# ClinVar Report

## James Diao

## November 4, 2016

## Contents

1	Col	lect and Merge ClinVar Data	<b>2</b>		
	1.1	Import ClinVar VCF	2		
	1.2	Merge ClinVar with 1000 Genomes and ExAC	2		
2	Sun	nmary Statistics	3		
	2.1	Fraction of Individuals with Pathogenic Non-Reference Sites	3		
3	Pen	netrance Estimates	4		
	3.1	Max/Min Penetrance as a Function of $P(D)$ and $P(V D)$	4		
		Penetrance Estimates by Ancestry			
	3.3	Empirical CDFs for All Penetrance Plots	7		
	3.4	Comparing Mean Penetrance between ExAC and 1000 Genomes	7		
$\mathbf{So}$	ourci	ng ClinVar input from: clinvar_2016-05-02.vcf			
Sending output to: Report_2016-05-02.pdf					

## 1 Collect and Merge ClinVar Data

## 1.1 Import ClinVar VCF

## Processed ClinVar data frame 100968 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	100968
LP/P-ClinVar	31247
LP/P-ClinVar & ACMG	6298
LP/P-ClinVar & ACMG & ExAC	956
LP/P-ClinVar & ACMG & 1000	167
Genomes	

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

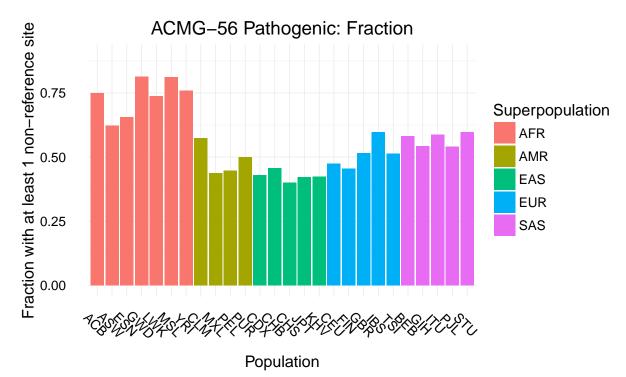
Subset_1000_Genomes	Number_of_Variants
Total 1000_Genomes & ACMG	139335
1000_Genomes & ACMG & ClinVar	4052
1000_Genomes & ACMG &	167
LP/P-ClinVar	

### ## Breakdown of ACMG-ExAC Variants

Subset_ExAC	Number_of_Variants
Total ExAC & ACMG	58873
ExAC & ACMG & ClinVar	8096
ExAC & ACMG & LP/P-ClinVar	956

