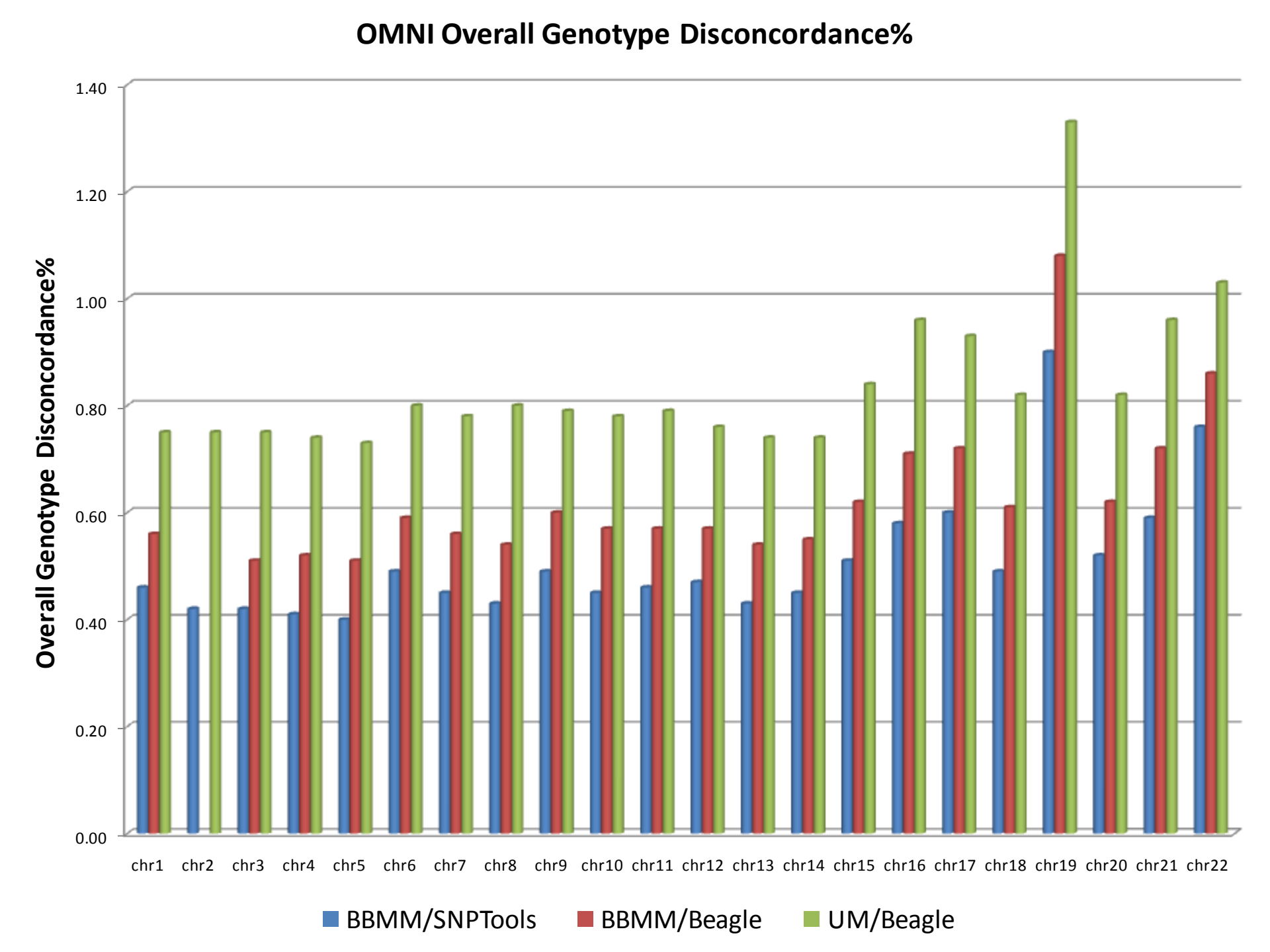
Phase I Low-Pass Whole Genome SNP Genotype Release

