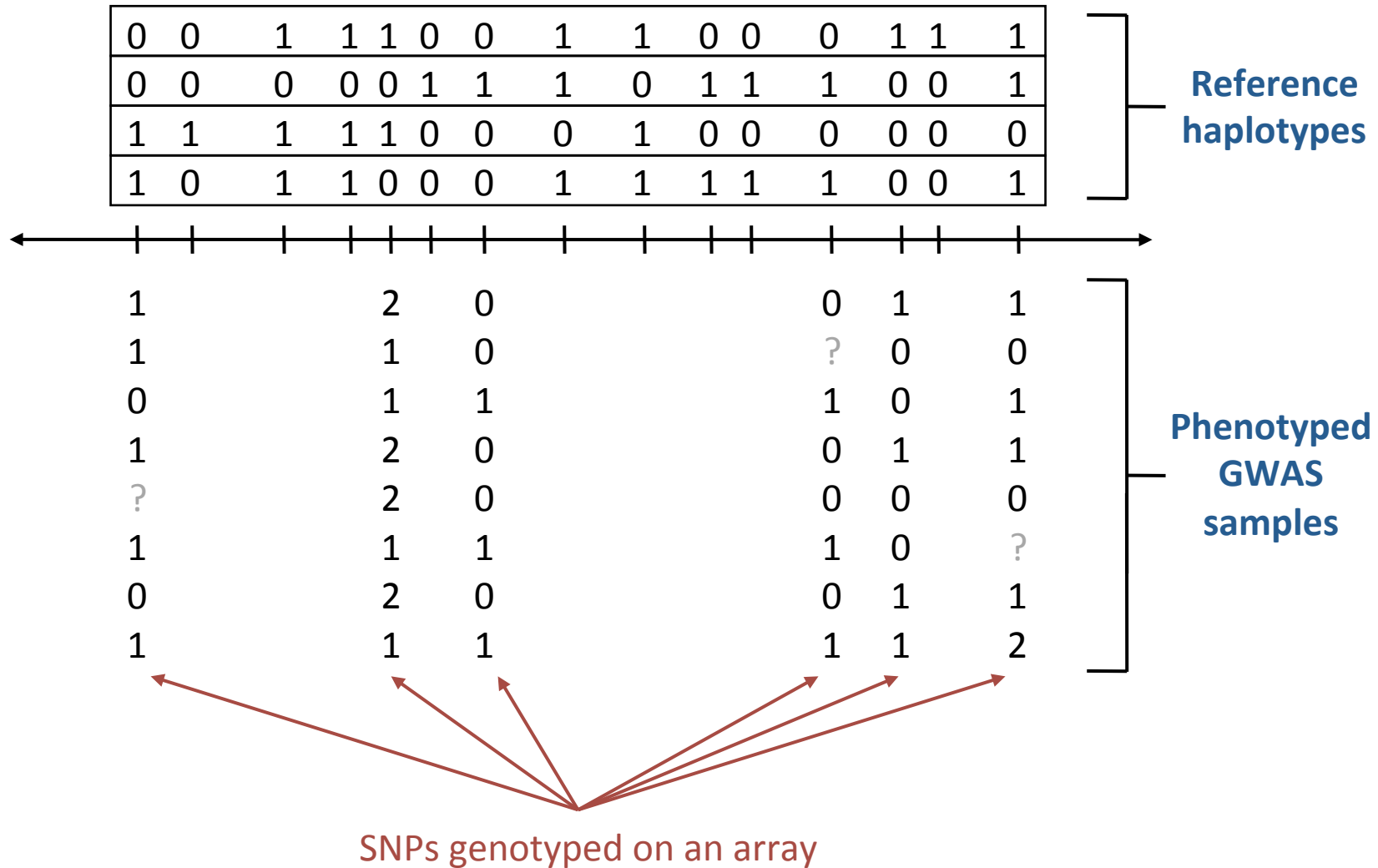


Using 1,000 Genomes data for imputation in genome-wide association studies

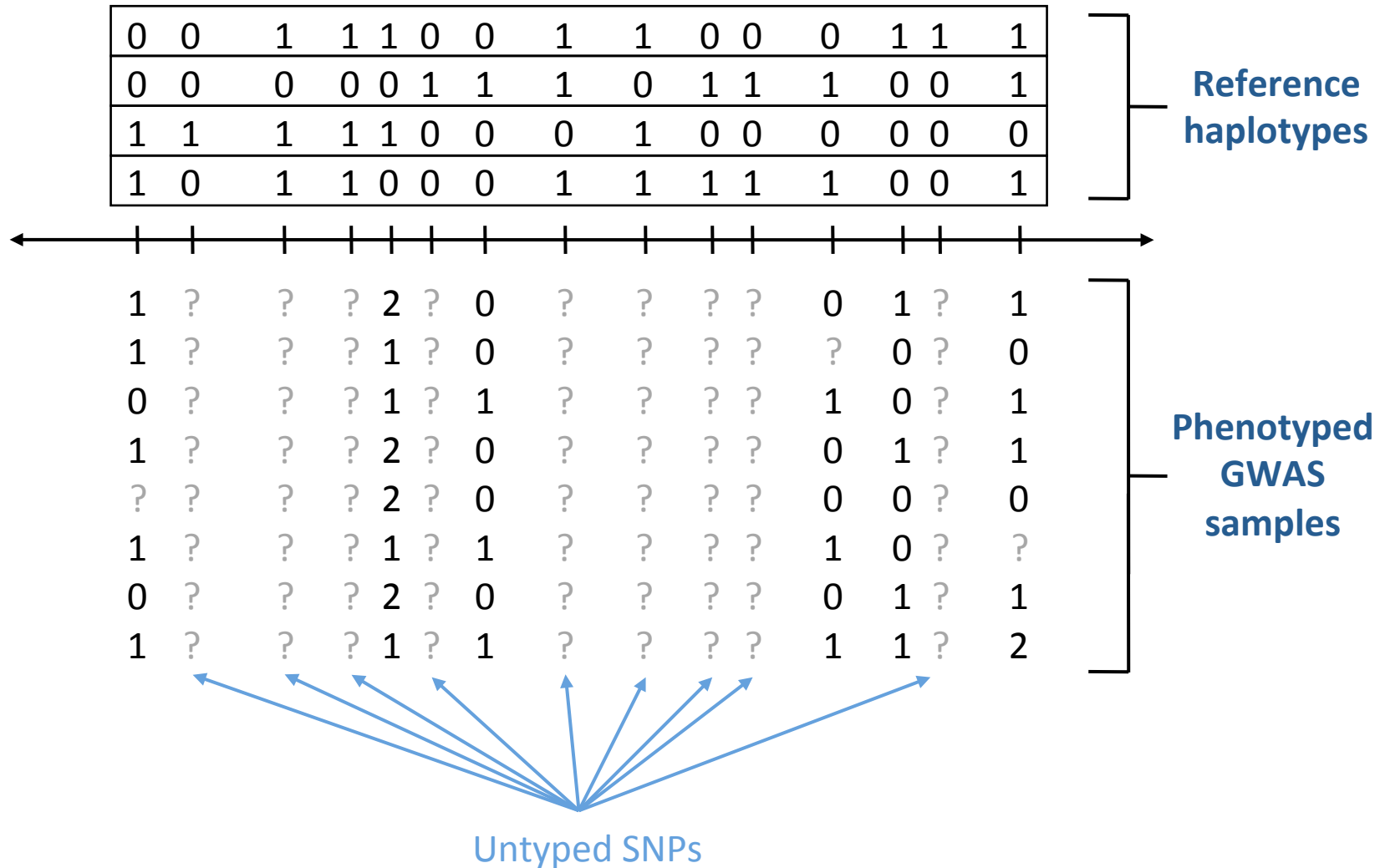
1,000 Genomes Data Tutorial
ICHG 2011, Montreal

