

Using Functional information

Michael Goddard
University of Melbourne

Introduction

We would like to find markers close to the causal variants for economic traits

Why?

Make genomic selection more robust

Understand what we are doing

How to find better markers

Genome sequence

Large datasets

Bayesian models

Multiple trait analysis

Functional data

Functional data

Data that describes the function of a site in the genome

Two kinds

Traits with genetic variation that are close to the primary action of the gene

Gene expression → eQTL

Sites that are annotated and therefore function is “known” without any genetic variation

Coding vs non-coding sites
ChIPSeq

Introduction

This talk

Functional data

gene expression

Chipseq

Gene expression

Amanda Chamberlain, Ben Hayes, Ruidong Xiang, Irene van den Berg, Christy van der Jagt

Many QTL are thought to affect a trait (eg milk yield) by altering the expression of a gene (eQTL)

We have looked at

- Gene eQTL

- Exon eQTL

- Splice eQTL

- ASE QTL

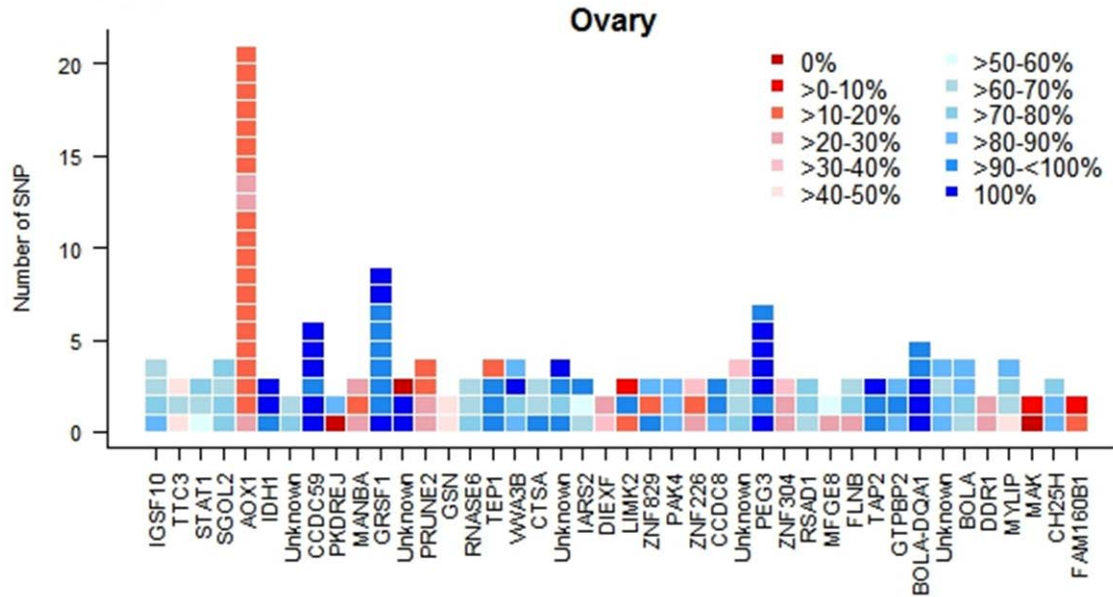
Look for a SNP close to the gene that affects its expression (cis eQTL)

