

# ClinVar Report

*James Diao*

*January 8, 2017*

## Contents

<b>1</b>	<b>Collect and Merge ClinVar Data</b>	<b>2</b>
1.1	Import ClinVar VCF . . . . .	2
1.2	Merge ClinVar with 1000 Genomes and ExAC . . . . .	2
<b>2</b>	<b>Summary Statistics</b>	<b>3</b>
2.1	Fraction of Individuals with Pathogenic Non-Reference Sites . . . . .	3
<b>3</b>	<b>Penetrance Estimates</b>	<b>4</b>
3.1	Max/Min Penetrance as a Function of $P(D)$ and $P(V D)$ . . . . .	4
3.2	Penetrance Estimates by Ancestry . . . . .	5

**Sourcing ClinVar input from:** clinvar\_2016-02-03.vcf

**Sending output to:** Report\_2016-02-03.pdf

# 1 Collect and Merge ClinVar Data

## 1.1 Import ClinVar VCF

## 1.2 Merge ClinVar with 1000 Genomes and ExAC

## Breakdown of ClinVar Variants

Subset_ClinVar	Number_of_Variants
Total ClinVar	97867
LP/P	30341
ACMG LP/P	6474
ACMG LP/P in gnomAD	1380
ACMG LP/P in ExAC	986
ACMG LP/P in 1000 Genomes	176

## Breakdown of ACMG-gnomAD Variants

Subset_gnomAD	Number_of_Variants
ACMG in gnomAD	96742
ClinVar-ACMG in gnomAD	9934
LP/P-ACMG in gnomAD	1380

