Association Mapping Across Sheep Breeds

- International Sheep Genomics Consortium
- Sheep HapMap project

ISGC
International Sheep Genomics Consortium

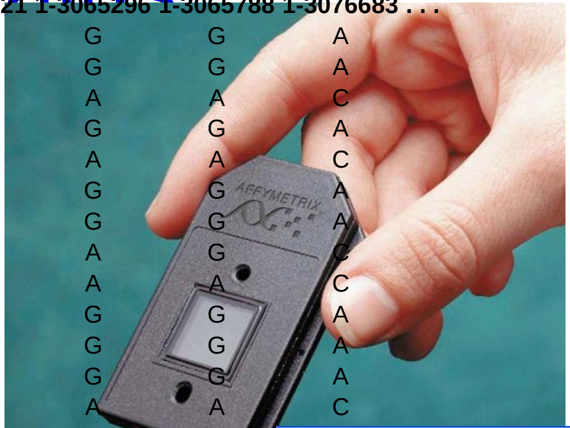
www.sheephapmap.org

- Herman Raadsma
- Peter Thomson
- Mehar Khatkar
- Lisa Jonas

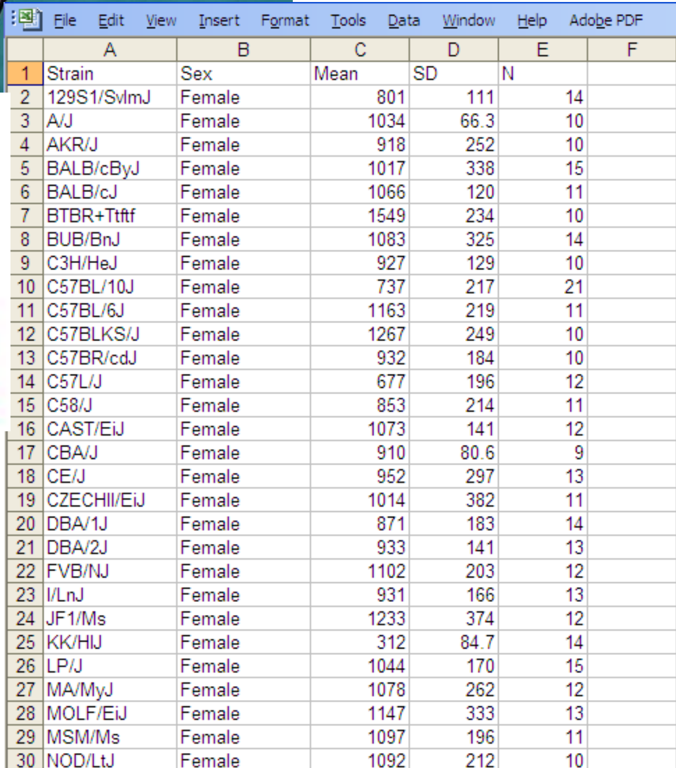
The University of Sydney

Association mapping across mouse strains

MARKER	1-3007431	1-3008615	1-3027623	1-3050112	1-3050199	1-3053121	1-3065296	1-3065788	1-3076683 . . .
129S1/SvImJ	G	G	T	C	G	N	G	G	A
129X1/SvJ	G	G	T	C	G	T	G	G	A
A/J	G	A	N	A	A	T	A	A	C
AKR/J	G	G	T	C	G	N	G	G	A
BALB/cByJ	G	N	A	A	A	T	A	A	C
BTBR+Ttftf	G	G	T	C	G	T	G	G	A
BUB/BnJ	G	N	T	C	G	N	G	G	A
C3H/HeJ	G	N	N	A	A	T	A	G	C
C57BL/6J	G	A	A	A	A	T	A	A	C
C57BLKS/J	G	G	T	C	G	T	G	G	A
C57BR/cdJ	G	N	T	C	G	T	G	G	A
C57L/J	G	G	T	C	G	T	G	G	A
C58/J	G	A	A	A	A	T	A	A	C
CAST/EiJ	A	G	T	A	A	N	G	G	
CBA/J	G	N	N	A	A	T	A	G	
CE/J	G	G	T	C	G	T	G	G	

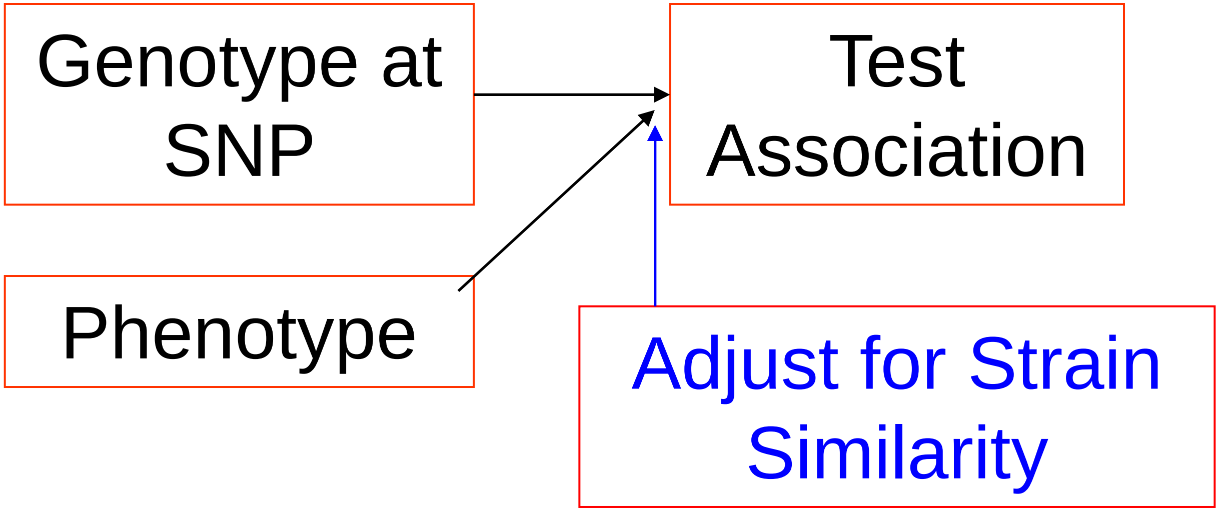


ISGC
International Sheep Genomics Consortium



	A	B	C	D	E	F
1	Strain	Sex	Mean	SD	N	
2	129S1/SvImJ	Female	801	111	14	
3	A/J	Female	1034	66.3	10	
4	AKR/J	Female	918	252	10	
5	BALB/cByJ	Female	1017	338	15	
6	BALB/cJ	Female	1066	120	11	
7	BTBR+Ttftf	Female	1549	234	10	
8	BUB/BnJ	Female	1083	325	14	
9	C3H/HeJ	Female	927	129	10	
10	C57BL/10J	Female	737	217	21	
11	C57BL/6J	Female	1163	219	11	
12	C57BLKS/J	Female	1267	249	10	
13	C57BR/cdJ	Female	932	184	10	
14	C57L/J	Female	677	196	12	
15	C58/J	Female	853	214	11	
16	CAST/EiJ	Female	1073	141	12	
17	CBA/J	Female	910	80.6	9	
18	CE/J	Female	952	297	13	
19	CZECHII/EiJ	Female	1014	382	11	
20	DBA/1J	Female	871	183	14	
21	DBA/2J	Female	933	141	13	
22	FVB/NJ	Female	1102	203	12	
23	I/LnJ	Female	931	166	13	
24	JF1/Ms	Female	1233	374	12	
25	KK/HIJ	Female	312	84.7	14	
26	LP/J	Female	1044	170	15	
27	MA/MyJ	Female	1078	262	12	
28	MOLF/EiJ	Female	1147	333	13	
29	MSM/Ms	Female	1097	196	11	
30	NOD/LtJ	Female	1092	212	10	

