International Sheep Genomics Consortium

www.sheephapmap.org

Physical Mapping and the Genome Assembly of Sheep

Noelle Cockett[1], Chunhua Wu[1], Jill Maddox[2], Thomas Faraut[3], Bertrand Servin[3],

and the ISGC

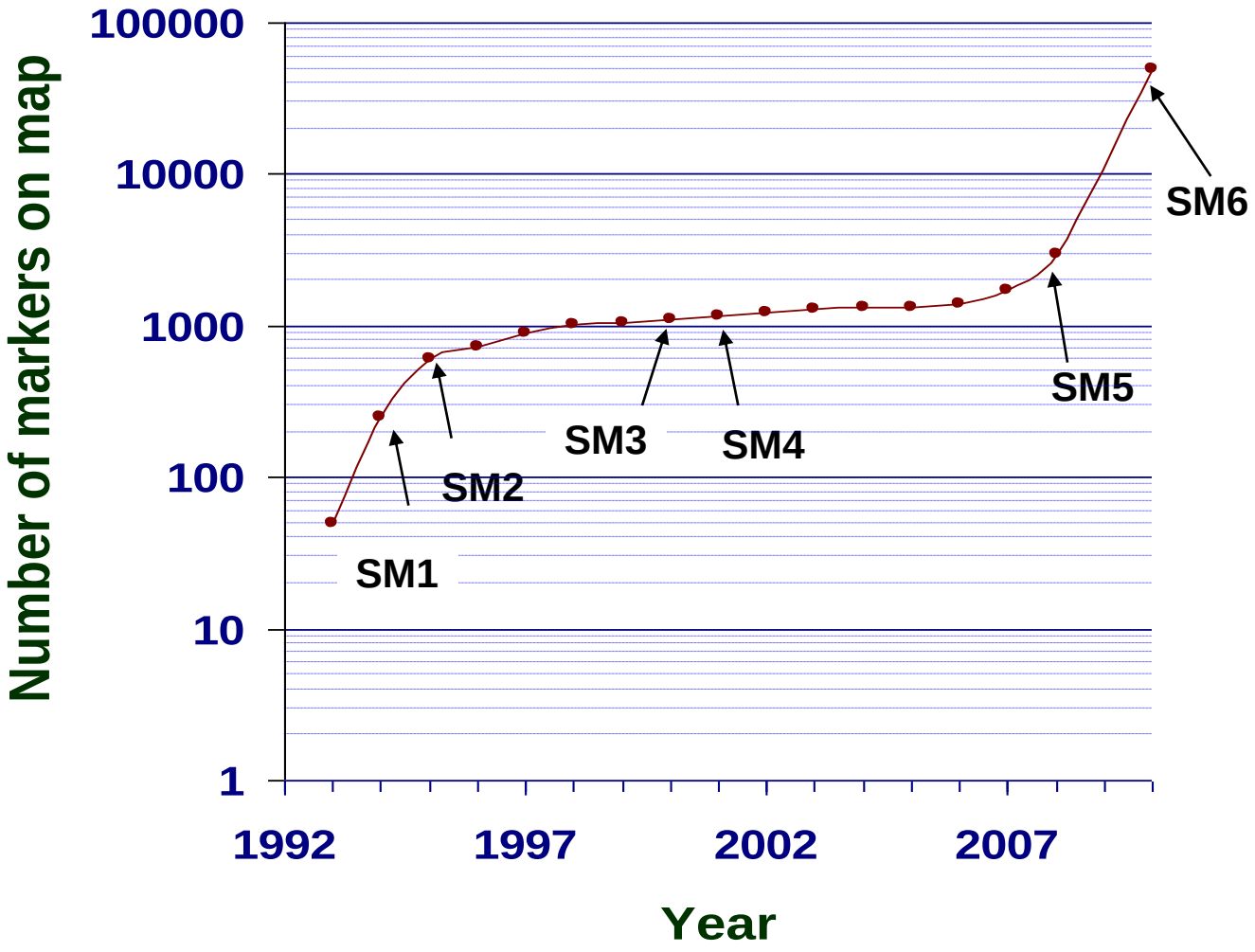
[1]Utah State University, [2]University of Melbourne, [3]INRA Toulouse

Outline of Today's Talk

- Linkage map – preSNP50
- Linkage map – SNP50
- RH map – preSNP50
- RH map – SNP50
- Linkage/RH maps assist the sheep genome assembly

