



EPIDEMIOLOGIA GENÉTICA E PREDIÇÃO DO IMPACTO FUNCIONAL DE VARIANTES DO *CFTR* EM INDÍGENAS

Ana Katarina Campos Nunes

1-INTRODUÇÃO

A farmacogenética pode ser definida como a ciência que se aprofunda no estudo das interações genéticas com a ação dos fármacos no organismo, ou seja, como a variabilidade genética pode influenciar a ação de medicamentos em qualquer tratamento farmacoterapêutico, prevenindo o surgimento de efeitos adversos e tornando as terapias farmacológicas mais eficazes, objetivando reduzir fracassos terapêuticos (SAINZ; SERRA, 2023).

Tendo em vista que o organismo humano pode ser comparado a uma rede metabólica complexa, cada indivíduo apresenta individualidades específicas em seu código genético, e portanto, as respostas a doses terapeuticamente recomendadas pode variar (LAUSCHKE; INGELMAN-SUNDBERG, 2020).

Quanto ao gene *CFTR*, as investigações acerca do fenótipo da Fibrose Cística (FC) visam uma farmacoterapia que possa restaurar a função da proteína de acordo com o mecanismo apresentado. Essa é uma patologia em que há uma disfunção na proteína *CFTR*, resultando em uma doença multisistêmica em que o desequilíbrio iônico provoca o aumento da viscosidade do muco produzido nas células epiteliais e, conseqüentemente, causa o congestionamento de ductos e propicia a proliferação de infecções por microrganismos (ANGYAL et al., 2021).

No estudo acerca da diversidade genética, principalmente em relação a distribuição do gene *CFTR*, nota-se a escassez de informações nos grupos sub-representados, como latino-americanos e povos indígenas e logo, a indisponibilidade de dados genômicos desses indivíduos é um grande obstáculo para investigações de caráter farmacogenômico (COHEN-PAES et al., 2022). O padrão de povoamento e distribuição da ancestralidade é heterogêneo e dispersivo. Quando se tem o estudo de populações indígenas temos a escassez de dados e informações, o que dificulta um painel farmacogenético mais conciso e completo no Brasil.

Dessa forma, este estudo irá abordar a análise de amostras de exomas de indígenas da região norte, no estado do Pará, visando analisar as variantes existentes no cromossomo 7 e como as mesmas estão relacionadas com o gene *CFTR* (Cystic Fibrosis Transmembrane Regulator), objetivando observar como a variabilidade genética dessa população interfere no fenótipo da Fibrose Cística.

2-METODOLOGIA

2.1 OBTENÇÃO DA AMOSTRA

Os dados analisados possuem carácter estruturado e contém o exoma de 110 indígenas do Pará, que foram retiradas de uma amostra já processada de variadas populações indígenas, a qual a extração e processamento de dados foi realizado no Laboratório de Genética Humana e Médica da UFPA (LGHM). Após coleta do sangue periférico dos indivíduos, as bibliotecas de sequenciamento de exoma completo foram preparadas utilizando o kit Nextera Rapid Capture Exoma (Illumina) e SureSelect Human XT all exon V6 (Agilent), de acordo com as orientações do fabricante. As bibliotecas foram sequenciadas na plataforma de sequenciamento NextSeq 550 (Illumina) em 4 execuções do NextSeq 500/550 High Output Kit com aproximadamente 16 amostras cada (SANTOS et al., 2020).

2.2 PRÉ-PROCESSAMENTO DE DADOS

Os dados do exoma foram pré-processados para filtragem de qualidade, alinhamento ao genoma GRCh38 e chamada de variantes com GATK. VCFTools e Plink (v1.9) foram empregados para filtragem de genótipos por coordenadas genômicas (GRCh38, 7: 117287120-117715971), avaliação da qualidade do genótipo (missing data, equilíbrio de Hardy-Weinberg). Realizamos o AlphaMissense para previsão de patogenicidade, bem como a ferramenta de anotação de variantes SNP Nexus para análise de epidemiologia funcional e genética.

Para o controle de qualidade do arquivo contendo as amostras das populações indígenas foi utilizado o VCFtools para filtragem do gene *CFTR* pelas coordenadas

genômicas obtidas no gnomAD (GRCh38, 7:117287120-117715971). Os demais parâmetros avaliados foram 'missing-indv' (utilizado para calcular a taxa de genotipagem faltantes por indivíduo), 'missing-site' (para verificar a taxa de genotipagem por posição genômica) , --singletons (para contagem de singletons e doubletons) e 'het' (para verificação de taxa de heterogosidade e observar se endogamia ou exogamia estão presentes nos dados) (Figura 1).

```
vcftools --vcf IND.vcf --from-bp 117287120 --to-bp 117715971 --chr chr7 --recode --out LOC_cftr-IND.vcf  
vcftools --vcf LOC_cftr_IND.vcf --missing-indv --out missing_ind_cftr  
vcftools --vcf LOC_cftr_IND.vcf --missing-site --out missing_site_cftr  
vcftools --vcf vcf LOC_cftr_IND.vcf --het --out heterogozidade_ind  
vcftools --vcf LOC_cftr_IND.vcf --singletons --out IND_singletons
```

Figura 1 – Etapas de processamento e controle de qualidade utilizando a ferramenta VCFtools.