Bryan Howie
University of Chicago

Genotype imputation background



Genotype imputation background



Genotype imputation background

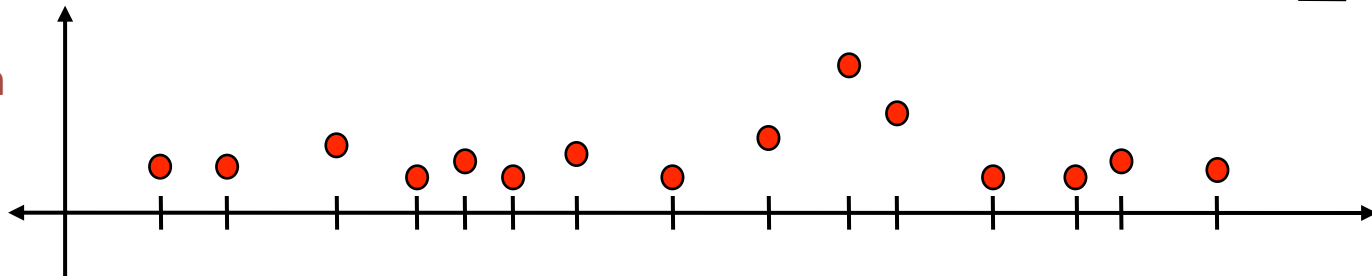
0	0	1	1	1	0	0	1	1	0	0	0	1	1	1
0	0	0	0	0	1	1	1	0	1	1	1	0	0	1
1	1	1	1	1	0	0	0	1	0	0	0	0	0	0
1	0	1	1	0	0	0	1	1	1	1	1	0	0	1

Reference
haplotypes



1	1	2	2	2	0	0	1	2	0	0	0	1	1	1
1	1	1	1	1	0	0	1	2	1	0	0	0	0	0
0	0	1	1	1	1	1	2	1	0	1	1	0	0	1
1	2	2	2	2	0	0	1	2	0	0	0	1	1	1
2	1	2	2	2	0	0	0	2	2	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
0	0	2	2	2	0	0	2	2	2	2	0	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	2

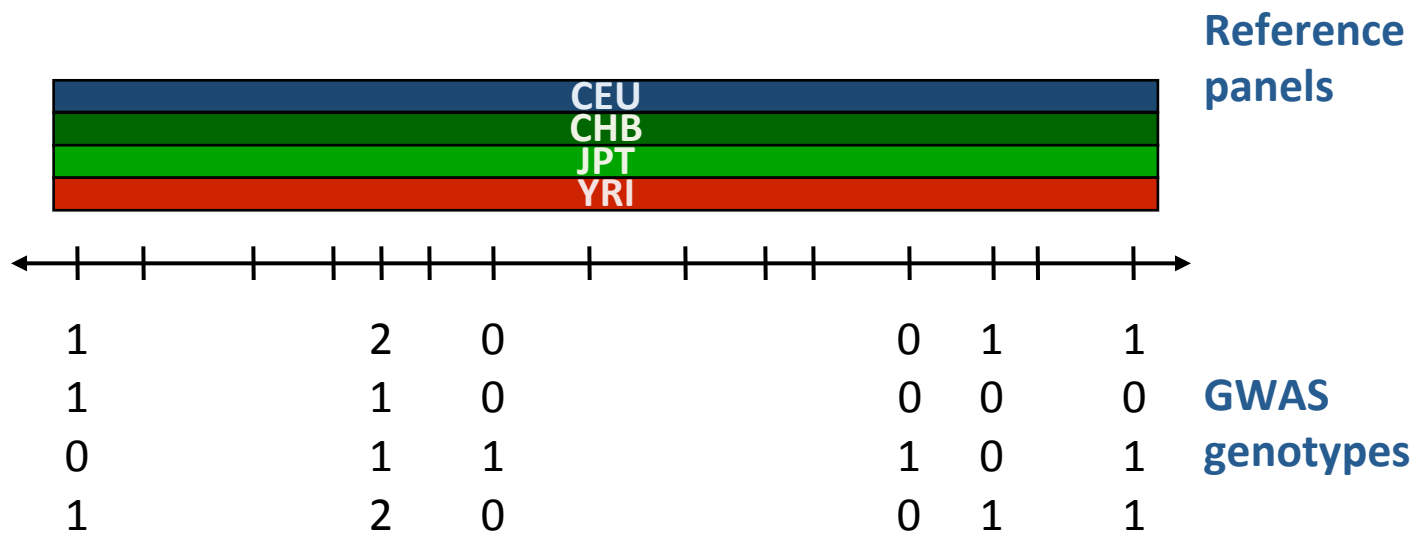
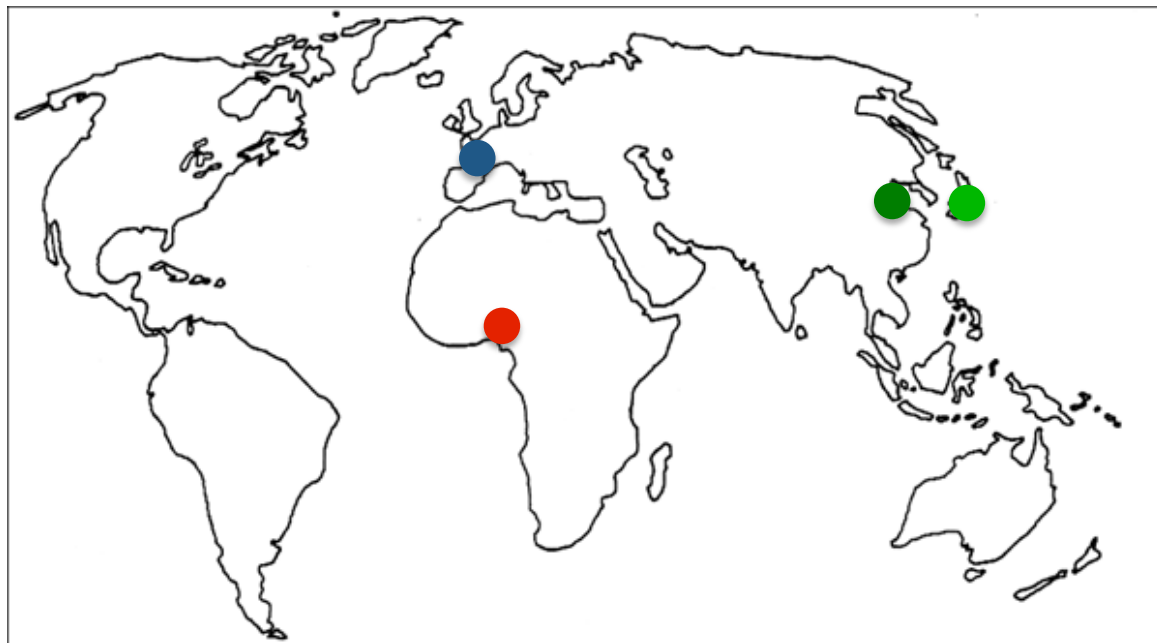
Phenotyped
GWAS
samples



A brief history of imputation reference panels:

HapMap 2, HapMap 3, and
the 1,000 Genomes Project

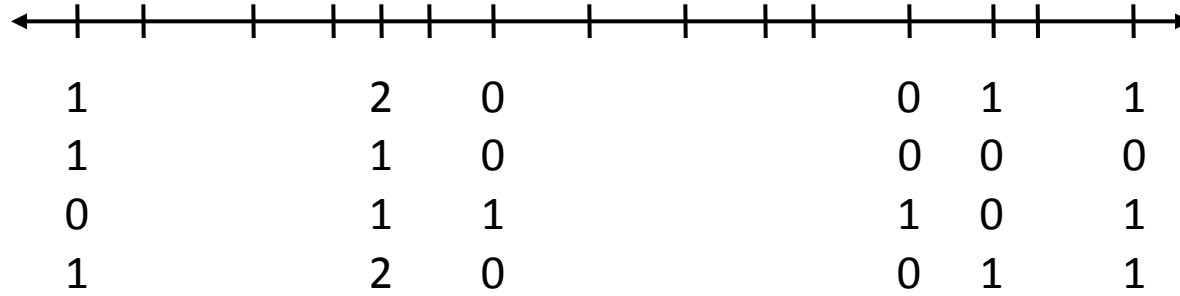
HapMap 2 (2007)



