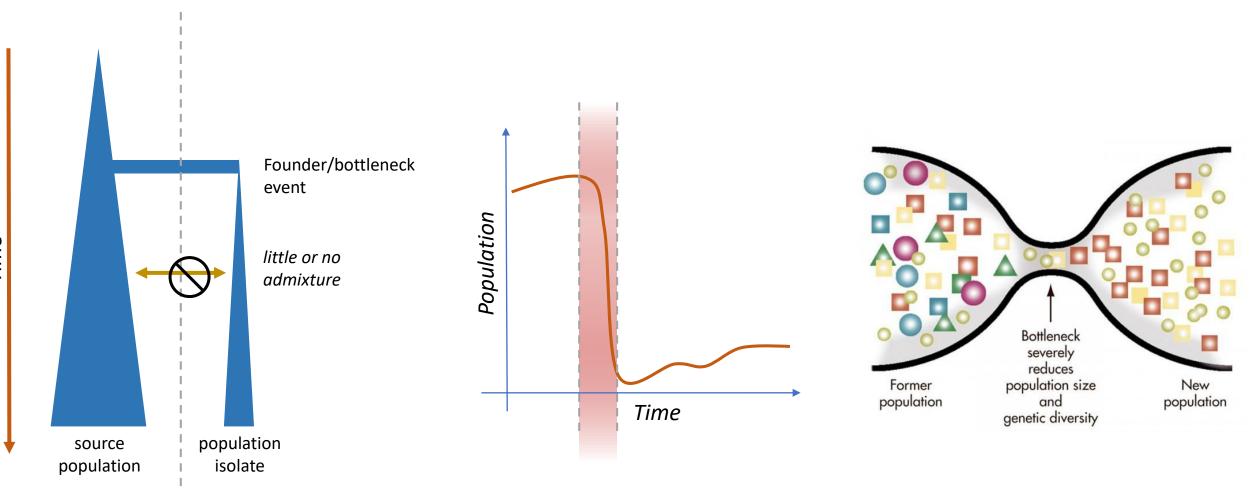
Next-generation sequencing studies in founder populations

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







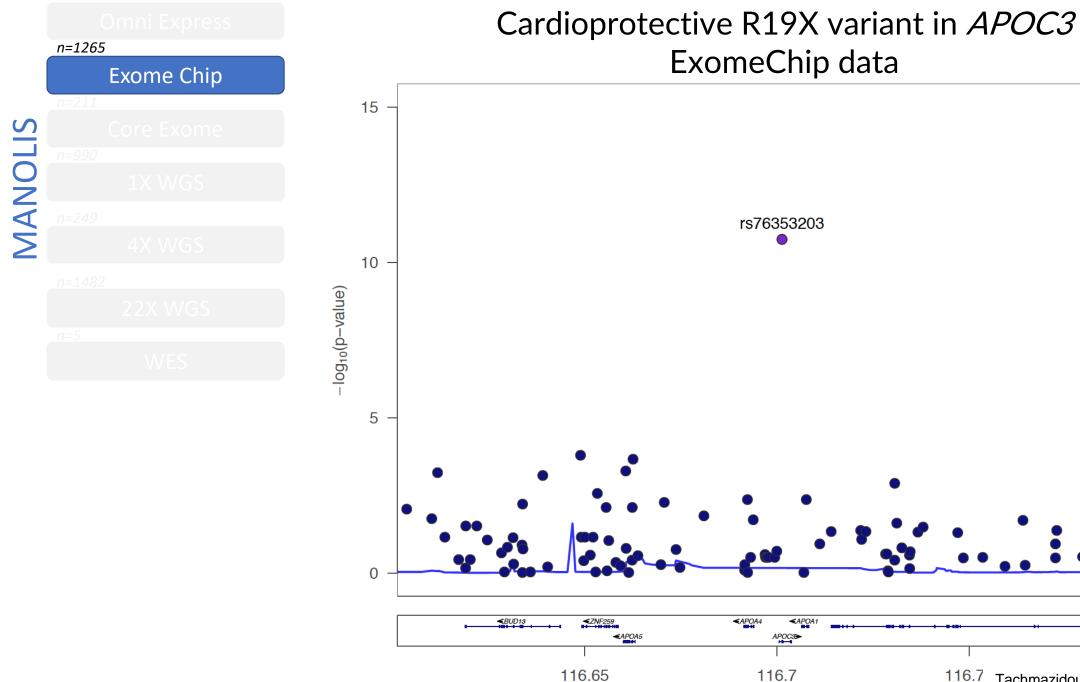
MANOLIS

HELIC datasets

Single-point association

GWAS, very-low-depth WGS

n=1265



^{116.7} Tachmazidou et al. 2013 Nat. Commun. 10.1038/ncomms3872

100

80

60

40

20

0

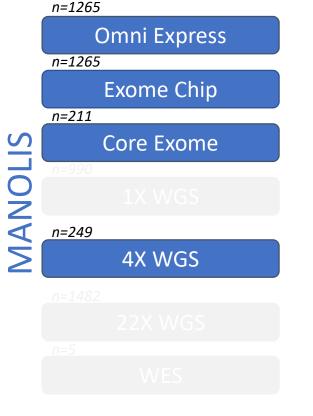
<SIK3

Recombination rate (cM/Mb)