Análise de Variantes Comuns e Raras

Na análise da quantidade de variantes raras e comuns, realizamos a conversão do arquivo VCF contendo as amostras para o formato Plink (v1.9) e, posteriormente, foi calculada a frequência alélica. As variantes foram filtradas no RStudio utilizando os valores-padrão: $f > 0.01$ (comuns), $f < 0.01$ (raras).

#Conversão do arquivo VCF contendo as Amostras para o formato Plink

```
plink --vcf IND.vcf --make-bed --out IND_plink
```

-> Como resultado, obtivemos três arquivos : IND_plink.fam , IND_plink.bed e IND_plink.bim;

#Cálculo de frequência alélica utilizando o Plink (v1.9)

```
plink --bfile IND_plink --from-bp 117287120 --to-bp 117715971 --freq --make-bed --out freq_plink
```

-> Utilizamos as coordenadas genômicas do CFTR para gerar a frequência exata no arquivo de saída "freq_plink.frq"

-> 1º Etapa : Carregamento do arquivo e remoção de dados ausentes (NA)

```
data_3 <- read.csv("freq_plink.frq", header = TRUE, sep = "")  
data3_semna <- na.omit(data_3)
```

-> 2º Etapa : Construção de um Data Frame com os dados

```
tabela_variantes <- data.frame(data3_semna)
```

-> 3º Etapa : Filtragem das variantes (raras - $F < 0.01$ e comuns $F > 0.01$)

```
var2 <- variantes_maiores_0.01 <- subset(tabela_variantes, tabela_variantes$MAF > 0.01)
```

```
var3 <- variantes_menores_0.01 <- subset(tabela_variantes, tabela_variantes$MAF < 0.01)
```

-> 4º Etapa : Criação de dataframe com as quantidades de variantes raras e comuns

```
data_variantes <- data.frame(  
  Variants = c("Rares", "Common"),  
  Quantity = c(214 , 35)  
)
```

```
data_var2 <- data.frame(  
  Variants_frequency = c("F>0.01", "F<0.01"),  
  Quantity = c(35 , 214)  
)
```

Figura 2 – Etapas de conversão do arquivo VCF para Plink (v1.9) para cálculo da frequência alélica. Após esse procedimento, as variantes foram filtradas utilizando o RStudio.

Taxa de genotipagem por indivíduos

Nas análises das taxas de genotipagem, observamos valores satisfatórios para a filtragem por indivíduos, predominando valores acima de 90% resultando em boas taxas de identificação e reprodutibilidade das variantes analisadas. A maioria dos valores obtidos estão próximos a 98%, como podemos observar nos picos do

histograma e gráfico de densidade. Dessa forma, os altos valores de genotipagem por indivíduo minimizam os dados faltantes.

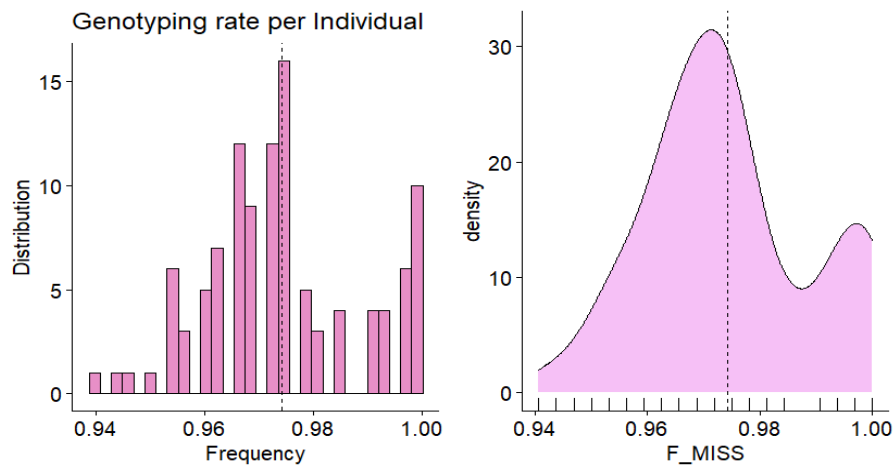


Figura 3 – Taxa de genotipagem por indivíduos.

Taxas de Genotipagem por localizações genômicas

Quanto a esta taxa de genotipagem, também foram obtidos resultados satisfatórios quanto a eficácia na obtenção de genótipos válidos. Neste caso, podemos observar que existem poucos locais com valores abaixo de 80% e o restante das amostras analisadas estão localizadas no pico mais alto corresponde a frequência 1.0 (100%) tanto no histograma quanto no gráfico de densidade.

