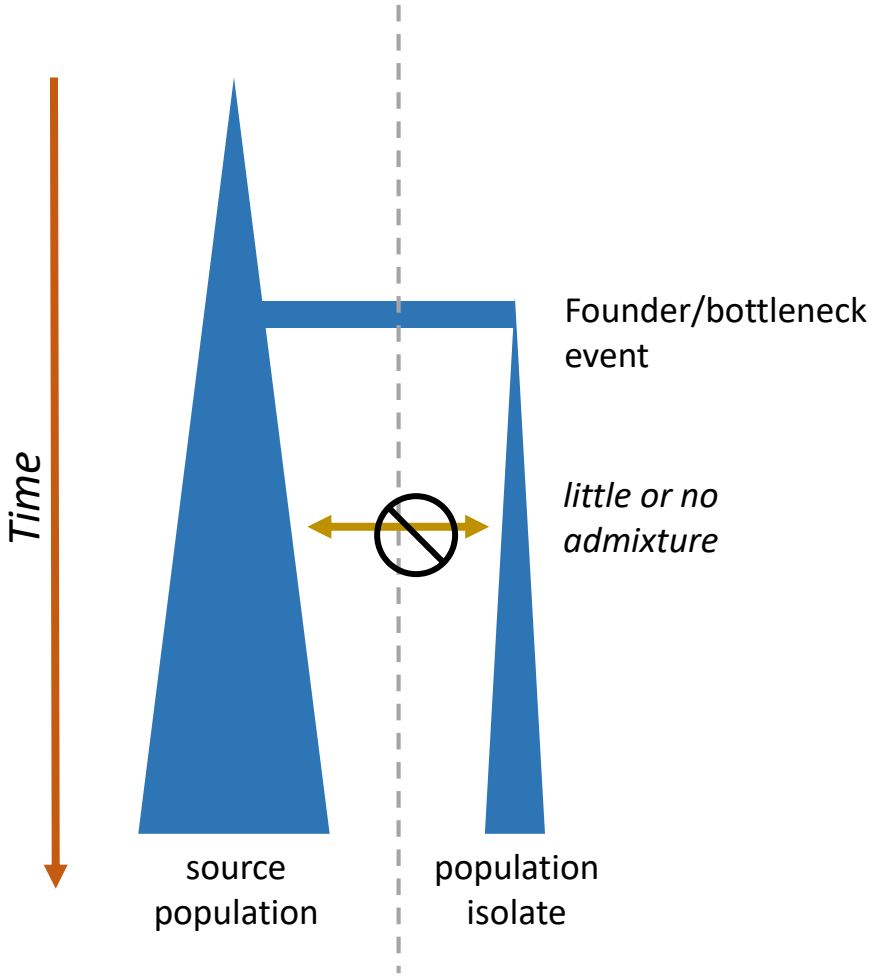
Next-generation sequencing studies in founder populations

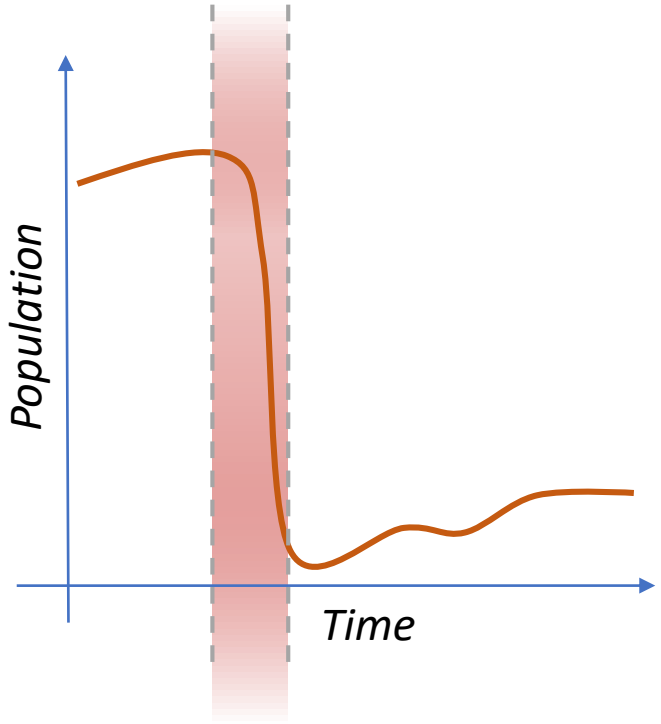
Arthur Gilly, Zeggini Team

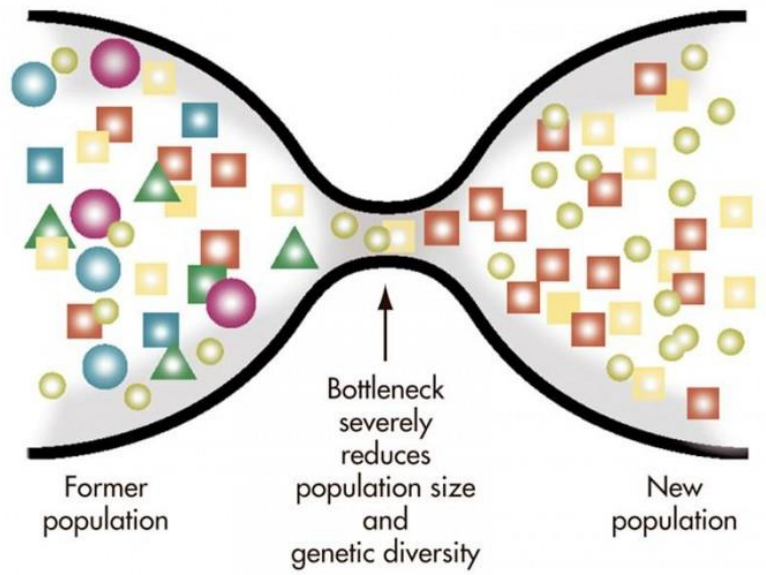
Wellcome Sanger Institute, Cambridge, UK

presented at the "New Statistical Methods for Family-Based Sequencing Studies" workshop in Banff, August 2018

Population isolates





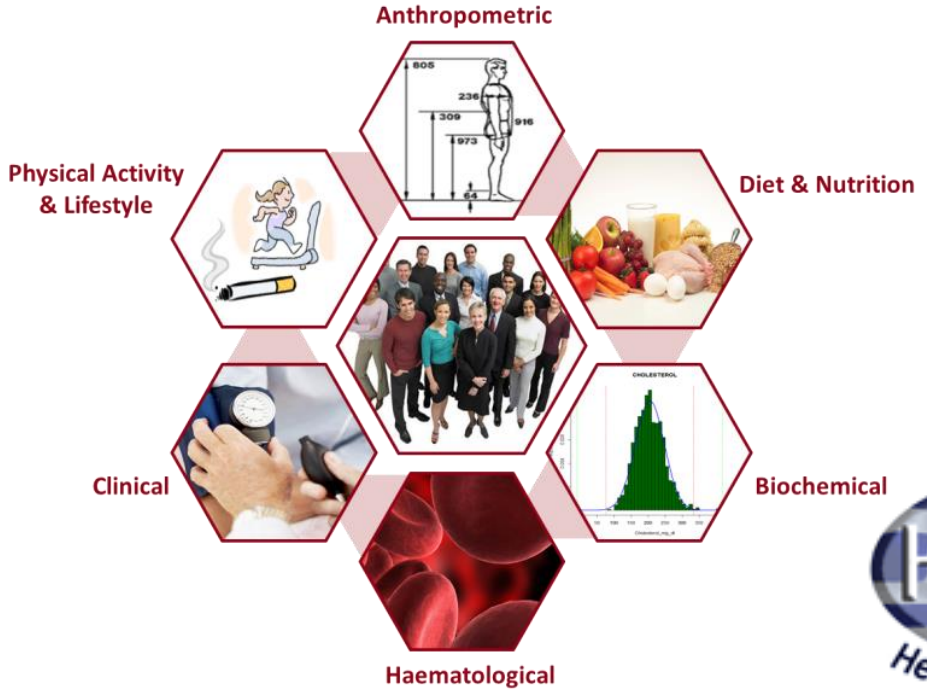


HELIC: Hellenic isolated cohorts

HELIC-MANOLIS (Minoan Isolates)

- Mylopotamos villages, Crete, Greece
- Geographically isolated
- Ancient Dorian dialect
- N~4,500 of which 1,800 collected

- Deeply phenotyped
- High fat content diet
- High rates of longevity
- Low rates of metabolic disease complications
- Ability to recontact individuals