*Association mapping across mouse strains/sheep breeds

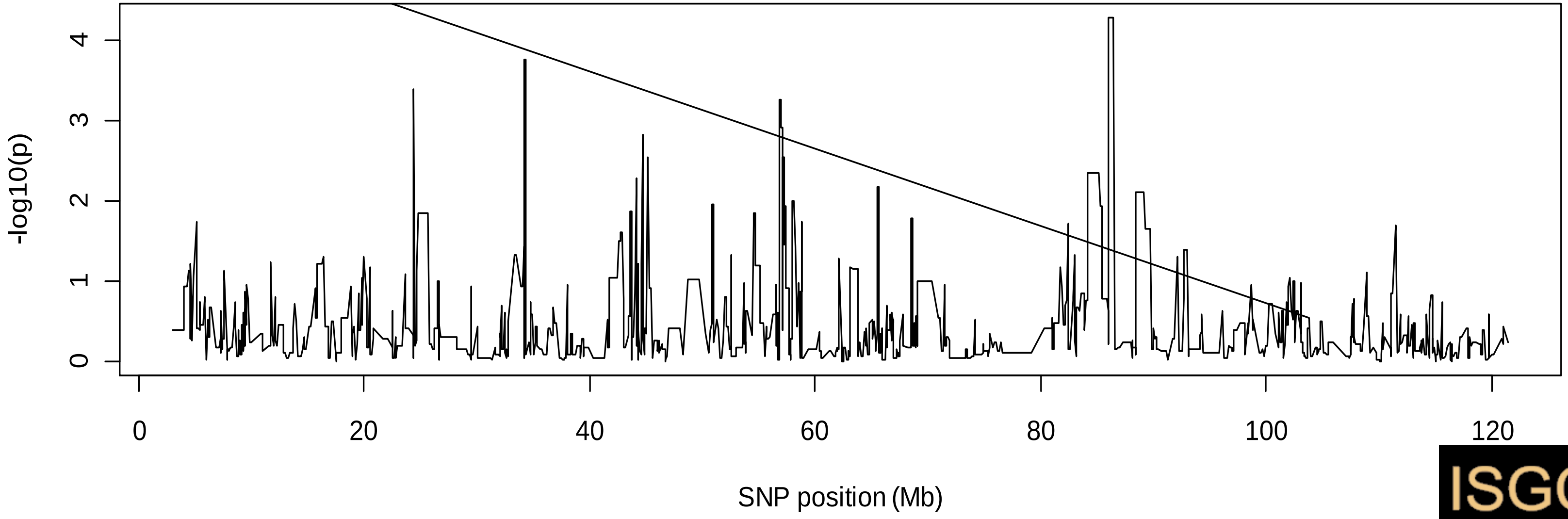


Genotype at SNP

Test Association

Phenotype

Adjust for Strain Similarity



ISGC International Sheep Genomics Consortium

Can we do something similar in sheep?

provided by Sally Anne Thompson

Provided by Chuck Dodd

provided by Gary Onan

provided by Ebrahim Nooriyan Sooror

provided by Sally Anne Thompson

from Lane d'Italia

provided by Ebrahim Nooriyan Sooror

International Sheep Genomics Consortium

The International Sheep Genomics Consortium is a partnership of scientists and funding agencies from Australia, Austria, Brazil, China, Finland, France, Germany, Greece, India, Iran, Israel, Italy, Kenya, New Zealand, Norway, Spain, Switzerland, Turkey, United Kingdom and United States to develop public genomic resources that will help researchers find genes associated with production, quality and disease traits in sheep.

The project commenced informally in 2002 with the creation of a high quality ovine BAC library, and was built on an existing collaboration for the International Mapping Flock that was created nearly a decade earlier.

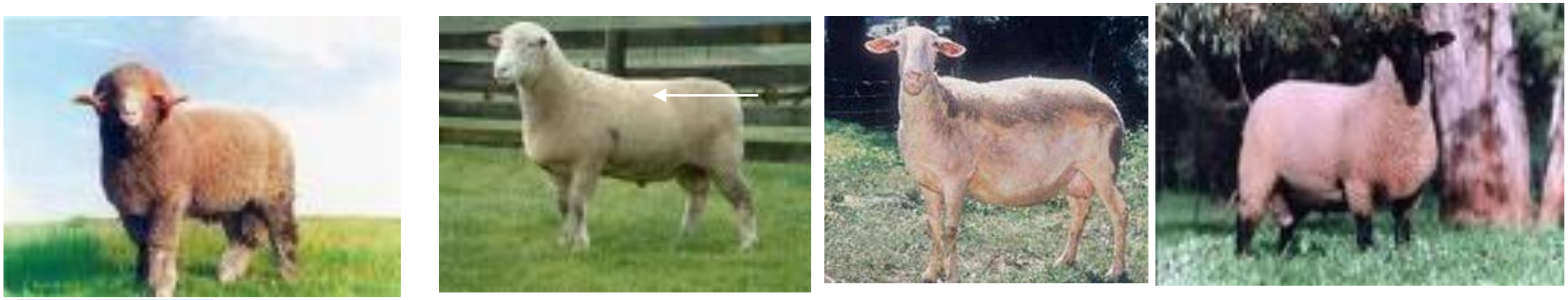
What's New

- Supplementary data released for the sheep HapMap project
- ISGC launches Newsletter to provide updates about our activities
- SNP Discovery in Deer (Cervus elaphus) Using The Illumina Genome Analyser IIx
- 2010 PAG Presentations
- Microphthalmia in Texel Sheep Is Associated with a Missense Mutation in the Paired-Like Homeodomain 3 (PITX3) Gene

http://www.sheephapmap.org/

Resources

- 73 breeds – wide range
- 2819 individuals
- 50K genotype array-
- 49,034 genotypes post QC
- Many phenotype characteristics



provided by Ebrahim Nooriyan Sooror

ISGC International Sheep Genomics Consortium

*Phenotype Database

- Unlike the mouse strain phenotypes, quantitative breed means not readily available
- Phenotypes took form of presence / absence or ordinal scale:
  - Horned: Y or N
  - Coat: Hair vs wool
  - Colour: Y or N
  - Fat tail: Y or N
  - Long tail: Y or N
  - Dairy : Y/N, meat : Y/N,
  - Adaptability: harsh or wet environments

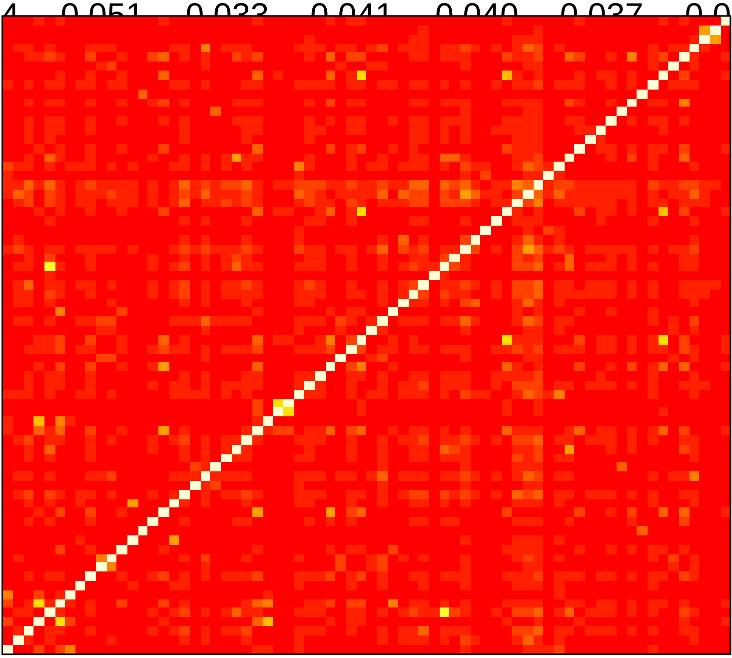
*Breed Similarity

- Necessary to adjust for breed similarity in association mapping
  - Genetically similar breeds are not independent replicates
  - Not adjusting can result in spurious associations
- GERMLINE used to determine number of shared IBD segments across genome for each pair of breeds
- Resultant 70 × 70 matrix re-scaled:
  0 = no similarity, 1 = complete similarity

