

Disclosure Slide

Financial Disclosure for:

Christian Fuchsberger

Sebastian Schönherr

Lukas Forer

Cassie Spracklen

Albert Smith

Sarah Hanks

We have nothing to disclose

Section 7

Imputation panels



Sarah Hanks
University of Michigan
schanks@umich.edu



MICHIGAN
IMPUTATION SERVER

Learning objectives

Participants will

1. Learn how to calculate and interpret empirical metrics of imputation quality using gold standard whole genome sequencing (WGS)
1. Understand the effects of genotyping array and reference panel choices on imputation quality compared to WGS

Array genotyping and imputation vs WGS

- Deep (>30X) WGS is gold standard for capturing single nucleotide variants (SNVs)
 - Large sample sizes often prohibitively expensive (\$600-1000 per genome)
- Genotyping arrays are less expensive (~\$50 to \$200 per sample) → larger samples
 - Limited to predetermined set of mostly common ($\text{MAF} > 0.05$) variants
 - Imputation estimates missing variants

How close are we getting to genotypes generated from WGS
with array + imputation?

Factors associated with higher imputation quality

Genotyping array	<ul style="list-style-type: none">• Higher density (more markers)
Reference panel	<ul style="list-style-type: none">• Larger panel size• Better sequencing quality/ higher sequencing depth• Greater genetic similarity between panel and sample

Imputation reference panels

Panel	N	Mean depth	Populations	Markers	Variants
1000G	2,504	7X	Multiethnic	~49M	SNVs + indels
HRC	32,470	4-8X	Mostly European	~39M	SNVs only
TOPMed	97,256	38X	Multiethnic	~308M	SNVs + indels

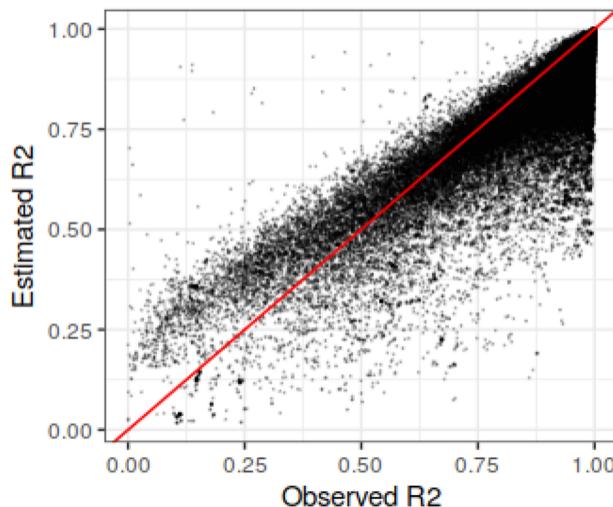
Other panels available on the Michigan Imputation Server include:

- Genome Asia (n=1739)
- CAAPA (n=883)
- HapMap 2 (n=60)

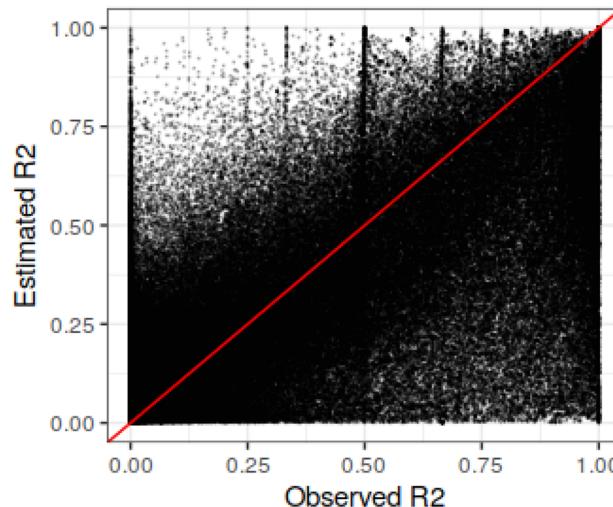
Estimated imputation r^2

SNP»	REF(0)»	ALT(1)»	ALT_Frq»	MAF»	AvgCall_Rsq»	Genotyped»	...»
20:61795:G:T»	G»	T»	0.26318»	0.26318»	0.88455»	0.54658»	Imputed»
20:63231:T:G»	T»	G»	0.03843»	0.03843»	0.98342»	0.67736»	Imputed»
20:63244:A:C»	A»	C»	0.16132»	0.16132»	0.91761»	0.49907»	Imputed»

