# Finding Rare Variants of Large Effects Related to Complex Disease

Eric Boerwinkle
NHGRI, DC

June, 2012

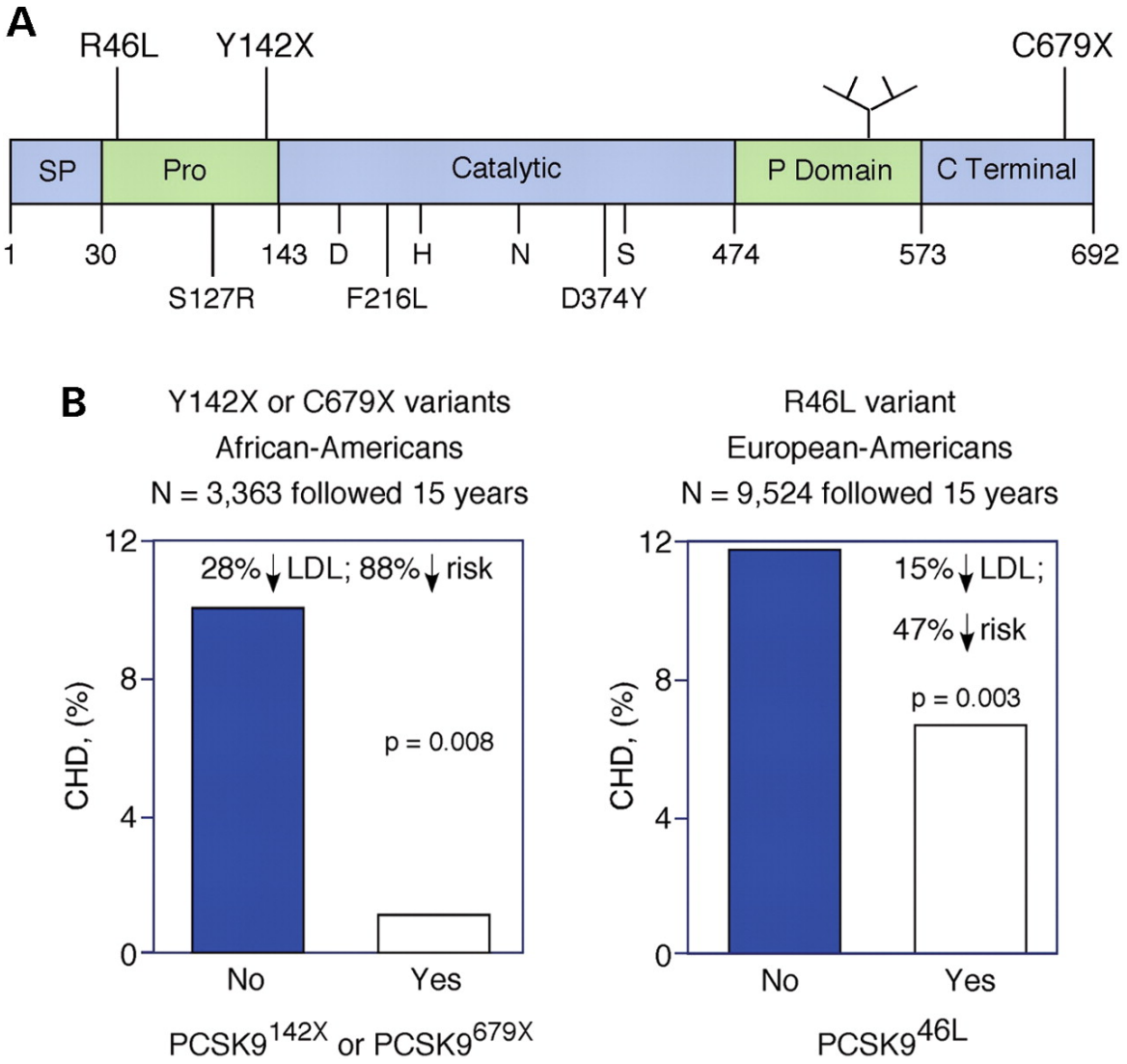
UTHealth
The University of Texas
Health Science Center at Houston