*Breed Similarity based on SNP data

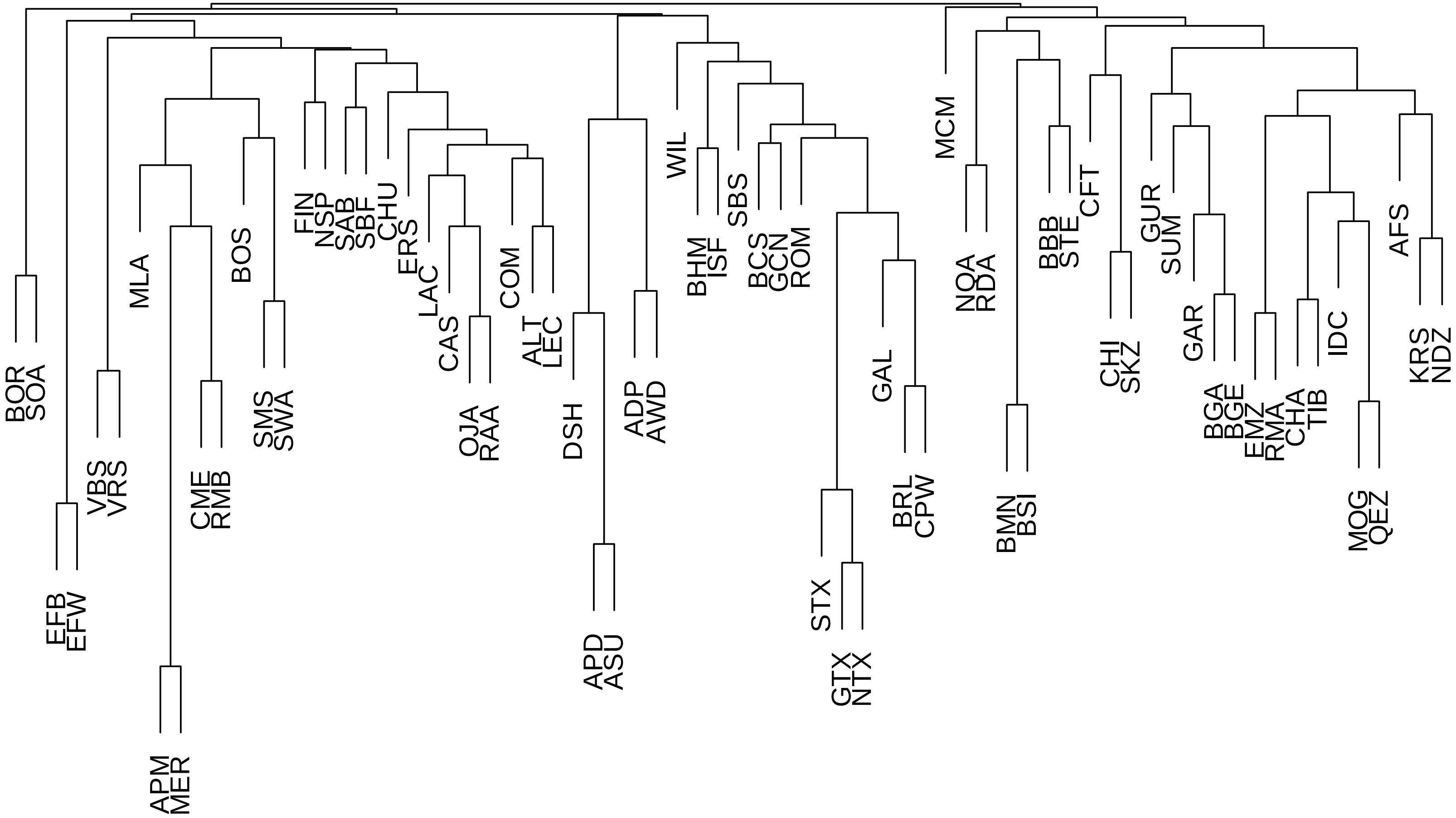
	ADP	AFS	ALT	APD	APM	ASU	AWD	BBB	BCS	BGA	BGE	BHM	BMN	BOR
ADP	1.000	0.067	0.069	0.250	0.086	0.207	0.343	0.062	0.065	0.044	0.055	0.047	0.046	0.024
AFS	0.067	1.000	0.102	0.042	0.099	0.064	0.051	0.073	0.087	0.091	0.103	0.047	0.065	0.036
ALT	0.069	0.102	1.000	0.070	0.161	0.103	0.050	0.080	0.127	0.076	0.087	0.081	0.074	0.053
APD	0.250	0.042	0.070	1.000	0.086	0.635	0.202	0.040	0.086	0.031	0.035	0.085	0.033	0.034
APM	0.086	0.099	0.161	0.086	1.000	0.132	0.058	0.083	0.161	0.070	0.080	0.091	0.078	0.051
ASU	0.207	0.064	0.103	0.635	0.132	1.000	0.172	0.062	0.131	0.047	0.052	0.182	0.054	0.049
AWD	0.343	0.051	0.050	0.202	0.058	0.172	1.000	0.044	0.051	0.033	0.041	0.040	0.037	0.017
BBB	0.062	0.073	0.080	0.040	0.083	0.062	0.044	1.00						28
BCS	0.065	0.087	0.127	0.086	0.161	0.131	0.051	0.07						48
BGA	0.044	0.091	0.076	0.031	0.070	0.047	0.033	0.05						24
BGE	0.055	0.103	0.087	0.035	0.080	0.052	0.041	0.05						28
BHM	0.047	0.047	0.081	0.085	0.091	0.182	0.040	0.04						37
BMN	0.046	0.065	0.074	0.033	0.078	0.054	0.037	0.09						23
BOR	0.024	0.036	0.053	0.034	0.051	0.049	0.017	0.02						00

. . .



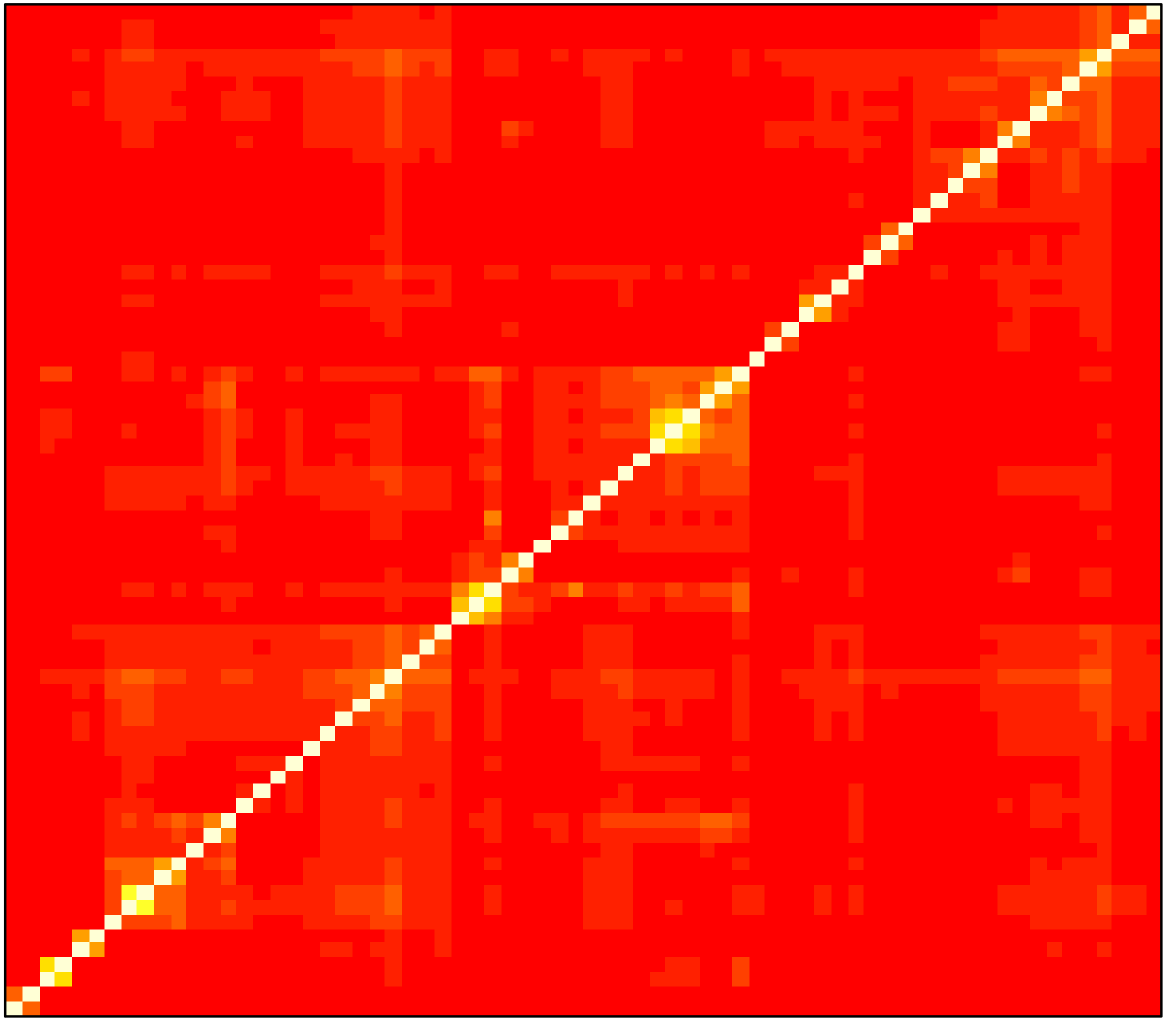
ISGC International Sheep Genomics Consortium

Cluster Dendrogram of Sheep Breeds



ISGC
International Sheep Genomics Consortium

Breed similarity
According to dendrogram…



International Sheep Genomics Consortium

*Computational Aspects of Model Fitting

- A separate model is fitted at each SNP
  ⇒ best done using R
- Continuous / LMM: `lme()` or `asreml()`
- Binary / GLMM: `asreml()`
- Efficient method for incorporating strain similarity matrix (**G**) (using Cholesky decomposition)