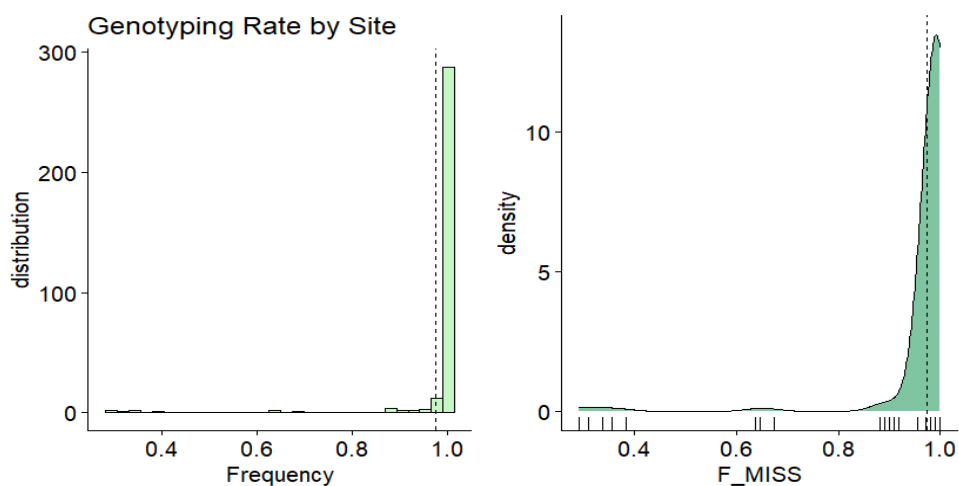


Figura 4 – Taxa de Genotipagem por Localização Genômica.

Análise de Heterozigidade

O coeficiente de endogamia é a probabilidade que dois alelos presentes em um determinado locus sejam idênticos por descendência. Se o coeficiente de endogamia for zero, irá indicar que os indivíduos da amostra não são endogâmicos. Por outro lado, se houver endogamia completa ou máxima ($F = 1$), a frequência de heterozigotos é zero, e a população é composta apenas por homozigotos, concluindo que aquela população é predominantemente endogâmica.

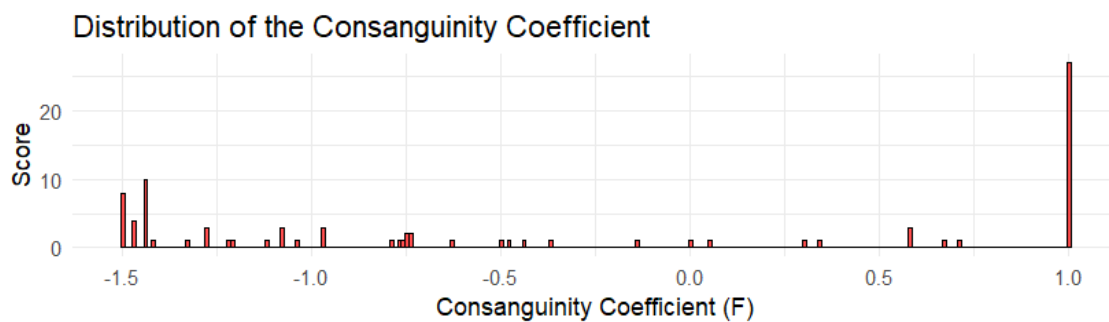


Figura 5 – Histograma de Distribuição de Consaguinidade (F)

Análise de Variantes pelo AlphaMissense

As predições de patogenicidade foram realizadas utilizando o arquivo VCF submetido ao The Ensembl Variant Effect Predictor através do plugin para o AlphaMissense. Este preditor forneceu duas anotações: uma com valores numéricos no intervalo de 0 a 1, interpretados como a probabilidade de uma variante ser patogênica, e outra em três categorias discretas baseadas nos thresholds numéricos da anotação anterior: likely_pathogenic ($> 0,564$), likely_benign ($< 0,34$) e ambiguous (uncertain) para os valores intermediários. Tais thresholds foram definidos para alcançar uma precisão de 90%. A análise foi conduzida considerando o genoma de referência GRCh38 (hg38).

Em seguida, foram baixadas todas as 8601 predições *in silico* do AlphaMissense para o *CFTR* (Uniprot ID P13569) e foi plotado um gráfico de dispersão com os scores de patogenicidade *versus* cada resíduo correspondente aos aminoácidos da estrutura proteica, utilizando o software OriginPro

Análise pela ferramenta de predição SNPnexus

As análises foram realizadas no site da ferramenta (<https://www.snp-nexus.org/v4/>), utilizando como referência o genoma GRCh38. Como formato de entrada, foram utilizadas as posições cromossômicas, o número do cromossomo, os alelos alternativos e de referência da amostra das populações indígenas do Pará. Os filtros selecionados foram Gene/Protein Consequences (Ensembl/RefSeq) para verificar a consequência das possíveis variantes de baseado nos dados genômicos do Ensembl e RefSeq, Population Data (1000 genomes, gnomAD/exomeData,) visando observar a distribuição das variantes nas populações.

4.RESULTADOS PRELIMINARES

Após o tratamento da amostra, 320 variantes foram encontradas, das quais 214 foram identificadas como raras ($F < 0.01$) e 35 variantes comuns ($F > 0.01$), visto que os dados ausentes (NA) foram omitidos.

Na ferramenta de anotação de predição (SNPnexus), o processamento das posições cromossômicas utilizando o filtro Ensembl e filtrando pela isoforma P13569 do *CFTR* (Transcript: ENST00000003084) dos quais resultou em 96 registros para esta isoforma, sendo: 50 intrônicas, 29 codificantes, 15 3'UTR e 2 de em sítio de Splicing.

Com relação a distribuição das variantes encontradas, dos 110 localizações de cada indivíduo processados pela ferramenta, apenas 8 foram encontradas no gnomAD/exome e 10 foram encontradas no 1000 genomes, com as frequências para população africana (AFR), americano misto (AMR), Judeu Ashkenazi (ASJ), leste asiático (EAS), europeu finlandês (FIN), europeu não-

finlandês (NFE), população não designada (OTH) e sul da Ásia (SAS) (Tabela 1) (Tabela2).

