Exploring genomic selection with the BovineHD

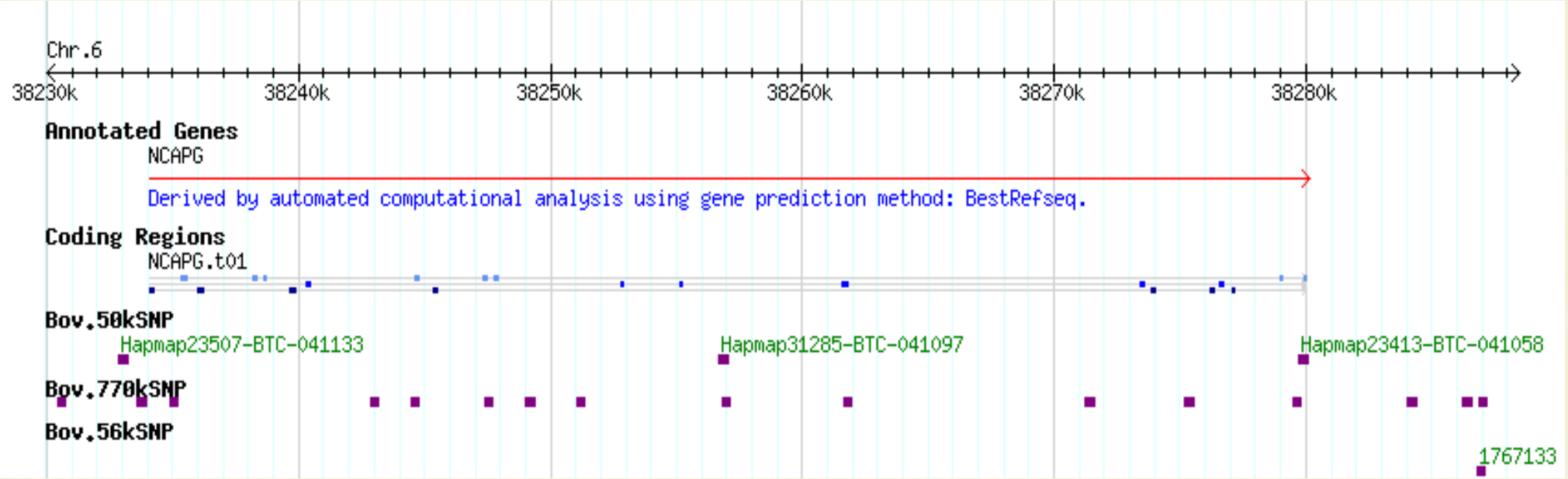
What is the BovineHD?

- Illumina BovineHD Beadchip
  >777,000 SNP
  ~3,500 bases between SNP

- Affymetrix Axiom Genome-wide BOS-1 Array Plate
  >640,000 SNP
  ~4,200 bases between SNP

How do the HD and 50K compare?

- BovineSNP50 v1 – 54,001
- BovineSNP50 v2 – 54,609
  - 52,340 common
  - 1,661 removed
  - 2,269 added



www.animalgenome.org

What can we expect from HD genotypes?

- whole-genome QTL “fine-mapping”
  - dense genotypes eliminate some SNP discovery and re-genotyping to refine QTL identified from microsatellite and 50K scans
    - re-genotyping might focus on SNP likely to have functional effects
      - SNP mined from public databases and next-generation sequence
        - 9.5 million *Bos taurus* in dbSNP
        - millions more in NGS data sets

What can we expect from HD genotypes?

- higher accuracy genomic evaluations
  - genomic selection accuracy affected by linkage disequilibrium between markers and QTL, and accuracy of marker effect estimates
    - HD SNP in higher LD with unknown causative mutations
    - complicated by more SNP in LD with unknown causatives

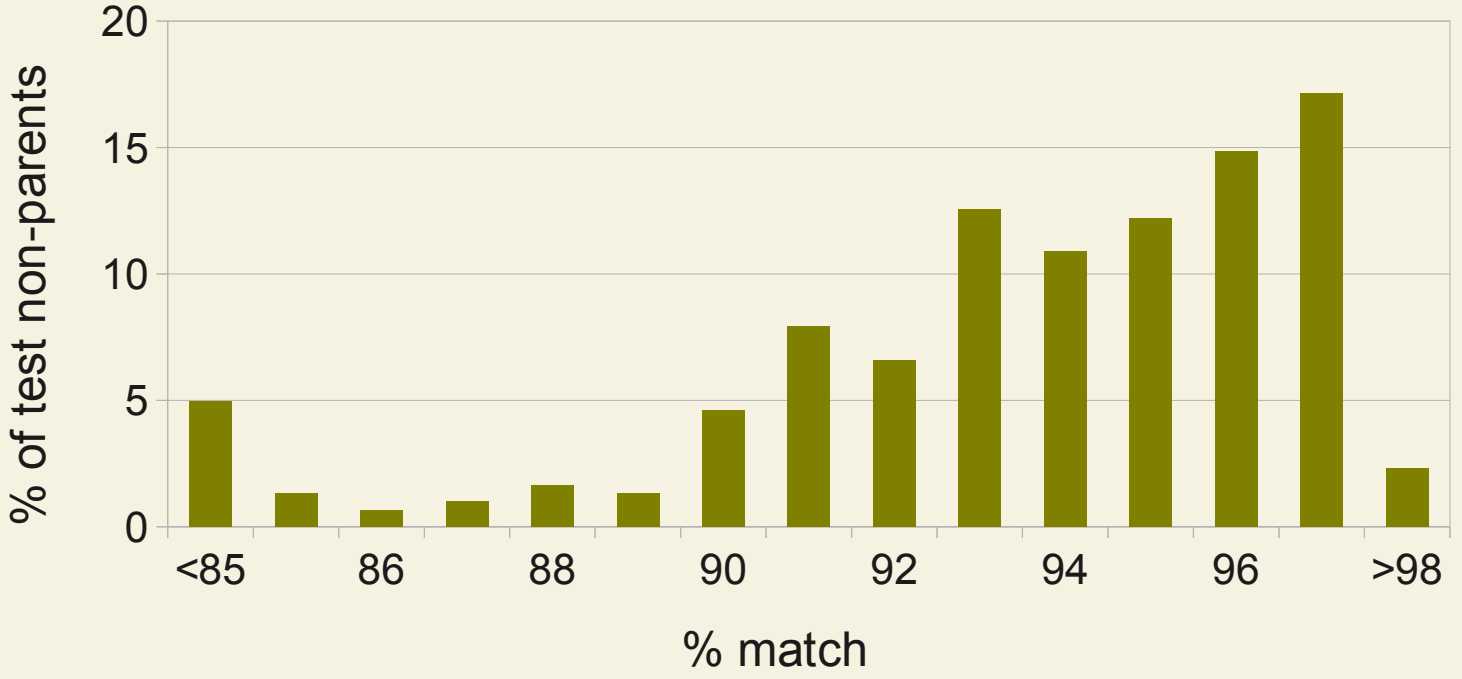
Do we need HD on everyone?