****Statistical Model for Association

- Binary trait:
- Generalised Linear mixed models used
- $\text{logit}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$ (logistic regression)

  where $\text{logit}(\pi) = \log_e[\pi/(1-\pi)]$

  $\pi = P(Y = 1)$

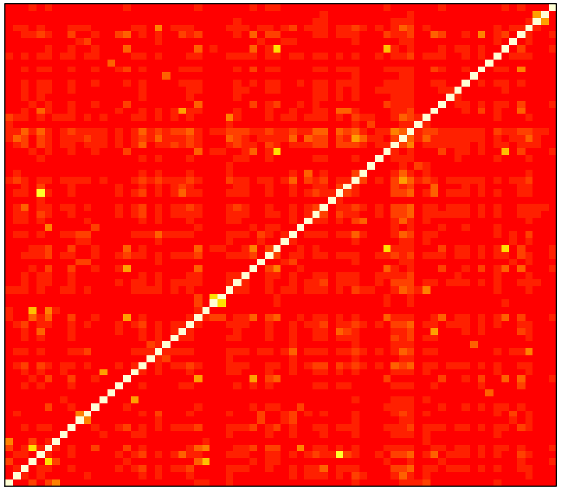
  e.g. $P(\text{Horned} = \text{“Y”}) = \pi$

  $P(\text{Horned} = \text{“N”}) = 1 - \pi$

***Statistical Model for Association

- Strain Relatedness ⇒ correlated phenotypes
- $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$
- Random effect **u** incorporates similarity

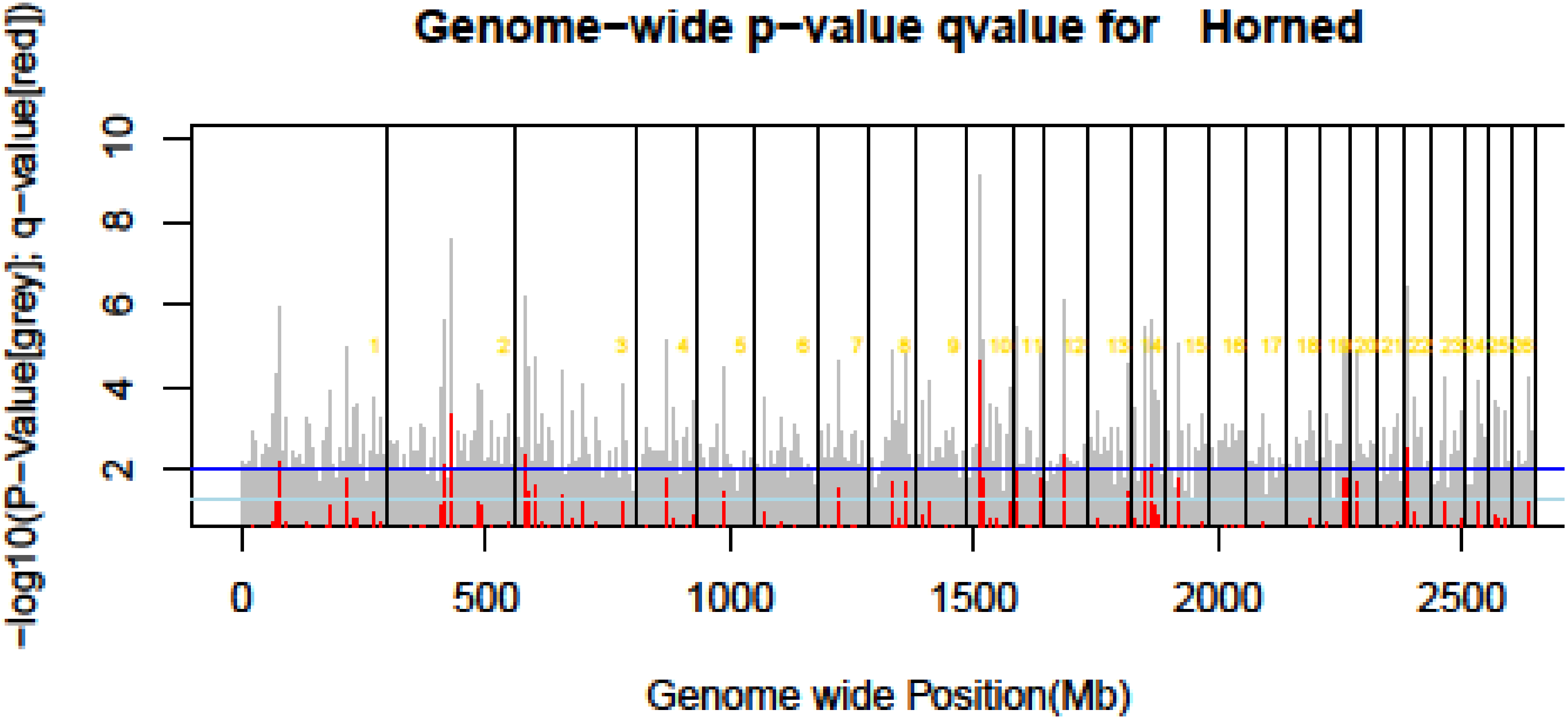
$\text{var}(\mathbf{u}) = \sigma^2\mathbf{G}$ $\mathbf{G} =$



- Many non independent tests
- Used FDR q value (Bio-conductor)