Variation ID	dbSNP	REF Allele	ALT Allele	AFR	AMR	ASJ	EAS	FIN	NFE	OTH	SAS
117297583:G/A	rs188541902	G	A	0.007173	0.000206	0.000206	None	None	0.000009	None	None
117595000:C/T:1	rs566908587	C	T	0.000123	0.000087	0.000087	None	None	0.000009	None	0.000098
117603609:C/T:1	rs121909034	C	T	0.000800	0.000261	0.000261	0.000489	0.000000	0.000396	0.001141	0.000555
117606754:G/A:1	rs75096551	G	A	0.001417	0.000058	0.000058	None	None	0.000009	None	None
117642590:A/G:1	rs1800130	A	G	0.149033	0.019531	0.019531	0.002852	0.125289	0.035352	0.042240	0.055875
117642615:G A/G:1	None	A	-	0.000000	0.000000	0.000000	0.000055	0.000000	0.000009	0.000000	0.000000
117652963:G/T:1	rs775555170	G	T	None	None	None	None	None	None	None	0.000098
117664769:G/A:1	rs201686600	G	A	None	None	None	0.000218	None	None	None	None

Tabela 1 – Distribuição de frequência populacionais por variantes no gnomAD/exome Data.

Variation ID	dbSNP	REF Allele	ALT Allele	Minor Allele	AFR	AMR	EAS	EUR	SAS
chr7:117297583:G/A:1	rs188541902	G	A	A	0.006800	None	None	None	None
chr7:117595000:C/T:1	rs566908587	C	T	T	0.002300	None	None	None	None
chr7:117603609:C/T:1	rs121909034	C	T	T	None	0.001400	None	None	0.002000
chr7:117606754:G/A:1	rs75096551	G	A	A	0.000800	None	None	None	None
chr7:117642590:A/G:1	rs1800130	A	G	G	0.177000	0.030300	0.001000	0.039800	0.048100
chr7:117652642:G/A:1	rs214164	G	A	A	0.180800	0.132600	0.012900	0.216700	0.188100
chr7:117664769:G/A:1	rs201686600	G	A	A	None	None	0.001000	None	None
chr7:117666798:G/C:1	rs568948497	G	C	C	None	None	None	0.001000	None
chr7:117668359:C/T:1	rs1042180	C	T	T	0.252600	0.152700	0.015900	0.252500	0.233100

Tabela 2 – Distribuição de frequência populacionais por variantes no 1000 genomas.

Com relação a frequência das variantes, foi possível observar após a filtragem uma grande quantidade de variantes raras ($n = 214$) e um total de 35 variantes comuns, de acordo com os parâmetros estabelecidos para as frequências. (Figura 6).

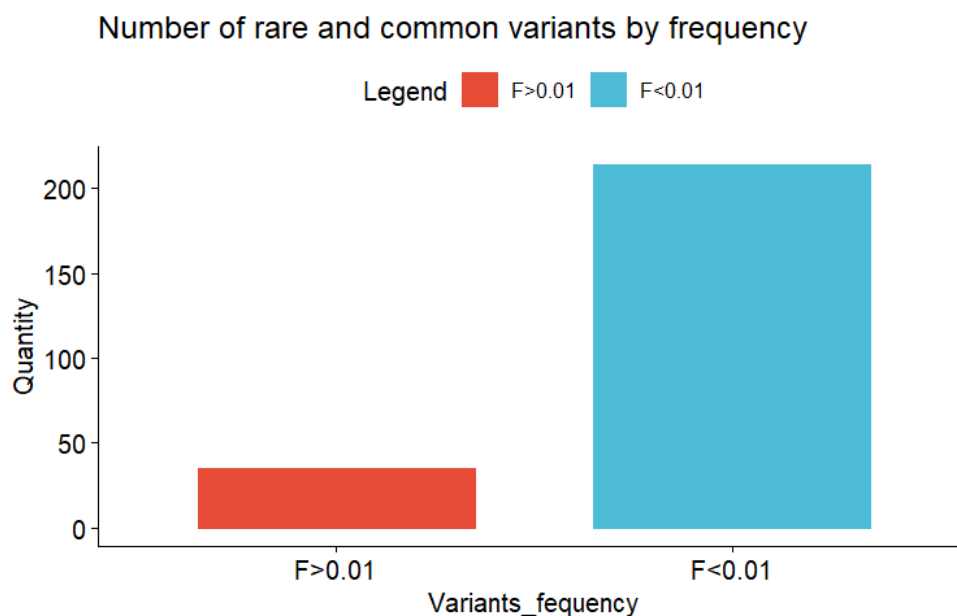


Figura 6 – Contagem de por frequência.

Contagem de Singletons e Doubletons

Nas amostras das populações indígenas, podemos observar que a maioria das variantes presentes são doubletons ($n = 200$), enquanto os singletons presentes estão em menor quantidade ($n = 35$), sendo que quatro variantes missense verificadas no AlphaMissense se encaixam como doubletons, apresentando alto score de patogenicidade, como será abordado em um tópico posterior (Figura 7).