In milk cells and white blood cells in 120 cows

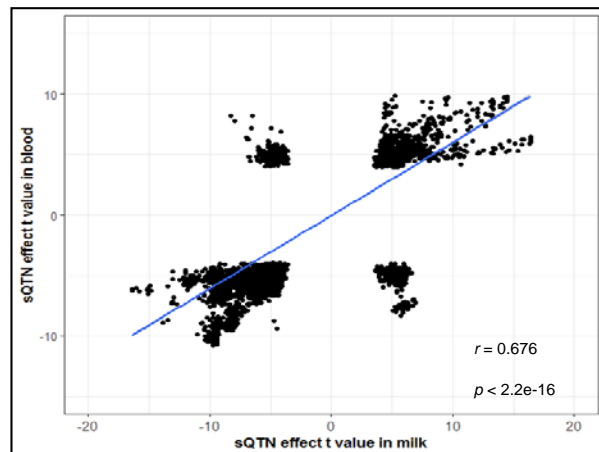
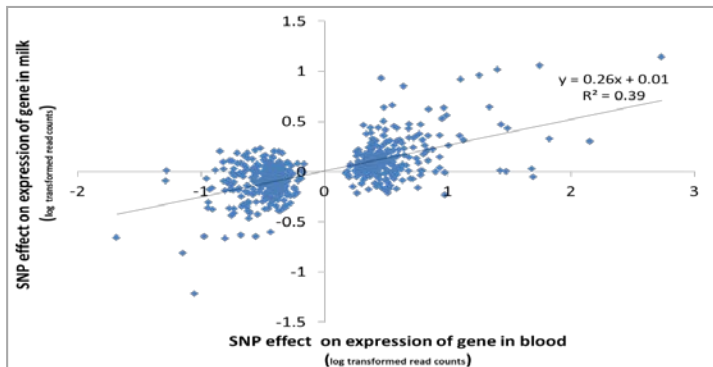
Gene Expression