Common: MAF>0.05



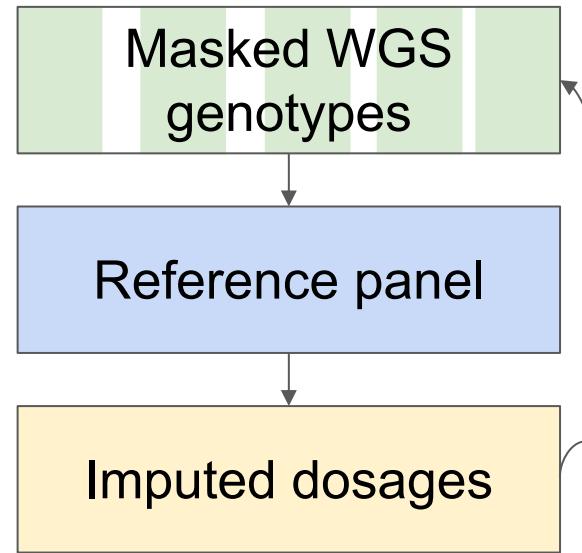
Rare: MAF<0.005



Measuring imputation quality with gold standard WGS

Ancestry	N	Mean depth
African	7,717	27X
European	2,987	39X
Hispanic/Latino	4,677	37X

Illumina Array	N variants
Core	0.3M
OmniExpress	0.7M
MEGA	1.7M
Omni 2.5M	2.4M

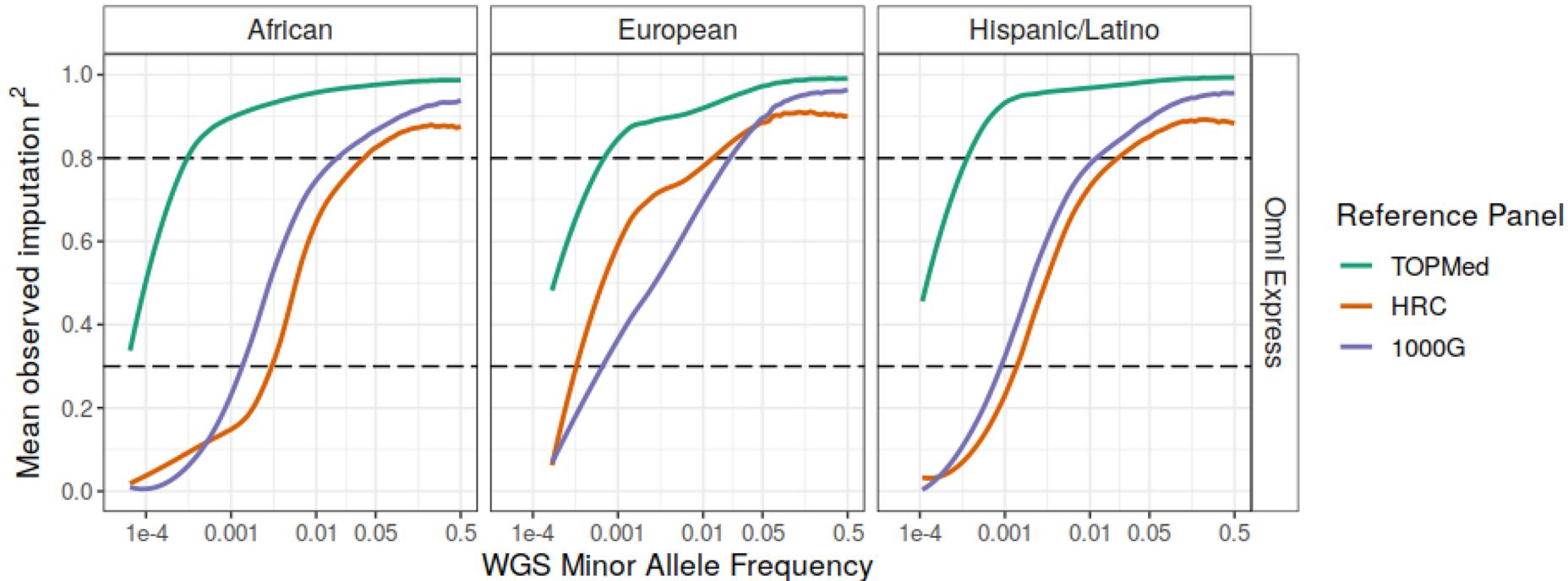


Observed/empirical imputation r^2 : squared Pearson correlation between WGS genotypes and imputed dosages
We set $r^2=0$ for any SNV not in reference panel