ISGC

Sheep Linkage Map – preSNP50

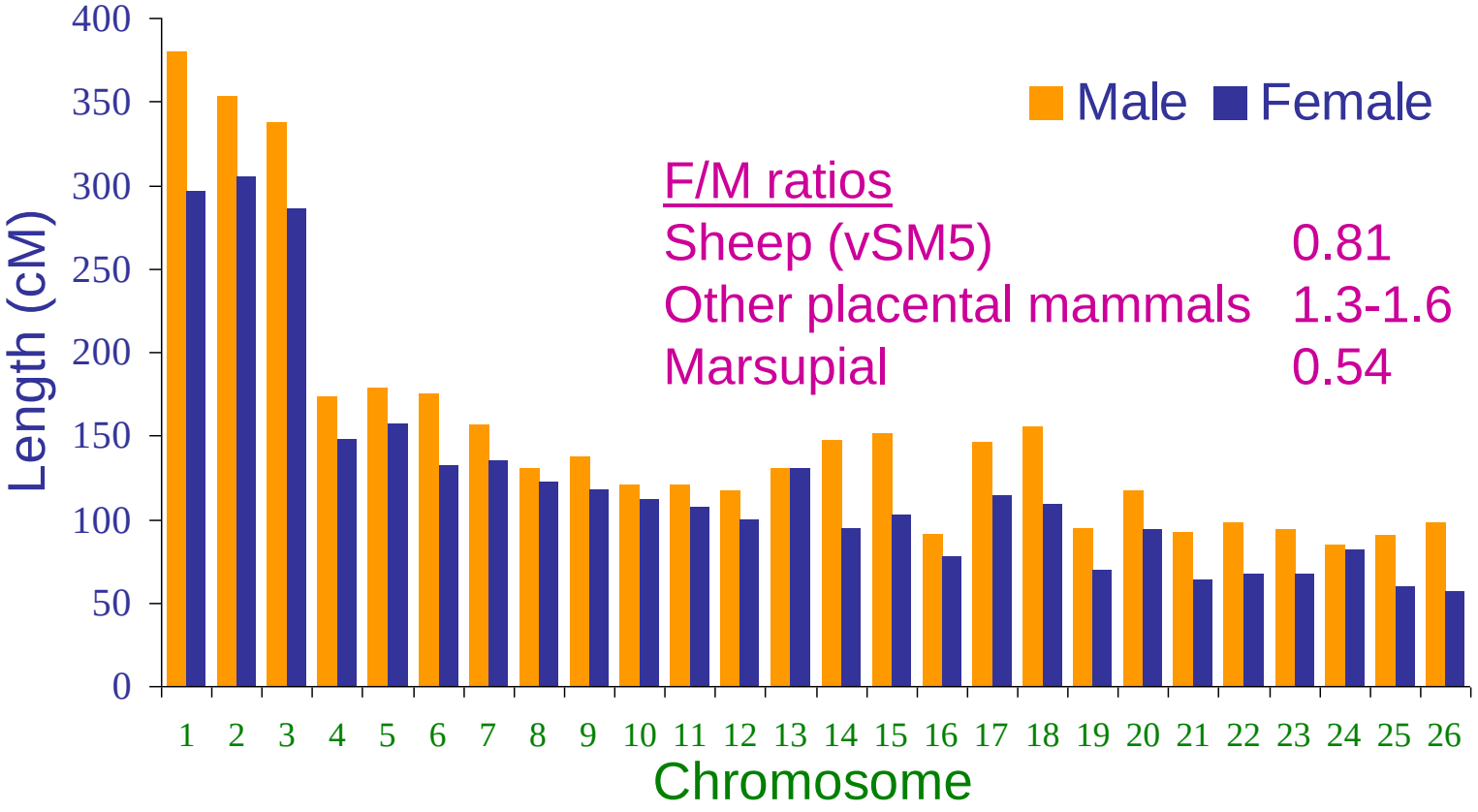


Sheep Linkage Map – vSM5

- Constructed with the International Mapping Flock (IMF)
- Sex averaged, spans 3,800 cM
- 2,528 loci representing 1,420 unique locations
- Mixed microsatellite and SNP map
- 1,100 loci from the ISGC pilot 1.5K sheep SNP chip

ISGC

Comparison of Male and Female Chromosome Length (vSM5)



Male map: 3,964 cM Female map: 3,202 cM

Sheep Linkage Map – SNP50 Analysis

- Pedigrees genotyped with the Illumina SNP50 Ovine BeadChip:
  - International Mapping Flock (IMF)
  - Falkiner Memorial Field Station (FMFS)
  - USDA – WSU
  - USU – LSU
- Software:
  - Modified version of CRI-MAP (Ian Evans and Jill Maddox)

ISGC

Sheep Linkage Map – SNP50 Populations

- International Mapping Flock (IMF)
  - Produced by AgResearch
  - 3 generation, 9 full-sib families, single common grandsire
  - 127 animals genotyped (sires, dams, and offspring)
  - Suitable for low density map on its own but adds microsatellites, etc. from previous linkage maps
- Falkiner Memorial Field Station (FMFS)
  - Produced by sheepGENOMICS
  - 2 generation, 20 half-sib families, industry sires, multiple breeds
  - 4,058 animals genotyped (sires and offspring)