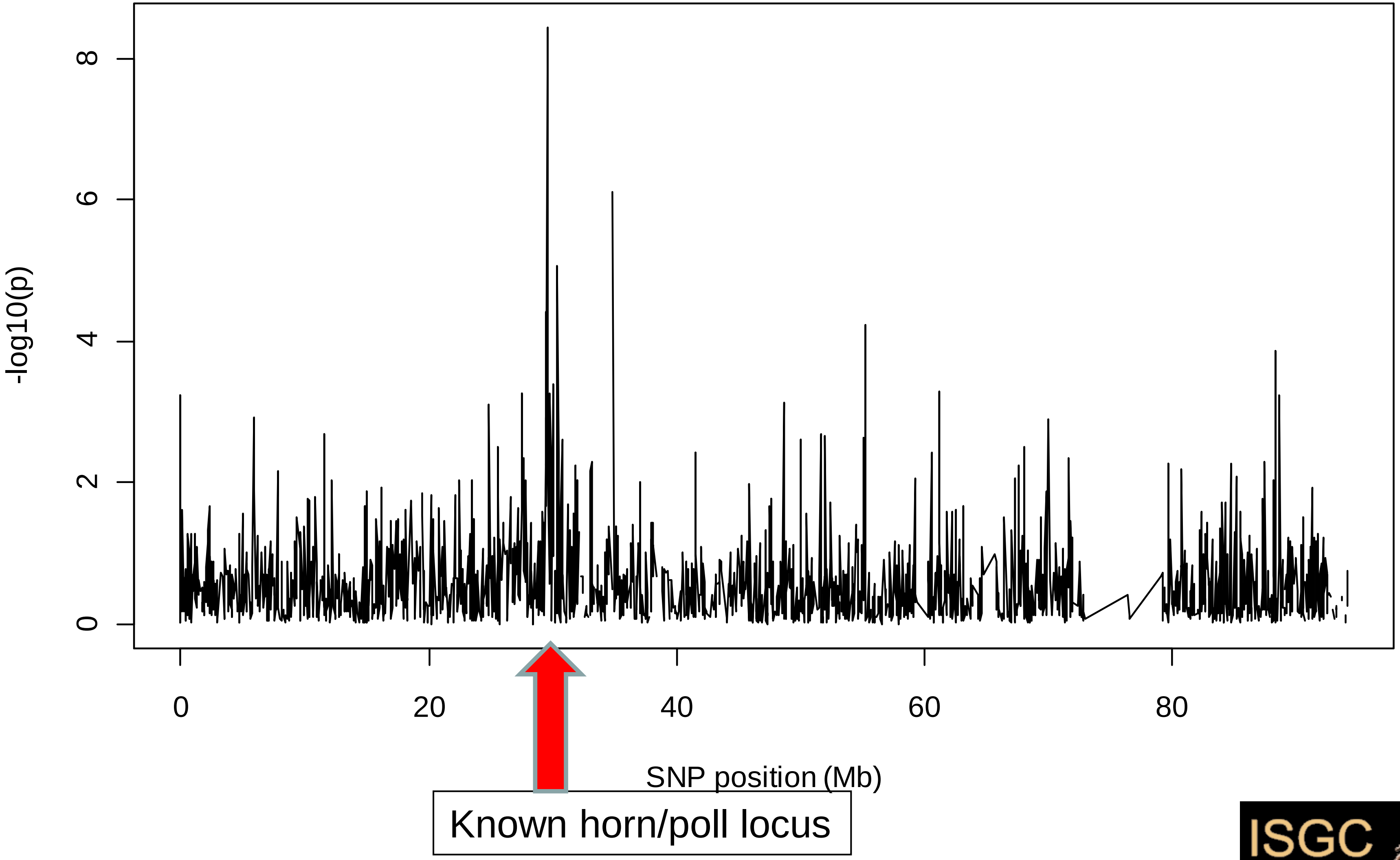
Horned vs polled

provided by Gary Onan



ISGC
International Sheep Genomics Consortium

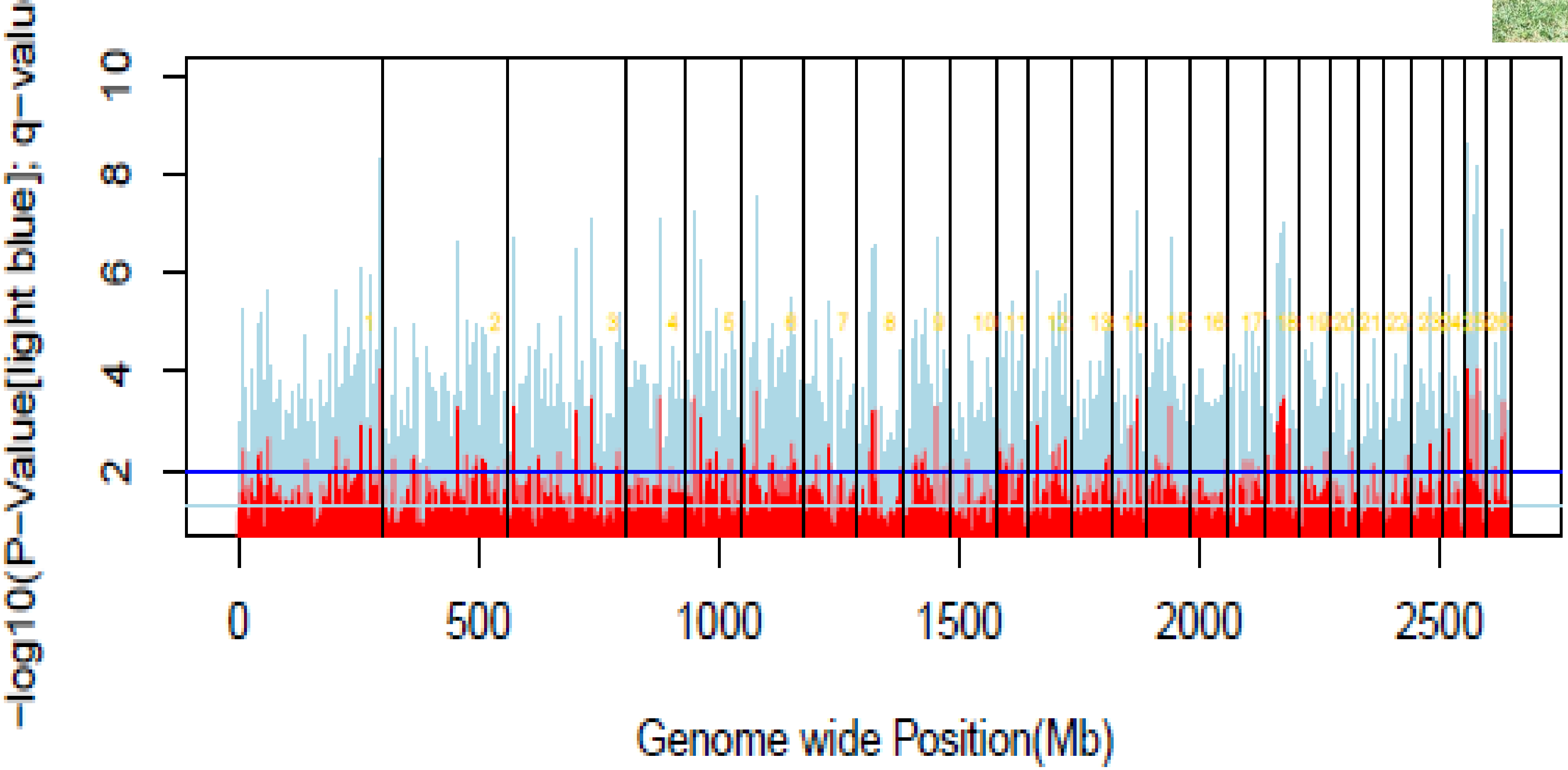
Horned (Y vs N) – OAR10



Known horn/poll locus

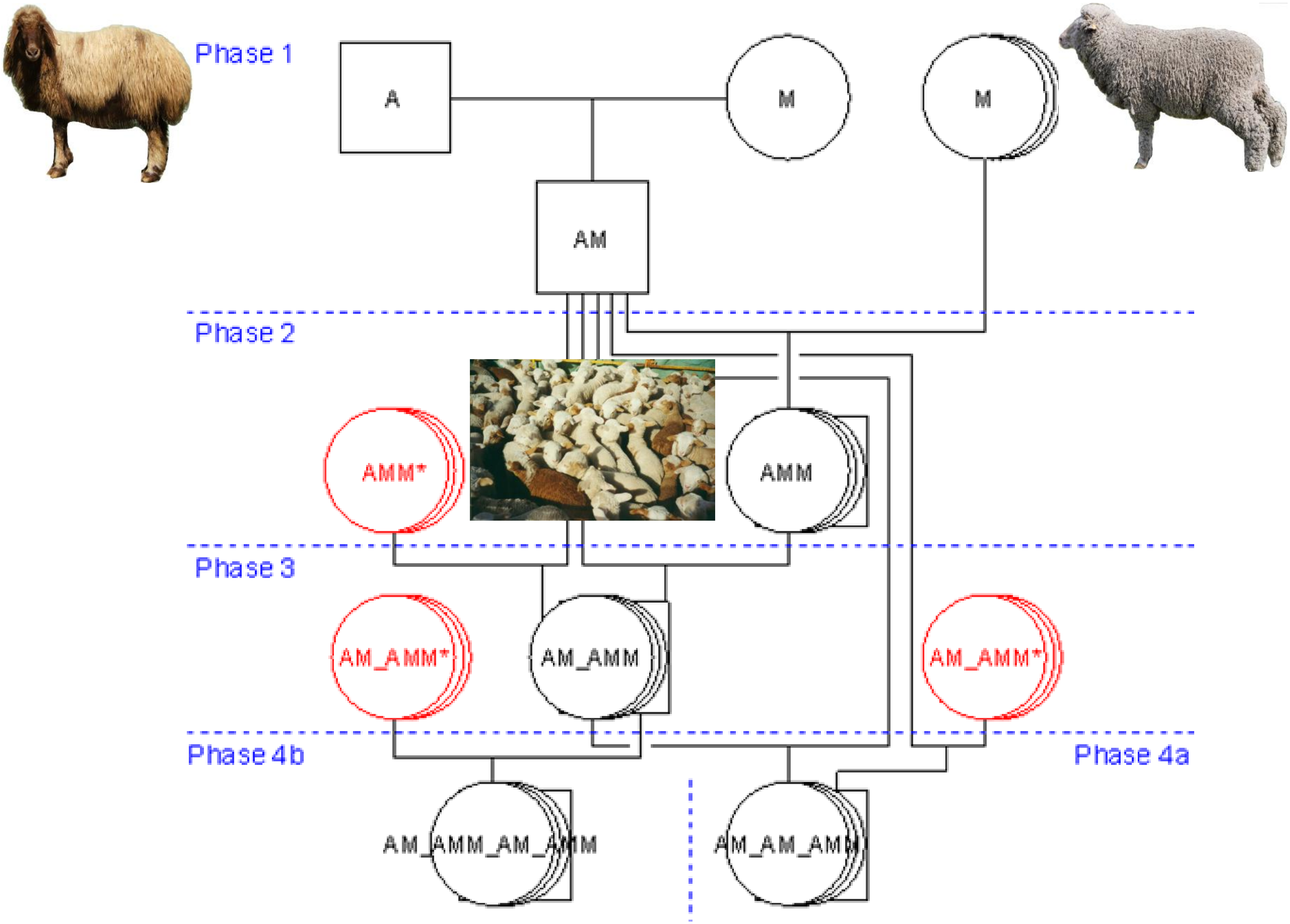
Hair vs wool

Provided by Chuck Dodd

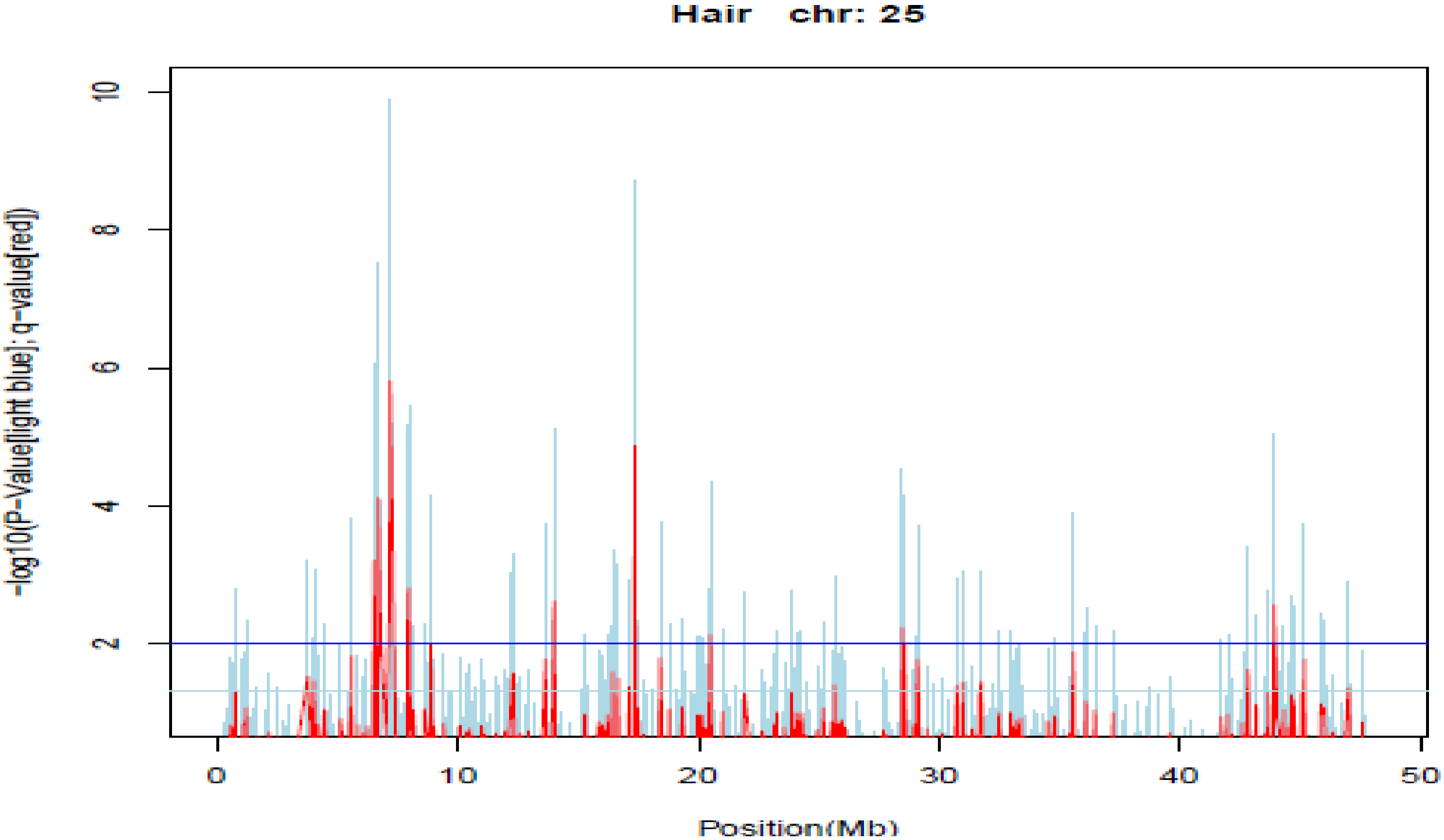


Genome-wide p-value qvalue for HvCW

mapping hair/carpet wool vs fine wool QTL



Hair vs Wool (OAR25)



