

Identification of Genetic Variation Regulating Gene Expression in Dairy Cattle with RNA Sequence Data

Amanda Chamberlain

Introduction

- Variants in cis-regulatory regions can cause variation in gene expression between individuals and between tissues
- Many mutations affecting complex traits reside in regulatory regions of genome
- Such mutations have been shown to increase the accuracy of genomic predictions
- This study uses an RNA sequence dataset to map:
 - Expression QTL (eQTL) – change total gene expression
 - Allele specific expression QTL (aseQTL) – change the allelic imbalance of expression
 - Splice QTL (sQTL) – change isoform expression

3 Analyses

112 Holstein MC and WBC
29 Jersey MC
(RIN>6)

RNA sequenced
(SureSelect Stranded + HiSeq3000)

Paired reads aligned (TopHat2, STAR)

Gene counts
(HTSeq-count,
DESeq)

Allele counts
(samtools)

Exon/Intron
counts
(FeatureCount,
Leafcutter)

109 Holstein +
29 Jersey BovineSNP50

1000 bull genomes
sequence genotypes
(run 5)

Imputed whole genome sequence data
(Fimpute) Phased with Eagle

Association testing with SNP within 1MB

eQTL analysis
(EMMAX)

Gene expression

aseQTL analysis
(EMMAX)

