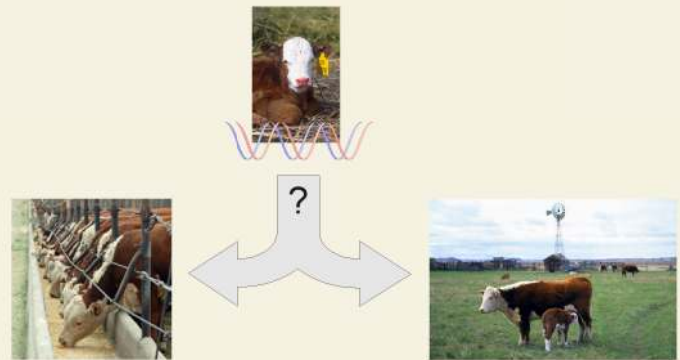


Pathways to genomic analysis of heifer puberty and pregnancy

Can information besides genotypes and phenotypes assist genomic selection for complex traits?

- enable robust DNA tests for fertility traits?



Sources of information for genomic analysis

- Individual genotypes and phenotypes
- Allele frequencies of DNA pooled by phenotype
- Gene expression
- Genome assembly and annotation

Integrated analysis

- examples applying supplemental information to imputed HD genotypes, puberty and pregnancy of GPE heifers

Information sources

- Individual genotypes and phenotypes



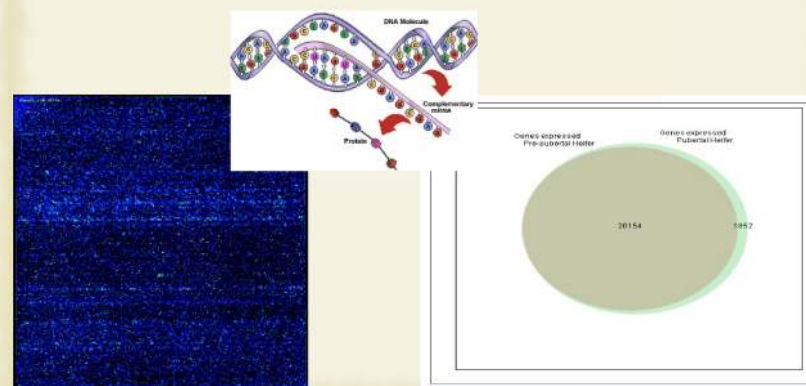
Information sources

- Allele frequencies of DNA pooled by phenotype



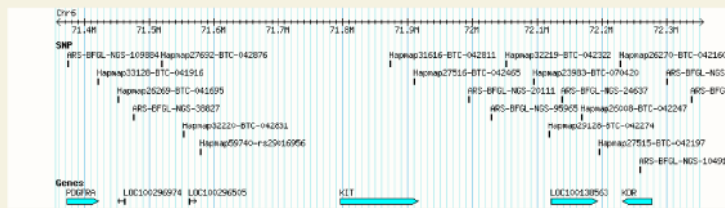
Information sources

- Gene expression
 - sequence and microarray technologies detect presence and abundance of mRNA coded by specific genes



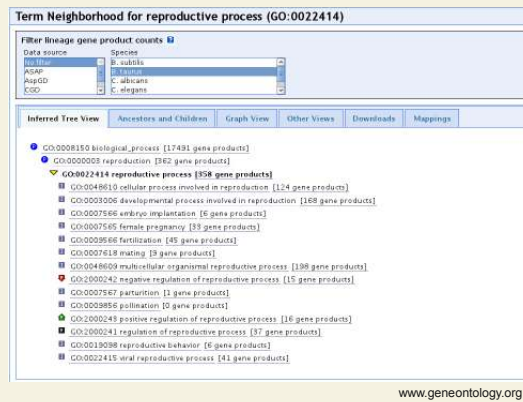
Information sources

- Genome assembly and annotation
 - base pair sequence, positions of SNP, genes, other features