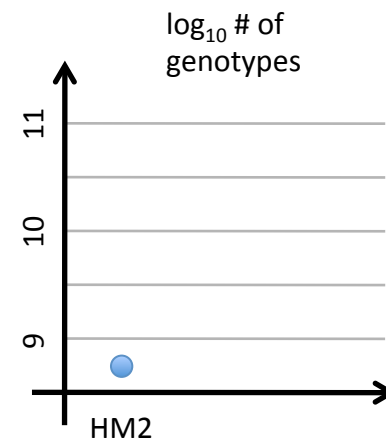
HapMap 2 (2007)



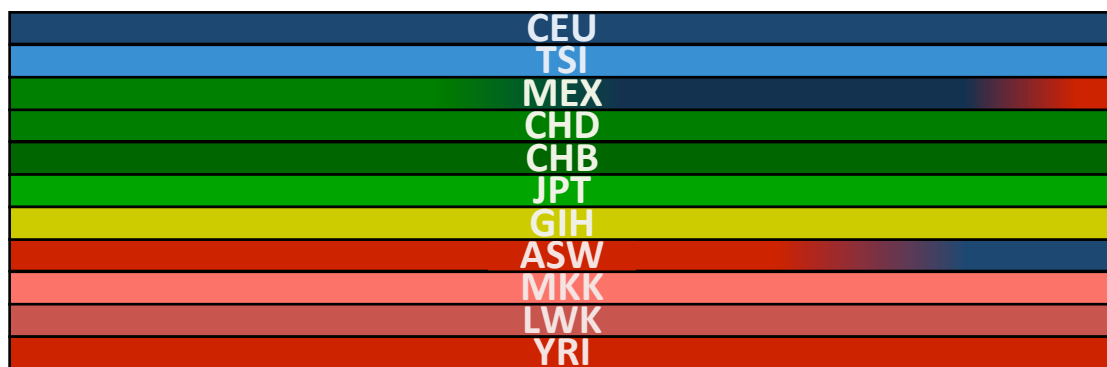
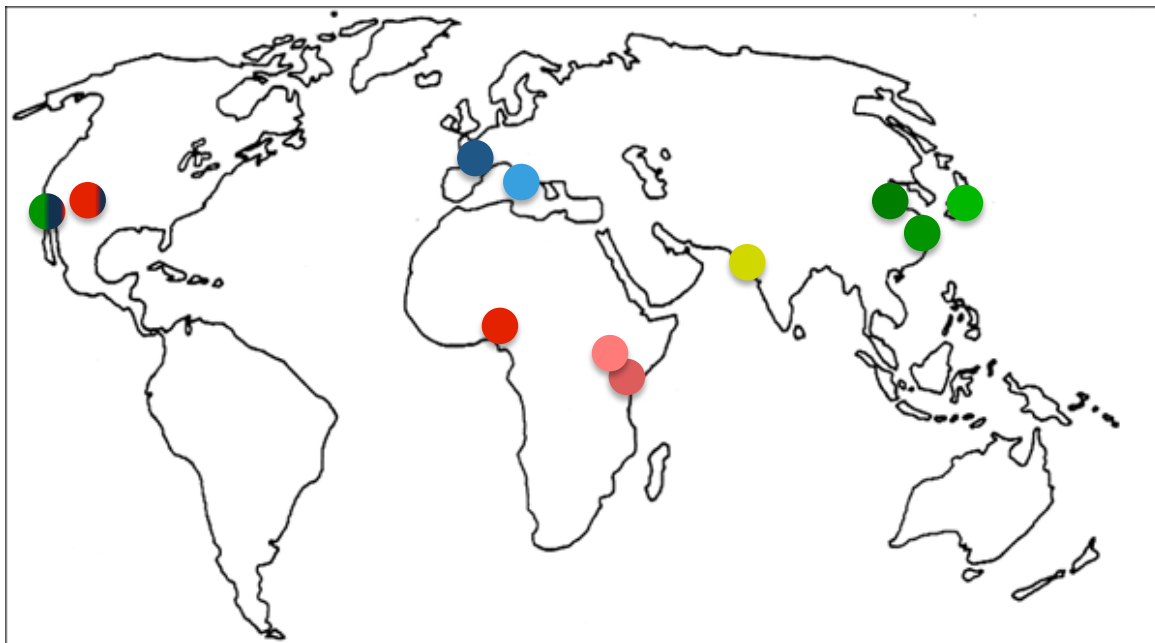
Reference panels



GWAS genotypes



HapMap 3 (2009)