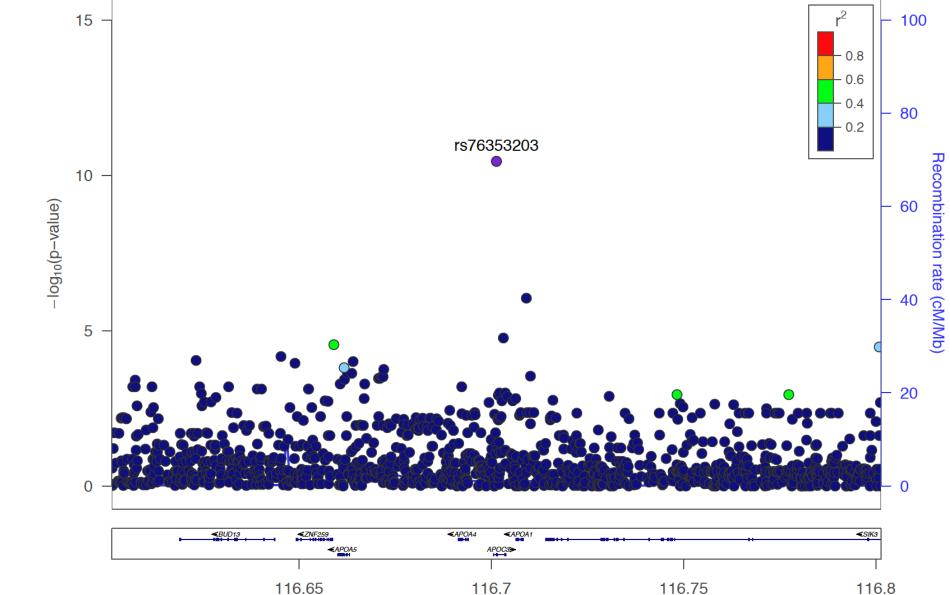
- 0.8 - 0.6 - 0.4

- 0.2

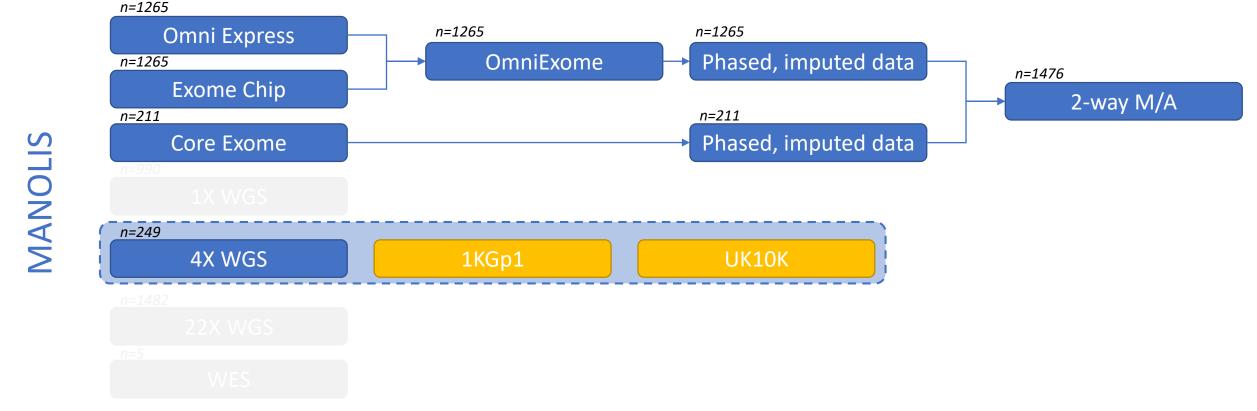
Position on chr11 (Mb)



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



Position on chr11 (Mb)

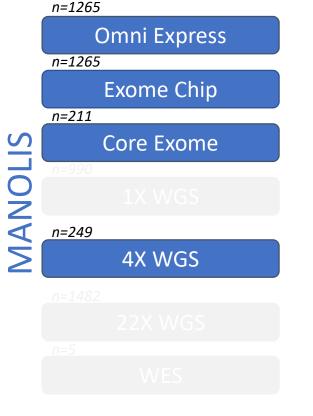


Method:

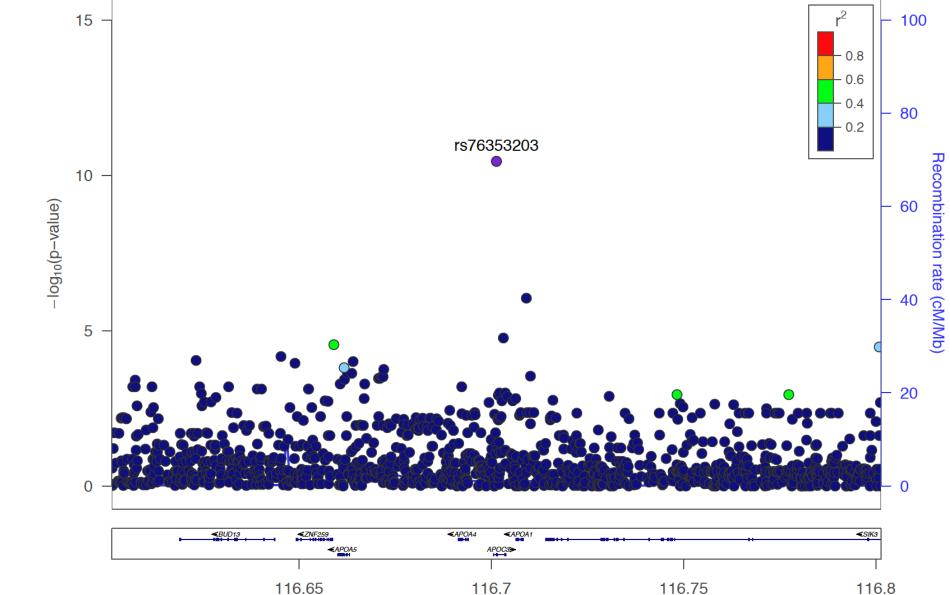
- GEMMA (with empirical relatedness matrix)
- METACARPA: meta-analysis accounting for sample relatedness