ISGC

Sheep Linkage Map – SNP50 Populations

- USDA – WSU
  - Animals from USDA, ARS Dubois, ID Sheep Station
  - 3 generation, 11 half-sib families, Rambouillet, Polypay, and Columbia breeds
  - 2,211 animals genotyped (sires, grandsires, some dams, and offspring)
  - Resource for X chromosome non-pseudoautosomal region
- USU – LSU
  - Louisiana State University parasite resistance flock
  - 3 generation, 5 F2 families, Suffolk x Gulf Coast Native
  - 503 animals genotyped (grandparents, parents, and offspring)
  - Resource for X chromosome non-pseudoautosomal region

ISGC

Sheep Linkage Map – SNP50 Results

- Currently only framework maps are constructed
- Using FMFS, maps for all chromosomes
  - 3,203 SNPs assigned positions spanning 3,331 cM
  - Lod > 6: chromosomes 9-11, 13, 14, 16-24
  - Lod > 3: X chromosome PAR
  - 38 discrepancies with cli6 genome sequence assembly (chromosome mismatches, missing SNPs, inversions)
- Using IMF, map for the X chromosome
  - Lod > 5
  - 4 discrepancies with cli6 genome sequence assembly (order)

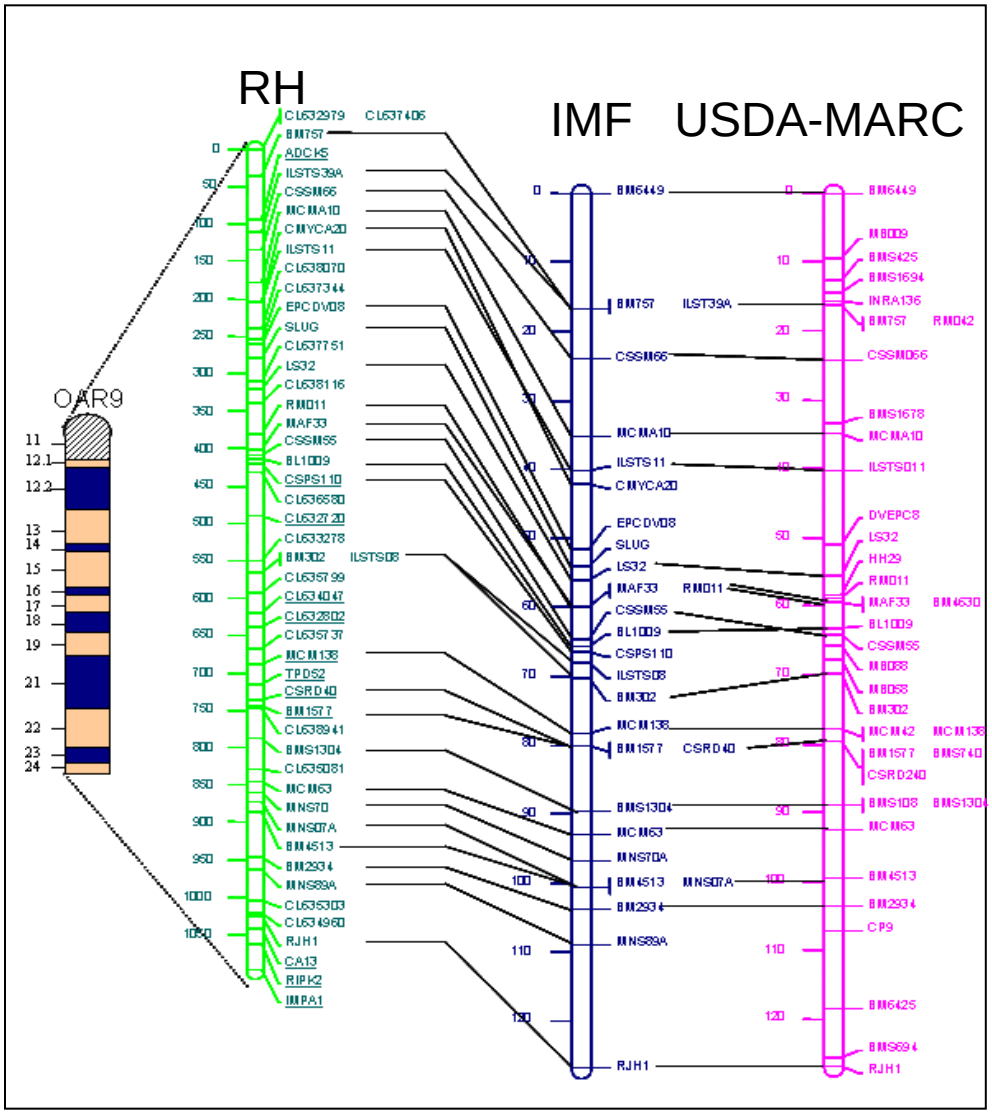
Sheep Linkage Map – SNP50
Future Work

- CRI-MAP needs modifications to handle the large amount of data from combined populations
- More sheep families are needed
- Desired pedigree structures:
  - half-sibling families with at least 20 offspring per sire and genotypes of sire's parents
  - full-sibling families with both parents and grandparents genotyped
- Contact Jill Maddox (jillian.maddox@alumni.unimelb.edu.au)