HELIC
Hellenic Isolated Cohorts

www.helic.org

MANOLIS

n=1265

Omni Express

n=1265

Exome Chip

n=211

Core Exome

n=990

1X WGS

n=249

4X WGS

n=1482

22X WGS

n=5

WES

HELIC datasets

Single-point association

GWAS, very-low-depth WGS

MANOLIS

n=1265
Omni Express

n=1265
Exome Chip

n=211
Core Exome

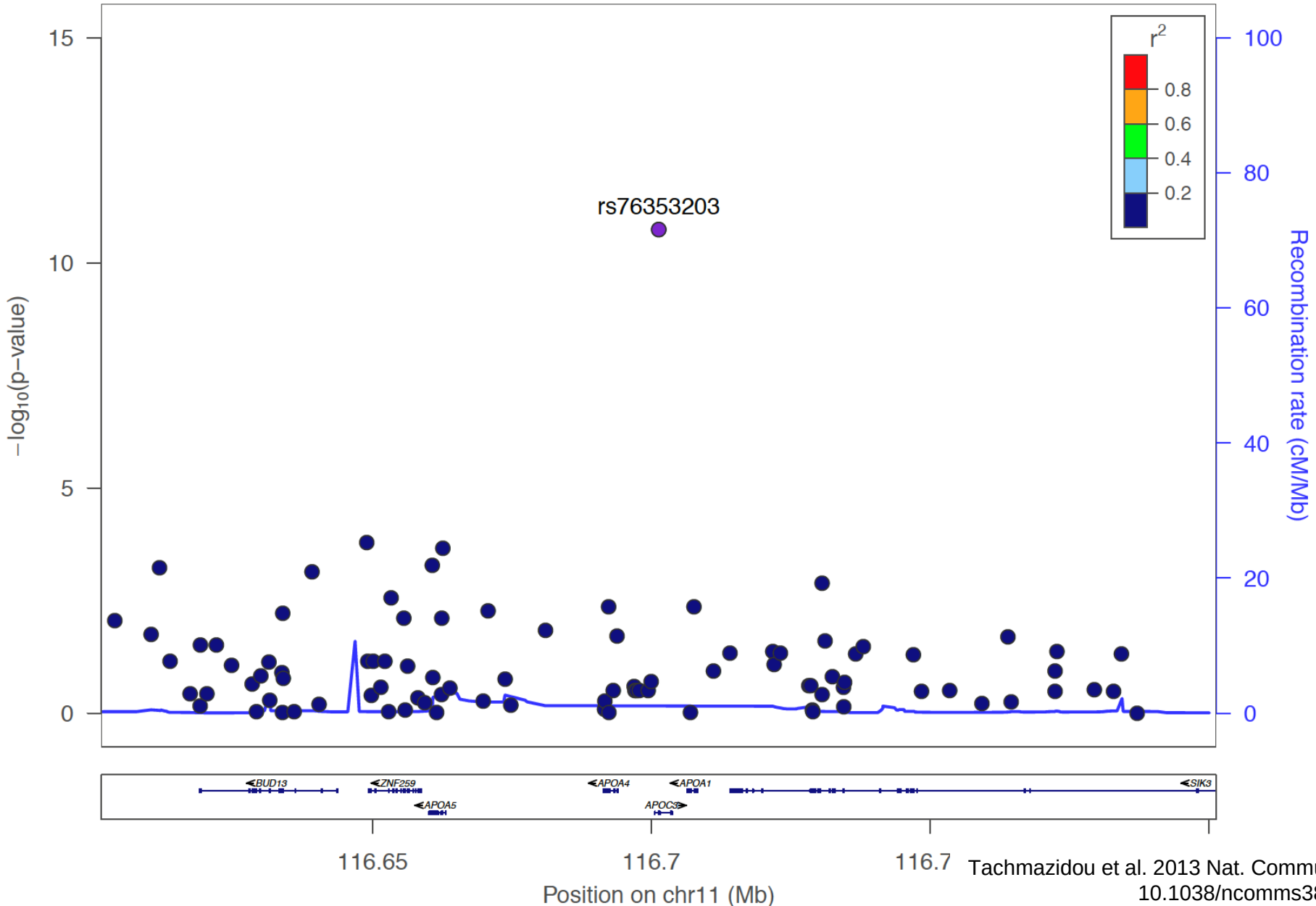
n=990
1X WGS

n=249
4X WGS

n=1482
22X WGS

n=5
WES

Cardioprotective R19X variant in *APOC3*
ExomeChip data



Tachmazidou et al. 2013 Nat. Commun.
10.1038/ncomms3872

MANOLIS

n=1265
Omni Express

n=1265
Exome Chip

n=211
Core Exome

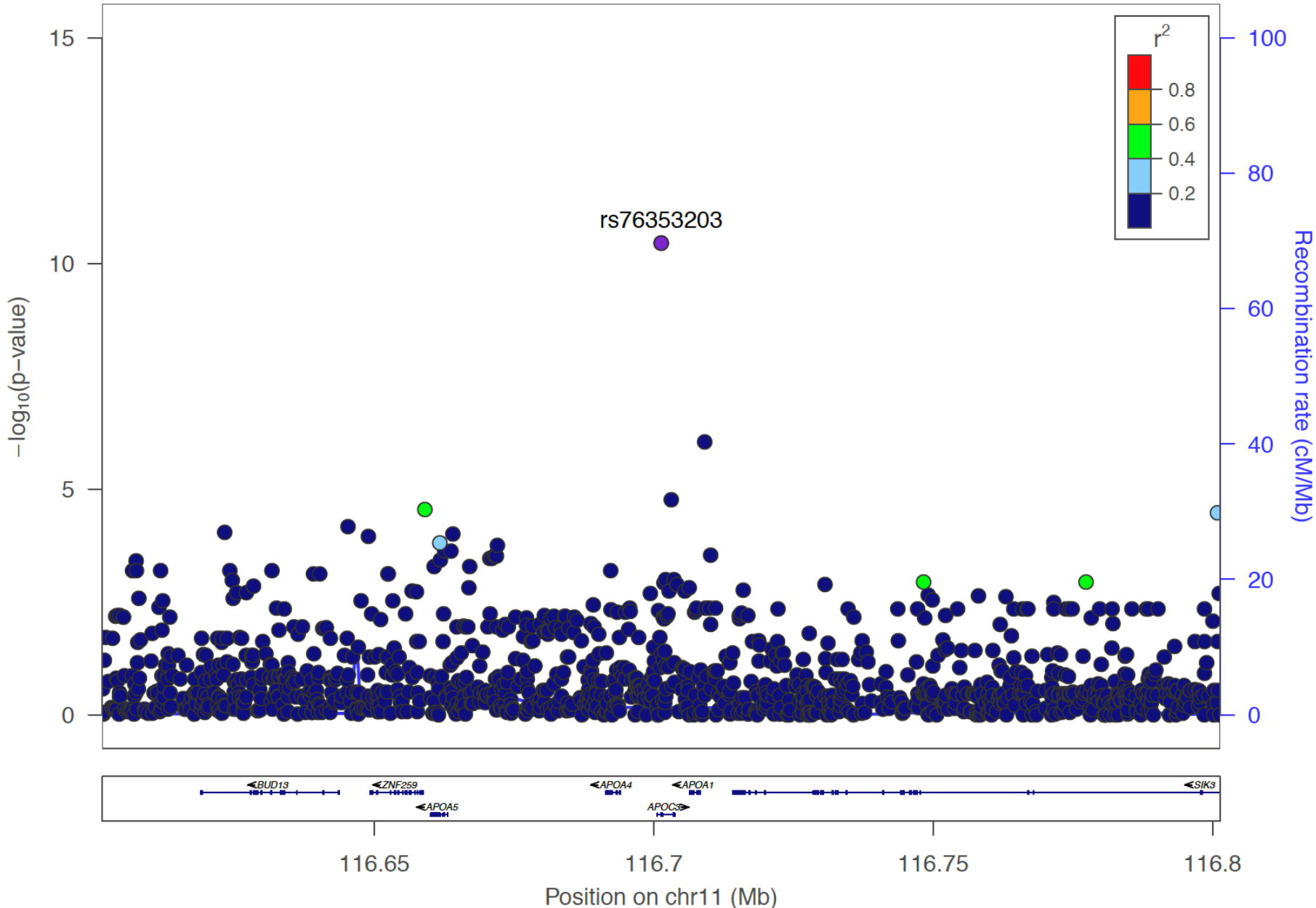
n=990
1X WGS

n=249
4X WGS

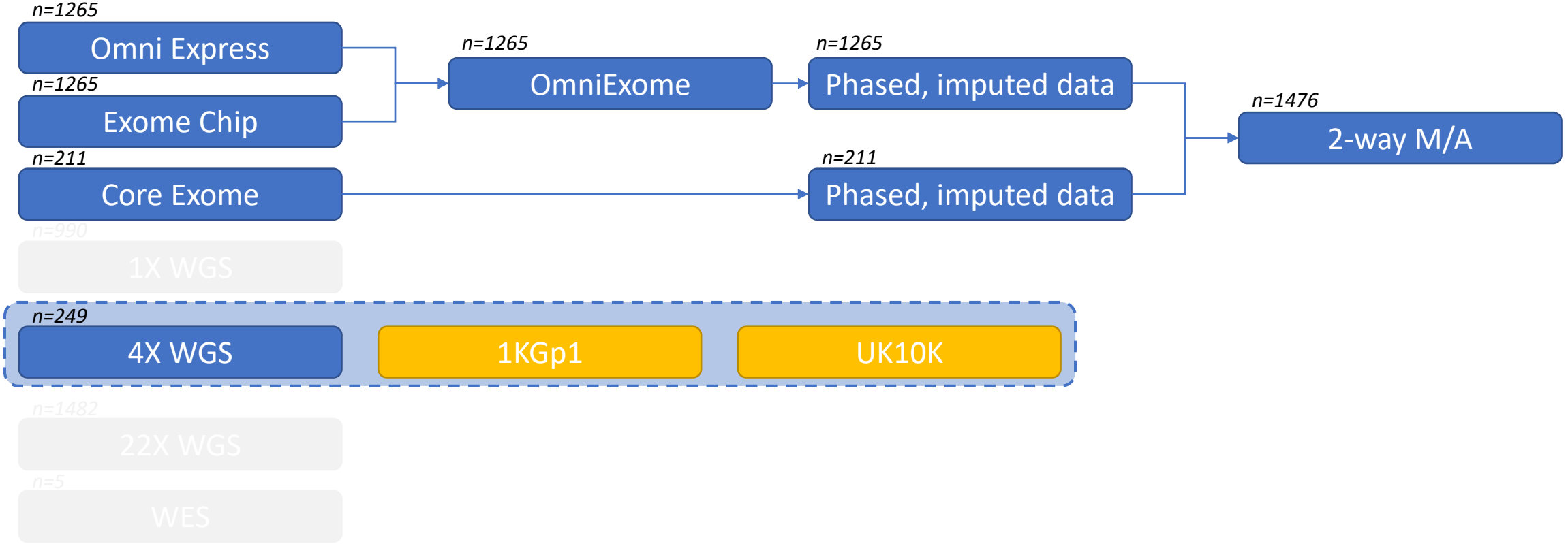
n=1482
22X WGS

n=5
WES

Cardioprotective R19X variant in *APOC3*
Imputed OmniExpress/ExomeChip data



MANOLIS



Method:

