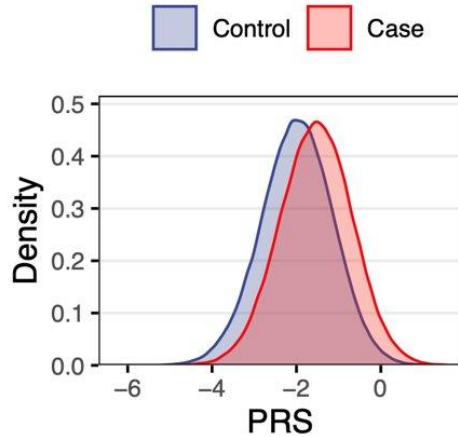


# Untitled01

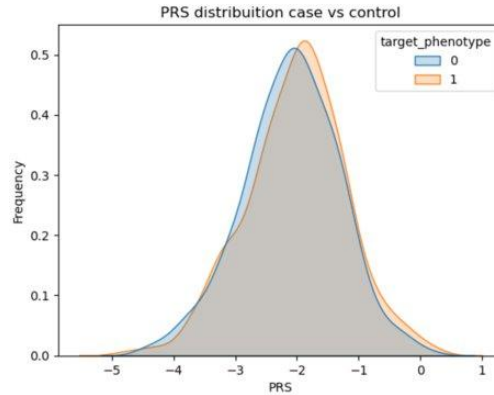
Datathon 2023 – Genética – Grupo 01

# Identificação do problema

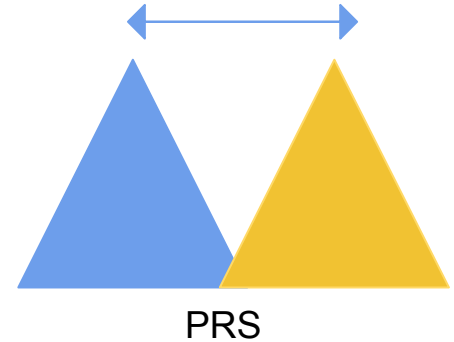
Onde funciona bem?



Nosso dado



O que precisamos fazer



# Questões de Estudo & Hipóteses

Incorporar features das variantes genéticas para melhorar o modelo

## Hipótese 1 FILTRAGEM

Filtrar variantes que possuem “impacto duvidoso” (provável não *casual variant*) melhore o modelo.

## Hipótese 2 NOVOS PESOS

Utilizar outras formas de gerar os pesos (que não apenas o beta do GWAS) pode melhorar o modelo

## Hipótese 3 FILTRAGEM + NOVOS PESOS

Combinar a H1 e H2 pode melhorar o modelo

# Metodologia Proposta

- Filtragem de variantes com base em características das variantes
- Adição de novos pesos de acordo com características das variantes

## **Crítérios de Filtragem**

- Variante presente em região gênica
- Variante presente com frequência semelhante em população AFR e EUR

## **Criar novos pesos**

- CADD
- Número de tecidos com expressão diferencialmente expressos

# Metodologia Proposta

Metodologia comum

Efeitos

VCF

Pesos genético  
(z-score)

Pesos clínicos

Regressão logística (só genético)  
Regressão logística (genético + clínico)  
(teste treino 0.66/0.33)

ROC (AUC)

# Resultados & Discussões / Filtragem



# Resultados & Discussões / Filtragem

SNP1		X		
SNP2		X		
SNP3	VCF (genótipos)	X	Efeitos/ Pesos (Beta GWAS)	NOVO SCORE GENÉTICO
SNP4		X		
SNP5		X		

# Resultados & Discussões / Filtragem

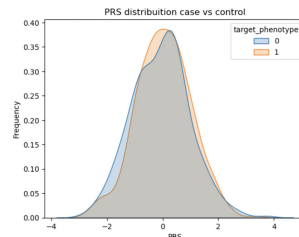
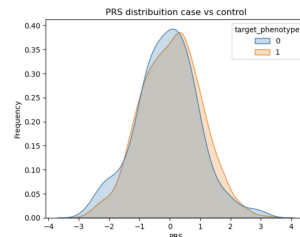
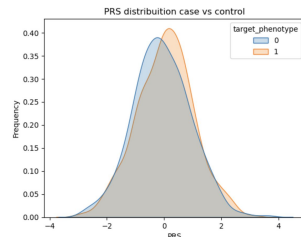
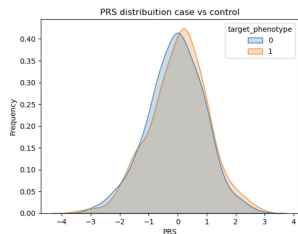
original

só r. genica  
(symbol != NA)