	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	

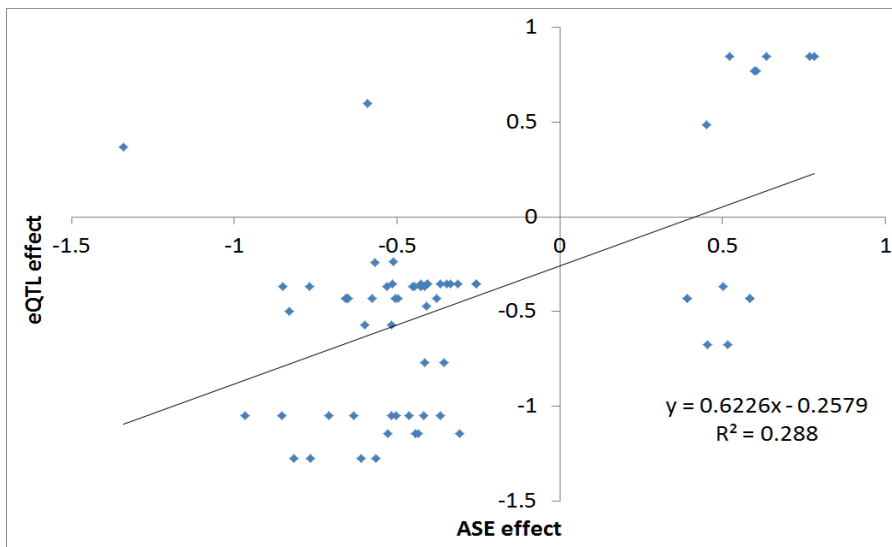
Divergent allele specific expression



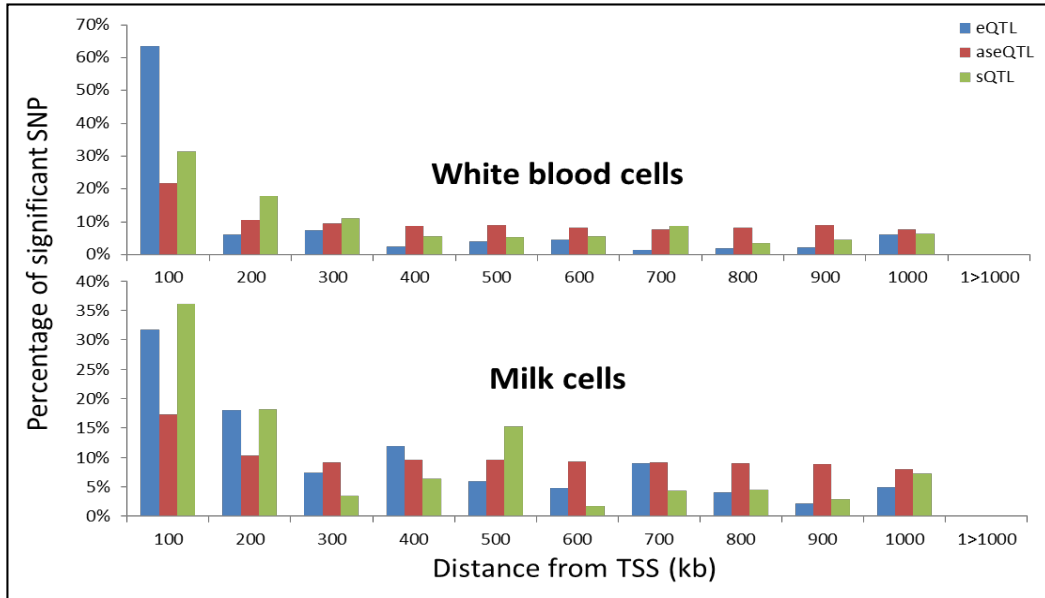
SNP effects in both cell types



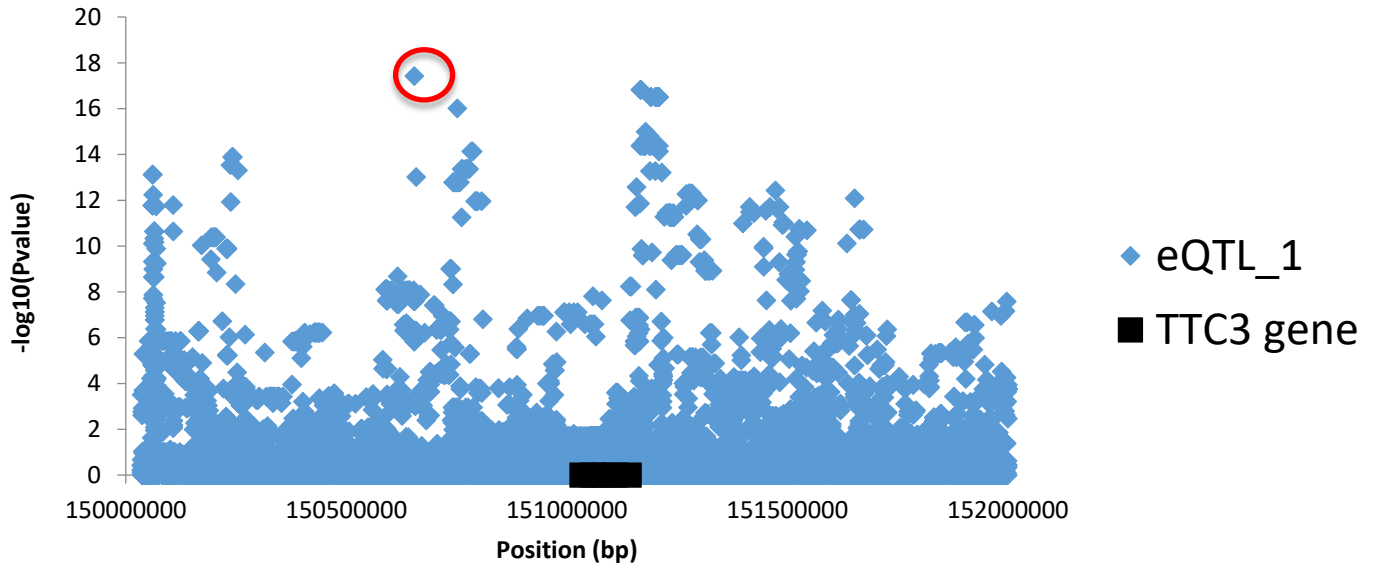
eQTL vs aseQTL effects



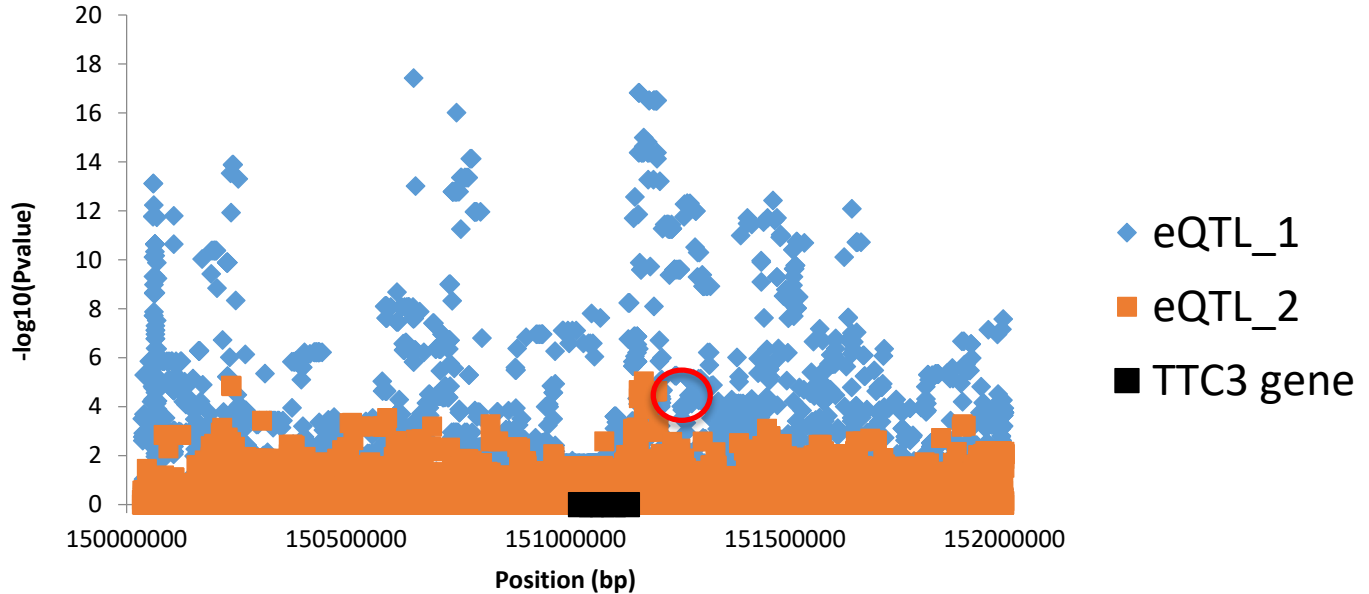
Proximity to gene



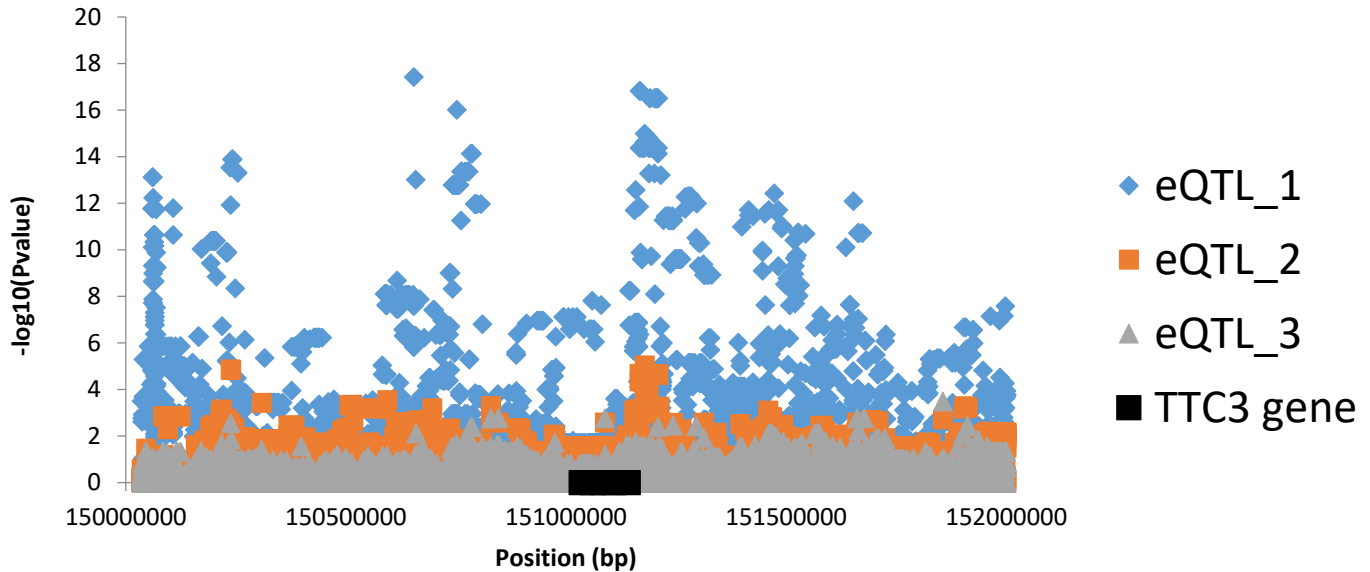
How many eQTL affect expression of a gene?



How many eQTL affect expression of a gene?



How many eQTL affect expression of a gene?



How many eQTL affect expression of a gene?

Gene	Number of eQTL
<i>TTC3</i>	2
<i>RAB6B</i>	1
<i>ITGB2</i>	1
<i>EHHADH</i>	3
