**Tutorial para análise com o Polygenic Risk Score (PRS) utilizando arquivo bruto de SNPs da Genera**

Elaborado por Luciano Menegaldo – PEB/COPPE/UFRJ – [lmeneg@peb.ufrj.br](mailto:lmeneg@peb.ufrj.br) – 8/11/22

https://prs.byu.edu/index.html

paper: <https://www.nature.com/articles/s42003-022-03795-x>

Page, M. L., Vance, E. L., Cloward, M. E., Ringger, E., Dayton, L., Ebbert, M. T., ... & Kauwe, J. S. (2022). The Polygenic Risk Score Knowledge Base offers a centralized online repository for calculating and contextualizing polygenic risk scores. *Communications biology*, *5*(1), 1-15.

O calculador do PRS baseado em web <https://prs.byu.edu/calculate_score.html> só permite poucos testes de cada vez e dá um monte de paus. Melhor não usar e ir para a linha de comando, como explicado abaixo.

1. Limpar e transformar e o arquivo bruto da Genera no formato do PRS

O arquivo da Genera apresenta vários tipos de problemas com alguns snps, que o programa abaixo filtra, isto é, elimina da amostra. Sem isso, o PRS não roda.

Utilizar o código matlab (R2020) genera2prs2.m, escrevendo o nome do arquivo bruto em arquivo='NNNNN.csv'; e o nome do arquivo de saída arq\_saida='SSSSS.txt';

1. Preparação do Windows – está explicado em

Para facilitar, instalar o Bash no Windows <https://windowsclub.com.br/como-ativar-ou-instalar-bash-no-windows-10/>

Eu instalei o Debian e funcionou

1. Baixar o programa do PRS em https://prs.byu.edu/cli\_download.html

(tem uns vídeos explicando como instalar e rodar, mas eu não consegui fazer igual ao vídeo)

No README.md (que está no pacote), é explicado como instalar o Phyton e o PRS

# Polygenic Risk Score Knowledge Base Command-line Interface Calculator

Instalar o Phyton

Na tela do bash:

pip install PyVCF

pip install filelock

pip install requests

pip install myvariant

pip install biopython

pip install biothings\_client

For certain tool functions accessed through the tool's menu, bash \*\*\*jq\*\*\* is required. To download on Ubuntu or Debian run:

sudo apt-get install jq

1. Escrever o comando para calcular o arquivo de saída

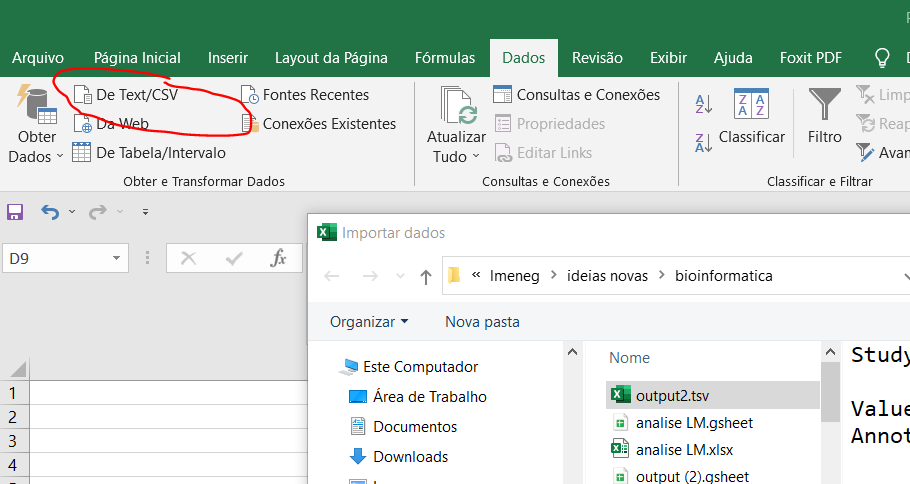
No exemplo, output.tsv a partir do arquivo de entrada (saida3.txt) usando o genoma de referência hg38, população europeia (ou outra)