HGSC
HUMAN GENOME SEQUENCING CENTER

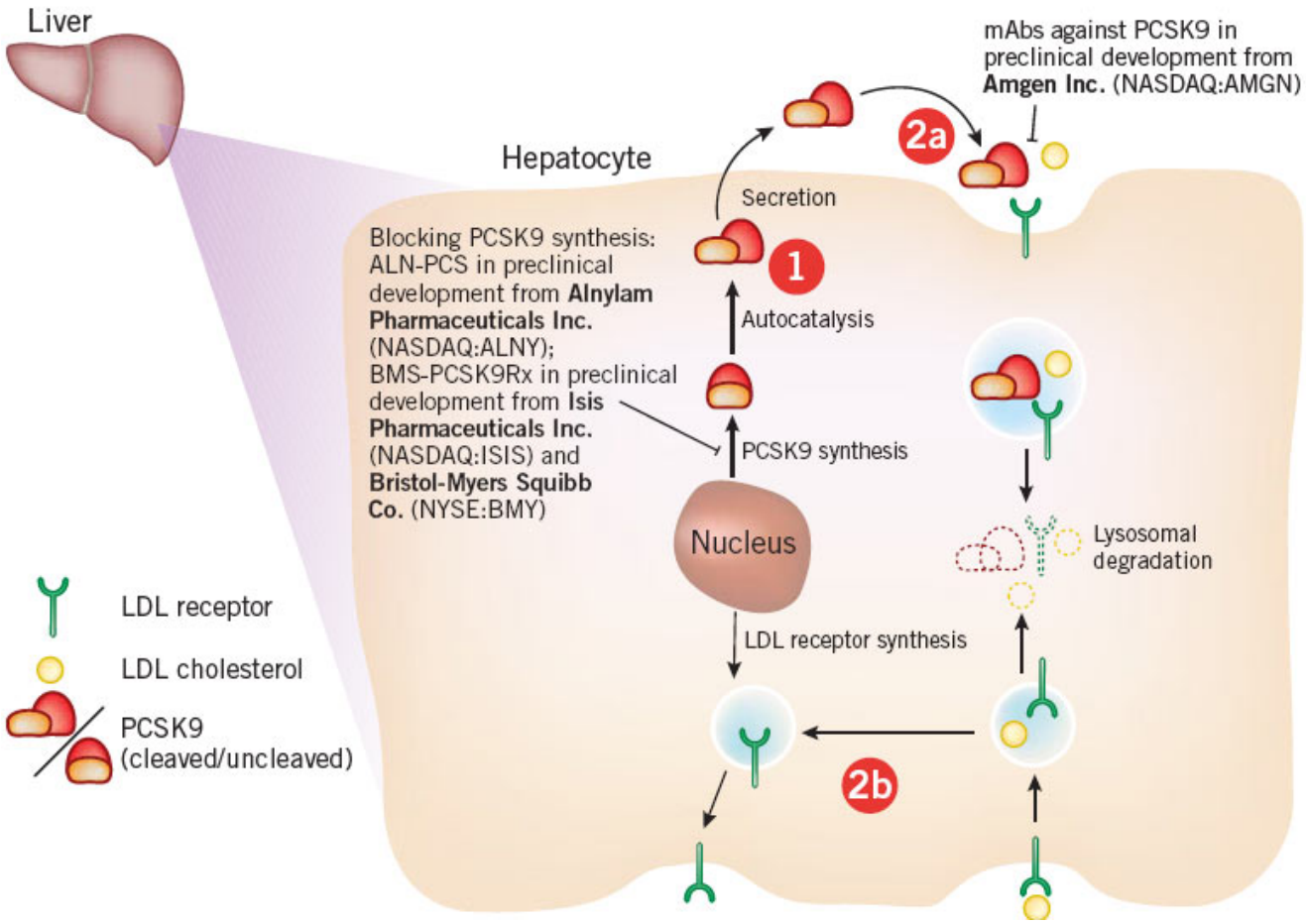
# Rare Variants

- Who cares?

