# Using DNA copy number aberrations to identify candidate drivers of carcinogenesis in naturally occurring canine cancers

## Daniel Rotroff PhD, MSPH

Postdoctoral Research Fellow, Bioinformatics Research Center, North Carolina State University

2nd International Conference on Genomics and Pharmacogenomics

**September 10, 2014**

# Introduction

❑ There are approximately 78 million domestic dogs residing in the USA.

❑ Cancer is one of the leading causes of death for domestic dogs, with popular breeds such as golden retrievers, Labrador retrievers and boxers, succumbing to cancer with frequencies of 50, 34 and 44%, respectively.

❑ Dogs exhibit a wide variety of spontaneous cancers that share clinicopathologic features with humans

Rotroff et al. (2013) Naturally occurring canine cancers: powerful models for stimulating pharmacogenomic advancement in human medicine. *Pharmacogenomics.*

Thomas et al. (2014*)* Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res.*

Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*

# Introduction

❑ The recent development of a high-quality canine genome sequence assembly has opened the door for researchers to identify key drivers of disease that may impact both canine and human patients.

❑ We have developed tumor-associated genomic DNA copy number aberration profiles for 75 canine hemangiosarcomas and more than 200 canine leukemias and lymphosarcomas using an oligonucleotide array comparative genomic hybridization (oaCGH) platform.

❑ We have mapped canine genes to available human homologues for pathway-based analyses, identified putative drivers of carcinogenesis, and have identified genes that may be useful as diagnostic tools for characterizing leukemia subtypes.

Rotroff et al. (2013) Naturally occurring canine cancers: powerful models for stimulating pharmacogenomic advancement in human medicine. *Pharmacogenomics.*

Thomas et al. (2014*)* Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res.*

Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*

# Methods

**Study Features**

❑ 123 leukemias comprised of ALL (28), AML (24), CLL-B (25), and CLL-T (46)

❑ 106 lymphosarcomas comprised of B-cell (74) and T-cell (32)

❑ 75 hemangiosarcomas  from 5 popular dog breeds

❑ ~180,000-feature Agilent technologies microarray design oaCGH platform. Array uses ~60-mer oligonucleotides distributed at approximately 13kb intervals

**Data Processing**

❑ Data was normalized and segmented using CBS.
❑ Gain/Loss/No change calls were made based on a 5 MAD cutoff per subject

**Modeling Approaches**

❑ Hierarchical clustering
❑ Feature formation

$$S_i = \ln\left( 1 + \left( X_i - \sum_{i=1}^{n} X_i \right)^2 \right)$$

❑ Regions were selected based on $S_i <$ 2.5 and were used as features for model development.
❑ A recursive random forest ensemble classification model
❑ Regions with the 100 highest Gini coefficients were used as features in a decision tree classification model
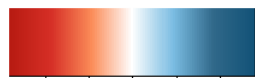
# CBS segmentation