Information sources

- Genome assembly and annotation
 - gene function, interactions and regulation

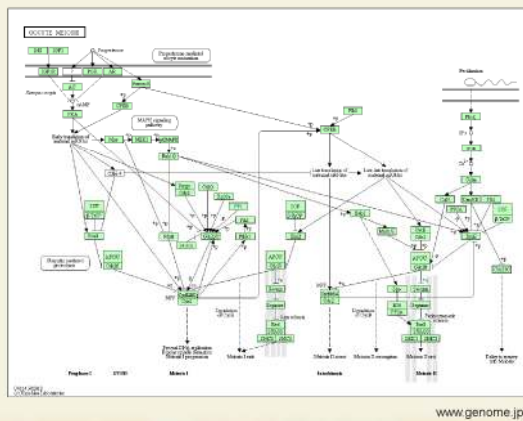


Gene ontology – controlled vocabulary of terms describing gene product characteristics

- cellular components
- biological processes
- molecular function

Information sources

- Genome assembly and annotation
 - gene function, interactions and regulation



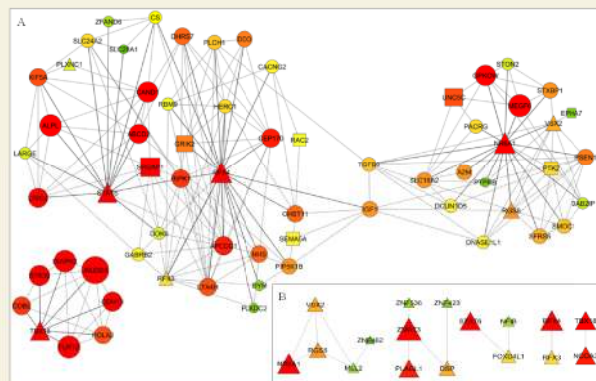
KEGG pathways - database representing current knowledge of molecular interactions and reactions

- metabolism
- genetic information processing
- environmental information processing
- cellular processes
- organismal systems
- human diseases

KEGG pathway bta04114 oocyte meiosis

Information sources

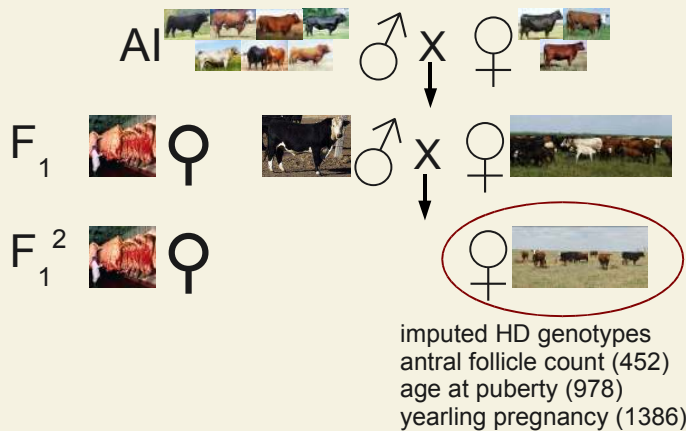
- Genome assembly and annotation
 - gene function, interactions and regulation



Key transcription factors and regulated genes in Brangus hypothalamic network (Fortes et al., in preparation)

Data for integrated 770K (HD) analysis

- GPE Cycle VII heifers



HD imputation

- 324 individuals with HD genotypes
 - AI founders & F₁ bulls, dams without 50K genotypes
- 4,525 individuals with 50K
- 10,899 individuals in pedigree
- findhap (Van Raden et al, 2010) results
 - 5,096 individuals imputed
 - 95% test bull agreement
 - imputed vs called HD genotypes of two bulls
 - 97% test bull progeny agreement
 - imputed vs called 50K of test bulls' progeny
 - >99% imputation rate