- 50K and lower density can be imputed to HD using HD genotypes of reference animals

- 50K->HD imputation for USMARC GPE
  - 18,182 animal pedigree
  - 9,644 genotyped - 950 HD, 8,694 50K
    - 2,418 sires – 482 HD, 358 50K
    - 8,029 dams – 143 HD, 924 50K
    - 7,739 non-parents – 325 HD, 7,411 50K
  - 9,777 imputed HD (findhap.f90, Van Raden)
    - 133 ungenotyped dams imputed from genotyped progeny & mates

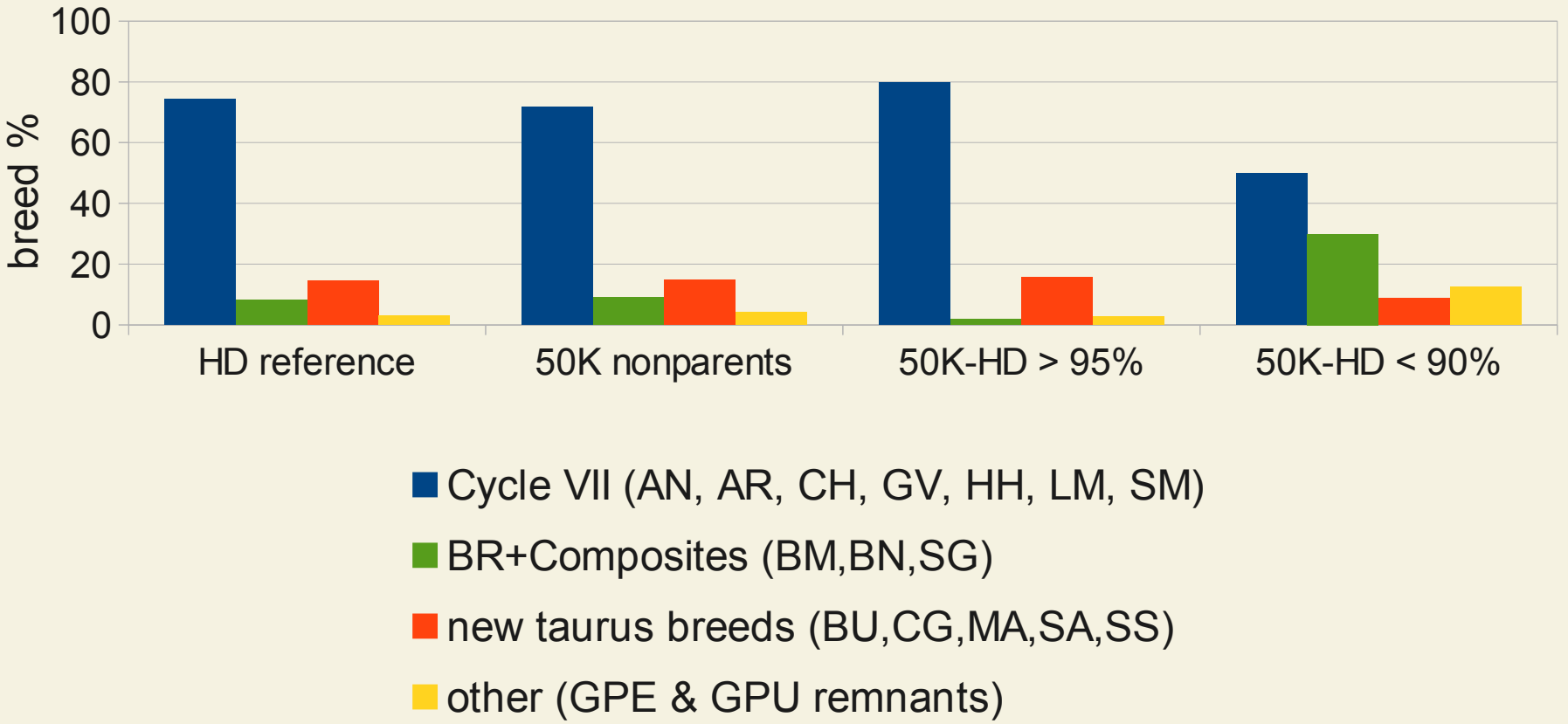
How accurate are imputed genotypes?

- 50K->HD in USMARC GPE
  - test using 50K of non-parents having HD genotypes
  - compared HD calls to imputed HD
    - 93% of called and imputed genotypes agree
    - individual agreement 27.4 to 98.3 %
      - 89% of test animals have >90% agreement



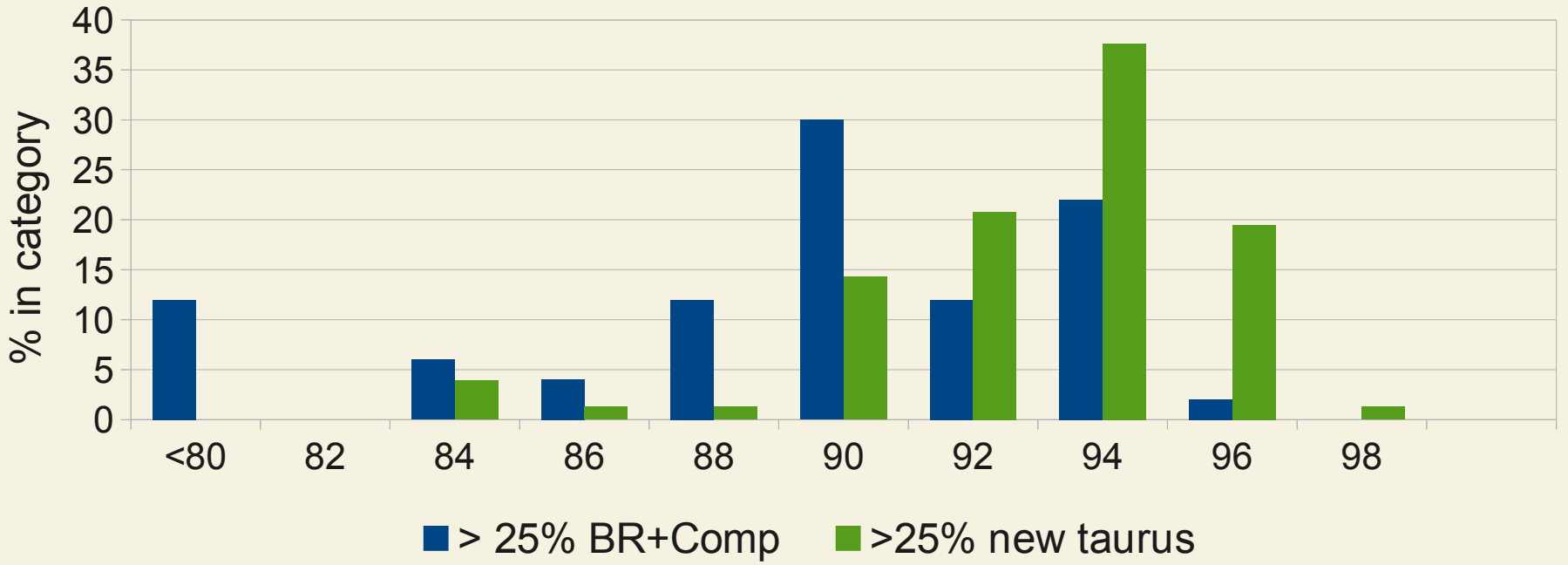
What influences imputation accuracy?