## Breakdown of ACMG-ExAC Variants

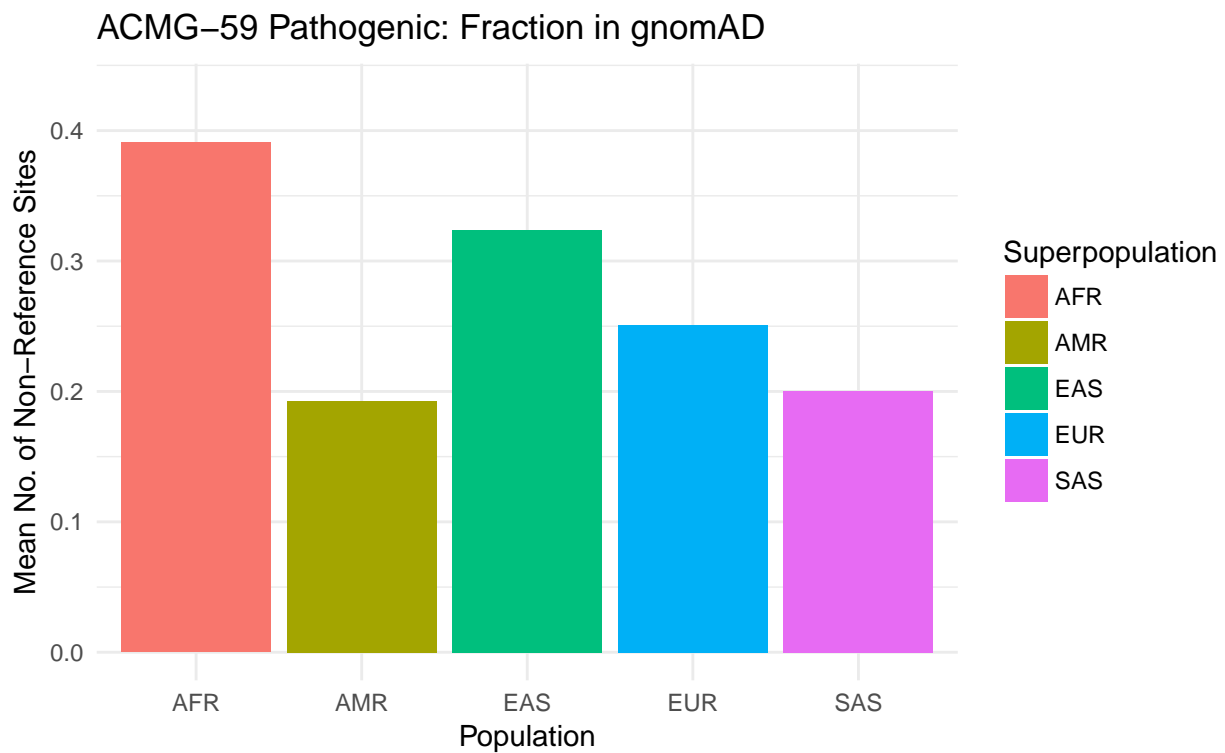
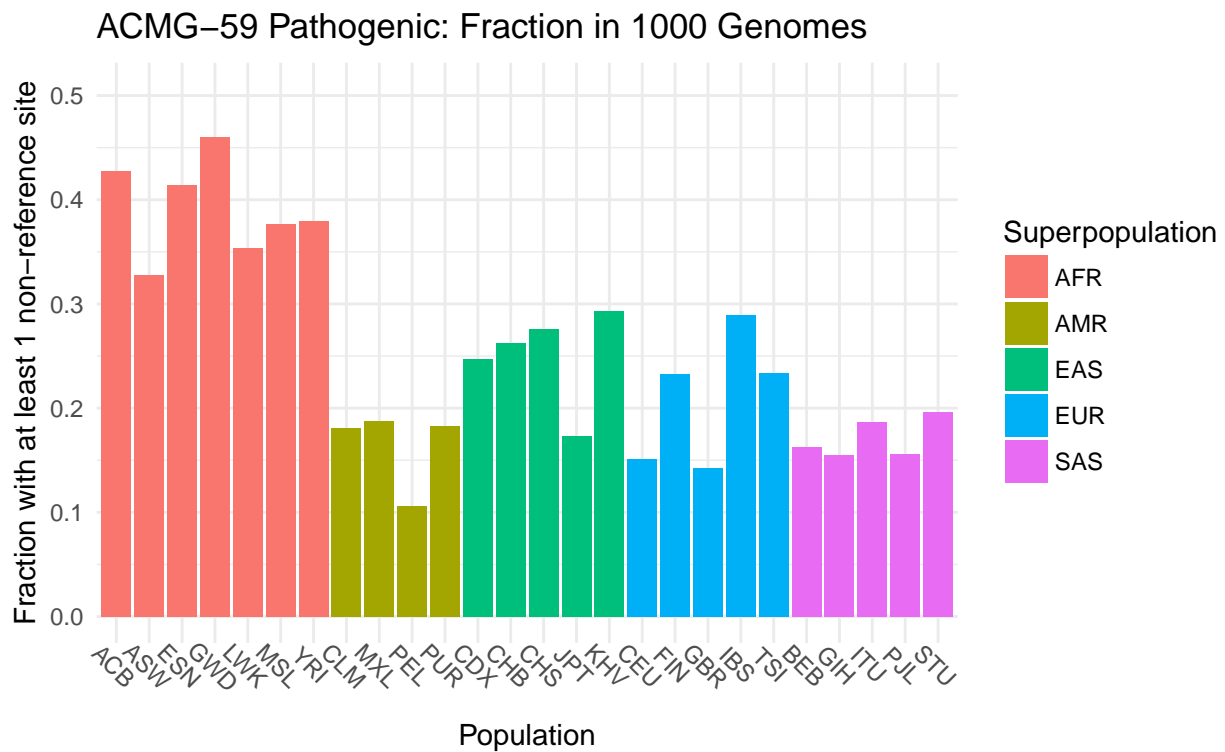
Subset_gnomAD	Number_of_Variants
ACMG in ExAC	59883
ClinVar-ACMG in ExAC	8138
LP/P-ACMG in ExAC	986

## Breakdown of ACMG-1000G Variants

Subset_gnomAD	Number_of_Variants
ACMG in 1000G	141466
ClinVar-ACMG in 1000G	3968
LP/P-ACMG in 1000G	176