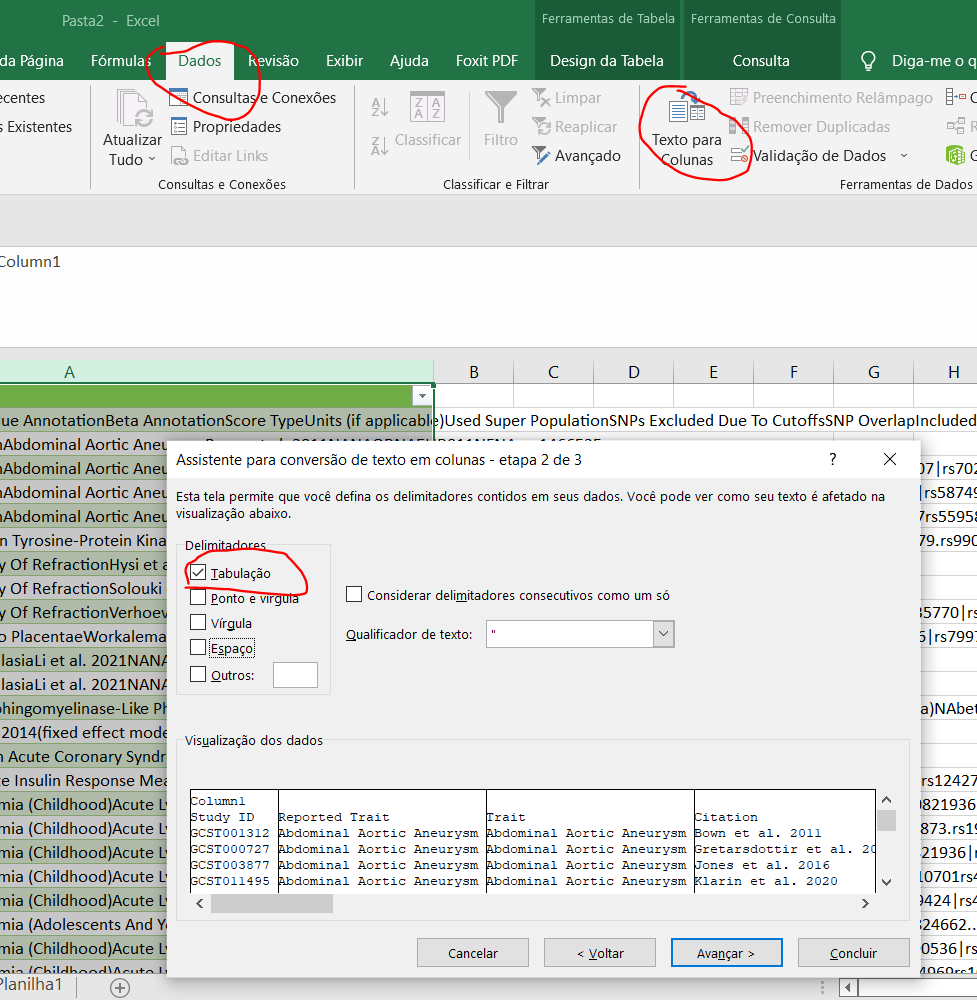
lmeneg@Titan:/mnt/c/Users/lucia/PrskbCLITool$ ./runPrsCLI.sh -f saida3.txt -o output3.tsv -r hg38 -c 0.05 -p EUR -v

