

# ClinVar Report

*James Diao*

*February 1, 2017*

## Contents

<b>1</b>	<b>Collect and Merge ClinVar Data</b>	<b>2</b>
1.1	Import ClinVar VCF . . . . .	2
1.2	Merge ClinVar with 1000 Genomes and ExAC . . . . .	2
<b>2</b>	<b>Summary Statistics</b>	<b>3</b>
2.1	Fraction of Individuals with Pathogenic Non-Reference Sites . . . . .	3
<b>3</b>	<b>Penetrance Estimates</b>	<b>4</b>
3.1	Max/Min Penetrance as a Function of $P(D)$ and $P(V D)$ . . . . .	4
3.2	Penetrance Estimates by Ancestry . . . . .	4

**Sourcing ClinVar input from:** clinvar\_2012-06-16.vcf

**Sending output to:** Report\_2012-06-16.pdf

# 1 Collect and Merge ClinVar Data

## 1.1 Import ClinVar VCF

## 1.2 Merge ClinVar with 1000 Genomes and ExAC

## **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  $P(V|D) = 0.01$ ,  
the bold line in the middle indicates  $P(V|D) = \text{point value}$ ,  
the right end of the boxplot indicates  $P(V|D) = 1$ .

Note: Some diseases have mean theoretical penetrance = 1 because the assumed allelic heterogeneity is greater than is possible, given the observed prevalence and allele frequencies.

#### 3.2 Penetrance Estimates by Ancestry