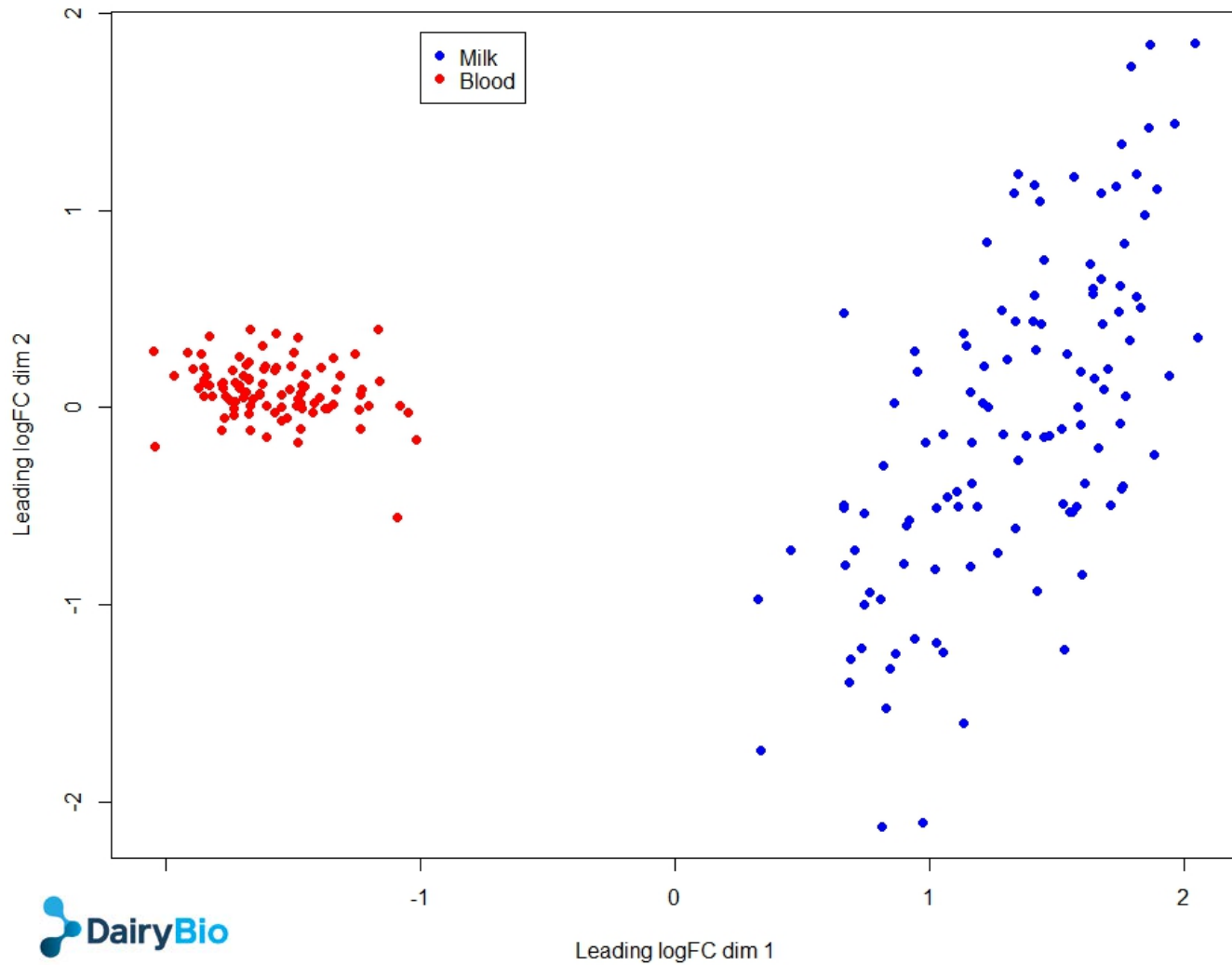
Allelic imbalance

sQTL analysis
(Matrix eQTL and
leafcutter)

Isoform expression

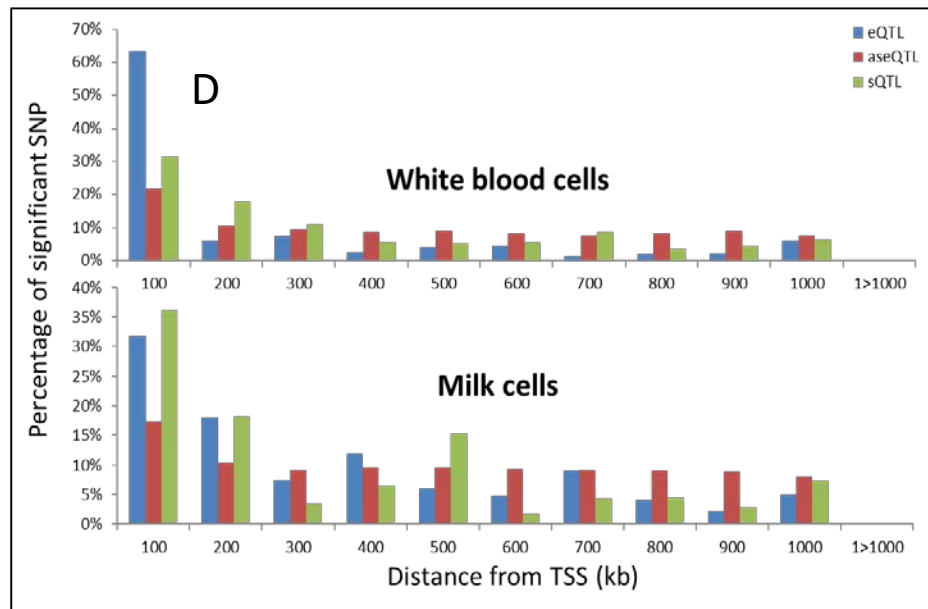