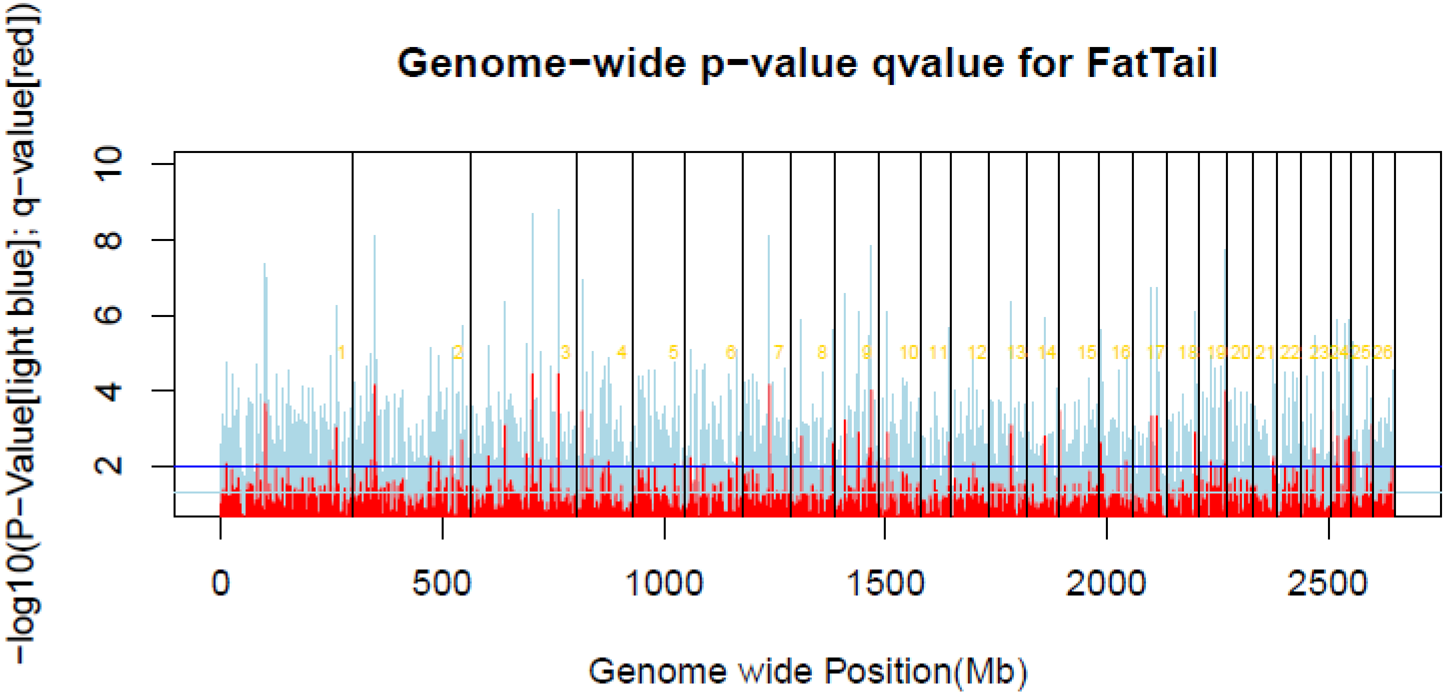
International Sheep Genomics Consortium

Fat vs non fat tail

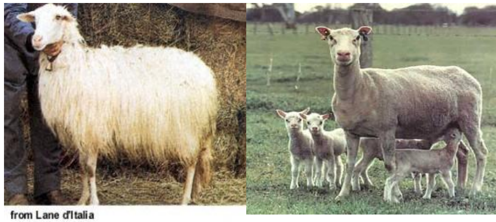
provided by Ebrahim Nooriyan Sooror

Provided by Chuck Dodd

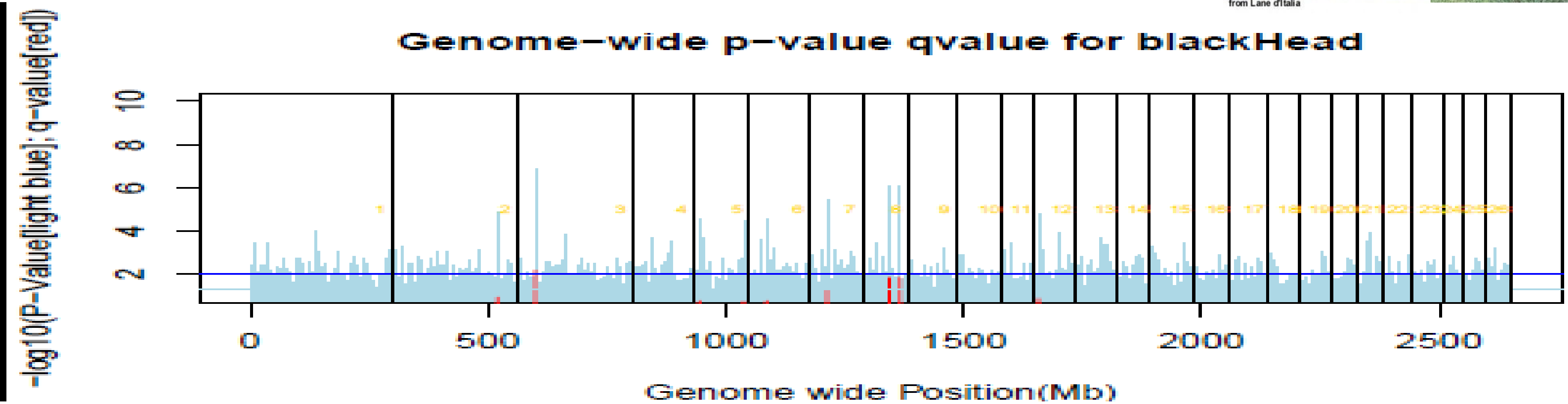
provided by Ebrahim Nooriyan Sooror

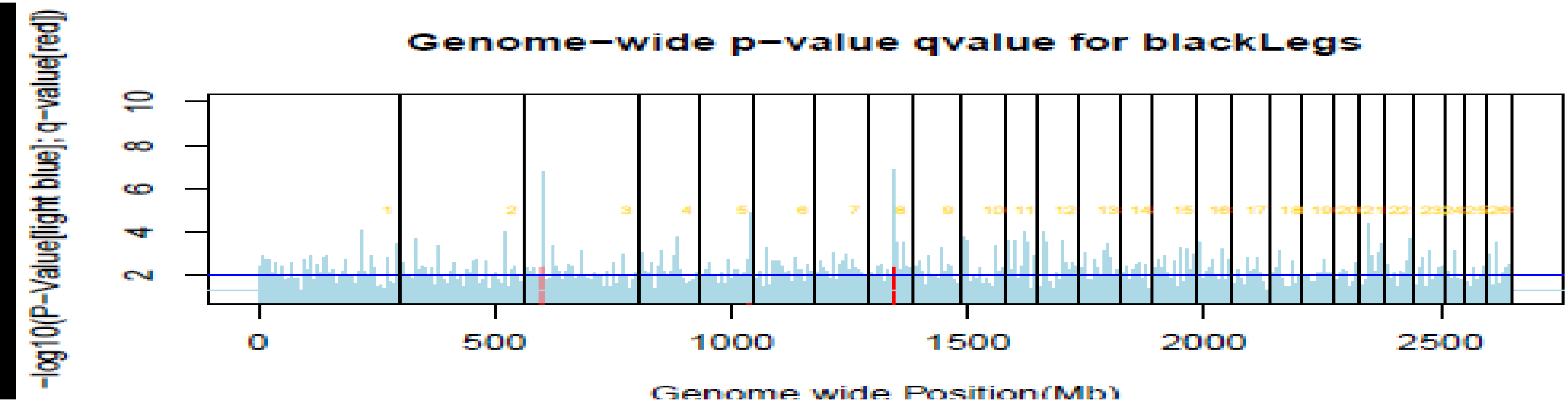


Colour markings



