# Genomic resources for canine cancer research

Jessica Alföldi
(for Kerstin Lindblad-Toh)
Broad Institute of MIT and Harvard

# Dog genome assembly – CanFam 3.1

## Comparison of most recent assembly versions

| | Sequence length (Gb) | Gaps as % of assembly | Scaffold N50 (Mb) | Contig N50 (kb) |
|---|---|---|---|---|
| **Human** | 3.23 | 0.5% | 59.4 | 56,413 |
| **Mouse** | 2.80 | 2.8% | 52.6 | 32,273 |
| **Dog** | 2.41 | 0.8% | 45.9 | 267 |
| **Rat** | 2.87 | 4.9% | 15.0 | 100 |
| **Rhesus** | 3.23 | 3.6% | 3.0 | 54.2 |

- Released in 2011 – 6 years of work
- 85 Mb added to genome assembly
    - – Went from 99.2 -> 99.6% euchromatic coverage
- Closed 1,044 gaps in promoters/first exons

# RNA-seq data for annotation

- 10 beagle tissues
  - blood, brain, heart, kidney, liver, lung, ovary, skeletal muscle, skin, and testis
  - Almost all single adult female beagle
- 3 sets of libraries
  - Strand-specific polyA isolated (mRNA)
  - DSN (mRNA + other transcripts)
  - microRNA

# Gene annotations - Ensembl

- Used our polyA and DSN RNA-seq for gene prediction/confirmation
- Very high quality – very conservative
- 19,856 coding genes, 3774 non-coding genes
- Orthology information: easily find equivalents in human and other genomes

