ClinVar Report

James Diao

January 8, 2017

Contents

1 Collect and Merge ClinVar Data **2**
1.1 Import ClinVar VCF 2
1.2 Merge ClinVar with 1000 Genomes and ExAC 2

2 Summary Statistics **3**
2.1 Fraction of Individuals with Pathogenic Non-Reference Sites 3

3 Penetrance Estimates **4**
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_2013-01-18.vcf
Sending output to: Report_2013-01-18.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	67042
LP/P	12647
ACMG LP/P	977
ACMG LP/P in gnomAD	312
ACMG LP/P in ExAC	227
ACMG LP/P in 1000 Genomes	72

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

Subset_gnomAD	Number_of_Variants
ACMG in gnomAD	96742
ClinVar-ACMG in gnomAD	2833
LP/P-ACMG in gnomAD	312

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

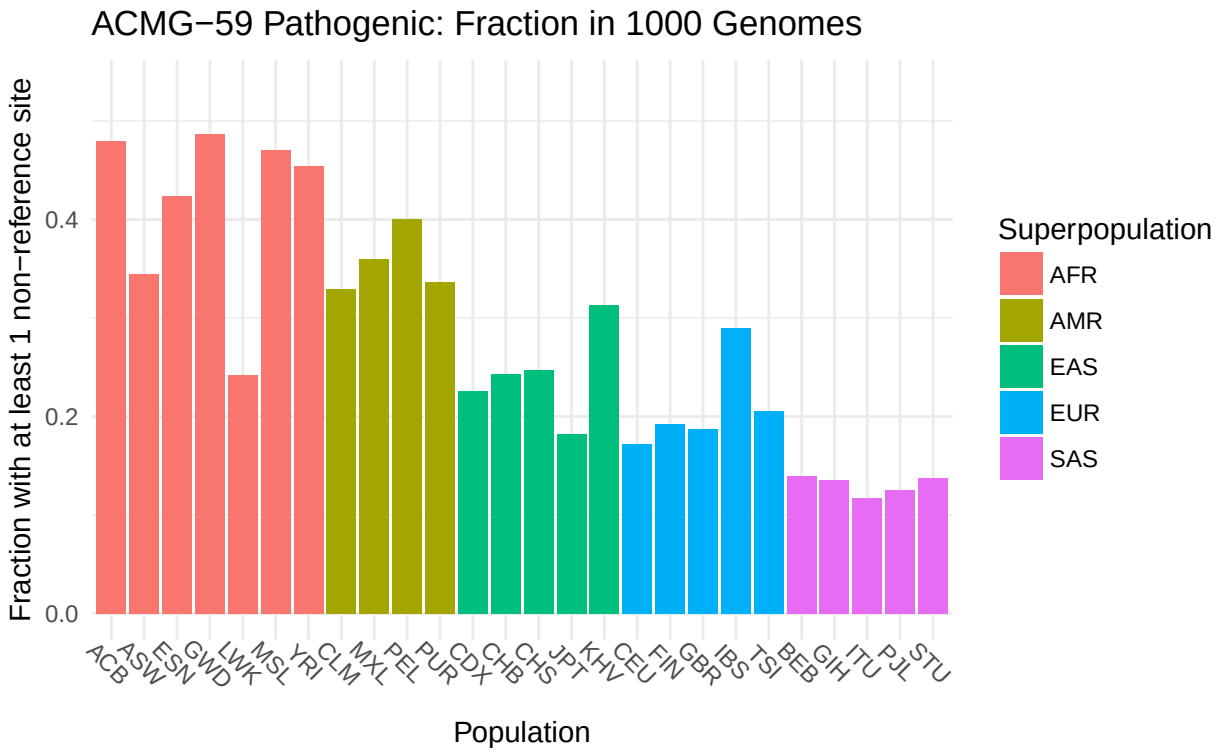
Subset_gnomAD	Number_of_Variants
ACMG in ExAC	59883
ClinVar-ACMG in ExAC	2303
LP/P-ACMG in ExAC	227

