ClinVar Report

James Diao

January 8, 2017

Contents

Sourcing ClinVar input from: clinvar_2014-04-01.vcf
Sending output to: Report_2014-04-01.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	69776
LP/P	16381
ACMG LP/P	2347
ACMG LP/P in gnomAD	481
ACMG LP/P in ExAC	335
ACMG LP/P in 1000 Genomes	84

`## Breakdown of ACMG-gnomAD Variants`

Subset_gnomAD	Number_of_Variants
ACMG in gnomAD	96742
ClinVar-ACMG in gnomAD	4178
LP/P-ACMG in gnomAD	481

`## Breakdown of ACMG-ExAC Variants`

Subset_gnomAD	Number_of_Variants
ACMG in ExAC	59883
ClinVar-ACMG in ExAC	3428
LP/P-ACMG in ExAC	335

`## Breakdown of ACMG-1000G Variants`

Subset_gnomAD	Number_of_Variants
ACMG in 1000G	141466
ClinVar-ACMG in 1000G	1666
LP/P-ACMG in 1000G	84

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

ACMG−59 Pathogenic: Fraction in 1000 Genomes

ACMG−59 Pathogenic: Fraction in gnomAD
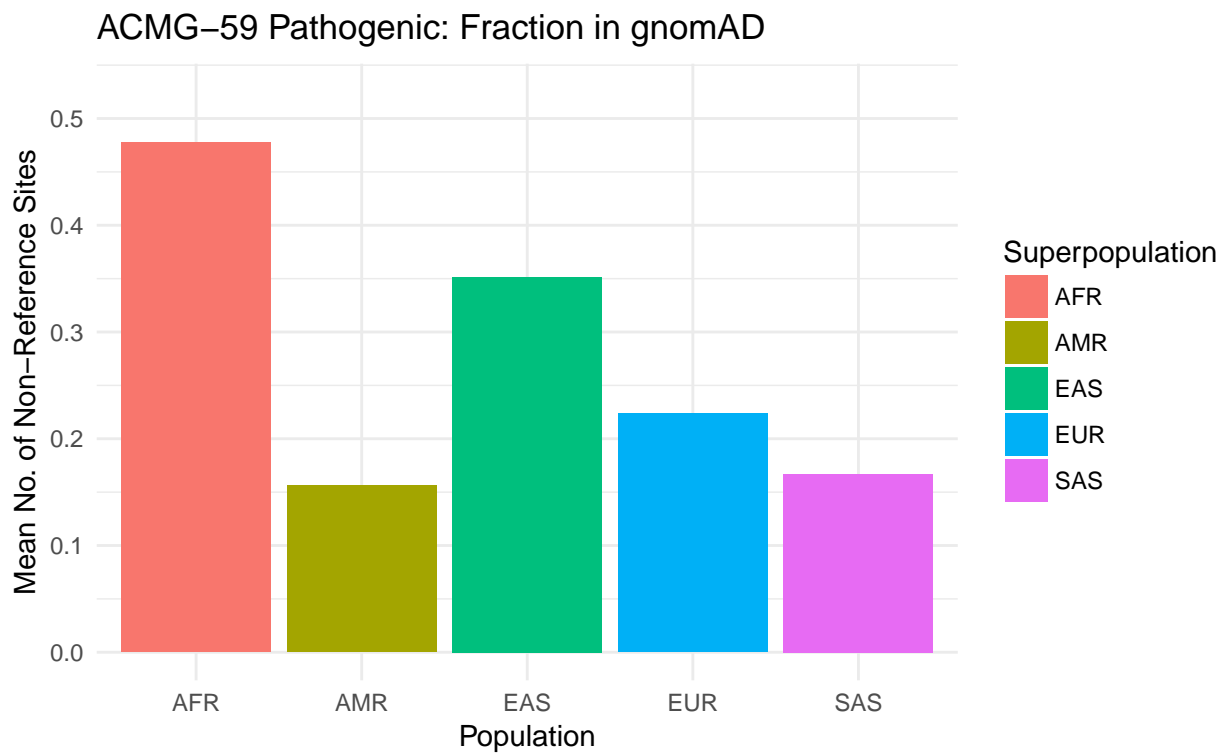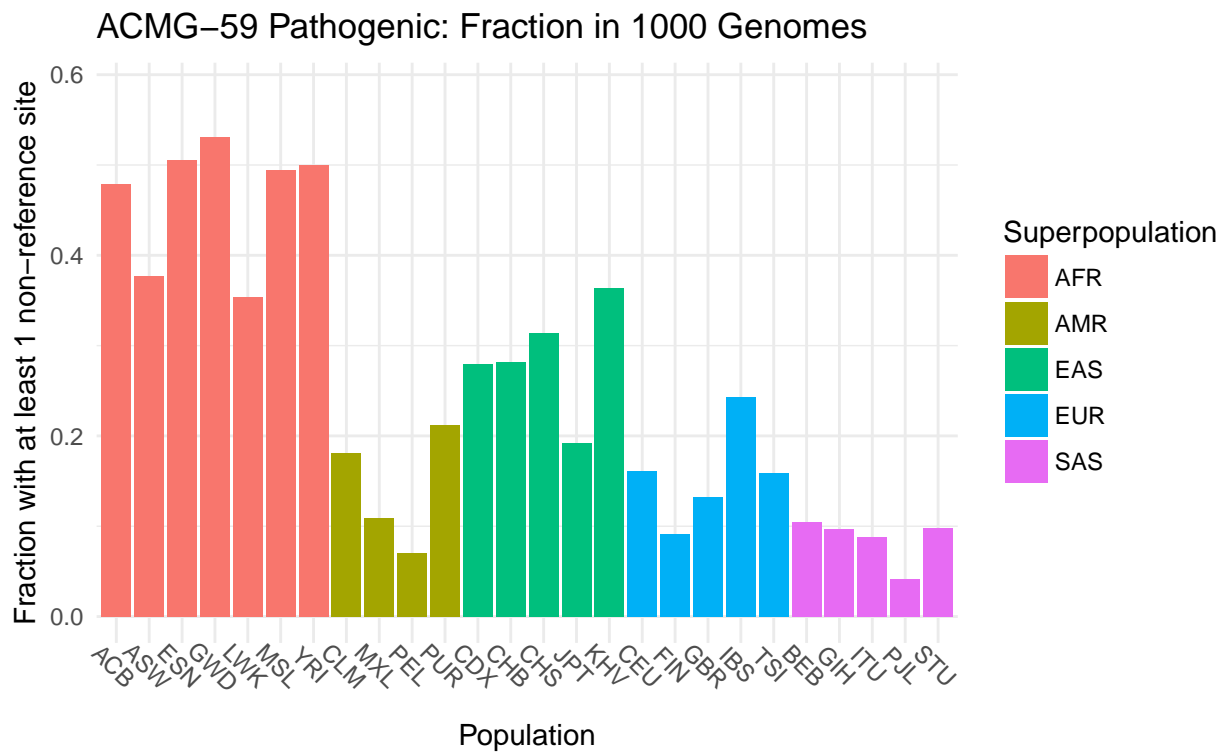
3 Penetrance Estimates
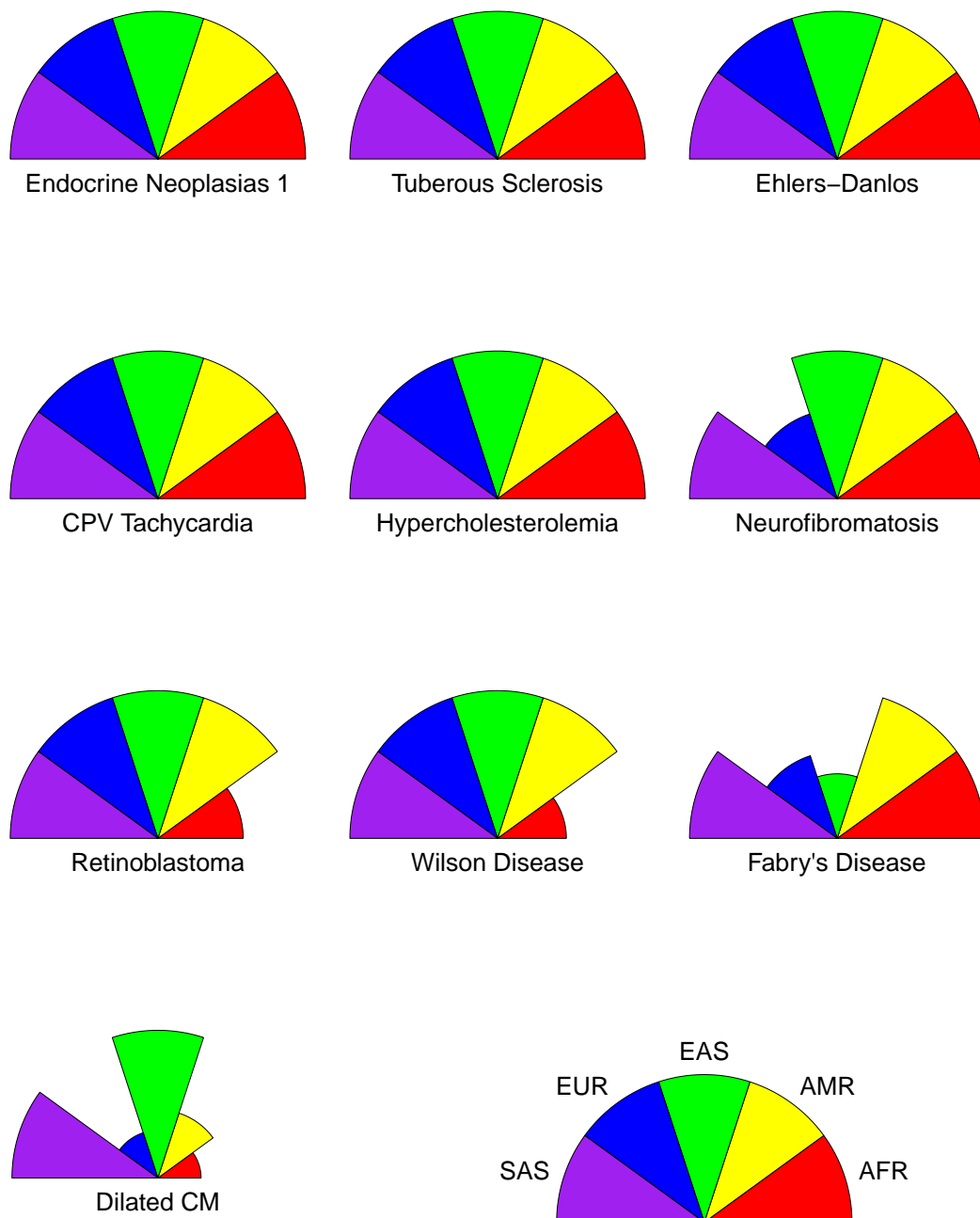
3.1 Max/Min Penetrance as a Function of P(D) and P(V|D)