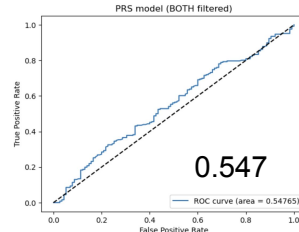
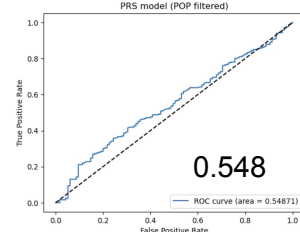
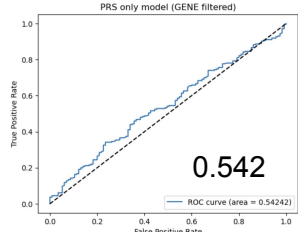
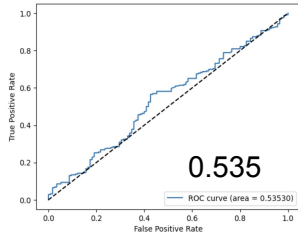
freq EUR AFR parecida  
(dif < 10%)

Both

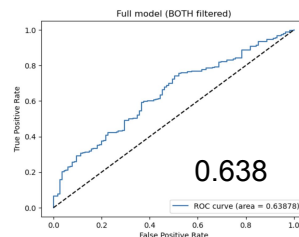
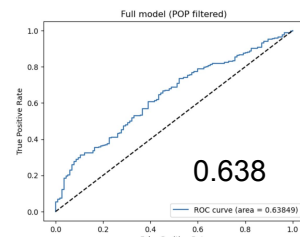
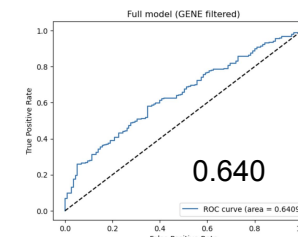
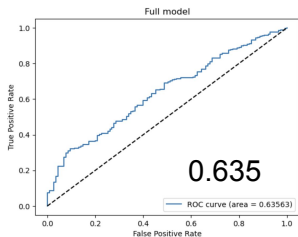
Distância



Genético



Genético +  
Clínico





# Resultados & Discussões / Novos Pesos

SNP1		X		
SNP2		X		
SNP3	VCF (genótipos)	X	Efeitos/ Pesos (Beta GWAS)	SCORE GENÉTICO
SNP4		X		
SNP5		X		

# Resultados & Discussões / Novos Pesos

SNP1		X	
SNP2		X	
SNP3	VCF (genótipos)	X	CADD score
SNP4		X	
SNP5		X	

NOVO  
SCORE  
GENÉTICO

# Resultados & Discussões / Novos Pesos

## CADD score PHRED

Como foi utilizado

Score como está

Pontuação que prevê a deleteriosidade de uma variante integrando anotações como conservação e informações funcionais do genoma. Como as pontuações brutas têm um significado relativo, as pontuações são normalizadas e ordenadas, expressando a classificação em termos de ordem de magnitude. Maiores pontuações CADD indicam variantes possivelmente mais deletérias.

## Expressão Gênica Diferencial

Como foi utilizado

Se está low ou high, score +1

**Artery\_Aorta, Artery\_Coronary, Artery\_Tibial,**

**Bladder, Heart\_Atrial\_Appendage,**

**Heart\_Left\_Ventricle, Kidney\_Cortex e**

**Whole\_Blood:** indica se o gene está diferencialmente expressos em cada um desses tecidos.

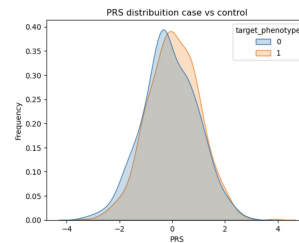
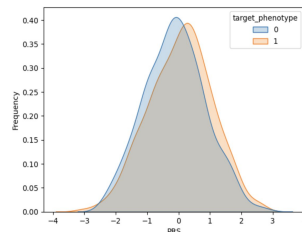
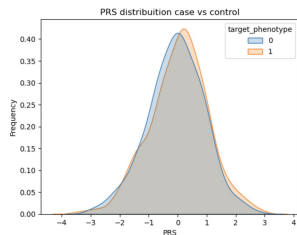
# Resultados & Discussões / Novos Pesos

original

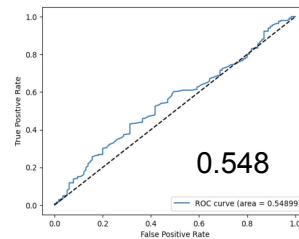
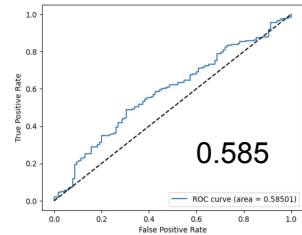
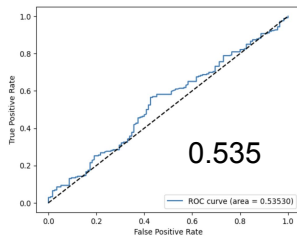
CADD PHRED