Tumor type: Leukemia | Subtype: ALL-T
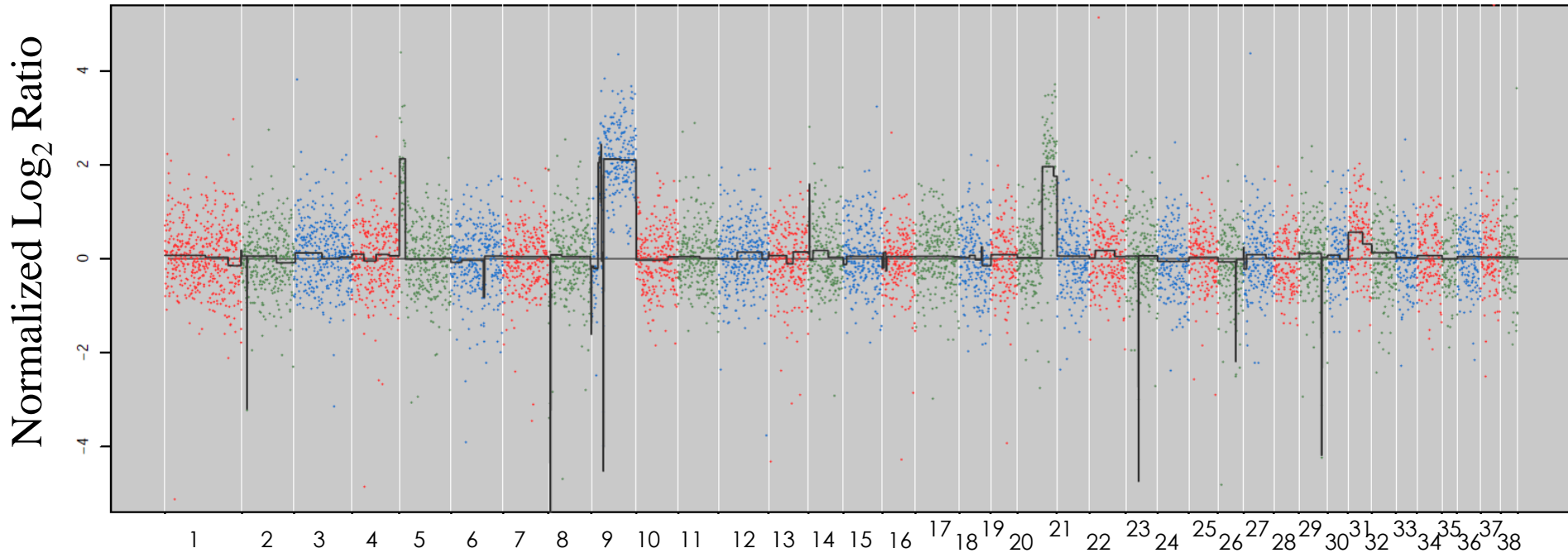
Example of an individual leukemia patient's oaCGH