- GEMMA (with empirical relatedness matrix)
- METACARPA: meta-analysis accounting for sample relatedness

Lorraine Southam

Southam, Gilly et al. Nat Comms. doi: 10.1038/ncomms15606

MANOLIS

n=1265
Omni Express

n=1265
Exome Chip

n=211
Core Exome

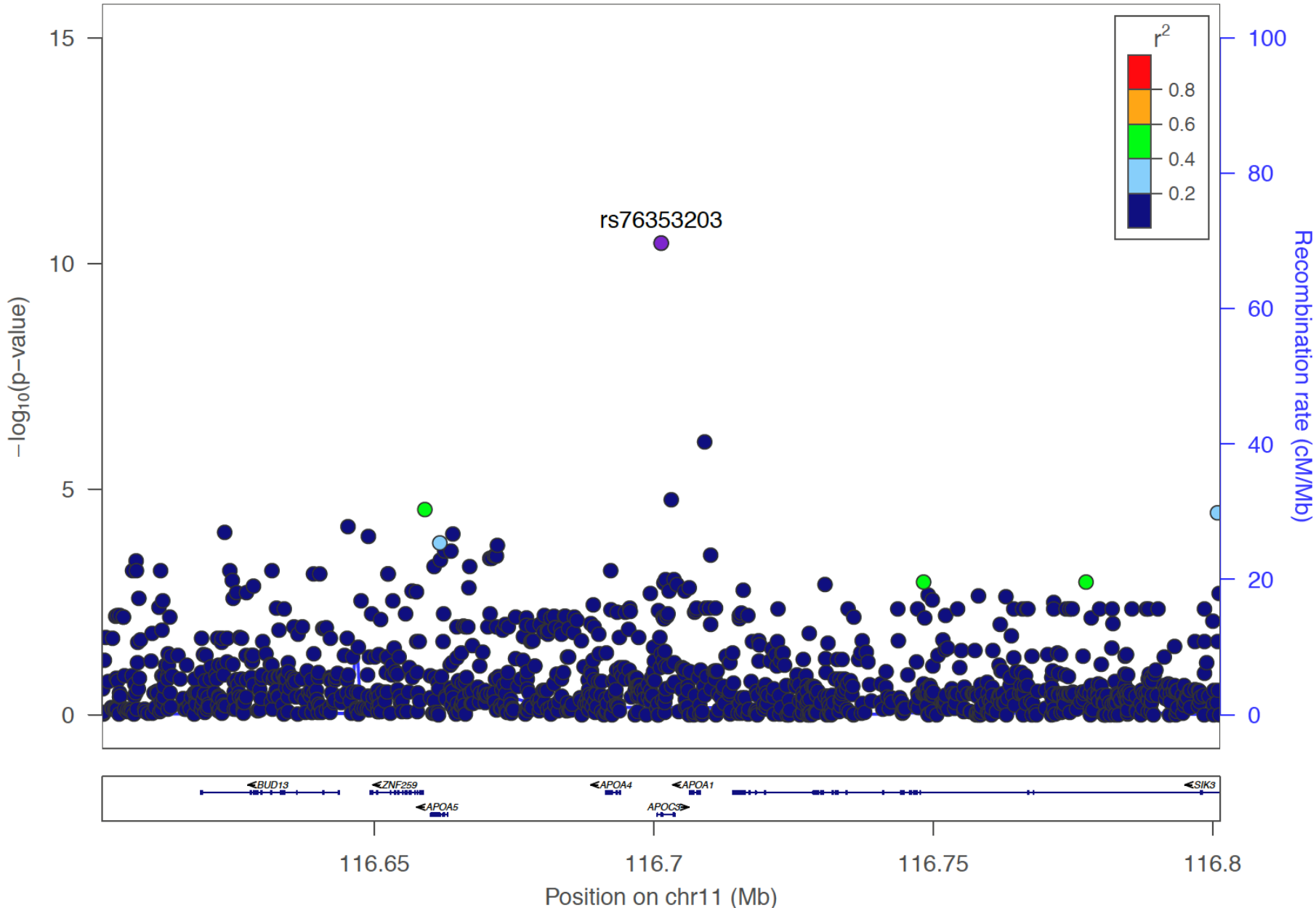
n=990
1X WGS

n=249
4X WGS

n=1482
22X WGS

n=5
WES

Cardioprotective R19X variant in *APOC3*
Imputed OmniExpress/ExomeChip data



MANOLIS

n=1265
Omni Express

n=1265
Exome Chip

n=211
Core Exome

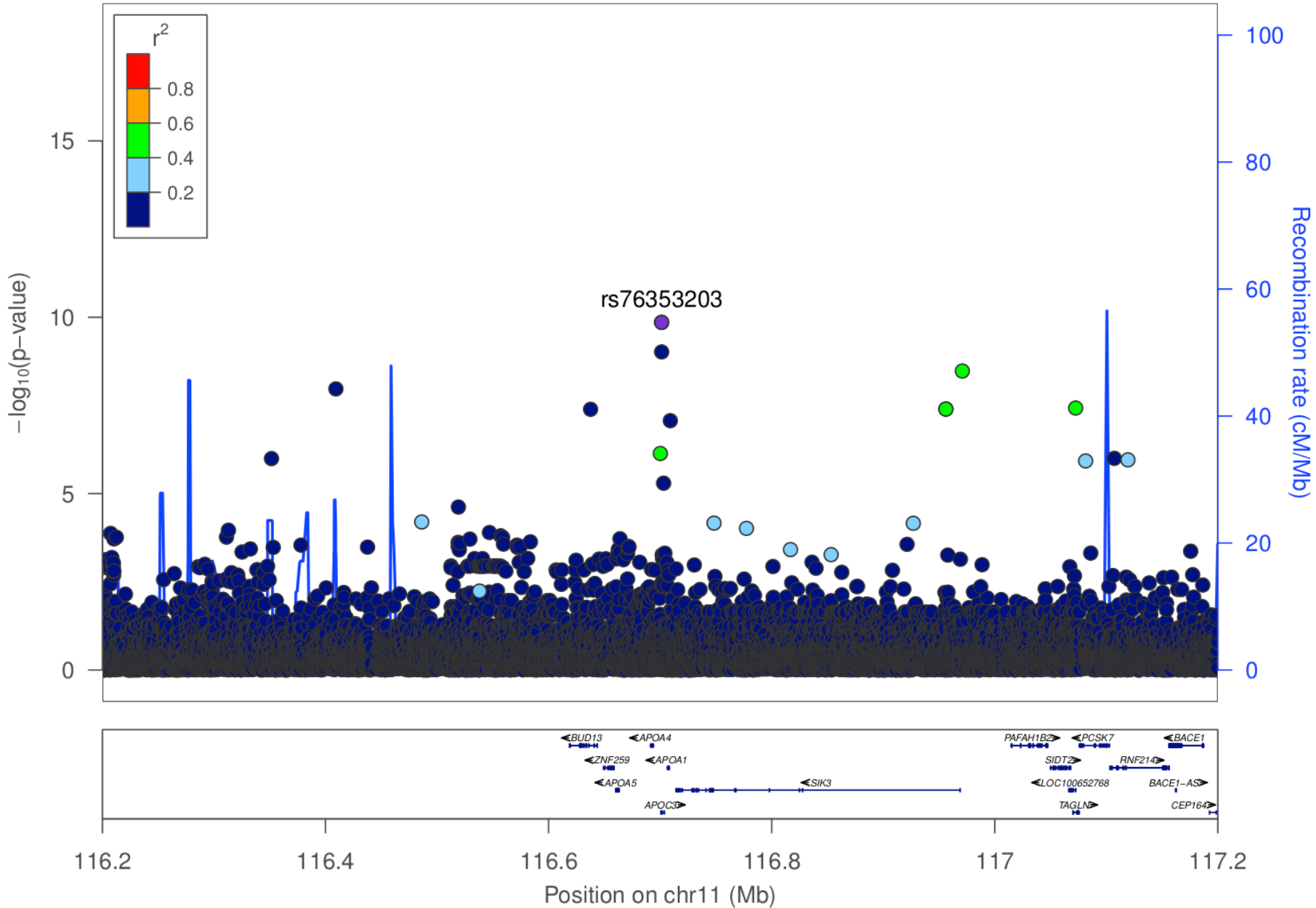