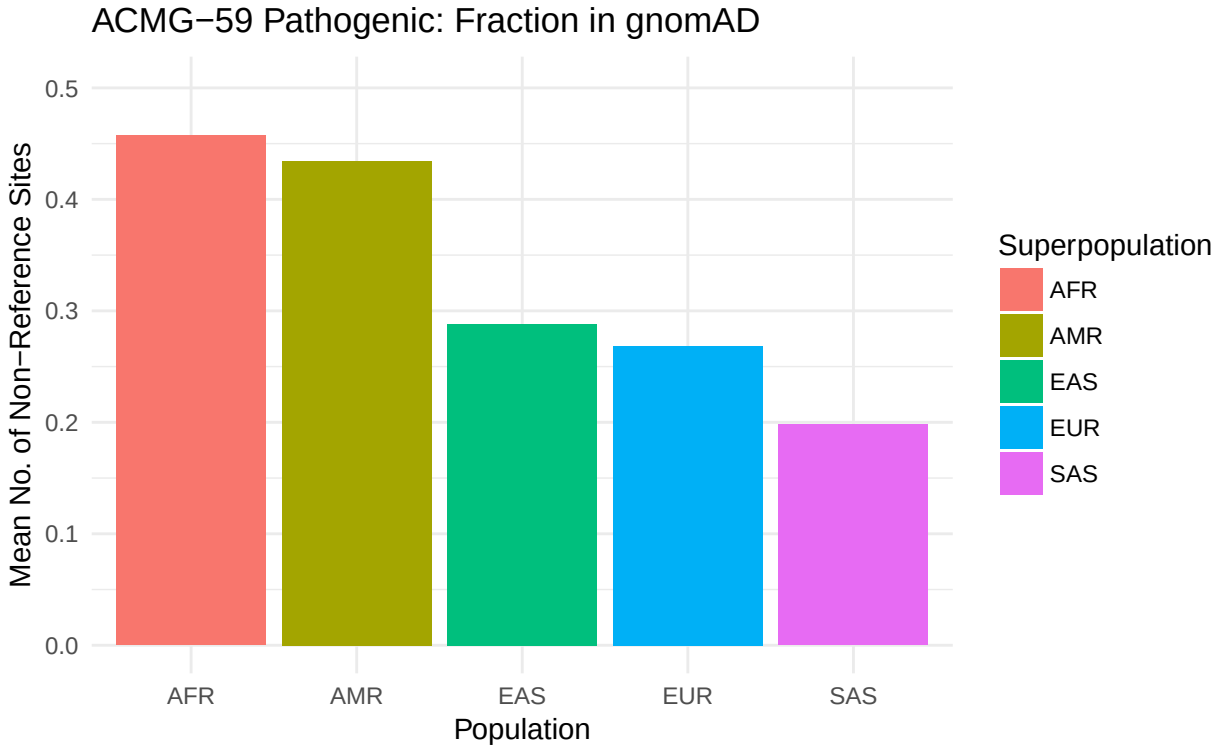
```
## Breakdown of ACMG-1000G Variants
```

Subset_gnomAD	Number_of_Variants
ACMG in 1000G	141466
ClinVar-ACMG in 1000G	1246
LP/P-ACMG in 1000G	72

