ClinVar Report

James Diao

November 4, 2016

Contents

Sourcing ClinVar input from: clinvar_2015-08-04.vcf
Sending output to: Report_2015-08-04.pdf

1 Collect and Merge ClinVar Data

1.1 Import ClinVar VCF

`## Processed ClinVar data frame 84688 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	84688
LP/P-ClinVar	27032
LP/P-ClinVar & ACMG	5800
LP/P-ClinVar & ACMG & ExAC	980
LP/P-ClinVar & ACMG & 1000 Genomes	182

`## Breakdown of ACMG-1000 Genomes Variants`

Subset_1000_Genomes	Number_of_Variants
Total 1000_Genomes & ACMG	139335
1000_Genomes & ACMG & ClinVar	3098
1000_Genomes & ACMG & LP/P-ClinVar	182

`## Breakdown of ACMG-ExAC Variants`

Subset_ExAC	Number_of_Variants
Total ExAC & ACMG	58873
ExAC & ACMG & ClinVar	7400
ExAC & ACMG & LP/P-ClinVar	980

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

ACMG−56 Pathogenic: Fraction

ACMG−56 Pathogenic: Mean in ExAC
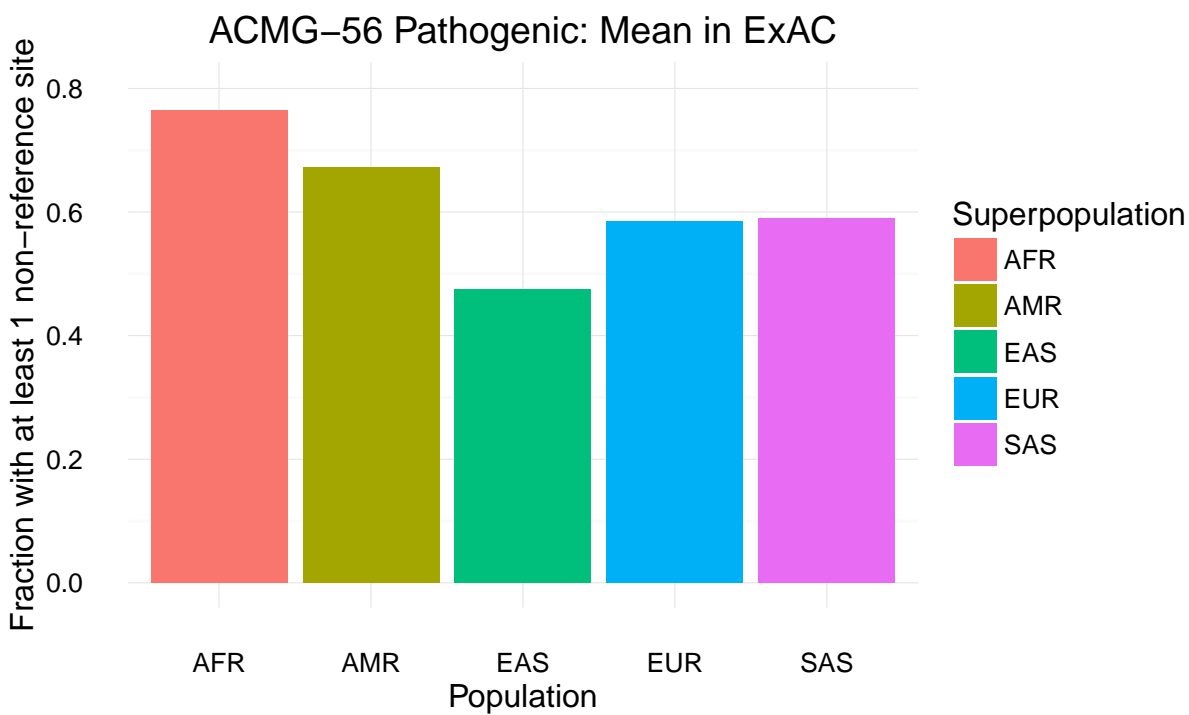
3 Penetrance Estimates