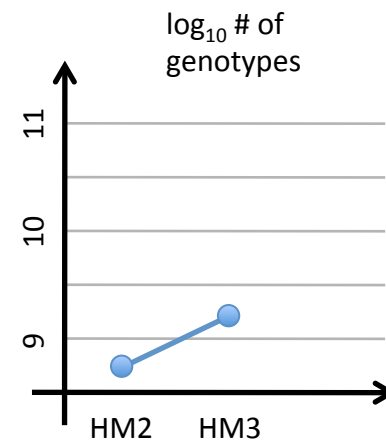


Reference panels

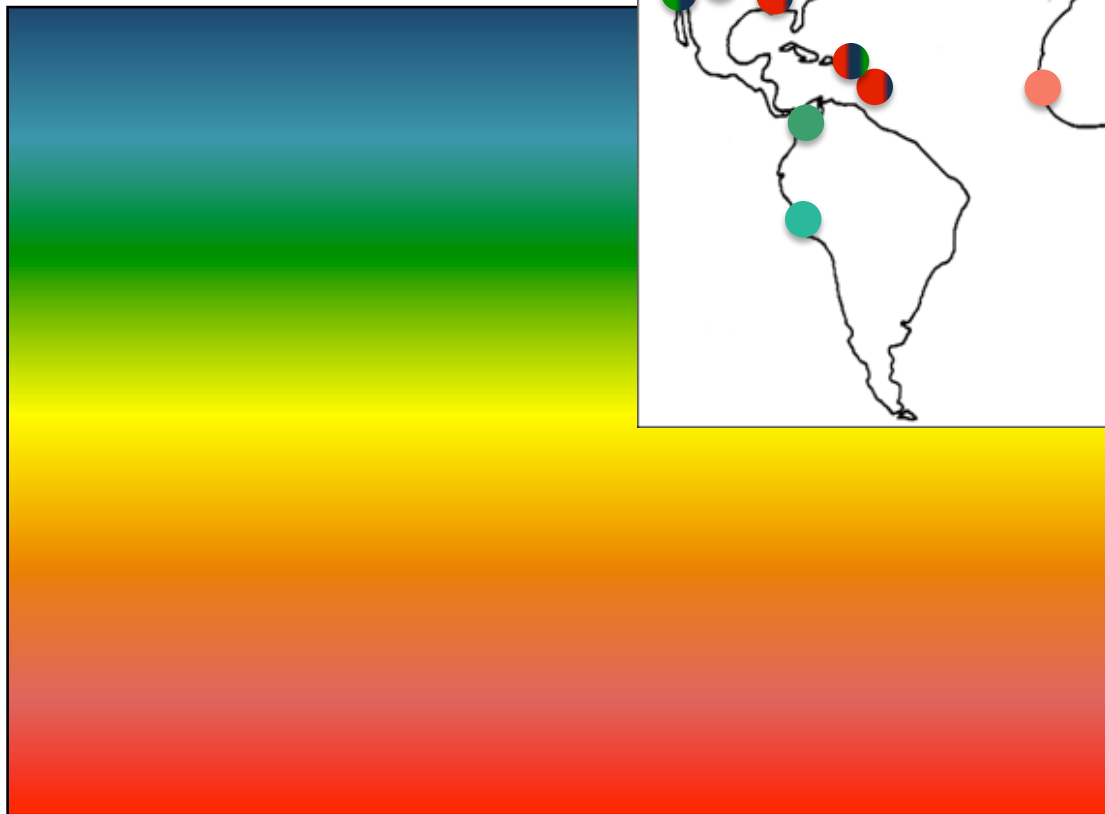
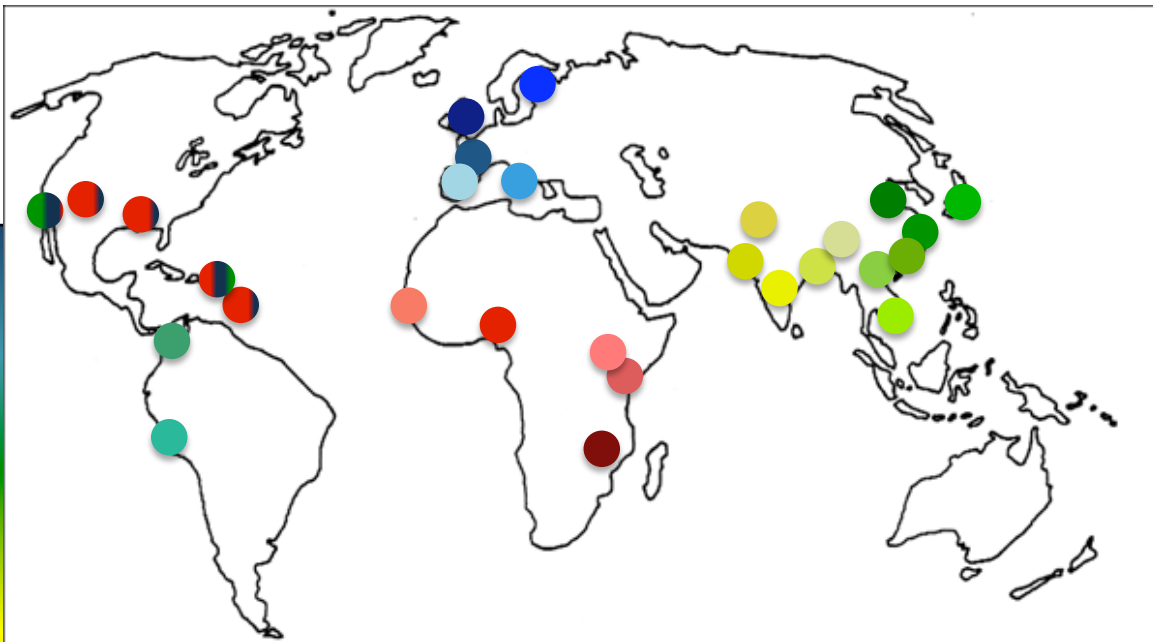


1	2	0	0	1	1
1	1	0	0	0	0
0	1	1	1	0	1
1	2	0	0	1	1

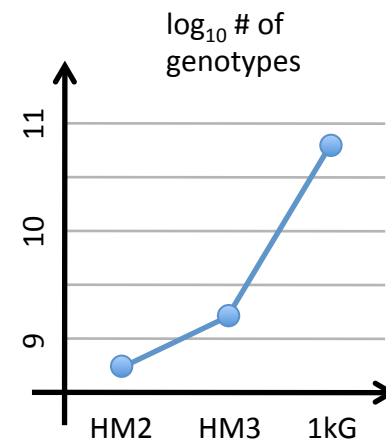
GWAS genotypes



1,000 Genomes (2010+)



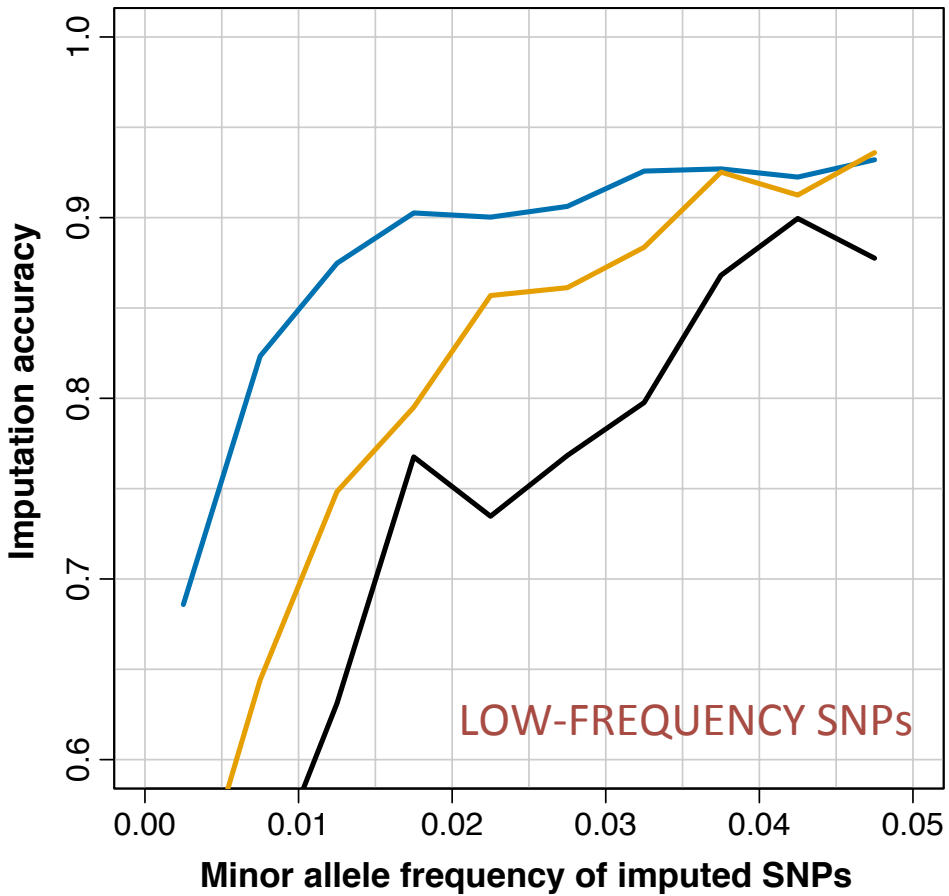
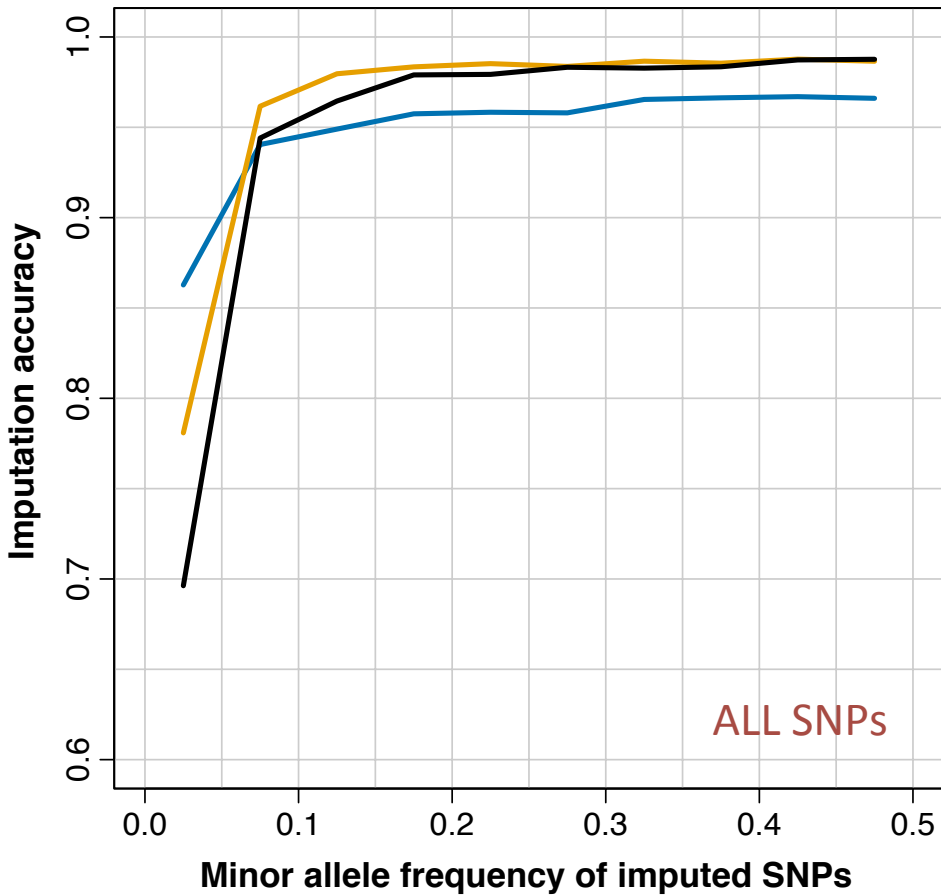
Reference panels



