Genetic structure and selection in the sheep examined using the SNP50 BeadChip

James Kijas on behalf of the International Sheep Genomics Consortium

ISGC Presentations

James Kijas, CSIRO Livestock Industries (james.kijas@csiro.au)

"Impact of Selection on the Sheep Genome: Findings from the ISGC HapMap Experiment"

John McEwan, AgResearch (john.mcewan@agresearch.co.nz)

"Putting the sheep SNP50 BeadChip to work: Case Studies in Gene Mapping and Genomic Selection"

Herman Raadsma, The University of Sydney (h.raadsma@usyd.edu.au)

"LD analyses in the ISGC sheep Hap-Map data set using a 50k SNP array"

ISGC Organisation and Role

- Open organisational structure (people come and go)

EVERYONE IS WELCOME

- Coordinate funding opportunities
- Avoid scientific redundancy, economy of scale
- Process: fortnightly phone conferences (secretary James.Kijas@csiro.au)

Next Meeting: 11:30 – 3:00 this Monday here at PAG, Sunset Room

All data generated is public domain to benefit the entire sheep genomics community

www.sheephapmap.org

ISGC Major Objectives

- **SNP discovery / chip development**
  - Three completed SNP discovery projects (2008)
  - Release of the *ovine* SNP50 BeadChip at PAG 2009

- **HapMap and Breed Diversity Experiment**
  - Genetic diversity, breed relationships
  - Linkage disequilibrium, implications for gene discovery and genomic selection
  - Impact of domestication and selection

- **Reference Genome**
  - Illumina and 454 blend sequencing of a single Texel ram (ongoing)

Development of the *ovine* SNP50 BeadChip

Human Genome Sequencing Center
Baylor College of Medicine

University of Otago
To Whare Wānanga o Otāgo
New Zealand

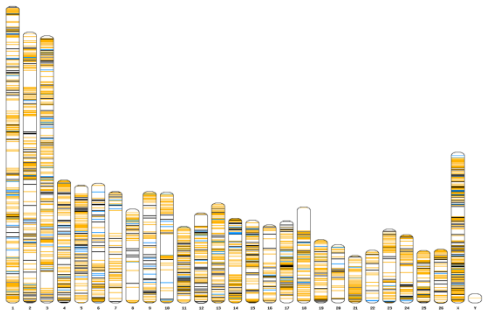
CSIRO

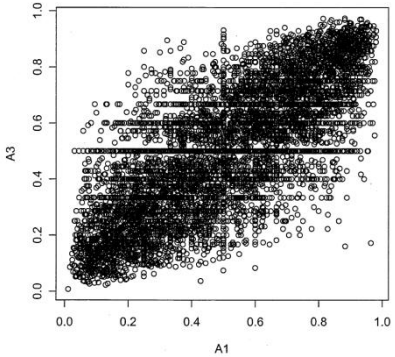
agresearch

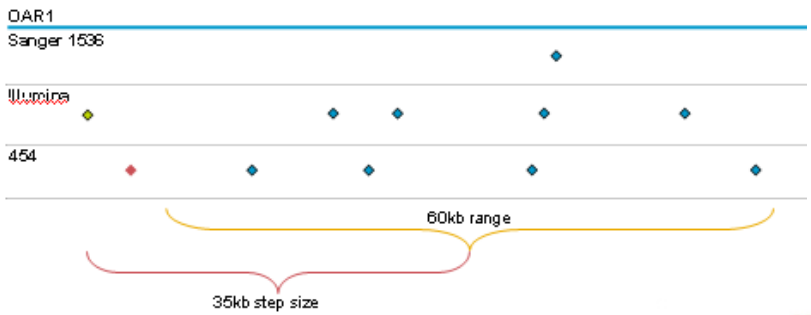
Sanger resequencing
9 animals, 9 breeds
6,021 SNP

3x genome using **454**
6 animals, 6 breeds
276,868 SNP

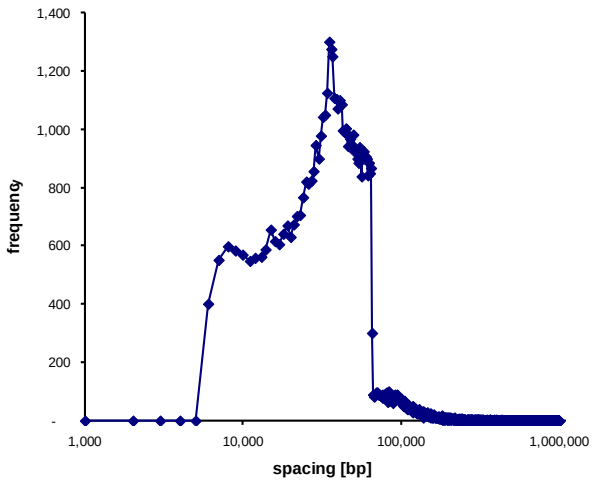
Illumina RRL
60 animals, 15 breeds
76,044 SNP





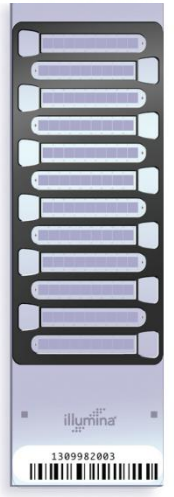


illumina

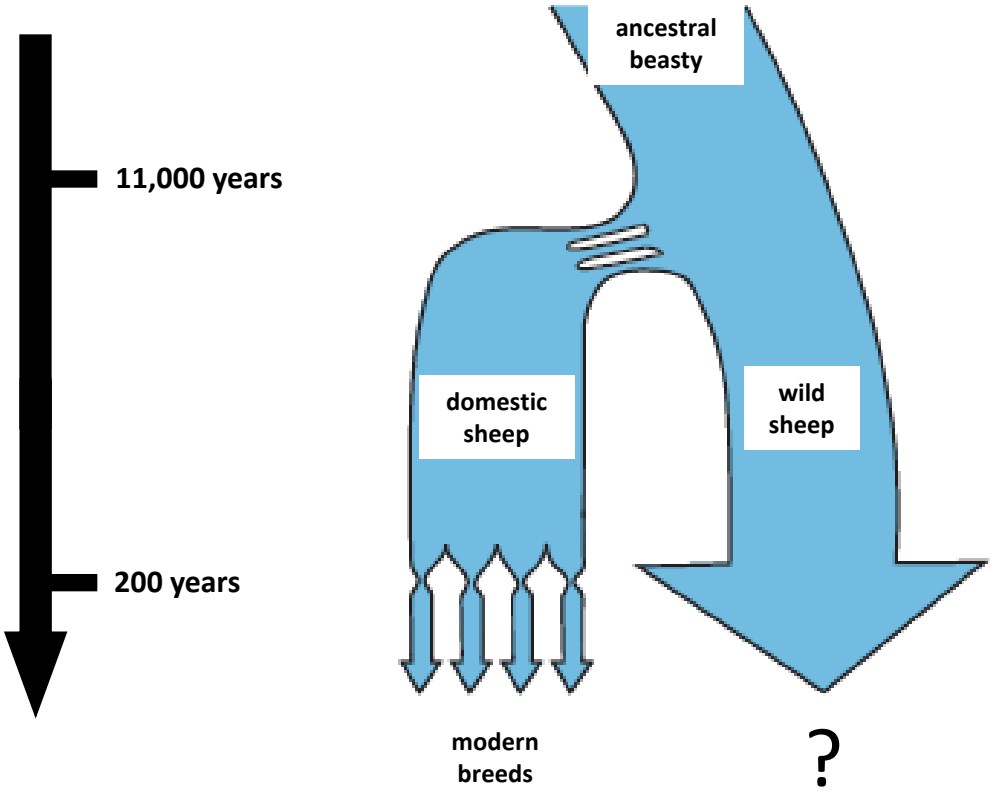


Illumina iSelect
***Ovine* SNP50BeadChip**
Launch Jan 2009

600	Sanger
40174	454
18669	Illumina GA
11	mtDNA
59454	Total SNP
49034	Passed manufacture and all QC



Aim: Understand the Genetic History of the Species



Adapted from Lindblad-Toh *et al.*, 2005

- Process of domestication (when, where, from what?)
- Genes which influence traits under selection
- Diversity within and between breeds
  - PCA, distance based networks, Structure

ISGC

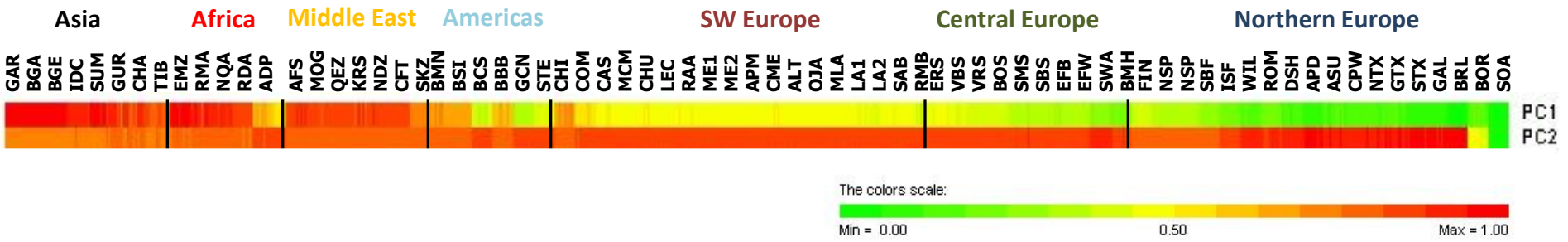
The sheep HapMap and Breed Diversity Experiment