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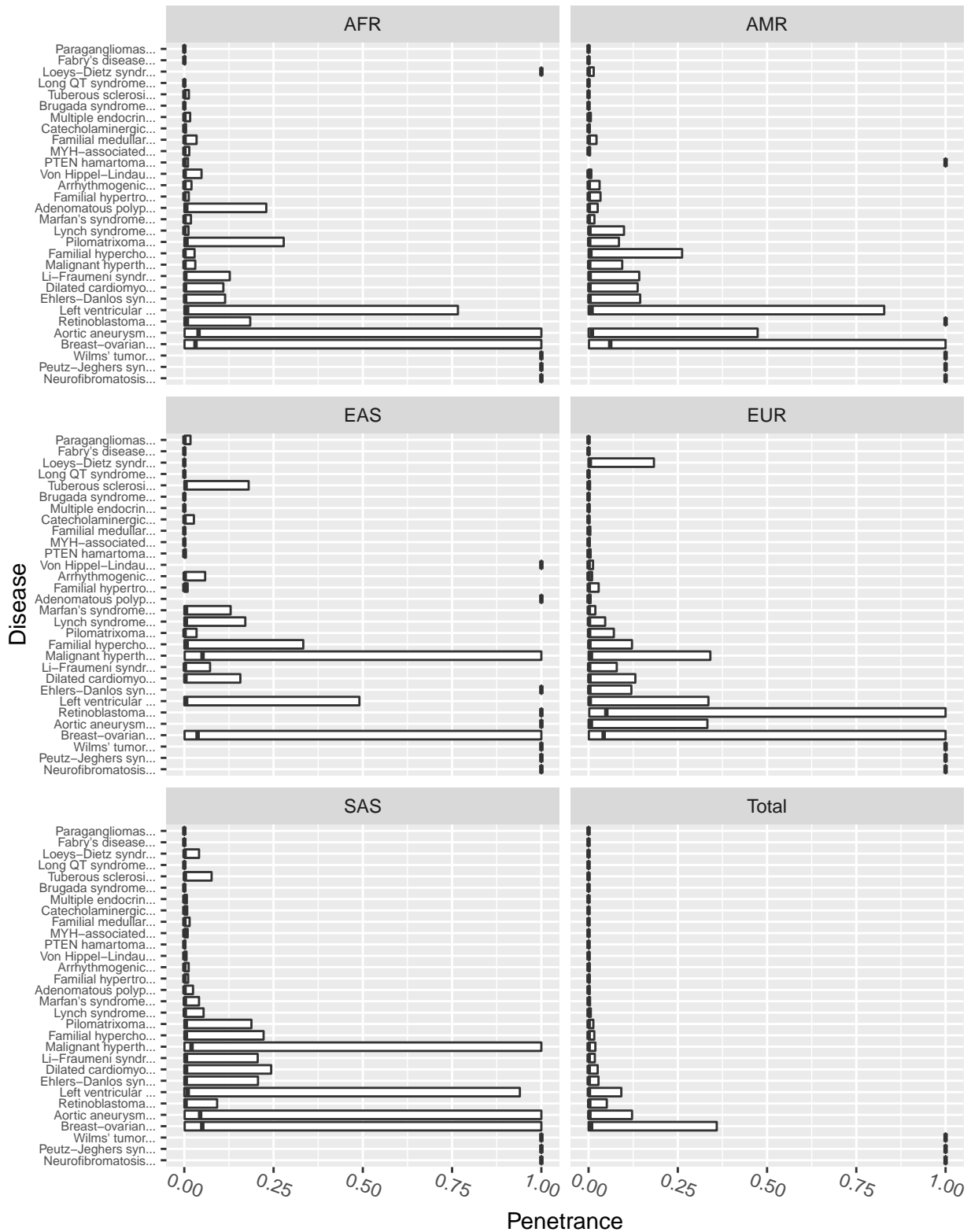the right end of the boxplot indicates P(D) AND P(V|D) = upper value.