1	2	0	0	1	1
1	1	0	0	0	0
0	1	1	1	0	1
1	2	0	0	1	1

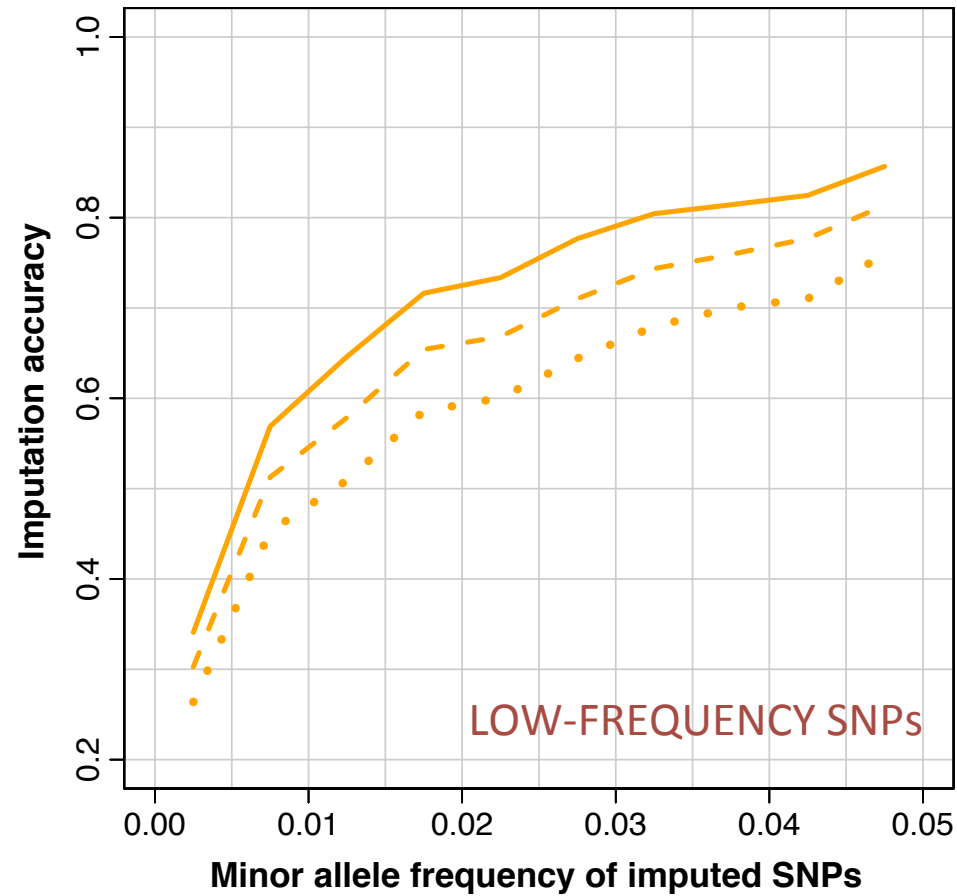
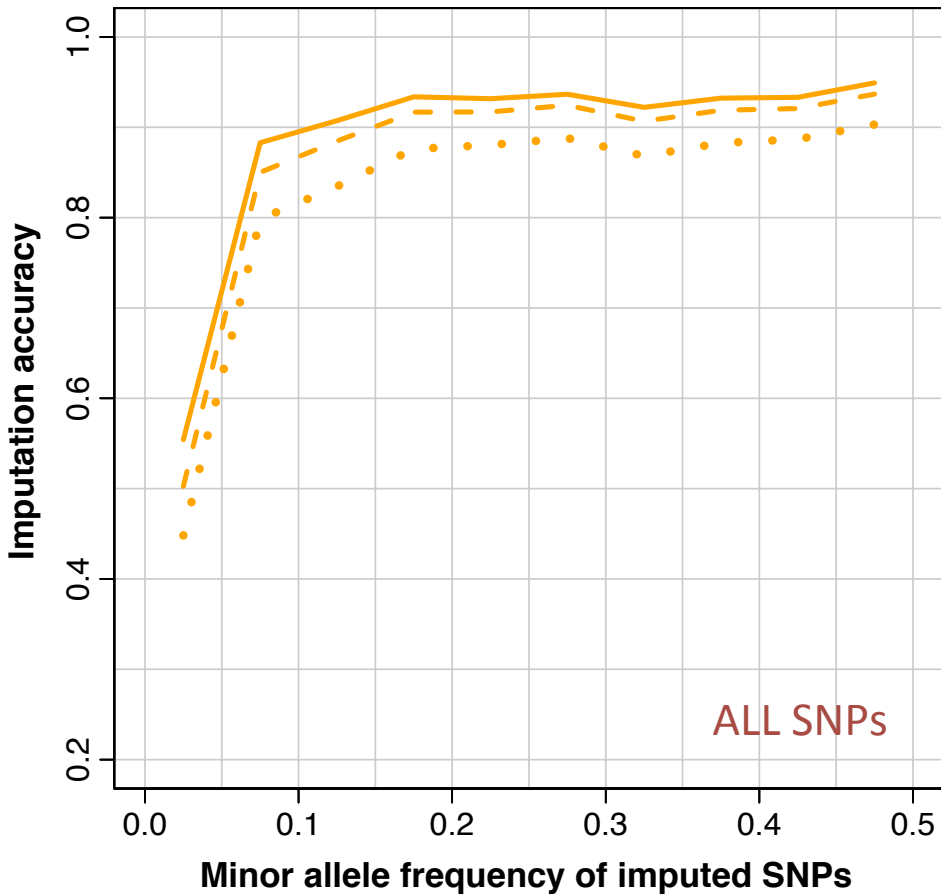
GWAS genotypes