Lorraine Southam

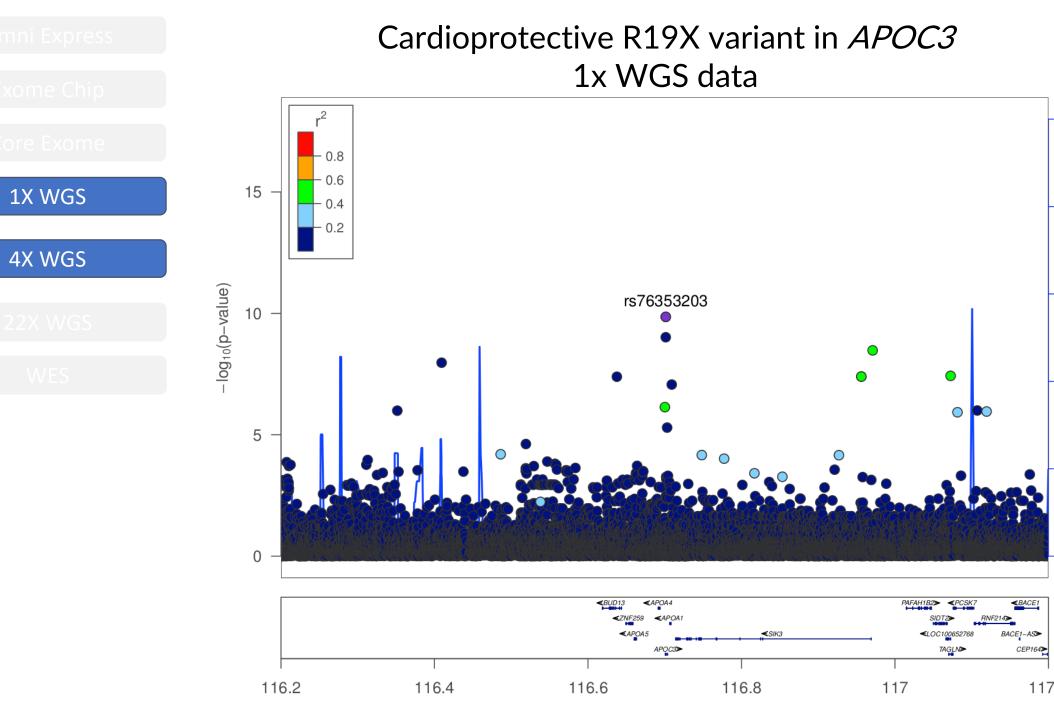
Southam, Gilly et al. Nat Comms. doi: 10.1038/ncomms15606



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



Position on chr11 (Mb)



MANOLIS

n=990

n=249

Position on chr11 (Mb)