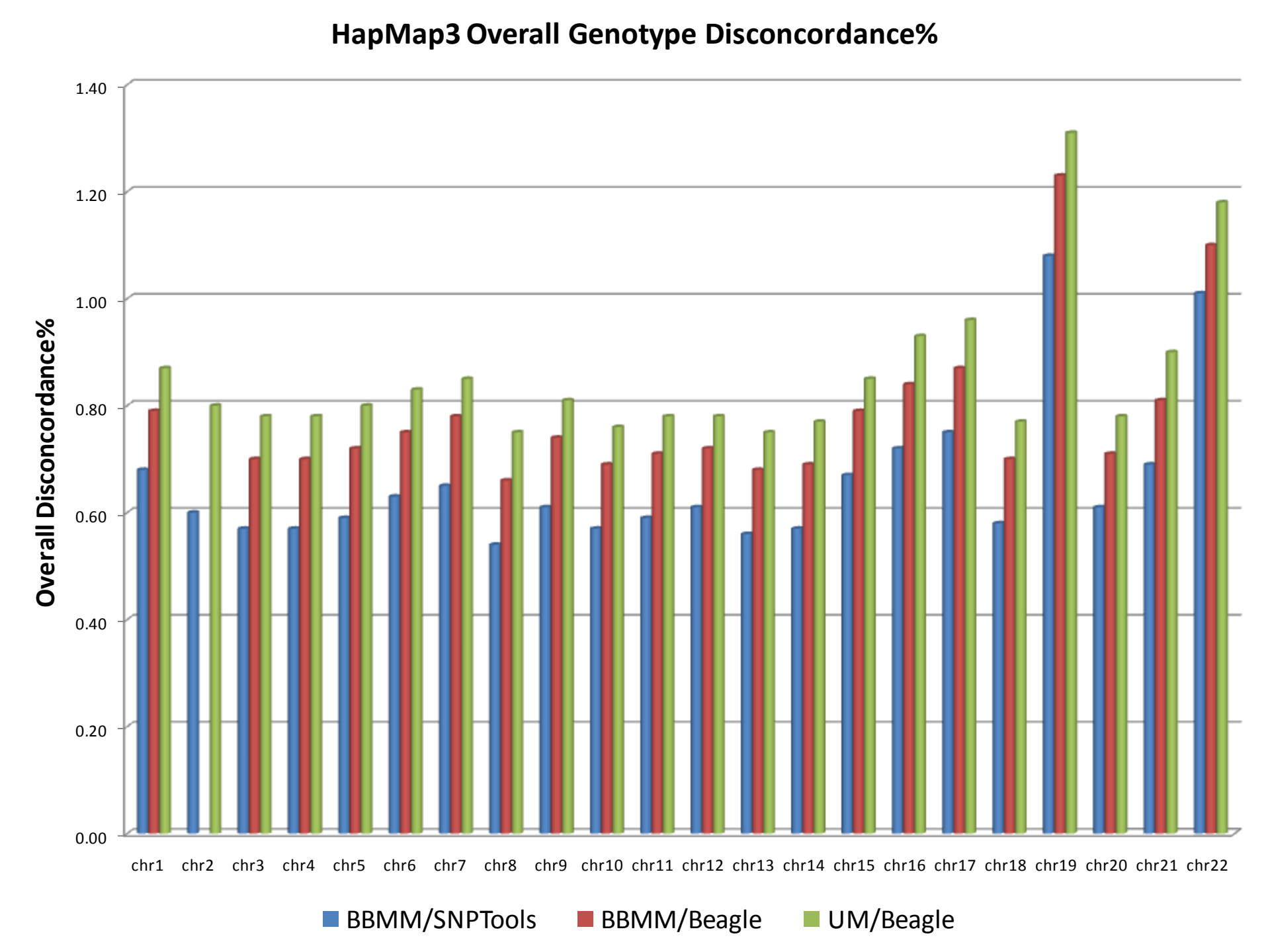
Yi Wang, Jin Yu, Richard Gibbs, Fuli Yu
BCM-HGSC
20110621