1,000 Genomes haplotypes are highly accurate



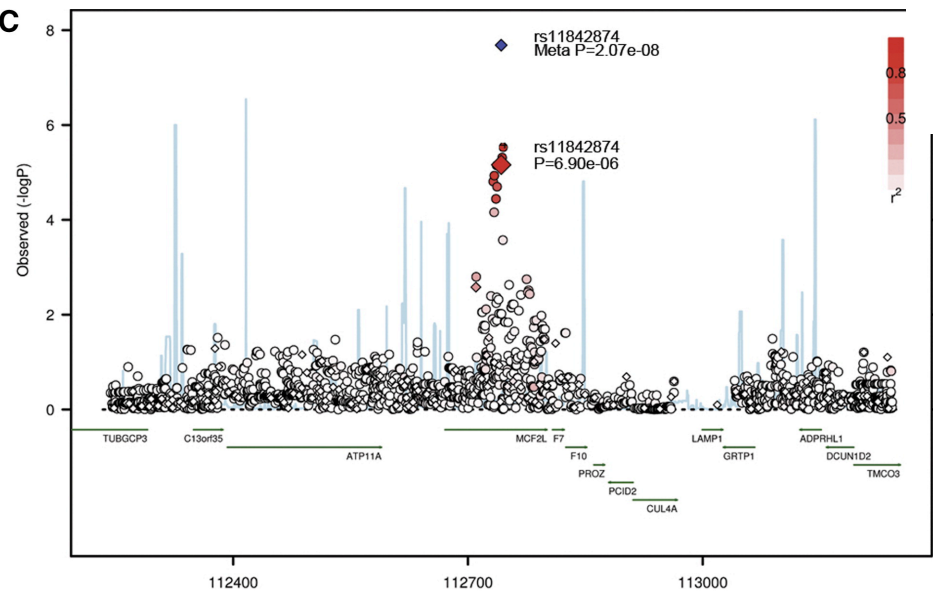
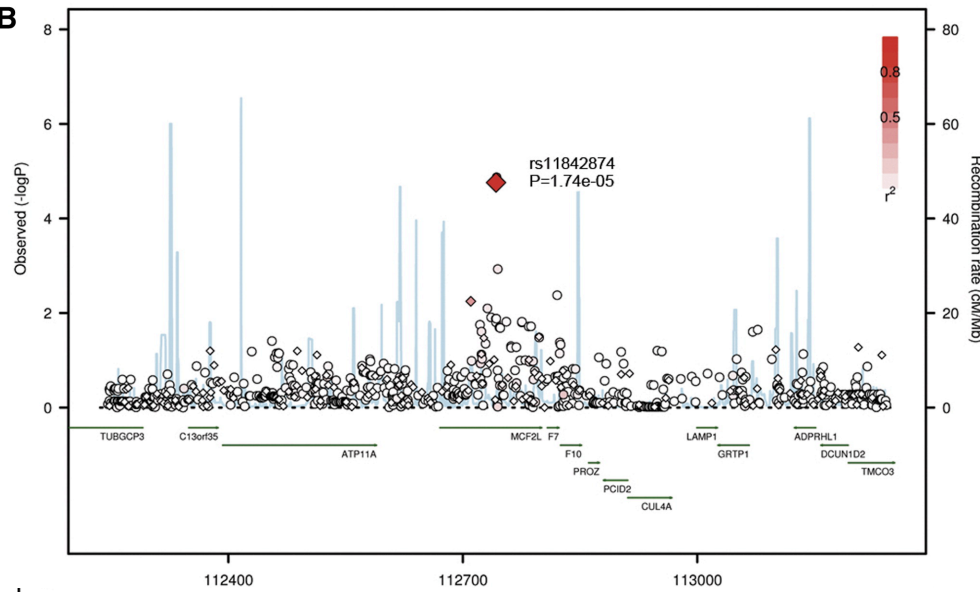
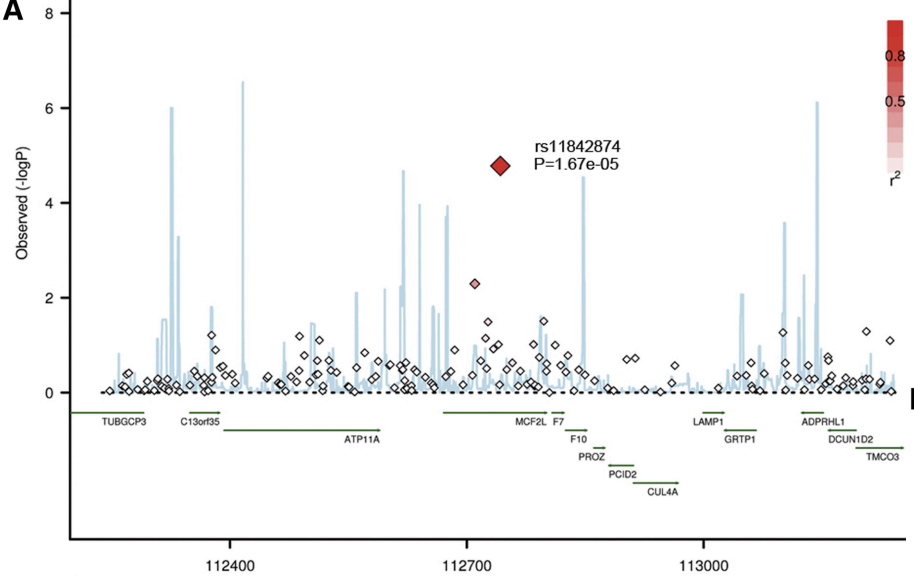
- European ancestry
- African ancestry
- Admixed (Americas)

Imputation accuracy depends on your GWAS chip



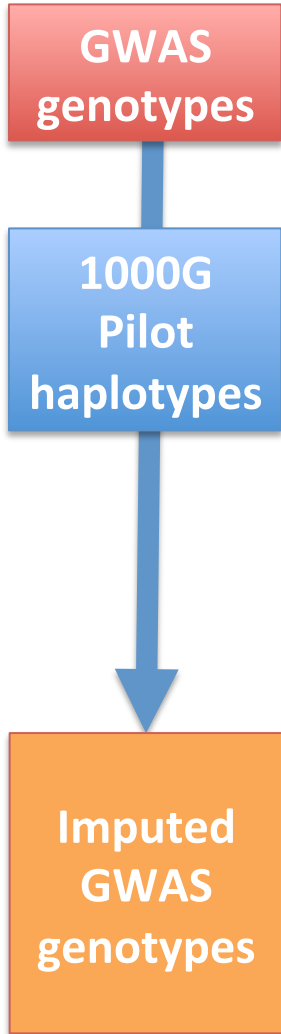
- Omni 2.5M
- - - Illumina 550k
- ... Affymetrix 500k

Imputation from 1,000 Genomes haplotypes can strengthen association signals.



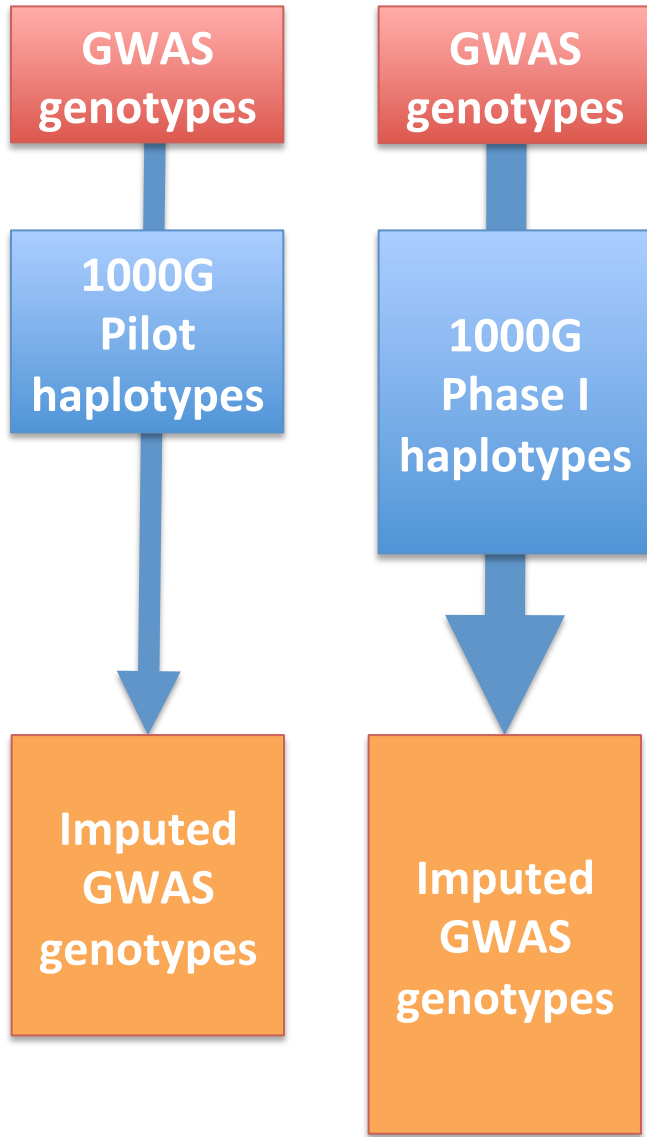
GWAS of Osteoarthritis
Day-Williams et al. (AJHG 2011)