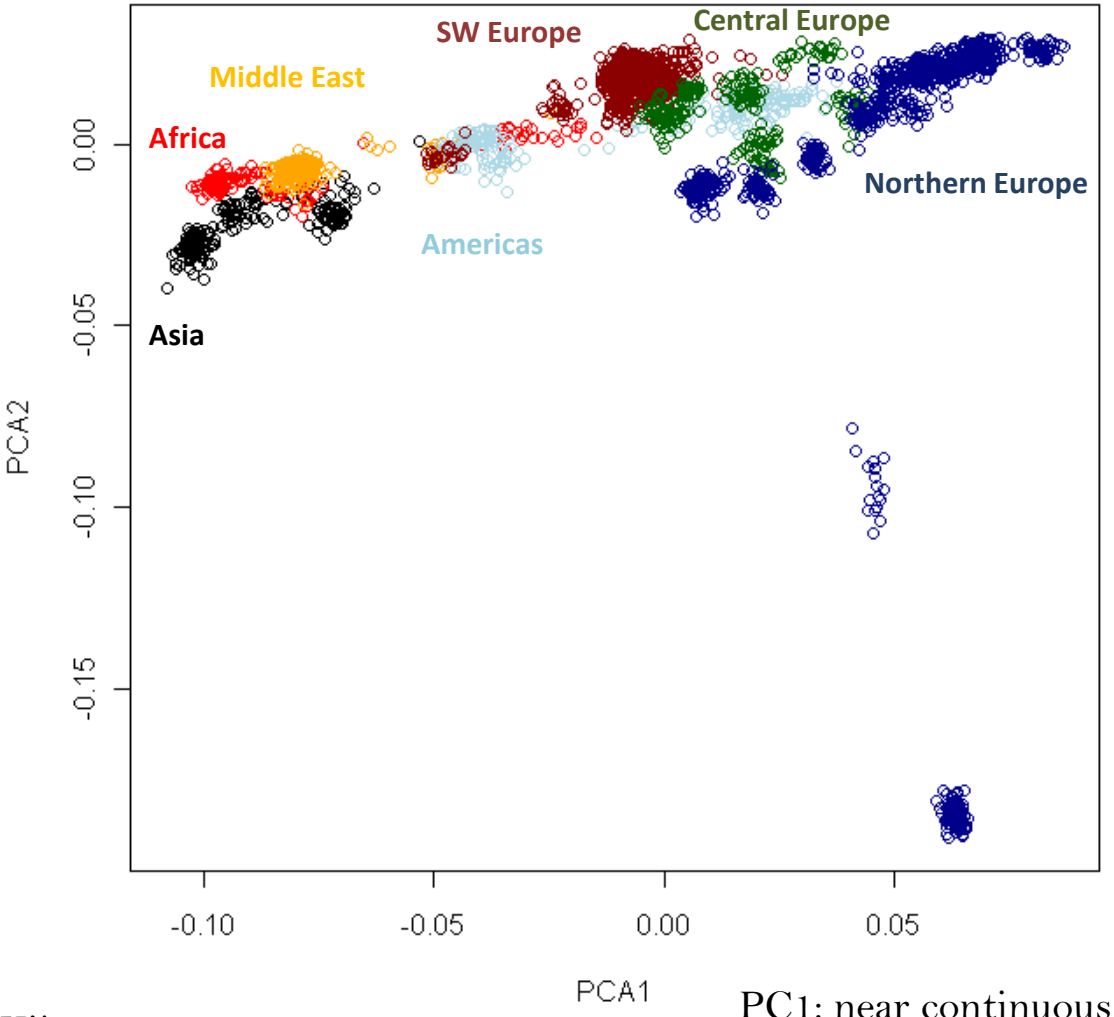
Total of 3400 animals genotyped using the SNP50 BeadChip

-74 breeds from Africa, Asia, Sth America, Nth America, Caribbean, Middle East, SW Europe, Central Europe, Northern Europe, Australia and New Zealand.

- selection of wild sheep (argali, urial, mouflon, bighorn, thinhorn)
- out group species (barbary, goat, thar, cattle..)

Lenstra and Kijas (workteam leaders for diversity)
Porto Neto, Boitard, Paiva, San-Cristobal..

