ClinVar Report

James Diao

November 4, 2016

Contents

Sourcing ClinVar input from: clinvar_2016-04-05.vcf
Sending output to: Report_2016-04-05.pdf

1 Collect and Merge ClinVar Data

1.1 Import ClinVar VCF

```
## Processed ClinVar data frame 100642 x 14 (selected rows/columns):
```

1.2 Merge ClinVar with 1000 Genomes and ExAC

```
## Breakdown of ClinVar Variants
```

Subset_ClinVar	Number_of_Variants
Total ClinVar	100642
LP/P-ClinVar	31042
LP/P-ClinVar & ACMG	6289
LP/P-ClinVar & ACMG & ExAC	955
LP/P-ClinVar & ACMG & 1000 Genomes	168

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

Subset_1000_Genomes	Number_of_Variants
Total 1000_Genomes & ACMG	139335
1000_Genomes & ACMG & ClinVar	4050
1000_Genomes & ACMG & LP/P-ClinVar	168

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

Subset_ExAC	Number_of_Variants
Total ExAC & ACMG	58873
ExAC & ACMG & ClinVar	8091
ExAC & ACMG & LP/P-ClinVar	955

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

ACMG−56 Pathogenic: Fraction

ACMG−56 Pathogenic: Mean in ExAC
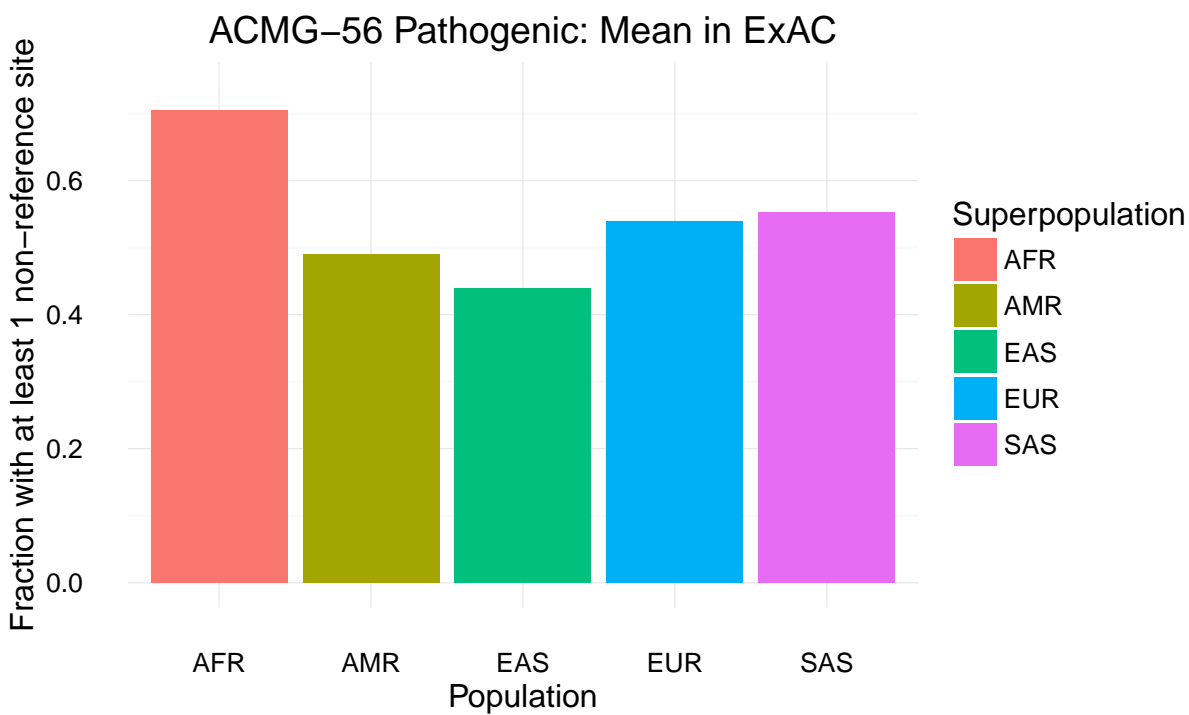
3 Penetrance Estimates