n=990
1X WGS

n=249
4X WGS

n=1482
22X WGS

n=5
WES

Cardioprotective R19X variant in *APOC3*
1x WGS data



MANOLIS

n=1265
Omni Express

n=1265
Exome Chip

n=211
Core Exome

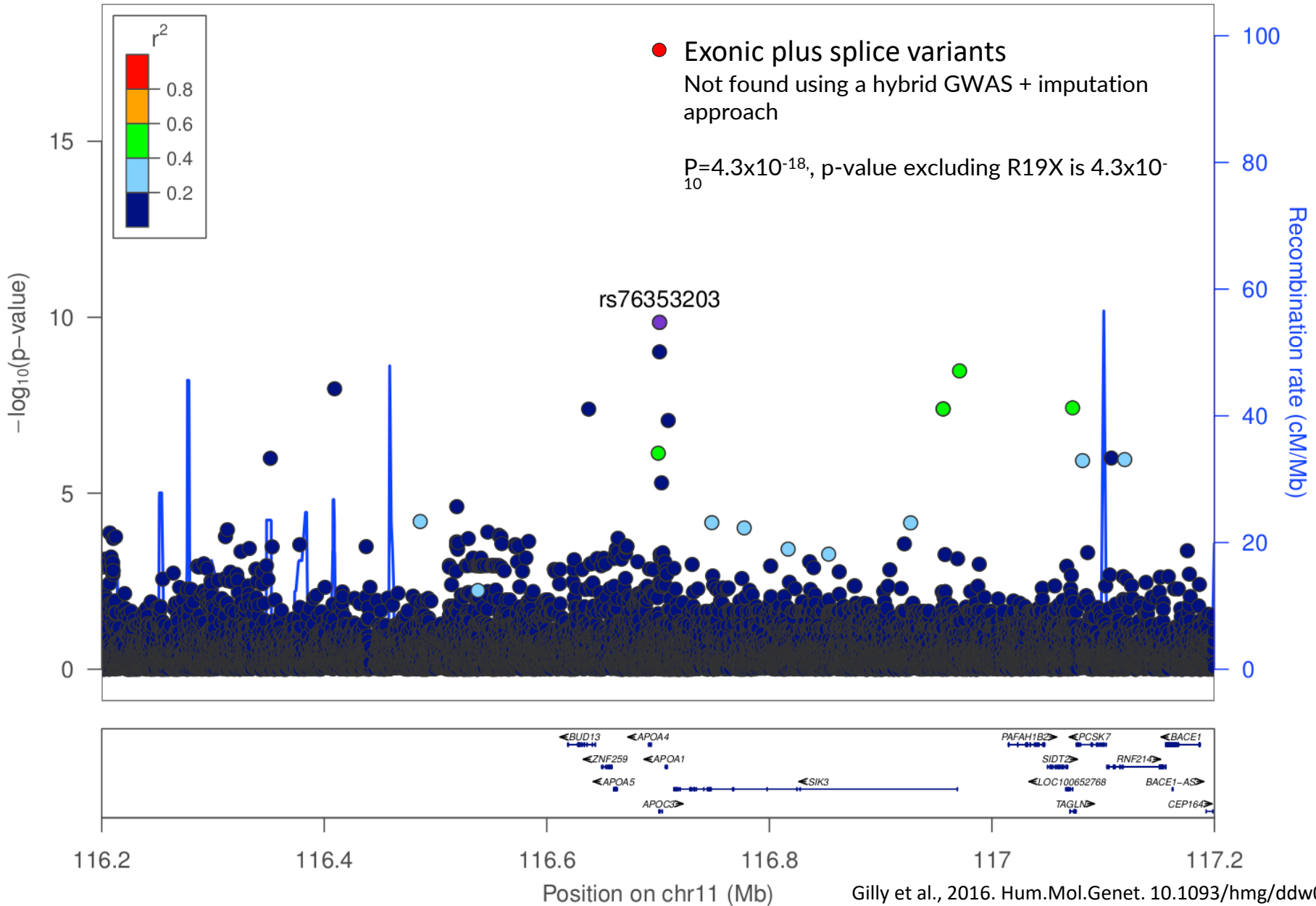
n=990
1X WGS

n=249
4X WGS

n=1482
22X WGS

n=5
WES

Cardioprotective R19X variant in *APOC3*
1x WGS data

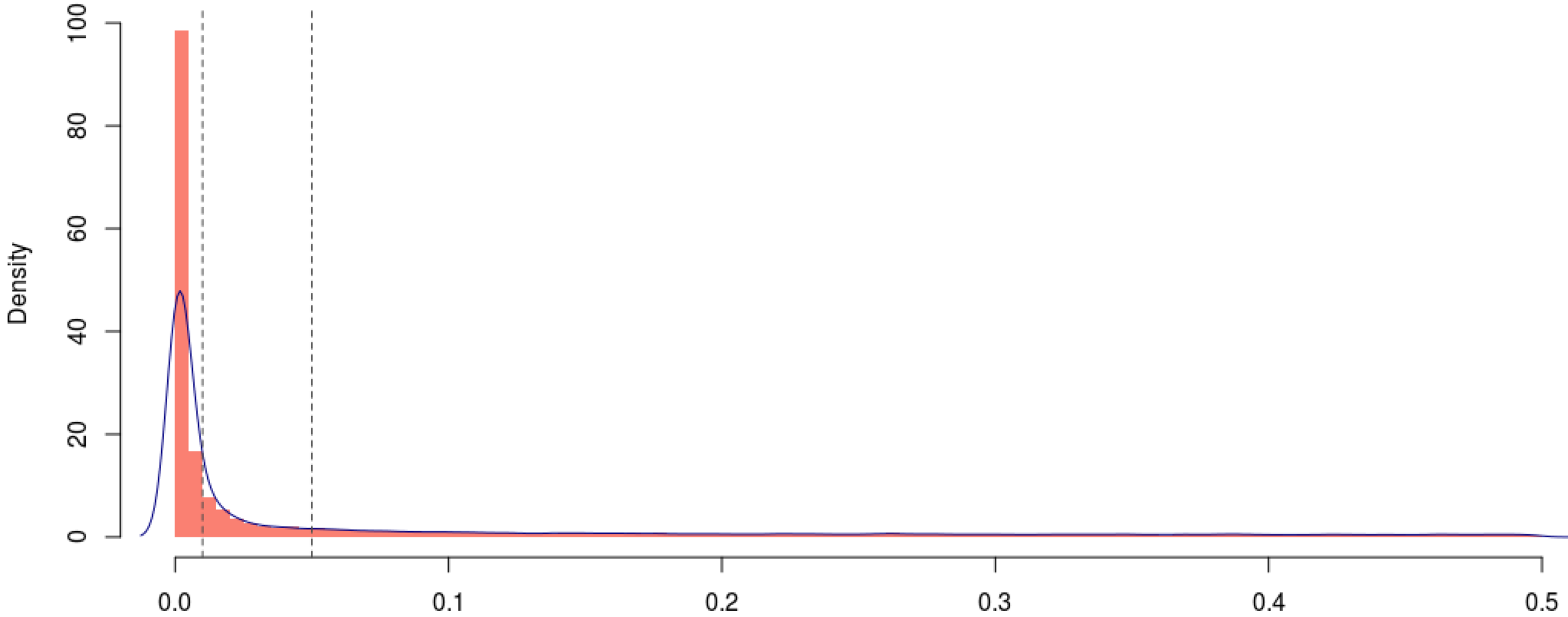


Gilly et al., 2016. Hum.Mol.Genet. 10.1093/hmg/ddw088

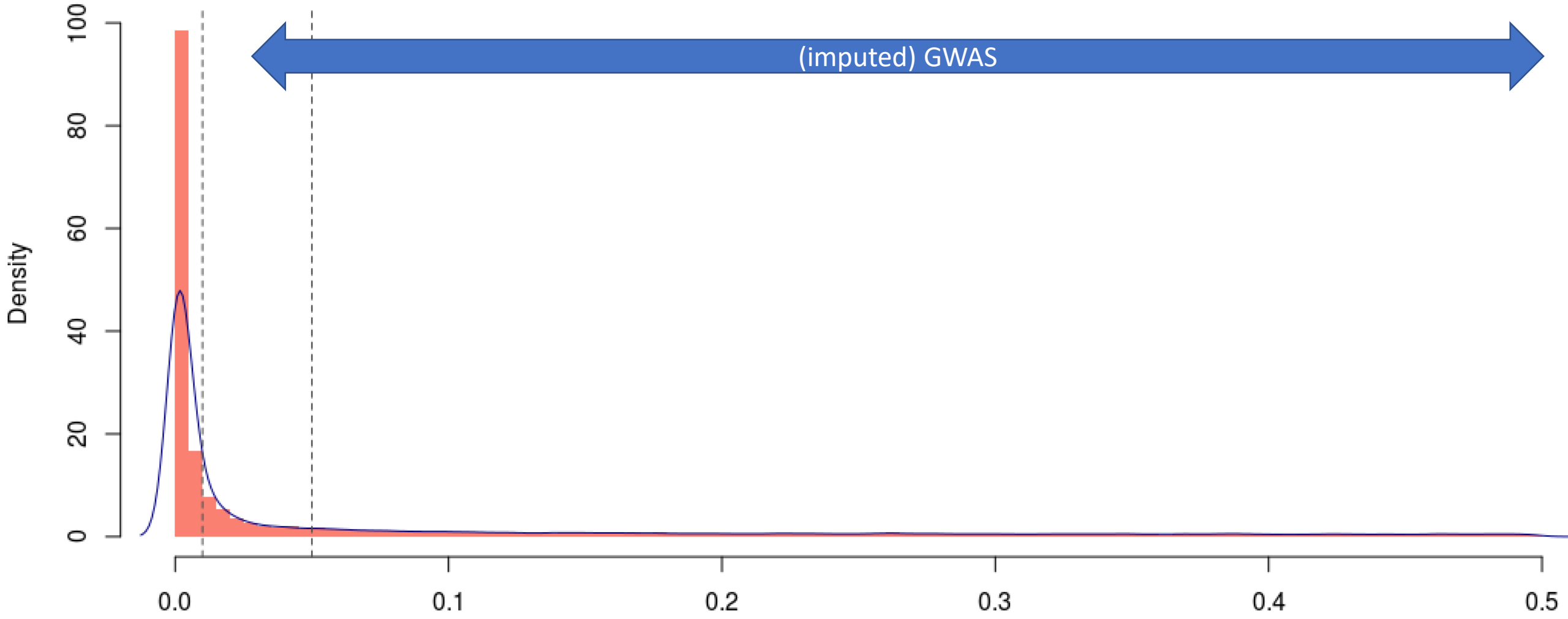
Rare variant aggregation tests

High-depth sequencing (18-22x WGS)