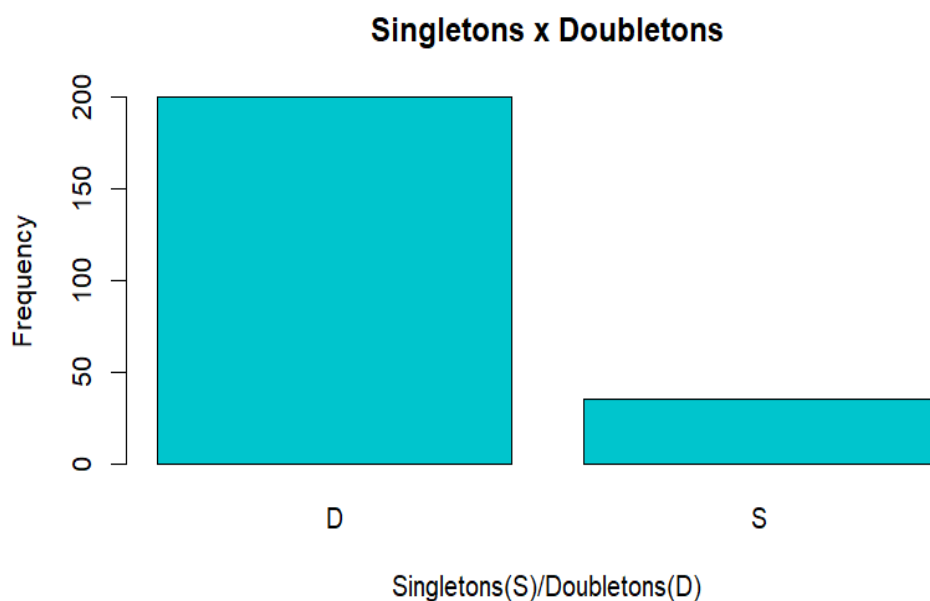


Figura 7 – Contagem de singletons e doubletons nas amostras de populações Indígenas.

As quatro variantes de interesse clínico classificadas como patogênicas, de acordo com escore de patogenicidade do AlphaMissense, foram identificadas como doubletons. Variantes doubletons são observadas exatamente duas vezes, estando presentes em apenas dois indivíduos dentro de uma população.

Na análise realizada pelo AlphaMissense, 4 variantes demonstraram relevância clínica, de acordo com o escore de predição de patogenicidade para a isoforma P13569 do gene *CFTR*. Duas variantes, 11.c.3728T>C e 11.c.2932A>G, apresentaram escores de patogenicidade elevados, sendo 0.87 e 0.89, respectivamente, e as outras duas variantes, 11:c.2588T>A e 11:c.4045G apresentam escores moderado, como 0,67 e 0,61, representados na imagem 1.

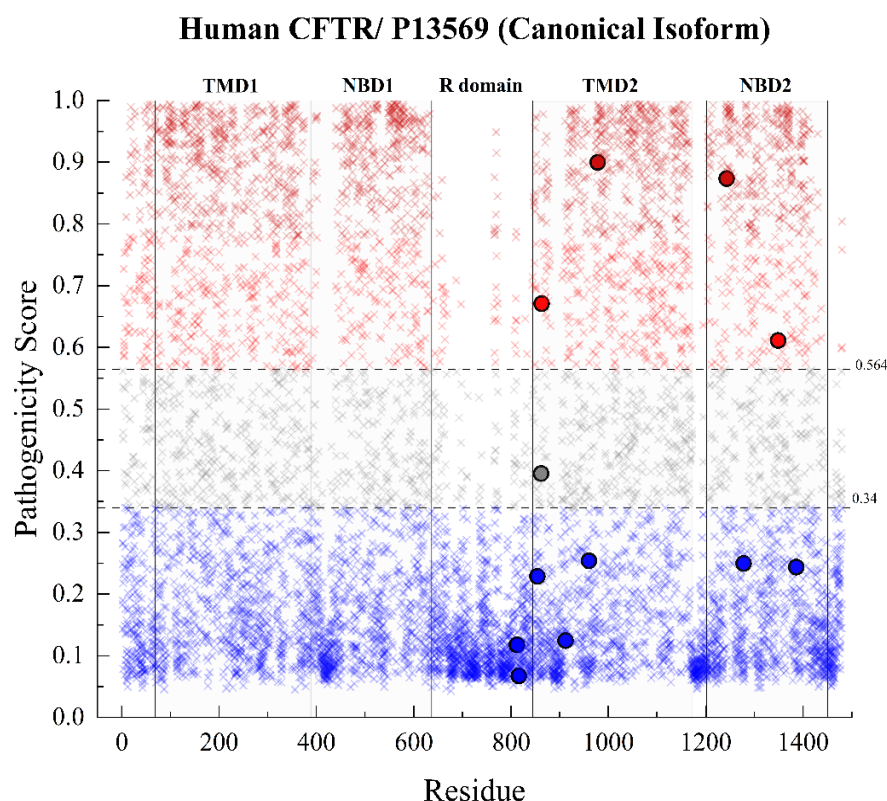


Figura 7 - Alphamissense possíveis previsões missense para a proteína CFTR humana estrutura prevista por AlphaFold – AF-P13569-F1-AM_v4 – (proteína de alta precisão previsão de estrutura com AlphaFold – <https://doi.org/10.1038/s41586-021-03819-2> ; Banco de dados de estrutura de proteínas AlphaFold: expandindo massivamente a estrutura cobertura do espaço de sequência de proteínas com modelos de alta precisão <https://doi.org/10.1093/nar/gkab1061>). Todas essas variantes são denotadas em 3 diferentes cores para expressar sua classificação: azul, cinza e vermelho (vermelho escuro para variante nitidamente patogênico) para benigno, ambíguo e patogênico (mais patogênico), respectivamente. Para o CFTR temos 4.870 (56%) Benignos, 1.100 (12,79%) Ambíguos e 2.631 (30,59%) Patogênico.