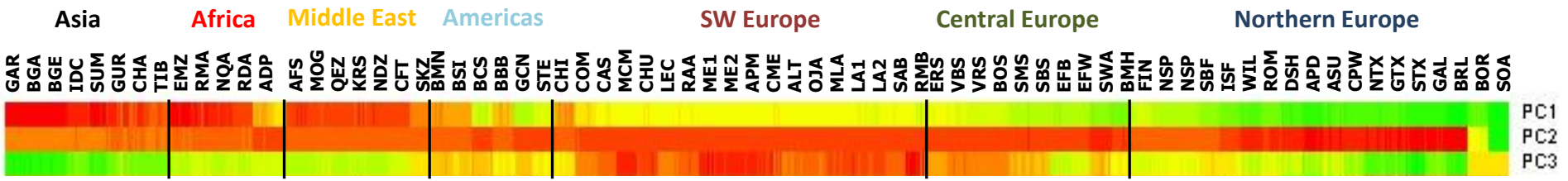




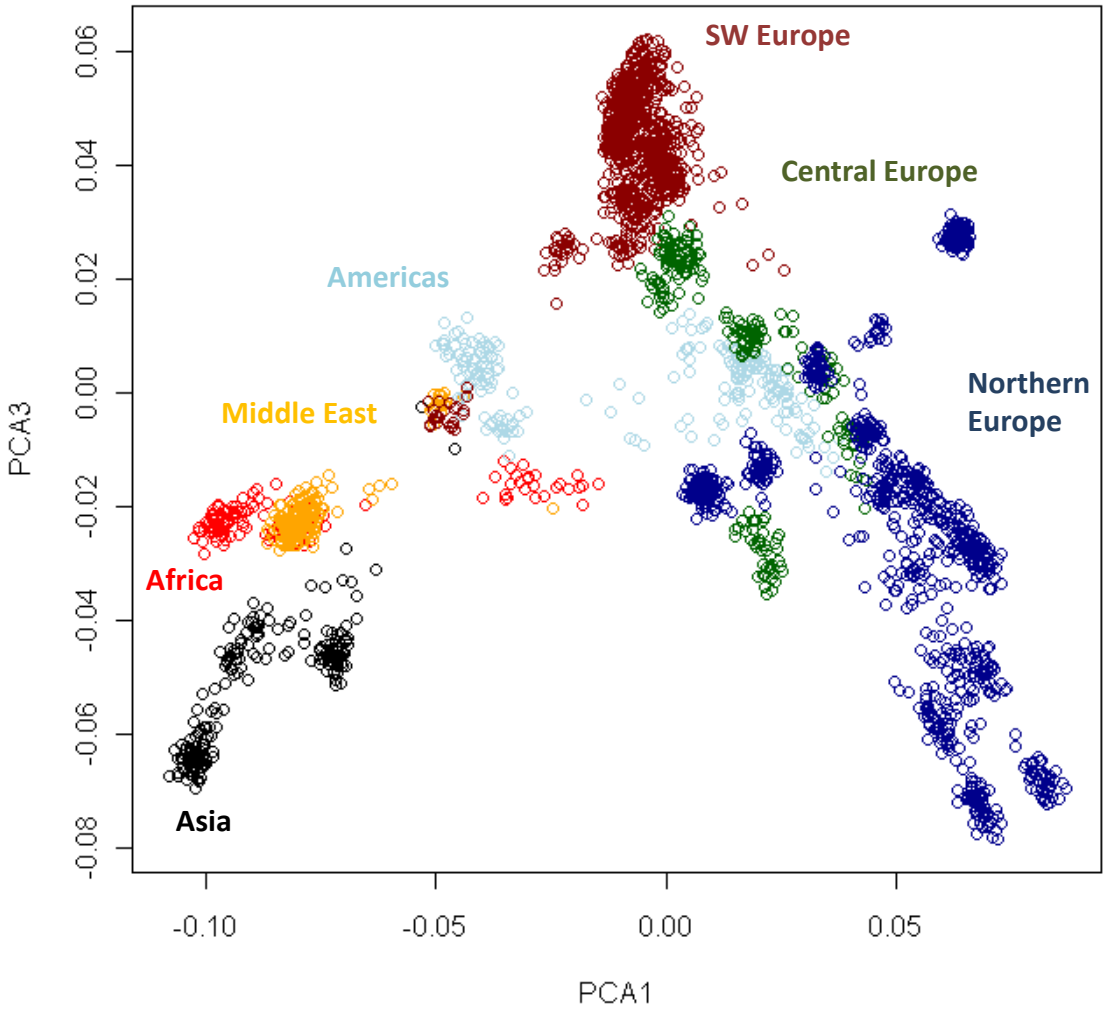
PC2:
Soay outgroup

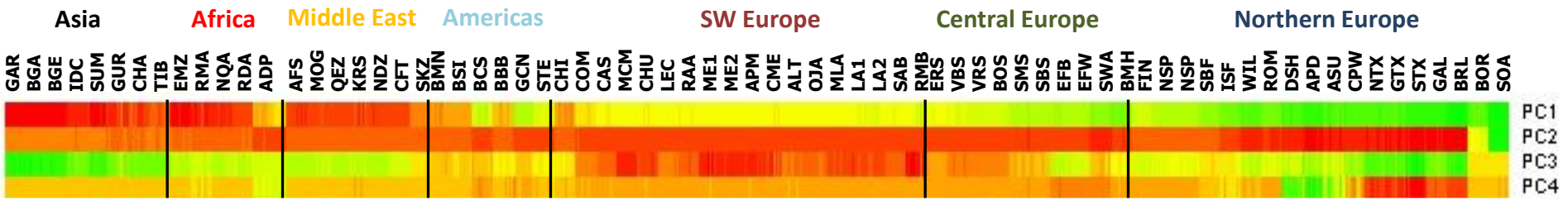
PC1: near continuous gradient

Porto Neto, Kijas



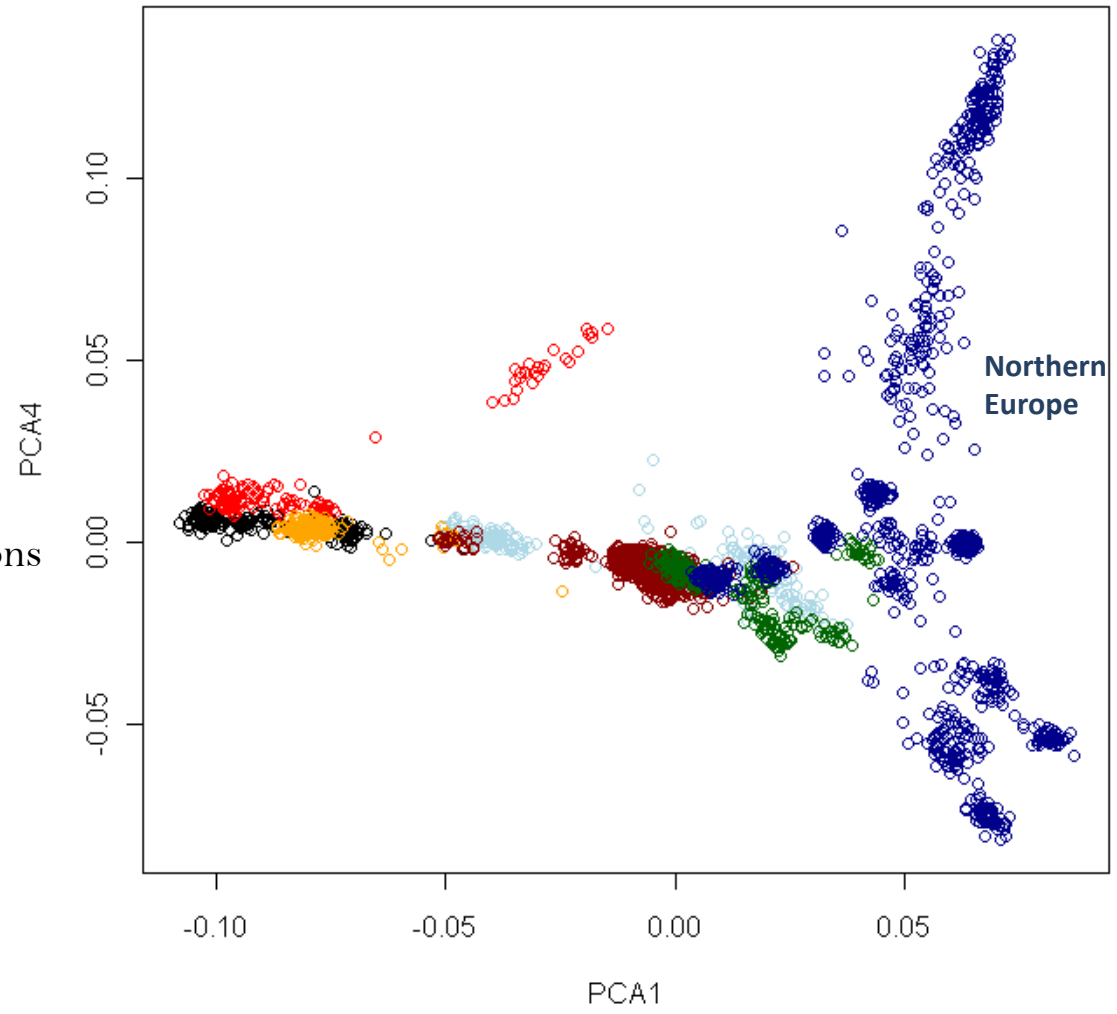
PC3:
Asian, African, Middle Eastern cluster out

