ClinVar Report

James Diao

January 8, 2017

Contents

Sourcing ClinVar input from: clinvar_2016-05-31.vcf
Sending output to: Report_2016-05-31.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	104647
LP/P	31287
ACMG LP/P	6593
ACMG LP/P in gnomAD	1395
ACMG LP/P in ExAC	996
ACMG LP/P in 1000 Genomes	165

```
## Breakdown of ACMG-gnomAD Variants
```

Subset_gnomAD	Number_of_Variants
ACMG in gnomAD	96742
ClinVar-ACMG in gnomAD	10732
LP/P-ACMG in gnomAD	1395

```
## Breakdown of ACMG-ExAC Variants
```

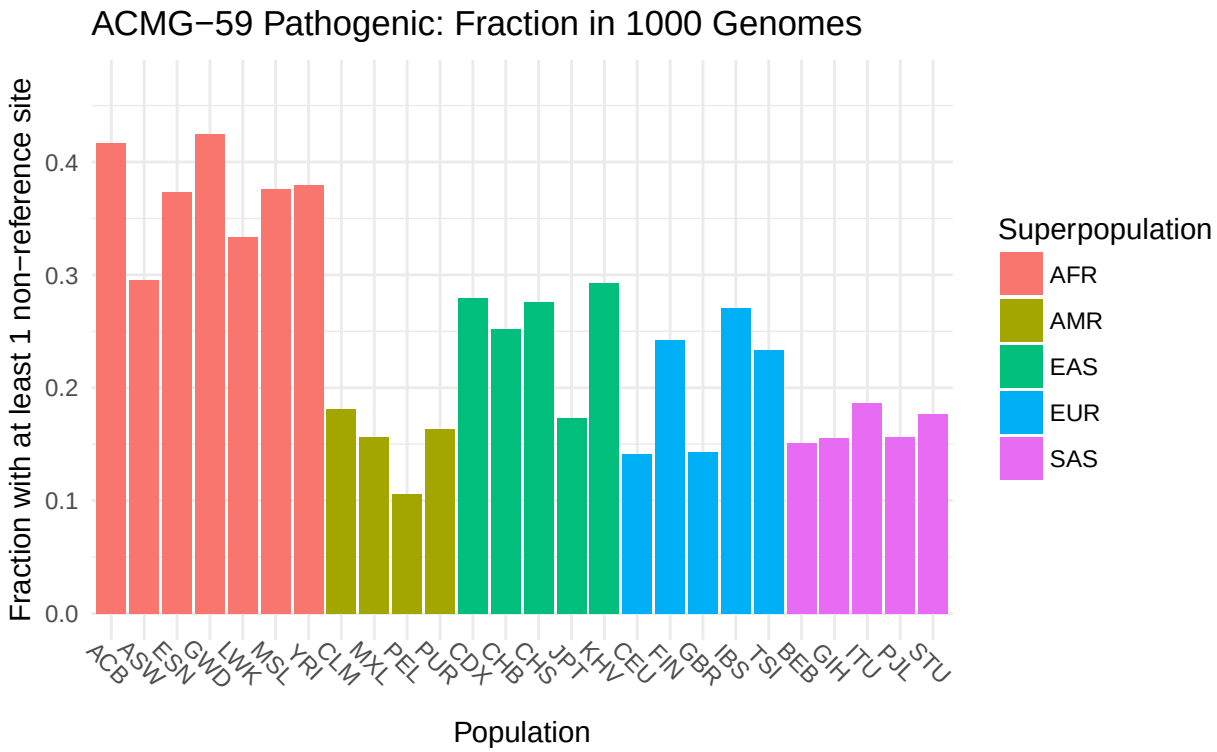
Subset_gnomAD	Number_of_Variants
ACMG in ExAC	59883
ClinVar-ACMG in ExAC	8851
LP/P-ACMG in ExAC	996