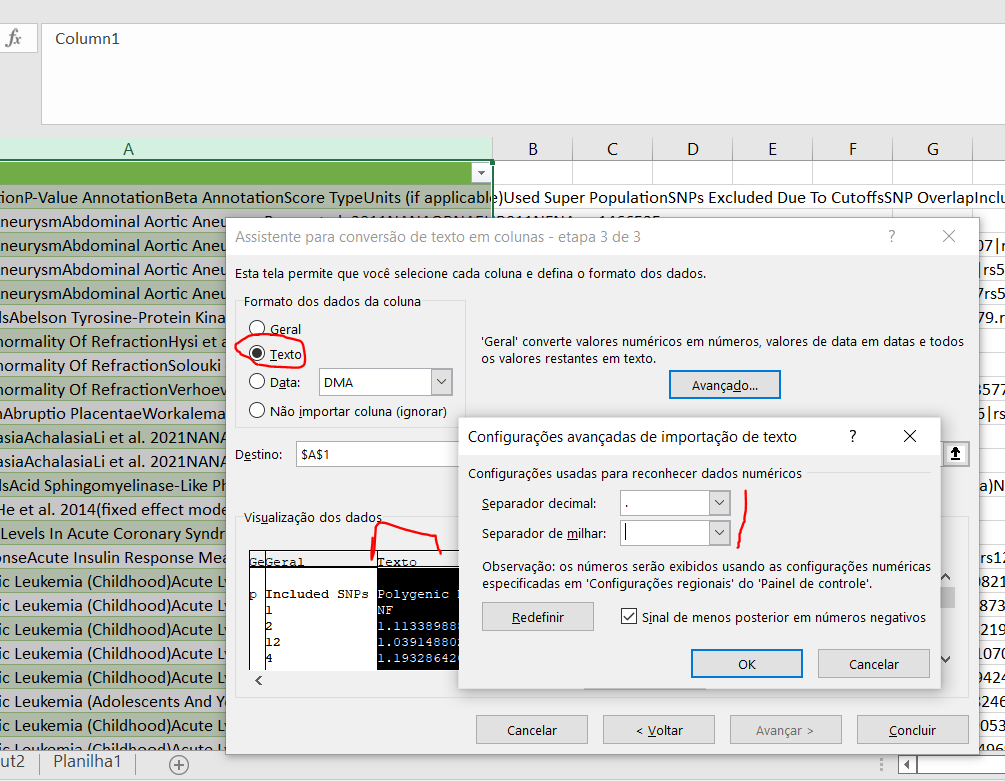
1. Importação no Excel

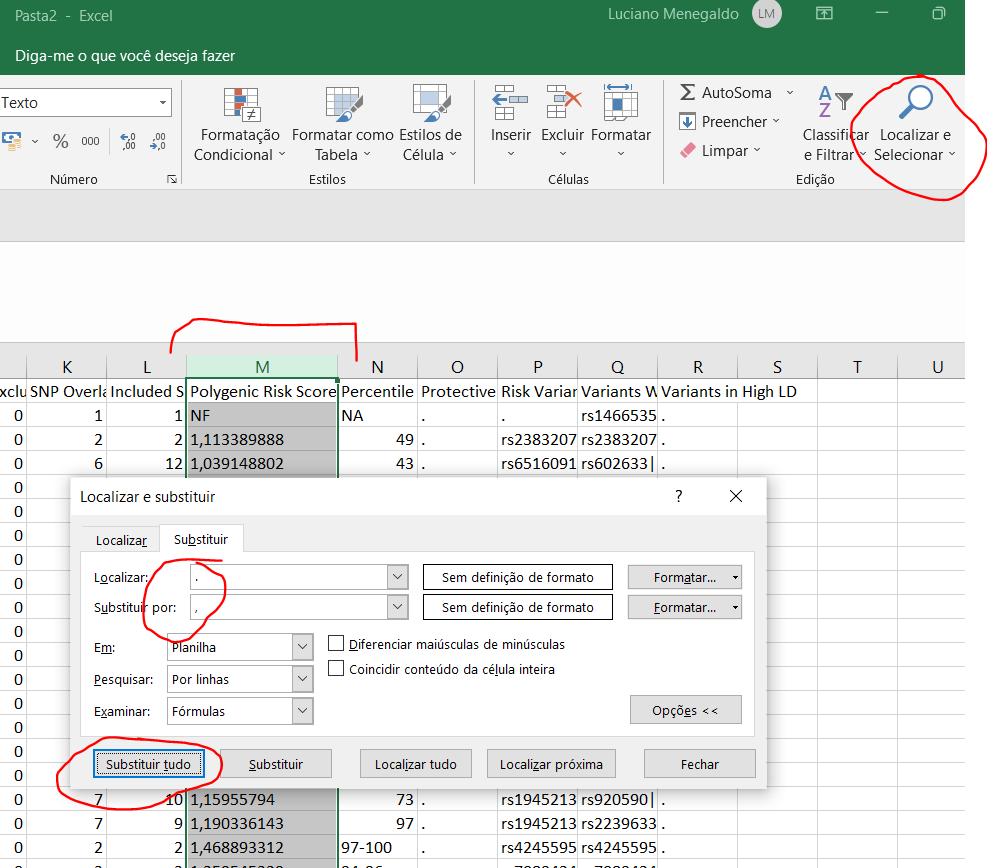
O problema acontece na coluna do PRS (que é a mais importante), com a confusão entre virgula e ponto que faz o Excel. Depois de muitas tentativas, a única maneira que eu achei para funcionar é a seguinte:

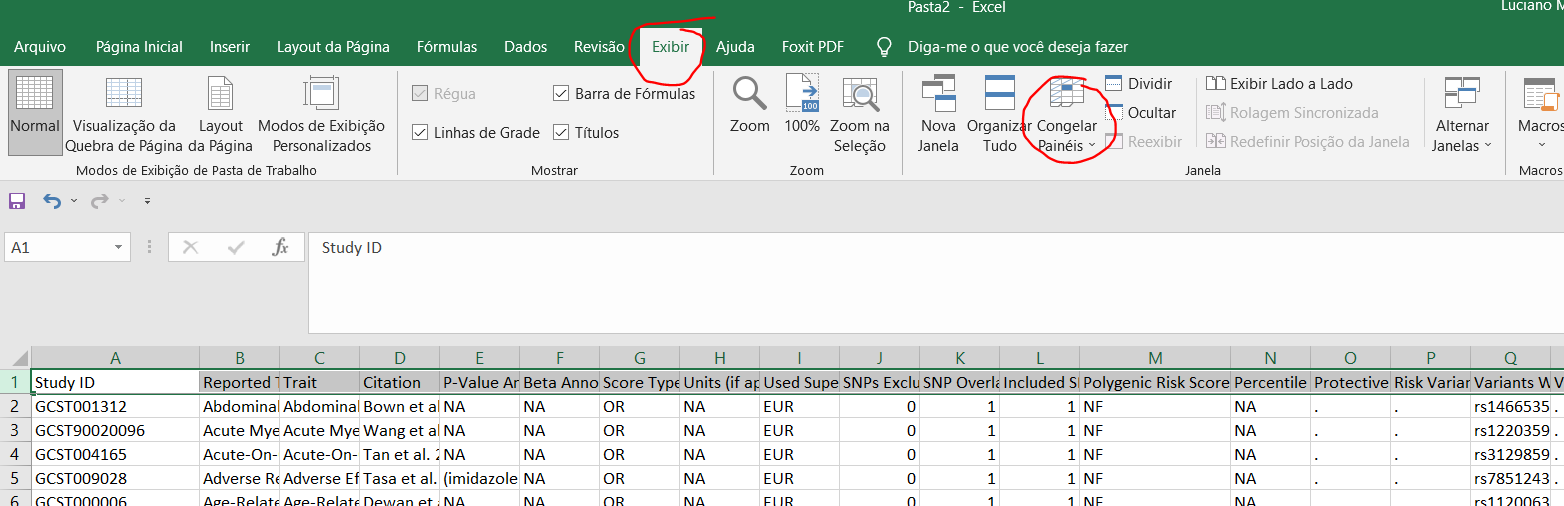
Importar os dados de saída do PRS como csv, seguindo as telas abaixo