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

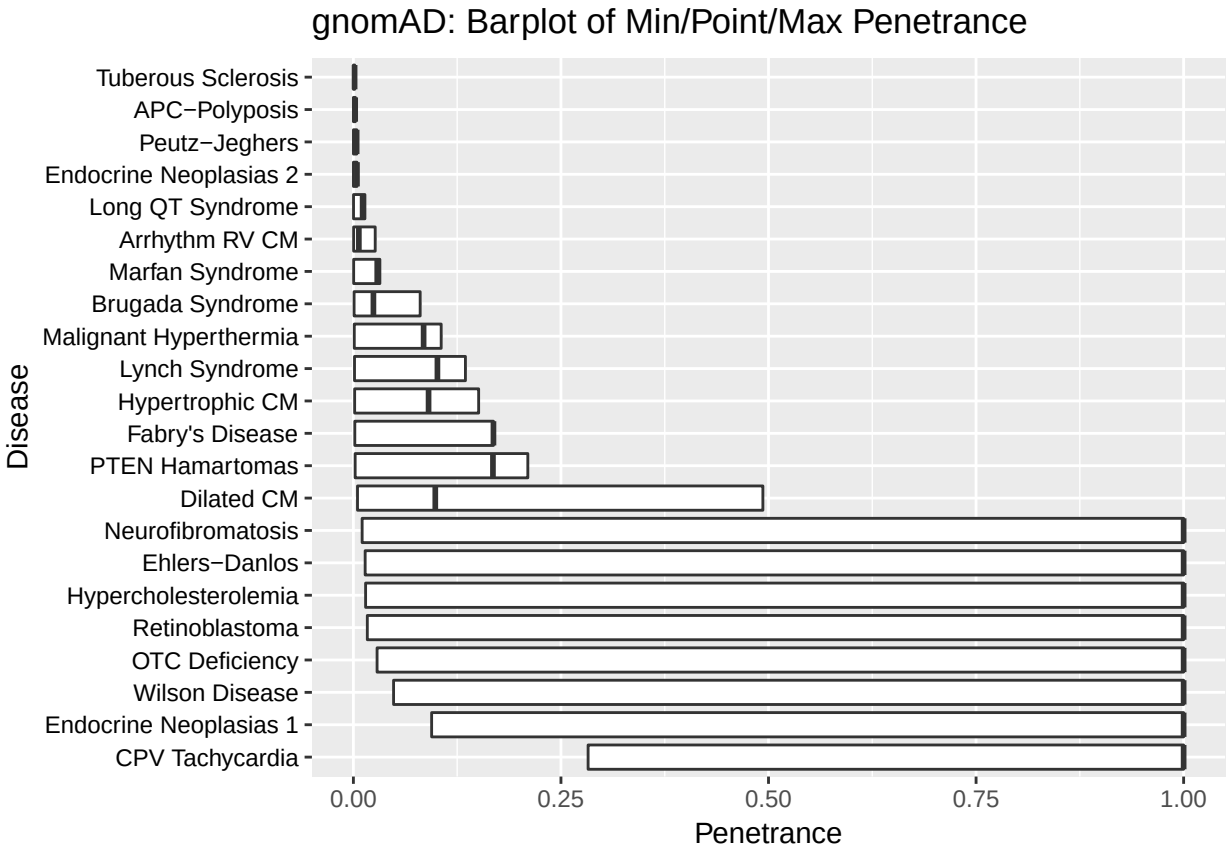




3 Penetrance Estimates

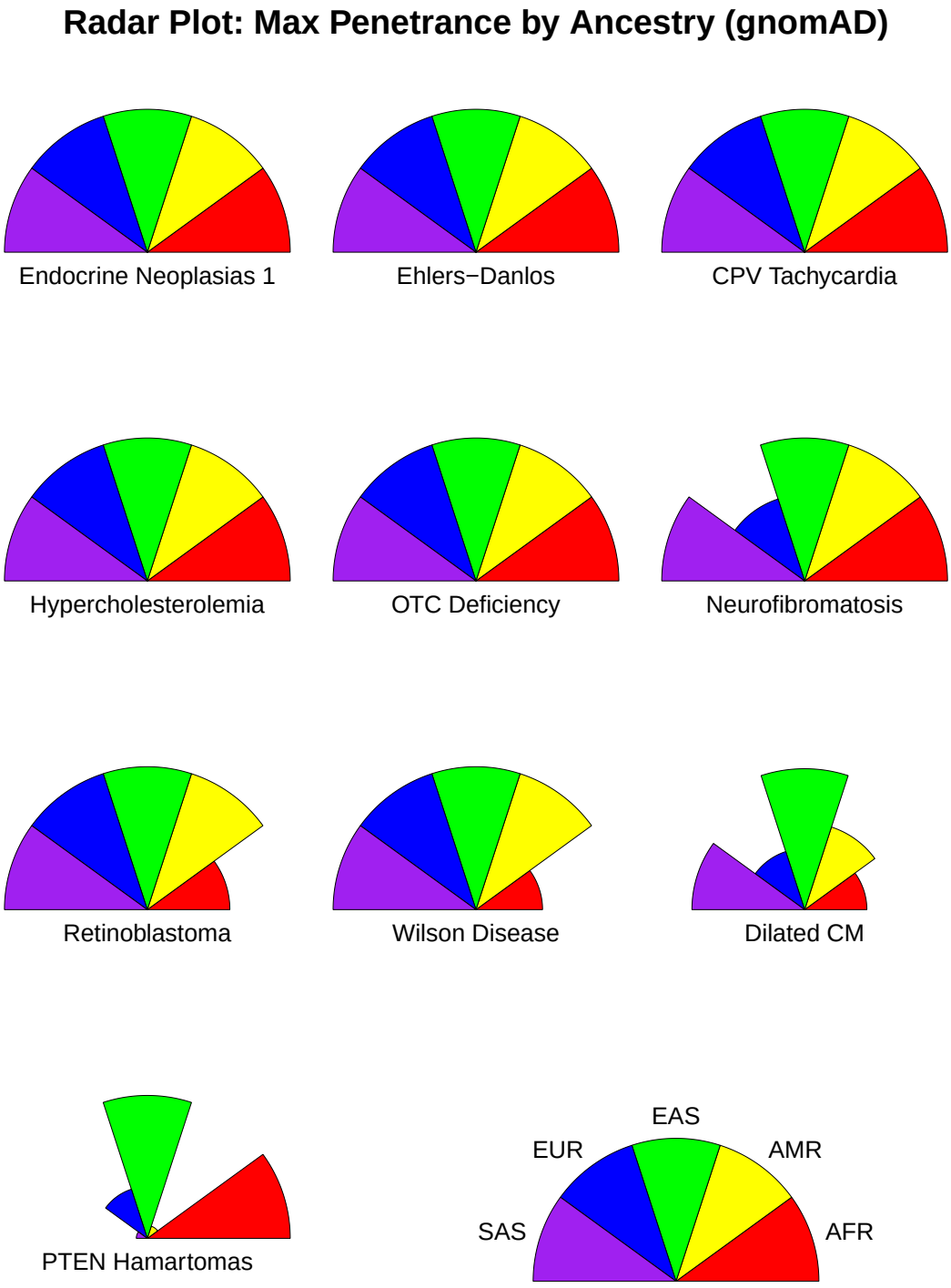
3.1 Max/Min Penetrance as a Function of P(D) and P(V|D)