# PCSK9 schematic of nonsense mutations associated with low LDL cholesterol



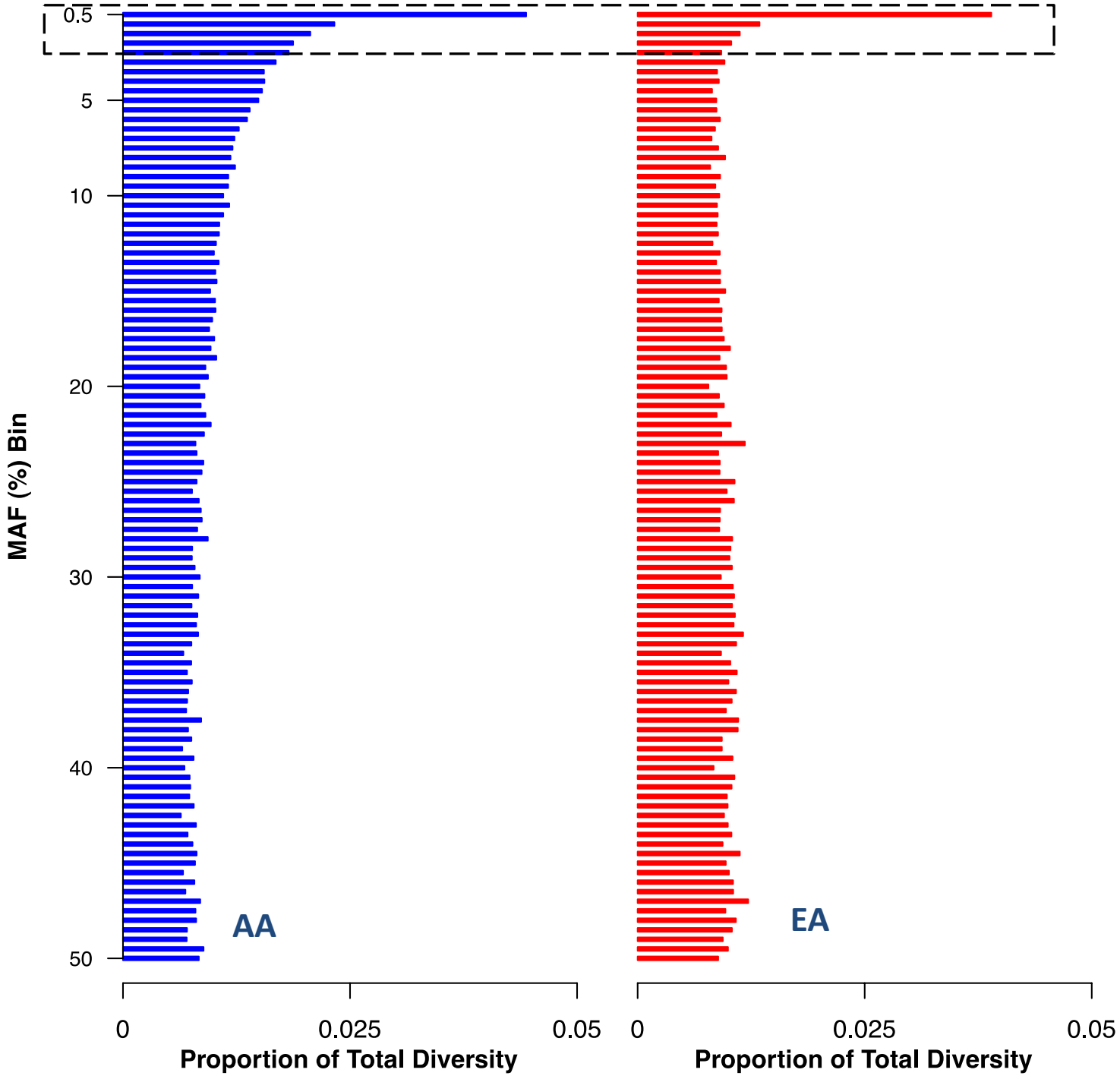
# PCSK9 as a Therapeutic



# Rare Variants

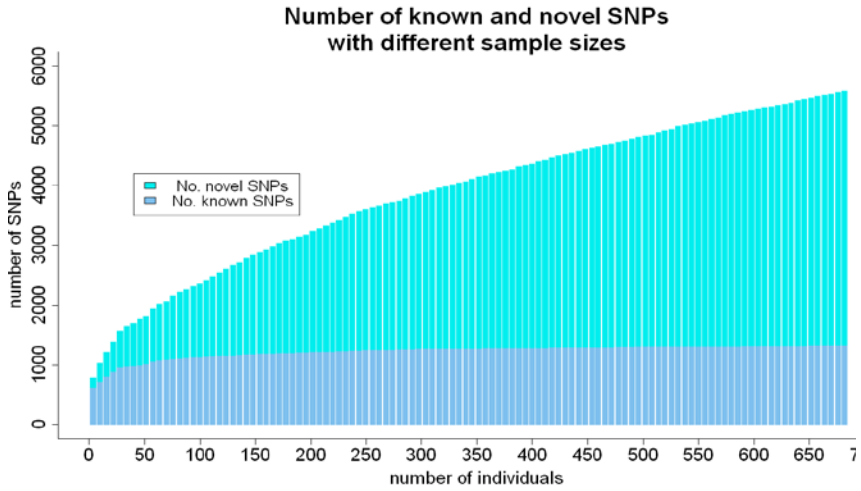
- Who cares?
- How many are there?