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

The left end of the boxplot indicates P(D) AND P(V|D) = lower value,
the bold line in the middle indicates P(D) AND P(V|D) = geometric_mean(values),
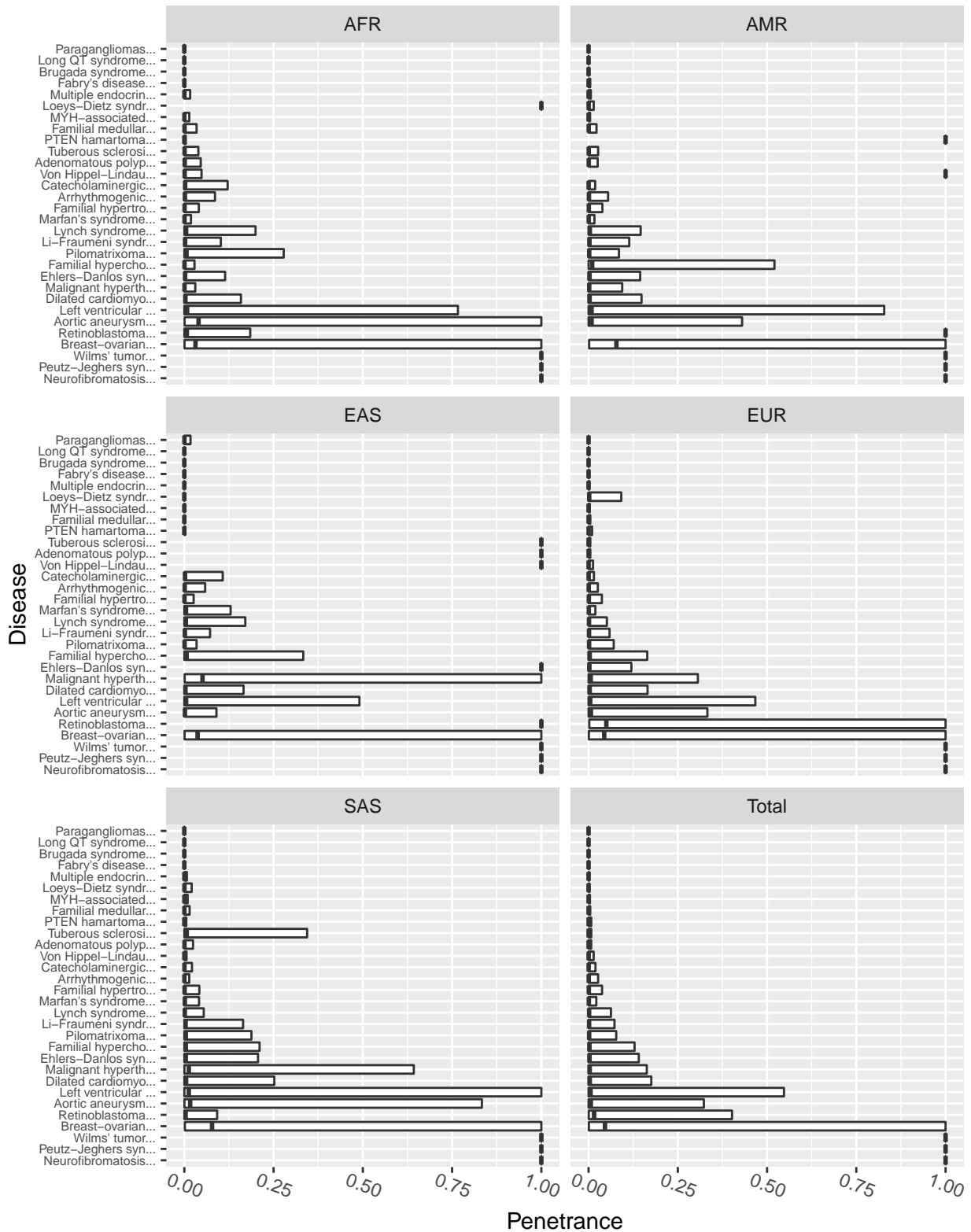the right end of the boxplot indicates P(D) AND P(V|D) = upper value.
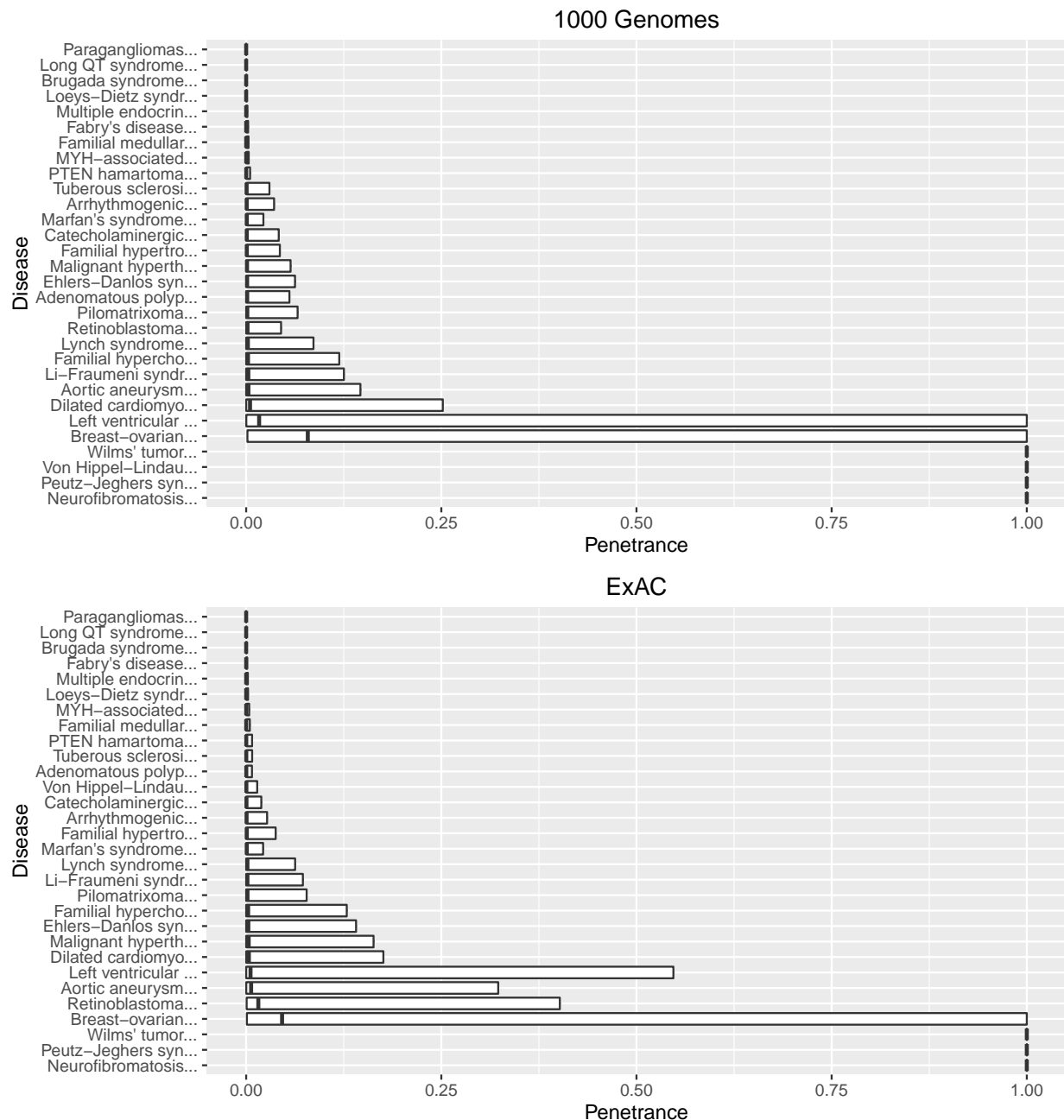
Note: Prevalence ranges of 5x were assumed for all point estimates of prevalence.
