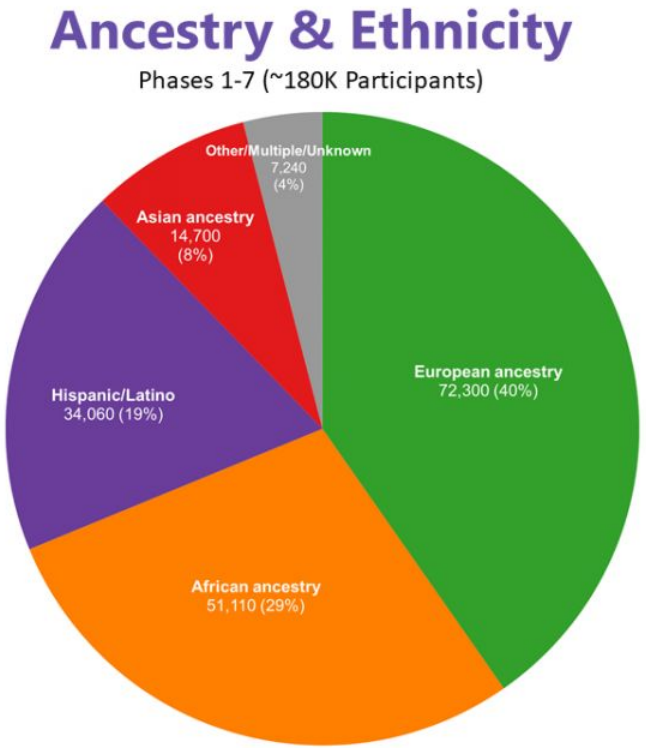
Section 6

The TOPMed Imputation Server

Albert Smith
University of Michigan
albertvs@umich.edu
@avsmith

TOPMed Program

- Trans-Omics for Precision Medicine (TOPMed) Program
- A Precision Medicine Initiative sponsored by National Heart, Lung and Blood Institute
- Integrating whole-genome sequencing and other omics data
- >180k participants from >90 studies



TOPMed Variant Call Set

Type	Category	PASS Variants	Singletons	Doubletons	AF > .0001	AF > .001	AF > .005	AF > .05
SNP	All	781M	46.4%	15.7%	4.50%	1.27%	1.06%	0.87%
	Synonymous	2.77M	42.2%	15.2%	5.23%	1.37%	1.06%	0.76%
	Missense	6.00M	46.4%	15.7%	3.96%	0.87%	0.56%	0.33%
	Stop Gain	197K	53.3%	16.0%	2.39%	0.44%	0.24%	0.12%
Indels	All	62.4M	49.7%	15.3%	4.22%	1.13%	0.90%	0.63%
	Inframe	112K	50.8%	15.5%	3.69%	0.70%	0.35%	0.16%
	Frameshift	271K	60.0%	15.5%	1.78%	0.31%	0.17%	0.09%

Stop-gain and frameshift variants progressively depleted among common variants

1/830 stop gain variants reaches MAF>5% vs. **1/115** among all SNPs, **1/303** among missense SNPs
1/1100 frameshift variants reaches MAF>5% vs. **1/159** among all Indels, **1/625** among inframe indels.

ASHG
American Society of Human Genetics

TOPMed Imputation

- Reference panel based on TOPMed Freeze 8 Calls
- Michigan Imputation Server ported to Amazon Web Services
- Released April 2020
- https://imputation.biodatacatalyst.nhlbi.nih.gov
- Registration as before, open access to TOPMed panel
  - (Michigan Imputation Server accounts *not* transferred)

TOPMed Panel

Variation type	Non-reference allele frequency bins				
	(0, 0.005]	(0.005, 0.01]	(0.01, 0.05]	(0.05, 1)	Totals
SNVs	270,352,495	3,365,284	5,330,340	7,020,861	286,068,980
Insertions	5,462,262	74,150	130,506	148,595	5,815,513
Deletions	15,406,052	185,606	297,186	333,748	16,222,592
Totals	**291,220,809**	**3,625,040**	**5,758,032**	**7,503,204**	**308,107,085**