Quanto a localização destas mutações na estrutura da proteína CFTR, as variantes 11.c.3728T>C e 11.c.4045G>A estão em NB2 (Nucleotide Binding Domain) e 11.c.2932A>G e 11.c.2588T>A em TMD2 (Domínio Transmembranar 2) (Figura 8) (Tabela 3). A localização destas variantes em ambos os domínios e a patogenicidade da alteração dos aminoácidos podem perturbar o tráfego iônico intracelular e extracelular, influenciando os níveis de muco celular. De acordo com os dados do genoma e exoma do gnomAD, as variantes c.4045G>A e c.3728T>C exibem frequências raras em nível continental ($<0,001$), e duas dessas variantes

nãopossuem frequências alélicas, sugerindo que são específicas para a população indígena do Brasil.

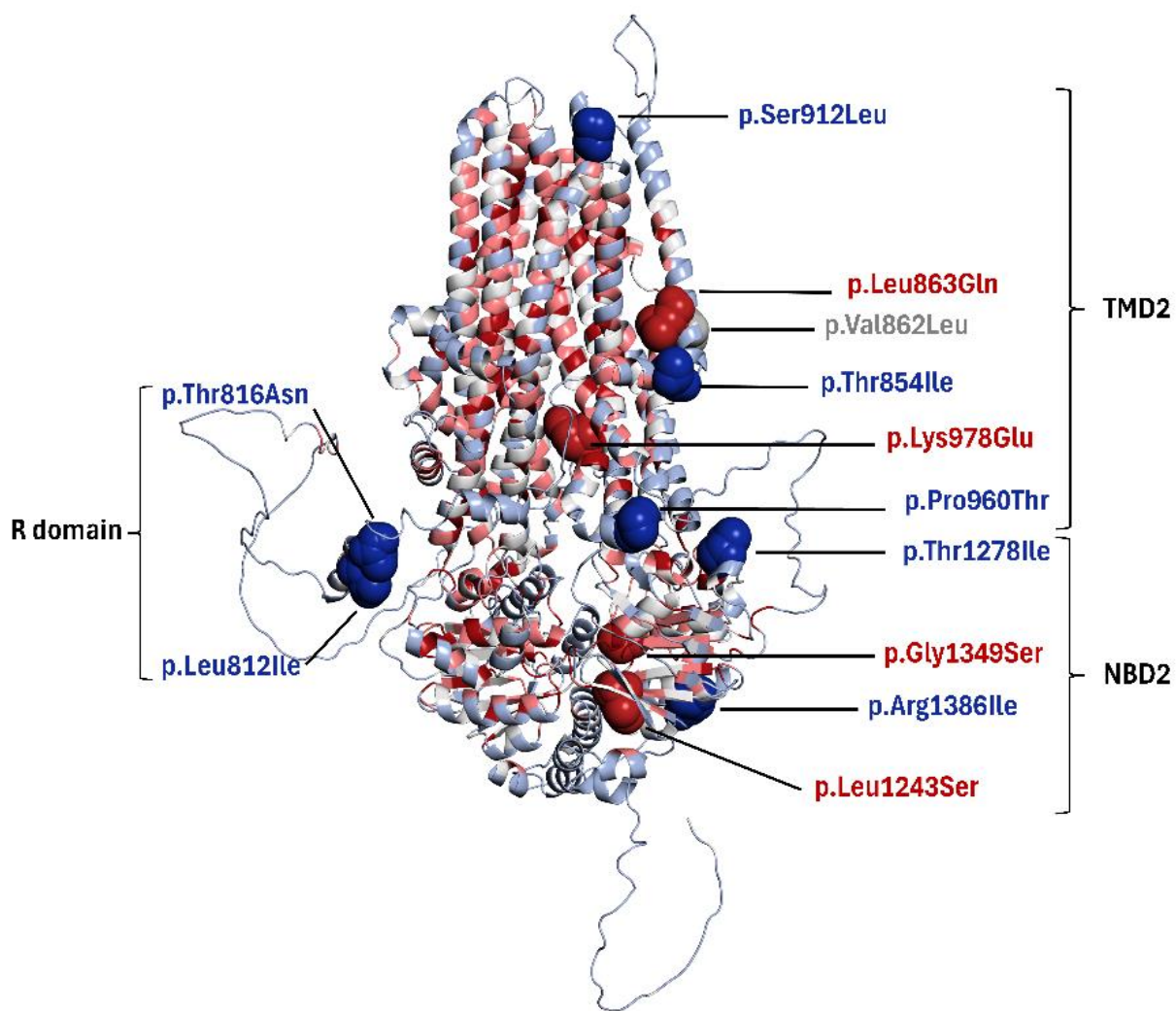


Figura 8 – Estrutura 3D do CFTR Humano .