Recombination rate (cM/Mb) 60 40

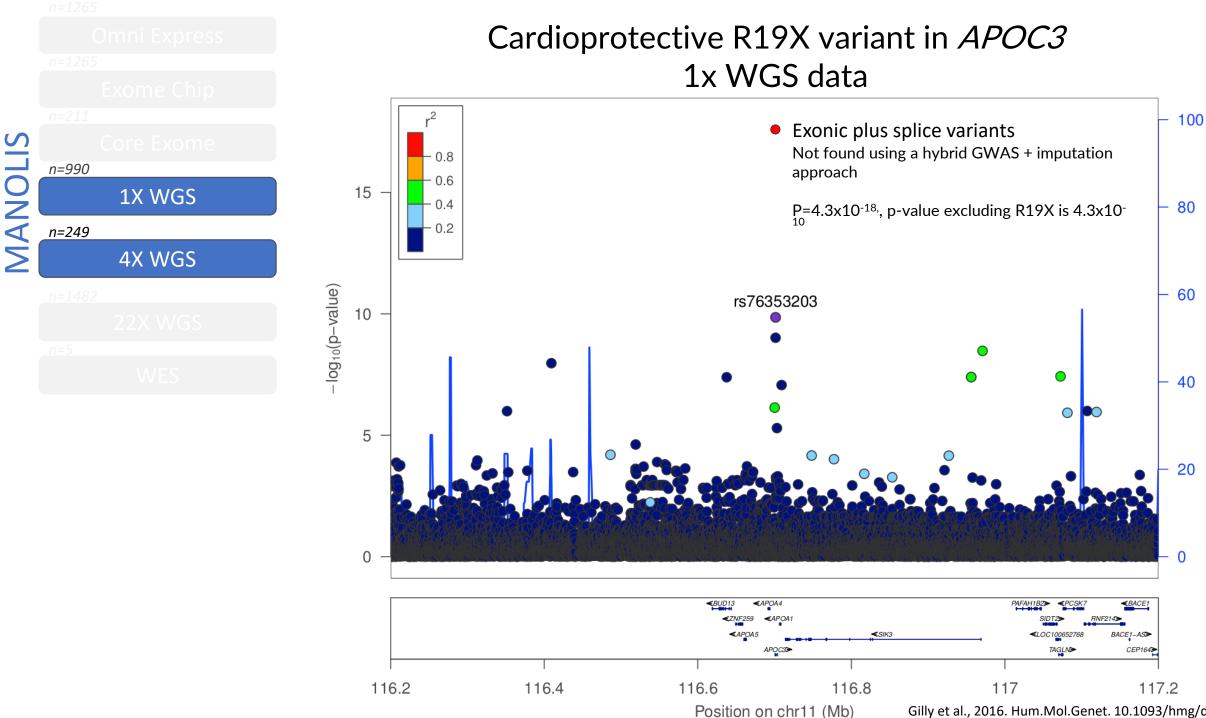
20

- 0

117.2

100

80



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

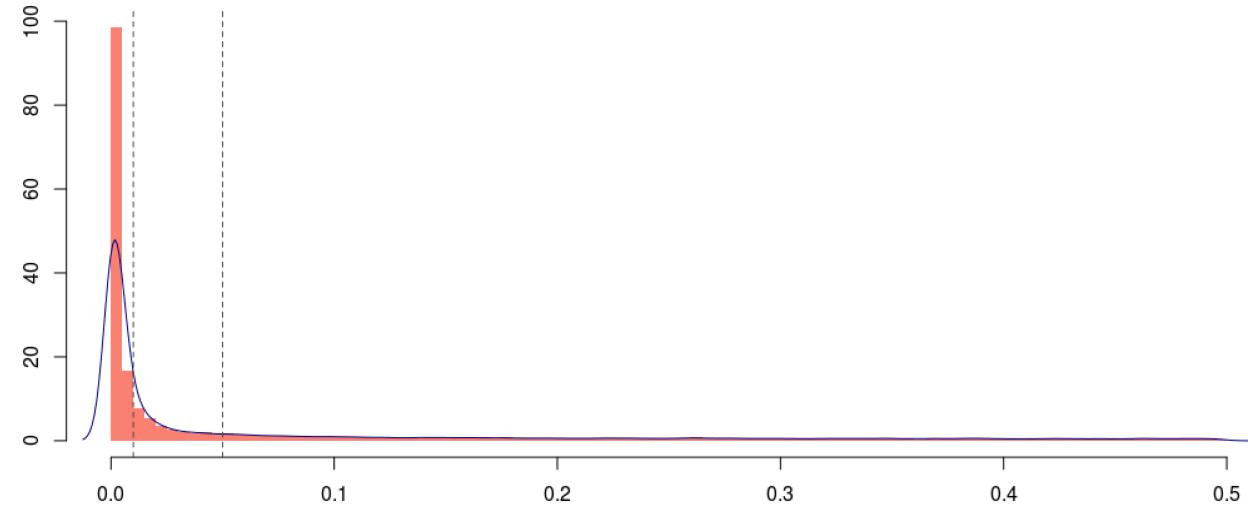
Recombination rate (cM/Mb)

Rare variant aggregation tests

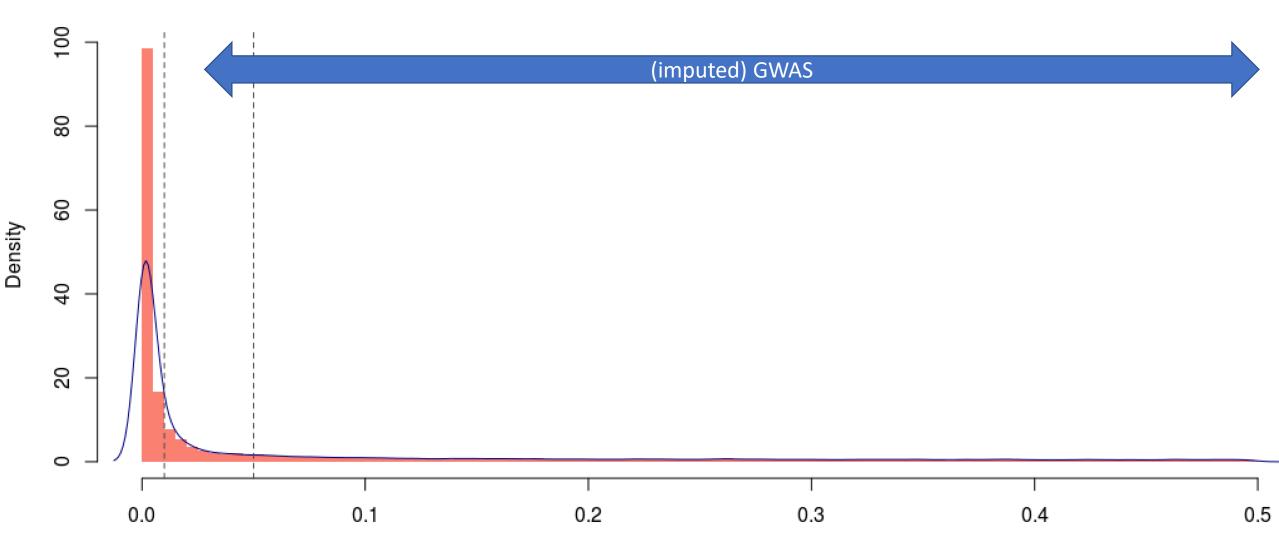
High-depth sequencing (18-22x WGS)