**Rare variation makes a significant contribution to genetic diversity**

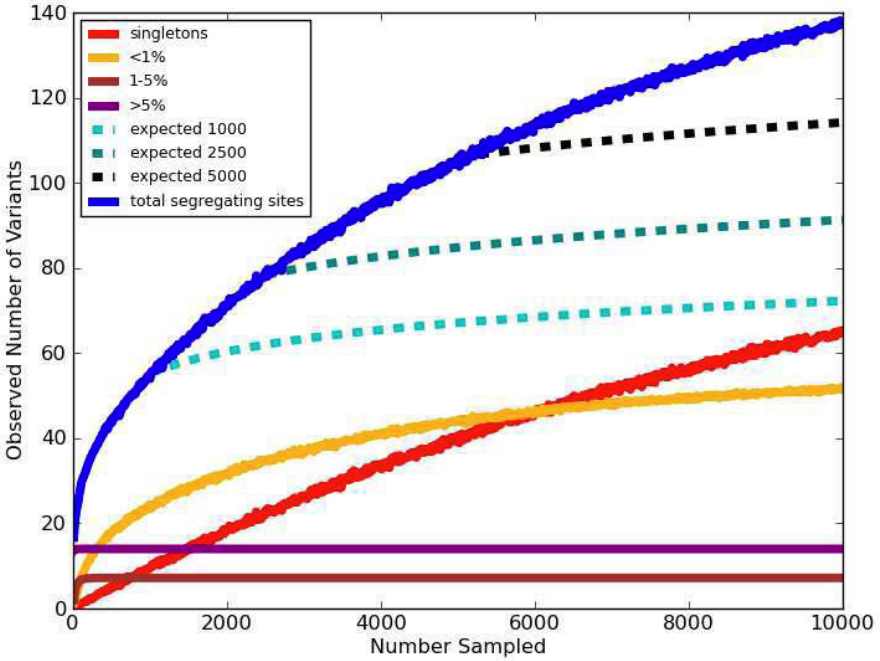


# How many rare variants?

## (Predicting $n^{th}$ Genome)




**ENCODE 3 Sequence Data**
**~ 700 samples. 1 Mb each**

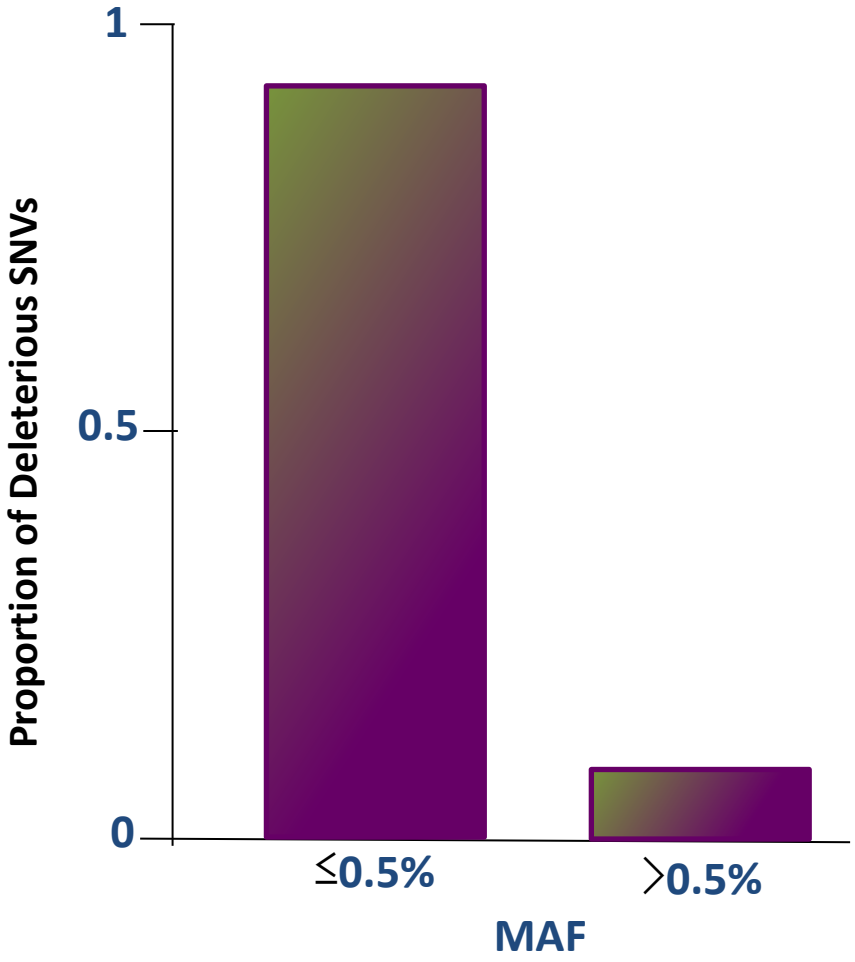



**ARIC:**
**~ 11,000 samples, ~ 25 kb**

# ~ 10-300 novel ns / genome!

# ~90% of Predicted Deleterious Alleles are Rare

- 15% of SNVs (77,746) are predicted to be functionally important

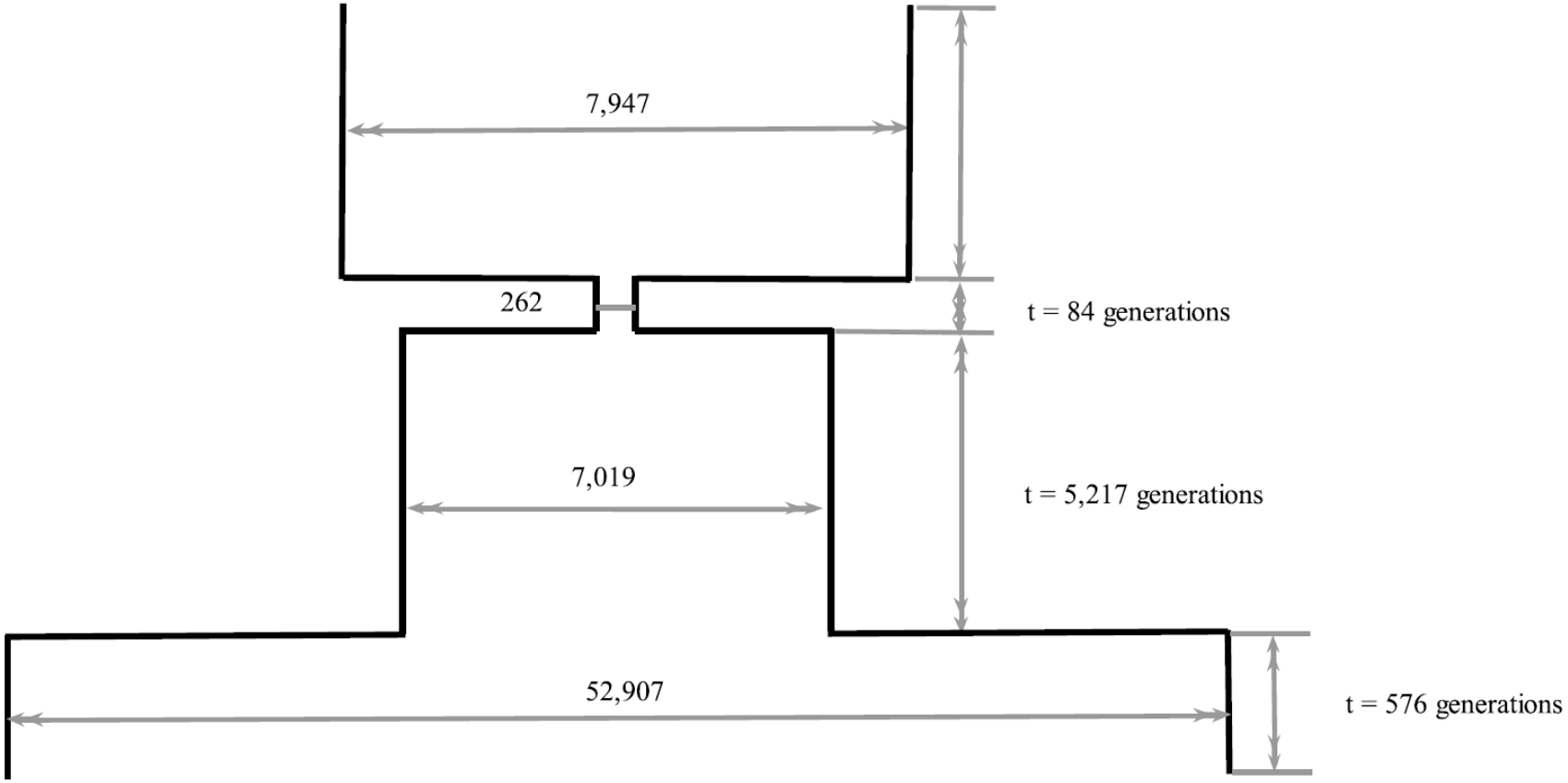


# Rare Variants

- Who cares?
- How many are there?
- Where did they come from?