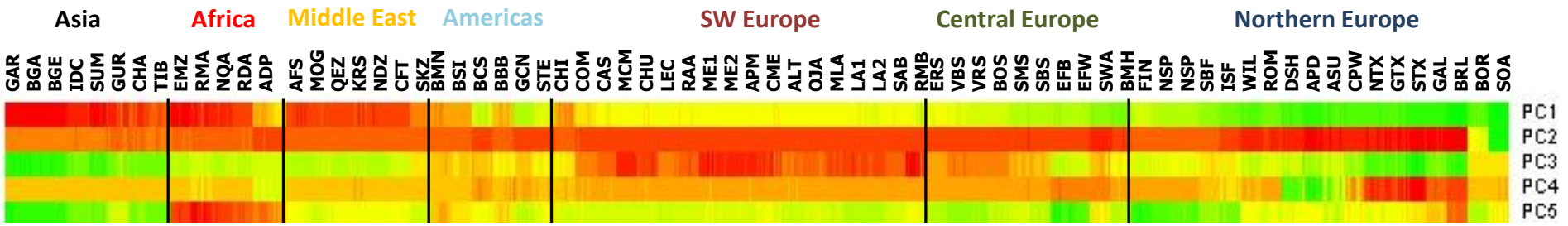




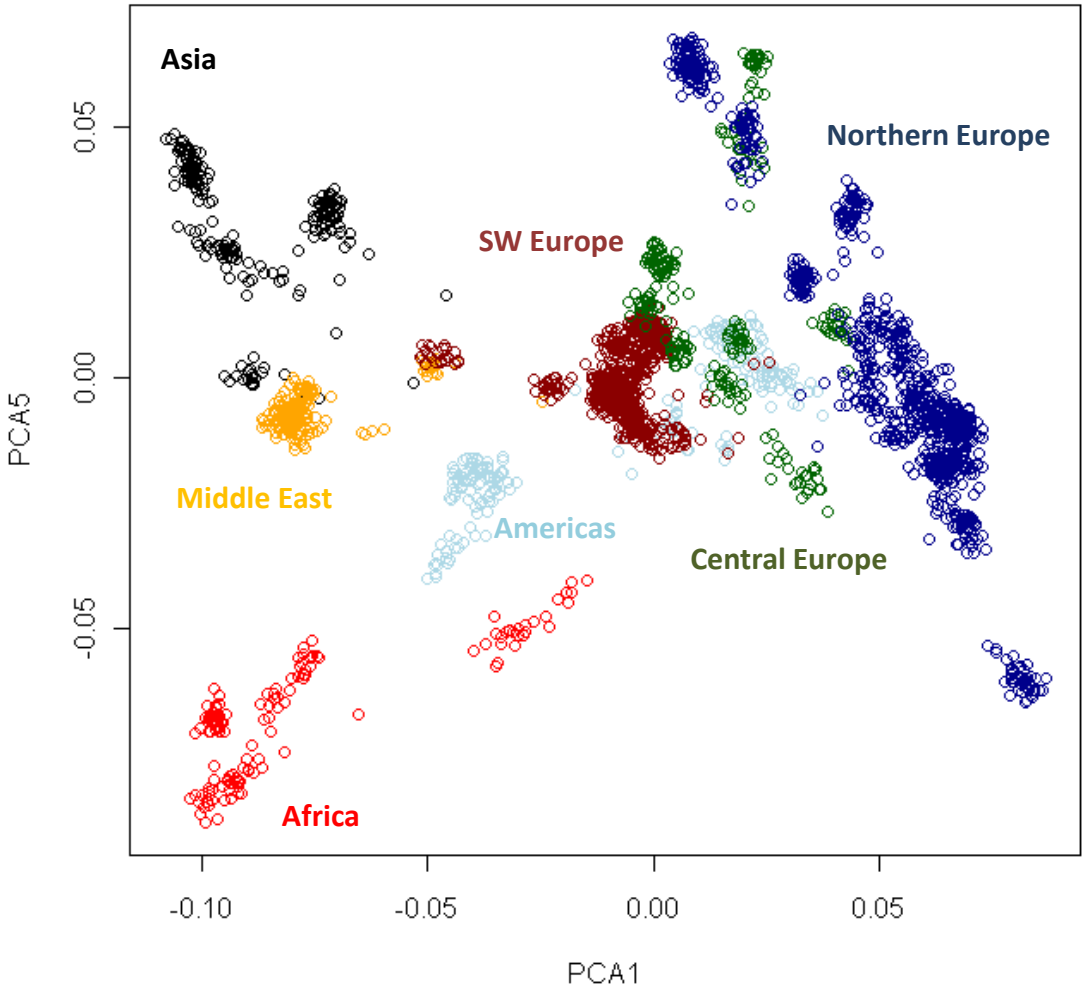
PC4:
Finescale clusters

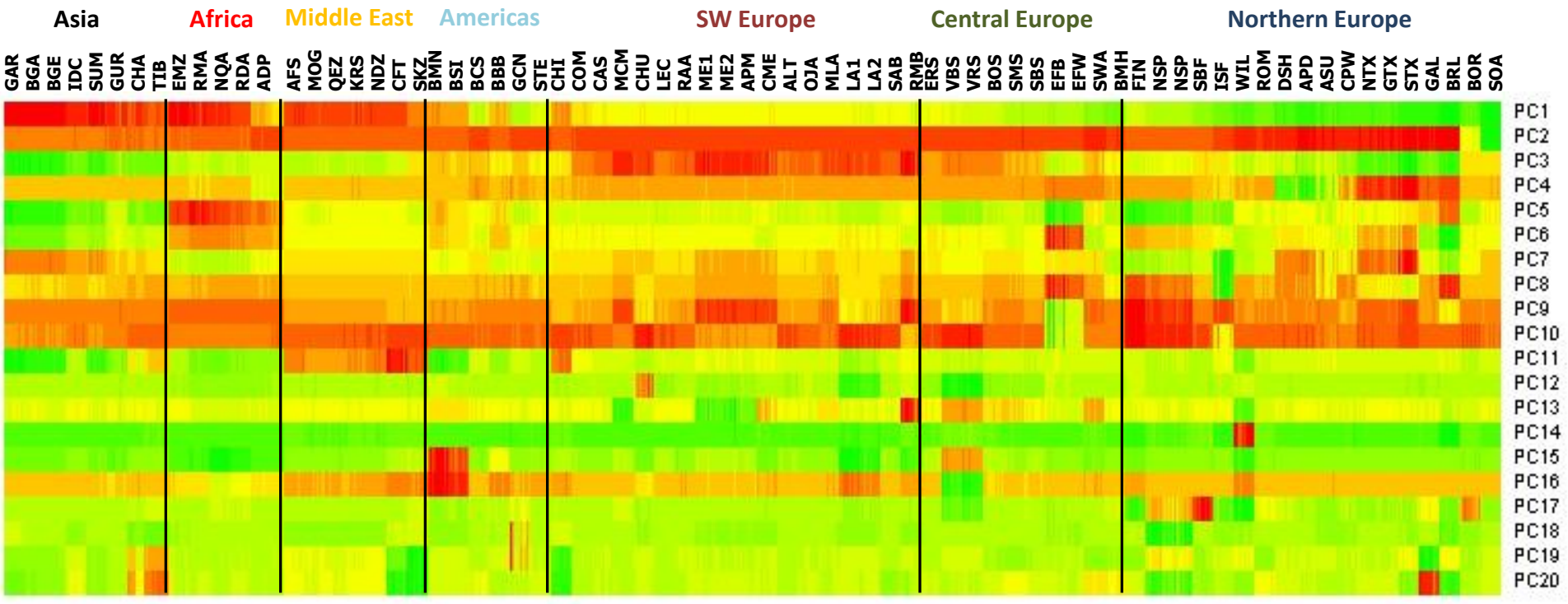
Admixed populations

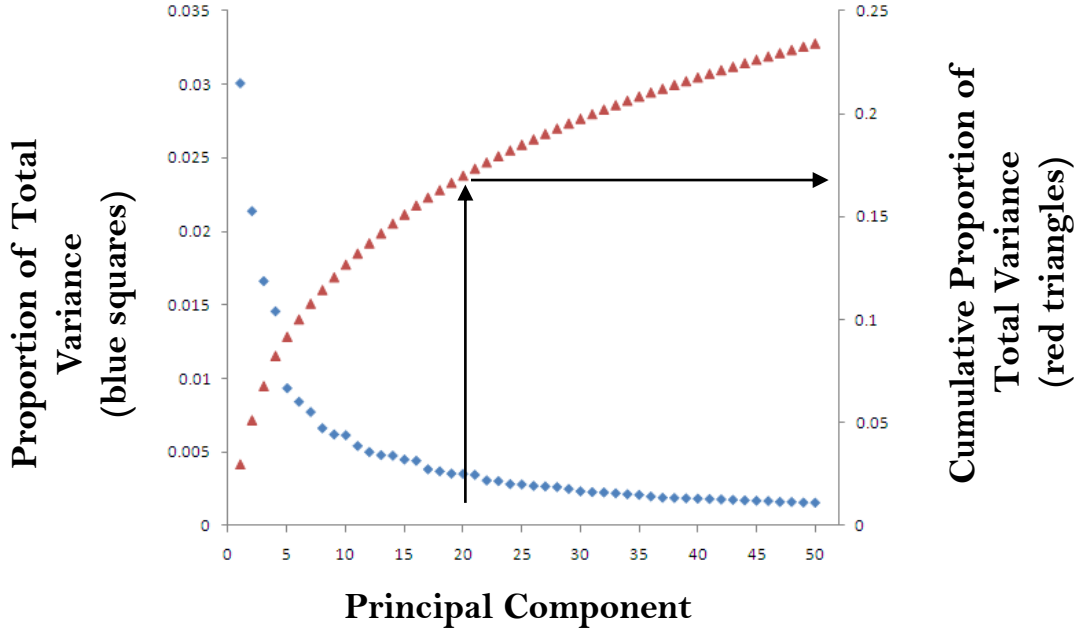




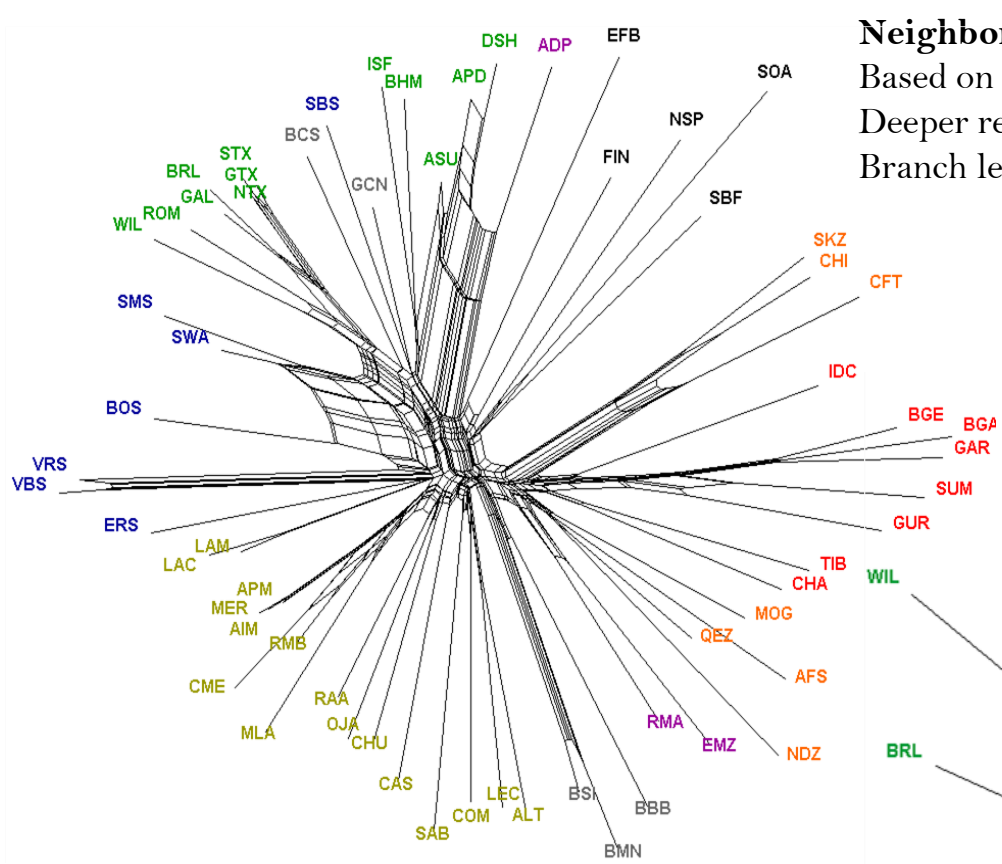
PC5:
Strong
phylogeography

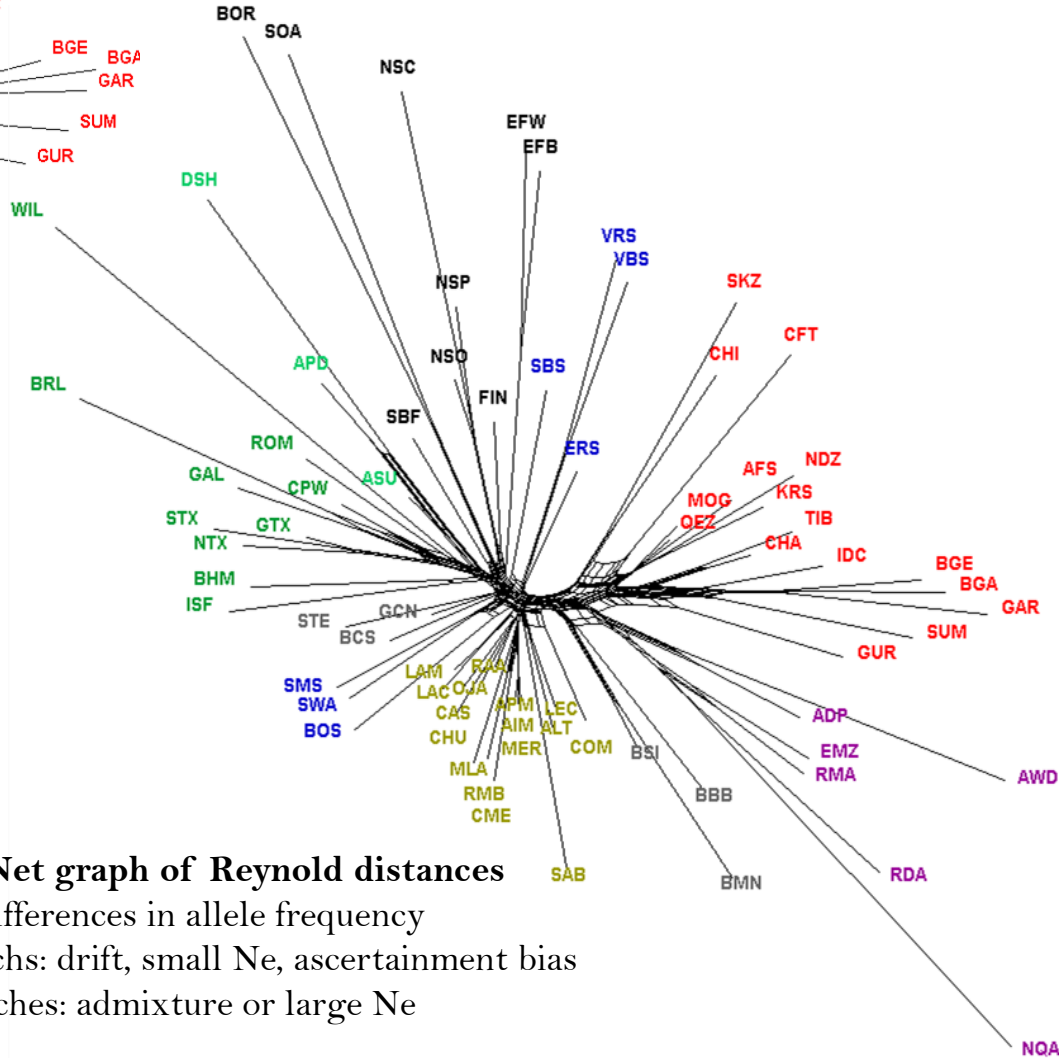



Asia
Africa
Middle East
Americas
SW Europe
Central Europe
Northern Europe
GAR BGA BGE IDC SUM GUR CHA TIB EMZ RMA NQA RDA ADP AFS MOG QEZ KRS NDZ CFT SKZ BMN BSI BCS BBB GCN STE CHI COM CAS MCM CHU LEC RAA ME1 ME2 APM CME ALT OJA MLA LA1 LA2 SAB RMB ERS VBS VRS BOS SMS SBS EFB EFW SWA BMH FIN NSP NSP SBF ISF WIL ROM DSH APD ASU CPW NTX GTX STX GAL BRL BOR SOA
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
PC11
PC12
PC13
PC14
PC15
PC16
PC17
PC18
PC19
PC20


Proportion of Total Variance (blue squares)
Cumulative Proportion of Total Variance (red triangles)
Principal Component
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30 35 40 45 50

NeighborNet graph of LD distances
Based on haplotype sharing
Deeper reticulations
Branch lengths less susceptible to drift, Ne, bias





NeighborNet graph of Reynold distances
Based on differences in allele frequency
Long branchs: drift, small Ne, ascertainment bias
Short branches: admixture or large Ne

Lenstra

Genetic Diversity

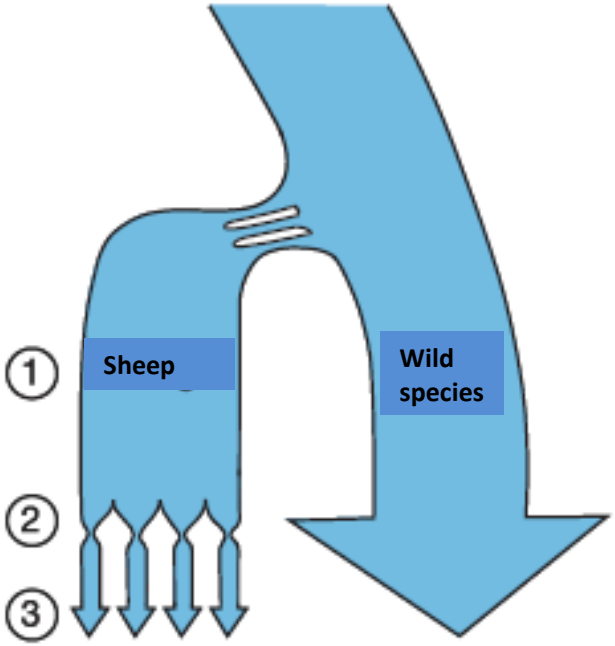