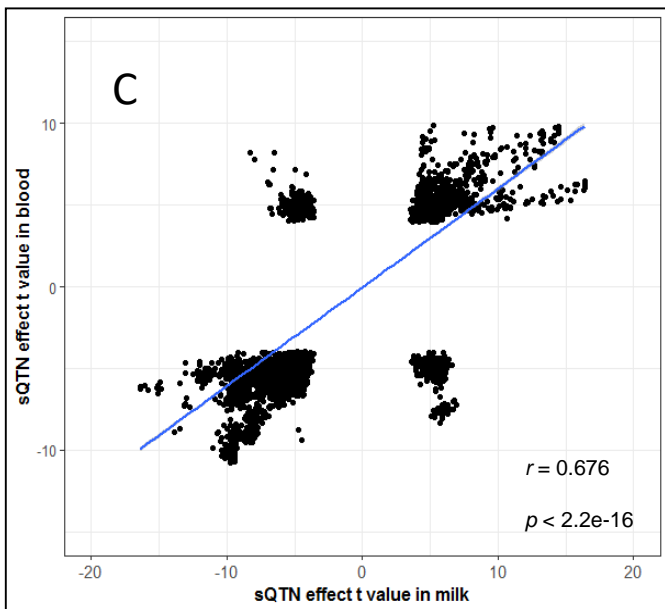
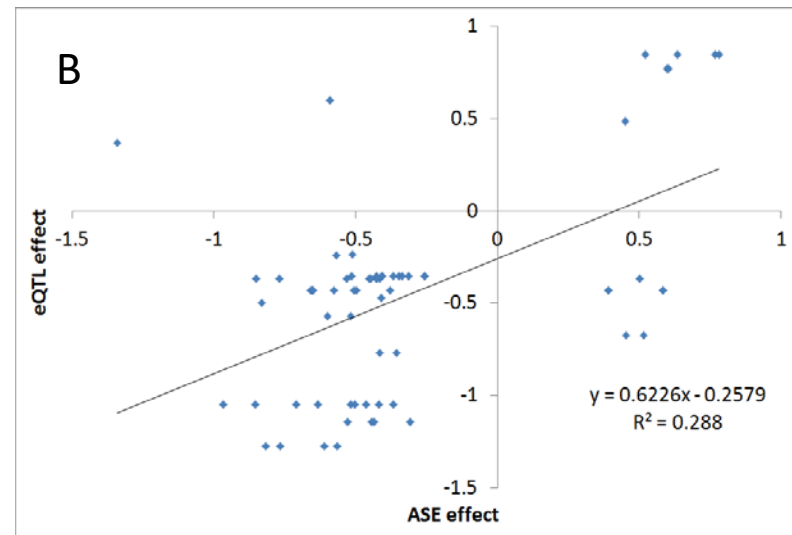
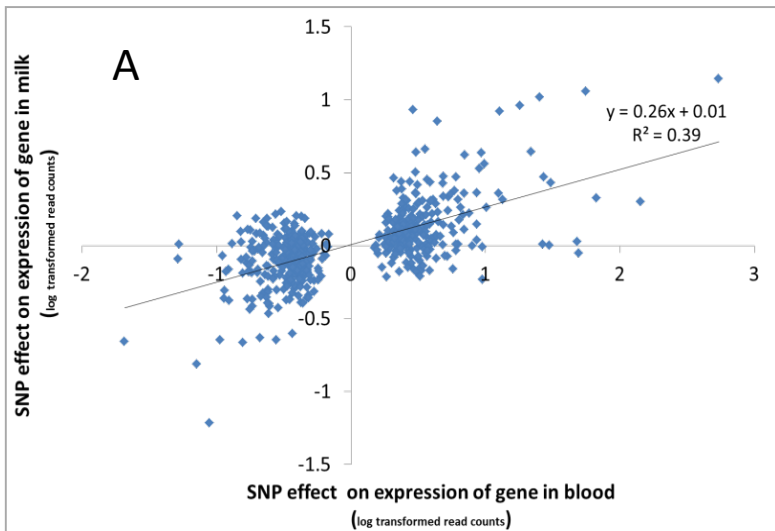
Gene expression