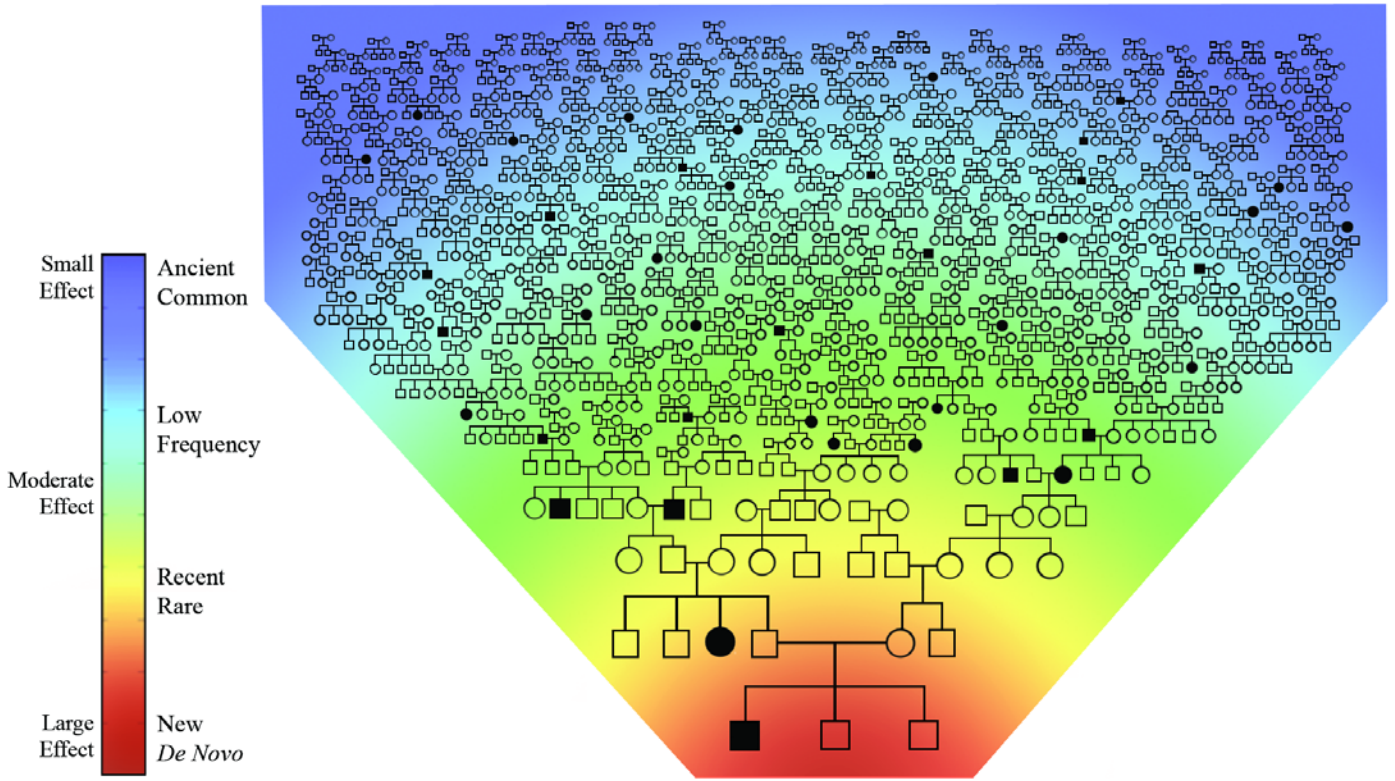
# Simulation of Variant Data

- Demographic history of European population




- Population genetic model incorporating demographic change (Boyko et al PloS Genet 2008)

# Genomics in a Historic Context




Lupski et al (2011) Cell

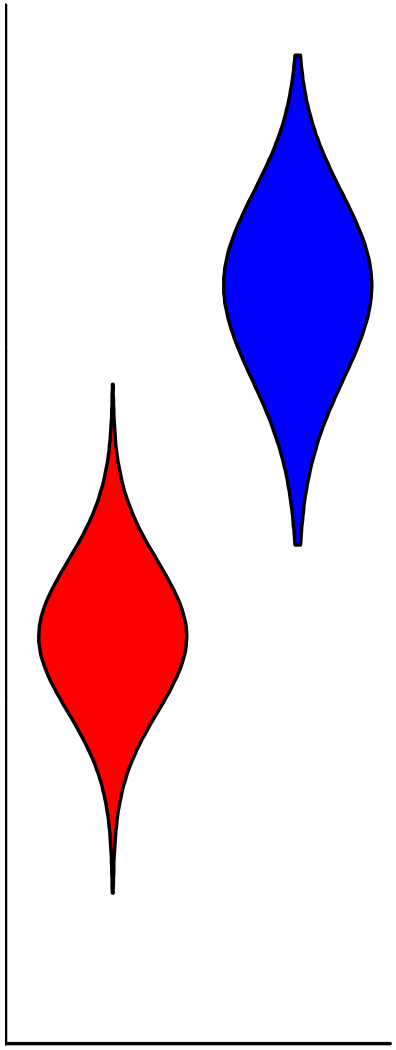
# Rare Variants

- Who cares?
- How many are there?
- Where did they come from?
- They are not in me, they are only in really sick people.

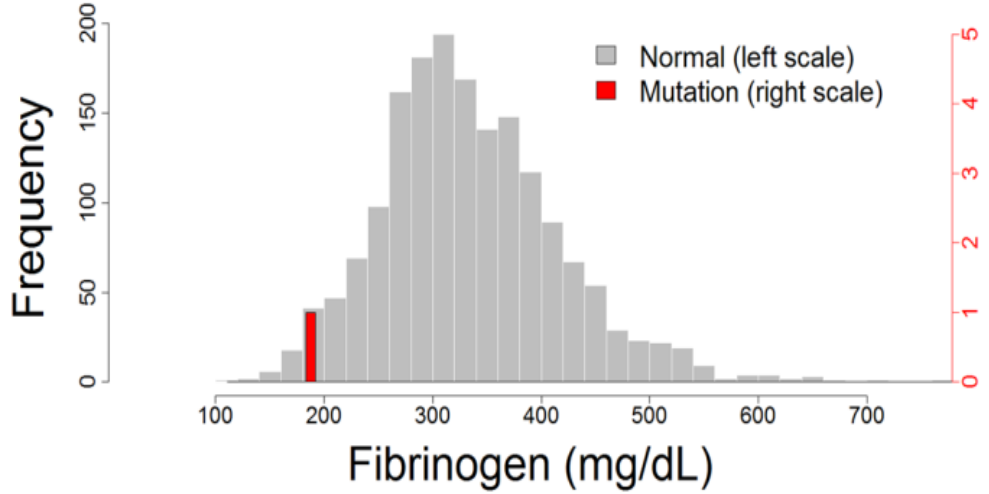
# How Many Deleterious SNVs Do We Carry?