Integrated analysis with Illumina HD SNP

One approach

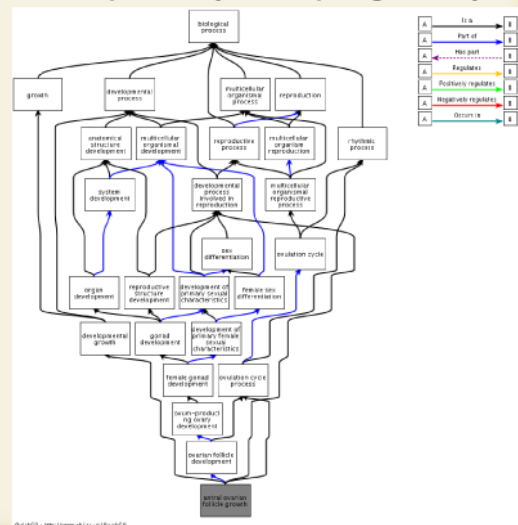
- select subset of HD (770K) using external information
- conduct genomic analysis
 - build and invert genomic relationship matrix (G)
 - replace A^{-1} with G^{-1} in mixed model equations
 - REML to estimate heritability due to SNP in subset
 - BLUP to predict individual animal gEBV
 - solve SNP effects from gEBV

$$\hat{g} = M' [MM']^{-1} \hat{u} \rightarrow \text{formula to predict interim gEBV as } \hat{u} = M\hat{g}$$
- (optional) select informative SNP and repeat

HD analysis of heifer puberty and pregnancy

Analysis 1

- GO term "antral ovarian follicle growth"
- 12,148 HD SNP near 273 genes
 - pure candidate gene set



HD analysis of heifer puberty and pregnancy

Analysis 2

Enriched GO terms and KEGG pathways from 6-trait 50K genomic analysis of GPE heifers

- Body traits (YW,PWG,BCS) 69,963 HD SNP near 3,570 genes
- Reproduction traits (AFC,AAP,Preg) 20,314 HD SNP near 695 genes
- All traits 80,214 HD SNP near 4,350 genes
- Positional candidate gene sets

Genomic heritabilities and correlations estimated from GPE heifers

	YW	PWG	AFC	AAP	BCS	Preg
YW	0.54	0.83	-0.16	0.30	0.73	-0.17
PWG	0.82	0.46	-0.26	0.26	0.52	-0.04
AFC	0.08	0.06	0.44	0.37	-0.63	-0.55
AAP	-0.01	0.06	0.02	0.14	0.15	-0.33
BCS	0.28	0.22	0.03	0.02	0.09	-0.07
Preg	0.04	0.05	0.00	0.00	0.12	0.11

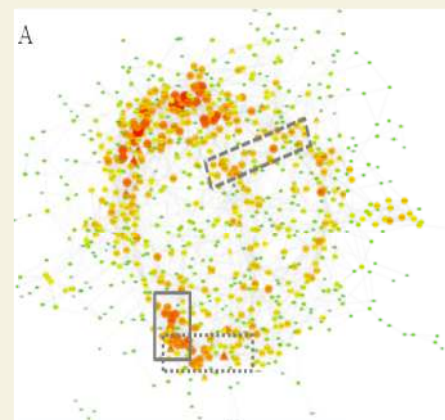
50K genomic relationship matrix genomic correlations above diagonal, phenotypic below

HD analysis of heifer puberty and pregnancy

Analysis 3

Hypothalamic gene network derived from Brangus 50K GWAS and hypothalamus transcriptomes (Fortes et al, in preparation)

- 60,650 HD SNP near 922 genes
- Positional candidate gene set from unrelated *indicus x taurus* population, supported by gene expression



HD analysis of heifer puberty and pregnancy

Analysis 4

HD SNP with different ($p < .05$) frequencies in females pooled by reproductive performance