- breed composition of HD reference and 50K



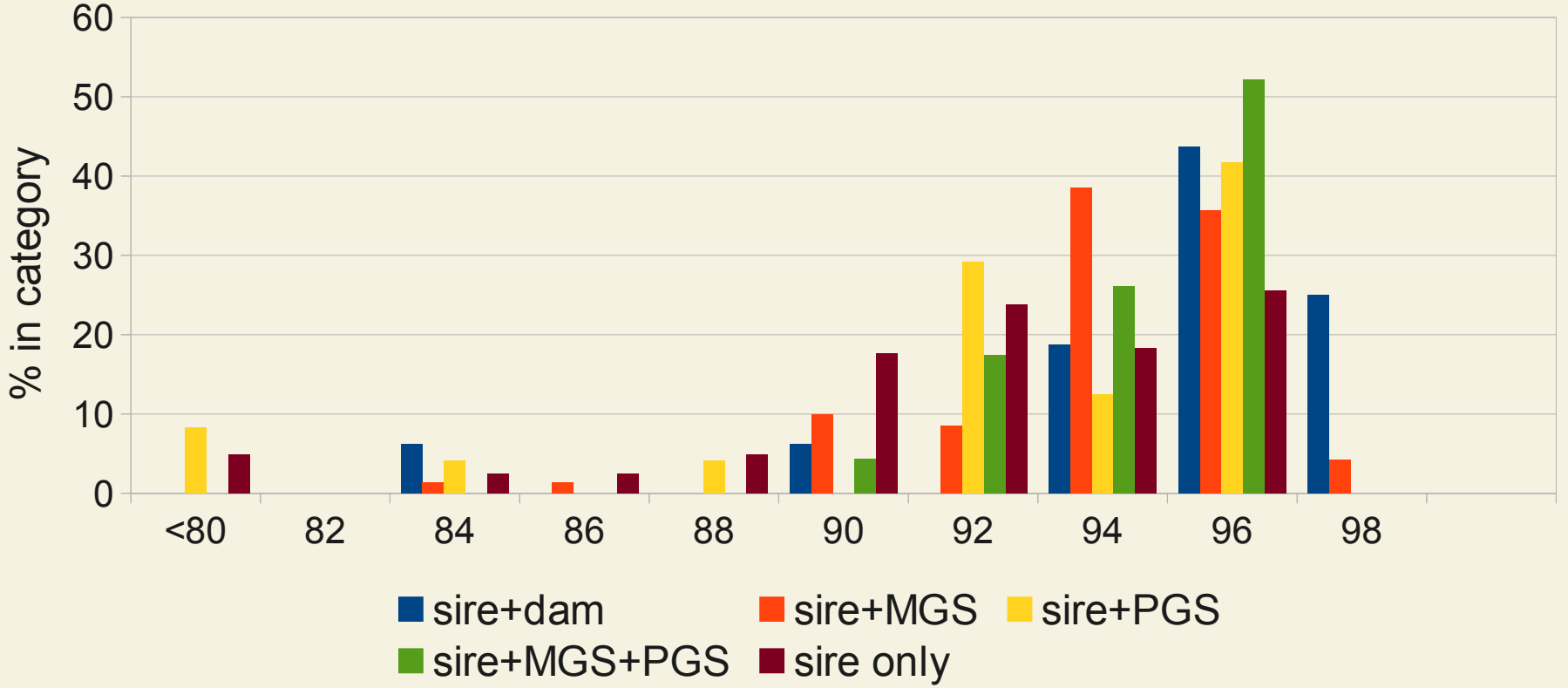
What influences imputation accuracy?

- breed composition of HD reference and 50K



What influences imputation accuracy?

- parents & grandsires in HD reference



Does the HD give different answers?

- Cycle VII and new GPE

Phenotypes

Data set	Birth weight	Ribeye area
Cycle VII	2940	1738
Cycle VII + new GPE	6752	2667

Marker sets

HD	630,571	BovineHD SNP on BTA 1-29, MAF > 0.05
50K	39,366	HD subset on both BovineSNP50 versions

Does the HD give different answers?

- Birth weight heritability

Data Set	Marker Set	h^2
Cycle VII	HD	.64 (.03)
	50K	.63 (.03)
	none	.60 (.04)
VII + new	HD	.64 (.02)
	50K	.58 (.02)
	none	.60 (.03)

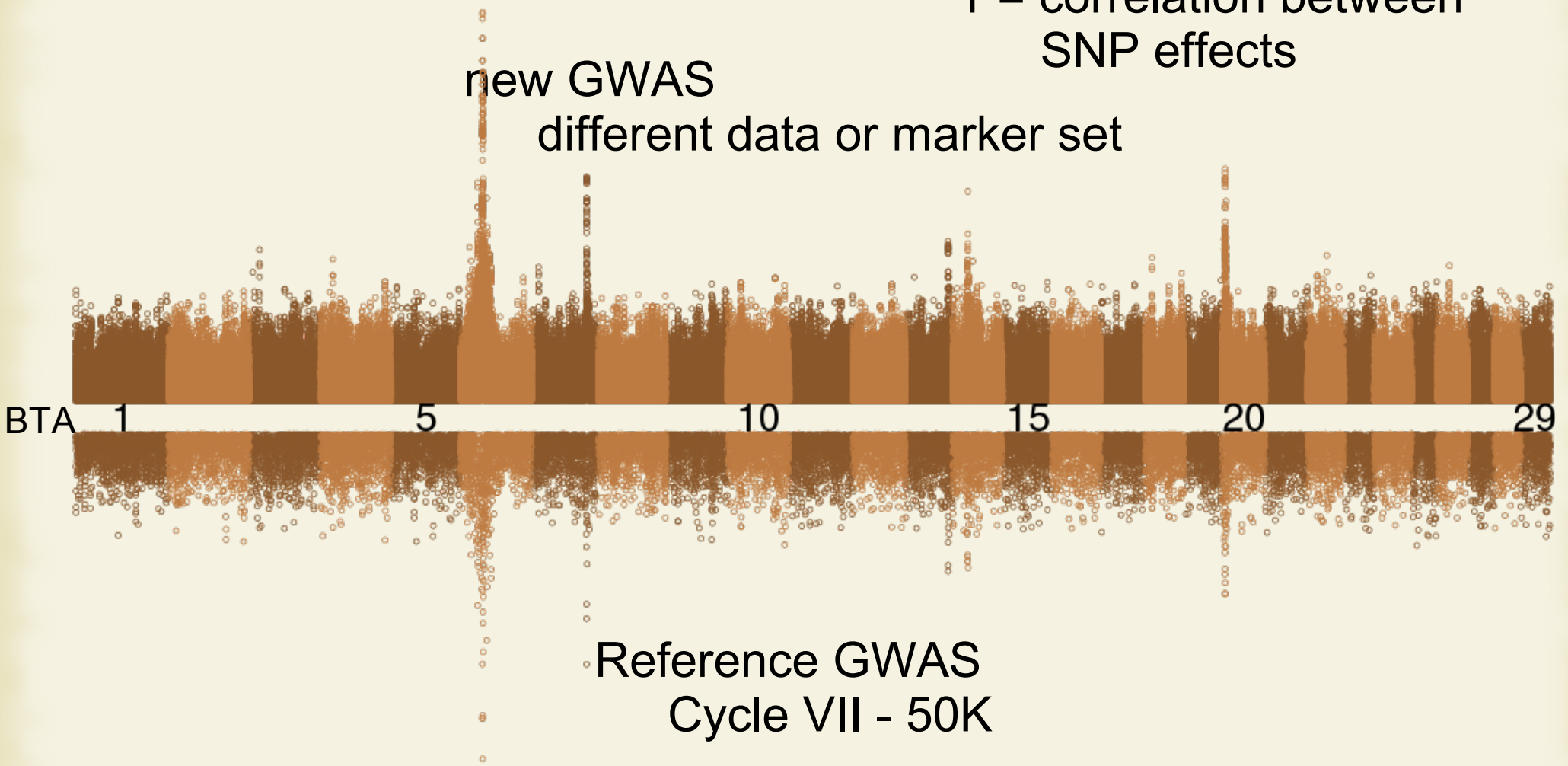
Does the HD give different answers?