profile. The x-axis contains genomic regions and the y-axis is the $\log_2$ ratio of the normalized fluorescent signal. The solid line represents the results of the CBS segmentation algorithm.
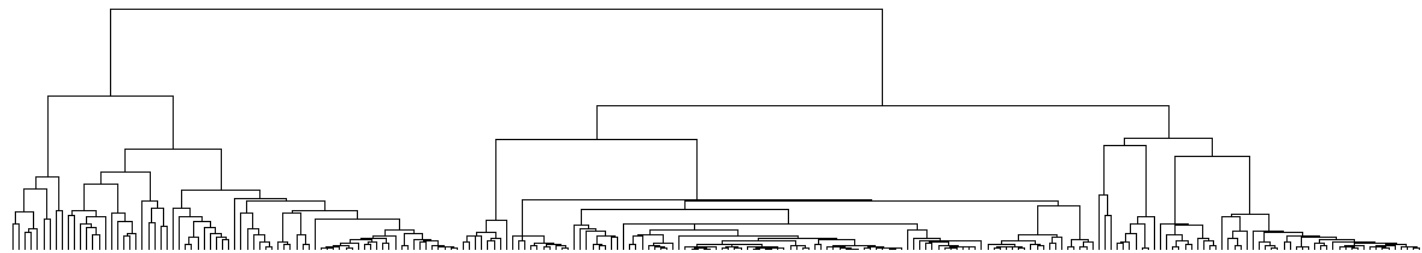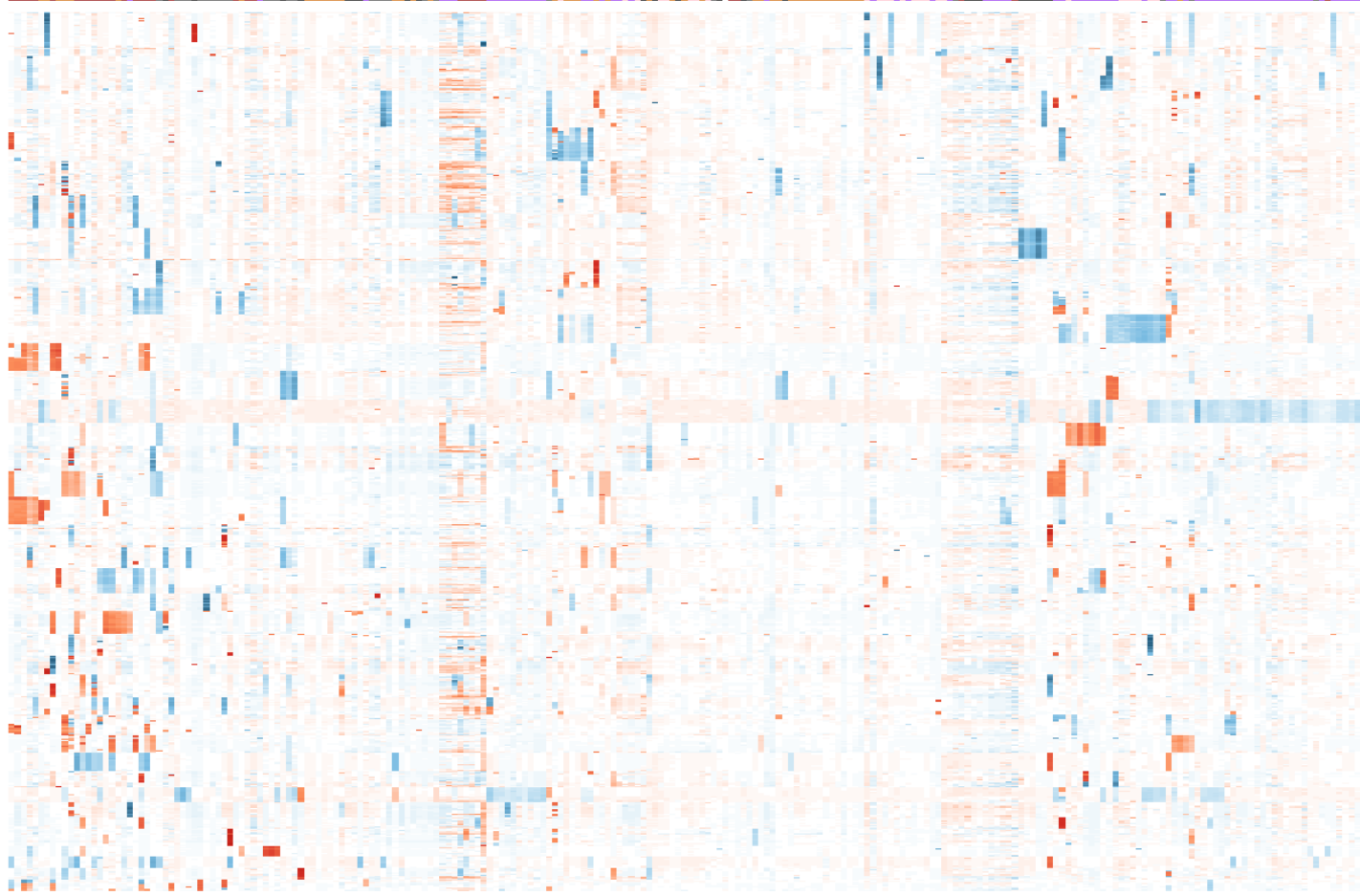