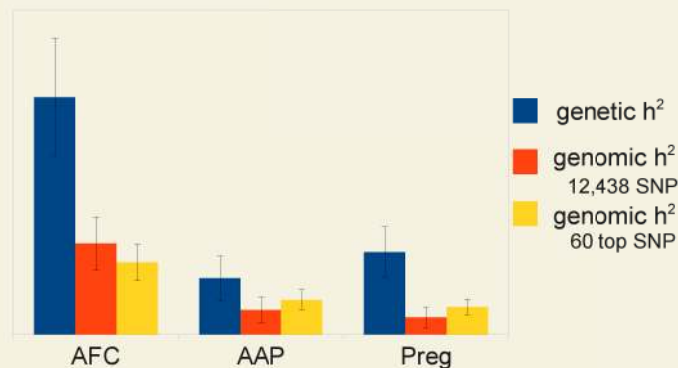
- MARC pools 20,883 SNP
- Commercial ranch pools 12,869 SNP

Reproductive performance pools

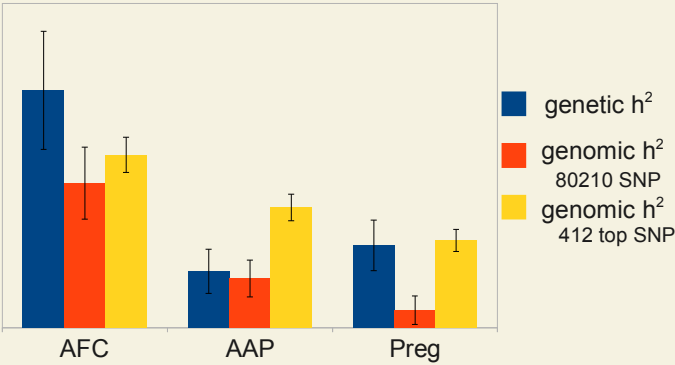
- MARC twice consecutive open culling repeat records ~15,000 ♀ pools with 1000 high & 1000 low (100/pool)
- Deseret Ranch yearling & 2-yr-old pregnancy ~3600 *indicus x taurus* ♀ pools with
 - 11% open, open
 - 30% open, pregnant
 - 16% pregnant, open
 - 43% pregnant, pregnant

Heritability estimates

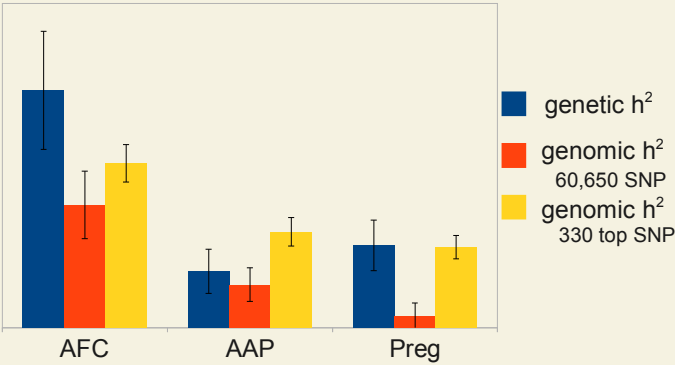
GO antral ovarian follicle growth



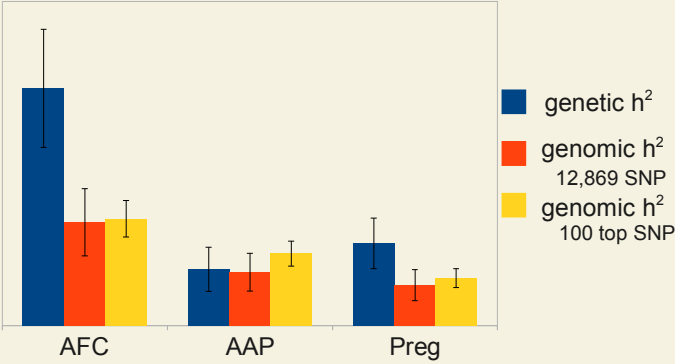
Heritability estimates
Enriched GO terms and KEGG pathways