TISSUE DEG

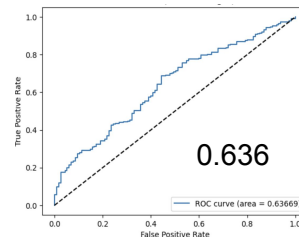
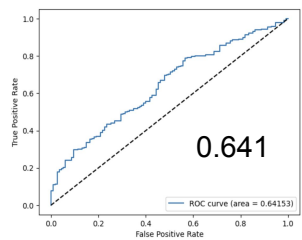
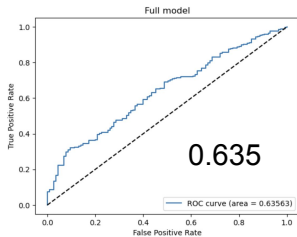
Distância



Genético



Genético +  
Clínico



# Resultados & Discussões / Novos Pesos

SNP1		X	
SNP2		X	
SNP3	VCF (genótipos)	X	Efeitos/ Pesos (Beta GWAS)
SNP4		X	
SNP5		X	

SNP1		X	
SNP2		X	
SNP3	VCF (genótipos)	X	CADD score
SNP4		X	
SNP5		X	

SNP1		X	
SNP2		X	
SNP3	VCF (genótipos)	X	Tissue DEG
SNP4		X	
SNP5		X	

SCORE  
GENÉTICO  
(BETA  
GWAS)

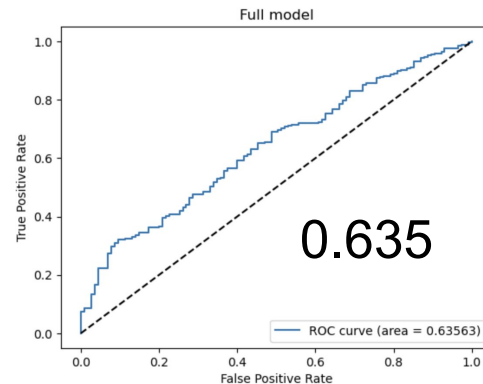
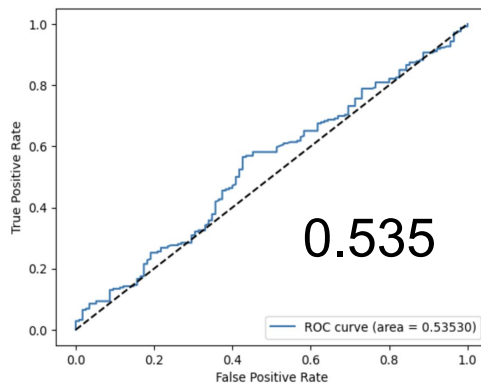
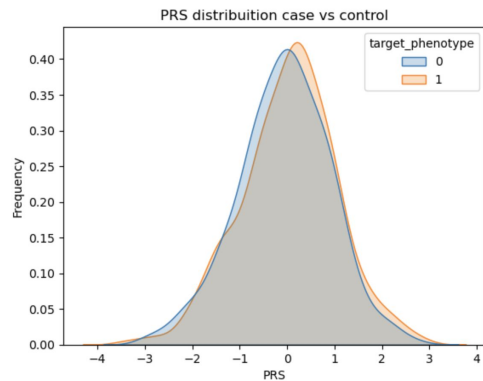
NOVO  
SCORE  
GENÉTICO