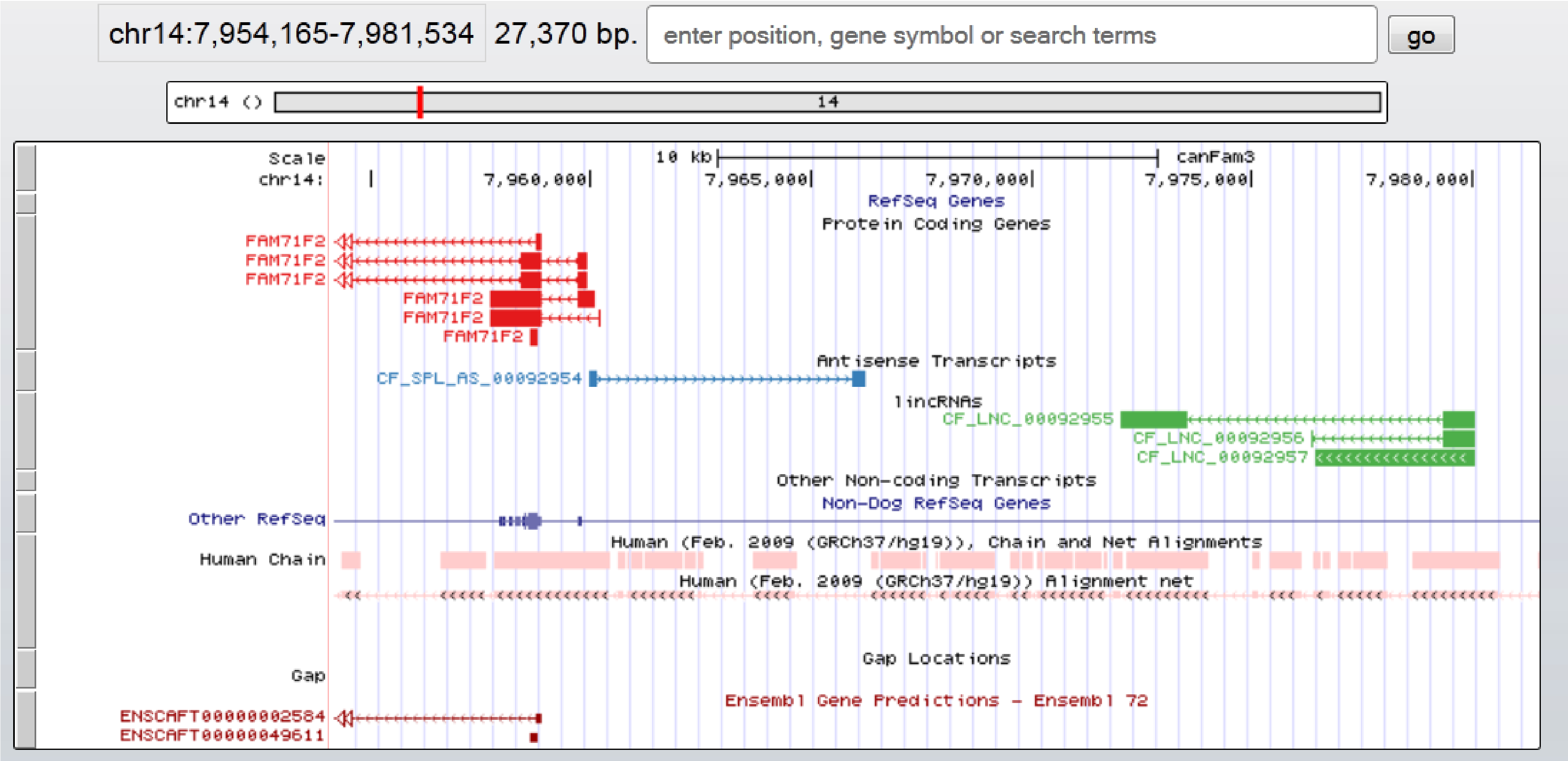
# Broad CanFam 3.1 annotation

- 20,657 high-confidence protein-coding genes
  - Based on Ensembl gene models + RNA-seq
  - 4.7 alternative transcripts/locus
- 2,141 low-confidence protein-coding genes
  - Ensembl (no RNA-seq confirmation)
  - Possible pseudogenes
  - Or only expressed in tissues we don’t have
    - Developmental
    - Specific brain regions

# New CanFam 3.1 annotation

- 5,301 Antisense transcripts (spliced + unspliced)
  - Overlap with protein-coding exon, on other strand
- 7,224 Long intergenic non-coding (lincRNAs)
  - No ORF, spliced, no overlap with pc genes
- >100K other transcripts (67% intronic)
  - 41,604 other intergenic non-coding transcripts

