# ClinVar Report

*James Diao*

*November 4, 2016*

## Contents



**Sourcing ClinVar input from**: clinvar_2013-02-26.vcf
**Sending output to**: Report_2013-02-26.pdf

# 1 Collect and Merge ClinVar Data

## 1.1 Import ClinVar VCF

```
## Processed ClinVar data frame 66988 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 | 66988 |
| LP/P-ClinVar | 12719 |
| LP/P-ClinVar & ACMG | 924 |
| LP/P-ClinVar & ACMG & ExAC | 208 |
| LP/P-ClinVar & ACMG & 1000 Genomes | 65 |

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

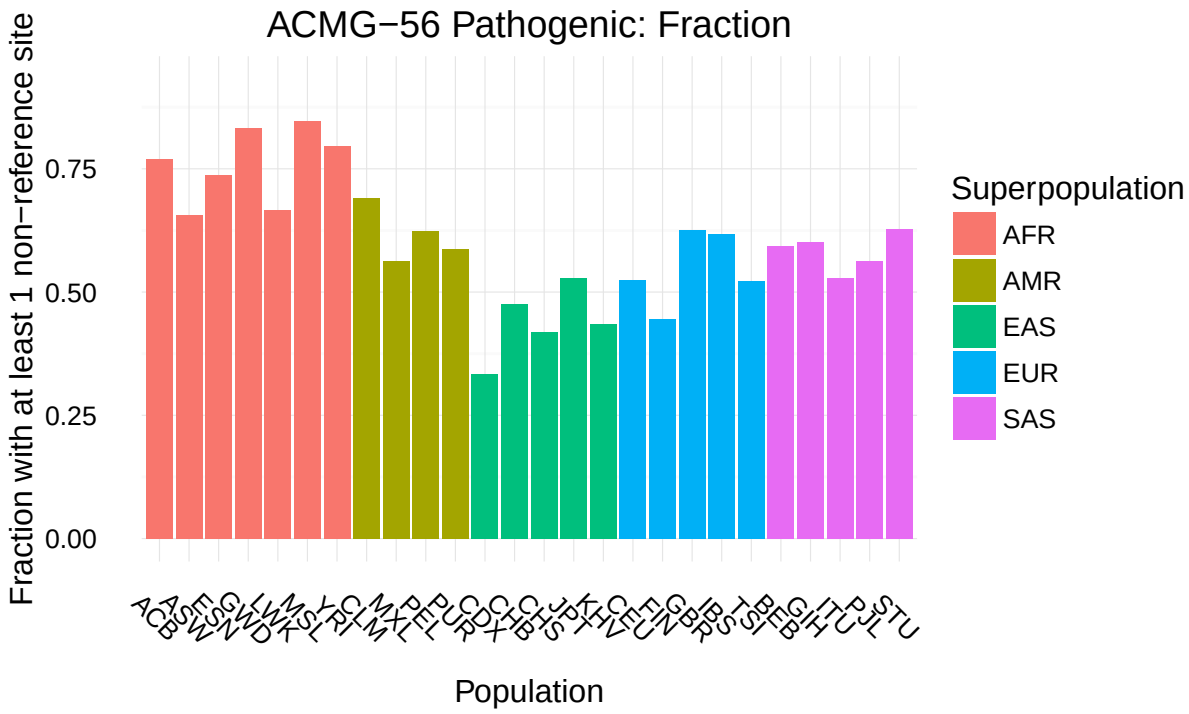
| Subset_1000_Genomes | Number_of_Variants |
|---|---|
| Total 1000_Genomes & ACMG | 139335 |