edgeR MDS Plot

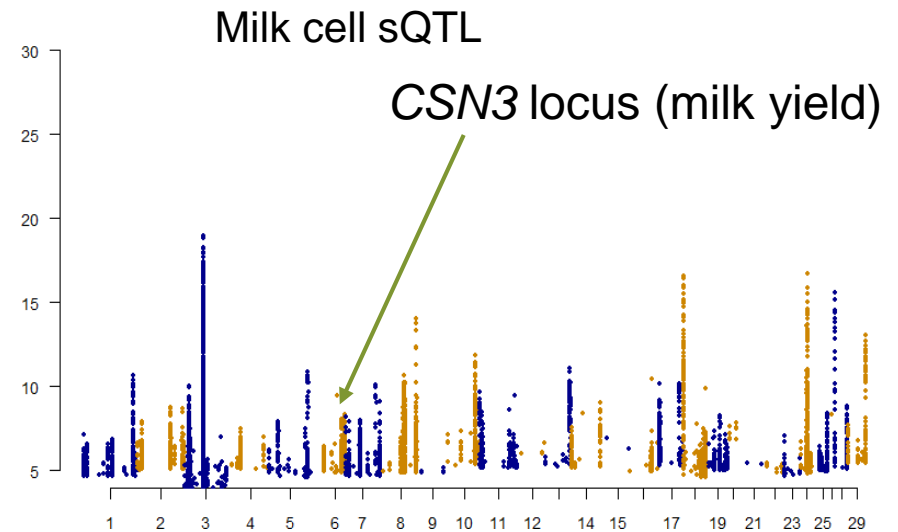
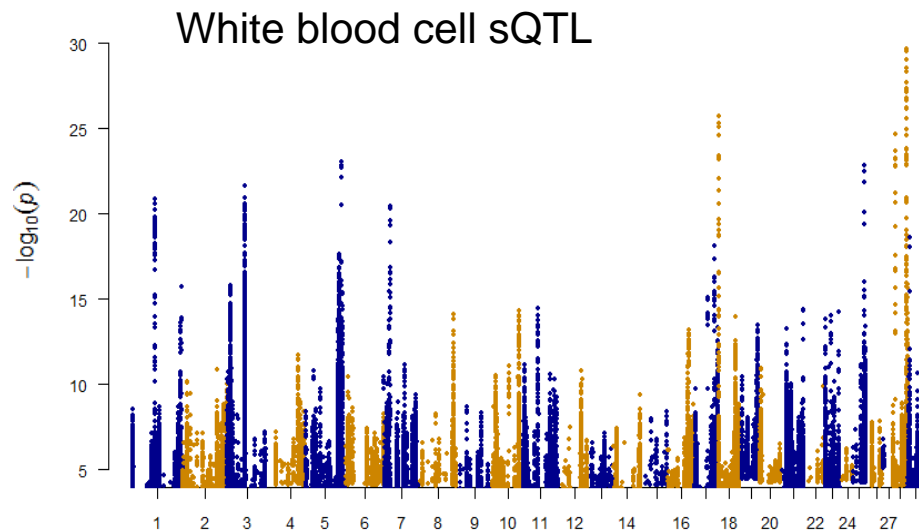


Results

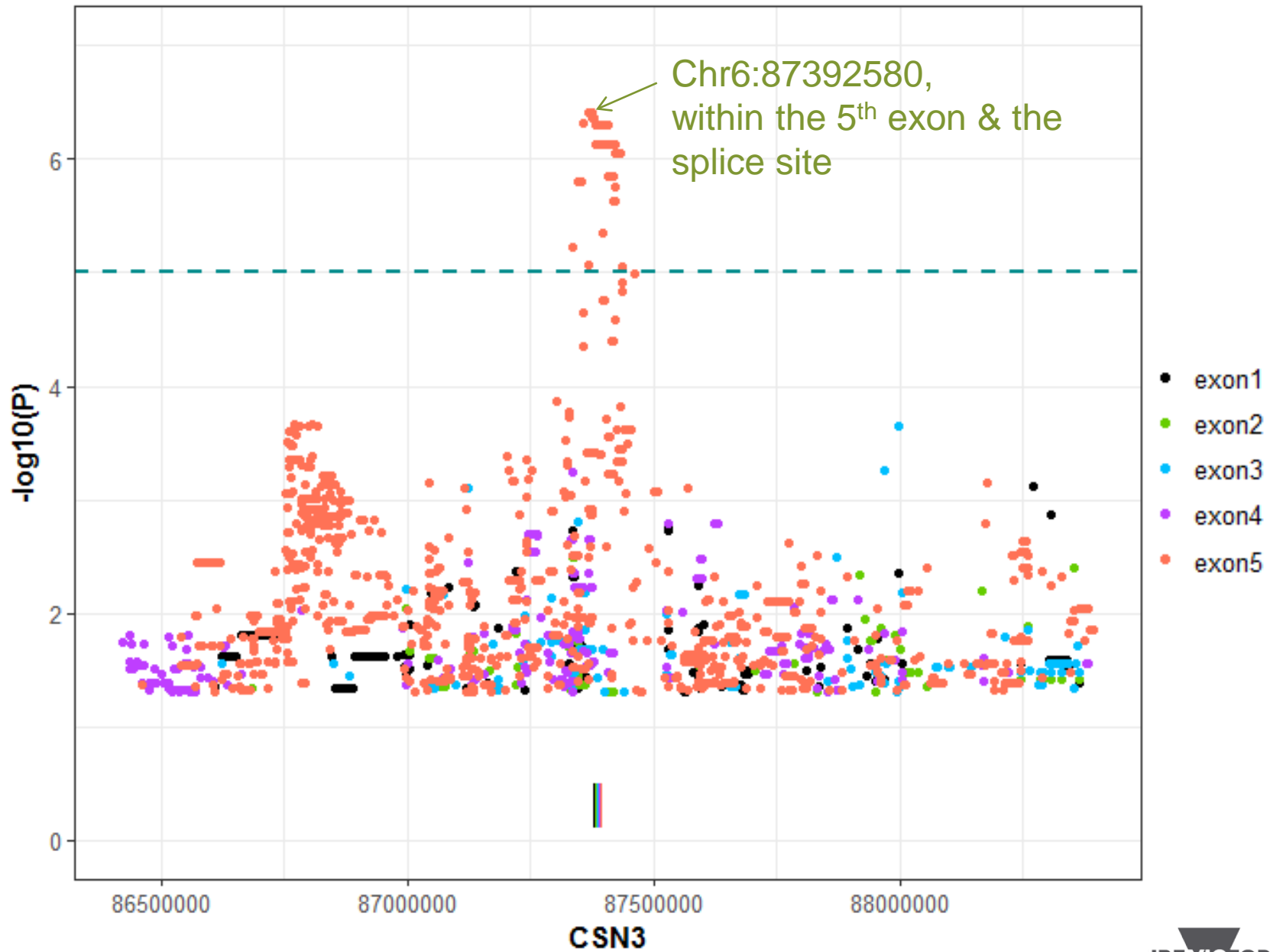
	eQTL		aseQTL		sQTL	
	MC	WBC	MC	WBC	MC	WBC
Variants tested	10.9M	10.4M	19.3M	17.8M	14.3M	14.3M
Genes/Positions/Exons	12,772	11,577	311,815	291,638	109,571	108,486
Sig variants ($P < 1 \times 10^{-4}$)	15,299	98,340	1.6M	2.9M	28,907	138,907
FDR	6%	0.8%	2.3%	1.7%	1%	1%
Genes	361	554	6,314	5,085	283	929
Gene overlap	60		3,701		107	