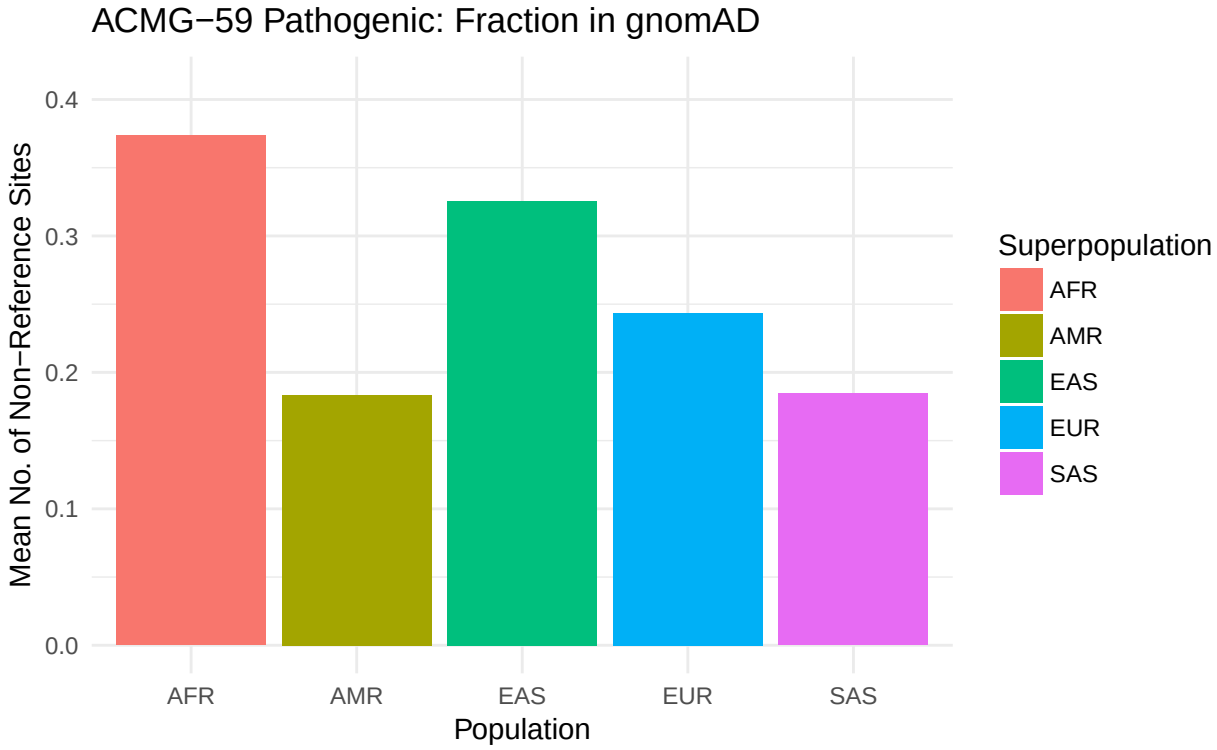
```
## Breakdown of ACMG-1000G Variants
```

Subset_gnomAD	Number_of_Variants
ACMG in 1000G	141466
ClinVar-ACMG in 1000G	4250
LP/P-ACMG in 1000G	165