- continuous gradient in genetic diversity
  - different from cattle, no evidence of *O. aries* subspecies
  - no clear signal of separate domestication events
  - additional mtDNA data required

- clear phylogeographic structure

  - mtDNA: not good at showing relatedness between breeds
  - 1536 SNP: detected high introgression, broad scale structure
  - 50K SNP: finescale structure

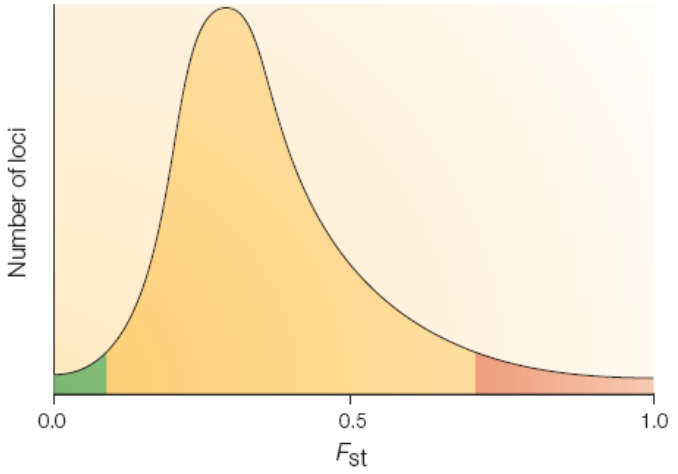
- identification of admixed populations
  - African Dorper, Gulf Coast Native, Merino derivatives

Selection



Distribution of *F*st

Fst gives the degree of allele frequency difference



Luikart, England, *et al.*, 2003
Concept of "Population Genomics"

Neutral loci
Genome wide forces

Bottlenecks, drift, Population diversity

Outlier loci
Locus specific forces

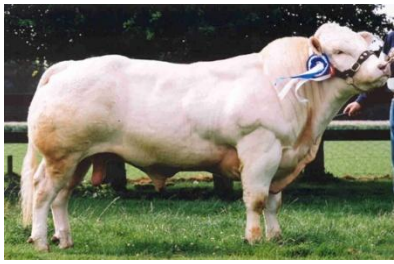
Selection following breed formation

Barendse, Hayes and McEwan (workteam leaders)
Gill, Kijas, Porto Neto, Scheet, Raadsma

Case 1: Outlier Loci and the Detection of Selection

Growth differentiation factor 8 (*myostatin*)

- negative regulator of muscle mass
- naturally occurring *myostatin* mutants identified in various species



Cattle: Belgian Blue

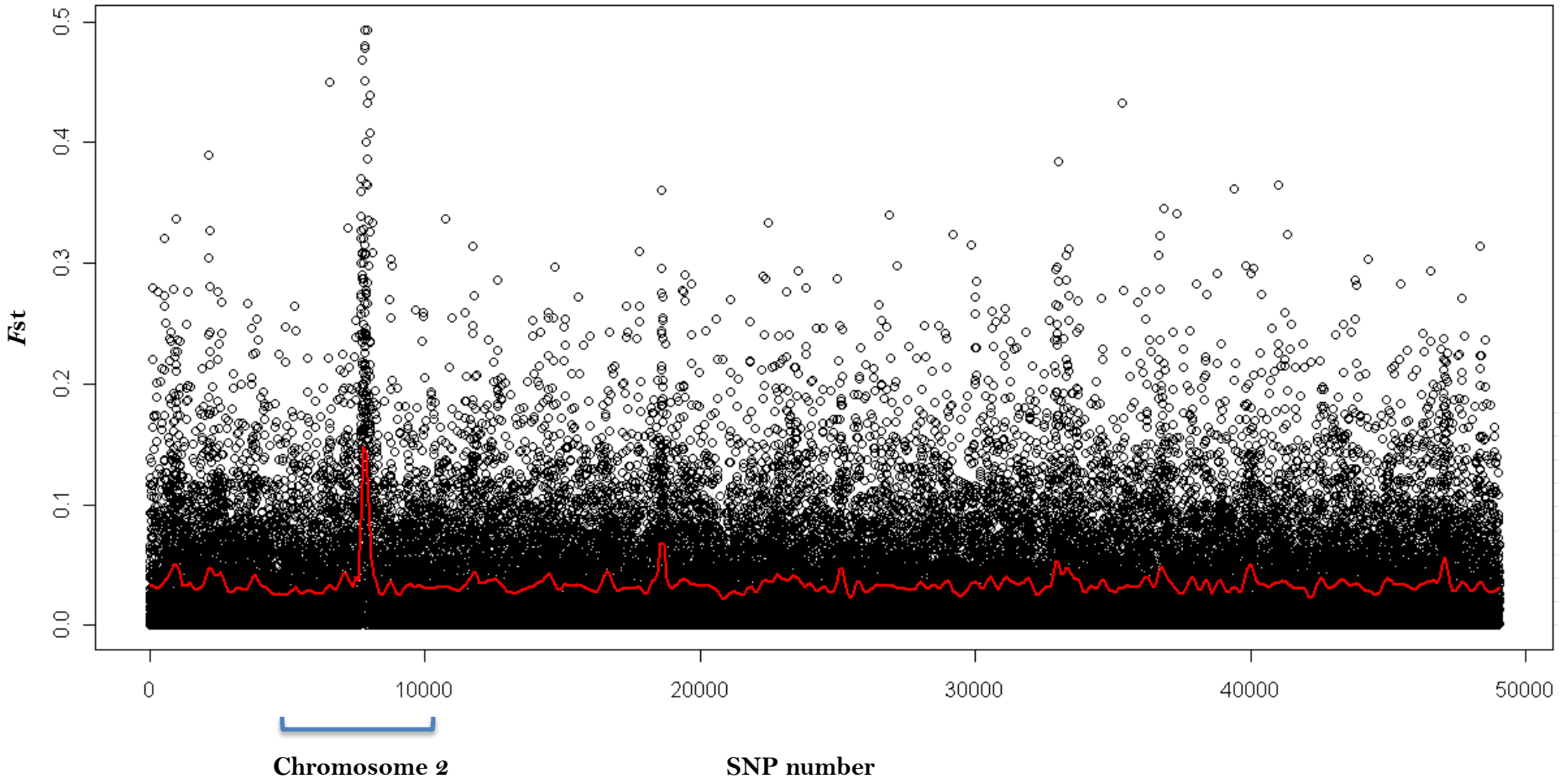
- located on sheep chromosome 2
- three distinct samplings of Texel in the experiment (German Texel, Scottish Texel and New Zealand Texel)