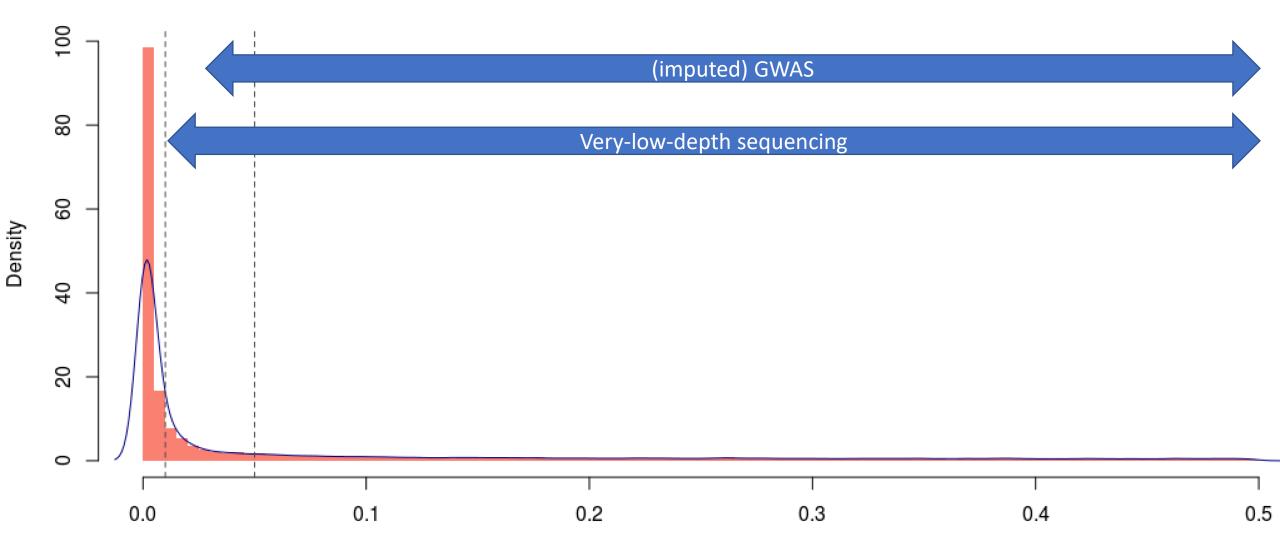
Density



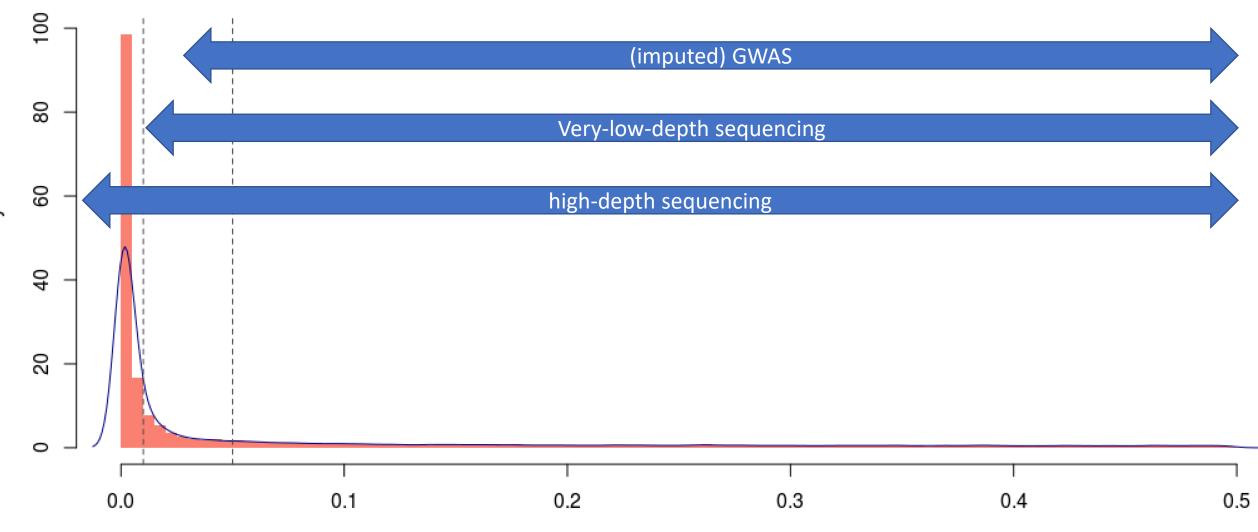
Allele frequencies



Allele frequencies

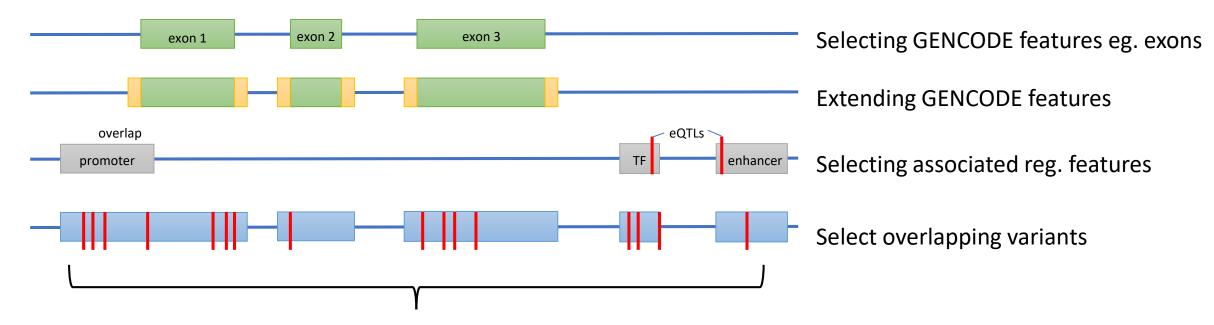


Allele frequencies



Density

Selecting variants for burden testing:



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?