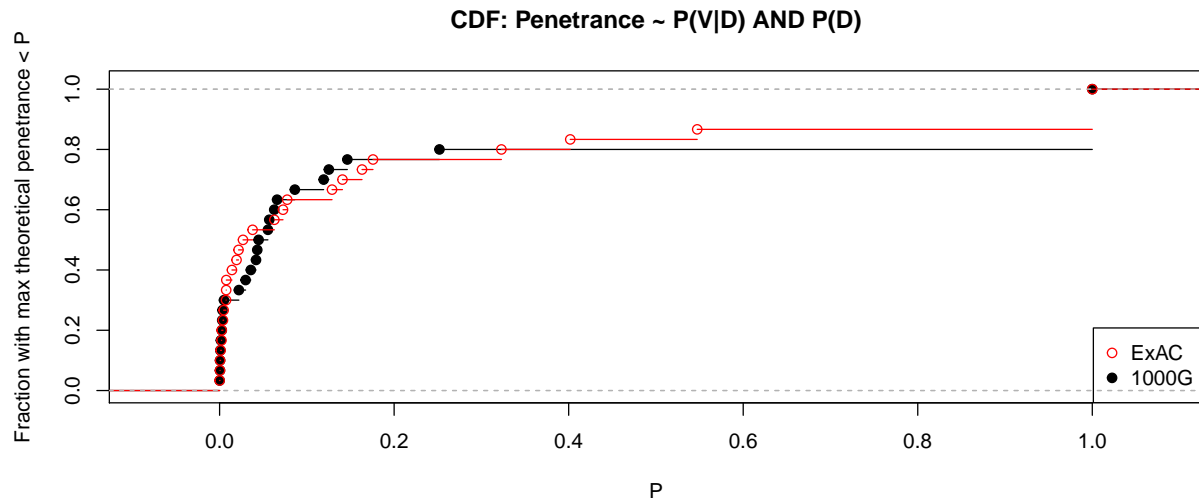For example: a point estimate of 0.022 would be given the range 0.01-0.05.
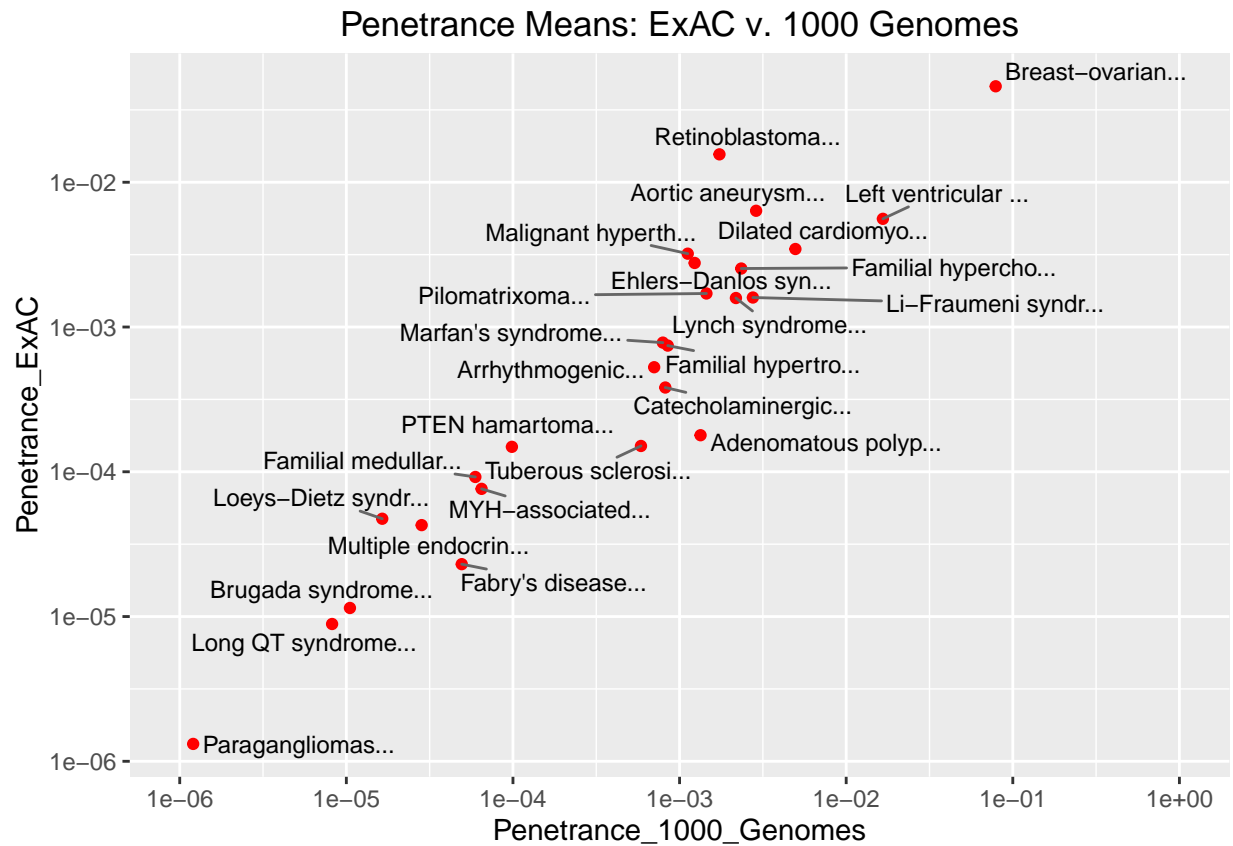
3.2   Penetrance Estimates by Ancestry