Standard Imputation



40 minutes
per genome

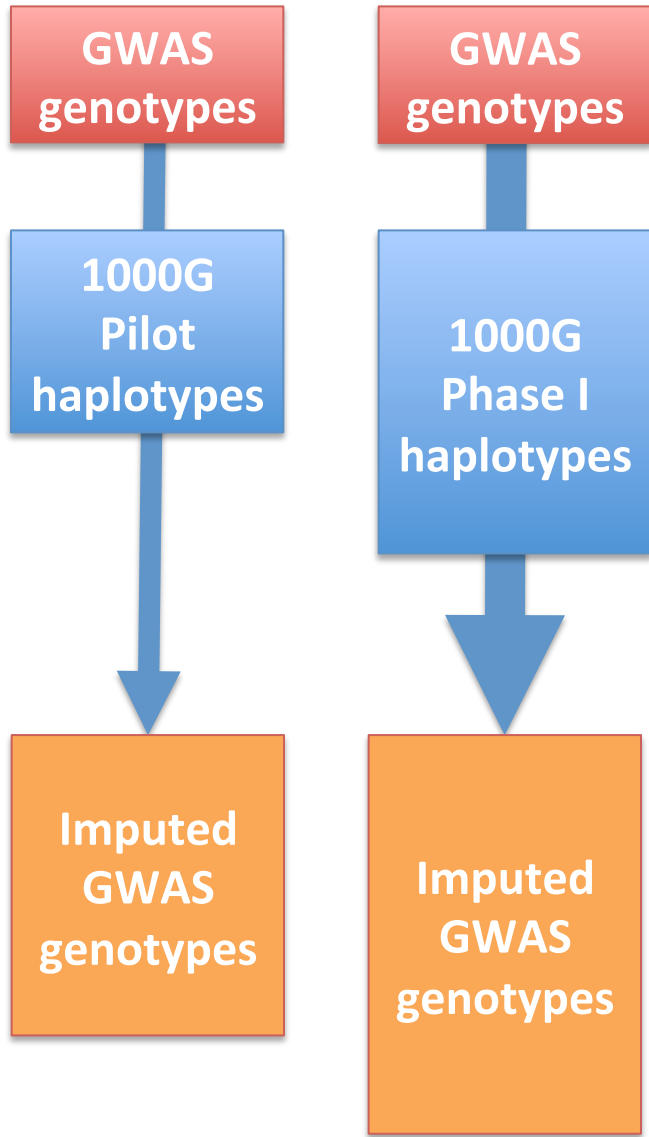
Standard Imputation



40 minutes
per genome

7800 minutes
per genome

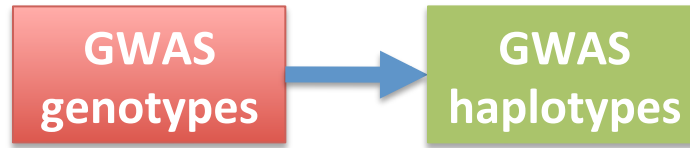
Standard Imputation



40 minutes
per genome

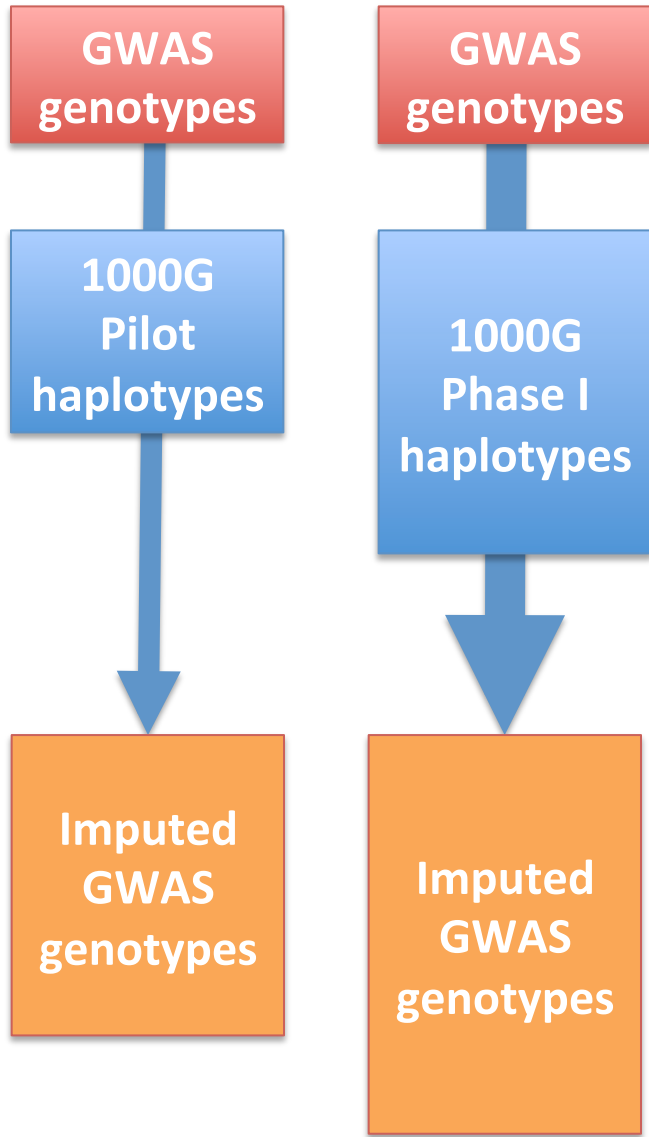
7800 minutes
per genome

Pre-phasing Imputation



25 minutes
per genome

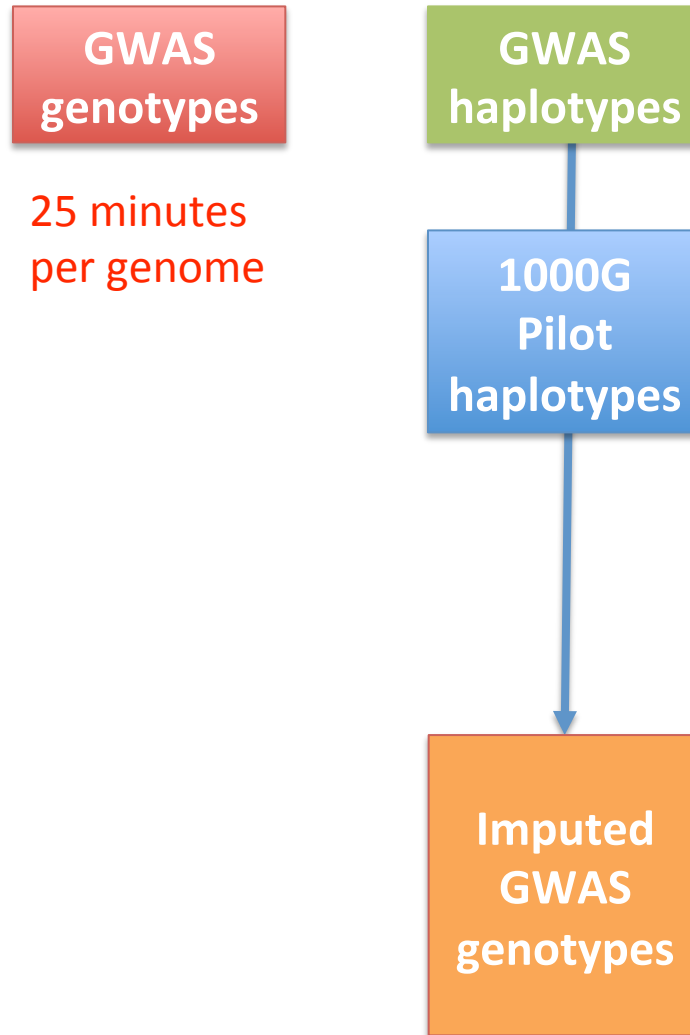
Standard Imputation



40 minutes
per genome

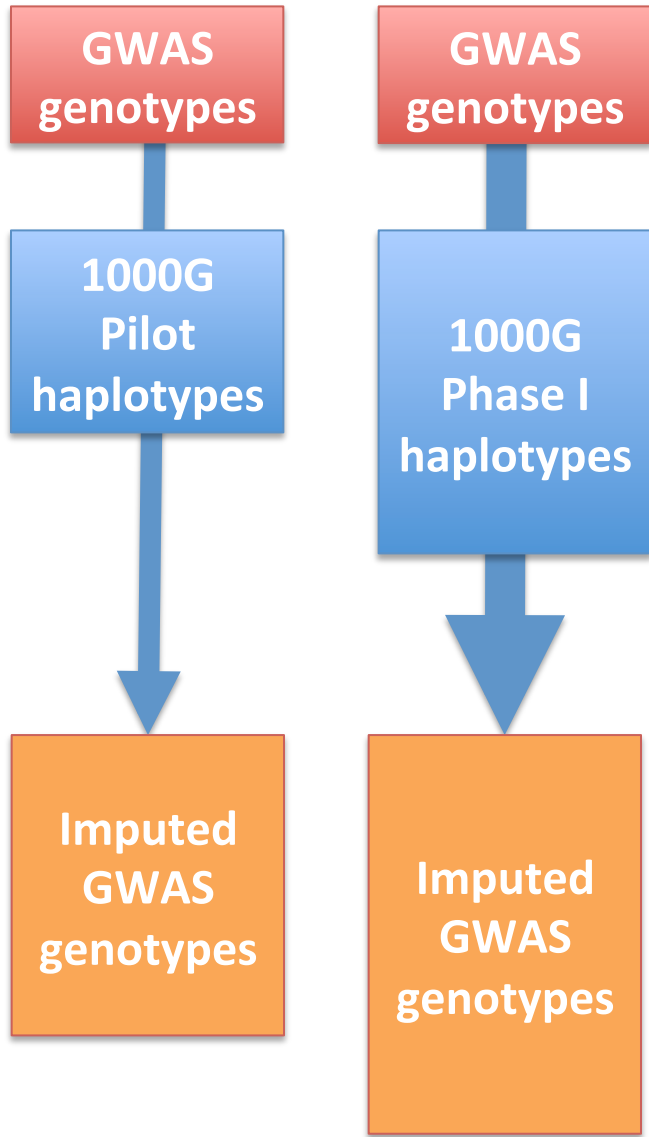
7800 minutes
per genome

Pre-phasing Imputation



1 minute
per genome

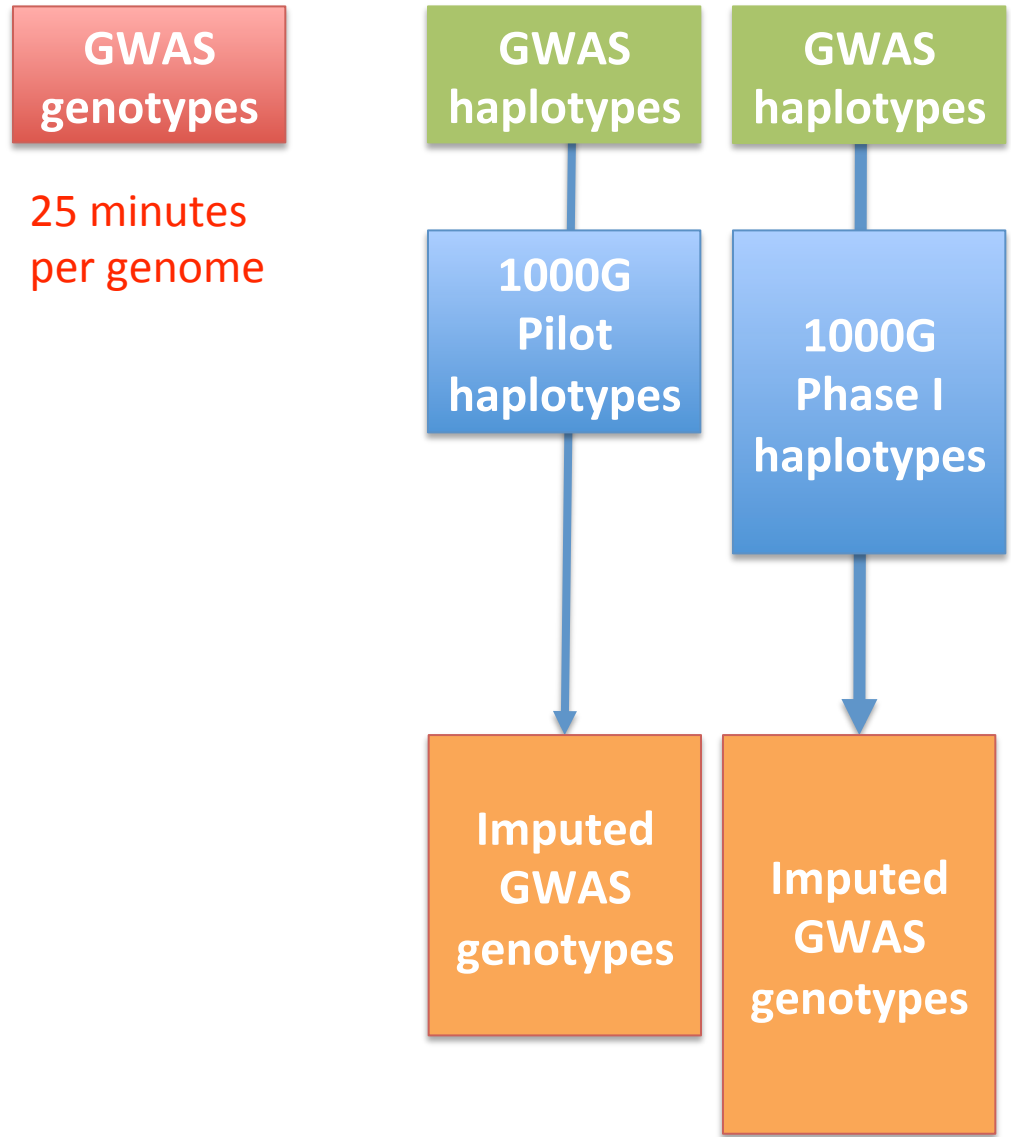
Standard Imputation



40 minutes
per genome

7800 minutes
per genome

Pre-phasing Imputation



1 minute
per genome

24 minutes
per genome

Standard Imputation

Pre-phasing Imputation

GWAS
genotypes

GWAS
genotypes

GWAS
genotypes

GWAS
haplotypes

GWAS
haplotypes

Imputation Accuracy (mean R^2)

1000G panel	MAF 1-3%	MAF 3-5%	MAF >5%
60 CEU	0.66	0.78	0.88
60 CEU	0.65	0.77	0.87
283 EUR	0.73	0.78	0.92
381 EUR	0.83	0.85	0.94

genotypes

GWAS
genotypes

genotypes

GWAS
genotypes

40 minutes
per genome

7800 minutes
per genome

1 minute
per genome

24 minutes
per genome

Getting the latest 1,000 Genomes haplotypes

- Phase 1 haplotypes now include SNPs, INDELS, and SVs!
- 1,000 Genomes haplotypes are available in the formats required by various imputation programs. For example:
 - Beagle:
<http://faculty.washington.edu/browning/beagle/beagle.html>
 - IMPUTE2:
http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
 - MaCH/minimac:
<http://www.sph.umich.edu/csg/abecasis/MACH/download/>
- Thanks for coming!