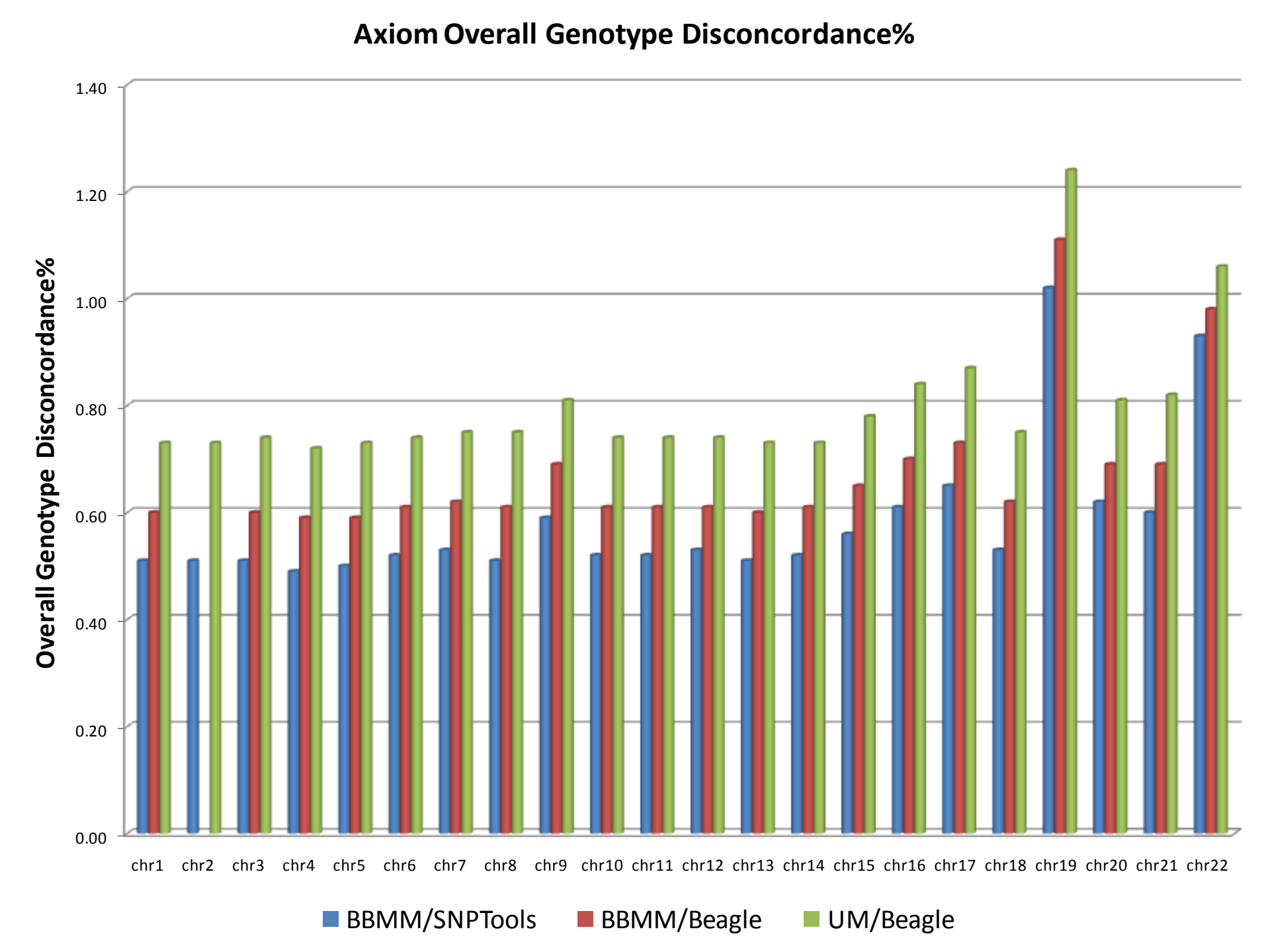
Summary

- **We finished Phase I Low-pass SNP calling task**
  - 1103 low-pass BAMs, 1094 individuals
  - VQSR v2 consensus site list
  - SNPTools BBMM consensus likelihoods
  - SNPTools consensus imputation engine