Panel based on TOPMed Freeze 8

ASHG
American Society of Human Genetics

TOPMed Panel Compared

	TOPMed_r2	HRC	1000G Genomes
N samples	97K	39K	2,500
Ancestry	Multiethnic	European	Multiethnic
N variants	308M	39M	88M
Avg. depth	38X	8X	4X
Genome build	b38	b37	b37

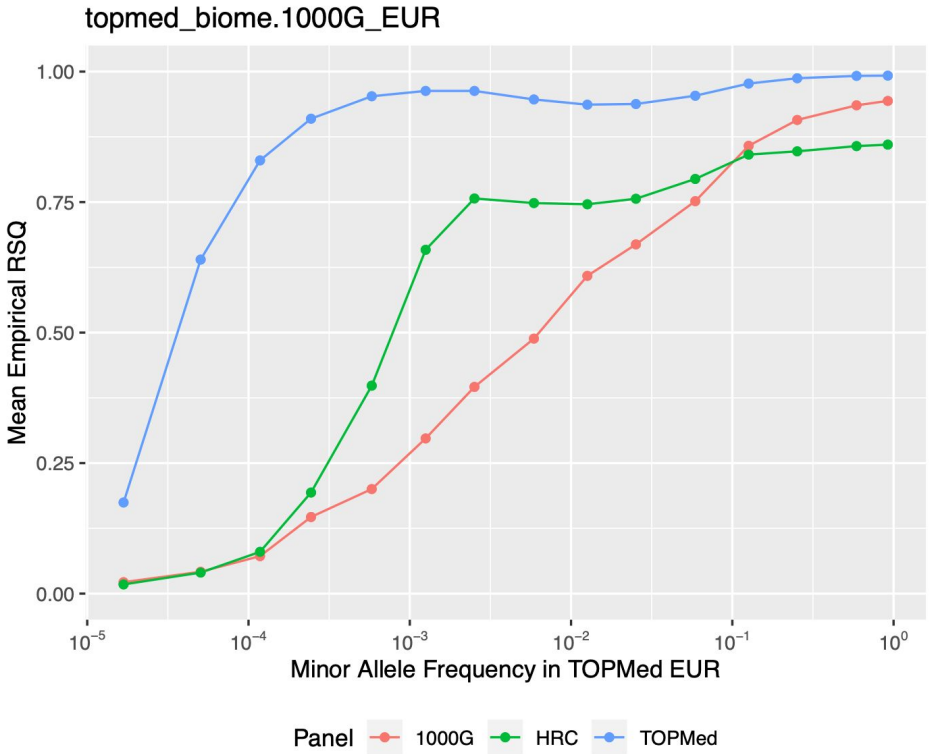
Contributing Cohorts

Cohort	Samples
Amish	1,109
ARIC	8,426
AustralianFamilialAF	120
BAGS	1,003
BioMe	11,570
CARDIA	3,449
CFS	1,291
CHS	3,528
COPDGene	10,514
DHS	376
FHS	4,146
GALAI	941
GALAII	4,663
GeneSTAR	1,759
GOLDN	942
HCHS_SOL	6,514

Cohort	Samples
HVH	696
HyperGEN	1,849
IPF	1,365
JHS	3,403
LTRC	1,388
MESA	5,347
MLOF	5,099
OMG_SCD	640
SAFS	1,776
SAGE	1,935
Sarcoidosis	633
VAFAR	173
VU_AF	1,133
walk_PHaSST	429
WHI	11,039
Total	**97,256**