- Ribeye area heritability – ribeye area

Data Set	Marker Set	h^2
Cycle VII	HD	.50 (.05)
	50K	.47 (.05)
	none	.54 (.07)
VII + new	HD	.50 (.04)
	50K	.47 (.04)
	none	.53 (.06)

Does the HD give different answers?

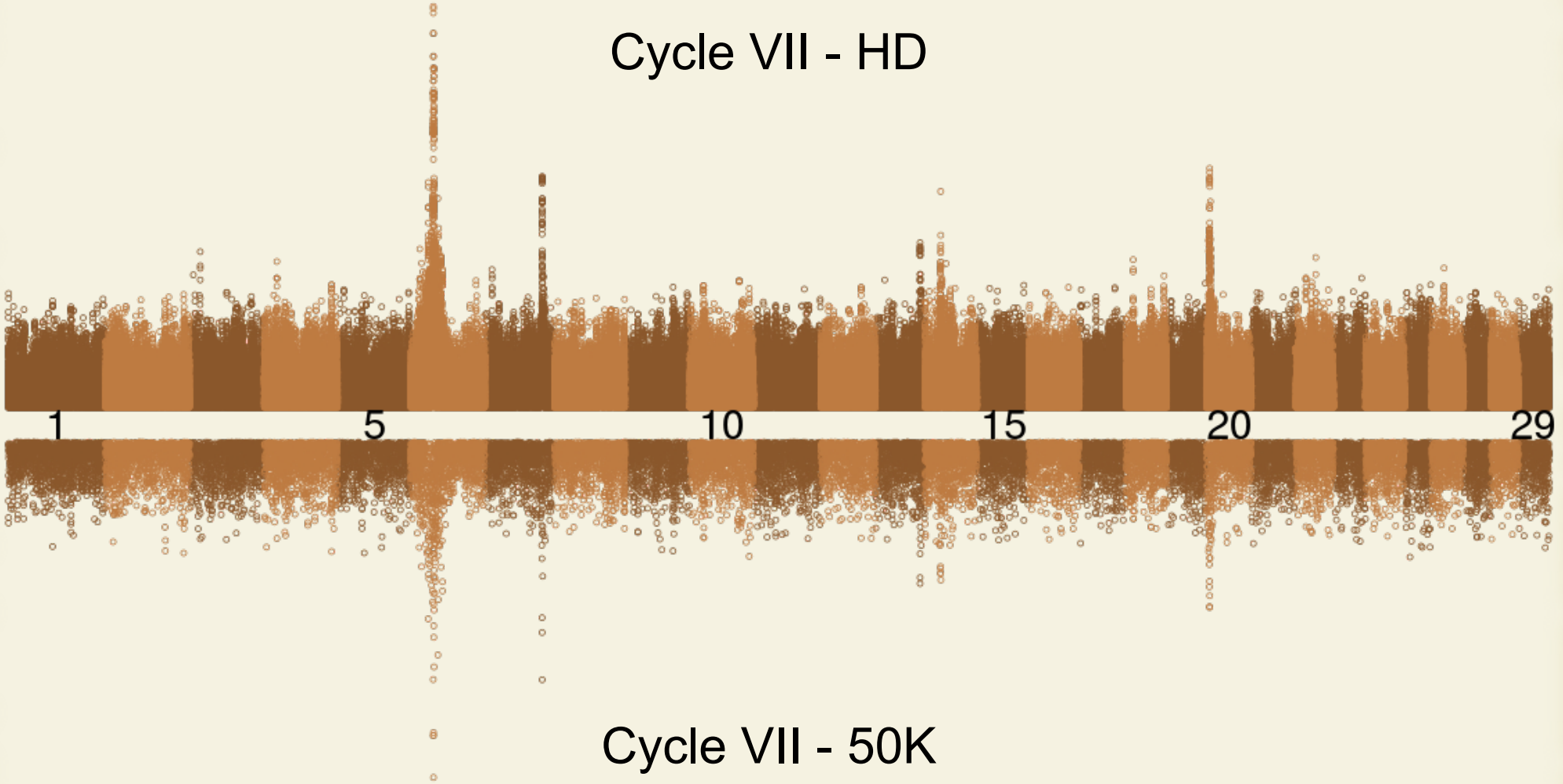
- Trait



Does the HD give different answers?