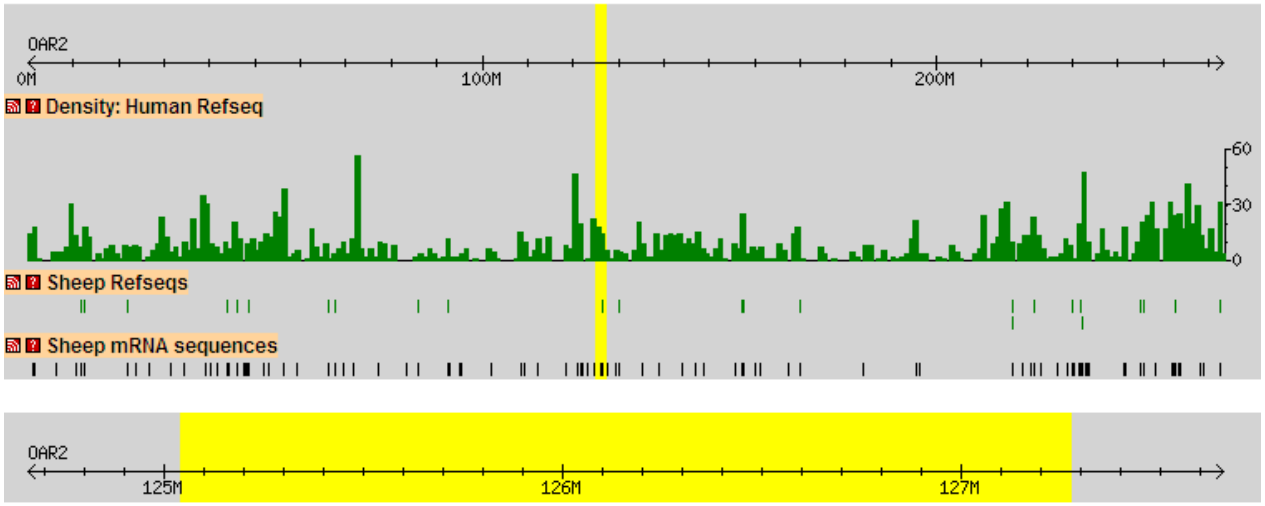
Dog: Whippet

- Estimated population pairwise *F*st between Texels and all others breeds for each of 49034 SNP. Looked for outliers.

Sheep:Texel
g+6723 G>A

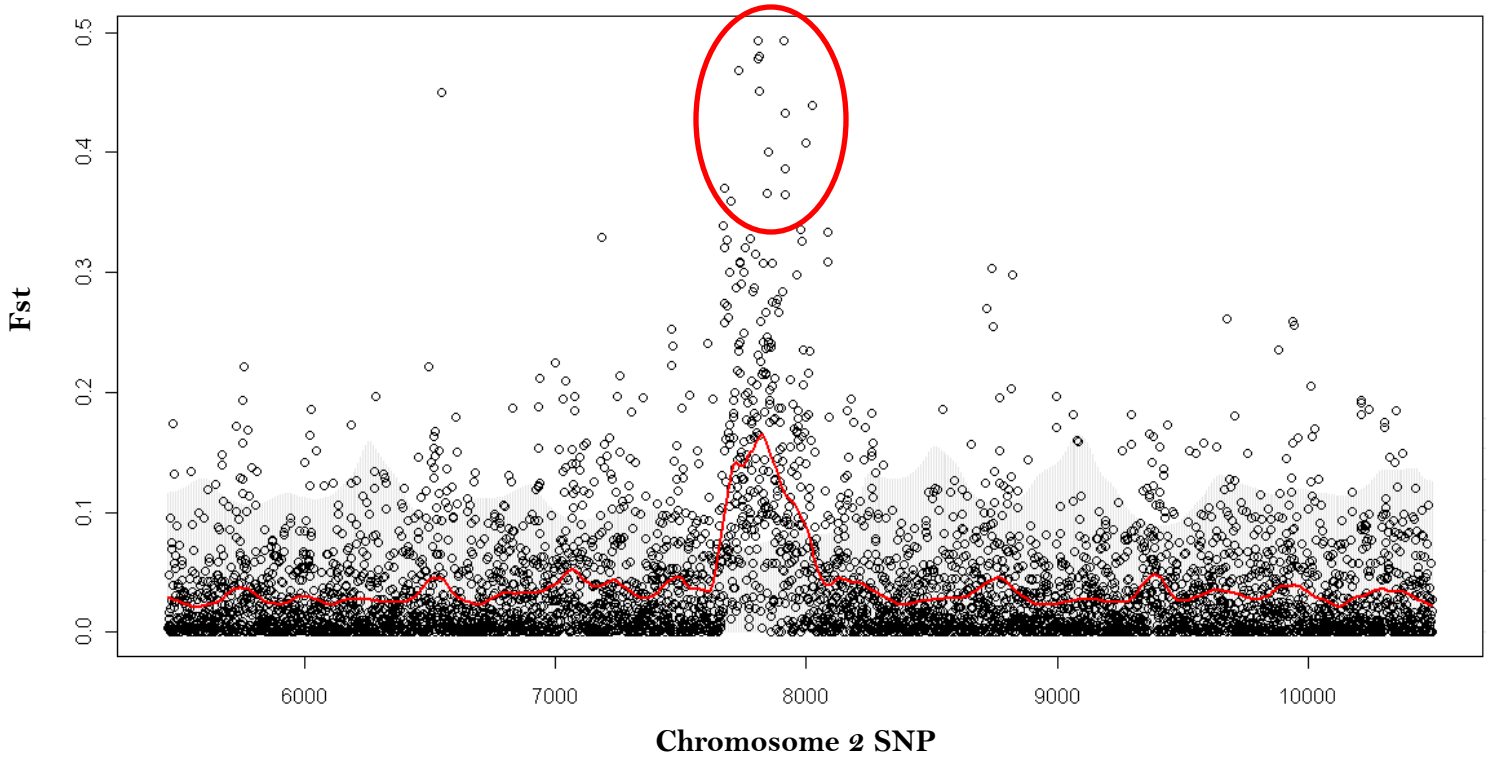
Genome Wide Scan for extreme *F*st in Texel





2.2 Mb region
Sheep chr 2

15 human Refseqs
including *myostatin*



Phased haplotypes for 74 sheep breeds

150 SNP window
Sheep chr 2
spanning *myostatin*

Scheet

Case 1: Growth differentiation factor 8 (*myostatin*)

- primary signal was both strong and wide
- haplotypes in affected breeds long and near fixation
- selection pressure likely to be recent

- outlier scan alone would not have been sufficient to incriminate the gene

Case 2:

The Horn Locus (*Ho*)

- wild sheep and many domestic breeds have horns
- selection for hornless (polled) breeds documented from 15th century
- many breeds polled, others mixed

- *Ho* mapped to sheep chromosome 10 (Montgomery *et al.*, 1996)
- ongoing positional cloning effort to identify the casual variant(s)

- Estimated population *F*st between:

  - group containing three European derived horned breeds
  - group containing eight European derived polled breeds