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.

gnomAD: Barplot of Min/Point/Max Penetrance
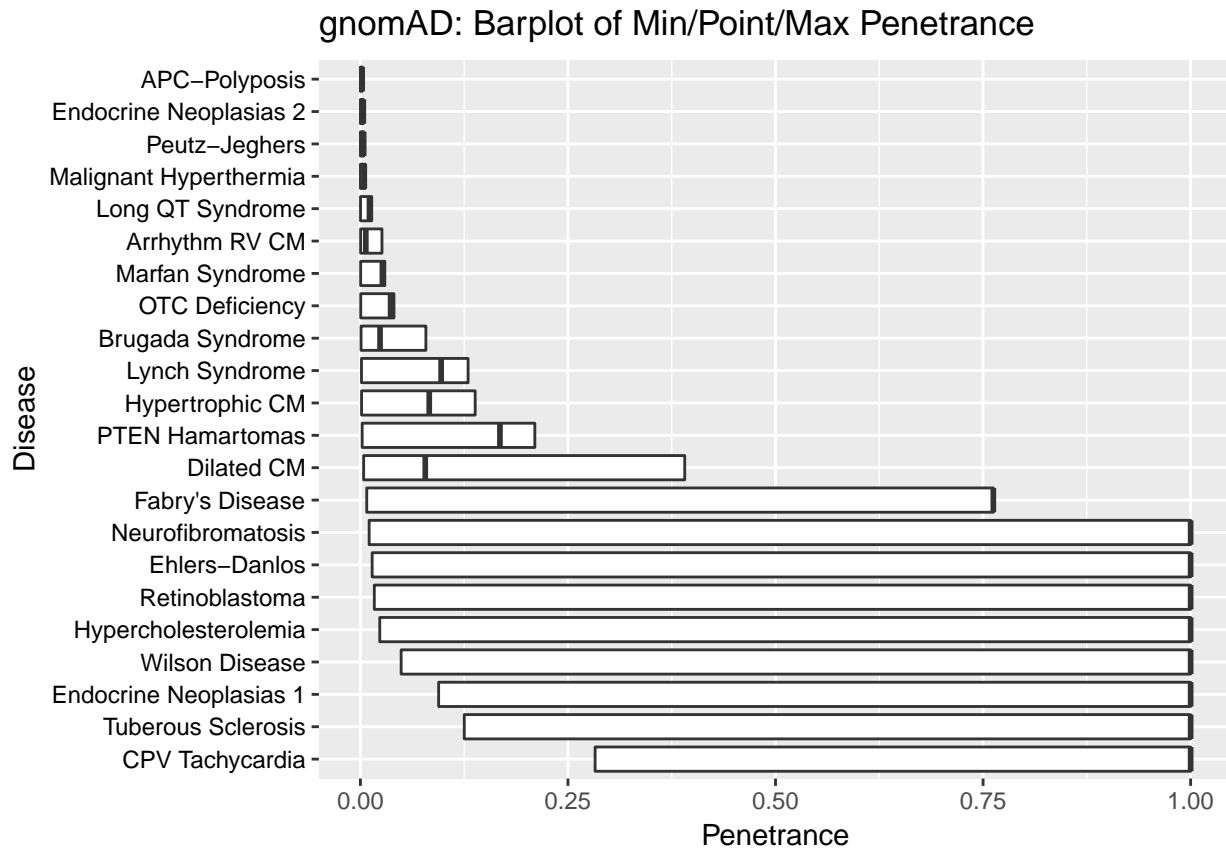
Note: Some diseases have mean theoretical penetrance = 1 because the assumed allelic heterogeneity is greater than is possible, given the observed prevalence and allele frequencies.