| 1000_Genomes & ACMG & ClinVar | 1216 |
| 1000_Genomes & ACMG & LP/P-ClinVar | 65 |

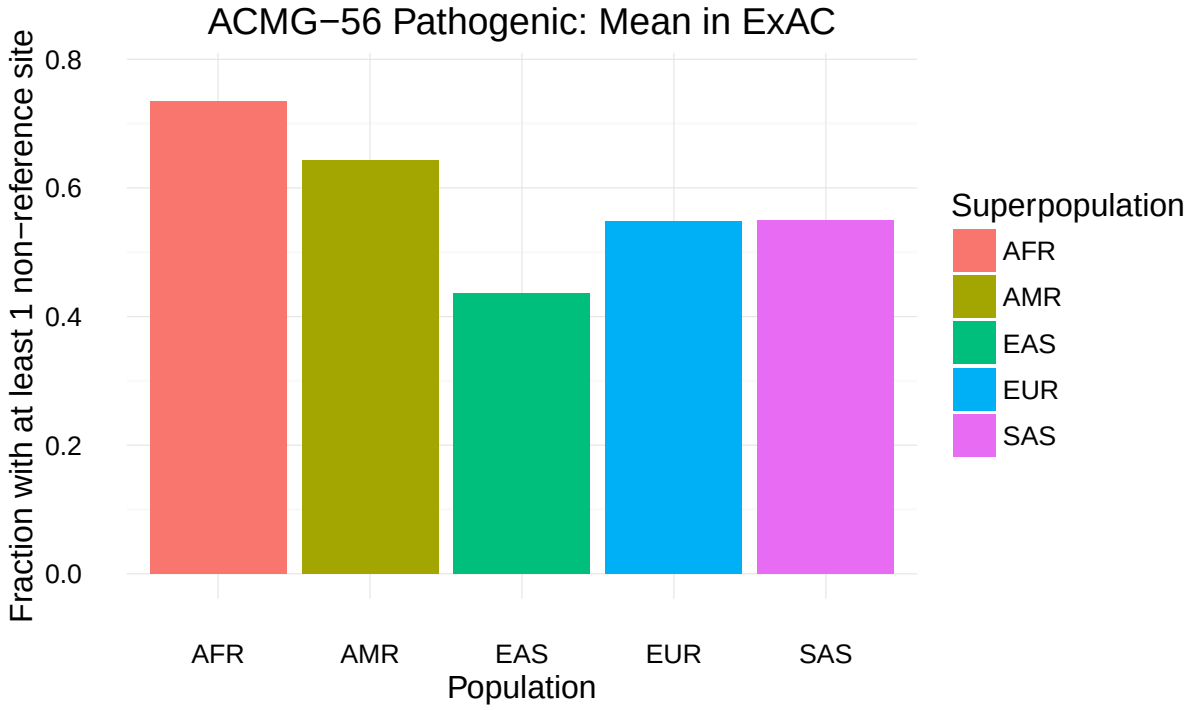
```
## Breakdown of ACMG-ExAC Variants
```

| Subset_ExAC | Number_of_Variants |
|---|---|
| Total ExAC & ACMG | 58873 |
| ExAC & ACMG & ClinVar | 2256 |
| ExAC & ACMG & LP/P-ClinVar | 208 |