Method:

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

Results:

Phenotypes:59 quantitative traits

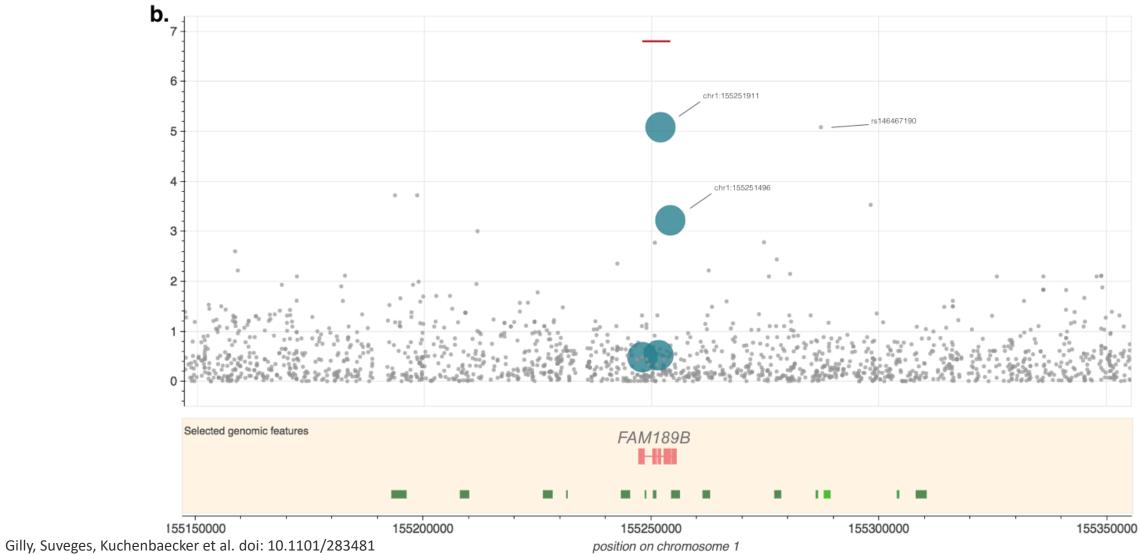
- 14 conditions
- At least 1 run is SW significant (p<1.3x10⁻⁷)
 - o 39 trait/gene pairs
 - o 24 unique genes