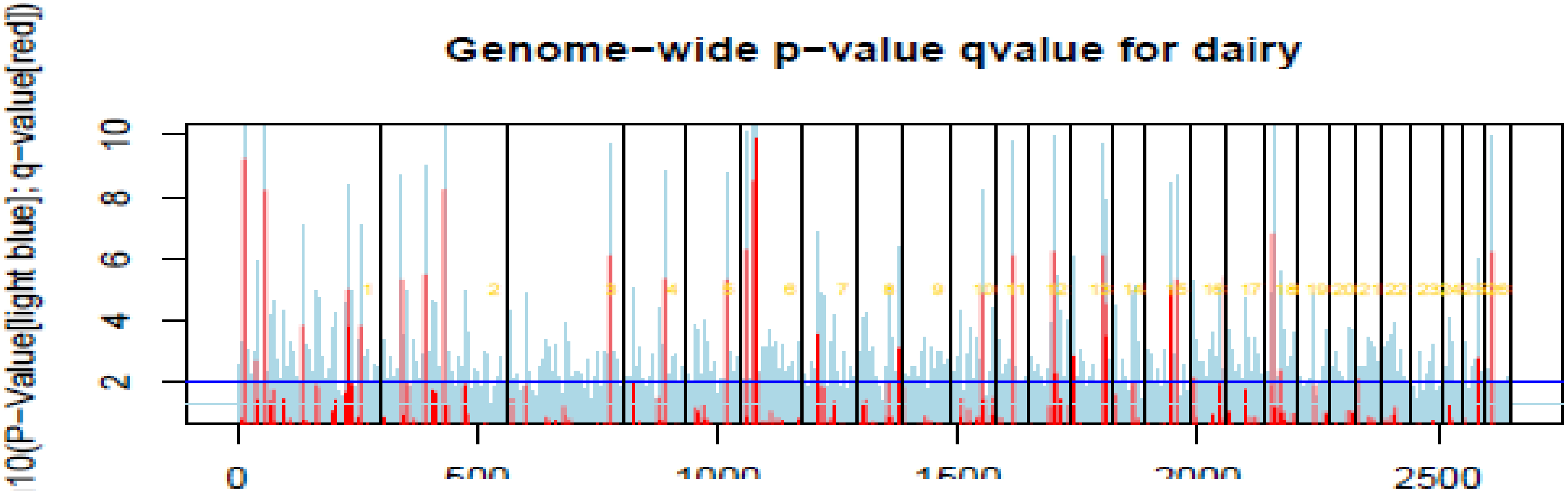
from Lane d'Italia


Genome−wide p−value qvalue for blackHead


Genome−wide p−value qvalue for blackLegs

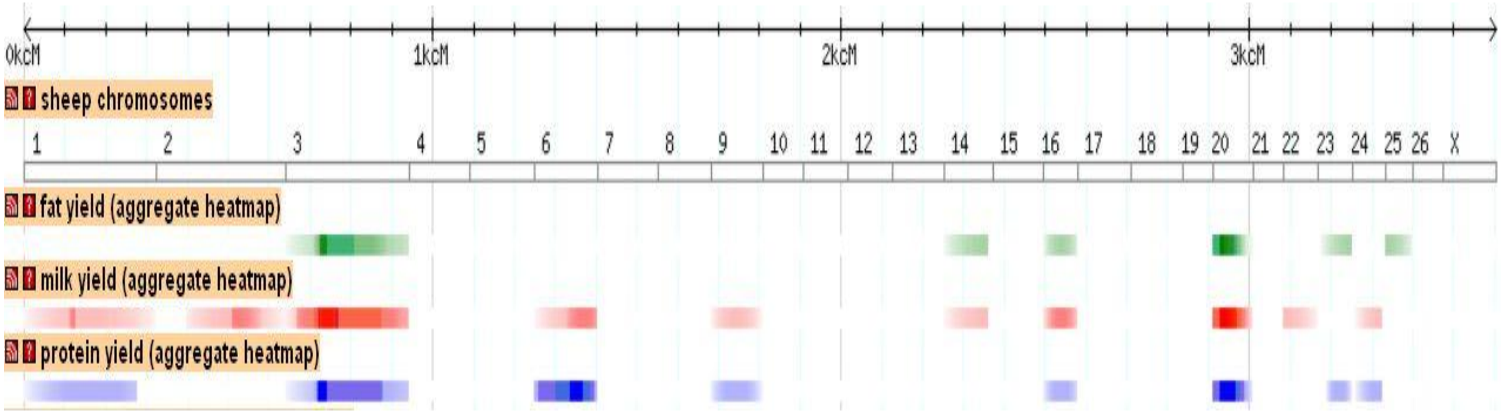
ISGC International Sheep Genomics Consortium

Dairy breeds vs non dairy



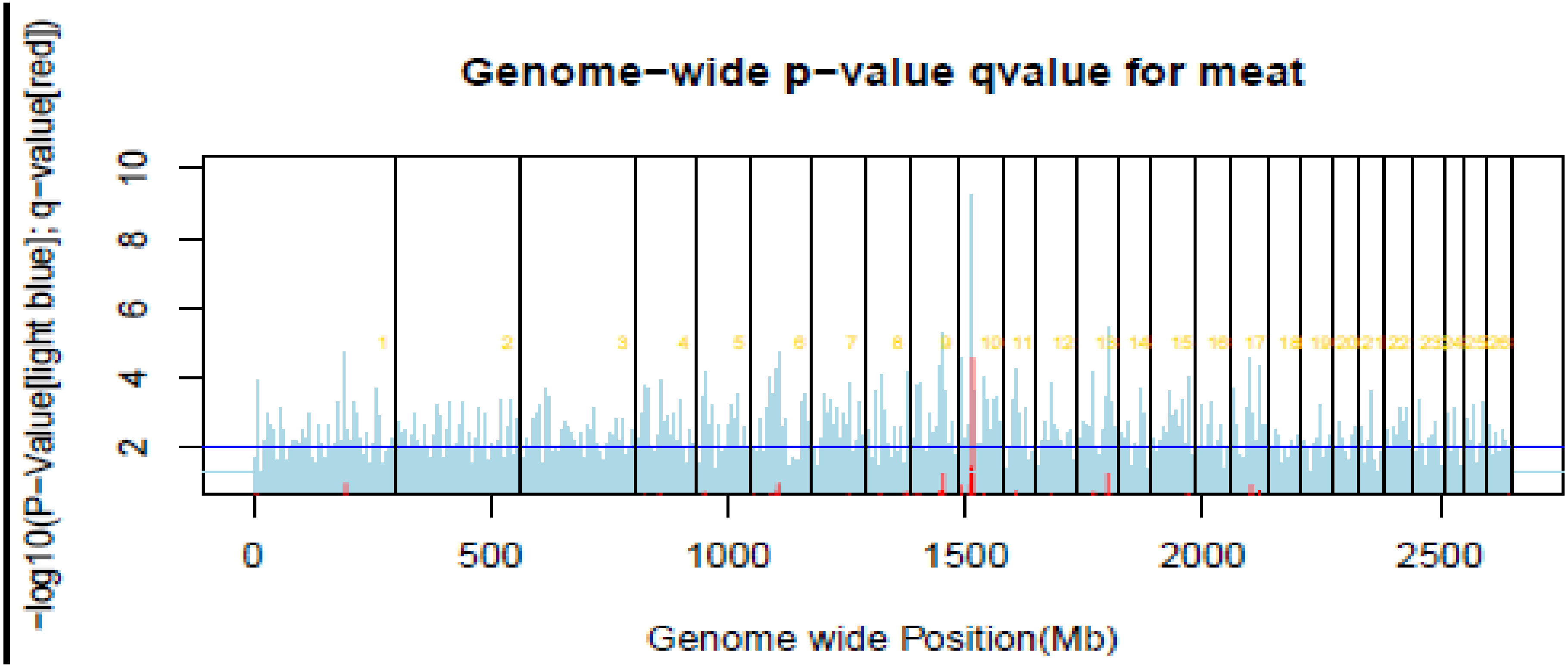
Compare genome wide QTL meta scores for dairy