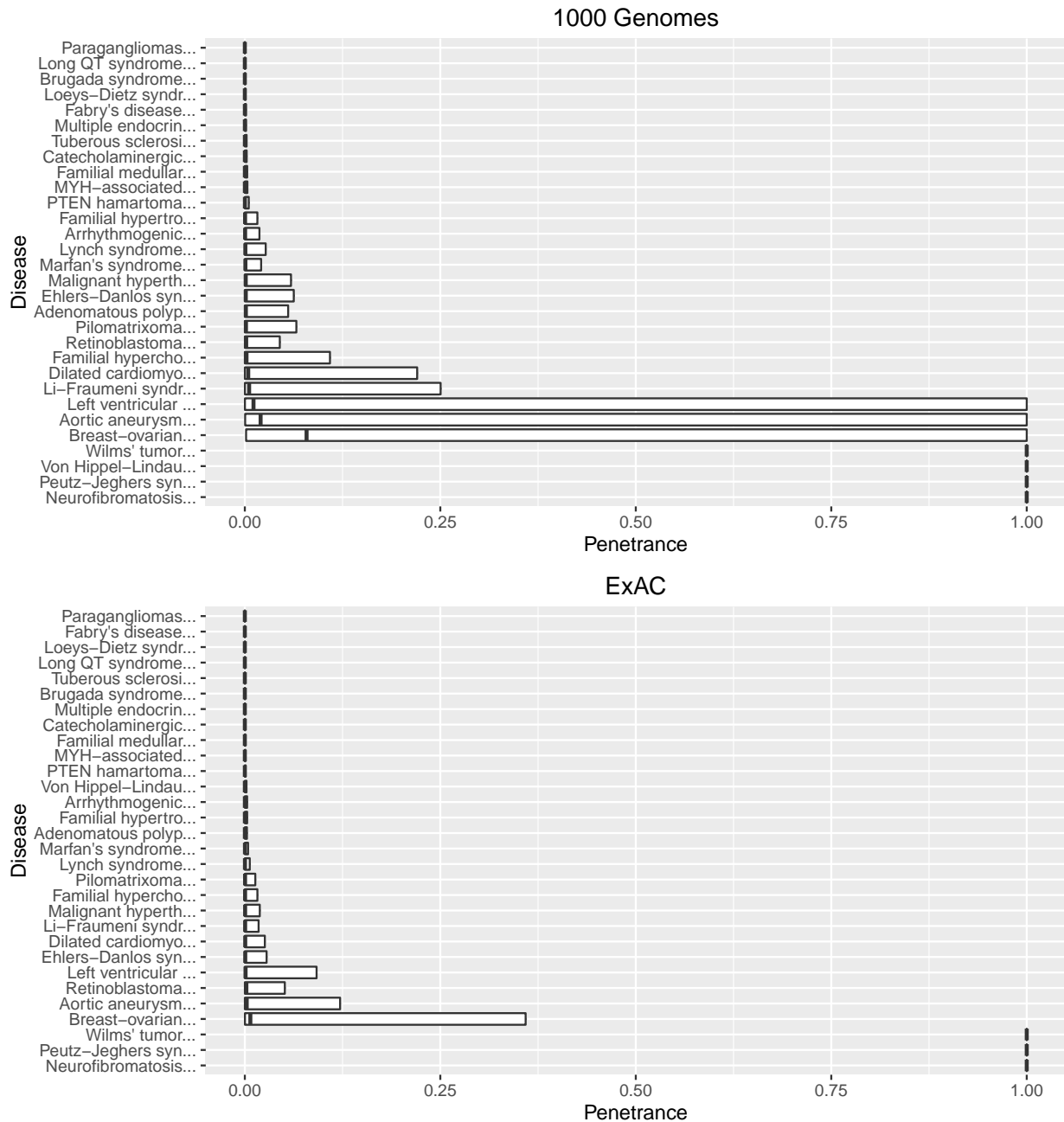
Note: Prevalence ranges of 5x were assumed for all point estimates of prevalence.
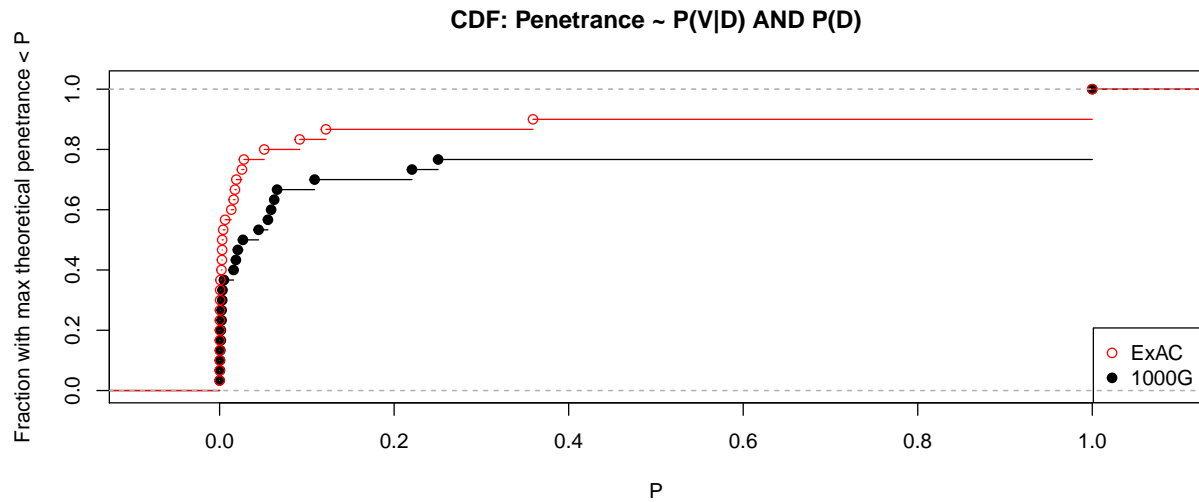For