- Birth weight

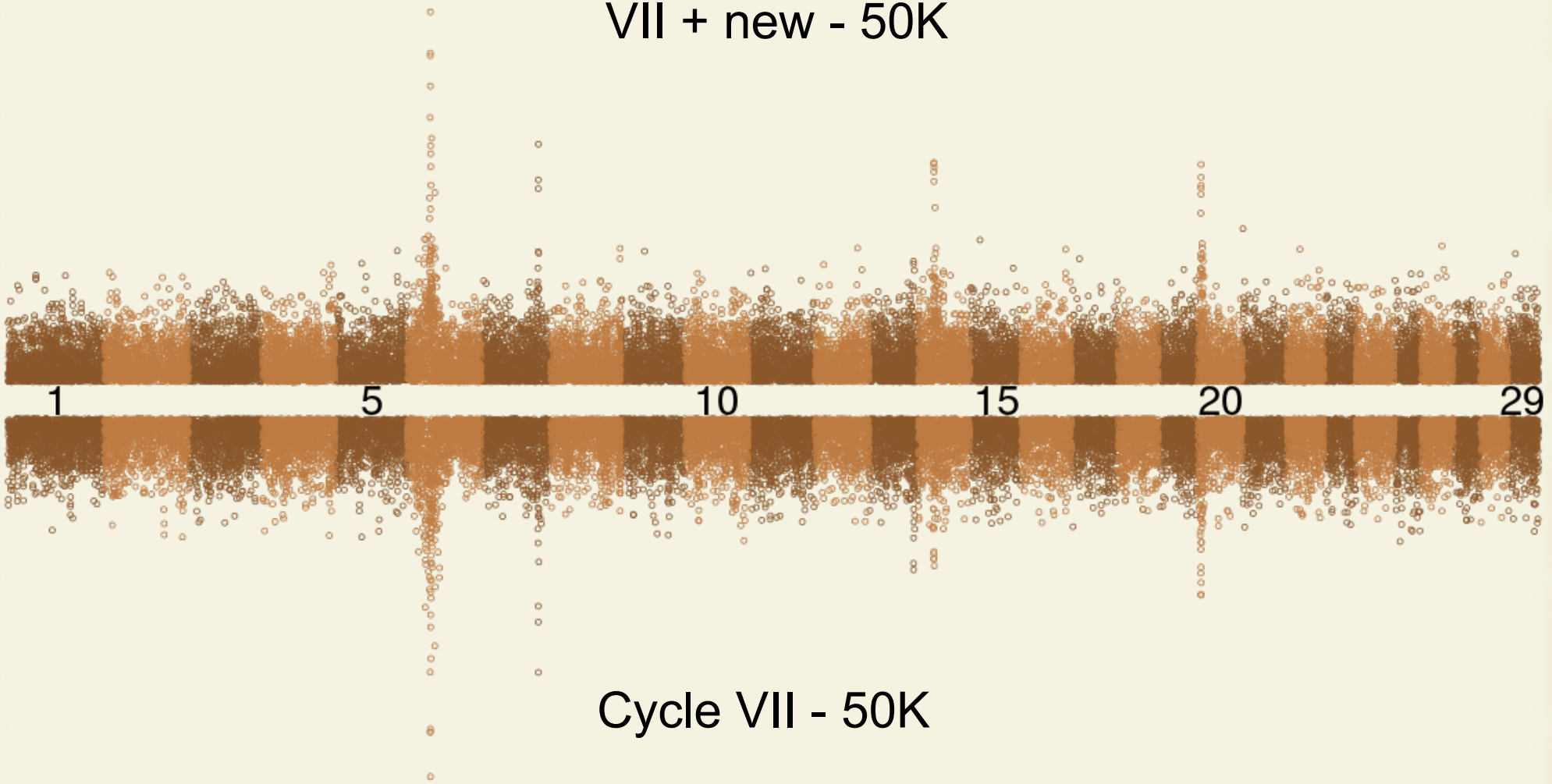
r = .989



Does the HD give different answers?

- Birth weight

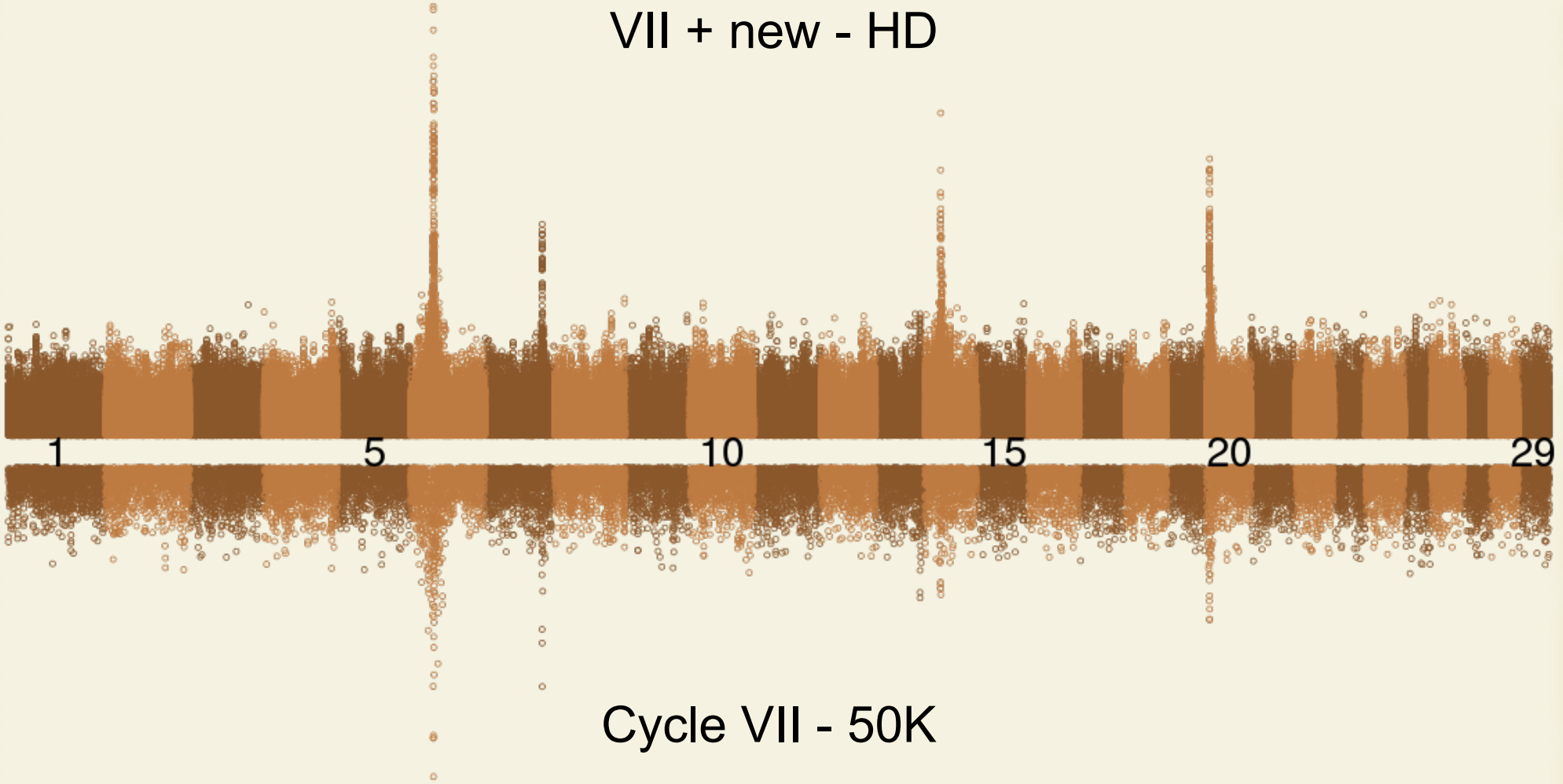
r = .664



Does the HD give different answers?