## 2 Summary Statistics

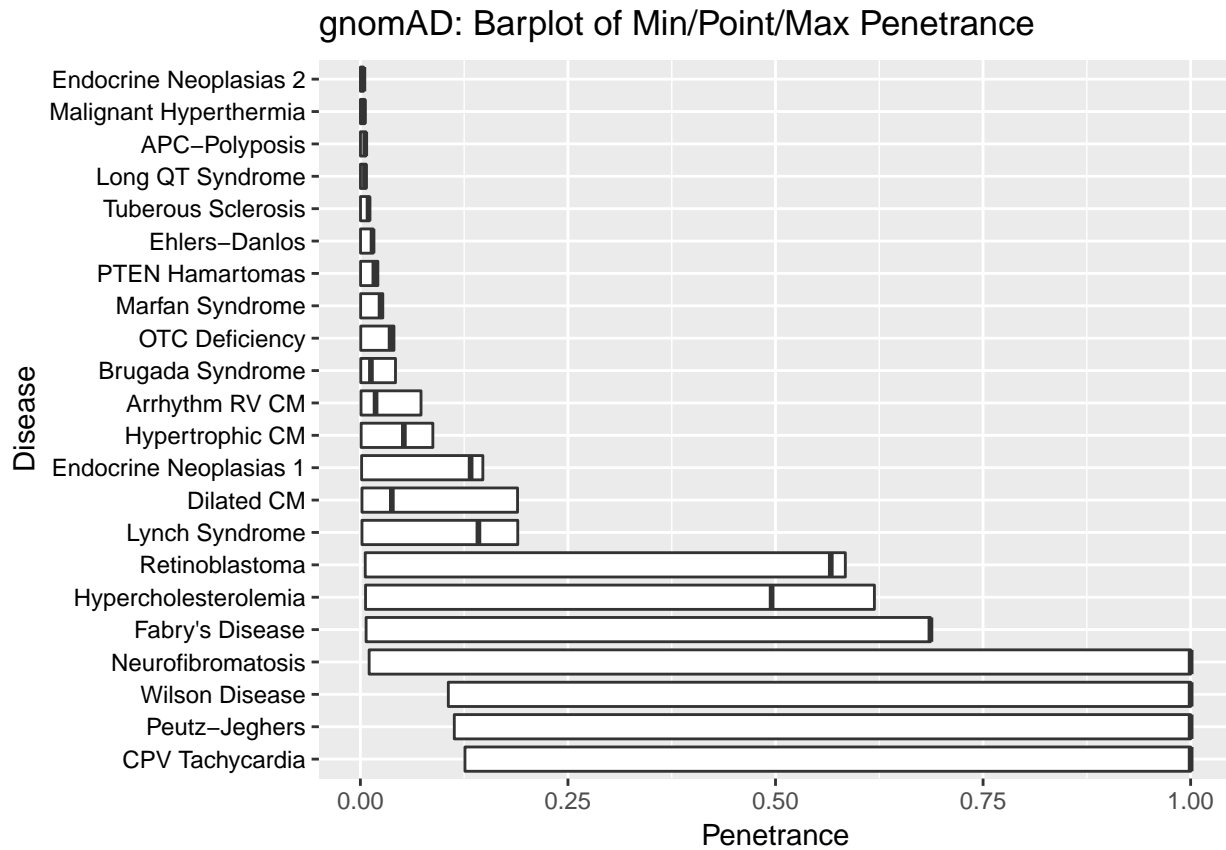
### 2.1 Fraction of Individuals with Pathogenic Non-Reference Sites