<i>PFN2</i>	2

Average across genes ~ 1.9 eQTL

Some QTL are eQTL

Cattle stature (Aniek Bouwman, Ben Hayes et al)

- Meta-analysis of stature
 - 1000 Bull Genomes
- 17 populations, 8 breeds and 58,265 cattle
- 163 lead variants
- Mostly non coding
- 10 were also eQTL in WBC

Annotation class	Number
intergenic_variant	83
upstream_gene_variant	11
5_prime_UTR_variant	1
intron_variant	55
missense_variant	5
downstream_gene_variant	8
ChiP-SEQ peaks*	8
WBC eQTL	10

	Effect	P-value	Prop. σ^2_p
<i>Additional traits</i>			
phosphorus conc.	41.8	1.10x10⁻¹¹	0.107
eSLC37A1	0.160	3.55x10⁻¹⁸	0.224
<i>Key production trait, milk yield</i>			
milk yield – Holstein cows	-37.6	2.19x10⁻³	0.001
milk yield – Holstein bulls	-40.3	3.17x10⁻³	0.003
milk yield – Jersey cows	-45.2	3.26x10⁻³	0.002

That is the allele that *increases* expression of SLC27A1 (an antiporter):

1. *Increases* phosphorus concentration
2. *Decreases* milk yield

Kath Kemper et al

eQTL and QTL

There are many eQTL (100,000?)

One near every QTL?

Are they the same?

Instead of GWAS

correlation (local GEBV, expression of a gene)

eQTL and QTL

Irene van den Berg

correlation (local GEBV for milk, expression of a gene)
for each 250 kb segment

420 segments have correlation with $p < 10^{-5}$
but they explain little variance in milk