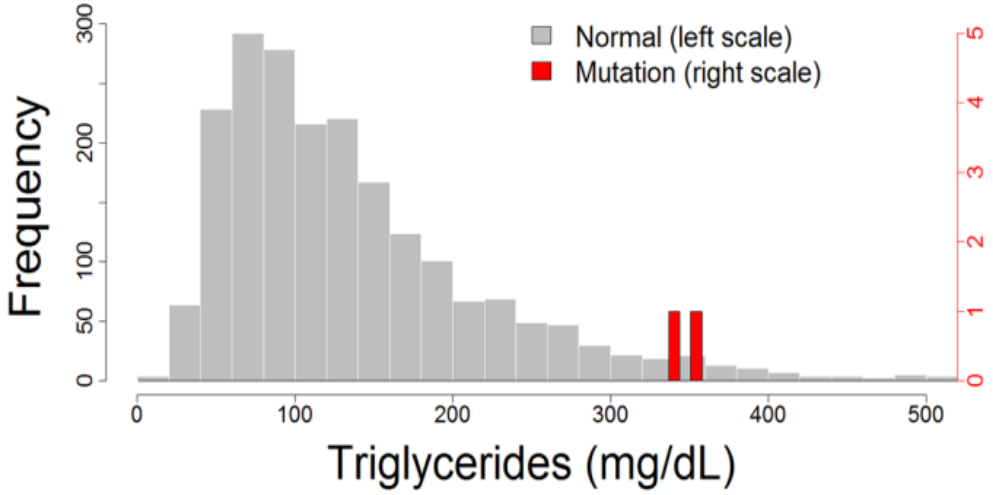
AA

EA



# Two Preliminary Examples



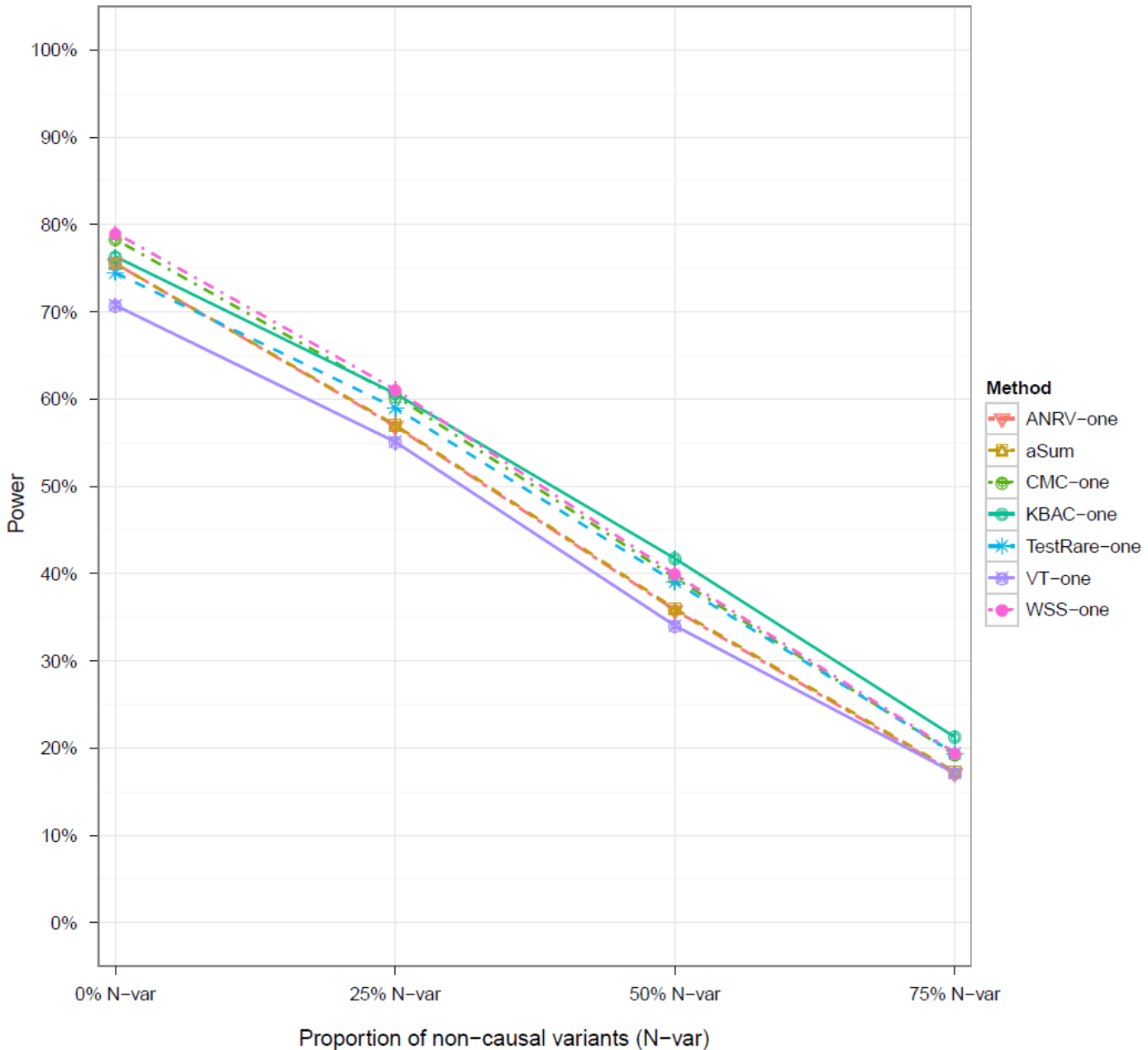
# Rare Variants

- Who cares?
- How many are there?
- Where did they come from?
- They are not in me, they are only in really sick people.
- How do we analyze rare variants?