### 3 Penetrance Estimates

#### 3.1 Max/Min Penetrance as a Function of $P(D)$ and $P(V|D)$

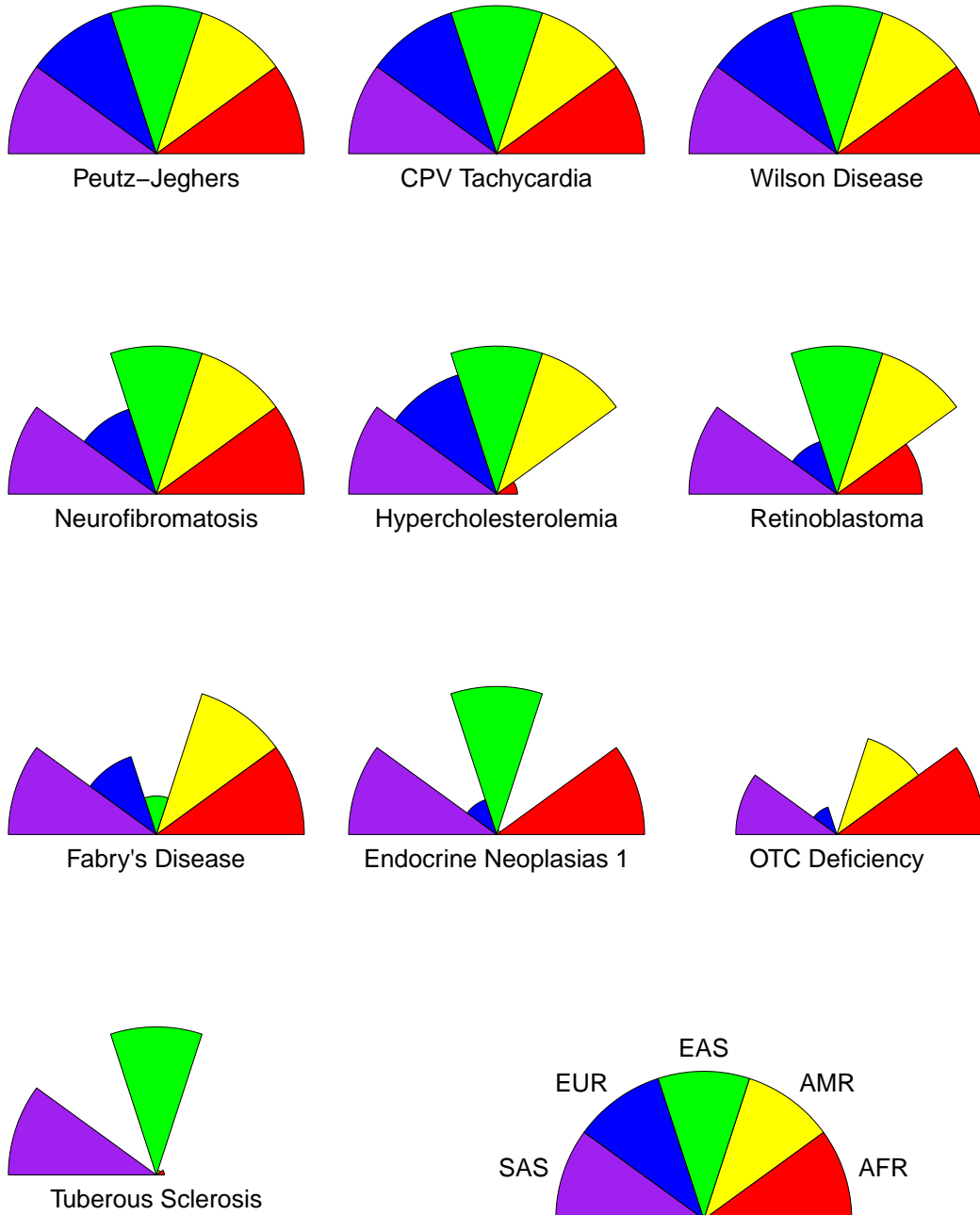
The left end of the boxplot indicates  $P(V|D) = 0.01$ ,  
the bold line in the middle indicates  $P(V|D) = \text{point value}$ ,  
the right end of the boxplot indicates  $P(V|D) = 1$ .



Note: Some diseases have mean theoretical penetrance = 1 because the assumed allelic heterogeneity is greater than is possible, given the observed prevalence and allele frequencies.