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

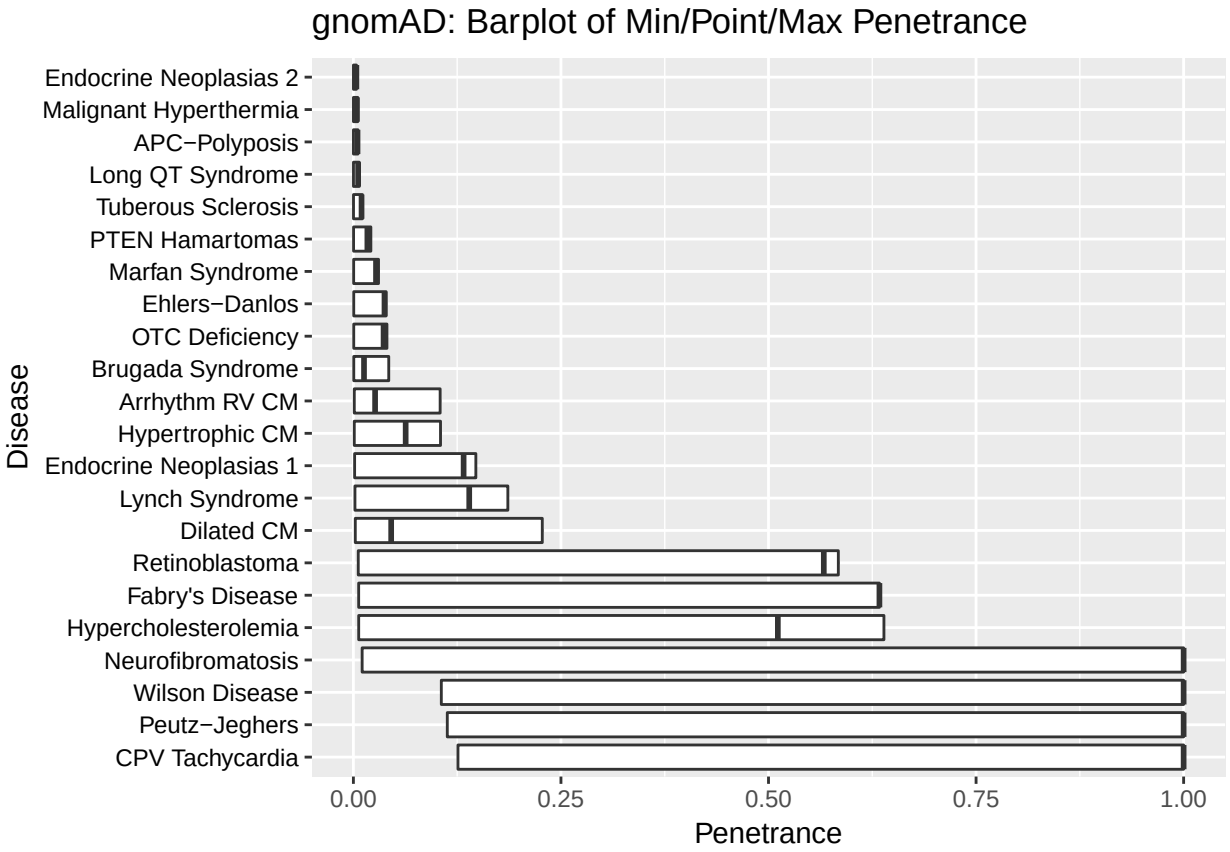