NOVO  
SCORE  
GENÉTICO

REG. LOGÍSTICA

# Resultados & Discussões / Filtragem + Novos Pesos

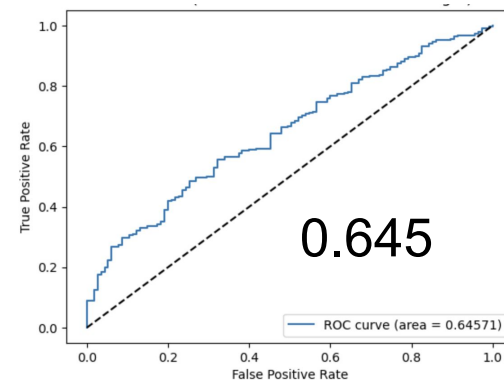
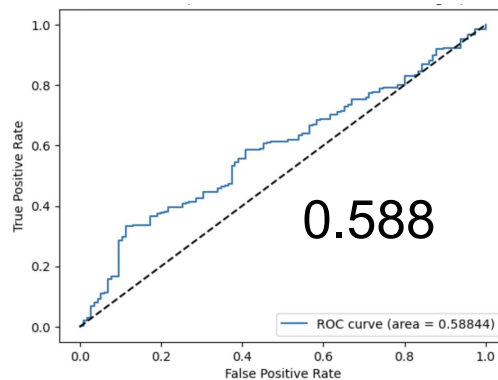
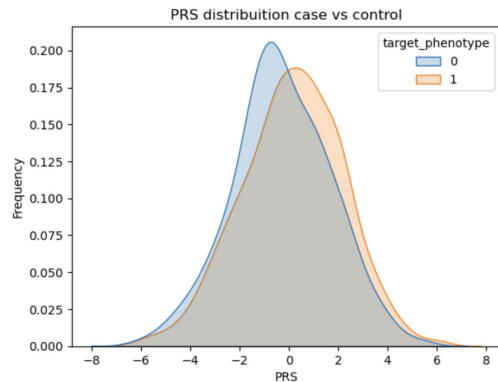
ORIGINAL (only BETA)



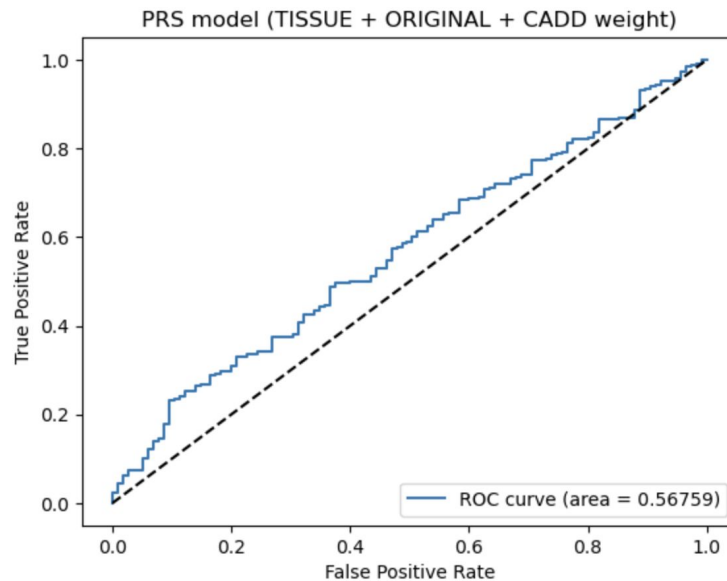
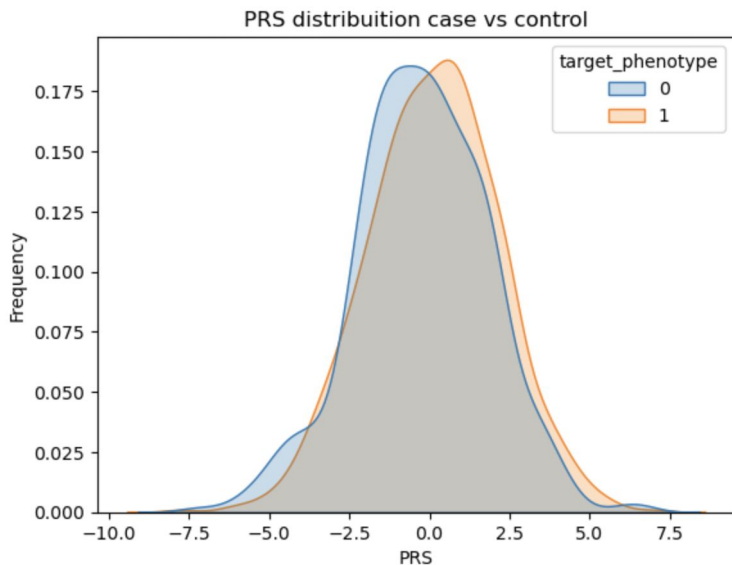
TISSUE

CADD  
PHRED

BETA



# Resultados & Discussões / Filtragem + Novos Pesos



# Aplicabilidade & Impacto

Nosso melhor modelo foi o que incorporou novos pesos baseados no impacto funcional da variante (CADD PHRED) e na expressão gênica diferencial em tecidos relevantes para o fenótipo (hipertensão);

O AUC original do modelo utilizando apenas preditores genéticos foi de 0.535 para 0.588, ainda inferior ao da população original (0.645);

A abordagem de filtragem também melhorou o poder de predição, porém de maneira inferior à utilização de novos pesos. A combinação dessas abordagens não melhorou o AUROC.



# Desafios & Próximos passos

- Podemos confiar em um acréscimo de 0.0X no AUC? É uma melhora real?
- Coortes da população alvo maior;
- Metodologia mais automatizada/sofisticada para seleção de features;