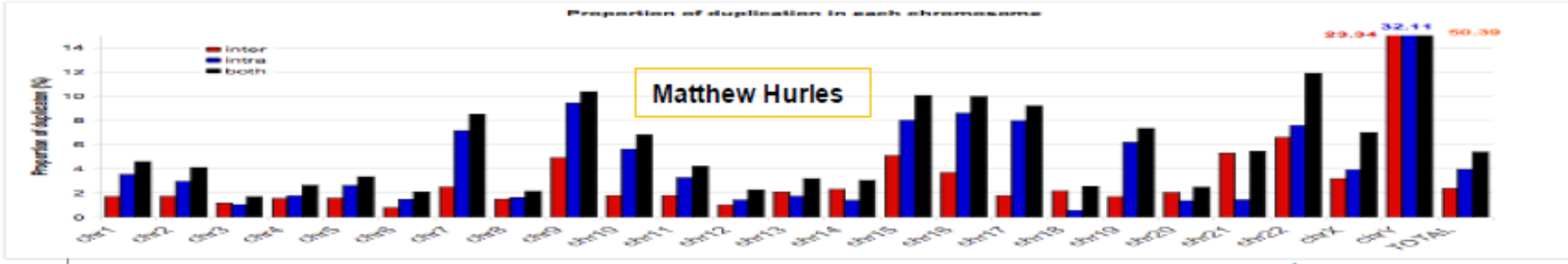
- **UM Beagle callsets and Broad validation results support the consensus methods**