ISGC

Sheep RH Map – preSNP50

- USUoRH$_{5,000}$
- INRA-SheepRH$_{12,000}$
- 3567 loci typed on USU panel, 2754 loci assigned to positions (W*u et al. in prep)*
- High density maps for all but OARY



Sheep RH Map – SNP50 Analysis

- RH panels typed with the Illumina SNP50 Ovine BeadChip:
  - $USUoRH_{5,000}$
  - $INRA\text{-}SheepRH_{12,000}$
- Method used for map construction:
  - Comparative mapping approach using a draft assembly order (VSG or cli6) as *a prior*
  - Construct a distribution of maps
  - Construct a "robust" map
  - Identify inconsistencies with the assembly

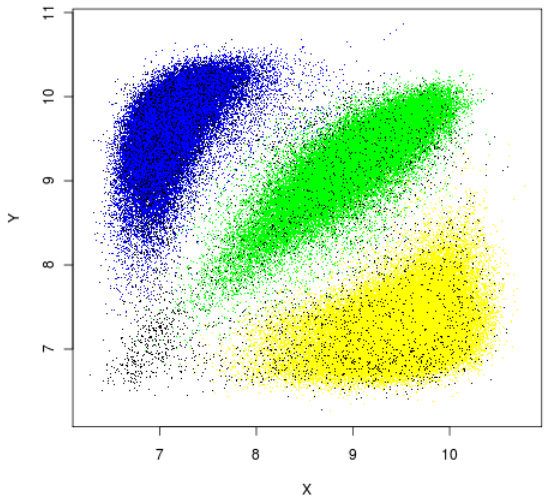
Servin and Faraut (2010)

ISGC

Sheep RH Map – SNP50

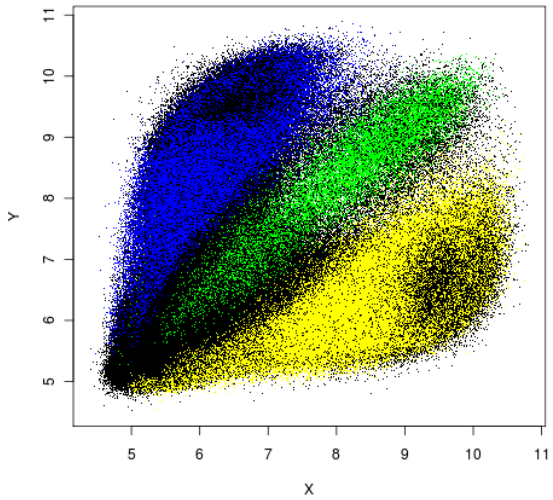
Calling SNP genotypes on a RH panel differs significantly from calling SNPs on genomic DNA