Permute the regression coefficients in local GEBV
348 segments have correlation with $p < 10^{-5}$

eQTL and QTL

Start with 300 segments with most variance for milk

54 contain an eQTL

300 random segments

36 contain an eQTL

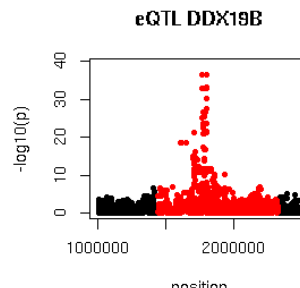
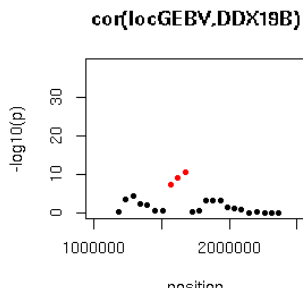
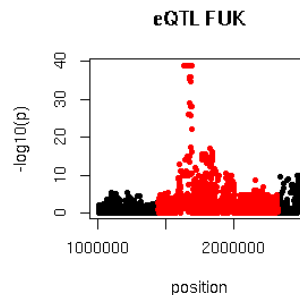
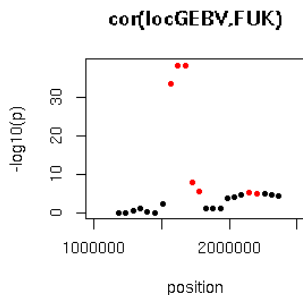
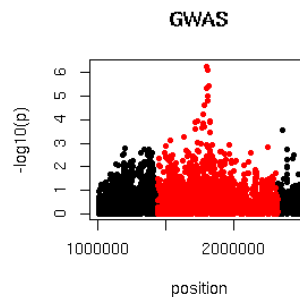
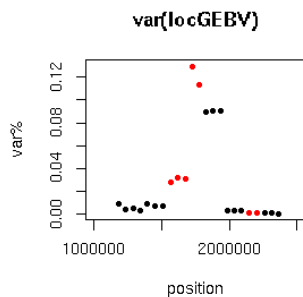
eQTL and QTL

Do eQTL and QTL colocalize?

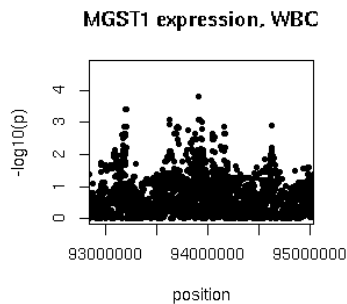
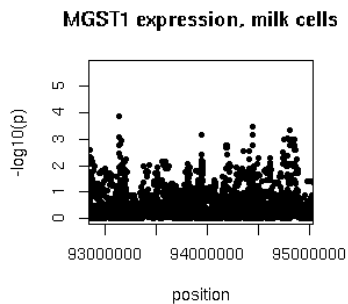
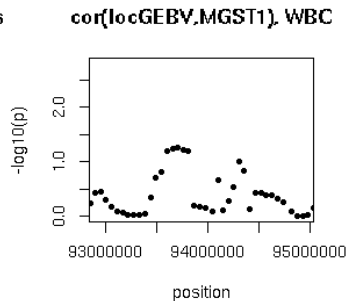
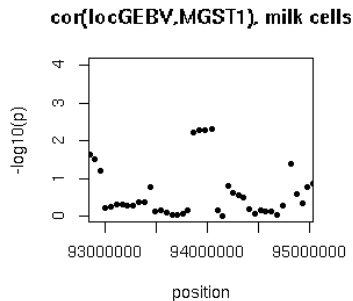
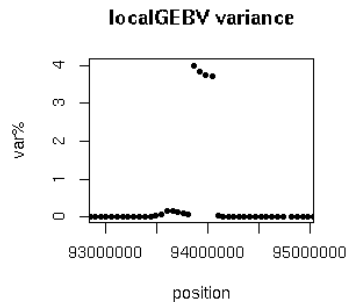
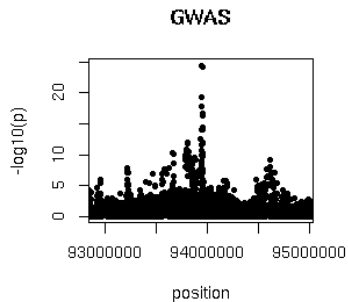
eCAVIAR