		Exon			Exon+regulatory						reg. only	
adiponectin-ADIPOQ	2.72	5.03	5.7	5.03	3.59	7.37	7.27	6.43	6.73	1.79	0.708	0.729
HDL-APOA1		0.377	0.453	0.377	0.332	1.72	1.38	0.683	19.9	5.98	1.47	5.99
TG-APOA1		0.884	0.87	0.884	0.118	4.21	4.08	2.67	25.1	11.5	4.32	11.5
HDL-APOC3	20.3 19.7	14.3	19.5	14.3	6.07	8.09	5.65	2.54	19.7	6.08	5.76	6.09
TG-APOC3	25.8 24.6	15.6	23.1	15.6	10.5	12.9	9.6	4.32	24.1	10.5	9.7	10.5
TG-FAM189B	6.8	0.902	0.867	0.902	0.373	0.906	0.742	0.844	0.976	1.92	0.46	1.28
RG-HIVEP1		0.137	0.096	B .137	0.11	1.32	2.16	1.28	2.13	7.12	2.22	7.14
HDL-PCSK7	0.352	20.311	0.327	0.311	2.46	1.65	1.29	1.23	8.28	12.2	1.33	15.9
TG-PCSK7	1.54	0.65	0.654	0.65	2.16	2.95	2.52	2.51	11.6	12.4	2.65	15.8
FT4-PTK2B	1.62	6.53	7.1	6.53	0.517	0.824	0.961	0.878	3.9	0.296	0.956	0.272
HDL-SIDT2	0.112	2 2.6	2.49	2.6	1.87	1.82	1.54	1.19	9.42	15.6	1.35	15.9
TG-SIDT2	0.39	5 3.4	3.15	3.4	2.44	3.25	2.83	2.47	12.5	15.8	2.65	15.7
TG-SIK3	1.83	1.28	1.48	1.28	2.12	3.02	2.88	3.1	14.8	7.37	3.88	8.29
HDL-TAGLN		2.25	2.12	2.25	14.1	1.67	1.35	1.21	9.03	15.9	1.36	15.9
TG-TAGLN		1.48	1.7	1.48	9.01	3.07	2.65	2.53	12.3	15.8	2.67	15.8
TG-TMPRSS3	0.320	3.32	3.32	3.32	2.97	3.37	2.01	2.66	3.55	6.78	2.26	5.9
Bilirubin-UGT1A10		0.238	1.3	0.238	3.96	7.69	7.24	6.66	6.89	8.84	7,47	8.4
Bilirubin-UGT1A3		5.85	5.76	5.85	6.23	5.64	5.55	6.86	5.97	6.98	5.77	7.8
Bilirubin-UGT1A4		0.107	3.83	0.107	6.52	5.61	0.767	0.557	5.44	7.7	5.77	7.8
Bilirubin-UGT1A5		0.095 [,]	43.45	0.0954	44.91	5.58	5.27	6.06	5.42	7.03	5.78	7.8
Bilirubin-UGT1A6		2.37	7.08	2.37	3.8	7.69	7.03	7.77	6.45	3.76	7.0	9.08
Bilirubin-UGT1A7		0.219	0.995	0.219	5.72	6.8	5.69	6.71	4.82	8.96	6.65	9.1
Bilirubin-UGT1A8		0.629	1.02	0.629	5.0	7.25	7.27	6.6	5.49	8.81	7.47	8.4
Bilirubin-UGT1A9		0.285	2.38	0.285	3.92	7.71	7.17	6.49	7.08	8.87	7.14	8.48
Bilirubin-USP40	0.1550.15	51.06	1.05	1.06	2.78	6.81	0.325	0.324	5.01	4.06	5.34	3.97
	Loftee_LC Exon_severe	Exon_CADD	Exon_CADD.50	Exon_CADD.median	Exon_Linsight	Exon_reg.Eigen	Exon_Reg.EigenPhred	Exon_Reg.EigenPCPhred	Exon_Reg.CADD_Eigen	Exon_Reg.Linsight	Regulatory_only.EigenPhred	Regulatory_only.Linsight

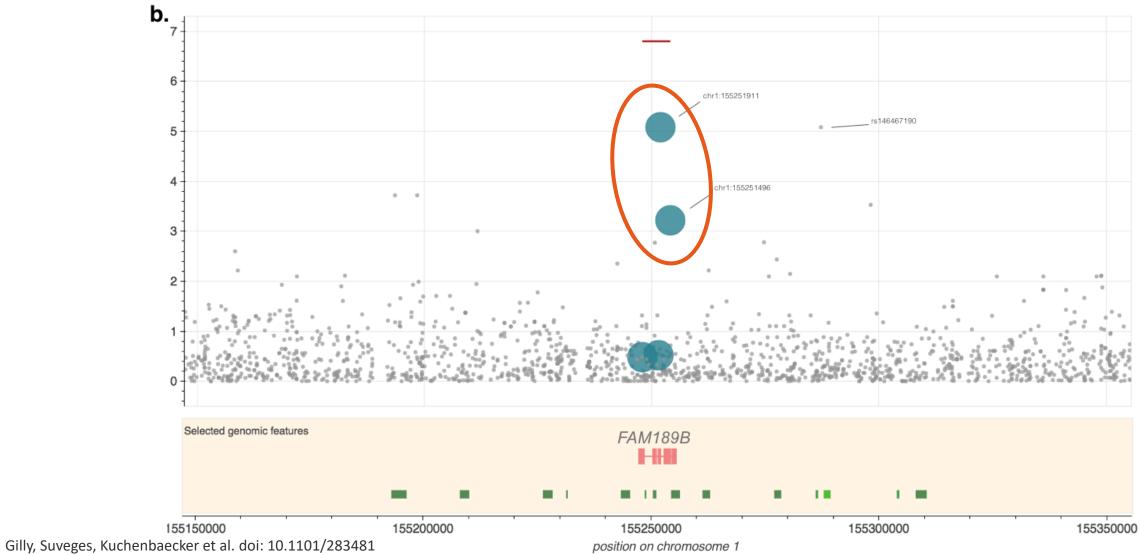
20

5

A burden driven by isolate-specific rare variants



A burden driven by isolate-specific rare variants

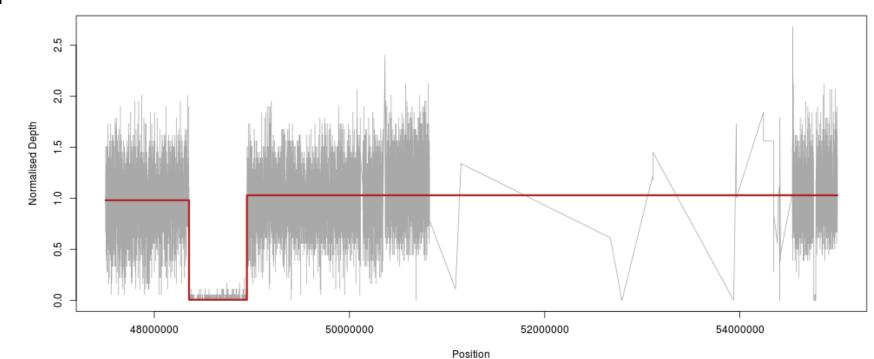


Large CNV calling from highdepth WGS variant calls

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

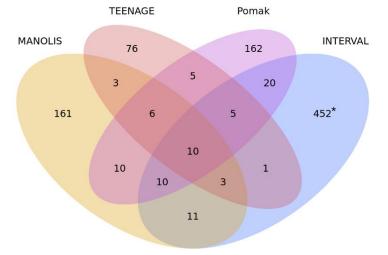
Large CNV calling based on WGS variant calls