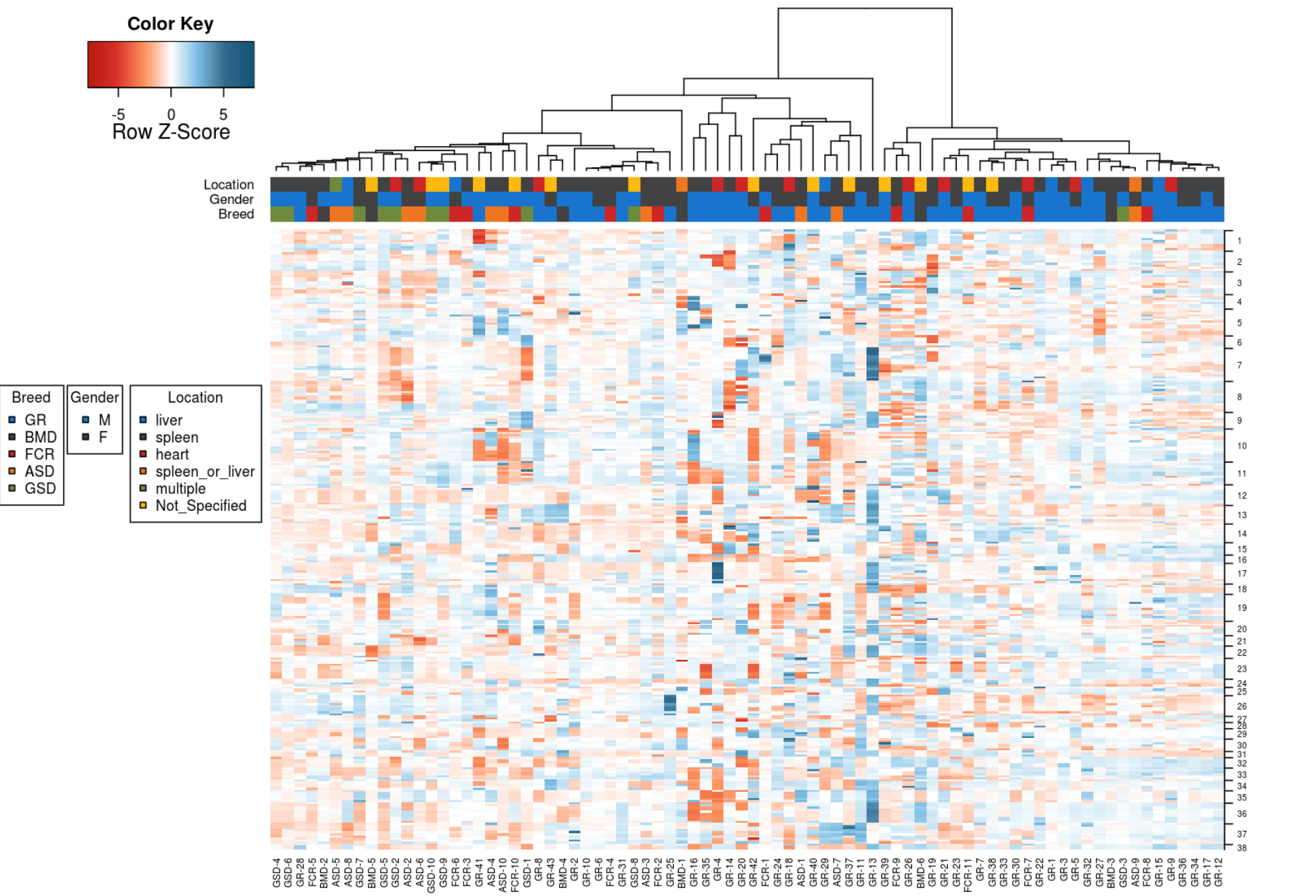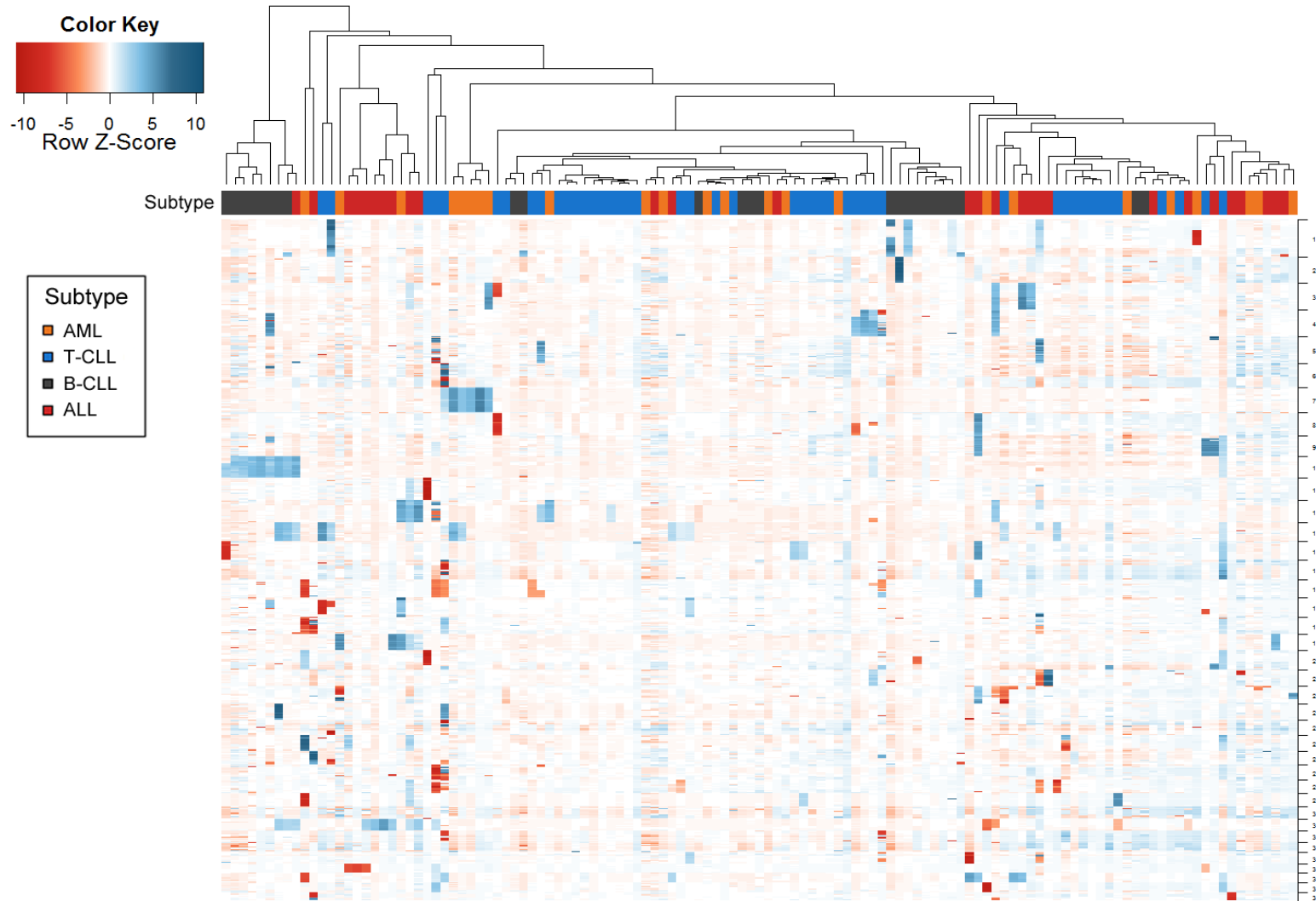
# Lymphosarcoma and Leukemia