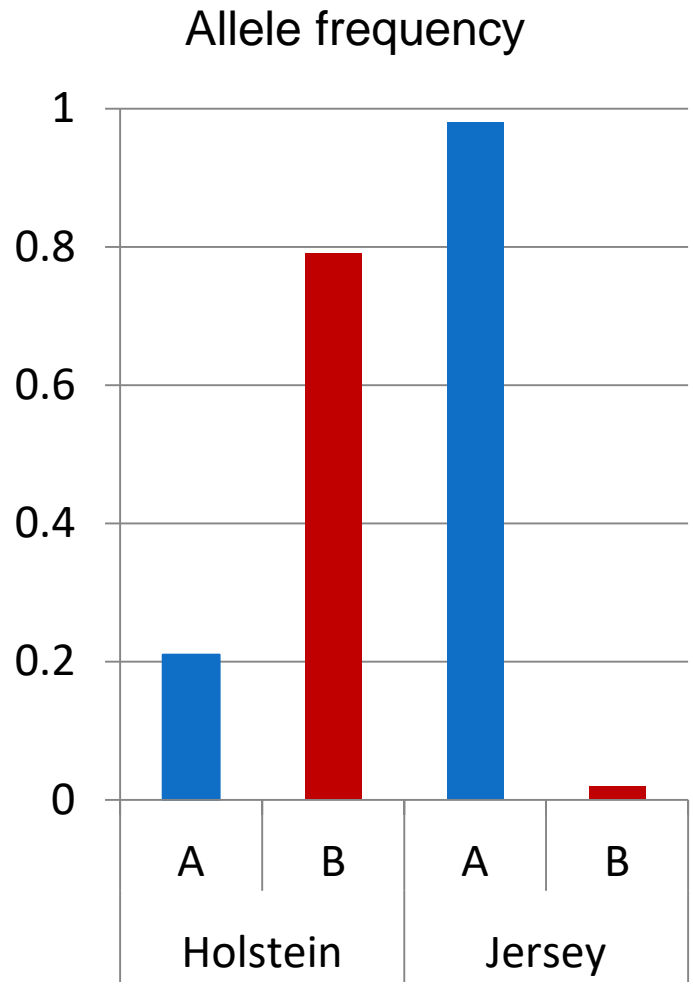
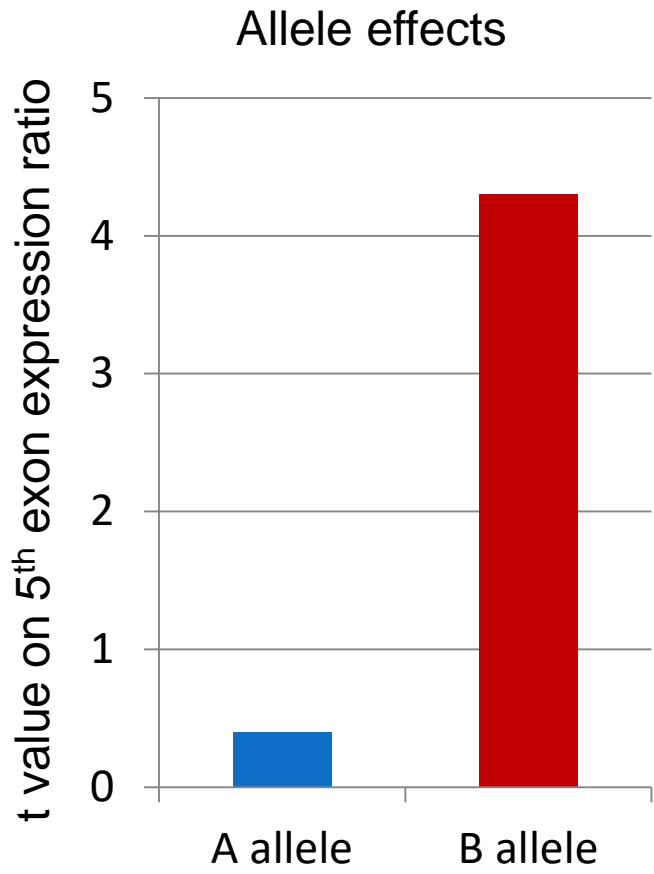