# New Annotation

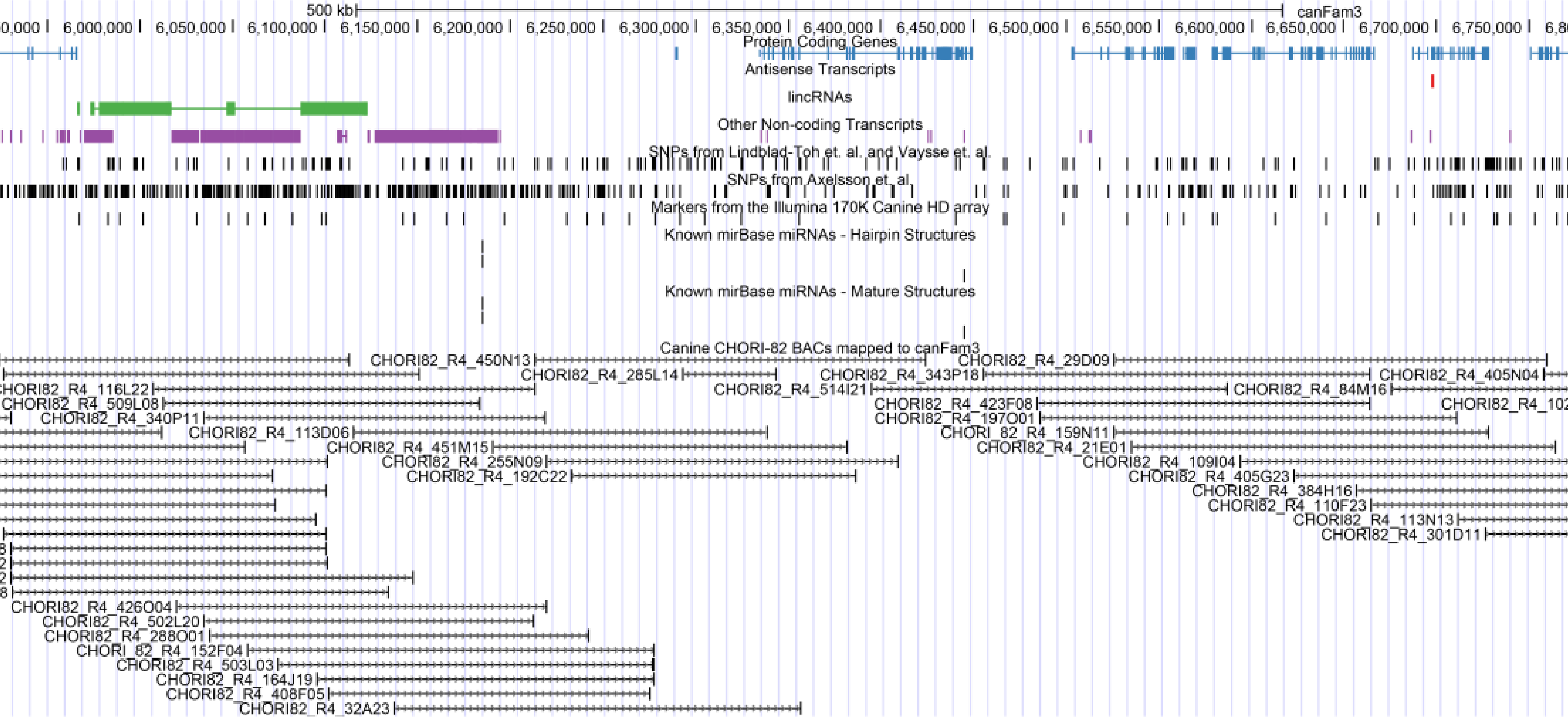


- Many spliceforms
- Spliced antisense transcripts and lincRNAs
- Antisense transcripts overlap protein-coding genes

# Broad – UCSC Canine Track Hub

- Data hosted by Broad integrated into the UCSC Genome Browser
- Annotation, RNA-seq data
- SNP sets including 170K chip
- CHORI-82 BAC ends
- Do you have data you'd like to have hosted?

# Broad-UCSC Canine Track Hub



# Illumina 170K Genotyping Chip

- Currently most high-density chip
- 175K SNPs evenly covered across genome
  - Including special efforts for regions >40kb without known SNPs
  - Wide variety of dog breeds + wolves
  - 99.2% of SNPs give reliable data
  - Mean spacing of 13 kb, only 21 gaps >200kb

# Osteosarcoma (OSA) in 3 breeds

- Most common bone cancer in dogs (~ 10,000 cases / year in USA )
- Highly metastatic, very high mortality
- Looks like human osteosarcoma (~ 1000 cases / year in USA)