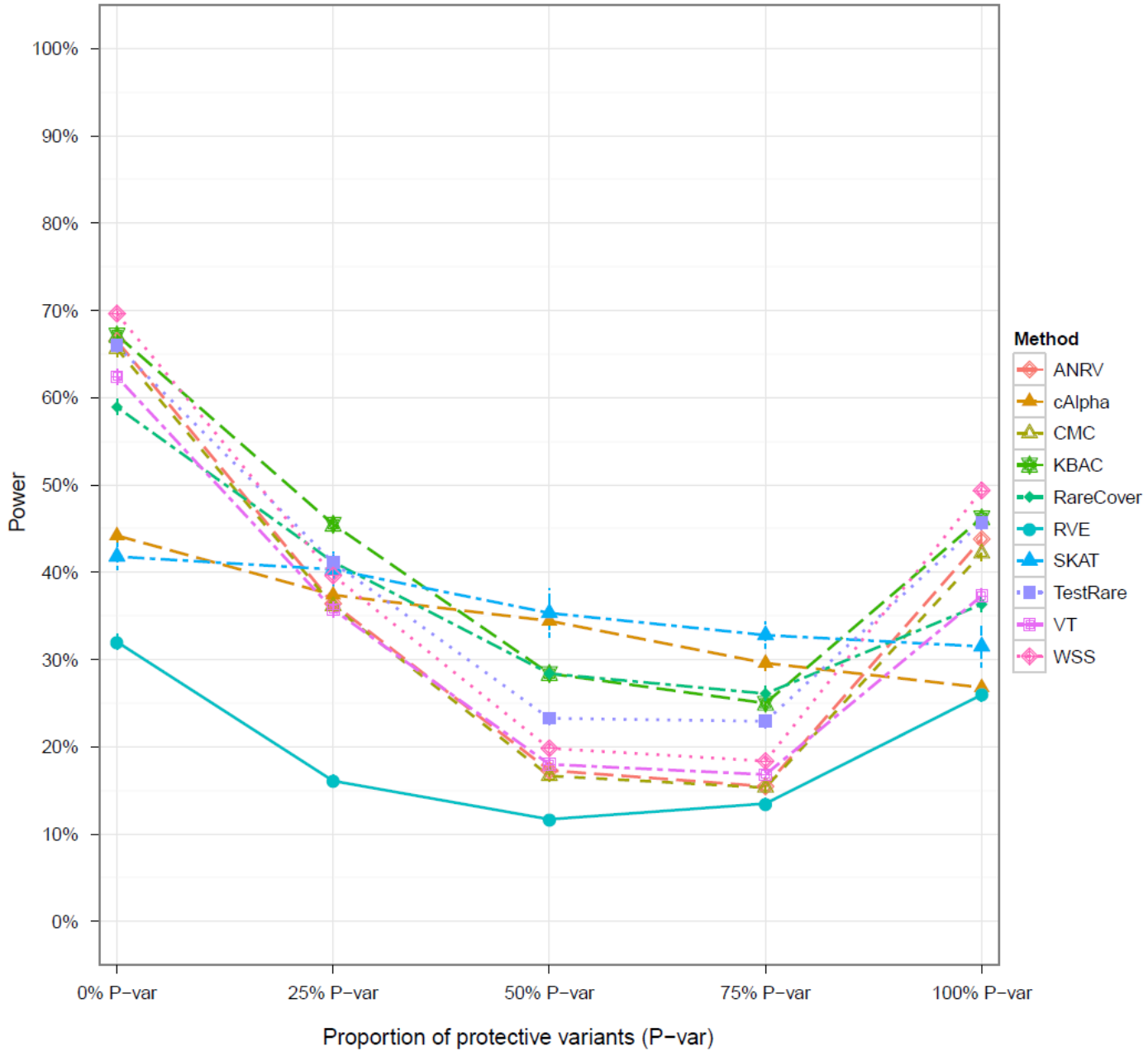
# NS Allele count for top protective gene: *NBEAL1*

| | Cases | Controls |
|---|---|---|
| European ancestry | 9 | 22 |
| African Americans | 1 | 8 |
| Uncorrected P | $2 \times 10^{-4}$ | |