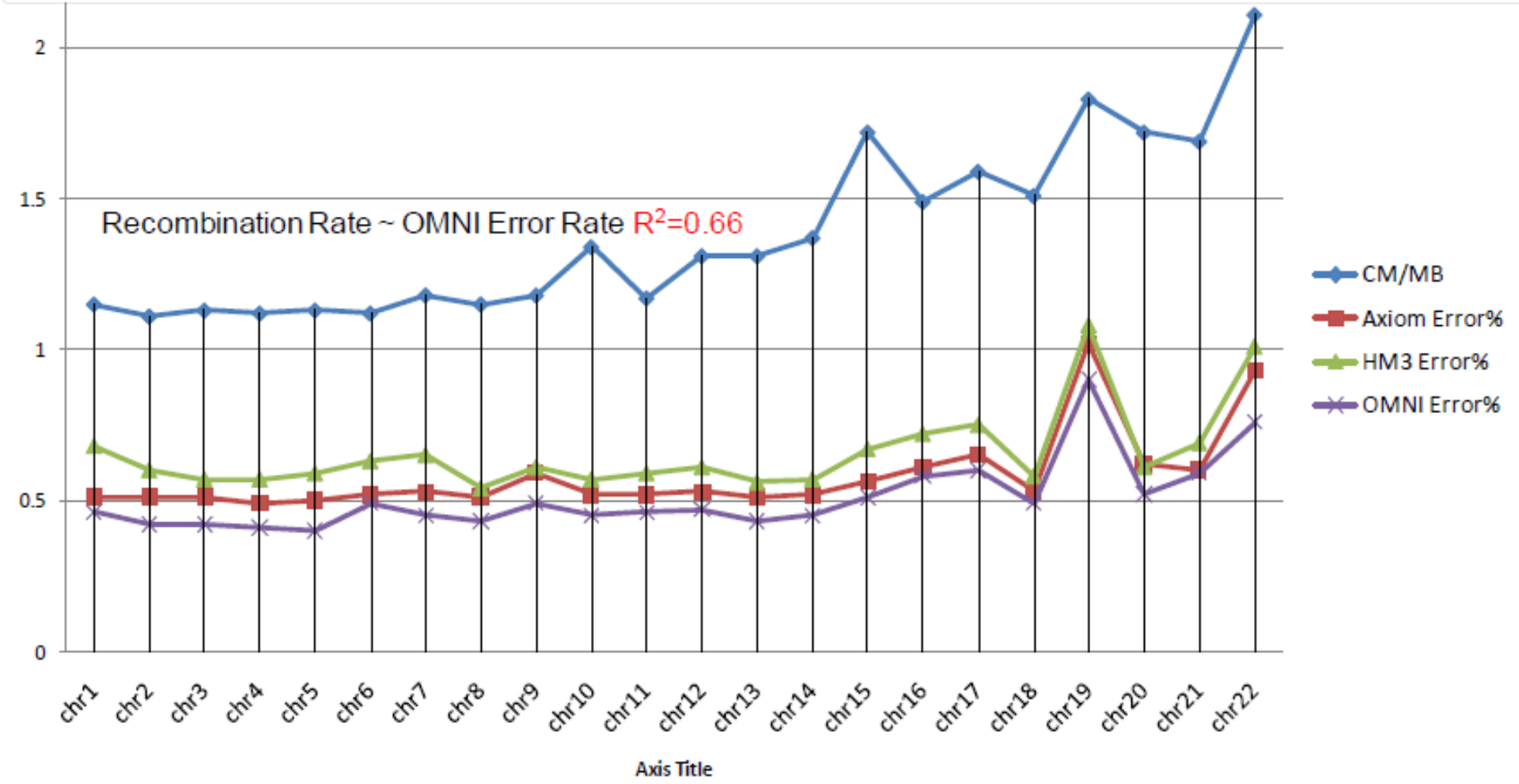

OMNI Overall Genotype Disconcordance%
Overall Genotype Disconcordance%
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
BBMM/SNPTools
BBMM/Beagle
UM/Beagle


HapMap3 Overall Genotype Disconcordance%
Overall Disconcordance%
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
BBMM/SNPTools
BBMM/Beagle
UM/Beagle


Axiom Overall Genotype Disconcordance%
Overall Genotype Disconcordance%
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
BBMM/SNPTools
BBMM/Beagle
UM/Beagle

Exploring Error Rate Variation





ChrX SNP Genotype Calls

ChrX-specific Genotype Calling Summary

1. Set male heterozygote genotype likelihoods to zero, except PAR
2. Normalize the male genotype likelihoods of two homozygote, except PAR
3. Apply SNPTools imputation engine on the modified GLs as autosome
4. In the best-guess stage, forbid male heterozygote calls, except PAR

PAR=*Pseudoautosomal Region*

female	alt/alt	ref/alt	ref/ref	total	non-ref
OMNI	1.29	1.25	0.23	0.59	1.70
HapMap3	3.69	1.38	0.26	1.21	2.74
Axiom	1.18	1.89	0.35	0.85	2.08

male	alt/alt	ref/alt	ref/ref	total	non-ref
OMNI	0.91	50.48	0.18	1.20	4.76
HapMap3	2.44	38.74	0.21	1.08	3.35
Axiom	0.91	3.87	0.24	0.43	1.50

ChrX Array Comparison Result

- Female's genotype dis-concordance rate is close to that of autosome
- Male's heterozygote dis-concordance rate is mainly contributed by non-specific hybridization of traditional technologies in X transposed region
- HapMap3's alt/alt dis-concordance rate for ChrX seems to be an outlier

Appendix: Utilizing AP field

- **The succinct Allele Probability (AP) field can be converted to various quality metrics as follow (autosomes):**
  - **Definition of AP:P(allele=1|haplotype)**
  - **e.g. 1|0:0.900,0.300**
  - **To genotype likelihoods (GL, before log10 scaling)**
    - *P(Ref/Ref)=(1-0.9)*(1-0.3)* *=0.07*
    - *P(Ref/Alt)=(1-0.9)*0.3+0.9*(1-0.3)* *=**0.66***
    - *P(Alt/Alt)=0.9*0.3* *=0.27*
  - **To genotype quality (GQ)**
    - *GQ=Round[-10*log(1-**0.66**)]* *=5*
  - **To allelic phasing confidence**
    - *P(1|0) : P(0|1)=0.9*(1-0.3) : (1-0.9)*0.3* *=21*

Acknowledges

- **IBM BlueBioU team at Rice University**
  - Dr. Kim Andrews
  - Roger Moye
  - Chandler Wilkerson

- **BCM IT related groups**
  - Jeffrey Reid
  - Clyde Lee
  - Garcia Eric