Mouflon


Australian Merino

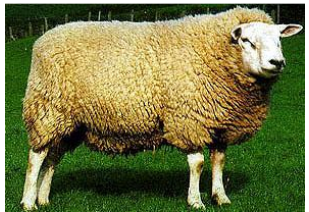

Gulf Coast Native


Black Welsh

White Suffolk

Romney

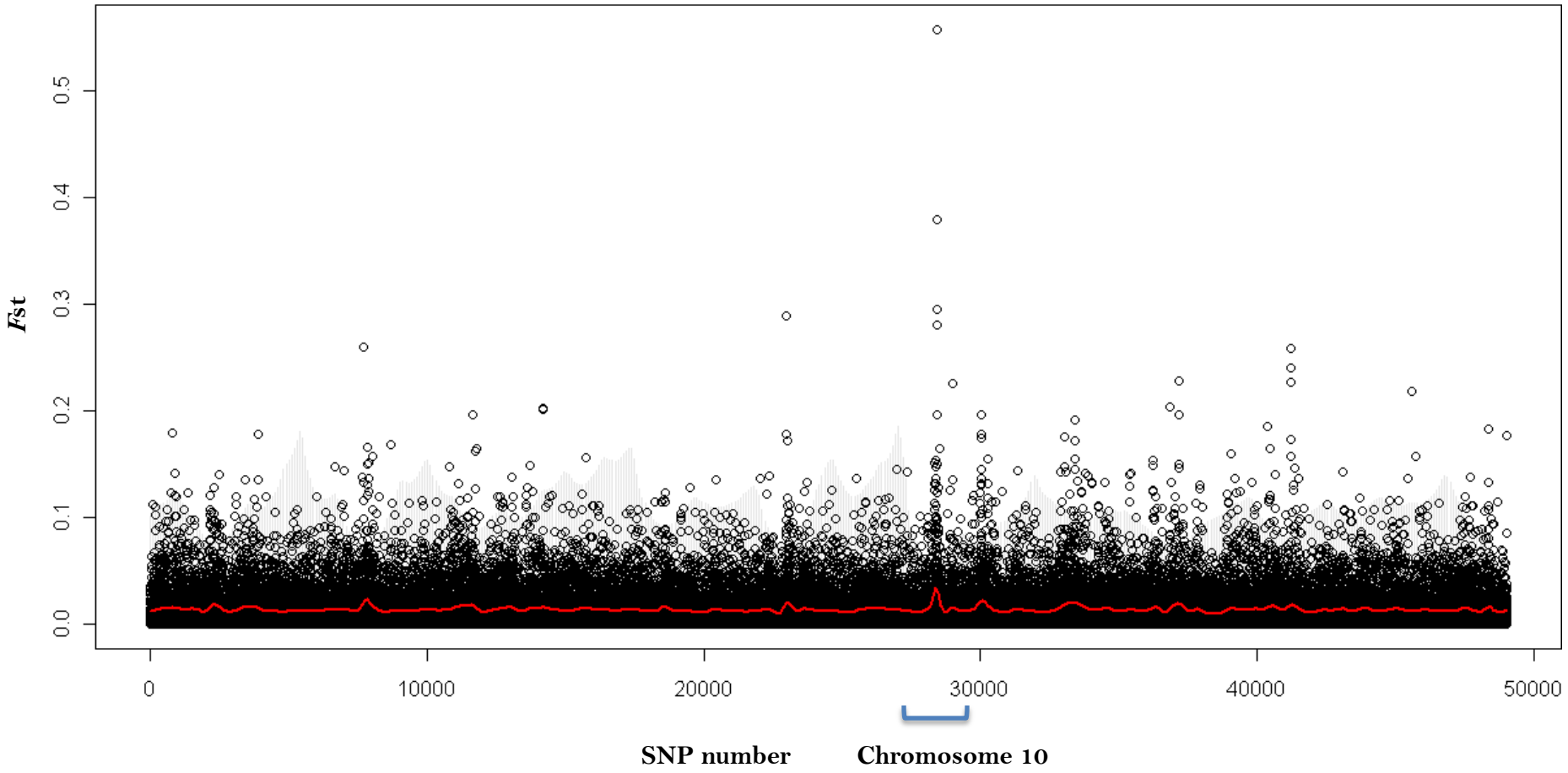
Texel

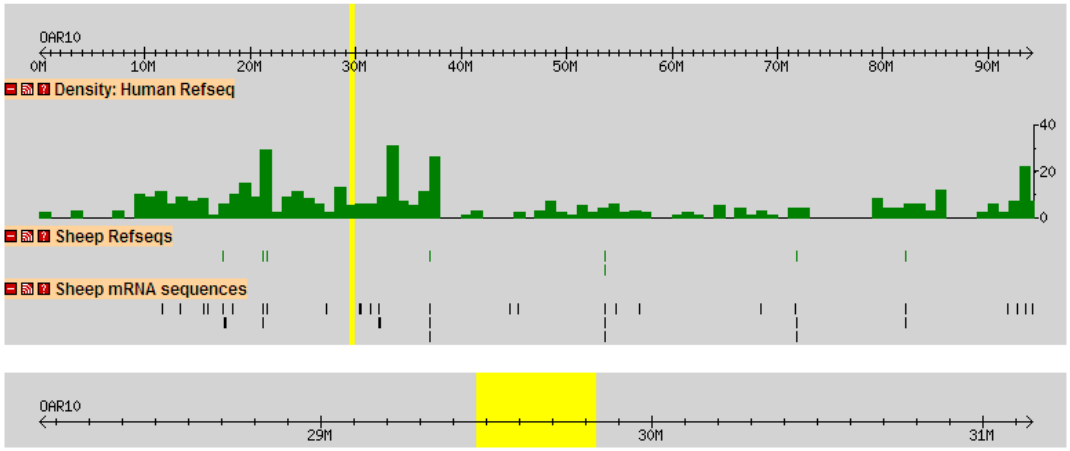


Poll Merino

Australian Poll Dorset

Genome Wide Scan for extreme *F*st in *Polled* animals

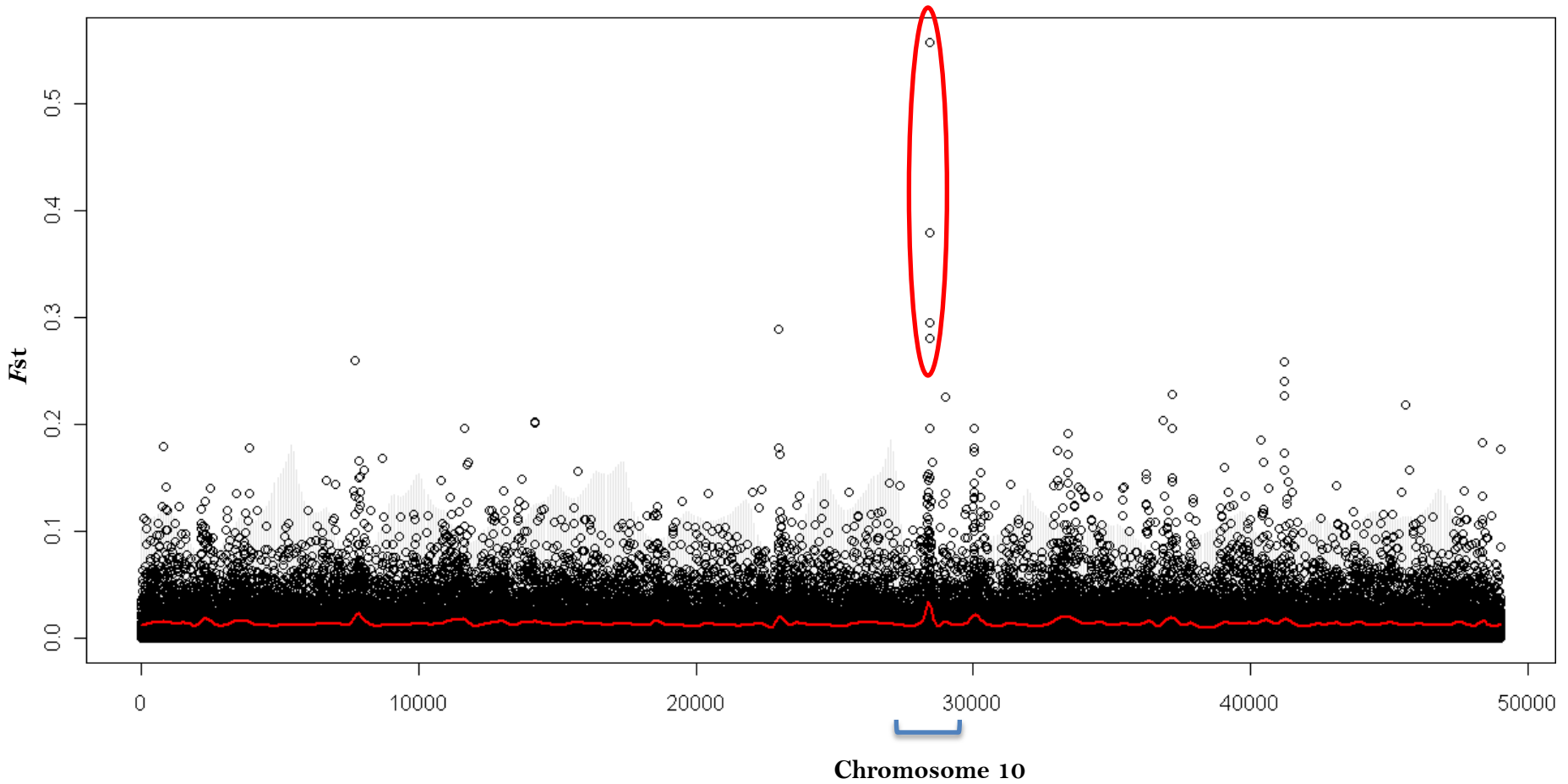


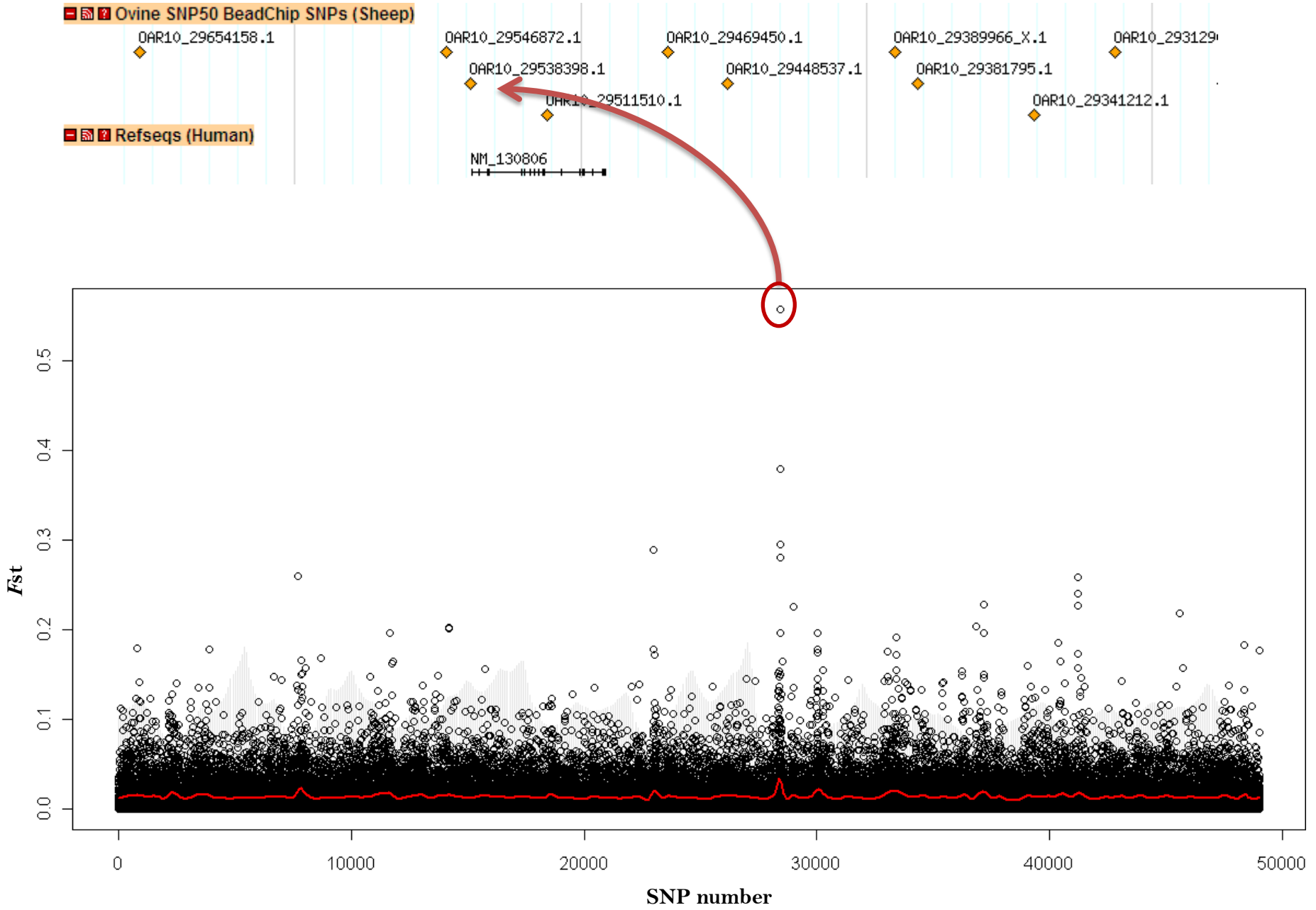


350 kB region
Sheep chr 10

1 gene within the interval

gene known to cause *Poll*
(McEwan et al., unpublished)




Ovine SNP50 BeadChip SNPs (Sheep)
OAR10_29654158.1
OAR10_29546872.1
OAR10_29469450.1
OAR10_29389966_X.1
OAR10_29538398.1
OAR10_29448537.1
OAR10_29381795.1
OAR10_29511510.1
OAR10_29341212.1
Refseqs (Human)
NM_130806
Fst
0.0
0.1
0.2
0.3
0.4
0.5
0
10000
20000
30000
40000
50000
SNP number

Case 2: The Horn Locus (*Ho*)

- primary signal narrow
- multiple breeds
- likely reflects an older mutation under selection for longer compared to case 1

- outlier scan alone took us directly to the causal gene

Additional Traits under investigation

- resistance to intestinal parasites
- meat versus milk production
- thin versus fat tail sheep
- wool type
- pigmentation



Additional Approaches

- iHS
- ROH
- breed specific *F*st
- browser for genome wide comparison of metrics

Acknowledgements

Brian Dalrymple
Wes Barris
David Townley
Sean McWilliam
Russell McCulloch

John McEwan
Rudi Brauning
Theresa Wilson
Alan McCulloch

Marylinn Munson
Kim Gietzen
Christian Haudenschild

Peter Wilson
Annette McGrath

Richard Gibbs
Donna Muzny

Jo-Ann Stanton

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……

All of the Project Participants who contributed Samples and Supporting Funds

Carole Moreno	INRA
Cord Drogemuller	Institute of Genetics, University of Berne
Cyril	Roberts University of the West Indies
David Machugh	University College Dublin
David Machugh	University College Dublin
Despoina Miltiadou	Cyprus University of Technology
Elena Ciani	University of Bari