- Birth weight

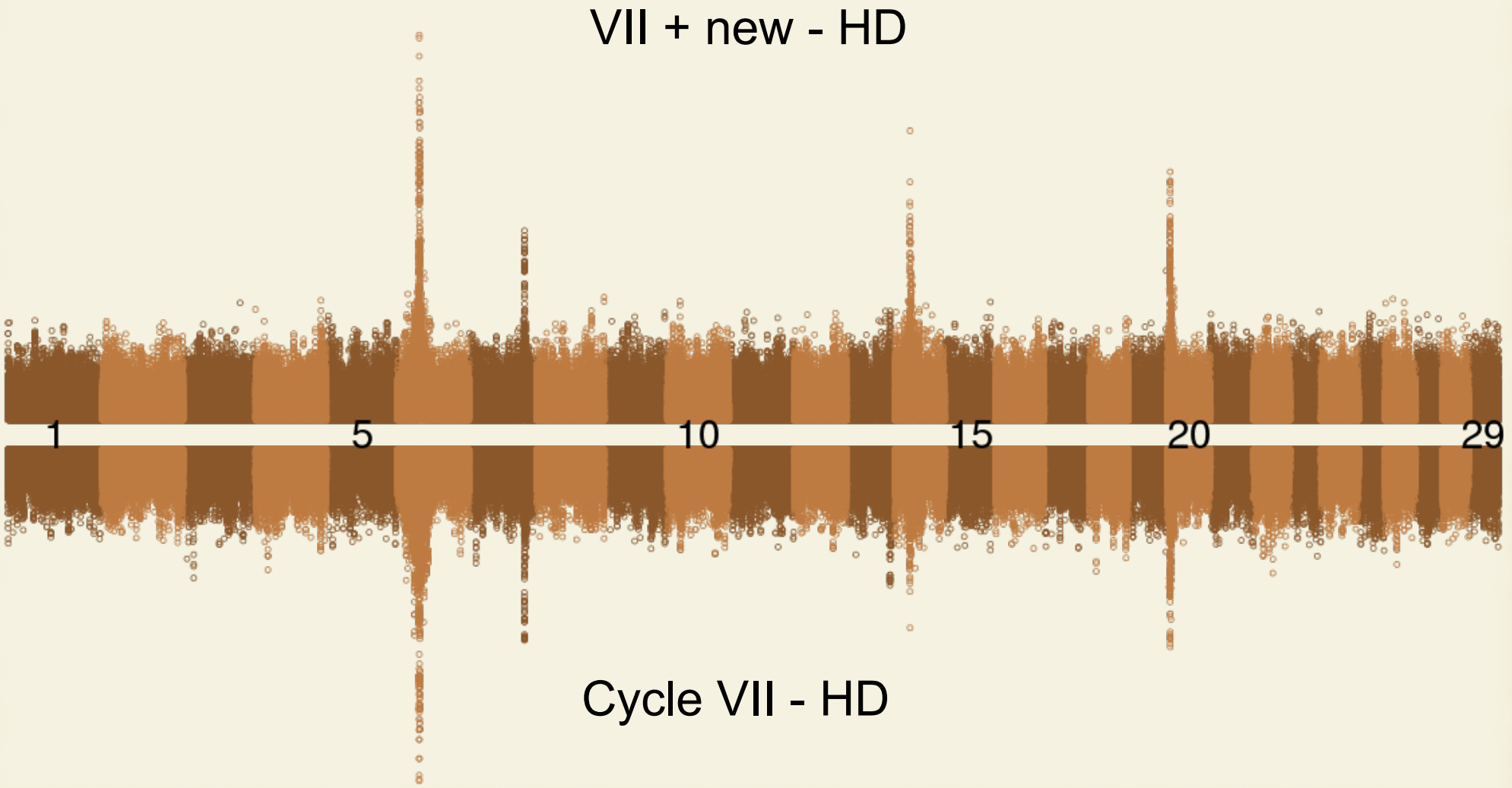
r = .658



Does the HD give different answers?

- Birth weight

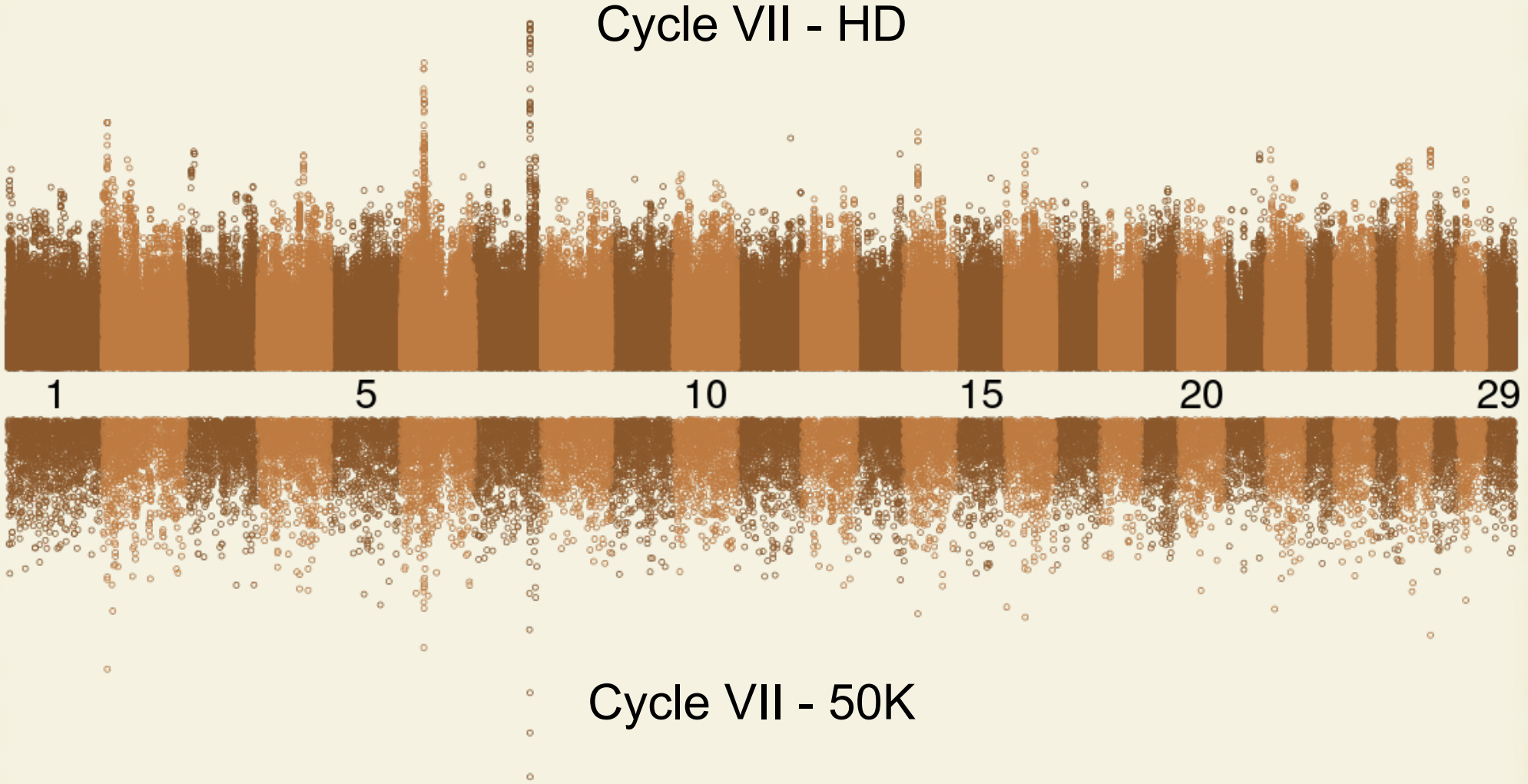
r = .617



Does the HD give different answers?