3.2 Penetrance Estimates by Ancestry

Radar Plot: Max Penetrance by Ancestry (gnomAD)

Endocrine Neoplasias 1

Tuberous Sclerosis

Ehlers–Danlos

CPV Tachycardia

Hypercholesterolemia

Neurofibromatosis

Retinoblastoma

Wilson Disease

Fabry's Disease
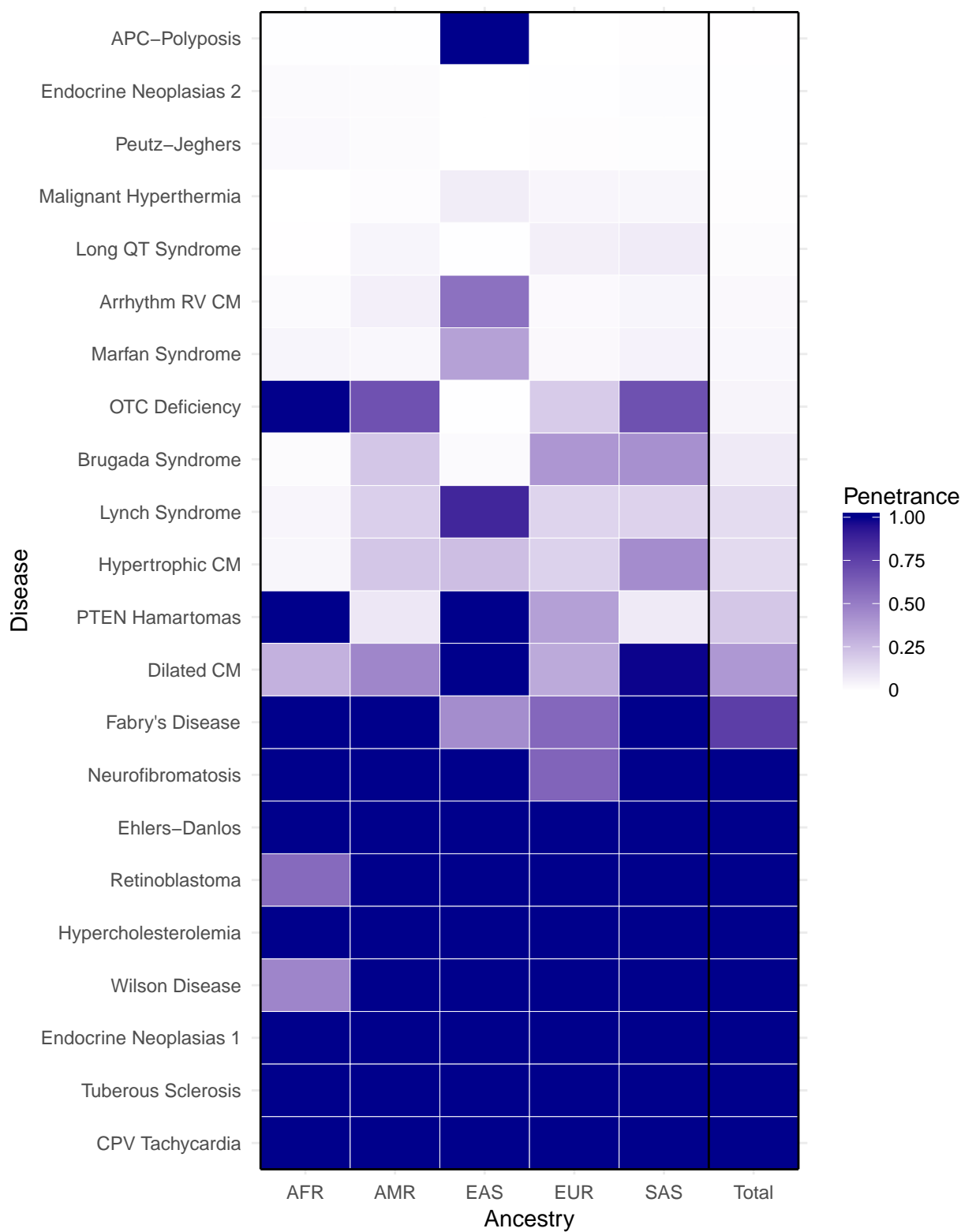
Dilated CM

EAS

EUR

AMR

SAS

AFR

```
## [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)
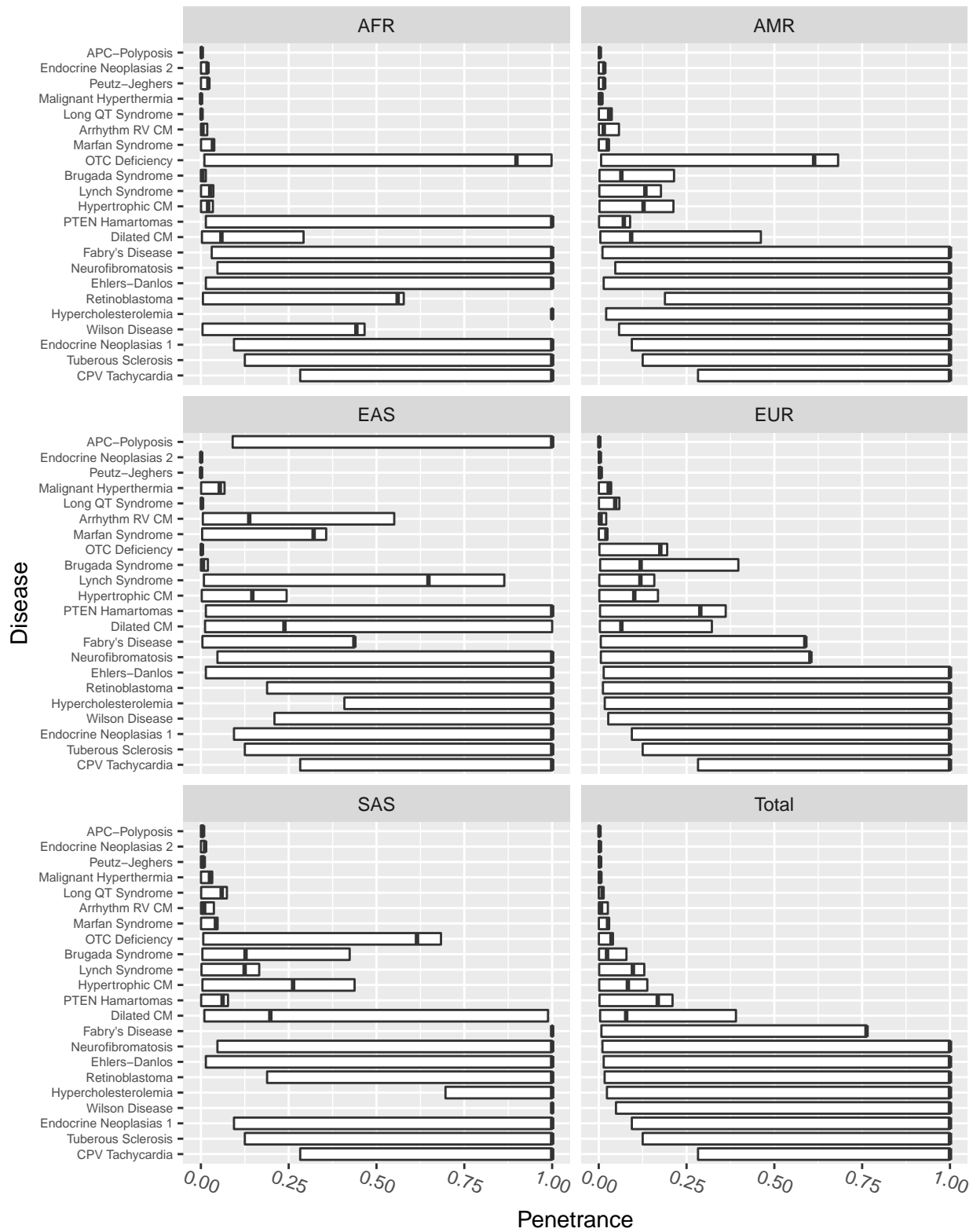
Heatmap: Max Penetrance by Ancestry (gnomAD)

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