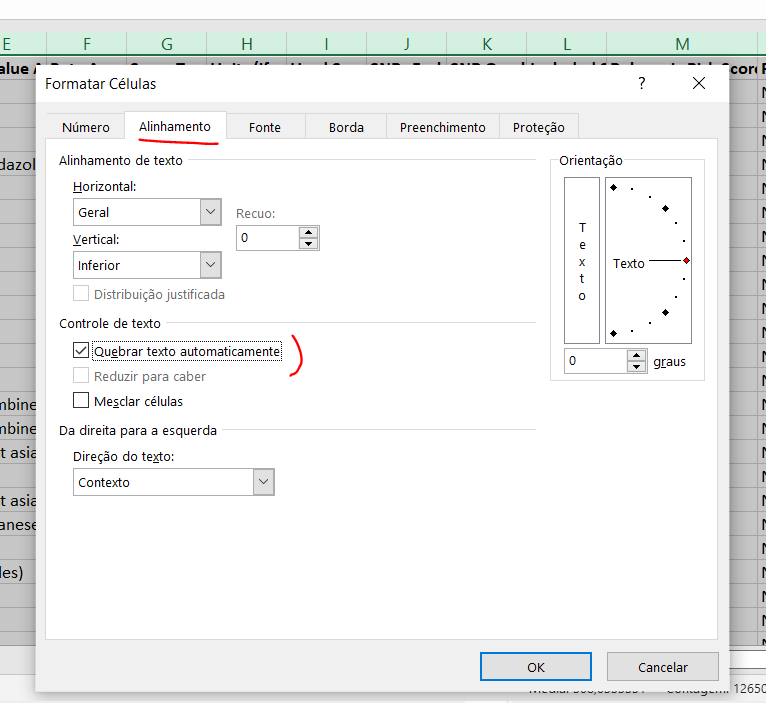






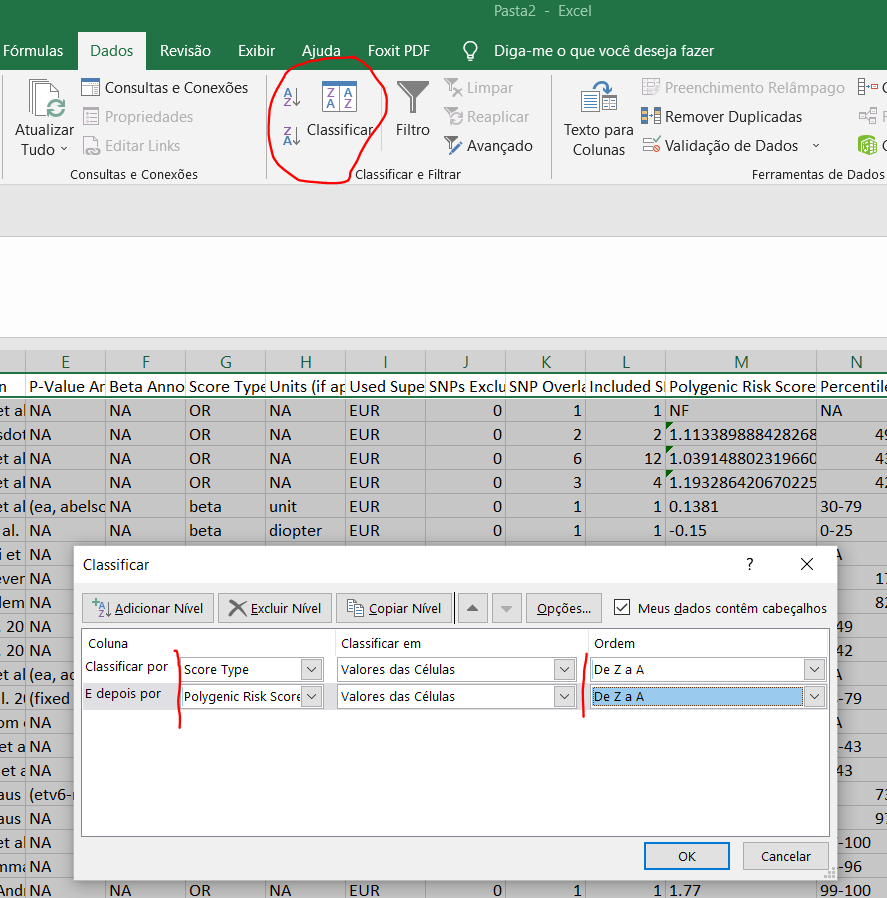


Formatar a coluna percentile como texto



Eu fiz uma classificação por score type (OD – odds ratio) e beta e depois pelo PRS, começando dos valores mais altos. O Excel vai colocar na ordem: snps sem risco e depois os com risco começando pelos riscos maiores.

As métricas do OR e do Beta são diferentes, por isso, ordenar separadamente. OD > 1 significa risco aumentado, quanto maior, maior o risco. Beta positivo, maior risco, negativo, menor.



Verificar, além do valor do PRS, o ‘percentile’, para entender como é a distribuição daquele PRS na população.

Pode-se verificar para cada trait o gráfico da distribuição, para cada artigo que serviu de base, em <https://prs.byu.edu/visualize.html>

Nos dois casos abaixo, o primeiro trait está em 52-53% da população, enquanto o segundo 89-98% (bem mais crítico)