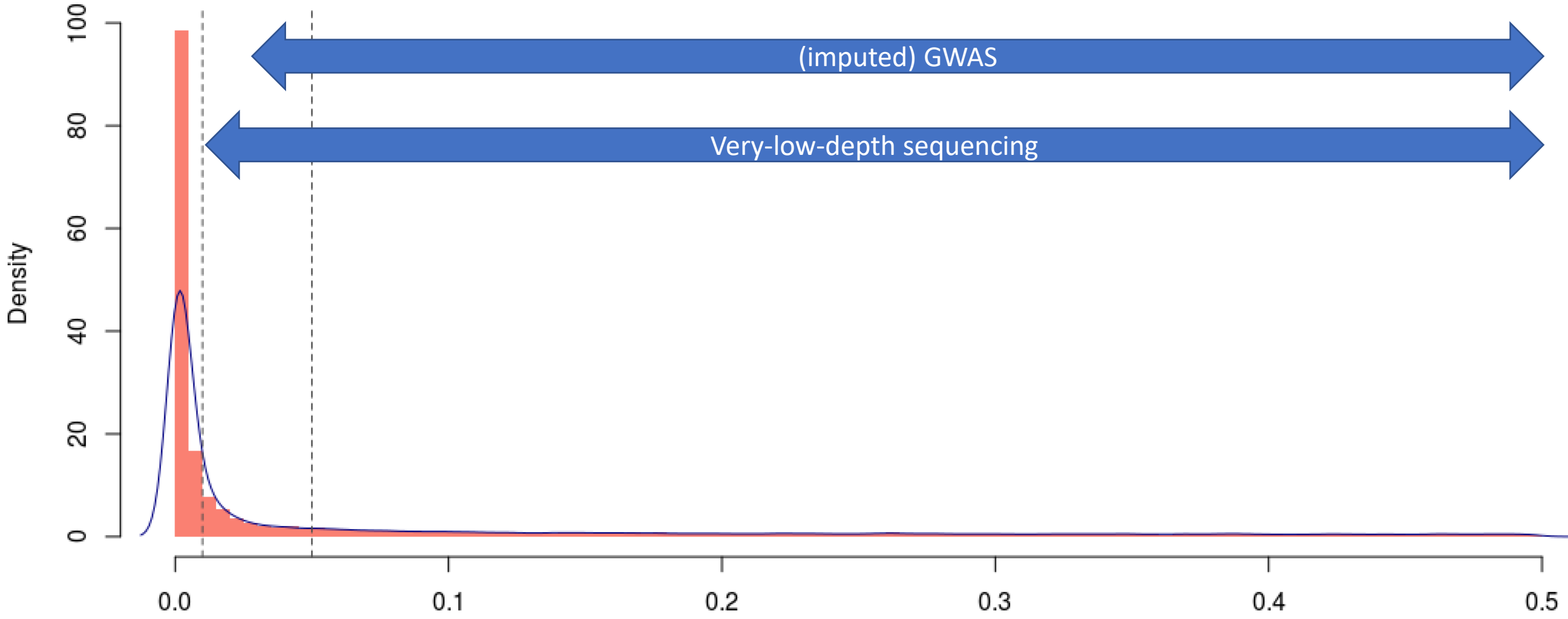
Allele frequencies



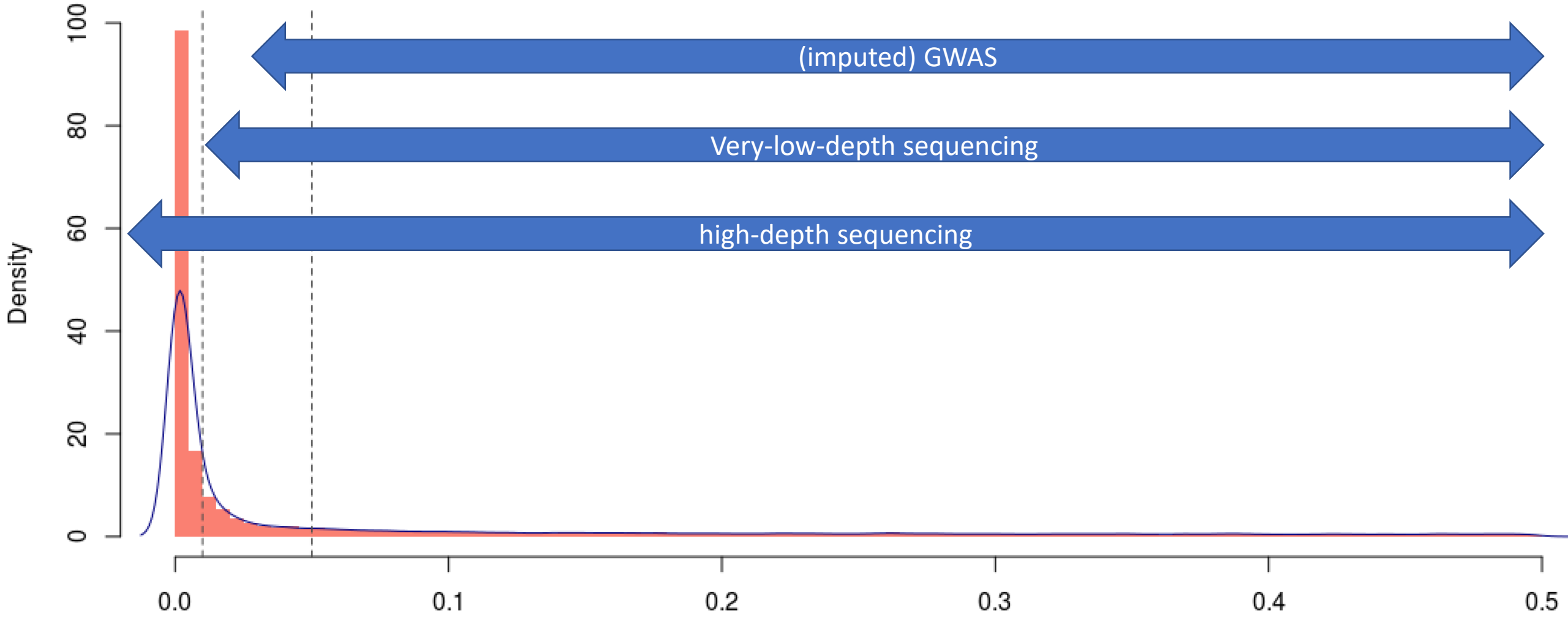
Allele frequencies



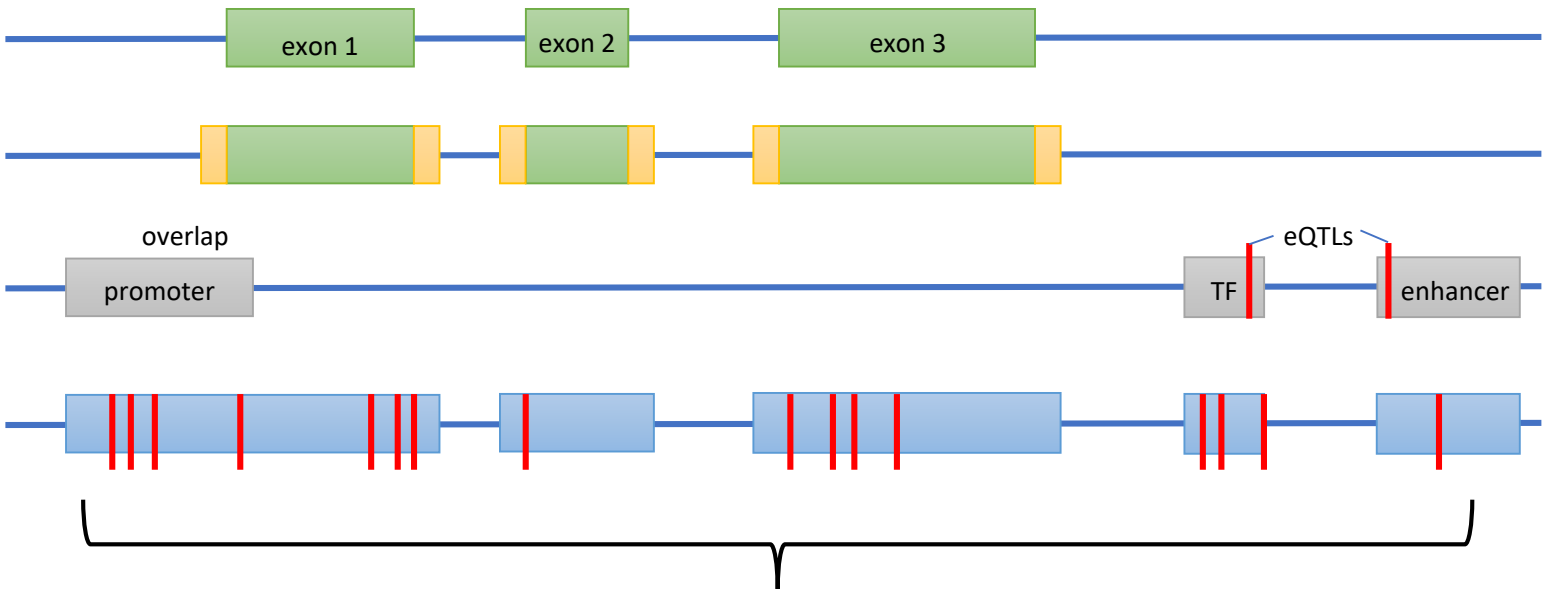
Allele frequencies



Allele frequencies



Selecting variants for burden testing:



Selecting GENCODE features eg. exons

Extending GENCODE features

Selecting associated reg. features

Select overlapping variants

Pooling variants and apply further filters:

- Upper MAF threshold: 5%
- Upper missingness threshold: 1% (remaining missingness is imputed)