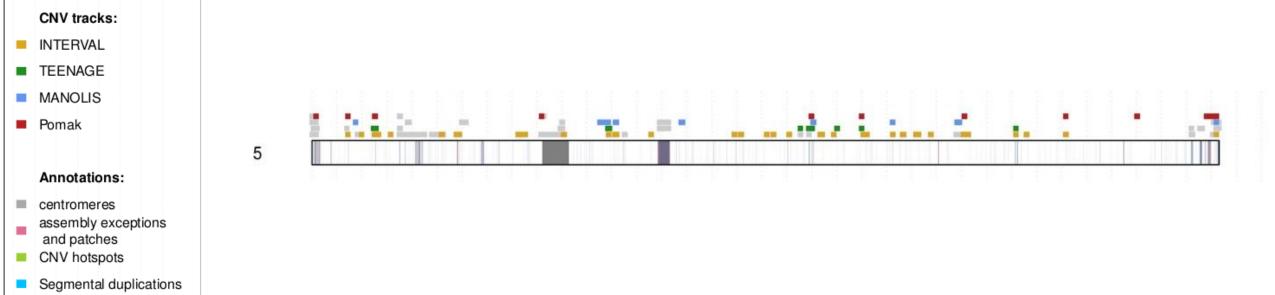
- UN-CNVc (written in R)
- Uses piecewise constant regression (regression trees) to smooth out noise and find "segments" of "constant" depth
- Uses only marker-level data



Large CNV calling based on WGS variant calls

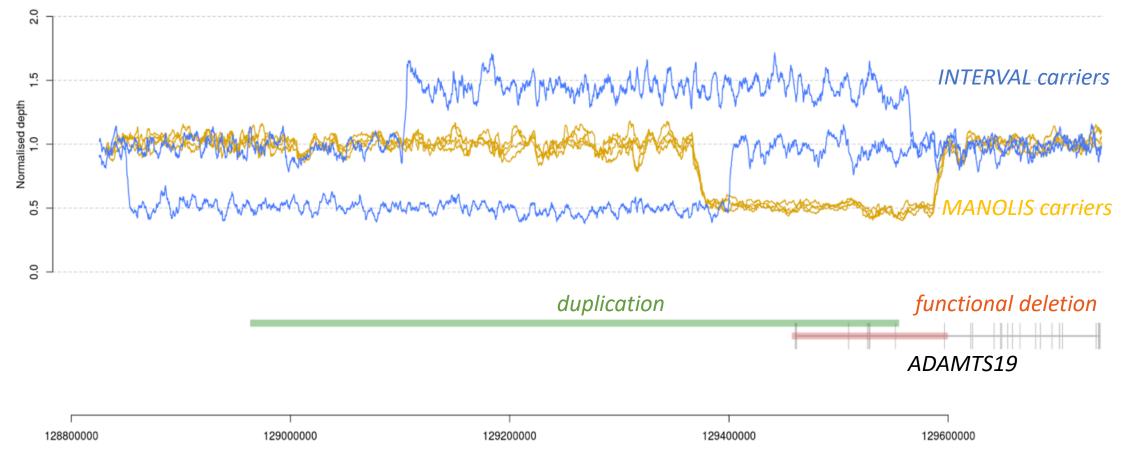
- The proportion of CNVs private to the isolates is similar to that in a cosmopolitan cohort
- 37% of shared CNVs had significant frequency differences between cohorts, mostly between isolates and cosmopolitan populations





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



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



Greece (Pomak and Mylopotamos villages),



Finland (general Finnish population cohorts and Northern Finland sub-isolates),



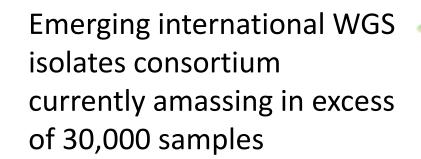
Italy (Carlantino, Val Borbera and Friuli Venezia Giulia villages, Sardinia),



UK (Orkney islands),



USA (Amish, Ashkenazi Jewish),

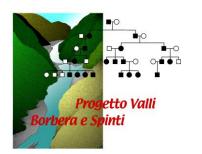












NATIONAL INSTITUTE FOR HEALTH AND WELFARE

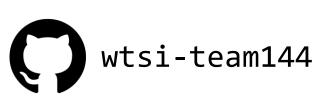




Greenland

<u>Sample collection</u> Alina Farmaki, Emmanouil Tsafantakis, Maria Karaleftheri, George Dedoussis

Thank you!



<u>Visiting students</u> Thea Bjørnland, Emil VR Appel, Elisabetta Casalone, Giorgio Melloni

> <u>Phenotype database</u> Britt Kilian, Nigel W. Rayner

Dániel Süveges

Grace Png