### 3.2 Penetrance Estimates by Ancestry

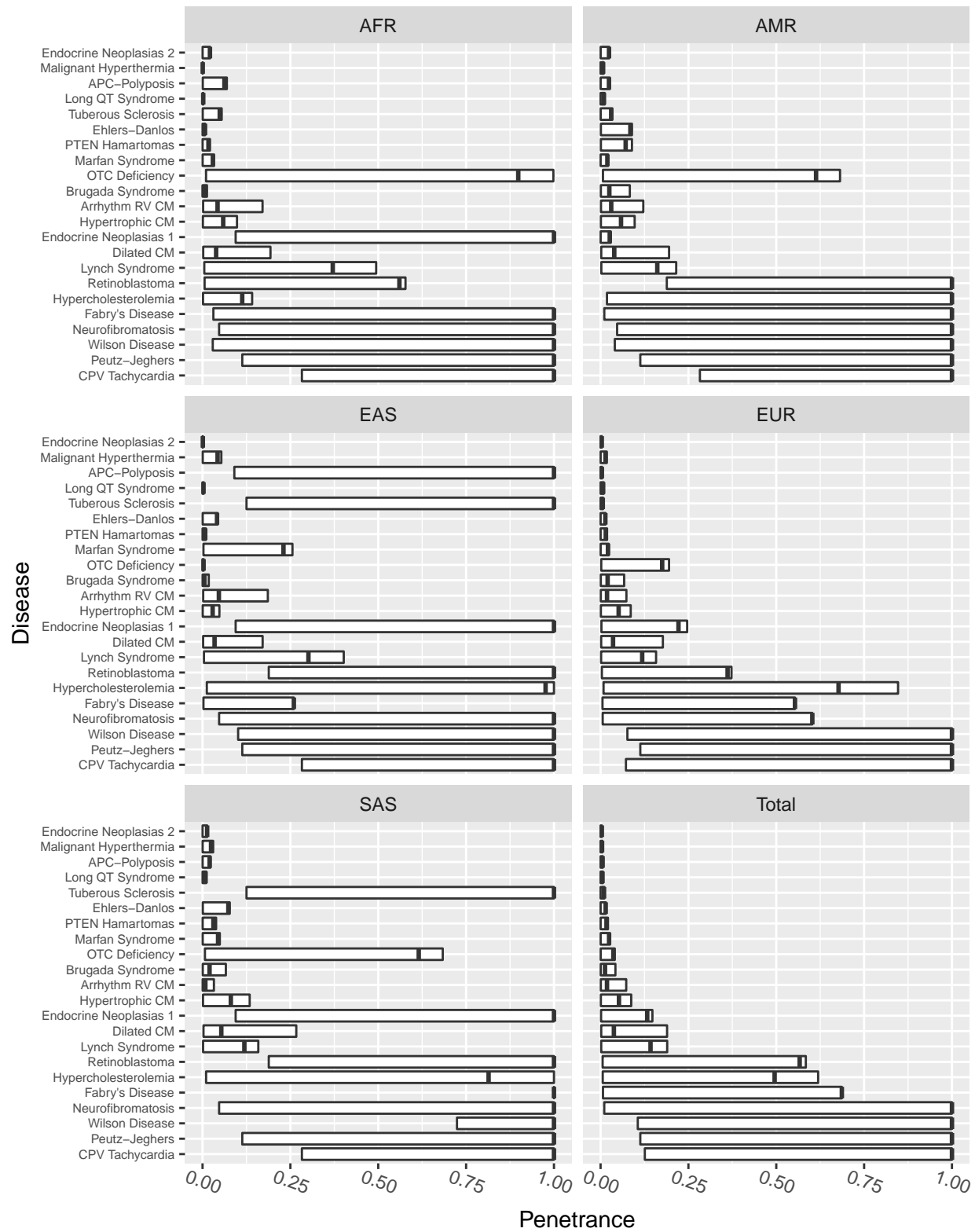
#### Radar Plot: Max Penetrance by Ancestry (gnomAD)



