ClinVar Report

James Diao

November 4, 2016

Contents

1	Collect and Merge ClinVar Data 1.1 Import ClinVar VCF	
2	Summary Statistics 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
	3.3 Empirical CDFs for All Penetrance Plots	 7
	3.4 Comparing Mean Penetrance between ExAC and 1000 Genomes $\ \ldots \ \ldots \ \ldots$	 7
	ourcing ClinVar input from: clinvar_2014-11-05.vcf ending output to: Report_2014-11-05.pdf	

1 Collect and Merge ClinVar Data

1.1 Import ClinVar VCF

Processed ClinVar data frame 83473 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	83473
LP/P-ClinVar	19973
LP/P-ClinVar & ACMG	3448
LP/P-ClinVar & ACMG & ExAC	567
LP/P-ClinVar & ACMG & 1000	121
Genomes	

Breakdown of ACMG-1000 Genomes Variants

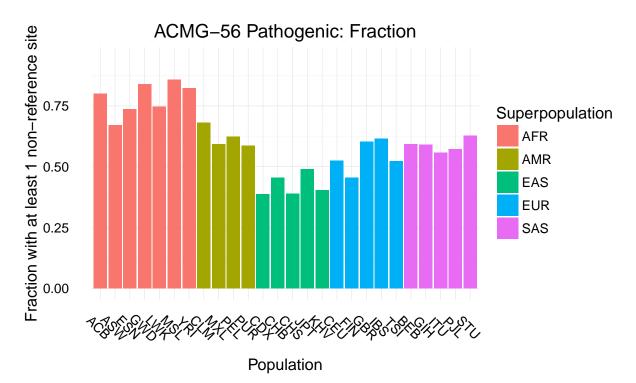
Subset_1000_Genomes	Number_of_Variants
Total 1000_Genomes & ACMG	139335
1000_Genomes & ACMG & ClinVar	1918
1000_Genomes & ACMG &	121
LP/P-ClinVar	

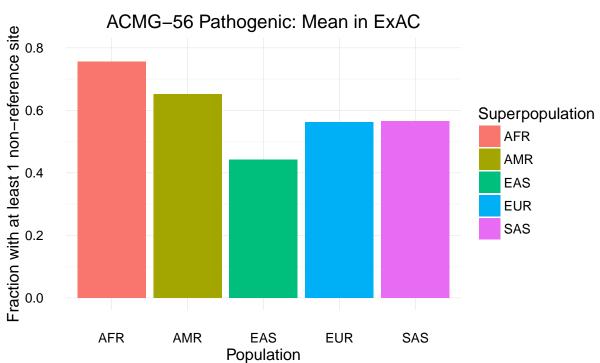
Breakdown of ACMG-ExAC Variants

Subset_ExAC	Number_of_Variants
Total ExAC & ACMG	58873
ExAC & ACMG & ClinVar	3669
ExAC & ACMG & LP/P-ClinVar	567