- Ribeye area

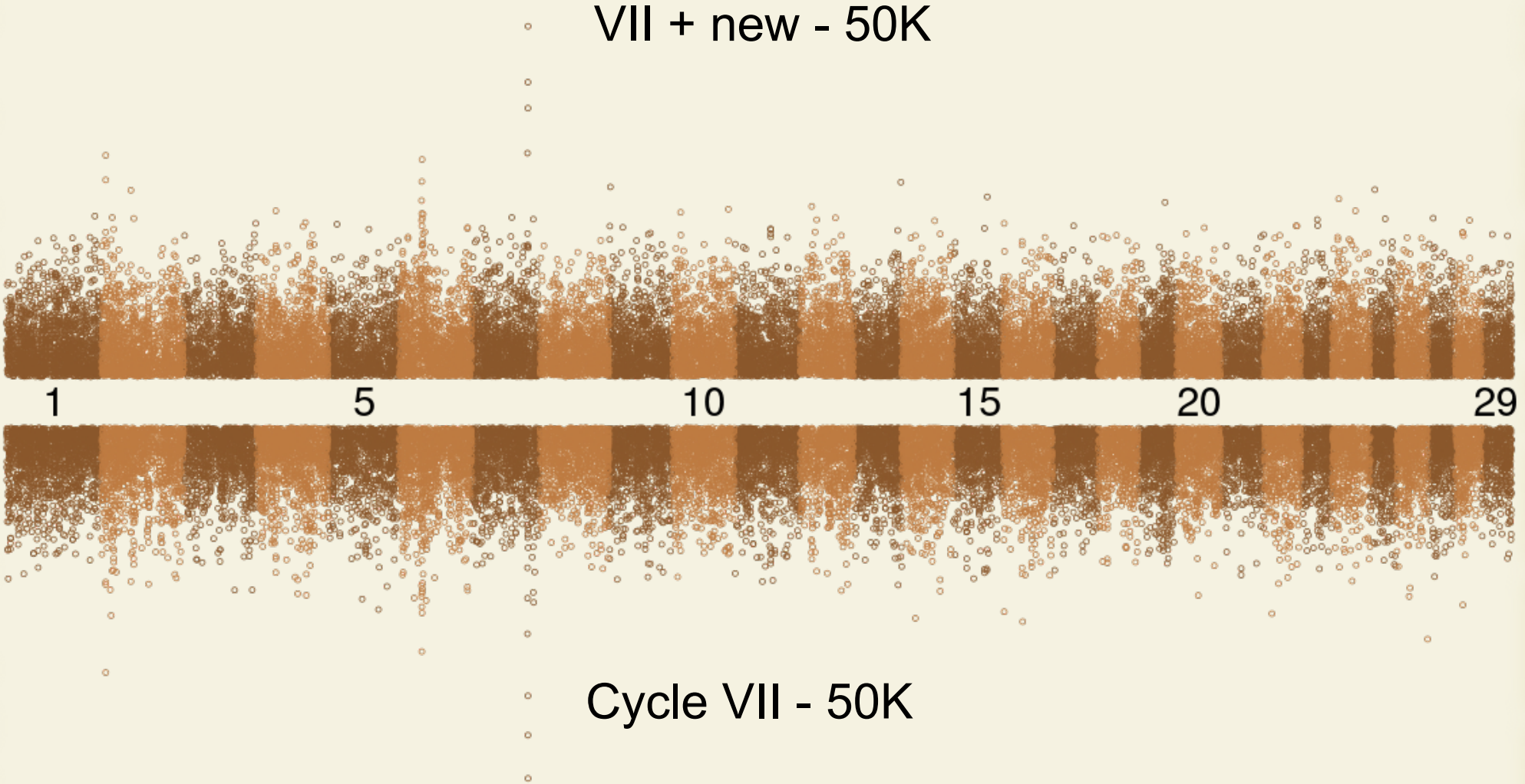
r = .995



Does the HD give different answers?

- Ribeye area

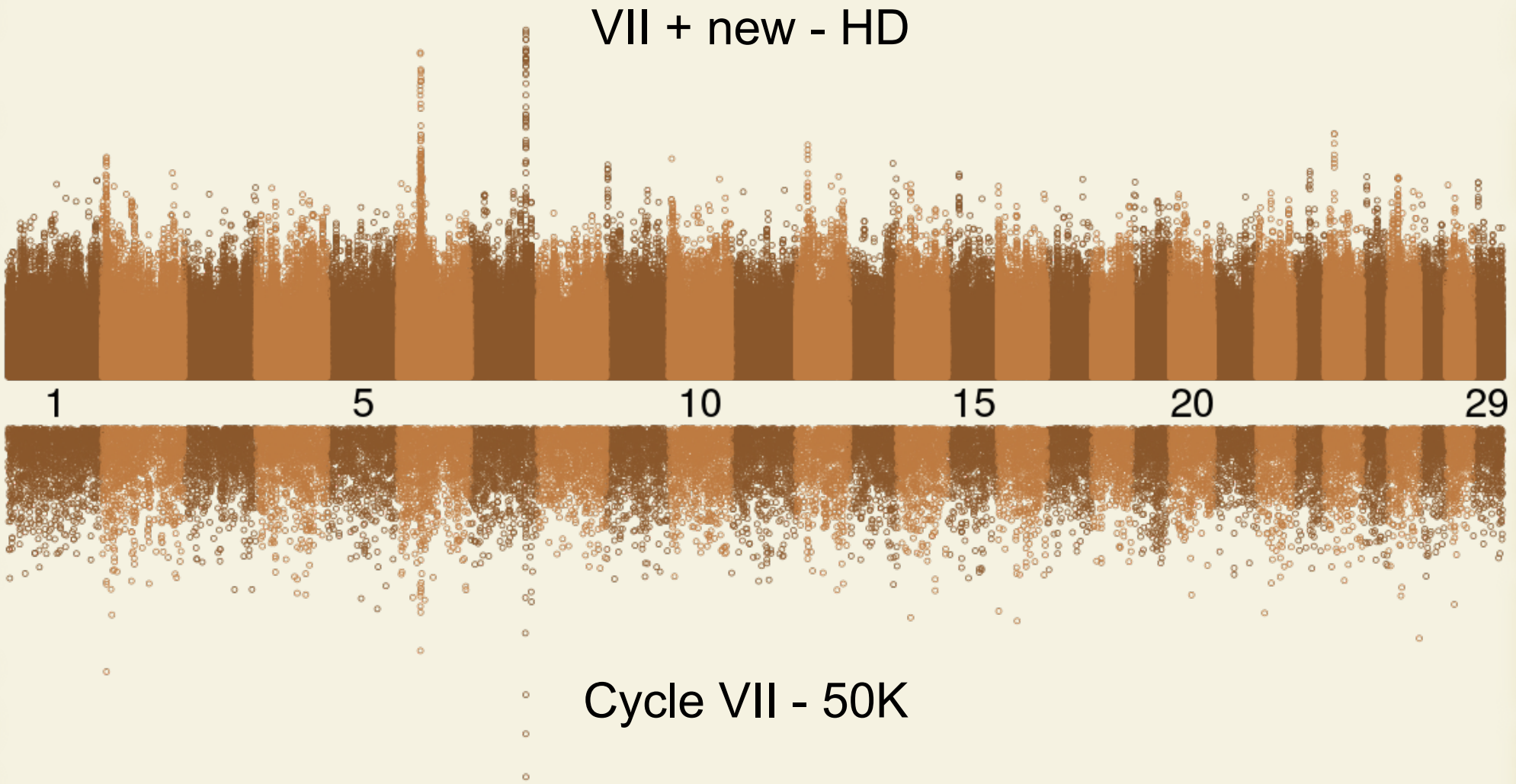
r = .764



Does the HD give different answers?