## 2 Summary Statistics

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



ACMG-56 Pathogenic: Mean in ExAC

Superpopulation

AFR

AMR

EAS

EUR

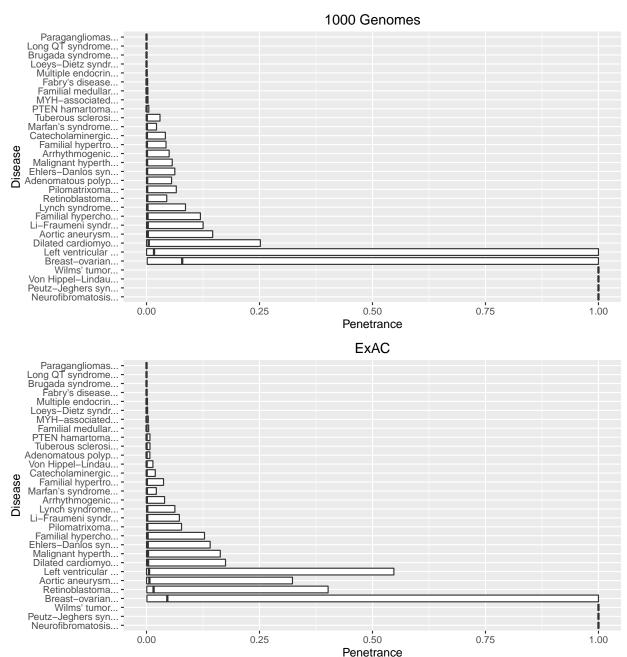
SAS

Population

### 3 Penetrance Estimates

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

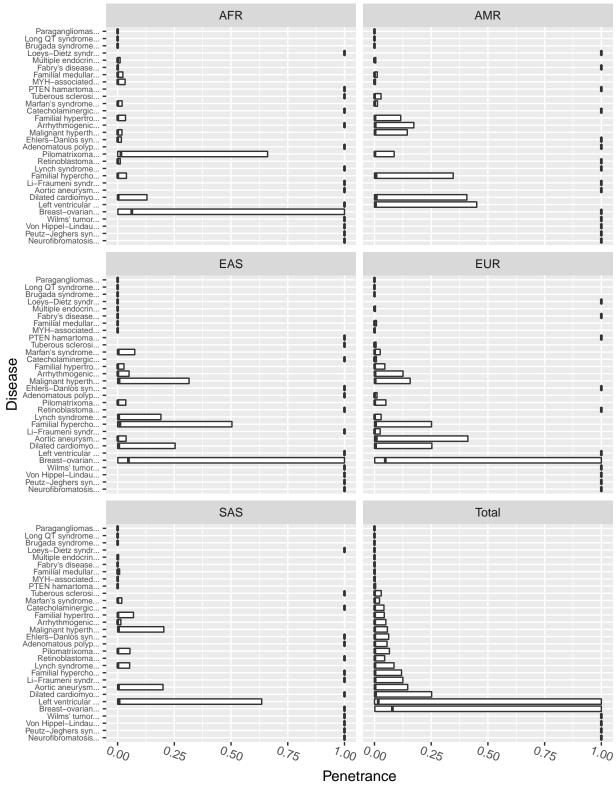
The left end of the boxplot indicates P(D) AND P(V|D) = lower value, the bold line in the middle indicates P(D) AND P(V|D) = geometric\_mean(values), the right end of the boxplot indicates P(D) AND P(V|D) = upper value.



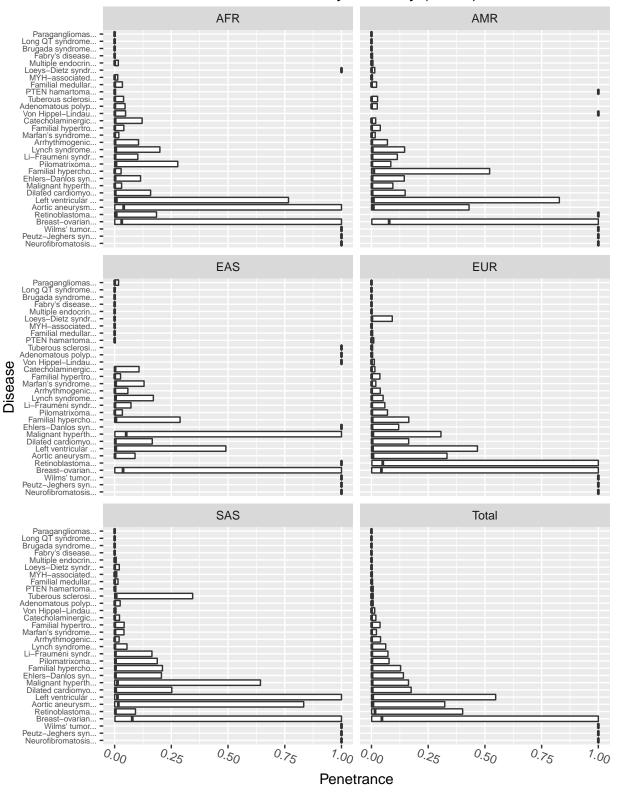
Note: Prevalence ranges of 5x were assumed for all point estimates of prevalence. For example: a point estimate of 0.022 would be given the range 0.01-0.05.

### 3.2 Penetrance Estimates by Ancestry

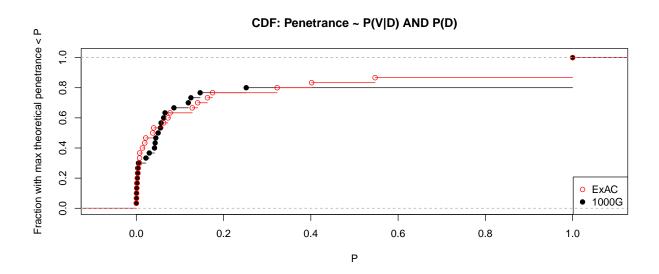




## Penetrance by Ancestry (ExAC)



### 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 -Aortic aneurysm... Left ventricular ... Dilated cardiomyo... Malignant hyperth. Familial hypercho... Penetrance\_ExAC Lynch syndrome... Arrhythmogenic... Catecholaminergic. Adenomatous polyp... PTEN hamartoma. Tuberous sclerosi... Familial medullar. Loeys-Dietz syndr.... MYH-associated... Multiple endocrin... Fabry's disease... Brugada syndrome. 1e-05 -Long QT syndrome... 1e-06 **-**1e-01 1e-06 1e-03 1e-02 1e+00 Penetrance\_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.