- Optional selection of variants based on predicted consequence

If variants are included, which scoring method should we use?

MANOLIS

n=1265
Omni Express
n=1265
Exome Chip
n=211
Core Exome
n=990
1X WGS
n=249
4X WGS
n=1482
22X WGS
n=5
WES

Method:

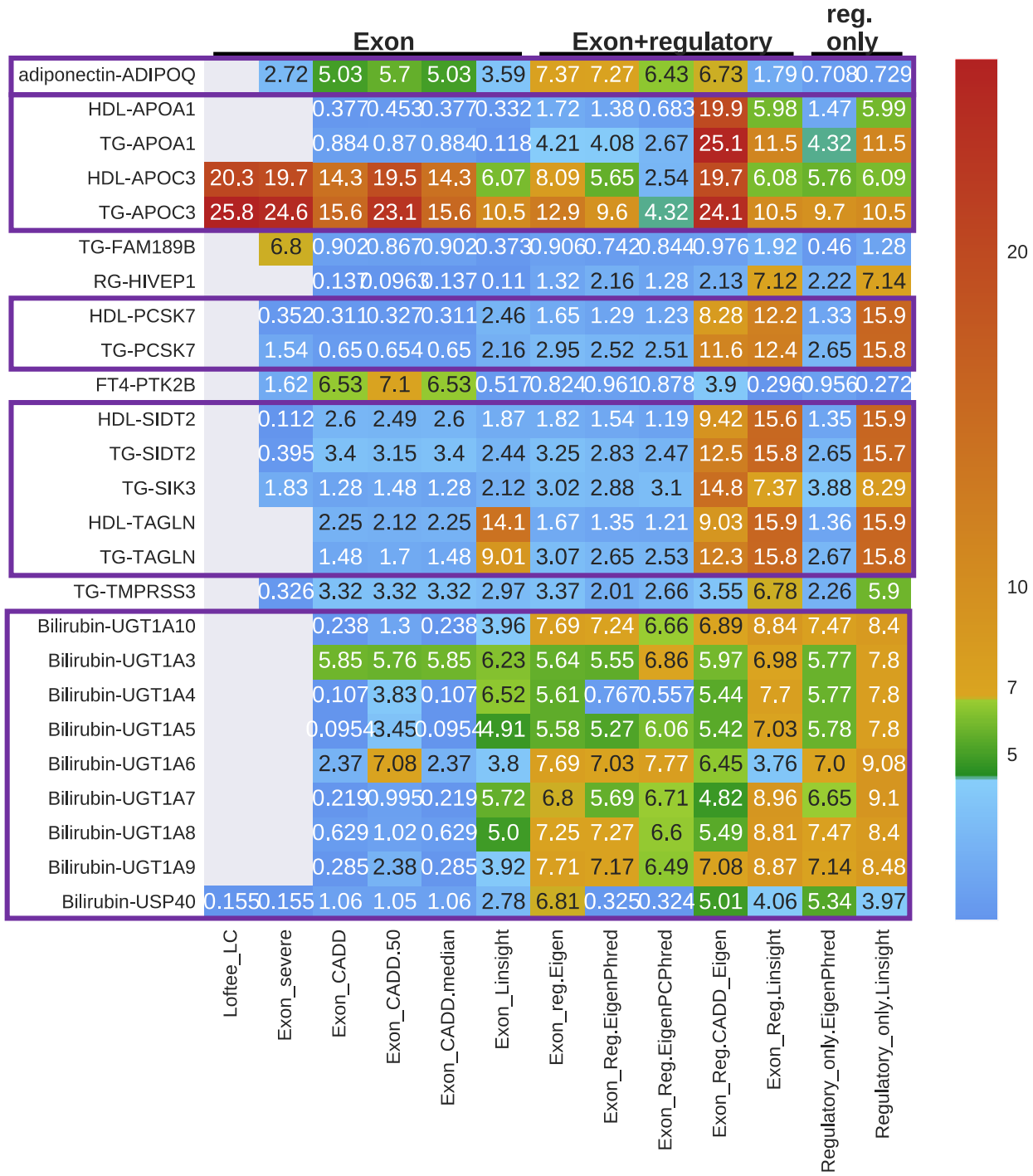
- MONSTER (SKAT-O-type unified test, with empirical relatedness matrix)
- Also accepts family information

Phenotypes:

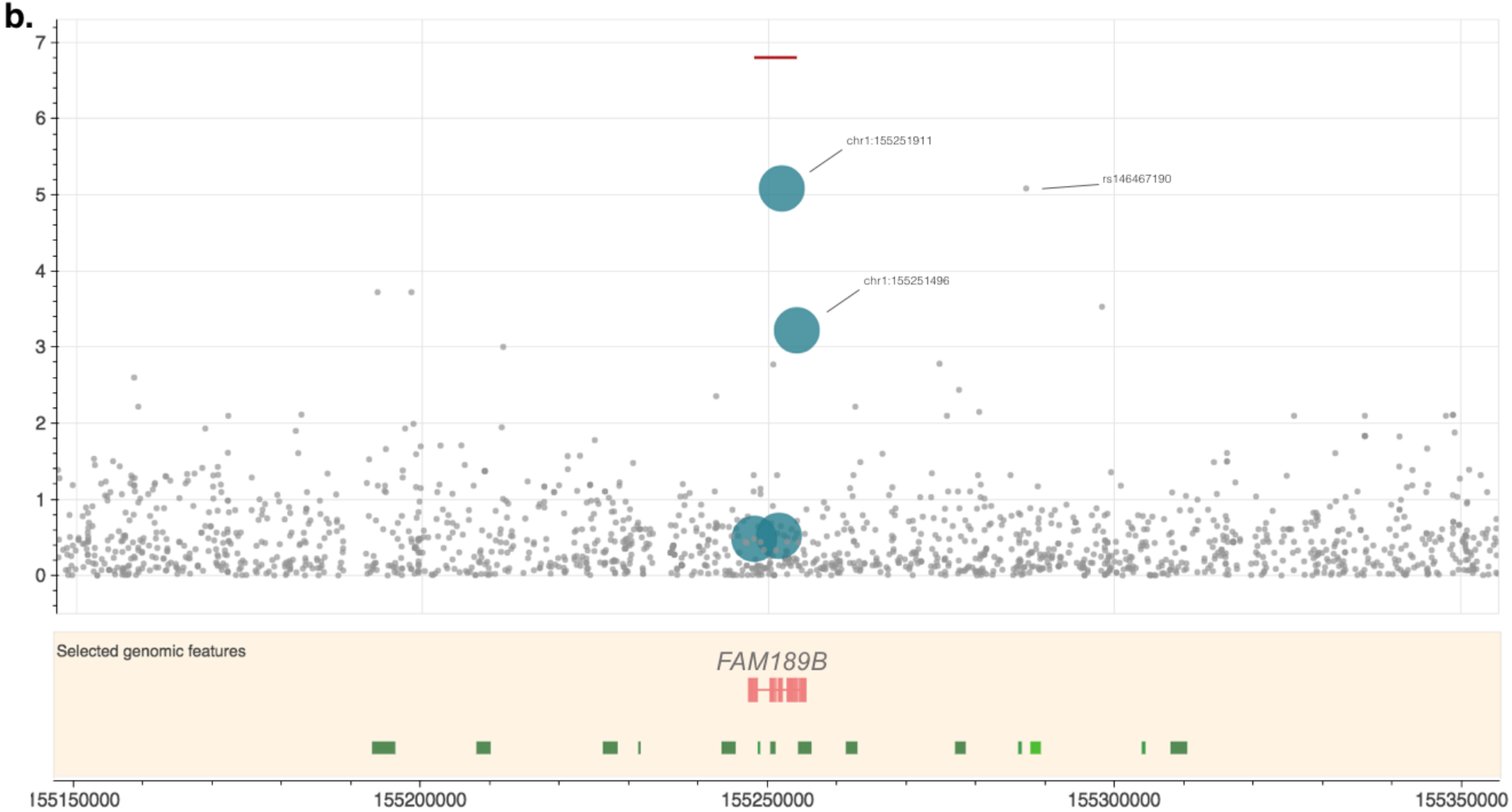
- 59 quantitative traits

Results:

- 14 conditions
- At least 1 run is SW significant ($p<1.3x10^{-7}$)
  - 39 trait/gene pairs
  - 24 unique genes

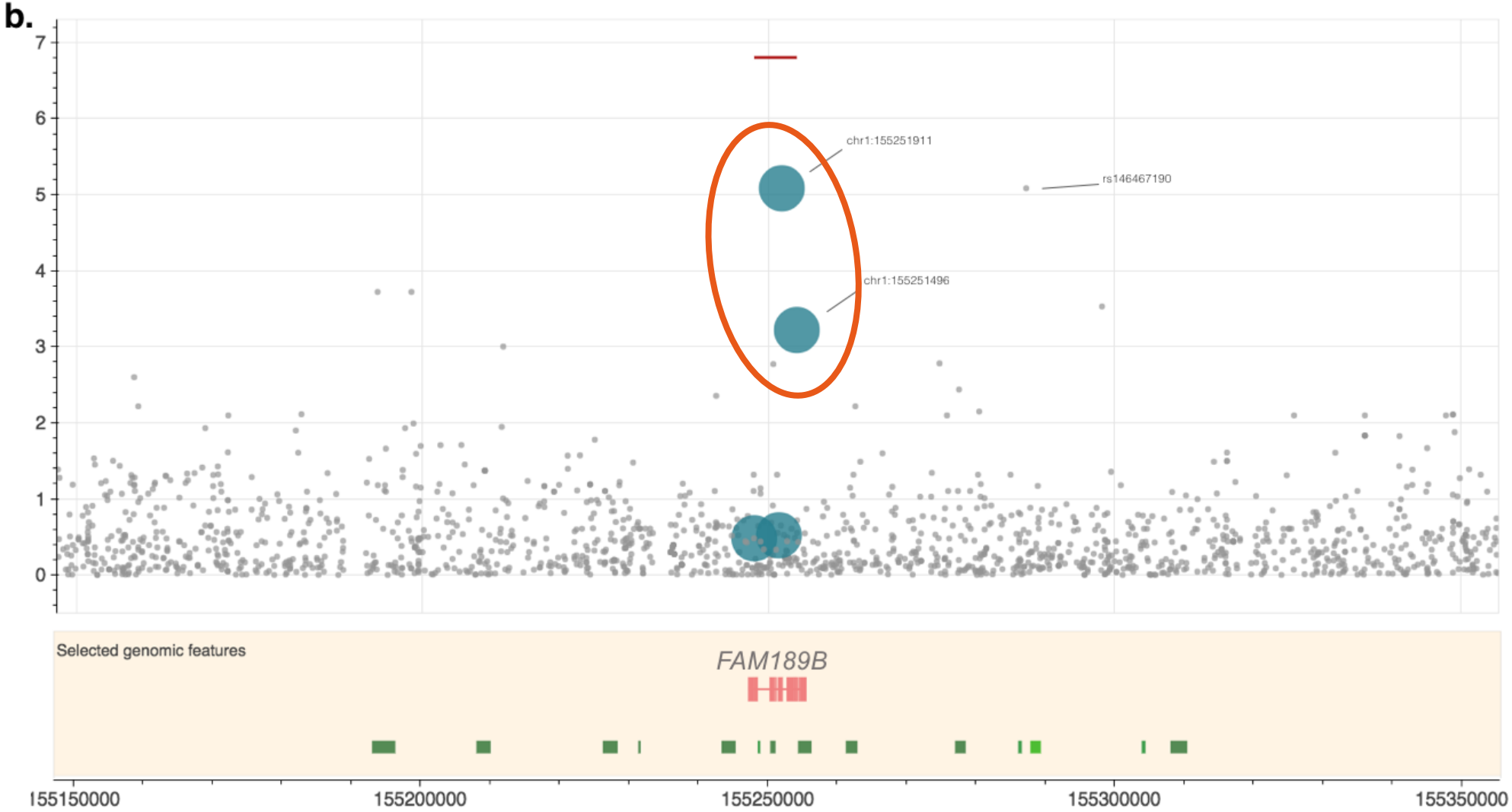


A burden driven by isolate-specific rare variants



Gilly, Suveges, Kuchenbaecker et al. doi: 10.1101/283481

A burden driven by isolate-specific rare variants



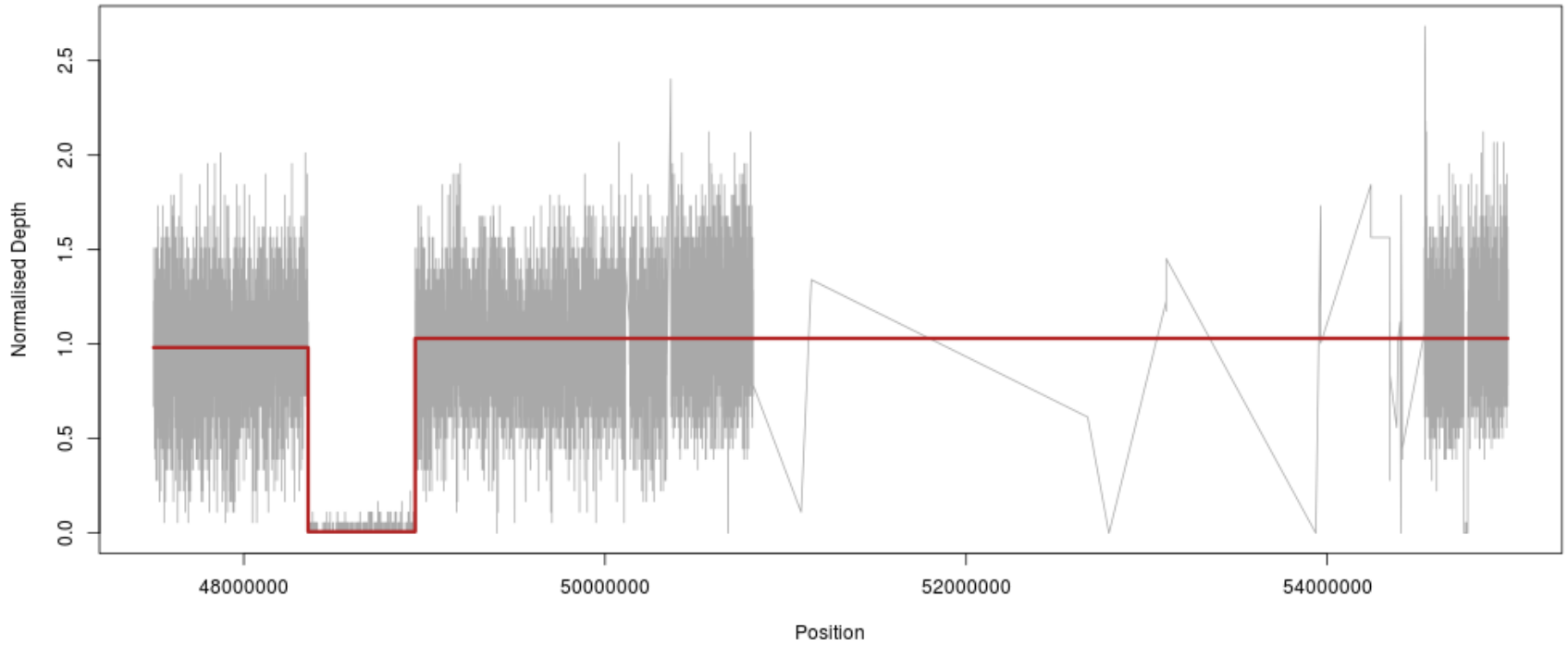
Gilly, Suveges, Kuchenbaecker et al. doi: 10.1101/283481

Large CNV calling from high-depth WGS variant calls

UN-CNVC (github.com/agilly/un-cnvc)

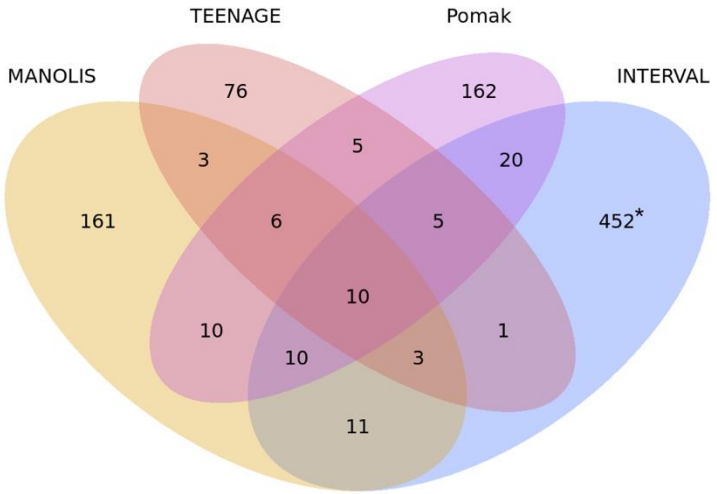
Large CNV calling based on WGS variant calls