# 2 Summary Statistics

## 2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

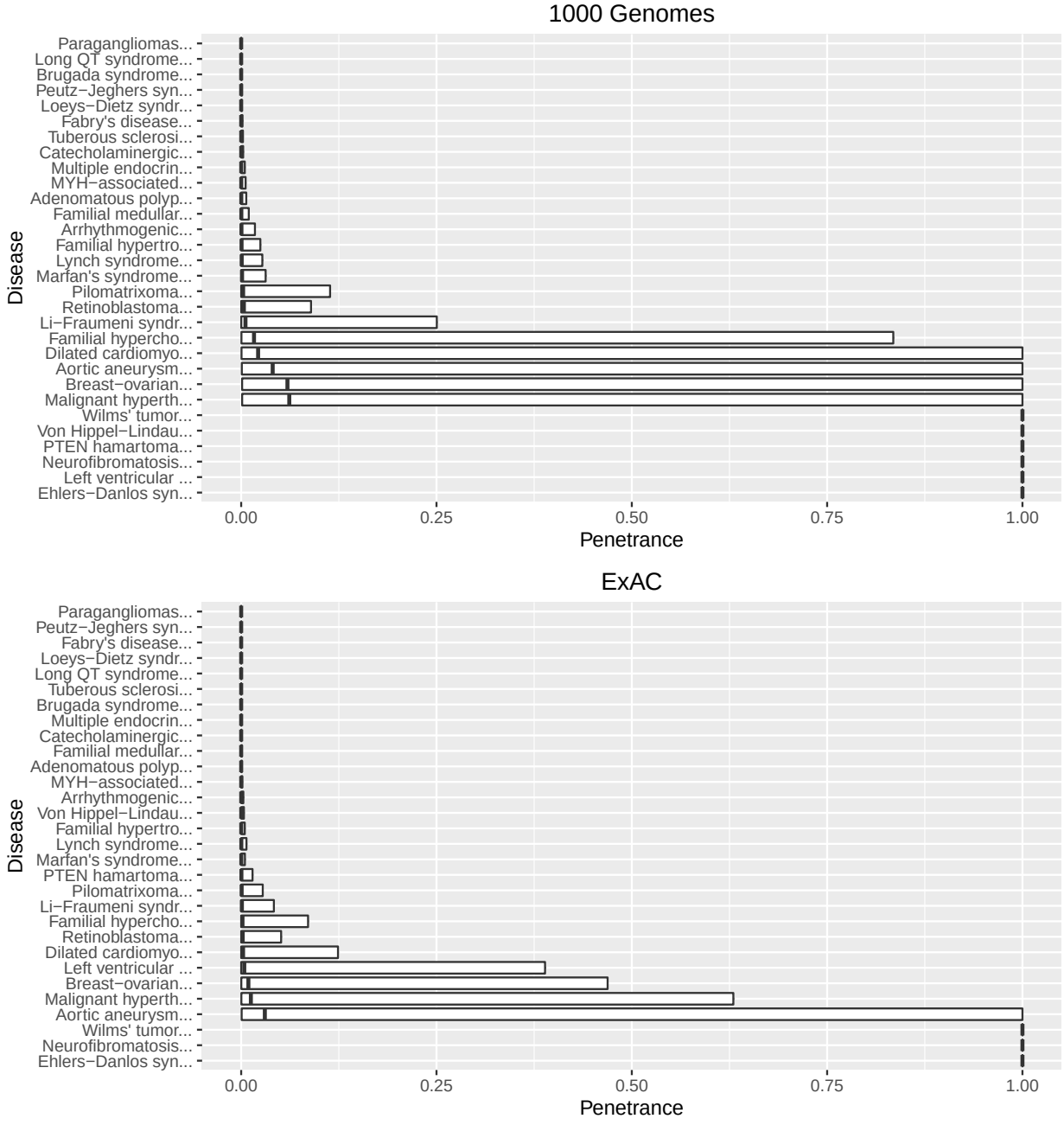


# 3 Penetrance Estimates

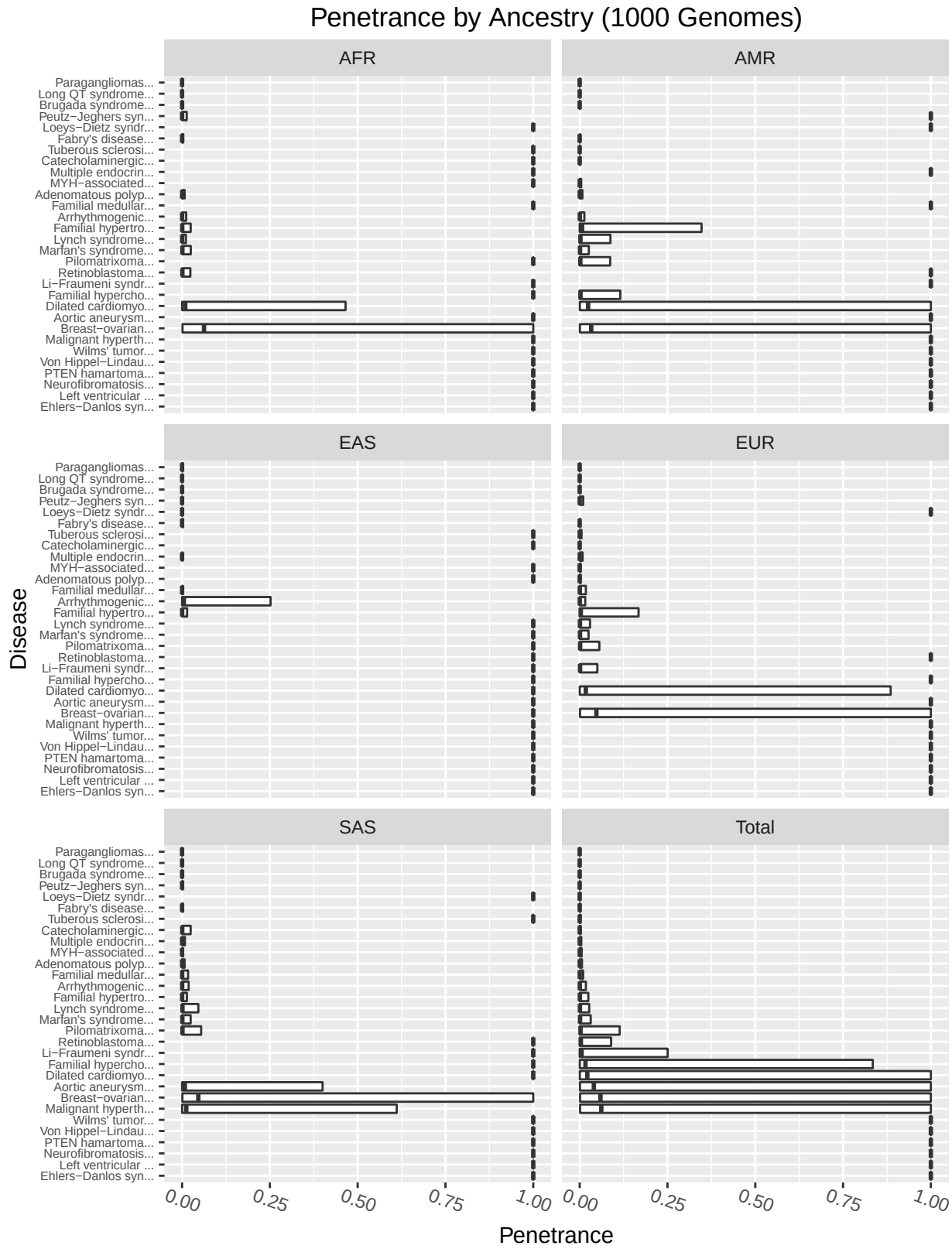
## 3.1 Max/Min Penetrance as a Function of P(D) and P(V|D)

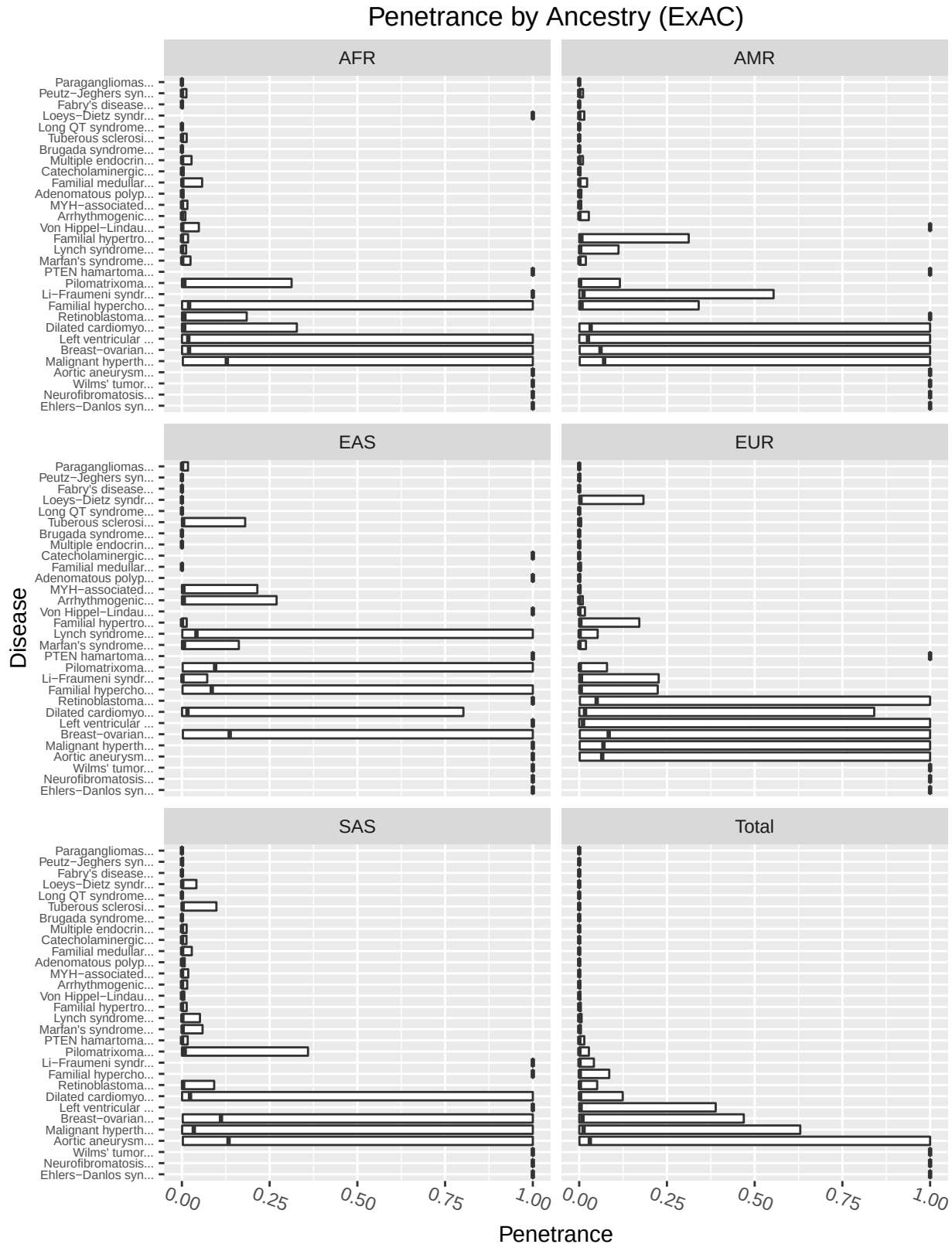
The left end of the boxplot indicates P(D) AND P(V|D) = lower value,