sQTL in milk and blood

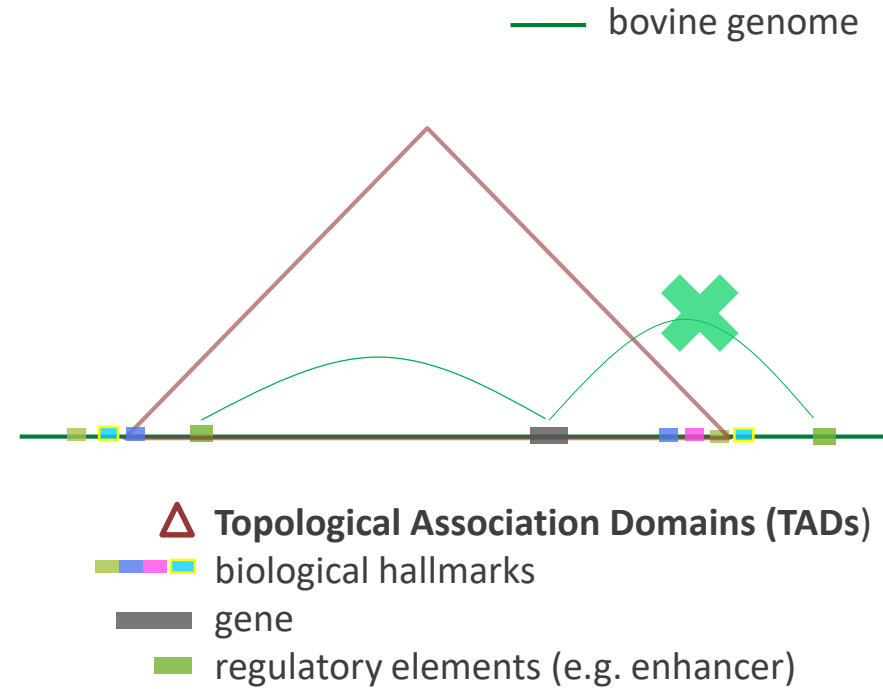
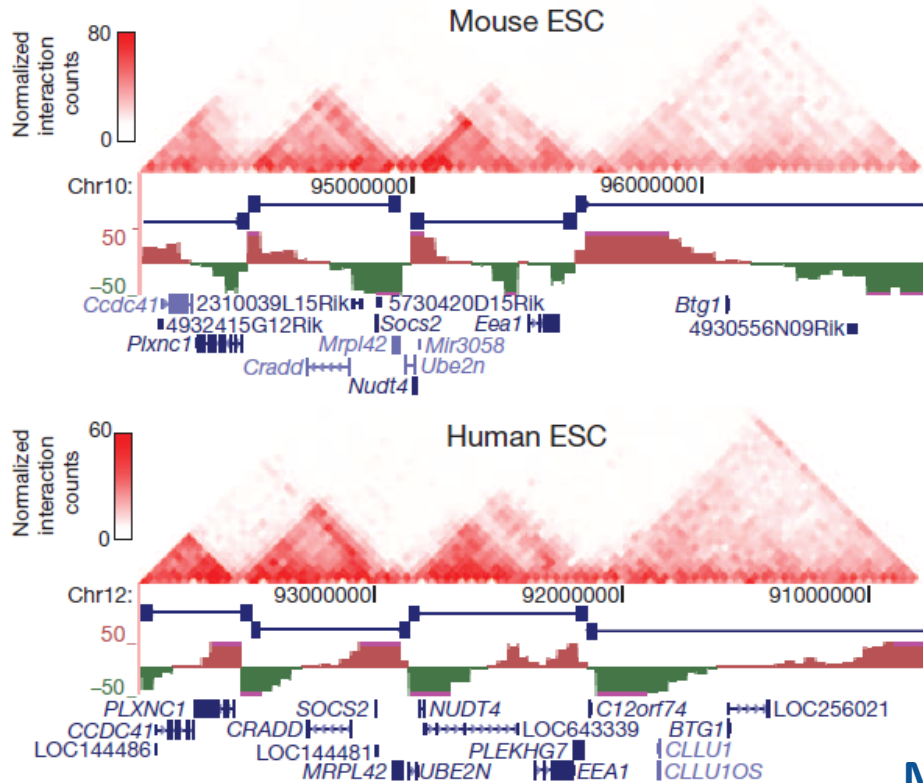


<https://www.biorxiv.org/content/early/2017/11/16/220251>





Bovine topologically associated domains



Map TADs from human, mouse and dog to cattle

Dixon *et al.* (2012)

<https://www.biorxiv.org/content/early/2018/01/04/242792>

Input TAD dataset from human ESC, Significance threshold $p \leq 10^{-8}$

QTL type	Cell type	# significant SNP	# within target gene TAD	# expected	FDR
aseQTL	WBC	13302	6375	0.0013302	0.0000%
	MC	9553	3862	0.0009553	0.0000%
eQTL	WBC	327	114	0.0000327	0.0000%
	MC	343	24	0.0000343	0.0001%

Same result for all 6 input TAD datasets and significance levels from $p \leq 10^{-5}$ to $p \leq 10^{-8}$

Conclusions

- There are many eQTL, aseQTL and sQTL that affect total gene expression, allelic imbalance or isoform expression.
- Many occur in both cell types, with similar effect.
- Most occur within 100 kb of TSS, but some are up to 1Mb away
- Most occur within the same bovine TAD as their gene targets.
- TADs provide a better search space than an arbitrary distance from the gene
- Many of the potential regulatory variants are now being genotyped in large populations to determine their affect on increasing the accuracy of genomic prediction

Acknowledgements

Sample collection, genotyping and library making

Dr Coralie Reich, Brett Mason, Claire Prowse Wilkins, Josie Garner, Meaghan Douglas, Dr Amanda Chamberlain

Sequencing

Dr Tracie Webster

Sequence imputation

Prof Ben Hayes, Dr Iona Macleod

aseQTL, eQTL and sQTL analysis

Prof Ben Hayes, Dr Ruidong Xiang, Dr Christy Vander Jagt, Dr Amanda Chamberlain, Prof Mike Goddard

TAD analysis

Min Wang, Dr Tim Hancock, Prof Ben Hayes, Dr Amanda Chamberlain, Prof Mike Goddard

Funding

Dairy Futures CRC, Dairy Bio