HGVSp	Mutation	Classification	Pathogenicity Score
p.Thr816Asn	T816N	likely_benign	0.0673
p.Leu812Ile	L812I	likely_benign	0.1178
p.Ser912Leu	S912L	likely_benign	0.1245
p.Thr854Ile	T854I	likely_benign	0.2286
p.Arg1386Ile	R1386I	likely_benign	0.2437
p.Thr1278Ile	T1278I	likely_benign	0.2496
p.Pro960Thr	P960T	likely_benign	0.254
p.Val862Leu	V862L	ambiguous	0.3955
p.Gly1349Ser	G1349S	likely_pathogenic	0.6111
p.Leu863Gln	L863Q	likely_pathogenic	0.6709
p.Leu1243Ser	L1243S	likely_pathogenic	0.8736
p.Lys978Glu	K978E	likely_pathogenic	0.8999

Tabela 3 – SNVS encontradas nos povos indígenas da Amazônia/Pará

5.DISSCUSSÕES

A construção da história dos povos indígenas americanos é extremamente complexa e, no entanto, existe uma grande sub-representatividade com relação aos estudos genéticos nesses grupos. Isso se deve, em grande parte, a negligência e priorização de estudos com populações de origem europeia nos estudos de diversidade genômica (SILVA, 2021).

De acordo com Silva (2021), os poucos estudos com as populações indígenas americanas podem levar ao risco na aplicação de farmacoterapias, visto que as informações genéticas são escassas e pouco documentadas nos painéis farmacogenômicos.

Dessa forma, analisamos durante o controle de qualidade dos dados, o fator de endogamia entre os indivíduos da amostra foi analisado e foi possível

observar que o pico máximo do histograma indica endogamia máxima entre grande parte dos indivíduos analisados na amostra. Silva (2021) ratifica que, apesar do processo de miscigenação durante o período da chegada e estabelecimento dos colonizadores na região, os grupos amazônicos são significativamente menos miscigenados do que os demais grupos, sendo indivíduos predominantemente endogâmicos, diminuindo assim a diversidade genética. Outro fator observado foi que, as 4 variantes apontadas como patogênicas se apresentam como doubletons, sendo que apenas uma (11:c.4045G>A) teve sua frequência encontrada na população do leste asiático (EAS), de acordo com o gnomAD/exome (0.0002177) podendo ser considerada como rara.

Verificamos também a predominância de variantes com localizações intrônicas para a isoforma estudada do *CFTR* (n= 50). Na literatura já é relatado a presença de variantes intrônicas e como essas são responsáveis por causar doenças graves. Ellingford et.al,2022, afirmam que variantes de regiões não codificantes que causam perturbações nos elementos não codificantes causam doenças graves, afetando assim o splicing, transcrição, tradução, processamento e estabilidade de RNA e interações de cromatina (ELLINGFORD et al., 2022).

Outras duas variantes splicing também foram contabilizadas para esta isoforma, sendo que as variantes que podem vir a afetar o splicing, na complexidade desse processo abre uma ampla gama de resultados , tanto em RNA quanto em proteína, dificultando a resposta moduladora de tratamentos já existentes para a Fibrose Cística (JOYNT; CUTTING; SHARMA, 2022). Outra importância observada na investigação dessas variantes trata-se de evitar a classificação das mesmas erroneamente como responsivas ao modulador de drogas de moléculas pequenas (JOYNT; CUTTING; SHARMA, 2022).

Comparando os resultados da ferramenta de predição SNP Nexus com as variantes identificadas como patogênicas pelo AlphaMissense, identificamos que apenas a variante Gly1349Ser (11:c.4045G>A) possui uma frequência de 0.000218 na população do leste asiático, enquanto as demais não possuem frequências em nenhuma dessas populações tanto no filtro do gnomAD/exoma Data quanto no 1000 genomas, sugerindo que possam estar presentes somente na população indígena do norte do Brasil que foi utilizada para este estudo.

Quanto a localização destas variantes na proteína CFTR, 11:c.3728T>C (Leu1243Ser) e 11:c.4045G>A (Gly1349Ser) estão em NB2 (Nucleotide Binding Domain), responsável por auxiliar o NB1 no processo de hidrólise de ATP. Pode-se supor algumas alterações na estrutura e funcionamento da proteína por analisando a estrutura dos aminoácidos e as possíveis interações que podem vir a ocorrer. A Leu1243Ser, o aminoácido leucina possui cadeia alifática, tendo em sua predominância ligações apolares (mais fracas) enquanto a serina é um aminoácido hidroxilado. A troca de leucina por serina pode indicar um aumento na formação de pontes de hidrogênio (ligações mais fortes) a partir da serina, levando a uma possível interação com moléculas próximas ou algum impedimento estrutural, visto que há o aumento de energia já que há a troca de ligações apolares por polares.

Na Gly1349Ser podemos observar uma possível redução energético, visto que o aminoácido glicina além de possuir cadeia alifática, tem um grupo amina (H₂N) em sua estrutura, o que permitirá a formação de mais pontes de hidrogênio, promovendo maior estabilidade a molécula. Já o aminoácido serina, como mencionado, é hidroxilado e portanto, a troca de glicina por serina implica na diminuição de pontes de hidrogênio podendo ocasionar uma possível desestabilização entre a interação desses dois domínios, visto a diminuição da energia e força das ligações intermoleculares.