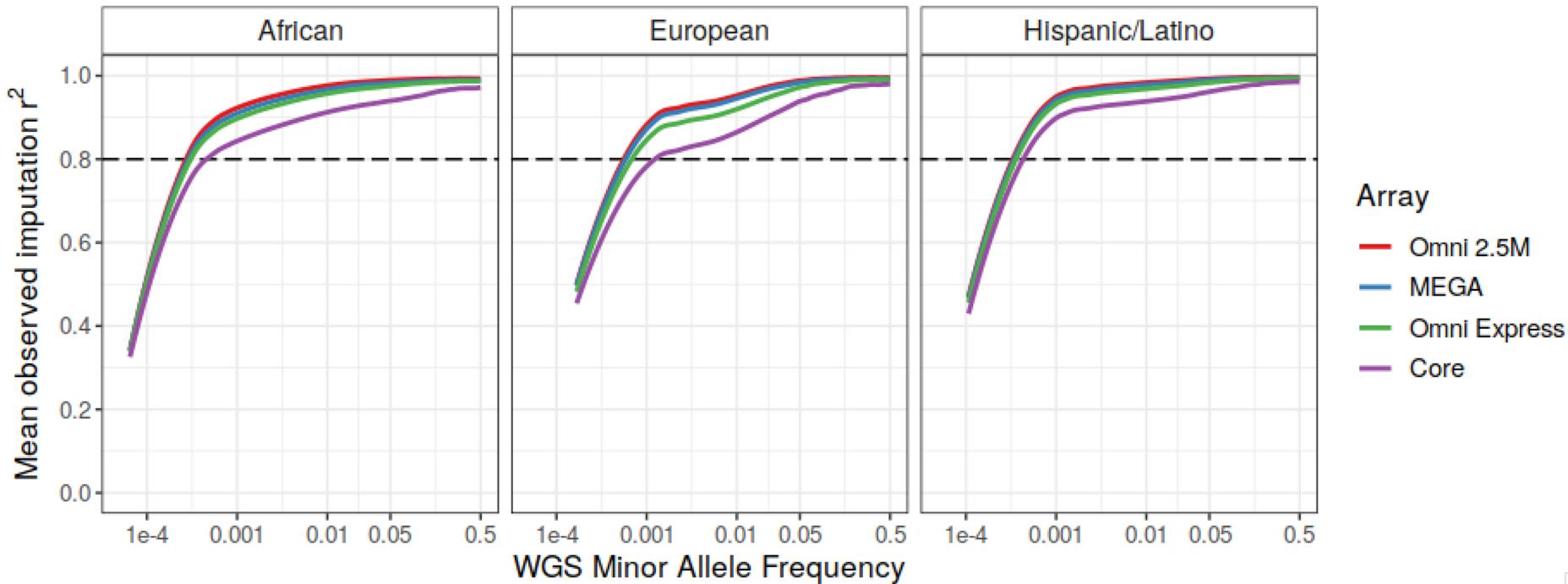
Panel
1000G
HRC
TOPMed

WGS datasets described in TOPMed paper proposal 7193

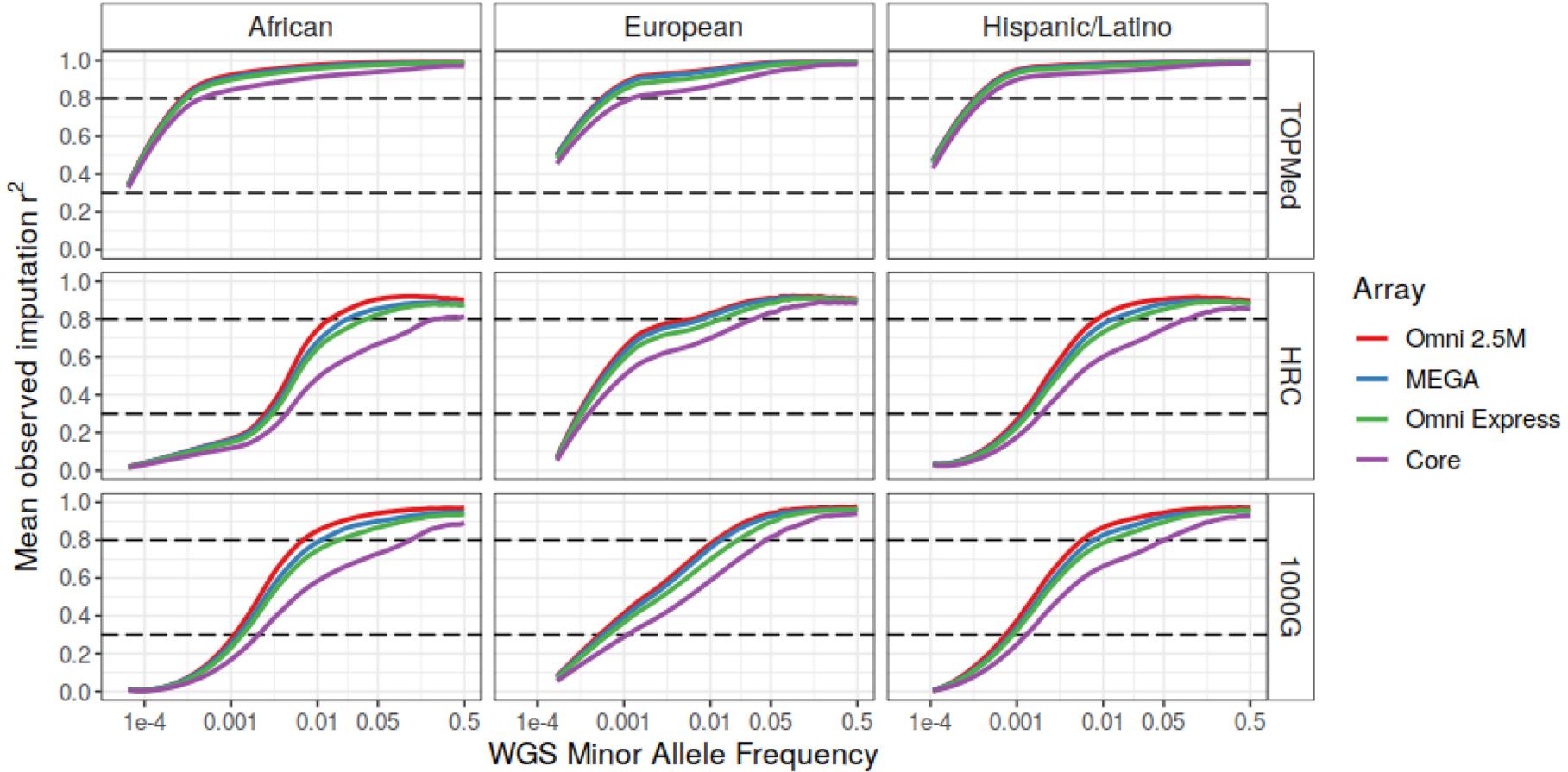
Higher imputation quality with TOPMed panel



Small effect of genotyping array on TOPMed imputation



Greater effect of array choice on 1000G, HRC imputation



Summary

- TOPMed panel gives highest imputation quality in samples of African, European, and Hispanic/Latino ancestry
 - SNVs with MAF 0.04-0.08% can be well imputed ($r^2 \geq 0.8$) in these ancestries
 - Other considerations (homogeneous QC) in ongoing meta-analyses
 - Imputation in samples of Asian ancestry may require other specialized panels (not shown)
- Small effect of genotyping array on TOPMed imputation quality
 - Omni 2.5M, MEGA, Omni Express arrays very similar
 - Greater effect of array on 1000G, HRC imputation

More info and FAQ can be found here:
<https://imputationserver.readthedocs.io>