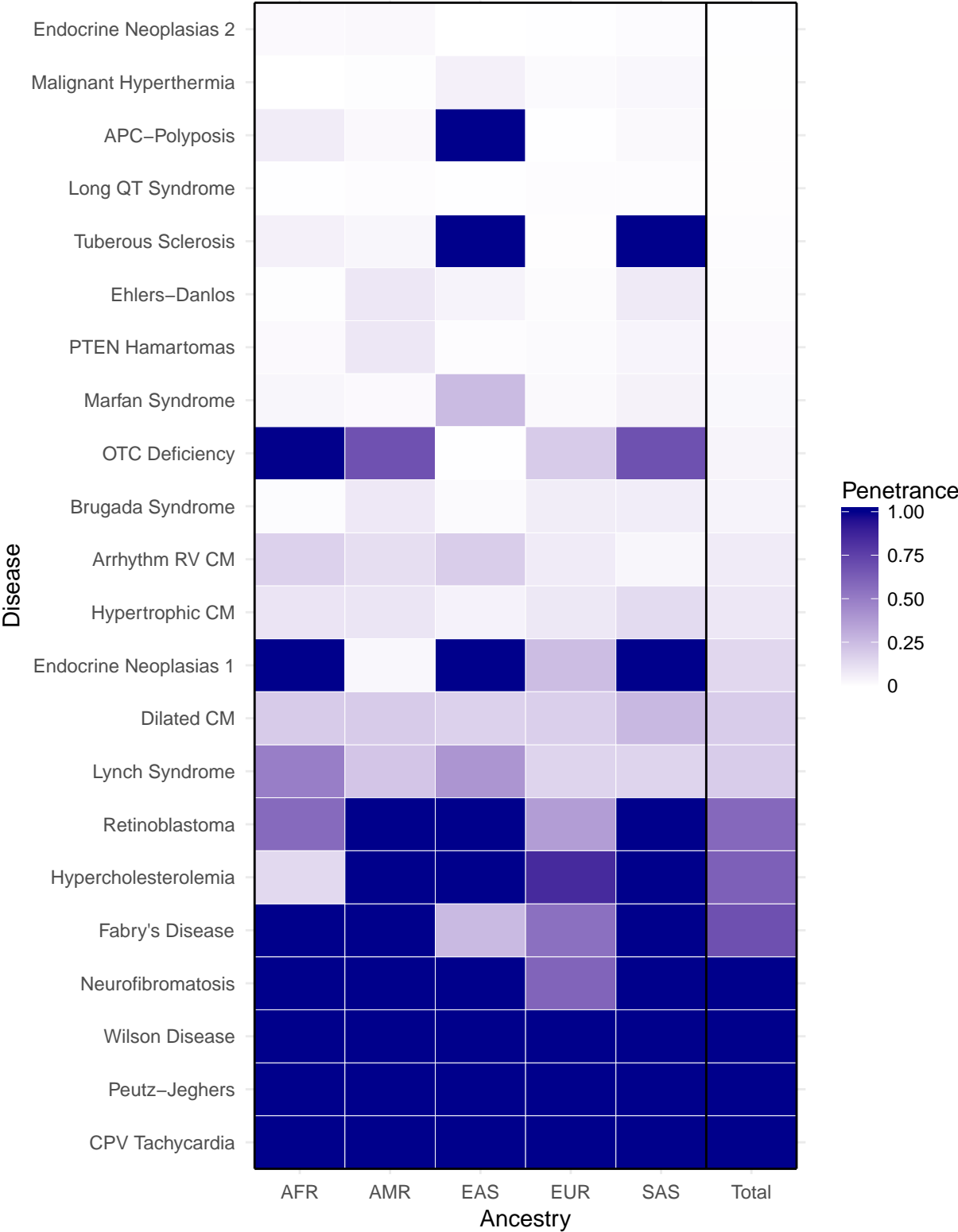
## [1] These are the top 10 diseases by summed allele frequencies. NULL values are not plotted.

## [1] Each radius is proportional to the penetrance of the disease in the given population.

Barplot: Penetrance by Ancestry (gnomAD)



Heatmap: Max Penetrance by Ancestry (gnomAD)



## Dark gray boxes are NA: no associated variants discovered in that ancestral population.