3 Penetrance Estimates

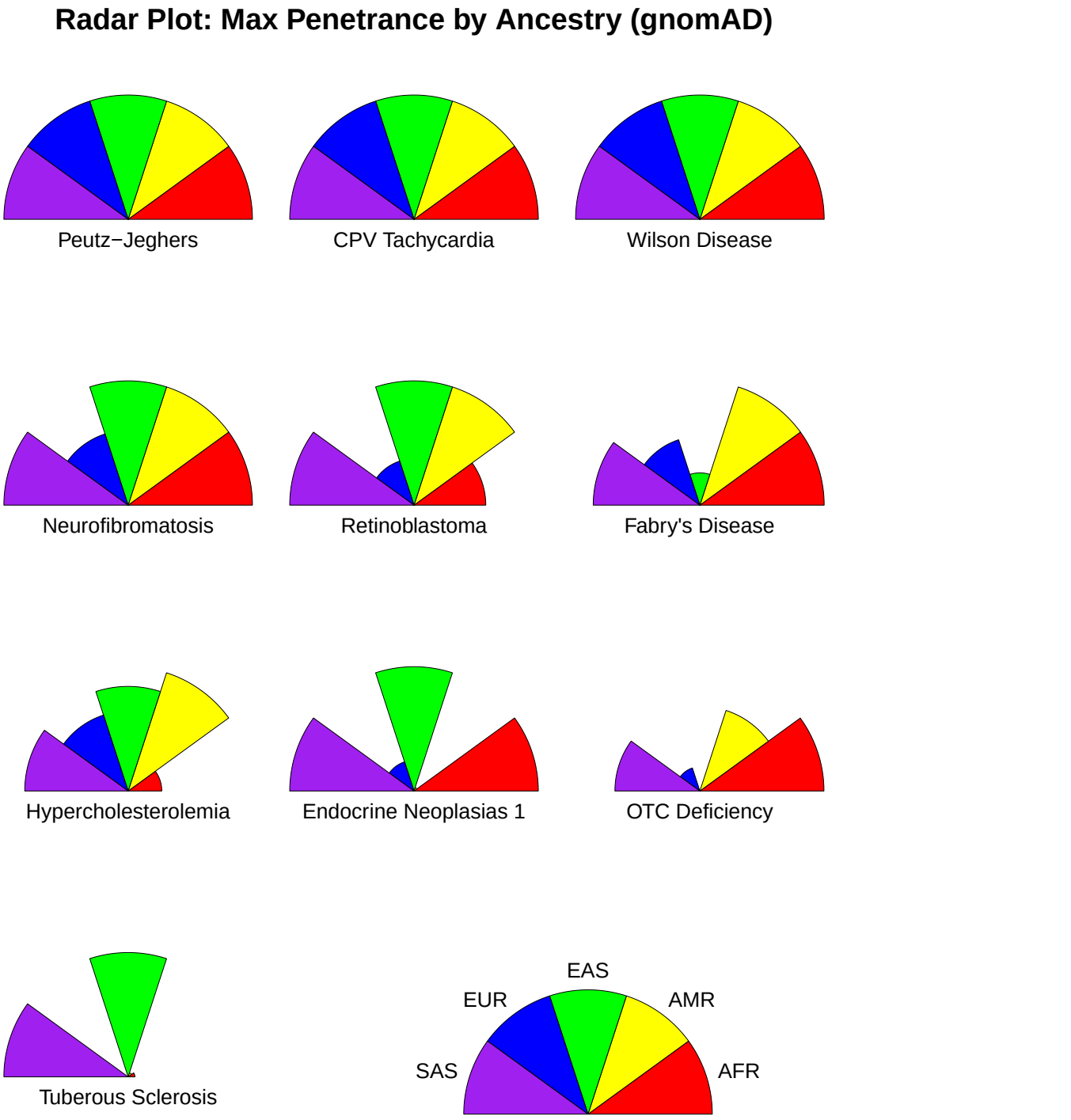
3.1 Max/Min Penetrance as a Function of P(D) and P(V|D)