http://crcidp.vetsci.usyd.edu.au/cgi-bin/gbrowse/oaries_genome/#search



provided by Gary Onan

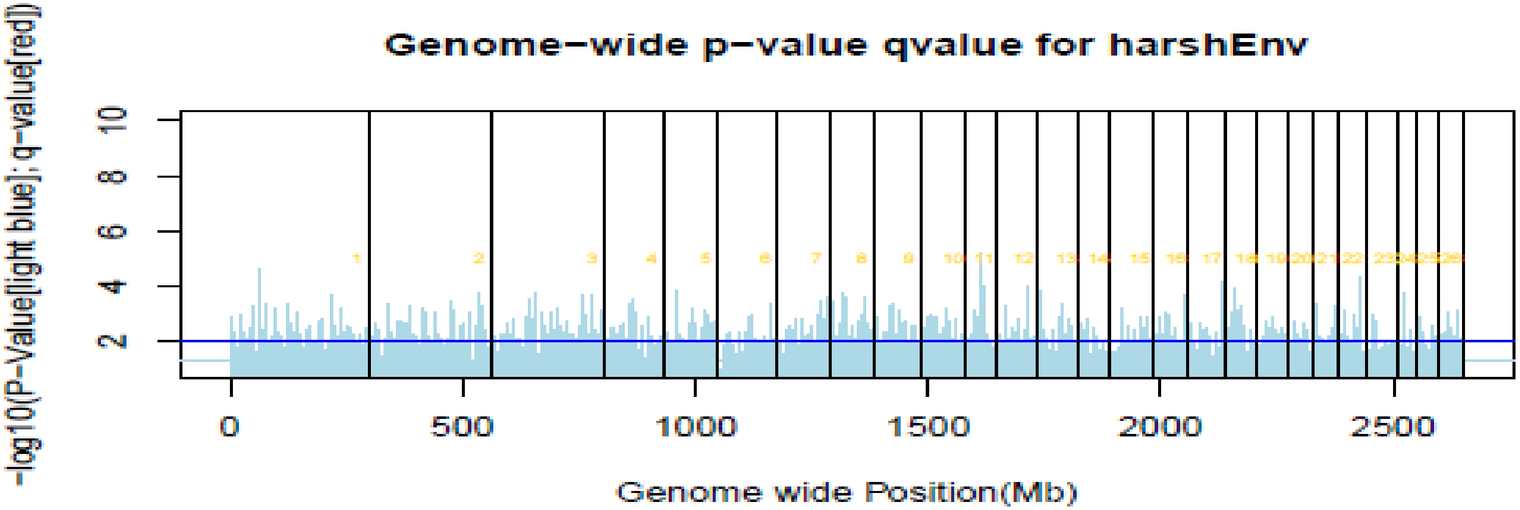
Meat type vs non meat

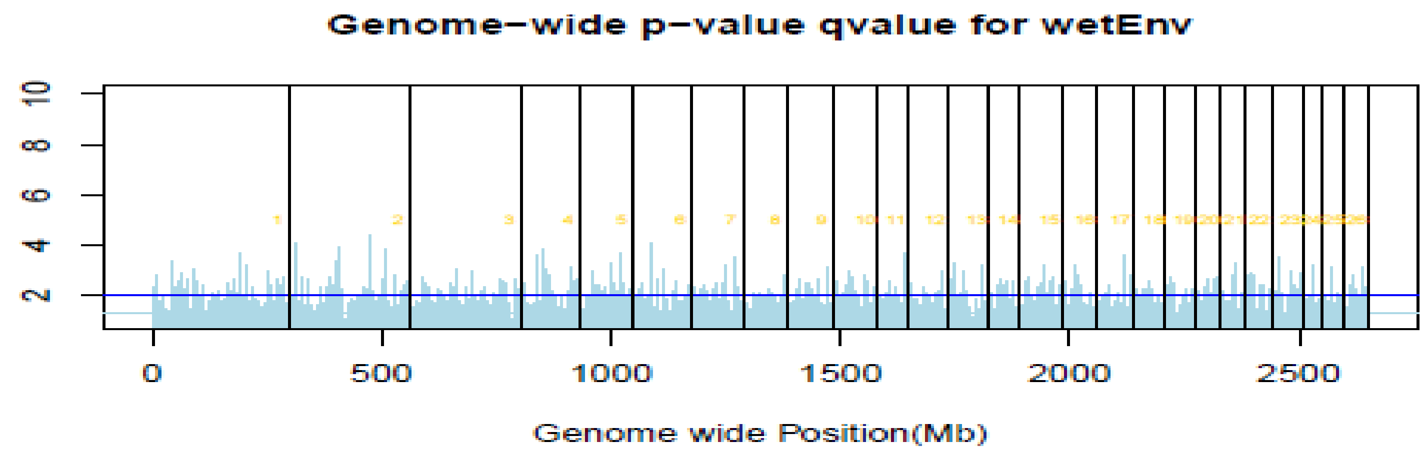


Confounding with other traits????????

adaptation..wet & harsh environment

Provided by Chuck Dodd





ISGC
International Sheep Genomics Consortium

Conclusions

- Strong associations with known major genes horned (OAR10)
- Strong associations and wool trait (OAR25) are consistent with QTL reports, but map to finer regions
- Potential to fine-map genes of major effect which have lead to breed formation
- Beware confounding
- Beware phenotype classification
- Enhancements: Use of trait means (continuous)
  - but may need to consider environment
- Further precision: inclusion of individual animal genotypes / phenotypes

ISGC

Acknowledgements

CSIRO

James Kijas
Brian Dalrymple
Wes Barris
David Townley
Sean McWilliam
Russell McCulloch

agresearch

John McEwan
Rudi Brauning
Theresa Wilson
Alan McCulloch

illumina

Marylinn Munson
Kim Gietzen
Christian Haudenschild

agrf
Australian Genome Research Facility Ltd

Peter Wilson
Annette McGrath

Human Genome Sequencing Center
Baylor College of Medicine

Richard Gibbs
Donna Muzny

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Jo-Ann Stanton

sheepGENOMICS

USDA United States Department of Agriculture

An Australian Government Initiative
Backing Australia's Ability

The University of Sydney

MEAT & WOOL
NEW ZEALAND

genesisfaraday

UtahState University

Jill Maddox (U. Melb)
Herman Raadsma (U. Syd)
Frank Nicholas (U. Syd)
Hutton Oddy (UNE)
Chris Warkup (GF)
Noelle Cockett (USU)

Tim Smith (USDA)
Mike Heaton (USDA)
Curt van Tassell (USDA)
Terry Longhurst (MLA)
Mohammad Shariflou (U. Syd)
Ben Hayes (Vic DPI)

And finally……..

Better, faster cheaper than Cattle Chicken and Pigs

Learn from others

Do what is best

www.sheephapmap.org

www.livestockgenomics.csiro.au/sheep/oar1.0.php

isgcdata.agresearch.co.nz

*Statistical Model for Association

- **X** = SNP predictor variable(s)
- For a SNP with genotypes AA AG GG
  p_{AA} = proportion of animals with genotype AA
  p_{AG} = proportion of animals with genotype AG
  p_{GG} = proportion of animals with genotype GG
- Then $x_1 = p_{\mathrm{AA}}$, $x_2 = p_{\mathrm{AG}}$
- Cannot include $x_3 = p_{\mathrm{GG}}$: $p_{\mathrm{AA}} + p_{\mathrm{AG}} + p_{\mathrm{GG}} = 1$