2 Summary Statistics

2.1 Fraction of Individuals with Pathogenic Non-Reference Sites

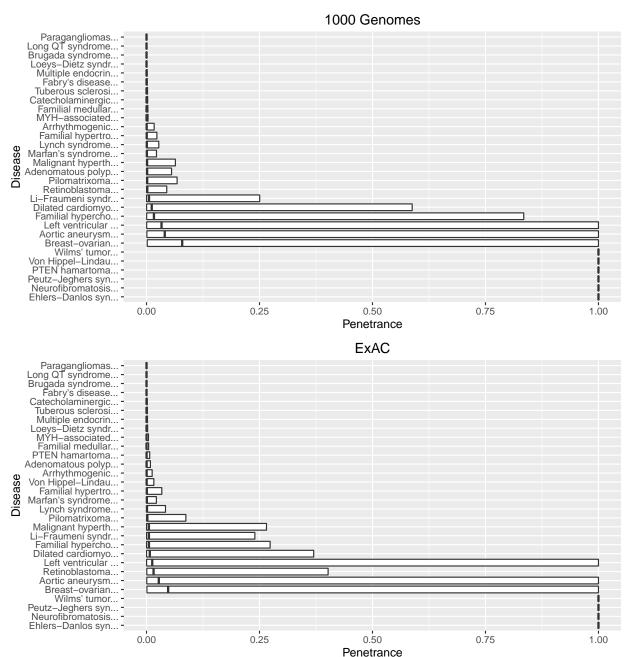




3 Penetrance Estimates

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

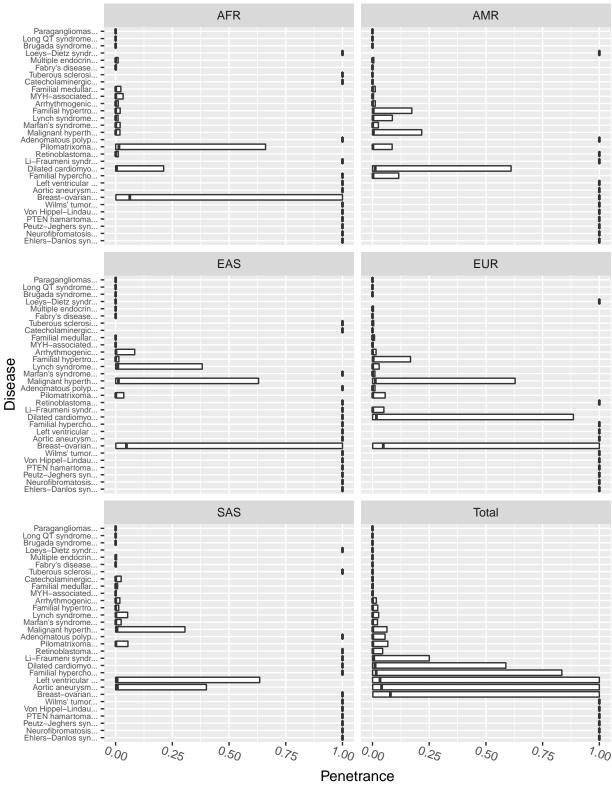
The left end of the boxplot indicates P(D) AND P(V|D) = lower value, the bold line in the middle indicates P(D) AND P(V|D) = geometric_mean(values), the right end of the boxplot indicates P(D) AND P(V|D) = upper value.



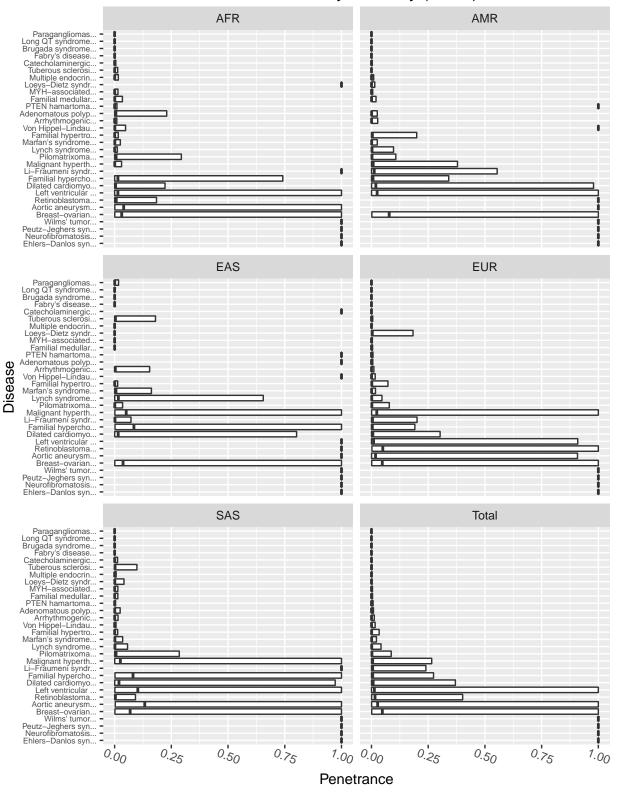
Note: Prevalence ranges of 5x were assumed for all point estimates of prevalence. For example: a point estimate of 0.022 would be given the range 0.01-0.05.

3.2 Penetrance Estimates by Ancestry

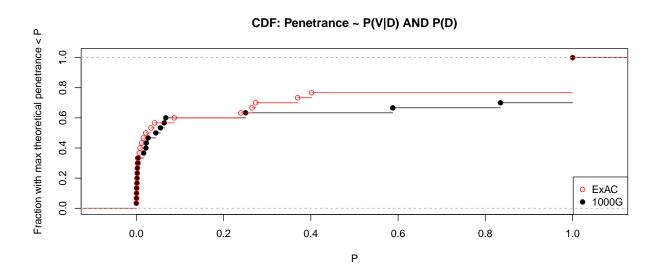




Penetrance by Ancestry (ExAC)



3.3 Empirical CDFs for All Penetrance Plots



3.4 Comparing Mean Penetrance between ExAC and 1000 Genomes

Penetrance Means: ExAC v. 1000 Genomes Breast-ovarian... Retinoblastoma... 1e-02 -Dilated cardiomyo. Li-Fraumeni syndr... Familial hypercho... Pilomatrixoma... Penetrance_ExAC Lynch syndrome Marfan's syndrome... Familial hypertro... Arrhythmogenic... Adenomatous polyp... Familial medullar... Loeys-Dietz syndr...Multiple endocrin... MYH-associated... Tuberous sclerosi Catecholaminergic... Brugada syndrome 1e-05 -Fabry's disease... Long QT syndrome... Paragangliomas... 1e-06 **-**1e-02 1e-01 1e+00 1e-06 1e-05 1e-03 Penetrance_1000_Genomes

The Pearson correlation is 0.95. Max penetrance values computed using 1000 Genomes are 1.6-fold larger than those computed using ExAC.