- Ribeye area

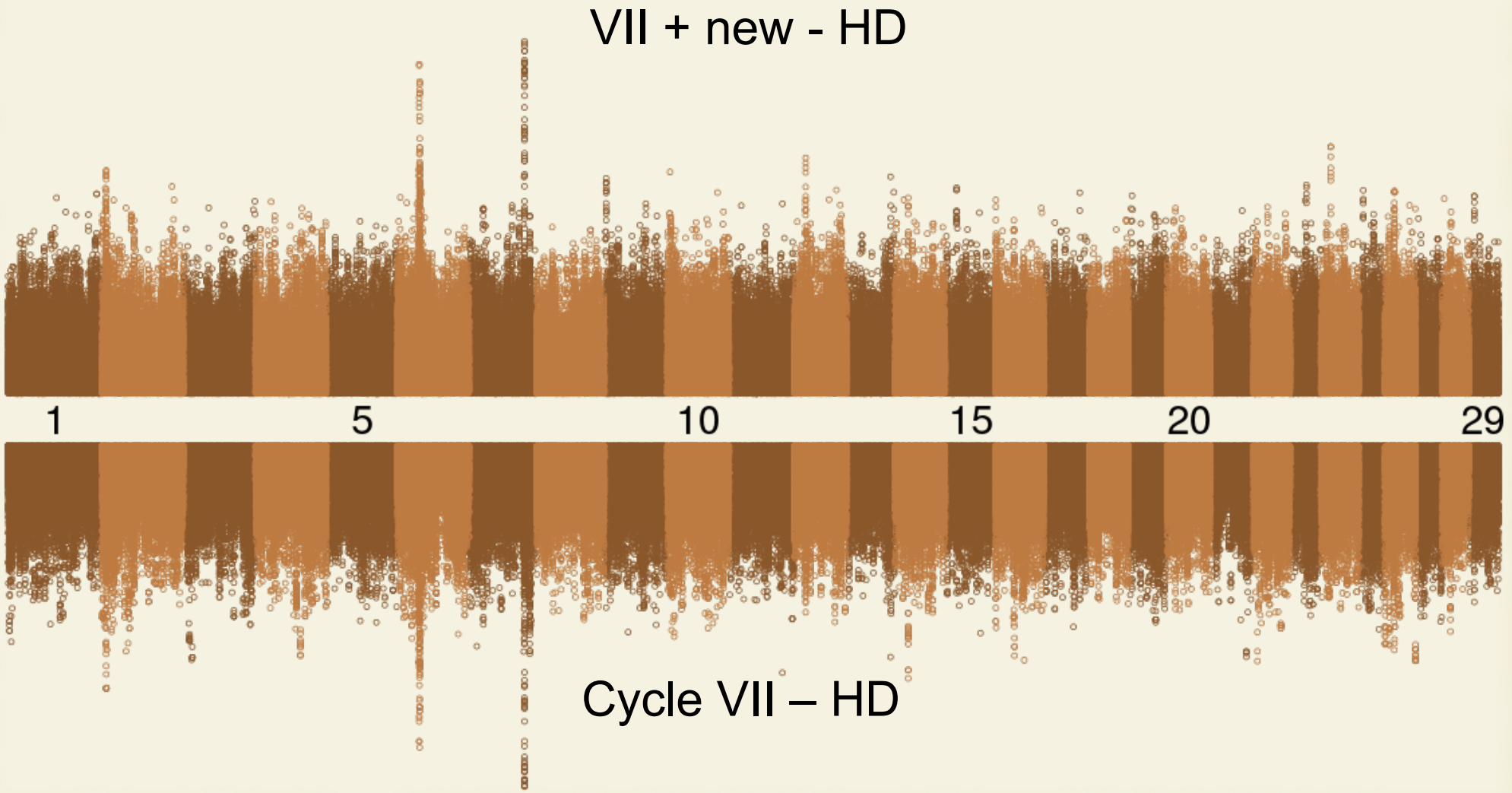
r = .763



Does the HD give different answers?

- Ribeye area

r = .719



Does the HD give different answers?

- Cycle VII trained predictions of new GPE
  - Cycle VII + new GPE pedigree
  - new GPE phenotypes
  - corresponding MBV or EBV
    - MBV = CycleVII SNP effects x newGPE genotypes
    - EBV = CycleVII pedigree estimates of newGPE

Does the HD give different answers?

- Cycle VII trained predictions of new GPE
  - Birth weight – BW MBV genetic correlations

	$r_{BW.MBV}$	r^2
HD	.52 (.05)	.27
50K	.52 (.05)	.27
EBV (pedigree est)	.24 (.05)	.06

n=3812

Does the HD give different answers?

- Cycle VII trained predictions of new GPE
  - Ribeye area – REA MBV genetic correlations

	$r_{REA.MBV}$	r^2
HD	.27 (.12)	.07
50K	.30 (.12)	.09
EBV (pedigree est)	.11 (.08)	.01

n=929

Does the HD give different answers?

- Agreement between newGPE predictions
  - Birth weight

	$r_{gEBV.MBV}$	$r_{gEBV.maMBV}$
HD	.57	.95
50K	.56	.94
EBV (pedigree est)	.31	.89
EBV (VII + newGPE)		.90

gEBV – from HD VII+newGPE
MBV – VII effects x newGPE genotypes
maMBV – 2-trait newGPE phenotypes, MBV

n=3812

Does the HD give different answers?

- Agreement between newGPE predictions
  - Ribeye area

	$r_{gEBV.MBV}$	$r_{gEBV.maMBV}$
HD	.40	.97
50K	.41	.96
EBV (pedigree est)	.22	.94
EBV (VII + newGPE)		.95

gEBV – from HD VII+newGPE
MBV – VII effects x newGPE genotypes
maMBV – 2-trait newGPE phenotypes, MBV

n=929

Are all HD SNP needed for evaluation?

- focus on important genes and regions
- polygenic selection for loci with subtle effects
  - eliminate noise from whole-genome SNP

Small HD sets evaluated in Cycle VII and new GPE

	n SNP	Source
GSA	4,795	HD subset in gene sets implicated by Cycle VII growth & carcass GWAS
LDa	105	HD subset spaced around Cycle VII growth and carcass GWAS peaks

Are all HD SNP needed for evaluation?

- Heritability estimates

Data Set	Marker Set	Birth Weight	Ribeye area
Cycle VII	GSA	.25 (.03)	.20 (.03)
	LDa	.11 (.02)	.07 (.02)
	none	.60 (.04)	.54 (.07)
VII + new	GSA	.22 (.02)	.19 (.03)
	LDa	.09 (.02)	.05 (.01)
	none	.60 (.03)	.53 (.06)

Are all HD SNP needed for evaluation?

- Cycle VII trained predictions of new GPE
  - Birth weight – BW MBV genetic correlations

	$r_{BW.MBV}$	r^2
GSA	.31 (.06)	.09
LDa	.38 (.05)	.14
EBV (pedigree est)	.24 (.05)	.06

n=3812

Are all HD SNP needed for evaluation?

- Cycle VII trained predictions of new GPE
  - Ribeye area – REA MBV genetic correlations

	$r_{REA.MBV}$	r^2
GSA	.18 (.12)	.03
LDa	.19 (.11)	.04
EBV (pedigree est)	.11 (.08)	.01

n=929

Are all HD SNP needed for evaluation?

- Agreement between newGPE predictions
  - Birth weight

	$r_{gEBV.MBV}$	$r_{gEBV.maMBV}$
GSA	.30	.91
LDa	.36	.91
EBV (pedigree est)	.31	.89
EBV (VII + newGPE)		.90

gEBV – from HD VII+newGPE
MBV – VII effects x newGPE genotypes
maMBV – 2-trait newGPE phenotypes, MBV

n=3812

Are all HD SNP needed for evaluation?

- Agreement between newGPE predictions
  - Ribeye area

	$r_{gEBV.MBV}$	$r_{gEBV.maMBV}$
GSA	.21	.94
LDa	.15	.94
EBV (pedigree est)	.22	.94
EBV (VII + newGPE)		.95

gEBV – from HD VII+newGPE
MBV – VII effects x newGPE genotypes
maMBV – 2-trait newGPE phenotypes, MBV

n=929

What's next?

- New developments that might increase portability of genomic predictions?
  - next-generation sequence, exon sequence, RNA-seq
    - identify and classify (likely) functional variants
    - reference to impute from panel genotypes (LD,50K,HD) to sequence
  - gene pathways and networks
    - integrate functional annotation, gene expression and other information to identify interacting genes likely to affect phenotype

Summary

- HD genotyping panels offer opportunity for high-resolution GWAS and QTL fine-mapping
- Existing 50K genotypes can be imputed to HD
  - need suitable HD reference genotypes
    - paternal and maternal haplotypes in reference
- Genomic selection similar using 50K or HD
  - limited portability
  - causal variants needed?
- Resources and methodology to better identify causal variants being developed
  - incremental advances or quantum leap?