemello@gmail.com

Telefone: 11 989841934 (24 horas)

Endereço: Hospital São Paulo – Rua Napoleão de Barros, 715

Vila Clementino, São Paulo/SP

Ou

Comitê de Ética da UNIFESP

Rua Botucatu, 740 Sala 557 – CEP 04023900 – Vila Clementino, São Paulo/SP

Contato: (11) 5571-1062 ou (11) 5539-7162

E-mail: cep@unifesp.br

Horário de atendimento telefônico e presencial:

Segundas, Terças, Quintas e Sextas, das 09:00 às 13:00hs.

Às Quartas não há atendimento telefônico ou presencial.

Eu, _____ recebi as informações verbais sobre o estudo acima e li as informações por escrito deste documento.

Eu tive a oportunidade de questionar sobre o estudo e esclareci todas as minhas dúvidas sobre minha participação no mesmo. Concordo em participar do estudo e estou ciente que minha participação é totalmente voluntária.

Eu estou ciente que posso retirar meu consentimento a qualquer momento, sem que isto afete o meu tratamento futuro. Sei que meu nome não será divulgado, que não terei despesas e não receberei dinheiro por participar do estudo.

São Paulo, de _____ de 20____.

Nome do participante de pesquisa: Rubrica:.....

Nome do representante legal: Rubrica: 	Nome do pesquisador ou investigador: Rubrica:
Nome da testemunha: Este formulário foi lido para (nome do participante de pesquisa) em/...../..... (data), pelo (nome do membro da equipe da pesquisa), enquanto eu estava presente. Rubrica:	