mainly low probability because of LD

eQTL and QTL



eQTL and QTL



eQTL and QTL

Often QTL are cis eQTL

Hard to prove individual eQTL and QTL are the same SNP

So many eQTL and QTL

Small effects

LD between them

Chip seq for histone modifications

Amanda Chamberlain, Claire Prowse-Wilkins

These histone modifications are associated with promoters and enhancers

Therefore likely to indicate position of eQTL

Don't need genetic variation in histone marks to be useful

H3K4me1 & 3, H3K27ac & me3 and CTCF

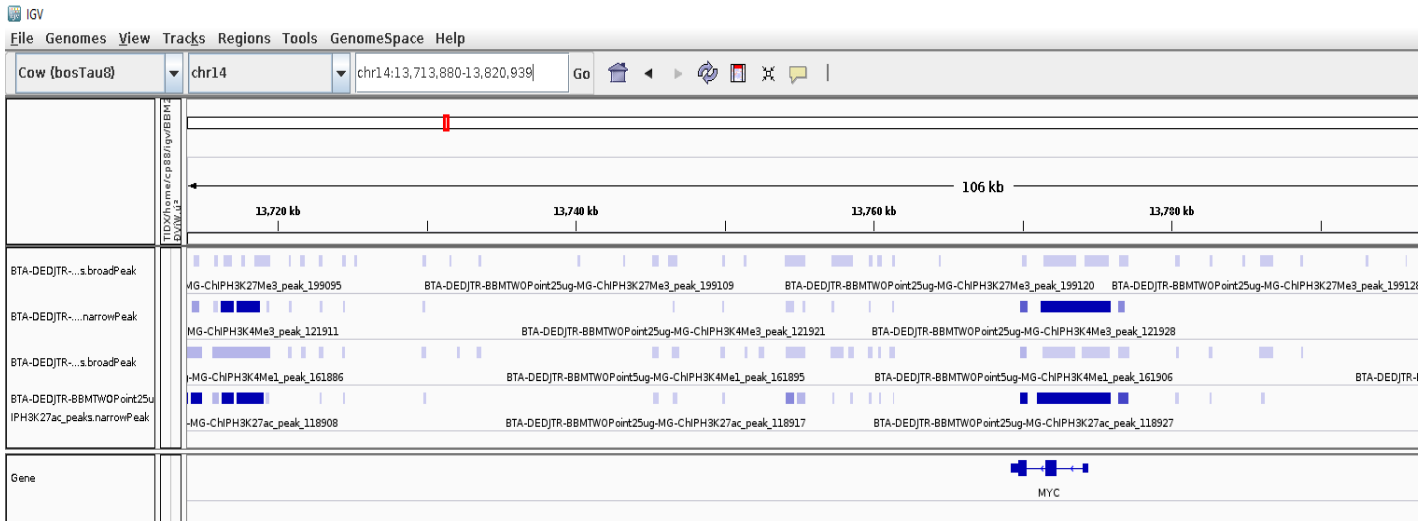
Mammary tissue, liver, spleen, heart, lung and kidney