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

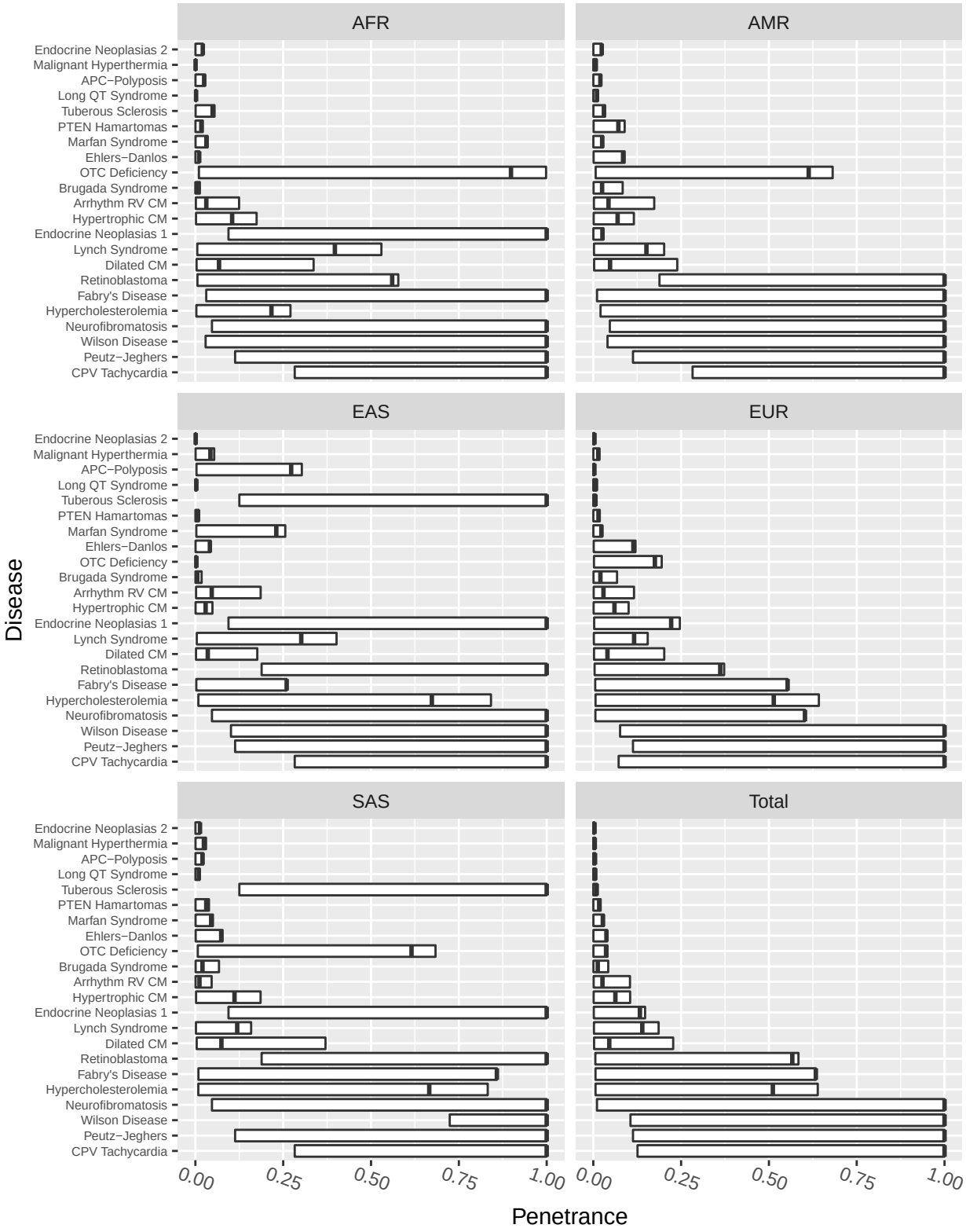


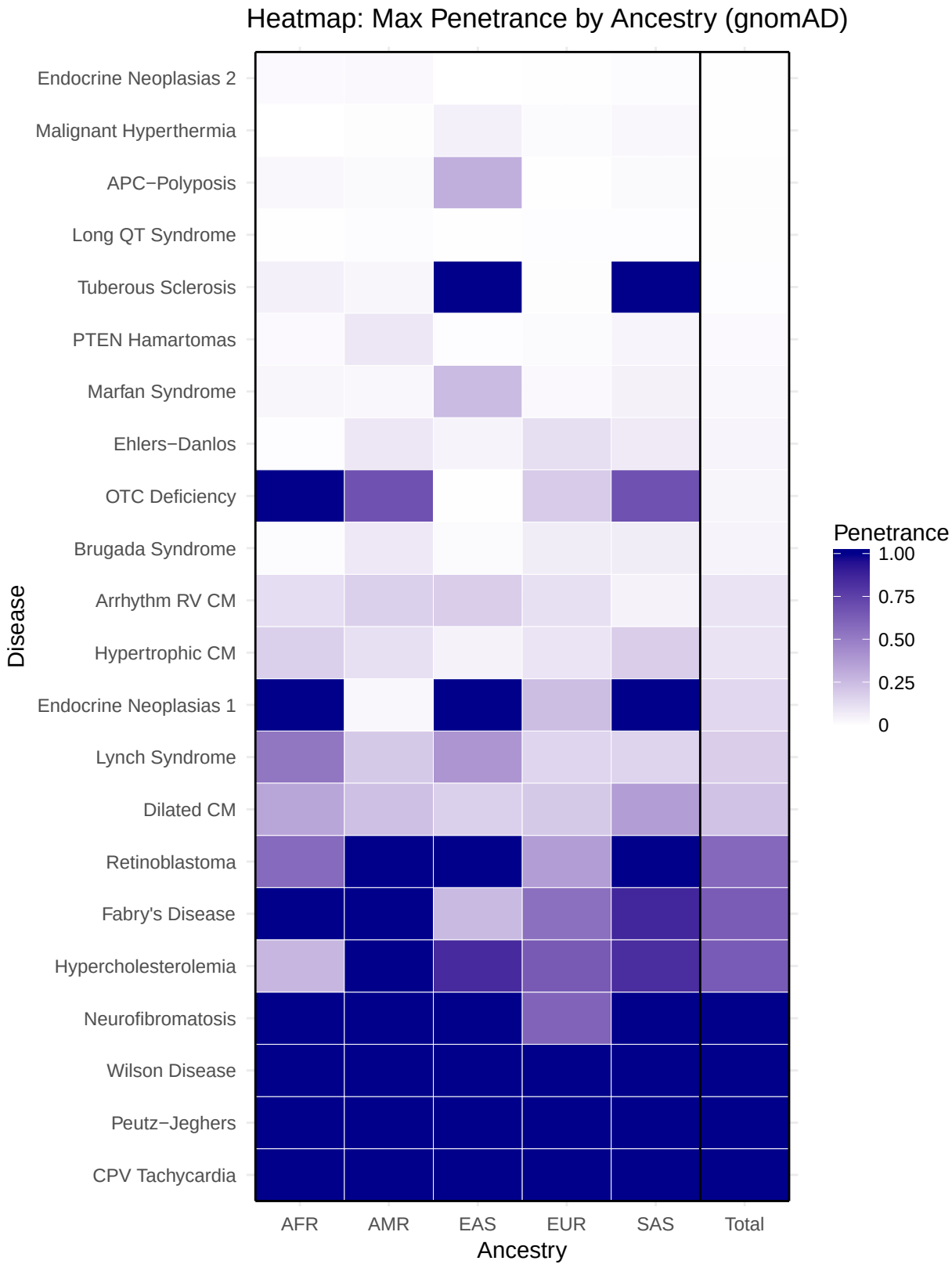
Note: Some diseases have mean theoretical penetrance = 1 because the assumed allelic heterogeneity is greater than is possible, given the observed prevalence and allele frequencies.

3.2 Penetrance Estimates by Ancestry




```
## [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)
AFR
AMR
EAS
EUR
SAS
Total
Endocrine Neoplasias 2
Malignant Hyperthermia
APC−Polyposis
Long QT Syndrome
Tuberous Sclerosis
PTEN Hamartomas
Marfan Syndrome
Ehlers−Danlos
OTC Deficiency
Brugada Syndrome
Arrhythm RV CM
Hypertrophic CM
Endocrine Neoplasias 1
Lynch Syndrome
Dilated CM
Retinoblastoma
Fabry's Disease
Hypercholesterolemia
Neurofibromatosis
Wilson Disease
Peutz−Jeghers
CPV Tachycardia
Disease
0.00
0.25
0.50
0.75
1.00
Penetrance



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