Genomic DNA



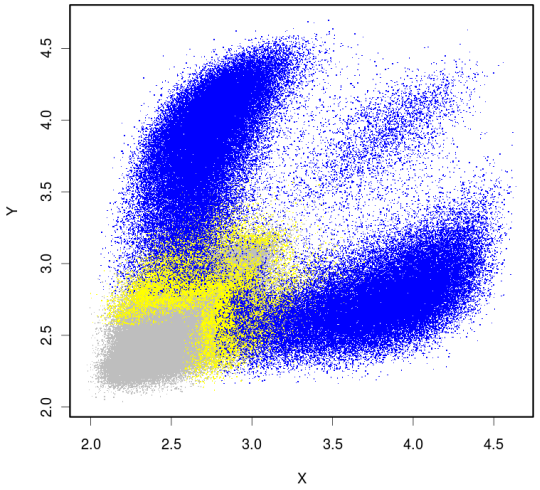
Radiation hybrids

Traditional algorithm



Radiation hybrids

Dedicated algorithm

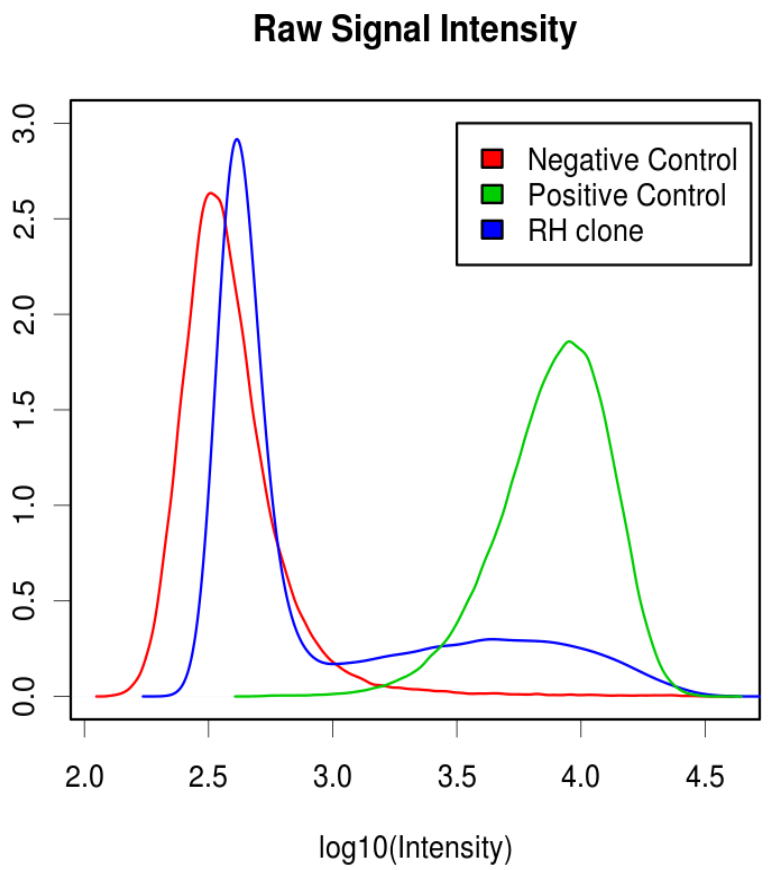


Homozygotes BB
Heterozygotes AB
Homozygotes AA

Present
Absent
Unknown

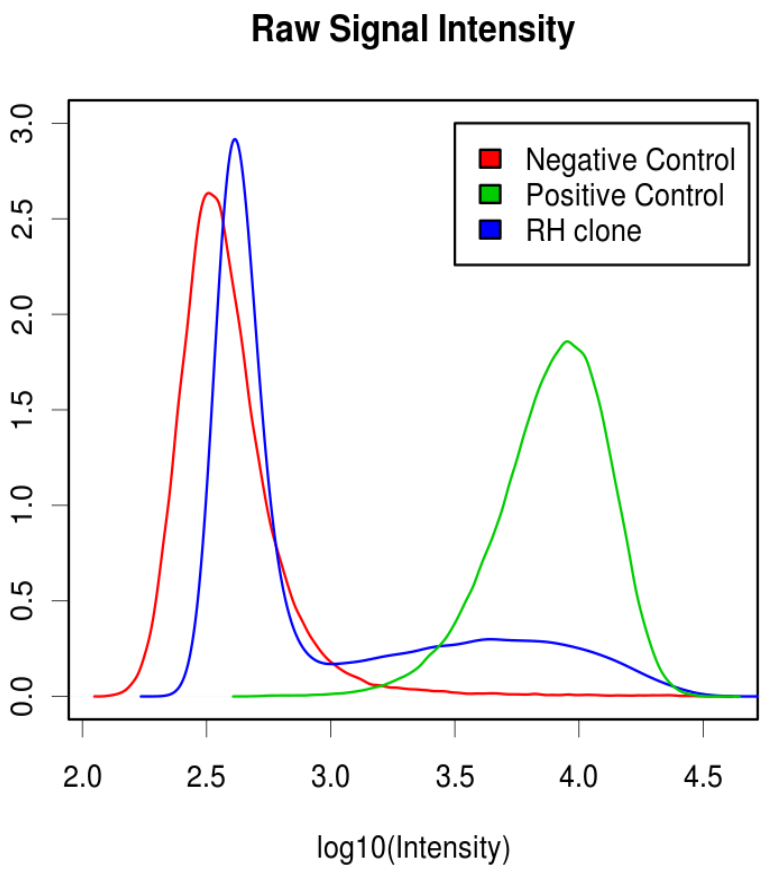
Sheep RH Map – SNP50

The Principle



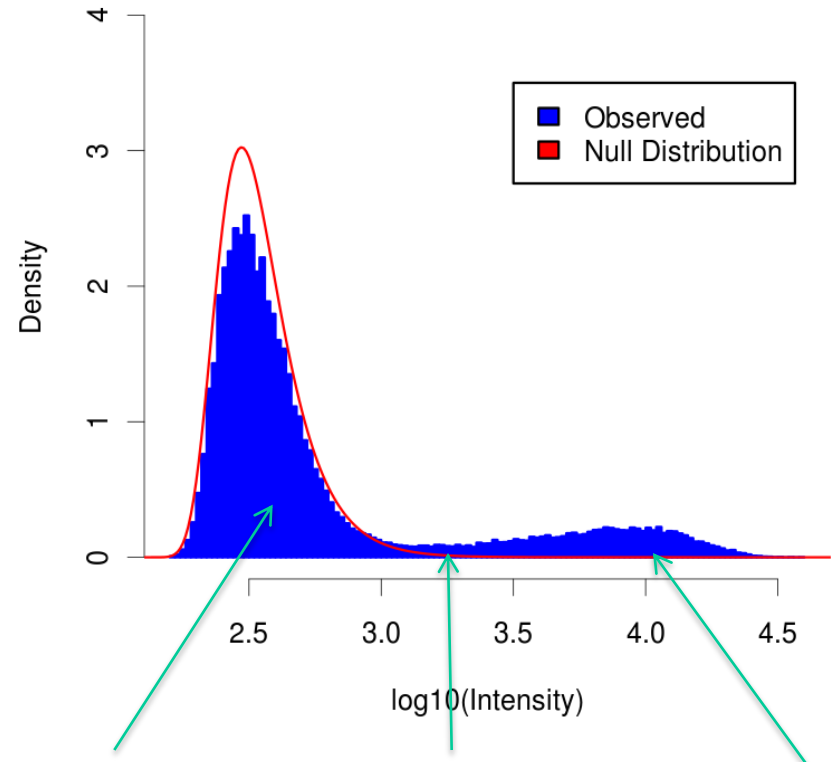
- For each SNP and each clone, the maximum intensity is recorded over the two possible alleles
- The signal distribution is a mixture of two distribution:
  - Absent SNPs (left)
  - Retained SNPs (right)

Sheep RH Map – SNP50
The Problem



- The positive and negative controls do not match the components of the mixture
- Each clone has different proportions of each component
  - Different retention rates per clone

Sheep RH Map – SNP50
The Solution



not retained unknown retained

- Fit a null distribution from the left part of the density
- Estimate a p-value that the observed signal comes from the null for each SNP in each clone
- Correct for multiple testing using FDR