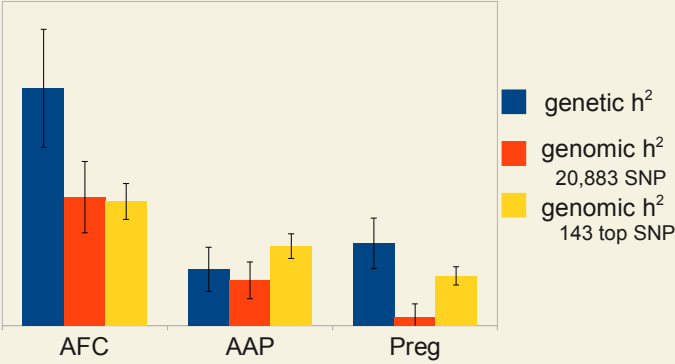
Heritability estimates
Hypothalamus gene network



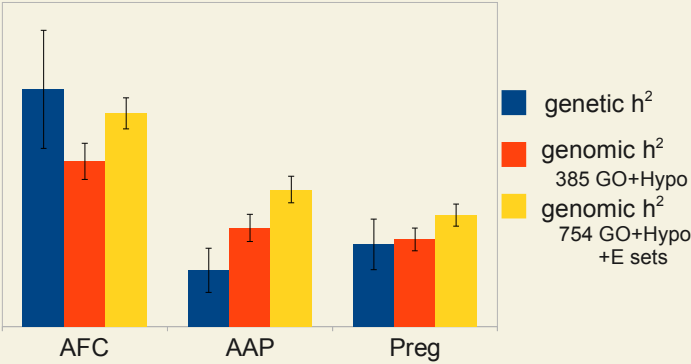
Heritability estimates
Deseret pools



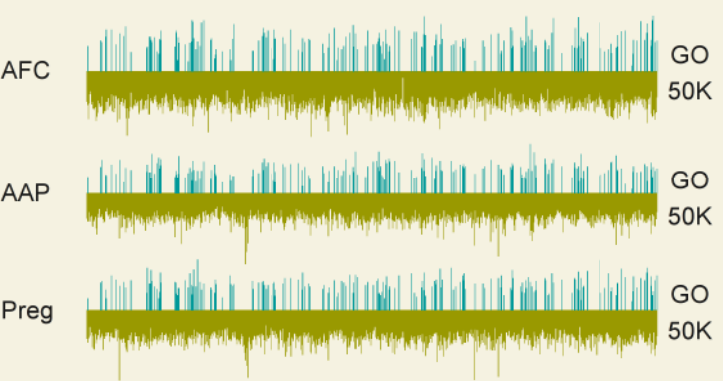
Heritability estimates
MARC pools



Heritability estimates
Top SNP combined

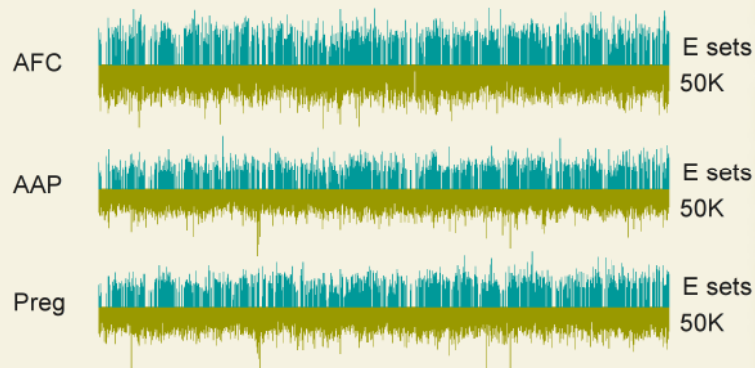


SNP effect estimates
50K & 12,438 GO antral ovarian follicle development



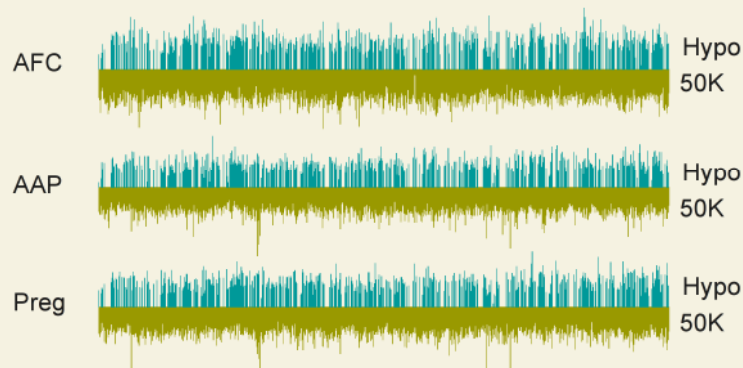
SNP effect estimates

50K & 80,210 Enriched GO and KEGG



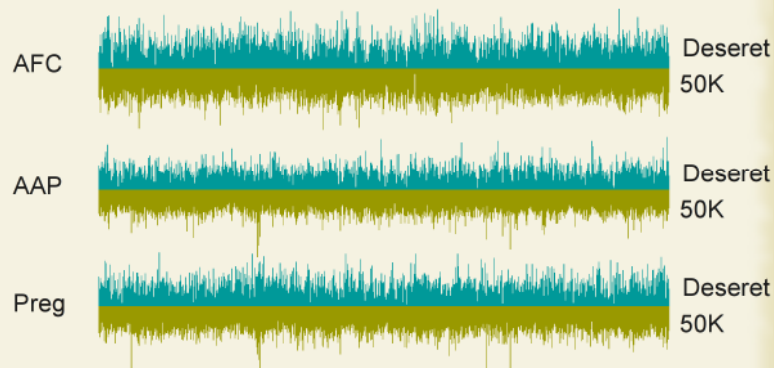
SNP effect estimates

50K & 60,650 Hypothalamus network



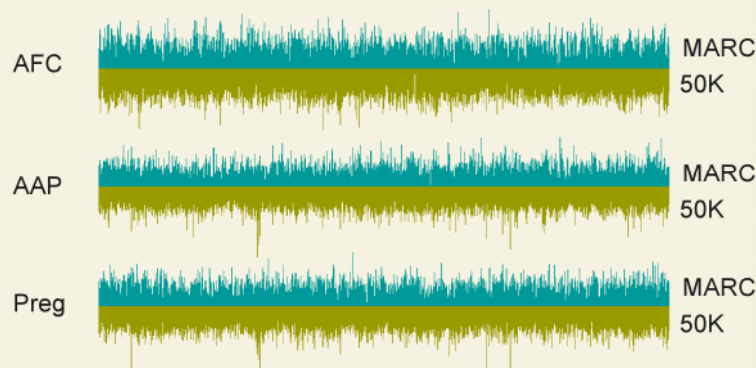
SNP effect estimates

50K & 12,869 Deseret pools



SNP effect estimates

50K & 20,883 MARC pools



Predicted yearling pregnancy – Deseret pools

- $pEBV = \sum(\text{pool allele frequency} \times \text{GPE effect})$
- pregnant pEBV > open pEBV
 - top Hypothalamus network (330 SNP)
 - top GO antral ovarian follicle + hypo (385 SNP)
 - top GO + hypo + enriched gene sets (754 SNP)
- pregnant pEBV < open pEBV
 - Deseret (all 12,869 & top 100 SNP)
 - all GO antral ovarian follicle (12,438 SNP)
 - all Hypothalamus network (60,650 SNP)
 - all enriched gene sets (80,210 SNP)

Can information besides genotypes and phenotypes assist genomic selection for complex traits?

- enable robust DNA tests for fertility traits?

preliminary results promising
• HD assisted by functional annotation and expression

robust results?

- SNP effects from small data set
- more data
 - genotypes & phenotypes
 - sequence, expression