Greyhounds

*~25% OSA*

153 cases

114 controls

*(Very little OSA in show Greyhounds)*

Irish Wolfhounds

*~20% OSA*

28 cases

62 controls

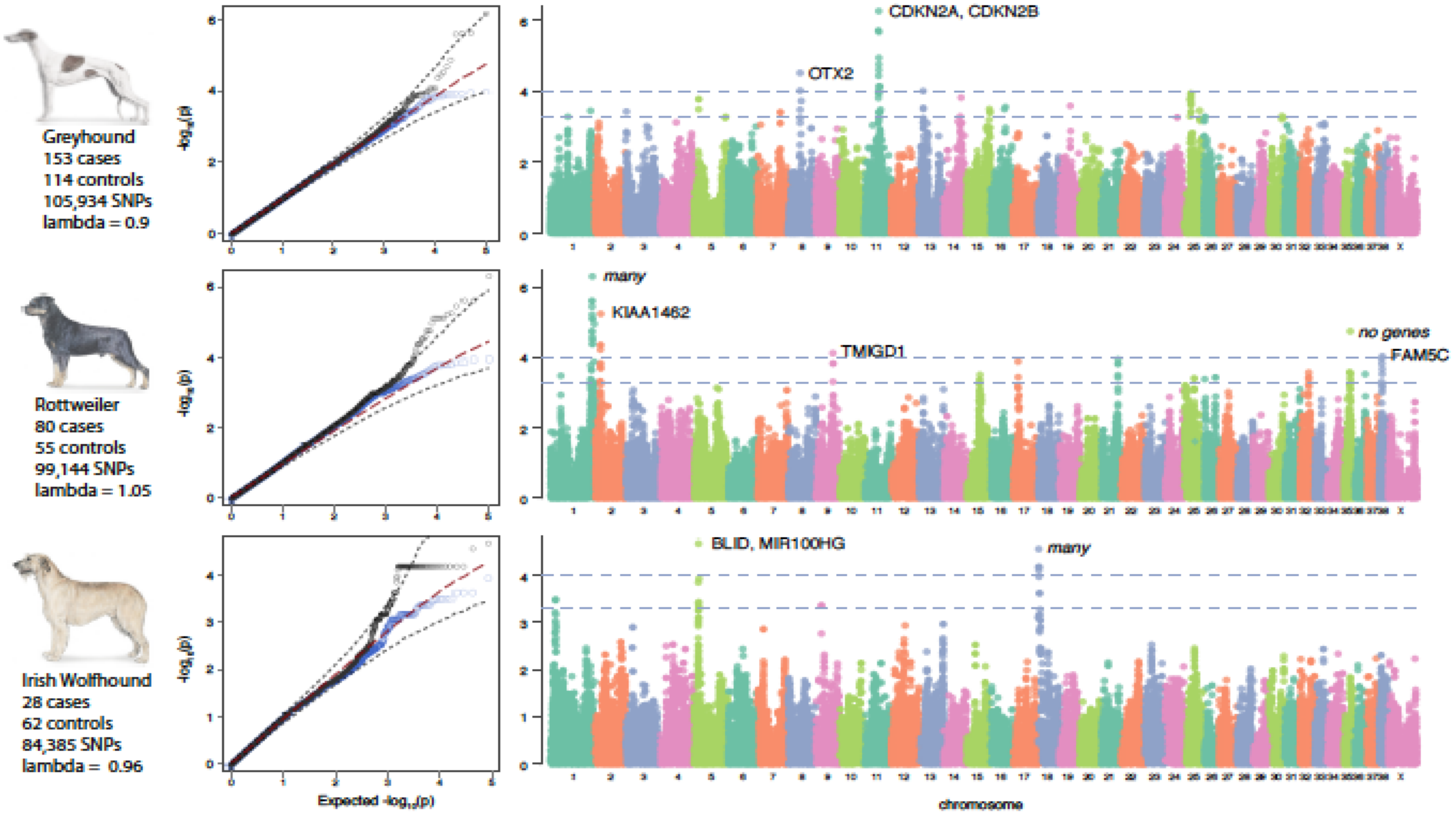
Rottweilers

*~15% OSA*

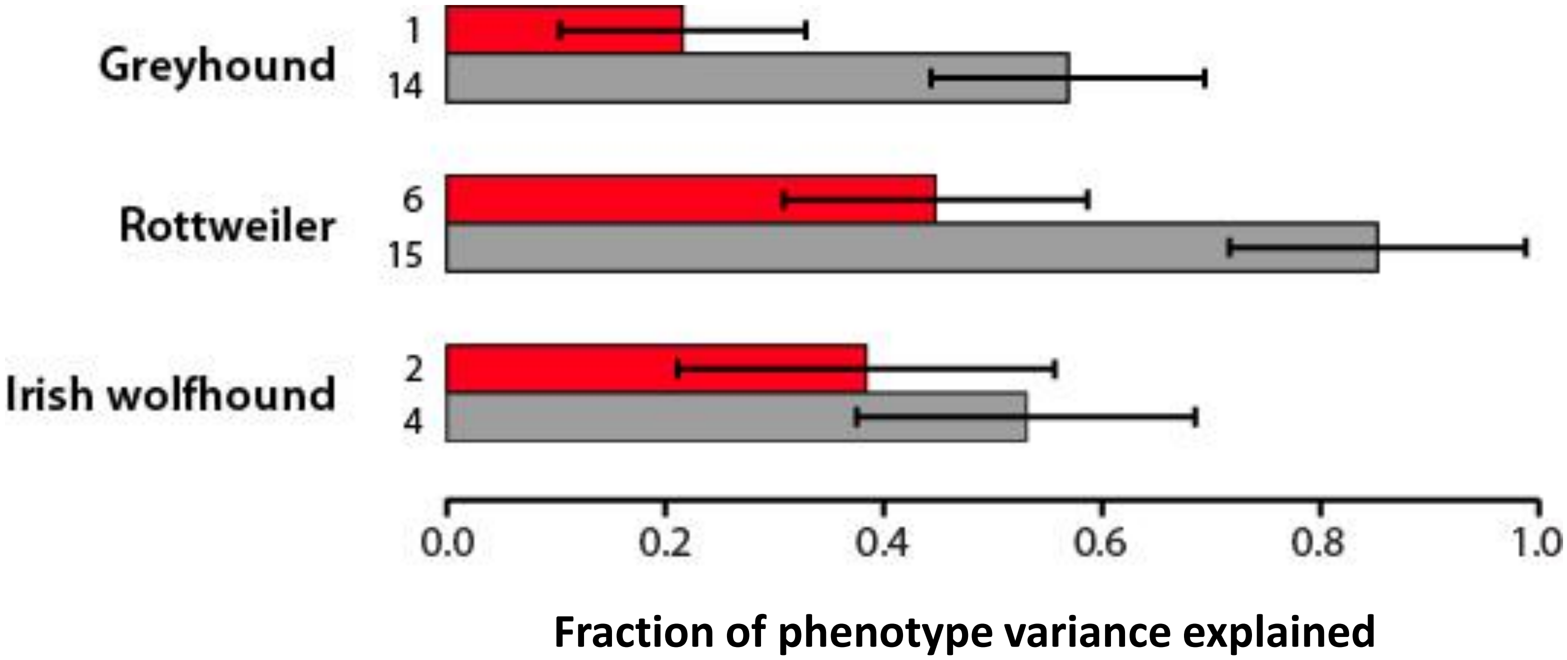
80 cases

55 controls