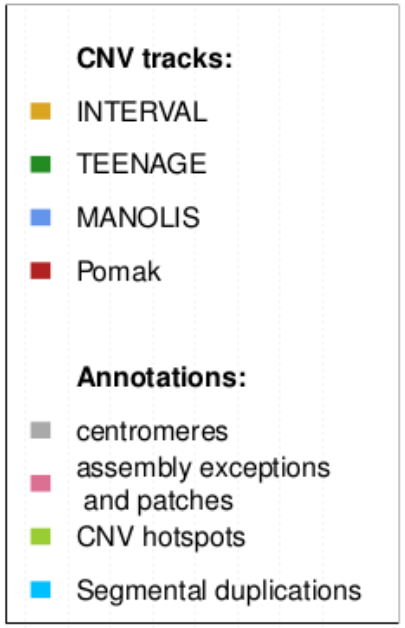
- UN-CNVc (written in R)
- Uses piecewise constant regression (regression trees) to smooth out noise and find “segments” of “constant” depth
- Uses only marker-level data

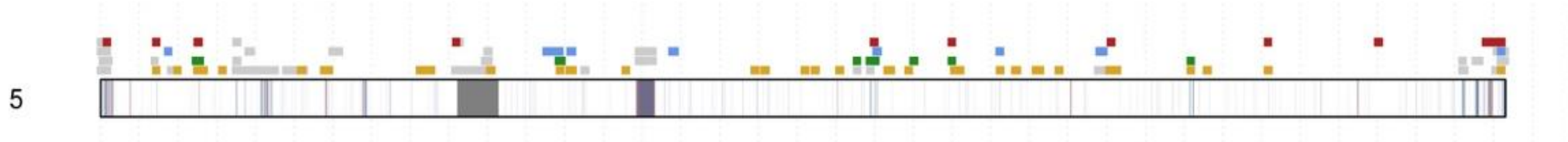


Large CNV calling based on WGS variant calls

- The proportion of CNVs private to the isolates is similar to that in a cosmopolitan cohort
- 37% of shared CNVs had significant frequency differences between cohorts, mostly between isolates and cosmopolitan populations

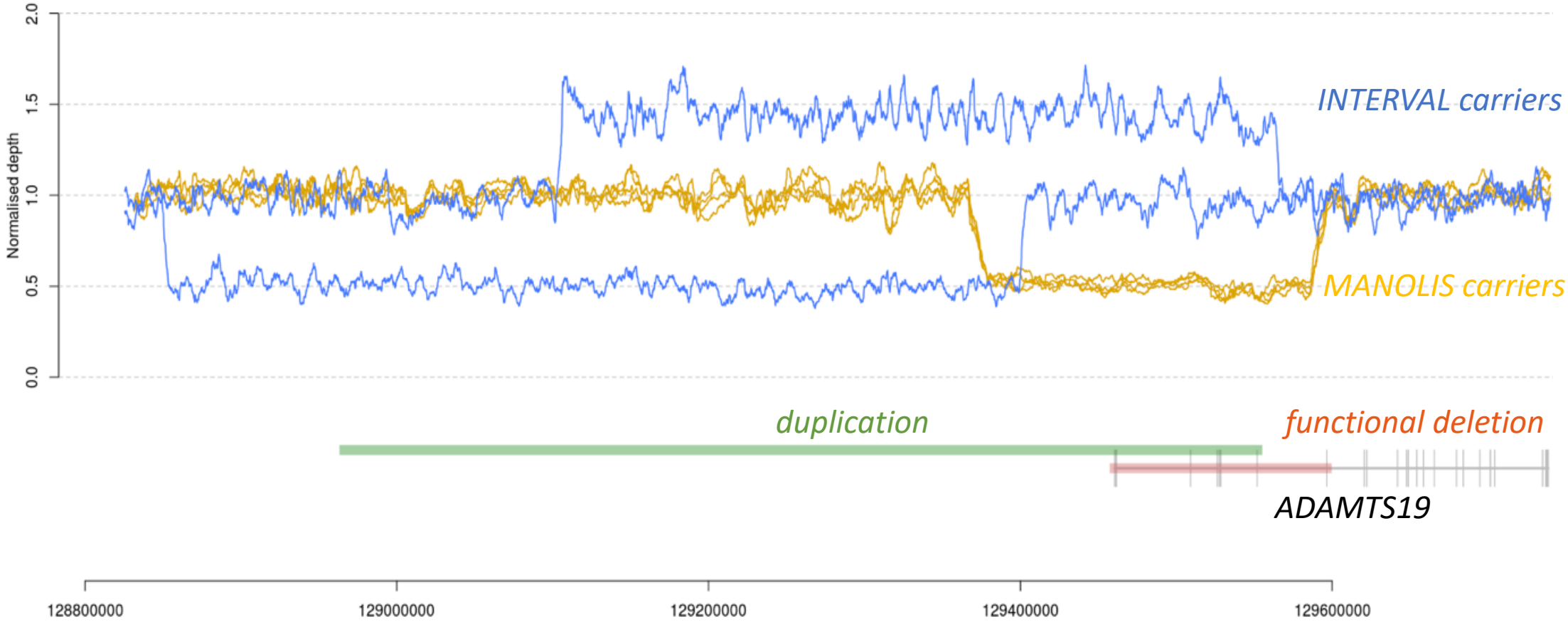






Population-specific structure in the *ADAMTS19 region*

- found by F. Wünemann's lab at the Cardiovascular unit at CHU Sainte Justine Research Center, Montreal
- four individuals from two consanguineous families with a rare form of cardiac valve disease
- also present in one AMR (Latino) sample in ExAc

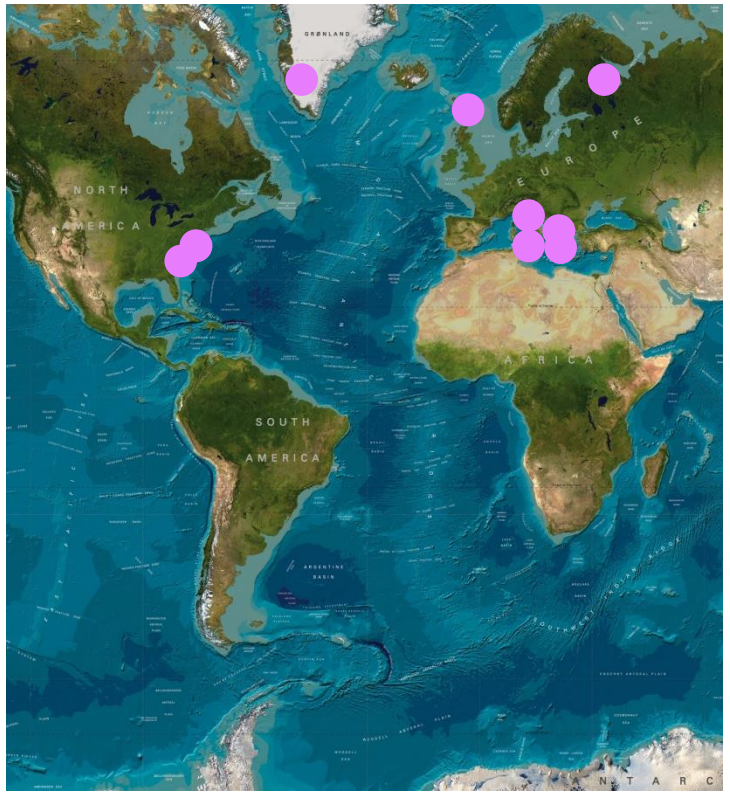


Network of isolated population cohorts

Sequencing Isolates Consortium

Over 15 well-phenotyped cohorts, including founder populations in:

- Greece (Pomak and Mylopotamos villages),
- Finland (general Finnish population cohorts and Northern Finland sub-isolates),
- Italy (Carlantino, Val Borbera and Friuli Venezia Giulia villages, Sardinia),
- UK (Orkney islands),
- USA (Amish, Ashkenazi Jewish),
- Greenland

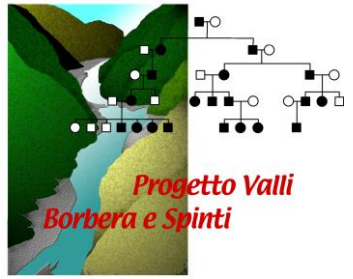


Emerging international WGS isolates consortium currently amassing in excess of 30,000 samples

University of Maryland School of Medicine

Uppsala Universitet

RGB Istituto di Ricerca Genetica e Biomedica



National Institute for Health and Welfare

Thank you!

wtsi-team144

Sample collection

Alina Farmaki, Emmanouil Tsafantakis, Maria Karaleftheri, George Dedoussis

Visiting students

Thea Bjørnland, Emil VR Appel, Elisabetta Casalone, Giorgio Melloni

Phenotype database

Britt Kilian, Nigel W. Rayner

Dániel Süveges

Grace Png