Sheep RH Map – SNP50
Results

- 41,999 SNP markers were successfully called on both panels
- 41,007 SNP markers included in the initial maps
- 32,879 markers included in the robust maps
- 27 RH maps - 1 for each chromosome
  - Only the first 70Mb of the X chromosome

ISGC

Linkage/RH Maps Assist the Sheep Genome Assembly

ISGC

Sheep Genome Sequence Assembly Issues

- Current assembly (cli6) contains many sequence discontinuities
- Current assembly (cli6) is based on 2 animals (4 haplotypes)
- Current approach uses the virtual sheep genome (VSG) to order scaffolds and contigs. The VSG order is based on ovine BAC-end alignments and the bovine genome sequence

ISGC

But a cow is not a sheep!

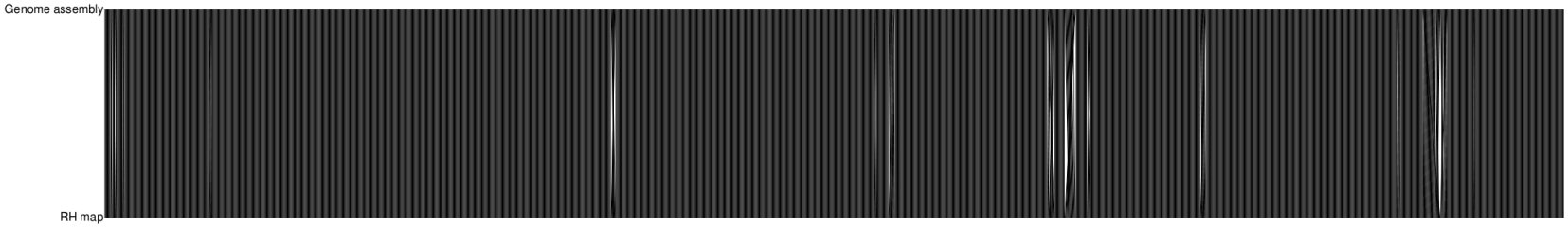
- The sheep has expansions/contractions of gene families
- The sheep and cow have chromosome and order differences
- There are discrepancies between bovine assemblies and all are incomplete

Need additional data to “sheepify” the sheep assembly and to incorporate all contigs