As variantes 11:c.2588T>A (Leu863Gln) e 11:c.2932A>G (Lys978Glu), localizadas no Domínio Transmembranar 2 (TMD2), próximo a membrana plasmática e ao canal iônico da célula também possuem possíveis interações que afetam diretamente o fluxo de substratos. Na Leu863Gln, a leucina é um aminoácido de cadeia alifática com maior quantidade de ligações apolares (mais fracas), enquanto a glutamina possui dois grupos amina (NH₂) em sua estrutura, favorecendo a formação de pontes de hidrogênio e fazendo com que a molécula fique mais polar. Nesta situação, uma possível consequência seria na interação com a parte apolar da membrana, que fica na parte interna da célula próximos ao TMD2.

Quanto a Lys978Glu, a lisina é um aminoácido de caráter básico e é substituída pelo ácido glutâmico. Esta mutação está localizada na parte interna do canal iônico, especificamente nas alfa-hélices. Dessa forma, há duas formas possíveis de interação do ácido glutâmico: a primeira é pela repulsão eletrostática,

visto que há a passagem do íon cloro (carga negativa) e o encontro com este aminoácido pode promover, possivelmente, a congestão ou algum dano ao canal iônico devido a repulsão contínua; a segunda forma seria pelo fator de desnaturação da proteína pela presença do ácido glutâmico, prejudicando a funcionalidade do canal e a entrada de sódio e água na célula.

6.CONCLUSÃO

Com os dados preliminares obtidos neste projeto, podemos inferir que detalhes mínimos como troca de aminoácidos em domínios-chave da proteína CFTR, por meio de interações intermoleculares e iônicas podem interferir na qualidade do funcionamento do fluxo de substratos e abertura do canal iônico, ratificando a importância da análise dessas variantes na população indígena visto que, apesar de serem doubletons, possuem um alto score de patogenicidade e seus efeitos patogênicos devem ser analisados mais profundamente. Além disso, outro fator importante a observar são as três variantes que não apresentaram frequências nas demais populações globais para confirmar se elas são exclusivas dos povos indígenas da região norte. Dessa forma, de modo geral, este estudo de diversidade genética aumenta nossa compreensão da genética de populações sub-representadas. Também identifica variantes genéticas com potenciais implicações clínicas, especialmente no campo da farmacogenômica para populações indígenas e mestiços latino-americanos.

REFERÊNCIAS BIBLIOGRÁFICAS

ANGYAL, D.; BIJVELDS, M. J.; BRUNO, M. J.; PEPPELENBOSCH, M. P.; JONGE, H. R. de. Bicarbonate transport in cystic fibrosis and pancreatitis. *Cells*, MDPI, v. 11, n. 1, p. 54, 2021. Citado na página 10.

COHEN-PAES, A.; ALCÂNTARA, A. L. de; MENEZES, E. de S.; MOREIRA, F. C.; FERNANDES, M. R.; GUERREIRO, J. F.; RIBEIRO-DOS-SANTOS, Â.; SANTOS, S. E. B. D.; SANTOS, N. P. C. d. Characterization of dna polymerase genes in amazonian amerindian populations. *Genes*, MDPI, v. 14, n. 1, p. 53, 2022. Citado 2 vezes nas páginas 10 e 11.

ELLINGFORD, J. M.; AHN, J. W.; BAGNALL, R. D.; BARALLE, D.; BARTON, S.; CAMPBELL, C.; DOWNES, K.; ELLARD, S.; DUFF-FARRIER, C.; FITZPATRICK, D. R. et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome medicine*, Springer, v. 14, n. 1, p. 73, 2022.

JOYNT, A. T.; CUTTING, G. R.; SHARMA, N. Genetics of cystic fibrosis: clinical implications. *Clinics in Chest Medicine*, Elsevier, v. 43, n. 4, p. 591–602, 2022.

LAUSCHKE, V. M.; INGELMAN-SUNDBERG, M. Emerging strategies to bridge the gap between pharmacogenomic research and its clinical implementation. *NPJ genomic medicine*, Nature Publishing Group UK London, v. 5, n. 1, p. 9, 2020

SANTOS, A. M. Ribeiro-dos; VIDAL, A. F.; VINASCO-SANDOVAL, T.; GUERREIRO, J.; SANTOS, S.; SANTOS, Â. Ribeiro-dos; SOUZA, S. J. D. Exome sequencing of native populations from the amazon reveals patterns on the peopling of south america. *Frontiers in Genetics*, Frontiers, v. 11, p. 548507, 2020. Citado na página 13.

SAINZ, J. I. Sainz de M.; SERRA, M. B. Influencia de la farmacogenética en la diversidad de respuesta a las estatinas asociada a las reacciones adversas. *Advances in Laboratory Medicine/Avances en Medicina de Laboratorio*, De Gruyter, v. 4, n. 4, p. 353–364, 2023.

SILVA, Marcos Araújo Castro. **Perspectiva genômica sobre a origem, história e diversidade dos povos indígenas da América do Sul: do povoamento inicial à colonização europeia**. 2021. Tese de Doutorado. Universidade de São Paulo.