# Impact of Non-Causal Variants (tests for detrimental variants)



# Detrimental and Protective Variants ("two-sided" tests)



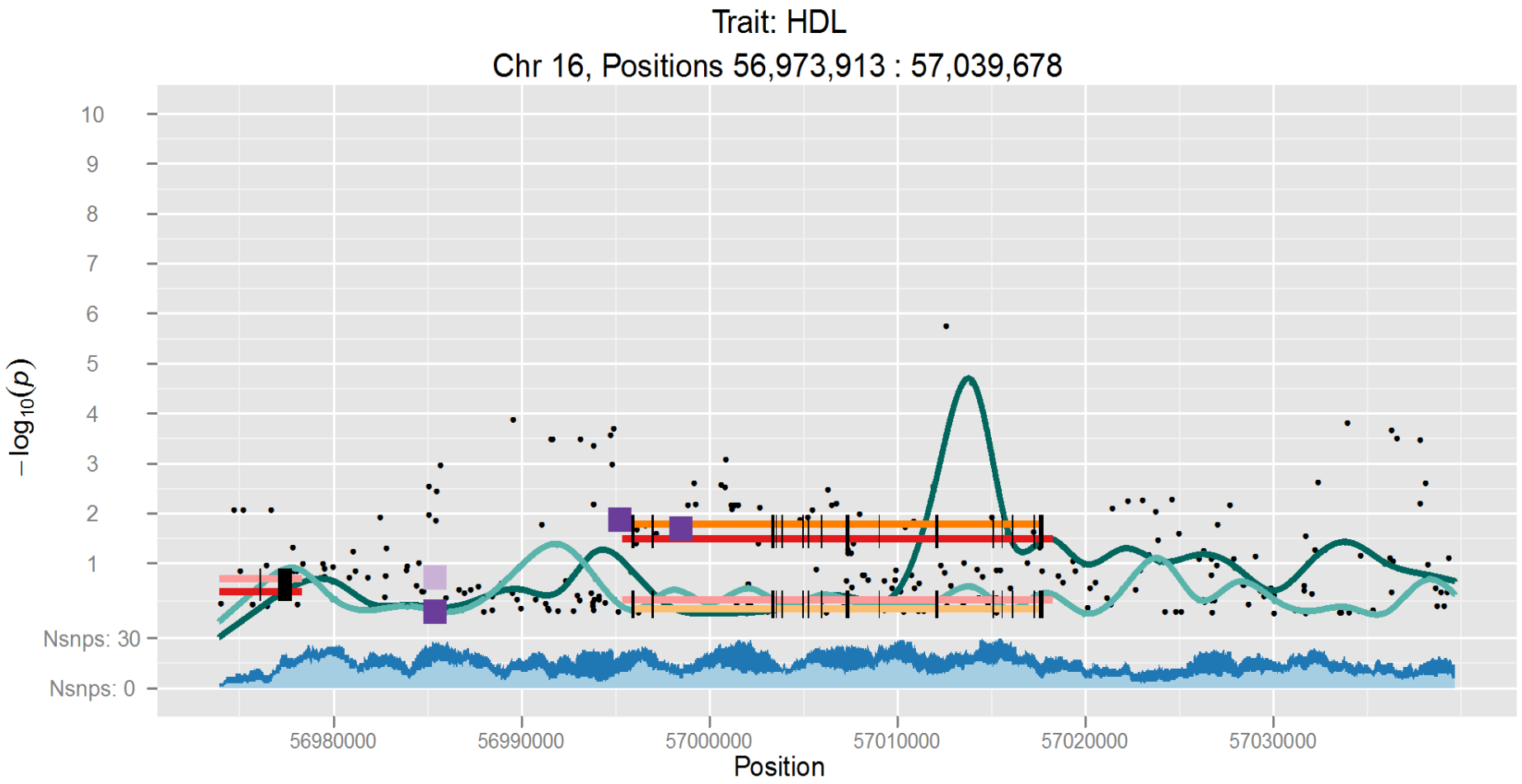
# Rare variants

- Who cares
- How many are there?
- Where did they come from?
- They are not in me, they are only in really sick people.
- How do we analyze rare variants?
- What are we going to do with WGS?

# Whole Genome Sequence Variation (N=990)

| Chr | #var | #mono | #single | #double | %MAF>.01 | %MAF>.05 | %Novel | TiTv | TiTv-singletons |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2093374 | 141080 | 775364 | 175532 | 0.378 | 0.378 | 0.544 | 2.07 | 2.019 |
| 2 | 2388603 | 172173 | 902211 | 200964 | 0.368 | 0.368 | 0.538 | 1.986 | 1.968 |
| 3 | 2000450 | 146724 | 754784 | 168949 | 0.368 | 0.368 | 0.533 | 1.973 | 1.982 |
| 4 | 2084913 | 165137 | 785378 | 174737 | 0.368 | 0.368 | 0.539 | 1.945 | 1.977 |
| 5 | 1827054 | 137193 | 691946 | 152439 | 0.37 | 0.37 | 0.539 | 1.973 | 1.978 |
| 6 | 1772868 | 124536 | 639915 | 143292 | 0.395 | 0.395 | 0.516 | 2.039 | 2.004 |
| 7 | 1614398 | 111886 | 594679 | 133168 | 0.386 | 0.386 | 0.531 | 2.019 | 1.97 |
| 8 | 1551600 | 105194 | 580819 | 131643 | 0.375 | 0.375 | 0.527 | 1.856 | 1.885 |
| 9 | 1233030 | 87525 | 456776 | 100802 | 0.38 | 0.38 | 0.539 | 1.914 | 1.91 |
| 10 | 1334721 | 84208 | 476347 | 107987 | 0.4 | 0.4 | 0.51 | 2.058 | 2.011 |
| 11 | 1347518 | 90082 | 498092 | 110821 | 0.386 | 0.386 | 0.524 | 1.988 | 1.973 |
| 12 | 1300915 | 86964 | 476183 | 109475 | 0.387 | 0.387 | 0.518 | 2.066 | 1.999 |
| 13 | 1025342 | 79869 | 386227 | 84437 | 0.375 | 0.375 | 0.539 | 2.02 | 2.011 |
| 14 | 899366 | 61431 | 329197 | 74412 | 0.386 | 0.386 | 0.525 | 2.038 | 1.978 |
| 15 | 784724 | 49117 | 282976 | 63790 | 0.395 | 0.395 | 0.52 | 2.003 | 1.961 |
| 16 | 830172 | 40649 | 294231 | 69934 | 0.403 | 0.403 | 0.501 | 1.846 | 1.796 |
| 17 | 710857 | 33871 | 249763 | 60821 | 0.406 | 0.406 | 0.5 | 2.254 | 2.044 |
| 18 | 780205 | 55099 | 286866 | 64334 | 0.384 | 0.384 | 0.522 | 2.043 | 2.02 |
| 19 | 550002 | 19172 | 178085 | 44026 | 0.454 | 0.454 | 0.472 | 2.186 | 1.957 |
| 20 | 570402 | 28374 | 201379 | 47880 | 0.412 | 0.412 | 0.485 | 2.186 | 2.041 |
| 21 | 373298 | 24933 | 133630 | 30665 | 0.396 | 0.396 | 0.521 | 2.089 | 2.003 |
| 22 | 342159 | 13967 | 113253 | 27538 | 0.442 | 0.442 | 0.479 | 2.285 | 2.017 |
| Total | 27415971 | 1859184 | 10088101 | 2277646 | 0.384 | 0.384 | 0.526 | 2.008 | 1.975 |

# Annotation and Analysis of Whole Genomes



# Thank You

- Who cares?
- How many are there?
- Where did they come from?
- They are not in me, they are only in really sick people.
- How do we analyze rare variants?
- What are we going to do with WGS?