1-3 cows

Preliminary data

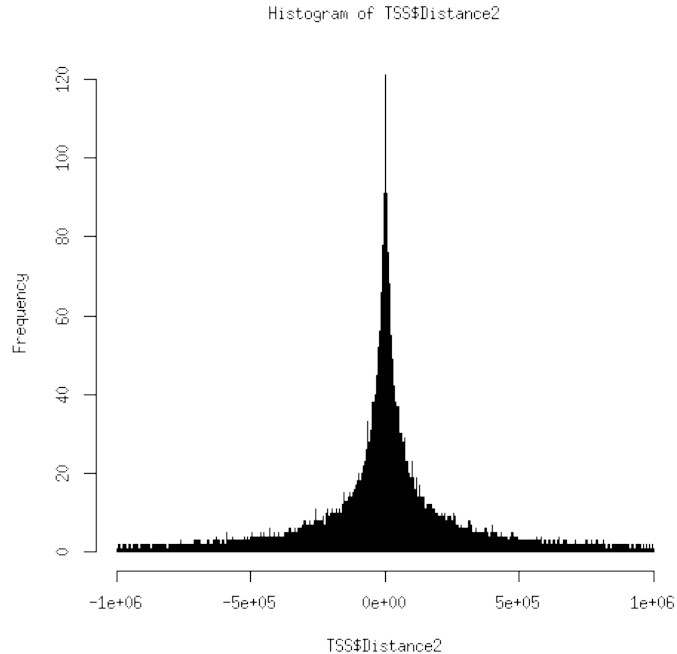
Chip seq

Histone marks H3k27ac, H3K27me3, H3K4me1, H3K4me3
mammary tissue



Chip seq

Position of peaks relative to TSS



Functional Annotation in Mammary Gland

Claire Prowse-Wilkins
Poster P26



Mammary and Liver

H3K4Me1/H3K4Me3
>Enhancers/Promoters

H3K27ac
> active

H3K27Me3
> silent



Mark	Animal	Tissue	Mapped Reads	Peaks	Average Size of Peaks
H3K4Me3	1	Mammary	214,126,761	842,020	360bp
	2		235,775,602	592,705	
	3		230,681,065	740,784	
	3	Liver	202,635,963	748,382	
	1		404,070,766	913,793	
	2		136,932,687	767,988	
H3K27Me3	3		313,557,543	1,052,701	
	2	Mammary	219,749,095	940,743	559bp
	1	Liver	243,279,828	1,003,463	
H3K4Me1	2	Mammary	161,157,266	752,310	560bp
	2	Liver	291,347,398	1,002,560	
H3K27ac	2	Mammary	230,646,070	564,609	394bp
	2	Liver	328,588,342	1,047,559	

Chip seq for histone modification

Typically 1m peaks of about 400 bp (10% of the genome)

Peaks overlap with eQTL

eQTL and Chip seq ideal combination?

eQTL data is single bp but confounded with LD

Chip seq data is multiple bp (ie a peak) but not confounded with LD

What sources of information about SNPs indicate potential QTLs?

Ruidong Xiang

Classify SNPs into categories

e.g. coding vs non-coding

Form a GRM for each category

Estimate variance of milk yield etc. associated with each category

Variance per SNP associated with categories of SNPs

Classification	h^2 per SNP ($\times 10^7$)
eQTL	110
SNP age	47
Slice site	45
UTR	25
Chipseq	22
Sel. Sig.	19
Coding	14
Conserved	13
Intergenic	12
CTCF	8
Intron	8

Conclusions eQTL

Many eQTL

2 per gene

eQTL are enriched for QTL

QTL for milk etc are often eQTL

Hard to prove eQTL and QTL are the same

Conclusions Histone modification (Chip seq)

1M peaks of around 400 bp each

Peaks for different modifications overlap

Peaks are enriched for eQTL?

Should help distinguish causal variant from variants in LD with cv

Acknowledgements

AgriBio


Amanda Chamberlain, Irene van den Berg, Ruidong Xiang, Iona MacLeod, Kath Kemper, Ben Hayes, Aniek Bouwman

Cattle GTEx project

- Many breeds – to increase precision
 - 100 lactating cows, 5 breeds (Jersey, Brahman, Wagyu, Angus, Reds), blood
- Lots of animals for few tissues – to increase power
 - Increasing our WBC dataset to >450 lactating cows
 - Increasing our Milk cell dataset to >220 lactating cows
- Lots of tissues from few animals
 - 2 lactating cows + fetus, 41 tissues
 - FAANG
- Deep sequencing (>40 million reads) – for aseQTL and sQTL
- Full sequence genotypes (imputed or WGS)

JOIN US

amanda.chamberlain@ecodev.vic.gov.au



Section Breaker