# Lots of loci (33) – but none shared



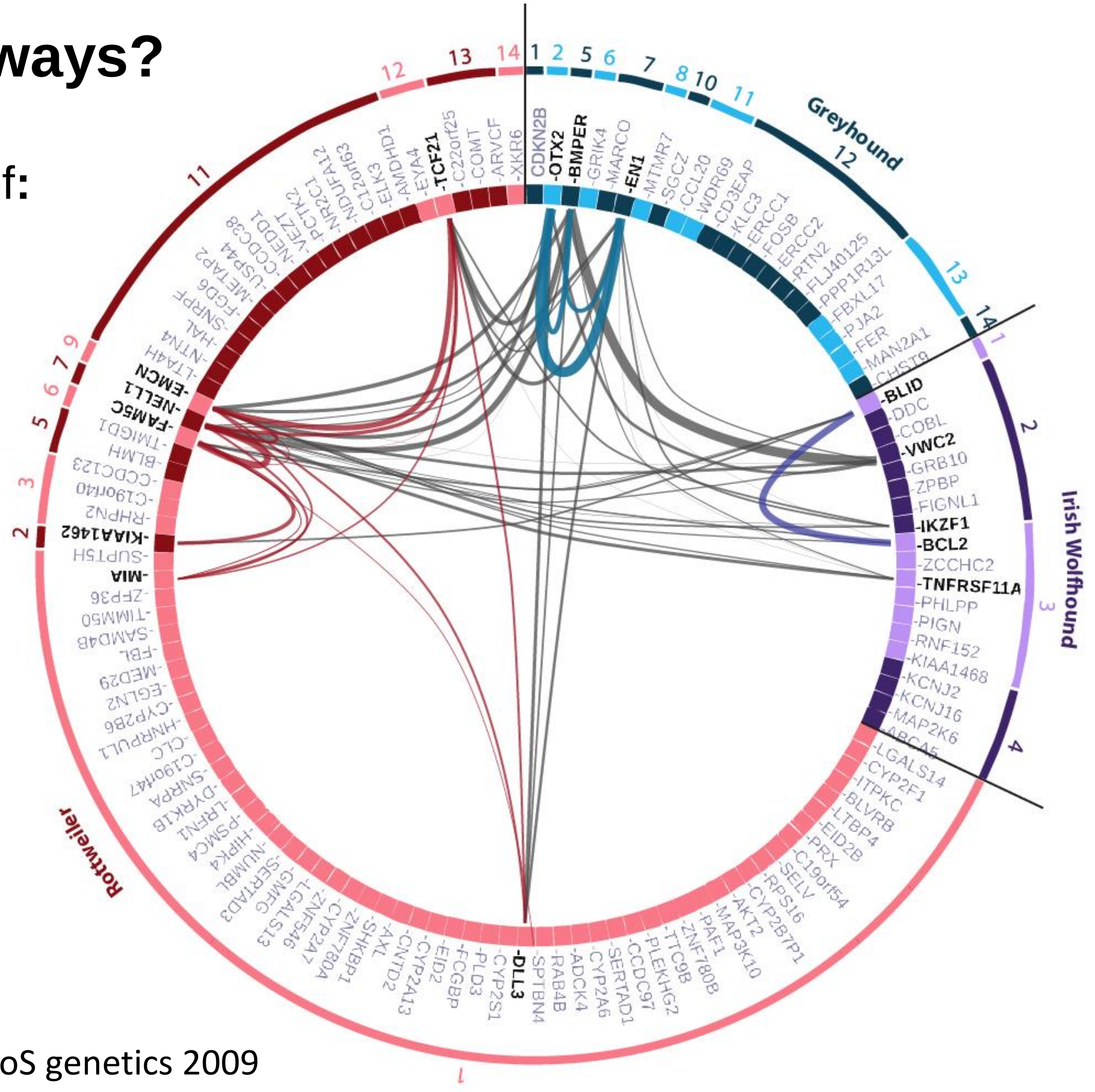
# Phenotype variance explained: high



# Shared pathways?

found regulators of**:**

- osteoblast differentiation, proliferation
- skeletal growth
- mesenchymal-epithelial cell transitions