Elisha Gootwine	The Volcani Center
Fabio Pila	Università degli Studi del Molise
Faruque Mdomar	Bangladesh Agriculture University
Georg Erhardt	Universität Gießen
Georgios Banos	Aristotle University of Thessaloniki
Han Jialin	ILRI and CAAS
Henner Simianer	University of Goettingen
Herman Raadsma	University of Sydney
Ibrahim Cemal	Adnan Menderes University

ISGC

James Kijas	CSIRO
JJ Arranz	Universidad de Leon
John McEwan	AgResearch
Jorge Calvo	CITA Spain
Jorn Benenwitz	University of Hohenheim
Josephine Pemberton	University of Edinburgh
Juha Kantanen	MTT Agrifood Research Finland
Lutz Bunger	Scottish Agricultural College
Matthew Kent	CiGene
Mikka Tapio	ILRI
Olivier Hanotte	formally ILRI
Noelle Cockett / Jim Miller	University of Utah
Ottmar Distl	University of Veterinary Medicine Hannover
Runlin Ma	Chinese Academy of Science
Samuel Paiva	Embrapa Recursos Genéticos e Biotecnologia
Steve Bishop	Roslin
Tiziana Sechi / Antonello Carta	AGRIS Sardegna
Vidya Gupta	NCL India

ISGC

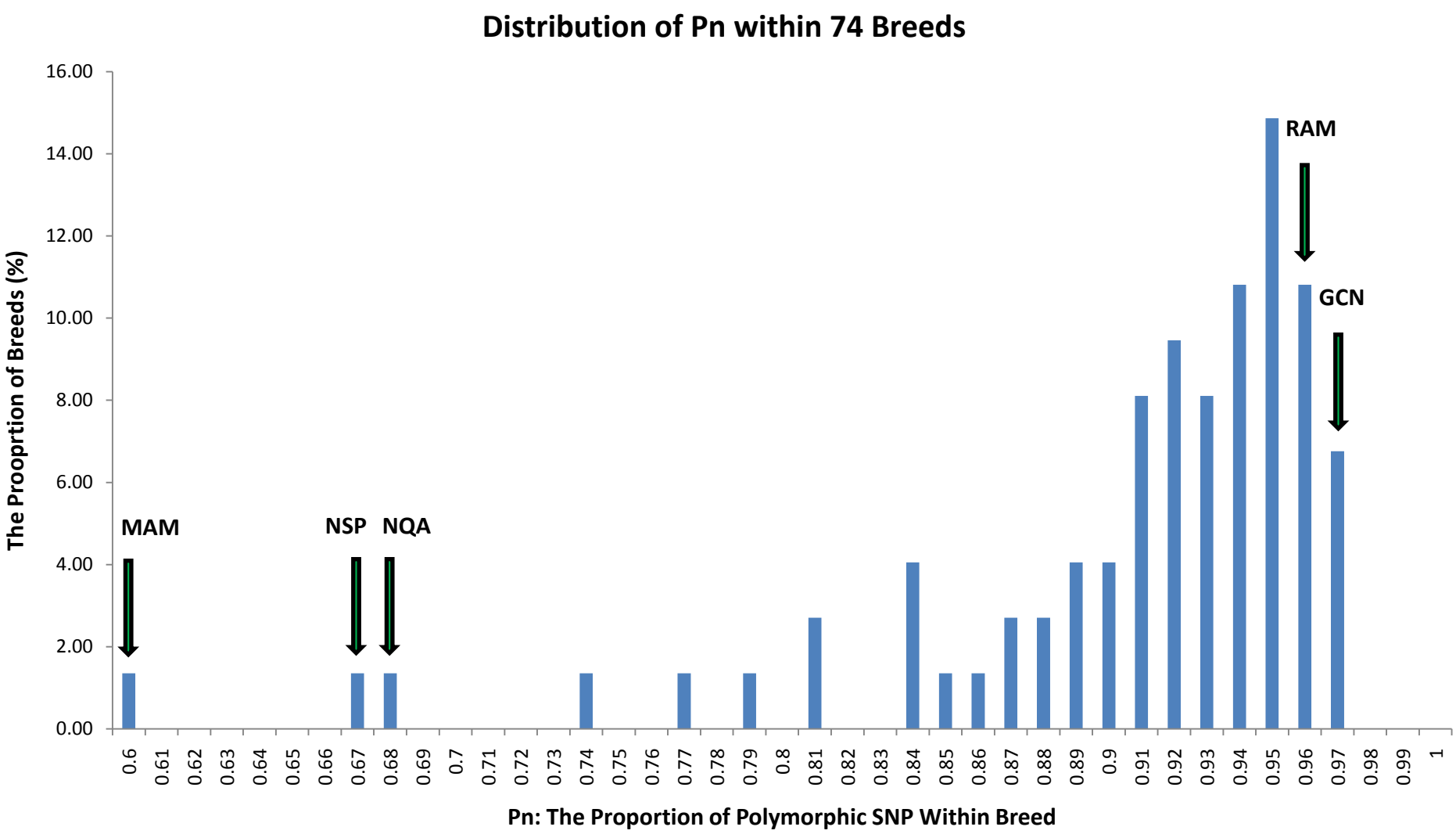
Summary of SNP Pruning

Total SNP **59,454**

- 5,322 failed the manufacture process
- 5,098 pruned during QC using HapMap animals

			Overlap Between SNP Sets				
Filter	**Criteria**	**SNP**	**1**	**2**	**3**	**4**	**5**
1	< 0.99 Call Rate	3612		3244 (89%)	433 (12%)	62 (2%)	4 (<1%)
2	Assay Abnormality	4101	3244 (79%)		433 (10%)	114 (3%)	4 (<1%)
3	MAF < 0.01	1165	433 (37%)	433 (37%)		0	0
4	Discordant Genotypes	125	62 (50%)	114 (91%)	0		0
5	Inheritance Problems	11	4 (37%)	4 (37%)	0	0	
All	Any filter	5098					

Total SNP remaining **49,034**


Distribution of Pn within 74 Breeds
The Prooprtion of Breeds (%)
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
MAM
NSP
NQA
RAM
GCN
0.6
0.61
0.62
0.63
0.64
0.65
0.66
0.67
0.68
0.69
0.7
0.71
0.72
0.73
0.74
0.75
0.76
0.77
0.78
0.79
0.8
0.81
0.82
0.83
0.84
0.85
0.86
0.87
0.88
0.89
0.9
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1
Pn: The Proportion of Polymorphic SNP Within Breed