the bold line in the middle indicates P(D) AND P(V|D) = geometric_mean(values),
the right end of the boxplot indicates P(D) AND P(V|D) = upper value.



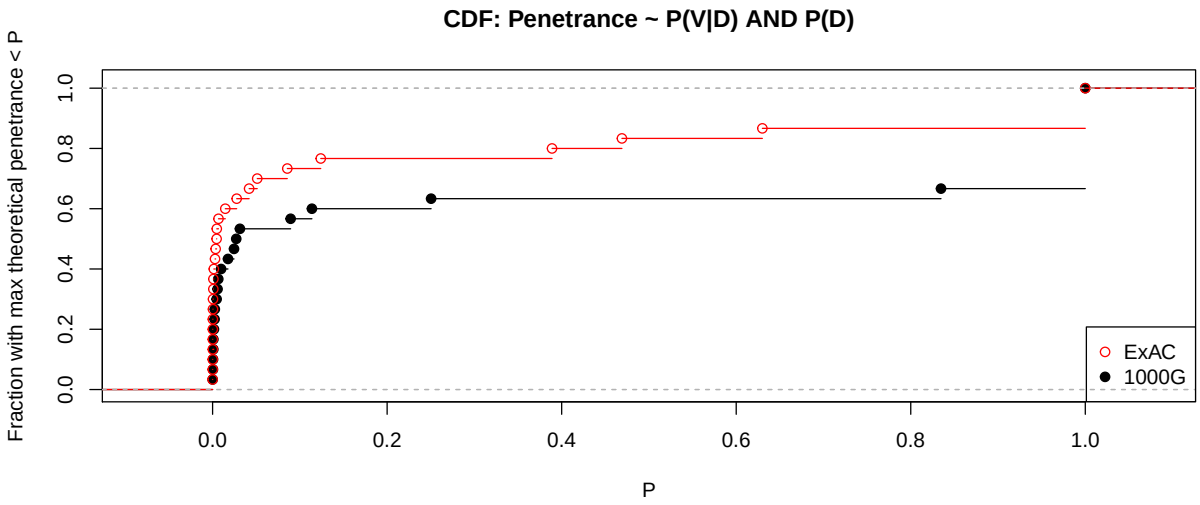

Note: Prevalence ranges of 5x were assumed for all point estimates of prevalence.
For example: a point estimate of 0.022 would be given the range 0.01-0.05.