Penetrance by Ancestry (1000 Genomes)
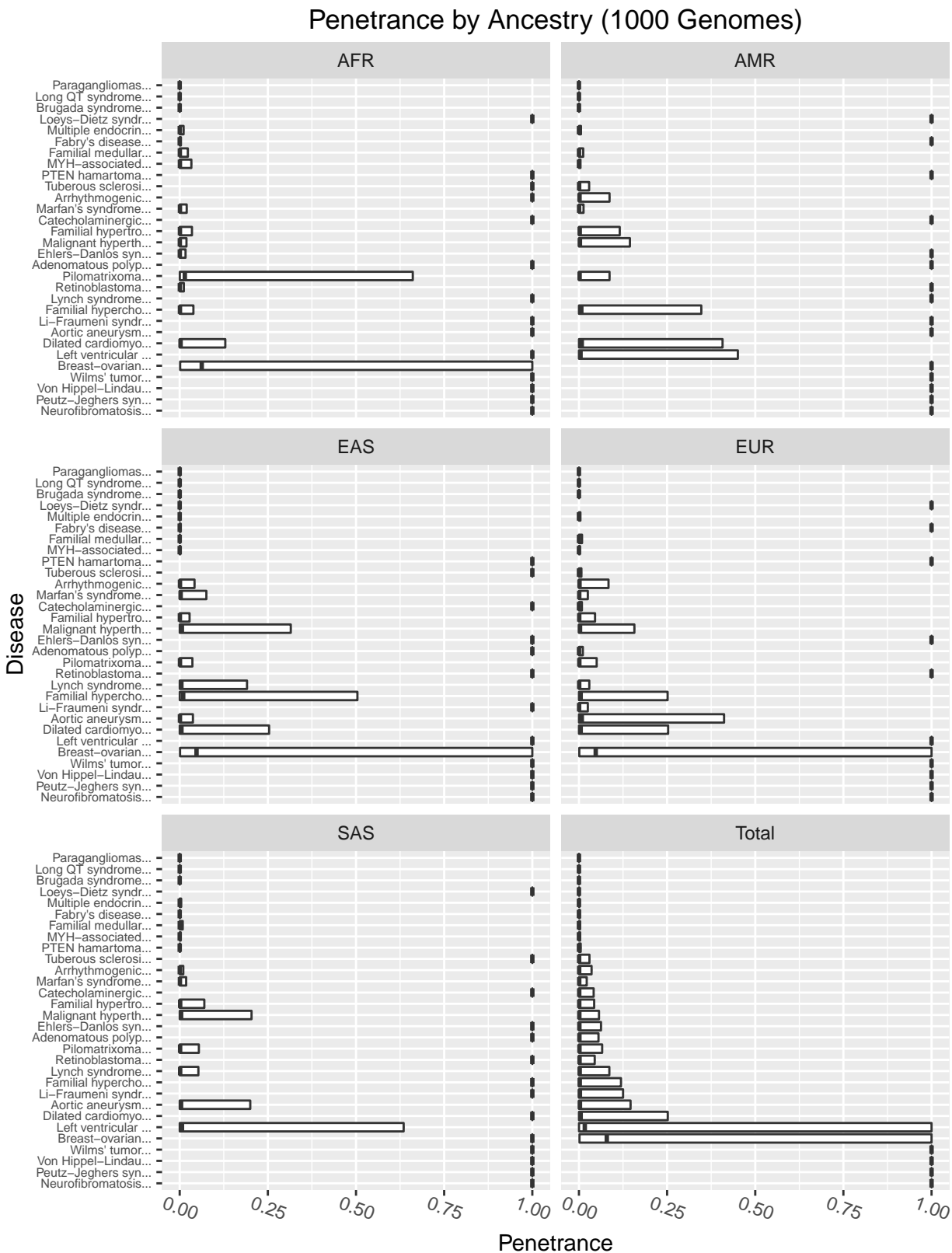
Penetrance by Ancestry (ExAC)

3.3 Empirical CDFs for All Penetrance Plots

CDF: Penetrance ~ P(V|D) AND P(D)

3.4 Comparing Mean Penetrance between ExAC and 1000 Genomes

Penetrance Means: ExAC v. 1000 Genomes

The Pearson correlation is 0.94.
Max penetrance values computed using 1000 Genomes are 1.6-fold larger than those computed using ExAC.