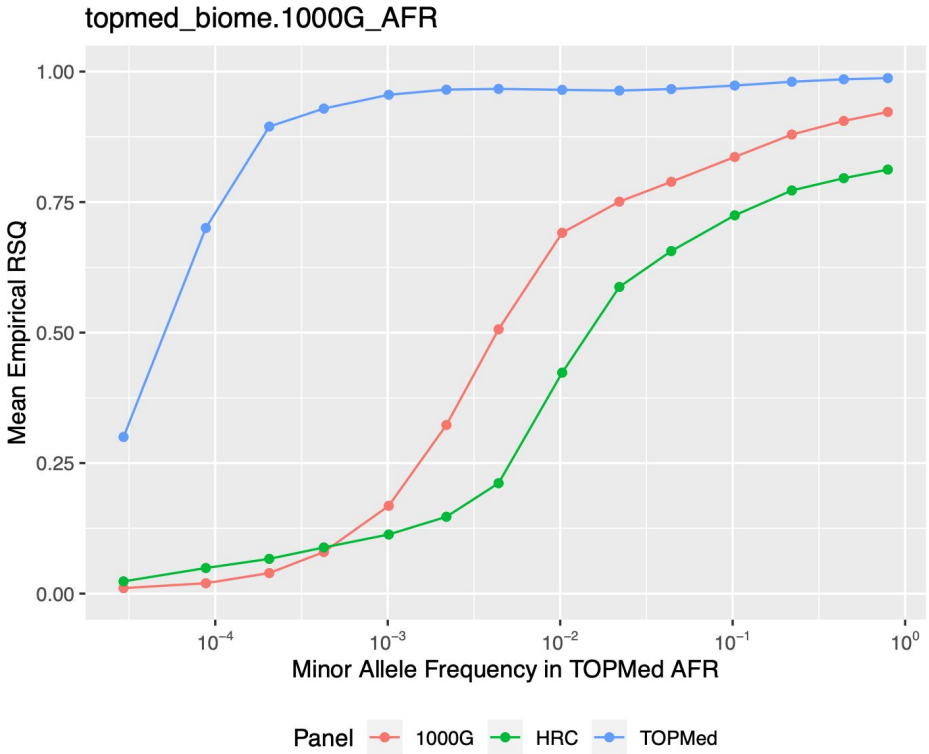
ASHG
American Society of Human Genetics

Imputation Panel Quality



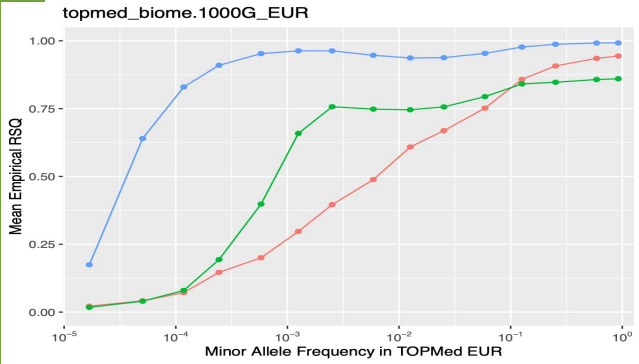
ASHG
American Society of Human Genetics

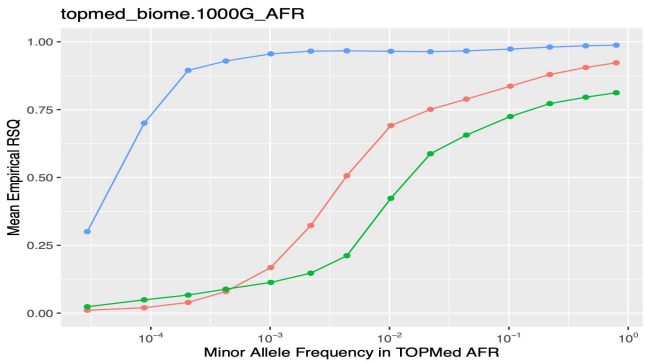
Imputation Panel Quality

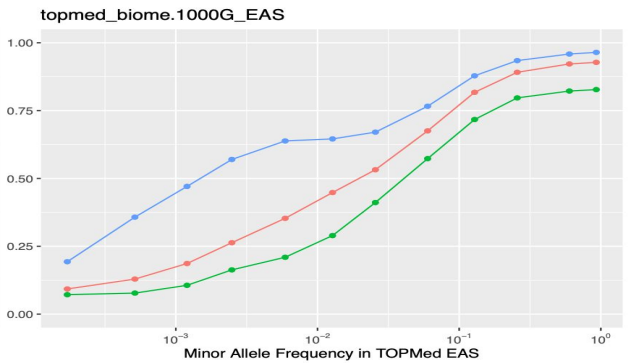


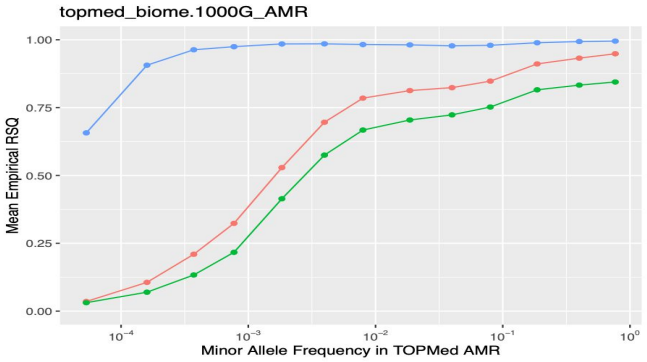
ASHG