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.

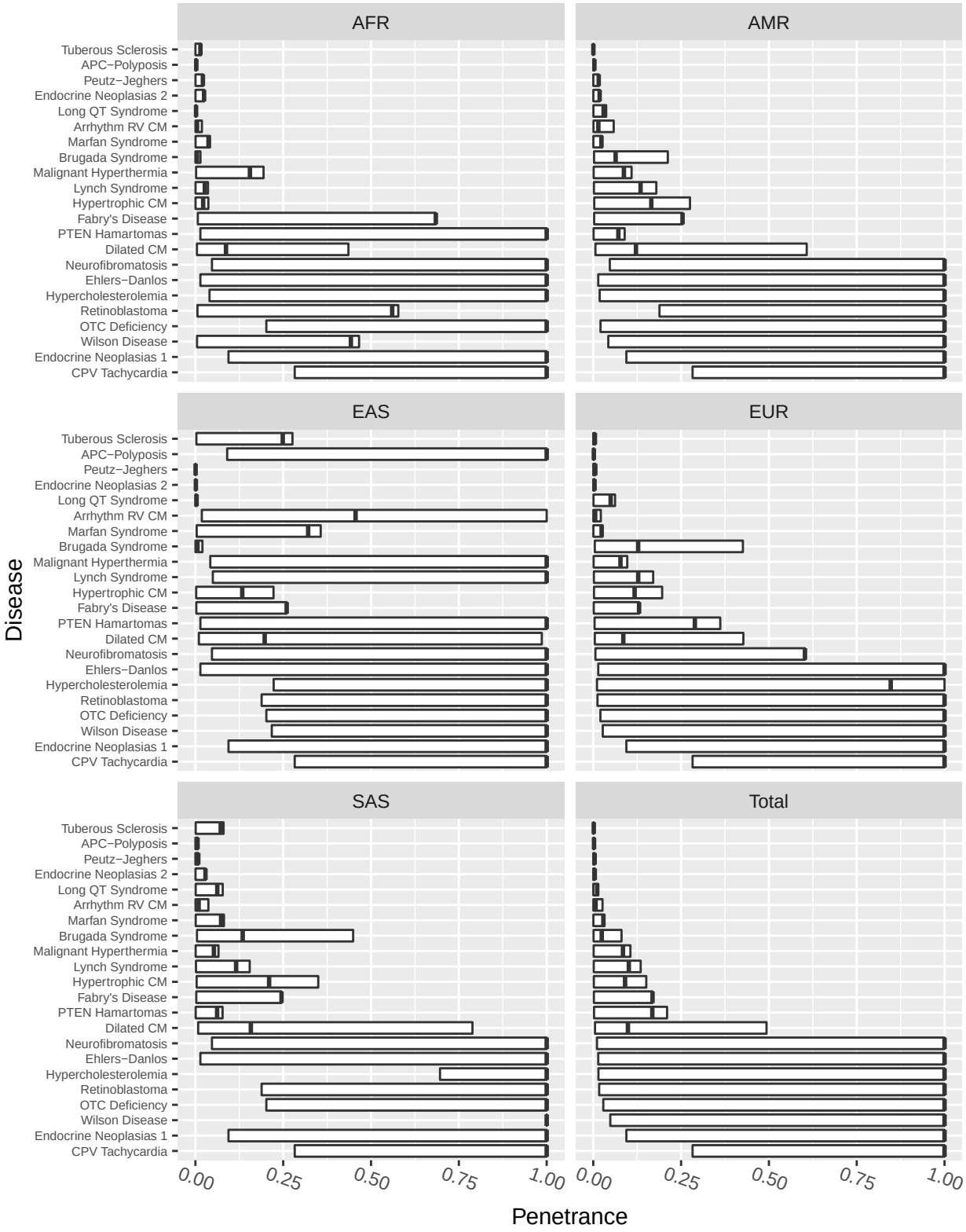


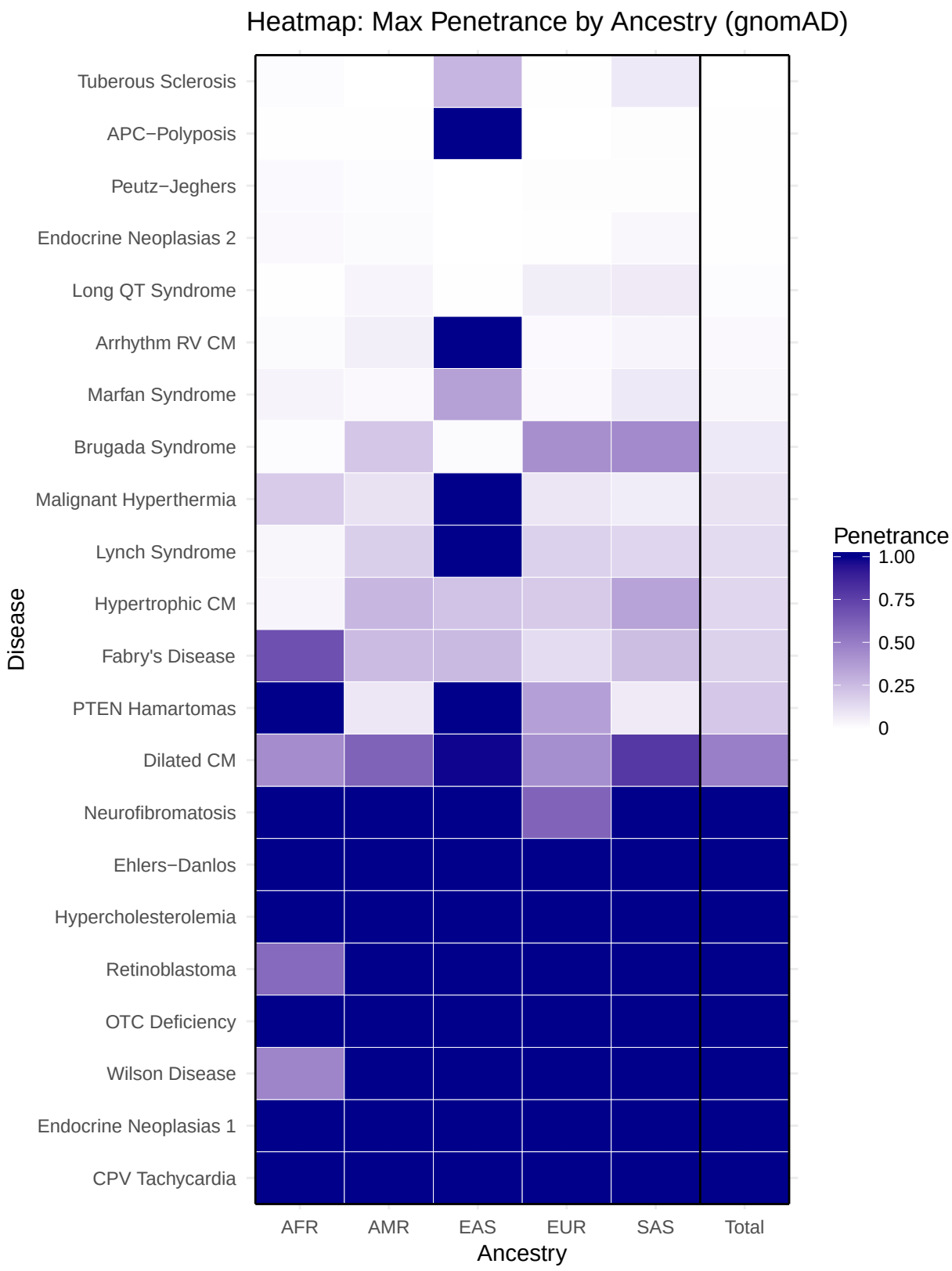
Note: Some diseases have mean theoretical penetrance = 1 because the assumed allelic heterogeneity is greater than is possible, given the observed prevalence and allele frequencies.

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



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