- RH maps
- Linkage maps
- FISH

ISGC

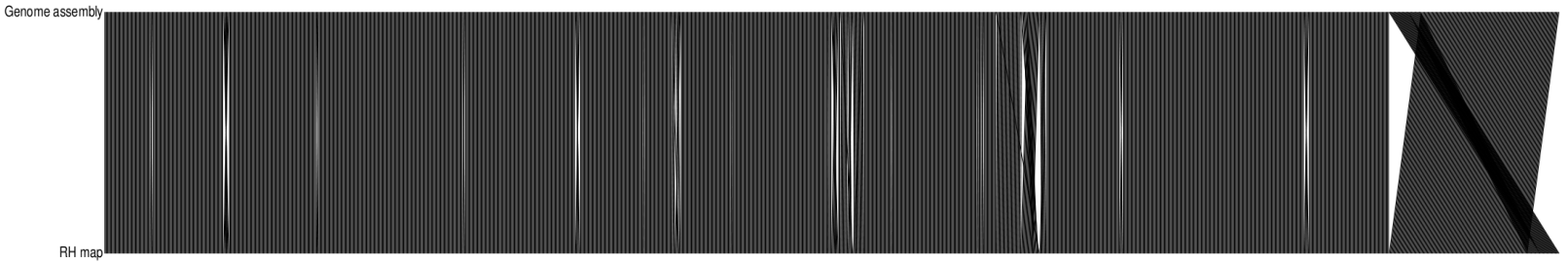
Using the SNP50 RH map to refine the sheep genome assembly (cli6)



Comparison of cli6 assembly and the RH map (OAR4)

- RH map order supports cli6 assembly

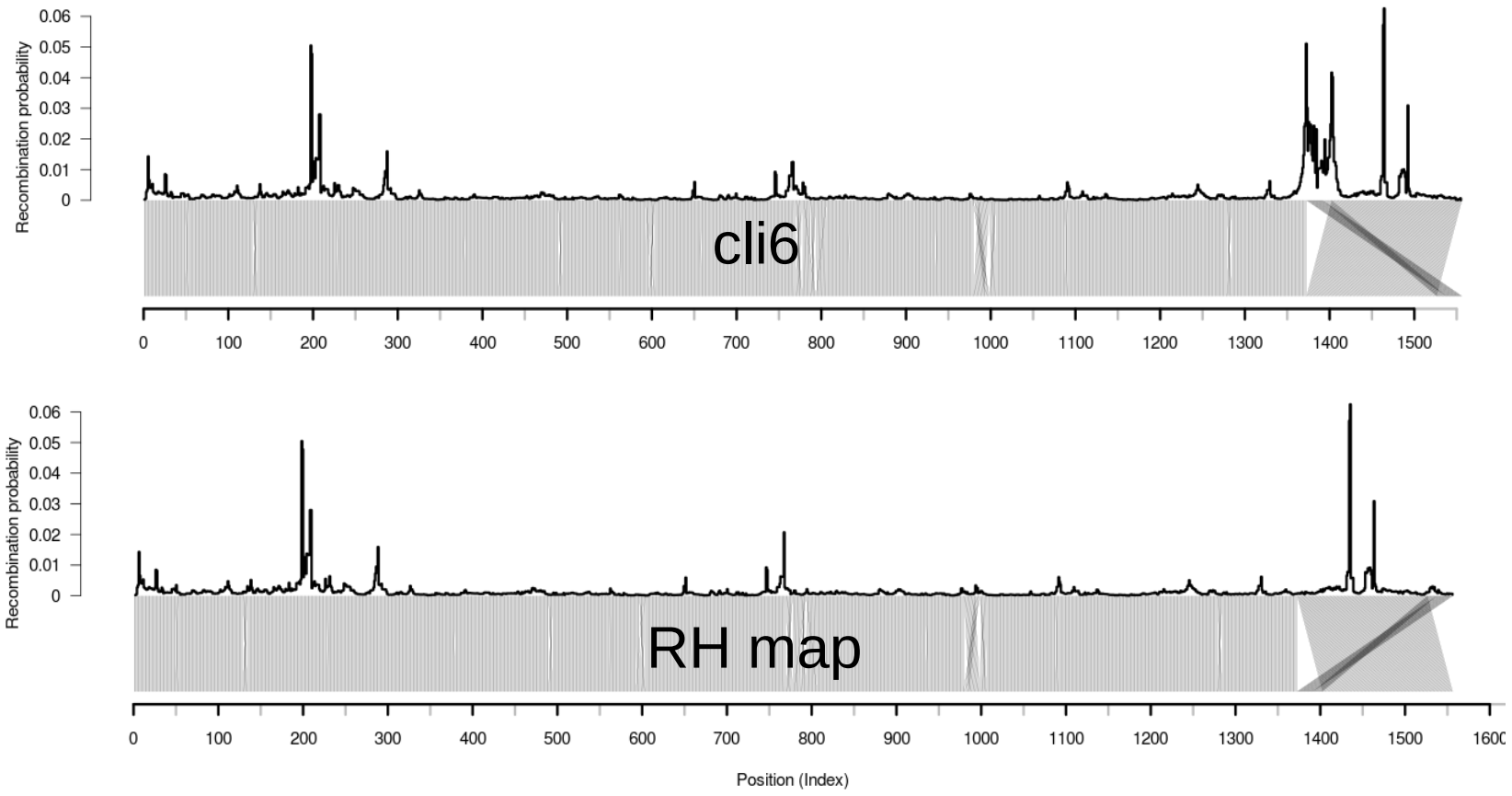
Using the SNP50 RH map to refine the sheep genome assembly (cli6)