American Society of Human Genetics

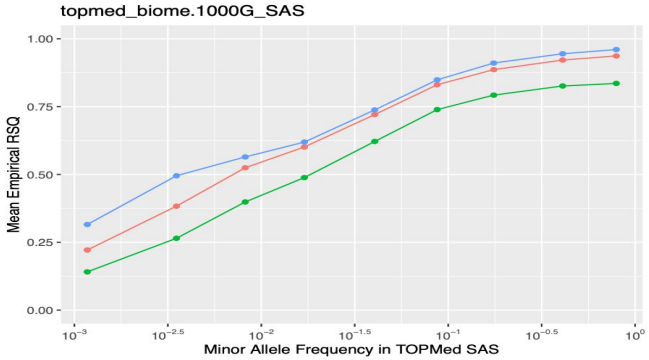
Imputation Panel Quality







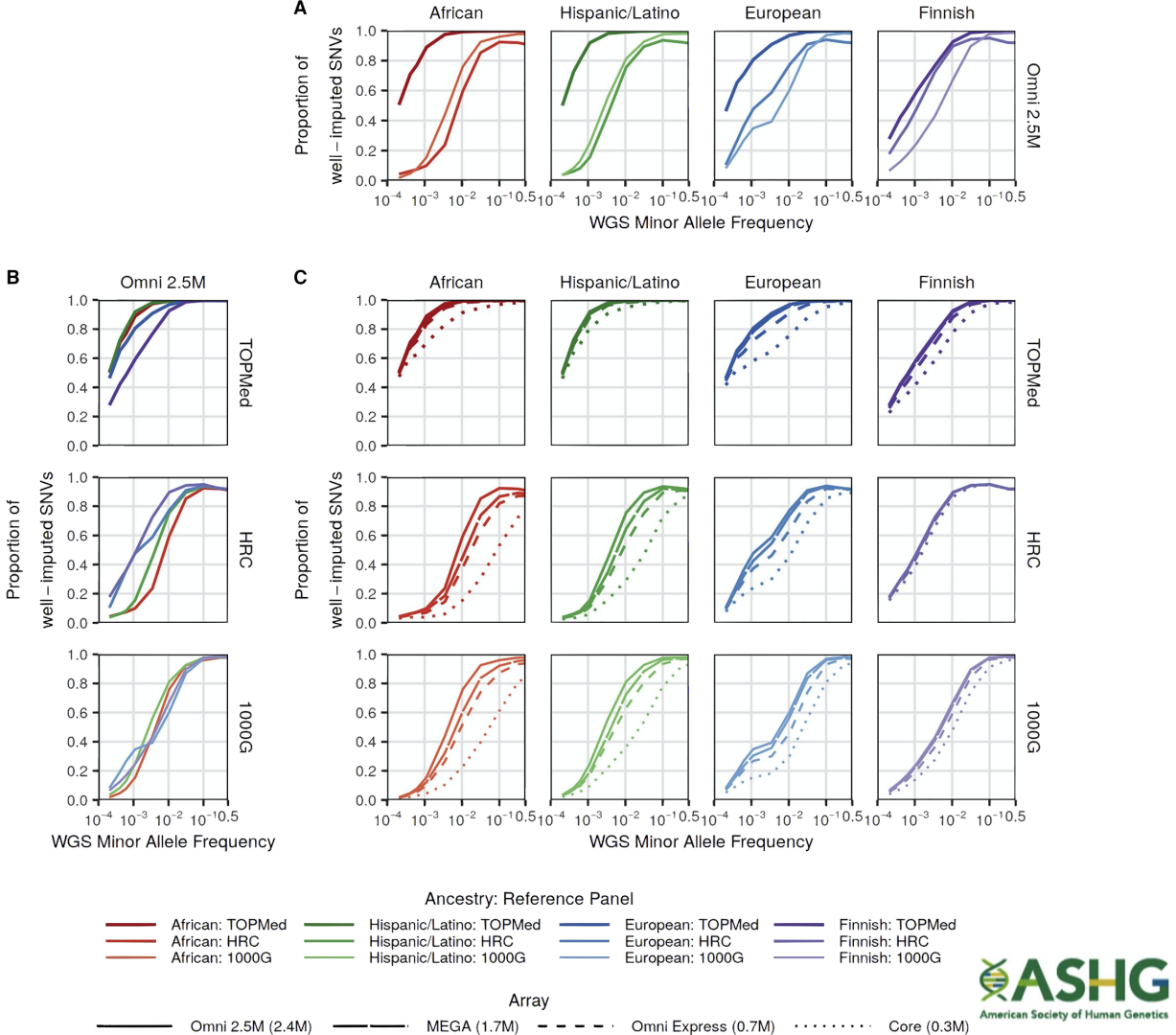


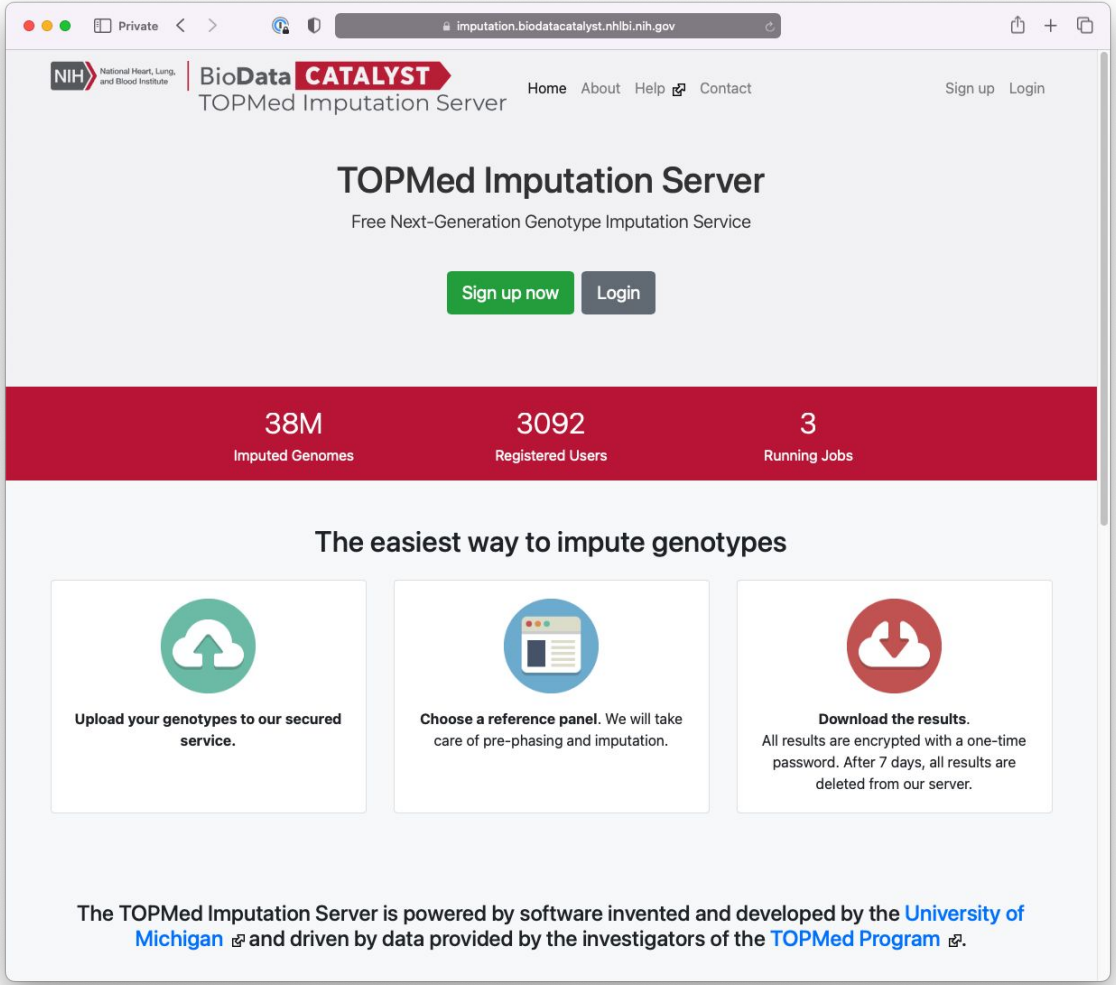


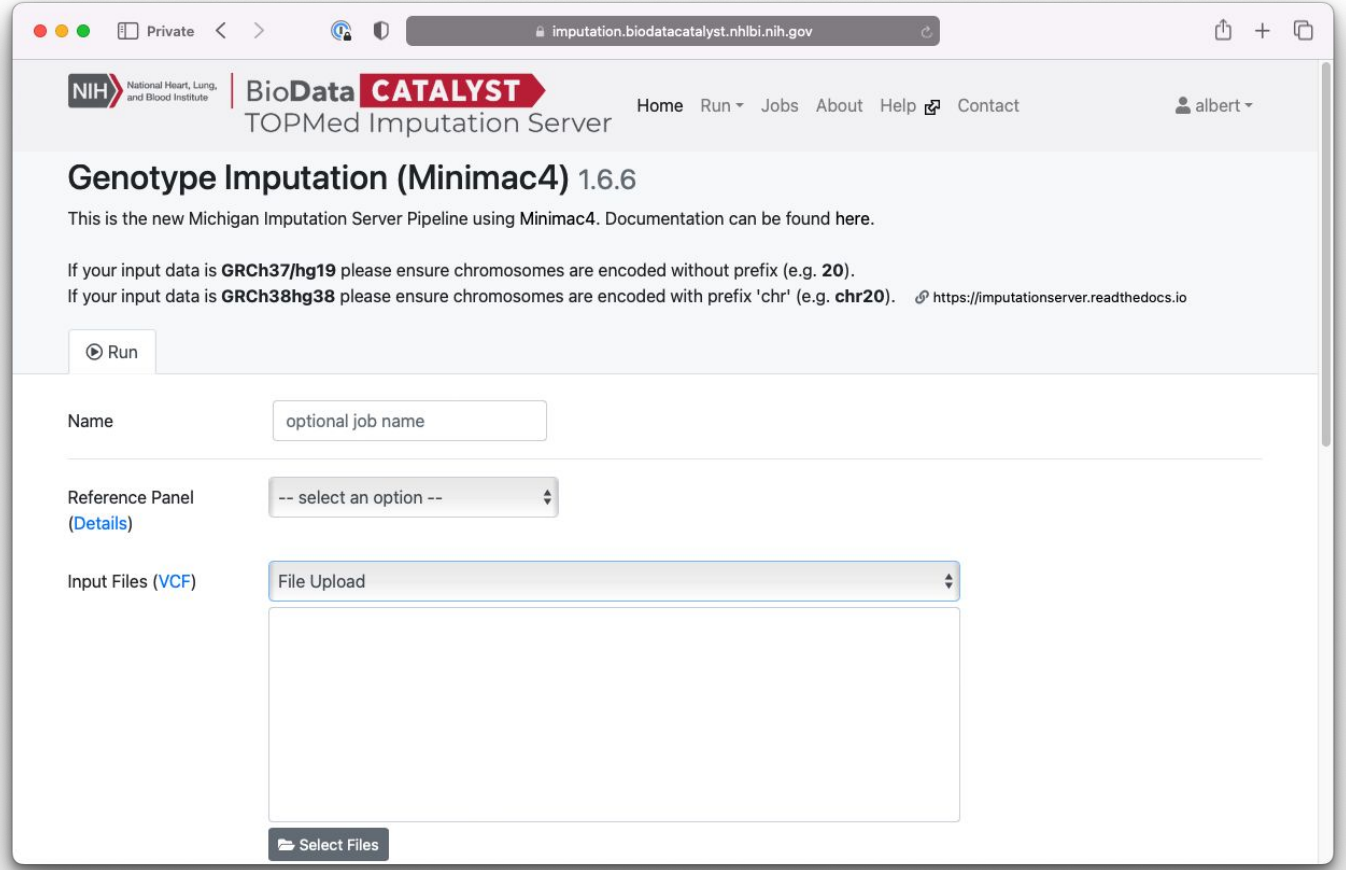
ASHG
American Society of Human Genetics

Imputation Compared to WGS

- Proportion well imputed (r2 > 0.8) down to MAF:
  - 0.14% in African
  - 0.11% in Hispanic/Latino
  - 0.35% in European
  - 0.85% in Finnish
- Similar performance for arrays with >700k variants
- Source: Hanks et al. https://doi.org/10.1016/j.ajhg.2022.07.012

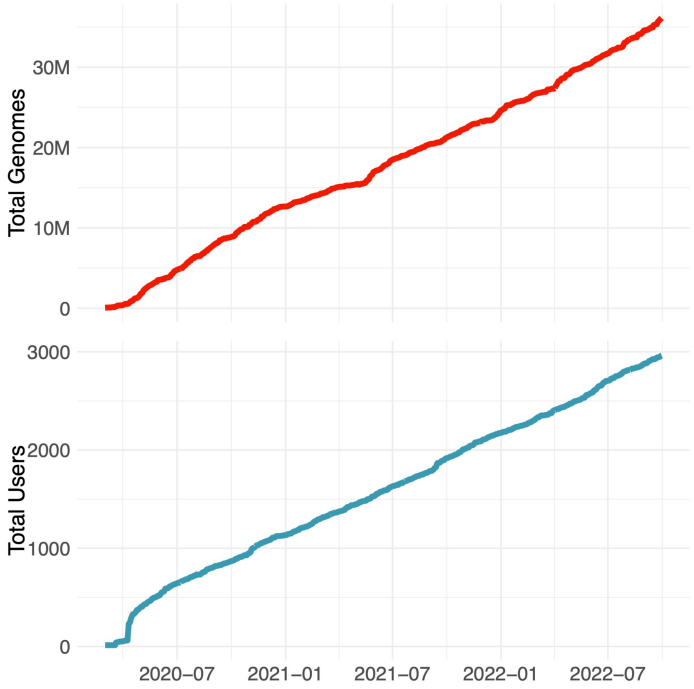






TOPMed Imputation

- Rapid uptake: 38M genomes imputed
- Largely supplanted 1000g & HRC imputation
- Particularly benefits ethnically diverse cohorts
- Satisfying GDPR-related concerns of European users remains a challenge



ASHG
American Society of Human Genetics

Imputation Resources

- Michigan Imputation Server

  https://imputation.sph.umich.edu/
- TOPMed Imputation Server

  https://imputation.biodatacatalyst.nhlbi.nih.gov/
- Documentation

  https://imputationserver.readthedocs.io/
  https://topmedimpute.readthedocs.io/
- TOPMed Imputation Contact

  imputationserver@umich.edu

Your questions

- Put in the chat
- Michigan Imputation Contact

  cfuchsb@umich.edu
- TOPMed Imputation Contact

  imputationserver@umich.edu

ASHG
American Society of Human Genetics