### 3.2 Penetrance Estimates by Ancestry

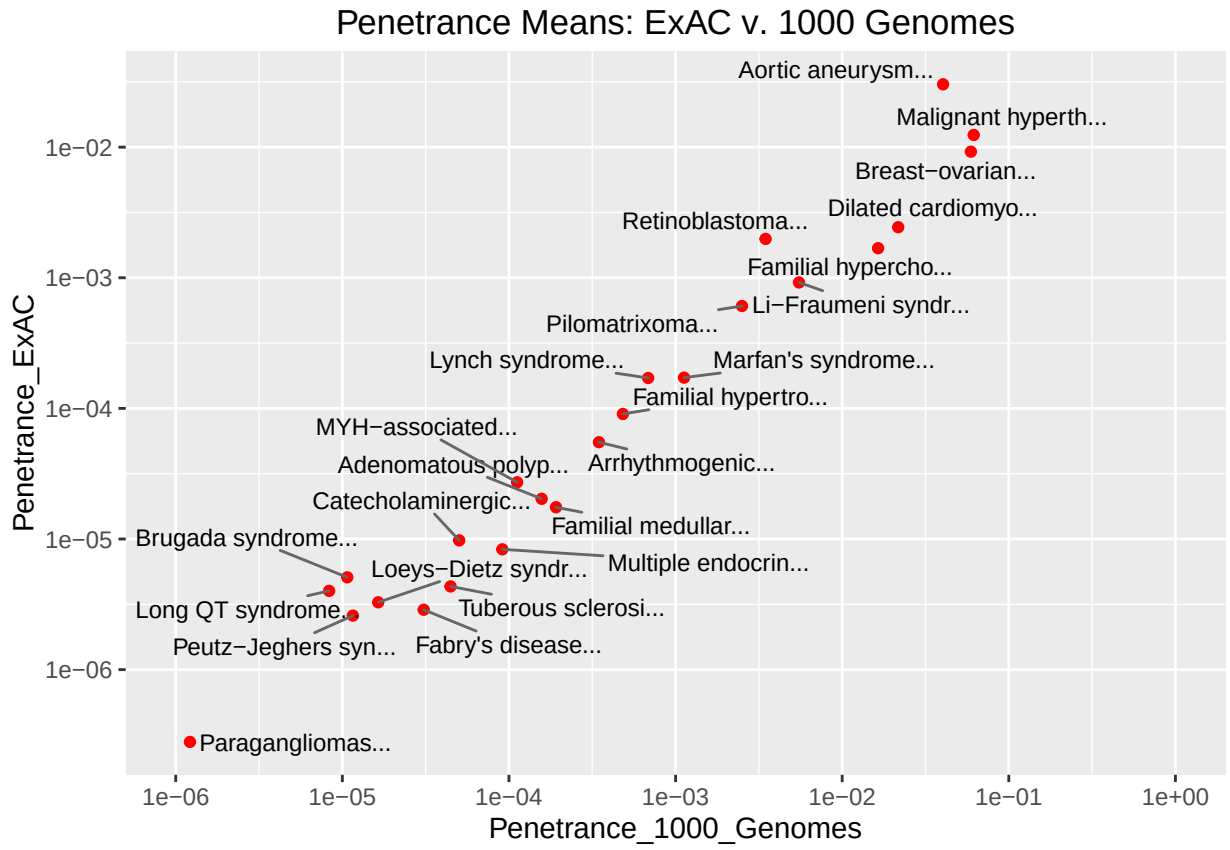



Penetrance by Ancestry (ExAC)
AFR
AMR
EAS
EUR
SAS
Total
Disease
Paragangliomas...
Peutz−Jeghers syn...
Fabry's disease...
Loeys−Dietz syndr...
Long QT syndrome...
Tuberous sclerosi...
Brugada syndrome...
Multiple endocrin...
Catecholaminergic...
Familial medullar...
Adenomatous polyp...
MYH−associated...
Arrhythmogenic...
Von Hippel−Lindau...
Familial hypertro...
Lynch syndrome...
Marfan's syndrome...
PTEN hamartoma...
Pilomatrixoma...
Li−Fraumeni syndr...
Familial hypercho...
Retinoblastoma...
Dilated cardiomyo...
Left ventricular ...
Breast−ovarian...
Malignant hyperth...
Aortic aneurysm...
Wilms' tumor...
Neurofibromatosis...
Ehlers−Danlos syn...
0.00
0.25
0.50
0.75
1.00
Penetrance

### 3.3 Empirical CDFs for All Penetrance Plots




### 3.4 Comparing Mean Penetrance between ExAC and 1000 Genomes




The Pearson correlation is 0.74.
Max penetrance values computed using 1000 Genomes are 2-fold larger than those computed using ExAC.