example: a point estimate of 0.022 would be given the range 0.01-0.05.
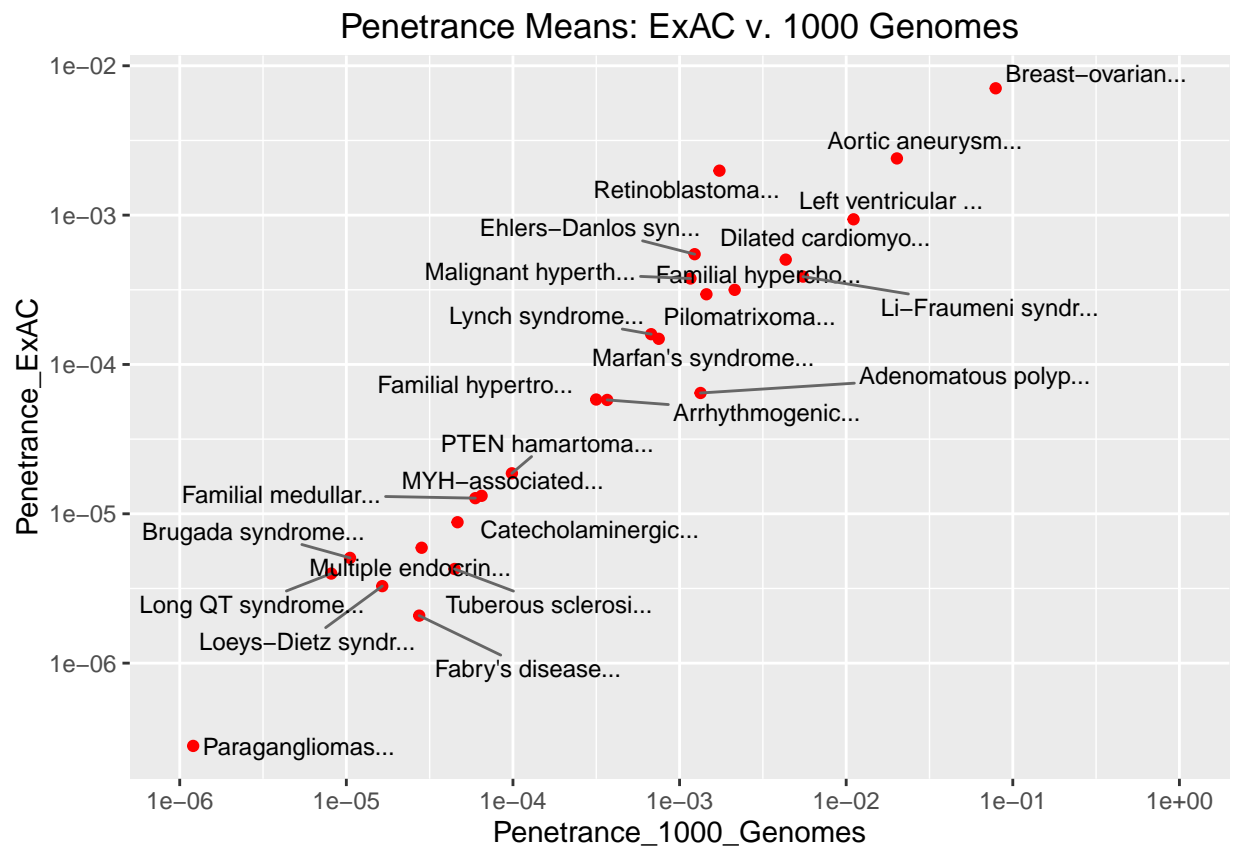
3.2 Penetrance Estimates by Ancestry