GRAIL Raychaudhuri PLoS genetics 2009

# Osteosarcoma conclusions

- Need to separate breeds for analysis
- 100 cases/controls yield high heritability
- Risk factors are entirely independent
  - Although in same pathways
- Non-coding mutations predominate
- Pathways align with somatic mutations
  - Breen CGH tumor data

# Canine exome designs

- Several Roche/Nimblegen designs available
- Ensembl based CanFam3.1 currently in use:
    - 49 Mb – protein-coding exons only
- Two new designs to be published shortly
    - Exome plus (based on Broad annotation)
        - 152 Mb – protein-coding and non-coding transcribed
    - Exome CDS (based on Broad annotation)
        - 71 Mb – Exome plus without UTRs

# Somatic mutations in canine lymphoma - methods

| Breed | B-cell pairs | T-cell pairs |
|---|---|---|
| Boxer | - | 16 |
| Cocker | 10 | - |
| Golden | 54 | 25 |

Exome Capture (Roche/Nimblegen)

↓

Illumina Sequencing
N > 60x, T > 100x

↓

Alignment to CanFam3.1 (bwa)

↓

Somatic Variant calling
(muTect, indeLocator)

↓

Variant Filtration

↓

Identifying significantly mutated genes
(MuSiC)

Tools chosen for canine compatibility

*UNPUBLISHED*

# Key genes in B-cell lymphomas

*Fraction of individuals with at least one mutation in a gene/pathway*

| Pathways/genes | All B | Cocker B | Golden B | All T | Boxer T | Golden T | Human Cancer |
|---|---|---|---|---|---|---|---|
| FBXW7 | **0.25** | 0.10 | 0.28 | **0.05** | 0 | 0.08 | Lymphoma |
| FBXW7-Myc | **0.28** | 0.30 | 0.28 | **0.05** | 0 | 0.08 | Lymphoma |
| TRAF3-TRAF2-cIAP1-cIAP2-MAP3K14 | **0.30** | 0.50 | 0.26 | **0** | 0 | 0 | Lymphoma |
| POT1 | **0.17** | 0.20 | 0.17 | **0** | 0 | 0 | Lymphoma |
| DDX3X | **0.11** | 0.30 | 0.07 | **0** | 0 | 0 | Breast |

*UNPUBLISHED*

# Key genes in T-cell lymphomas

*Fraction of individuals with at least one mutation in a gene/pathway*

| Pathways/genes | All B | Cocker B | Golden B | All T | Boxer T | Golden T | Human Cancer |
|---|---|---|---|---|---|---|---|
| SATB1 | **0** | 0 | 0 | **0.17** | 0.25 | 0.12 | Lymphoma |
| PTEN | **0** | 0 | 0 | **0.10** | 0.25 | 0 | Lymphoma |
| PTEN-PI3K-Akt-mTOR | **0** | 0 | 0 | **0.22** | 0.44 | 0.08 | Lymphoma |
| LTA4H | **0.02** | 0 | 0.02 | **0.10** | 0 | 0.27 | Carcinoma |

*UNPUBLISHED*

# Lymphoma somatic mutations - lessons

- Separating by breed is essential
- Underlying risk factors likely result in greater tumor homogeneity within breeds
- Thorough phenotyping of tumors increases power
- Dog breeds can act as models for subtypes of human cancers

*UNPUBLISHED*

# Acknowledgments

Dog genome sequencing consortium
Matthew Breen
Jaime Modiano
Cheryl London
Guillermo Cuoto
Mike Starkey
Kim Cronin
Claire Wade
Kenine E. Comstock
Barb Davis
Kristine Burgess
Lisa Barber
John Berg
Dieter Deforce

NHGRI (NIH)
American Kennel Club
Morris Animal Foundation
European Research Council
Swedish Research Council FORMAS