Comparison of cli6 assembly and the RH map (OAR6)

- RH map order suggests change in cli6 assembly

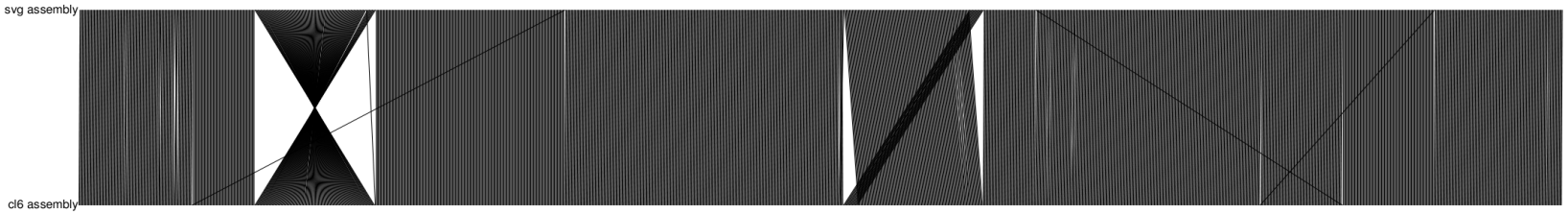
Comparison of recombination frequencies for cli6 order and RH map



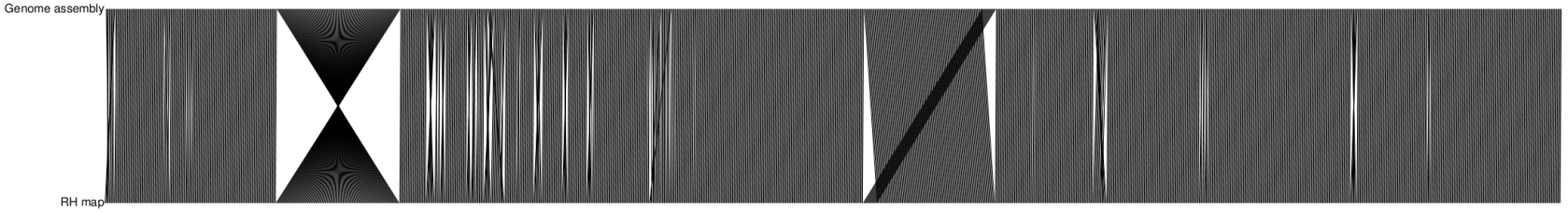
OAR6

Using the SNP50 RH map to refine the sheep genome assembly (cli6)

Comparison of SVG to the cli6 assembly (OAR13)



Comparison of SVG to the RH map (OAR13)



- RH map and cli6 orders in agreement, suggests change in SVG assembly

Using the SNP50 linkage map to refine the sheep genome assembly (cli6)

- Linkage assignment of 49,220 SNPs to chromosome positions
- SNPs missing from assembly
  - 782 SNPs with linkage assignment but no cli6 assignment
  - 5,637 SNPs with cli6 assignment but no linkage assignment
- SNPs incorrectly assigned in assembly
  - 111 (0.2%) discordant between cli6 and linkage map
- SNPs incorrectly ordered in assembly
- SNPs incorrectly duplicated in assembly

ISGC

Conclusions

- The current sheep reference genome assembly is strongly supported by the linkage and RH maps

- Assembly issues are being identified through comparisons of SNP50 positions on the linkage/RH maps

International Sheep Genomics Consortium

- Alan Archibald, Roslin Institute
- Steve Bishop, Roslin Institute
- Noelle Cockett, Utah State University
- Brian Dalrymple, CSIRO
- Thomas Faraut, INRA
- John Gibson, ILRI
- Clare Gill, Texas A&M University
- James Kijas, CSIRO
- Jill Maddox, University of Melbourne
- John McEwan, AgResearch
- Sean McWilliam, CSIRO
- Hutton Oddy, University of New England
- Herman Raadsma, University of Sydney
- Betrand Servin, INRA
- Ross Tellam, CSIRO
- Wen Wang, Kunming Institute of Zoology
- Chris Warkup, Genesis Faraday Partnership
- Jiang Yu, Kunming Institute of Zoology
- Wenguang Zhang, Inner Mongolia Ag Univ

ISGC

International Sheep Genomics Consortium

11:30 a.m. – 3:00 p.m.
Monday, January 17, 2011
Sunset Room in the Meeting House

- Update on Reference Genome Project
- **Reference Genome Project – Annotation Teams and Genome Interpretation**
- Proposed New Re-sequencing Project
- Ordering Illumina SNP50 BeadChip
- New Synthesis and Content of Illumina SNP50 BeadChip
- Proposed 6K SNP Chip
- Proposed High Density SNP Chip
- Parentage SNP Panel
- The Way Forward

ISGC