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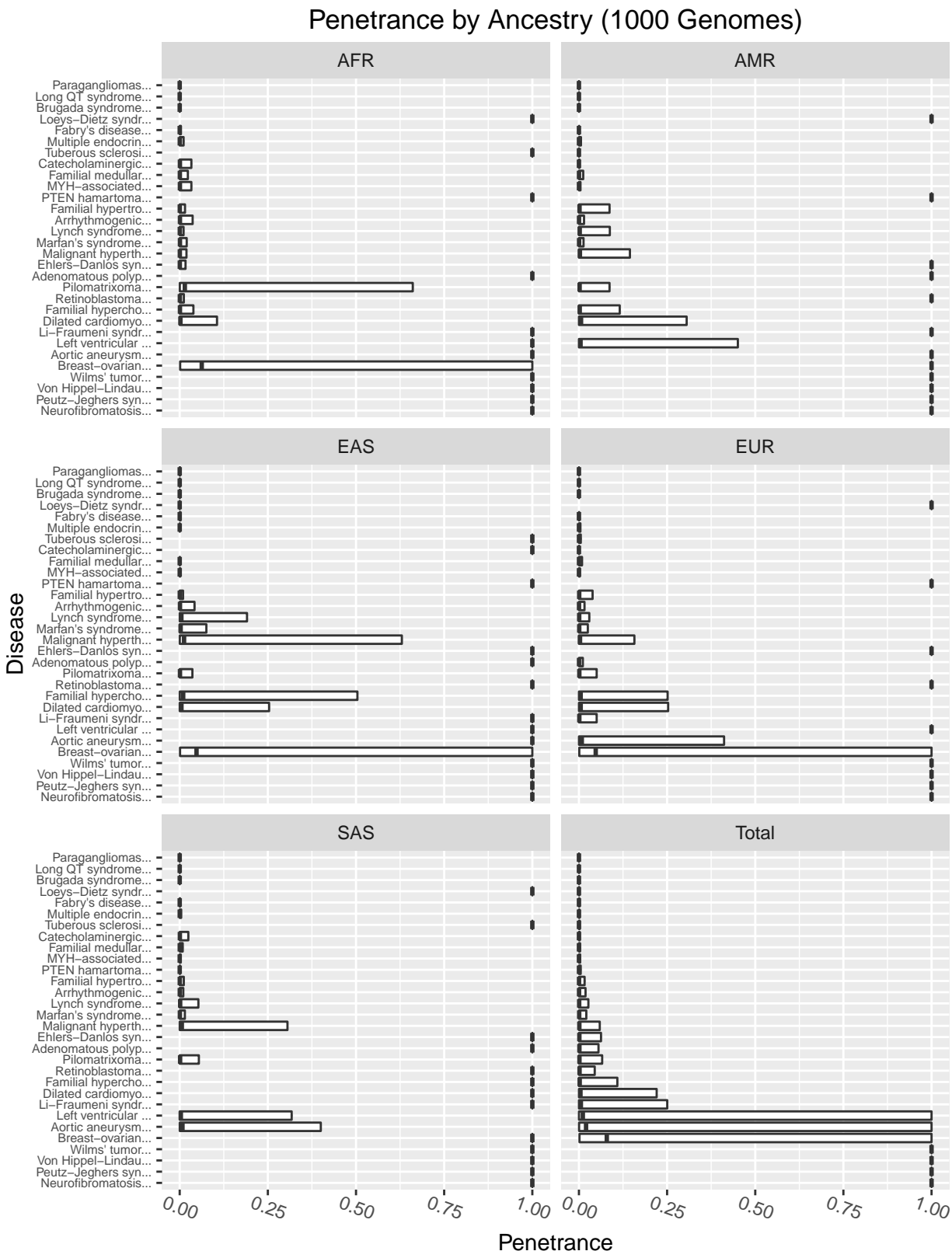
Penetrance by Ancestry (ExAC)

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 0.97.
Max penetrance values computed using 1000 Genomes are 10-fold larger than those computed using ExAC.