# Canine Hemangiosarcoma

# Canine Leukemia

Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*

Leukemia Penetrance Plot

Lymphosarcoma Penetrance Plot
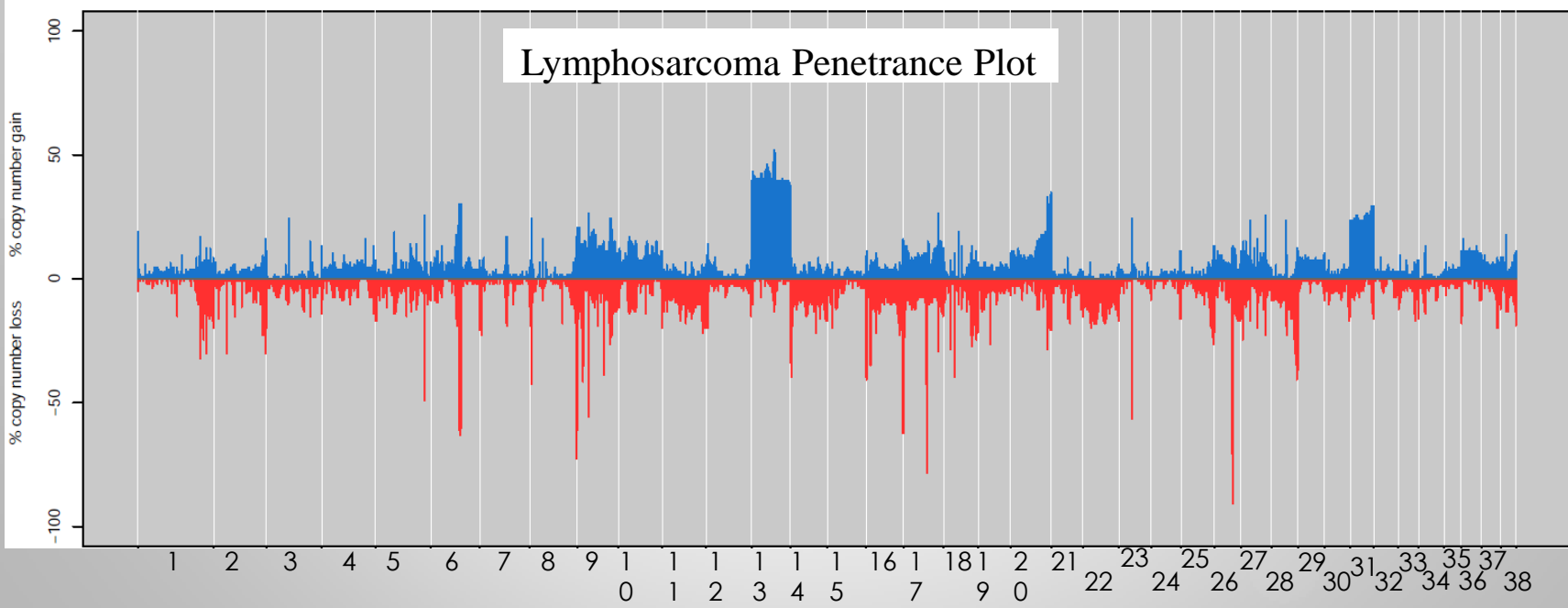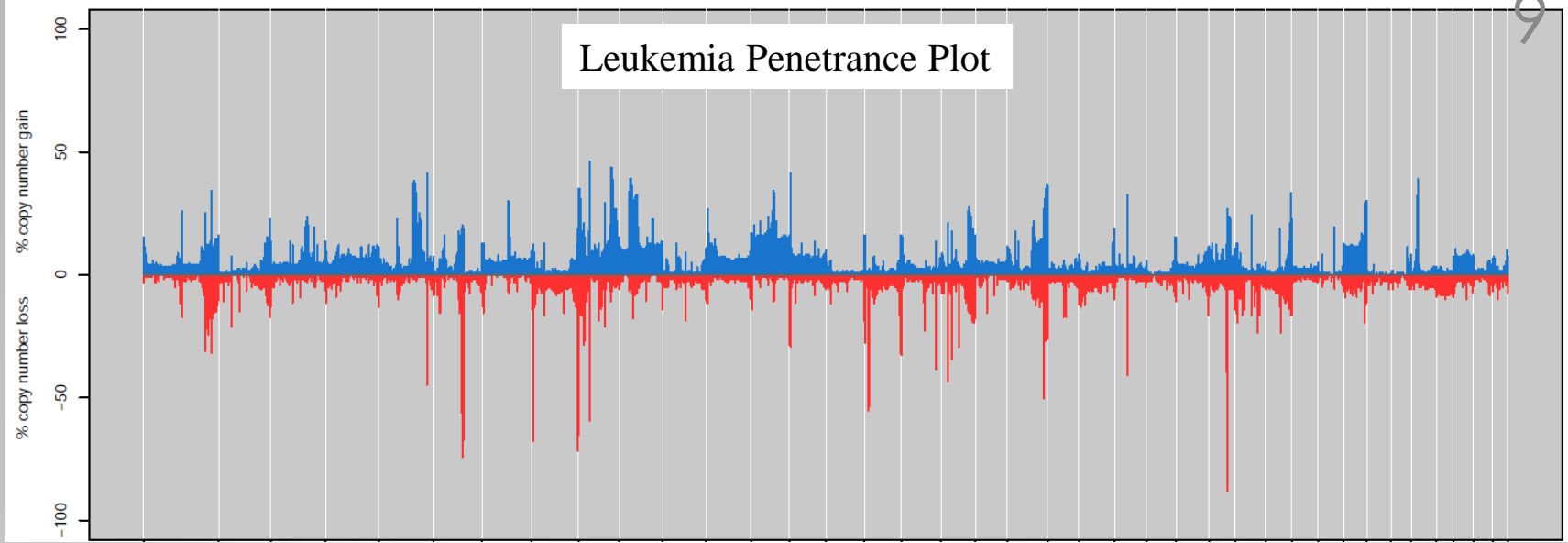
# Leukemia Subtype Penetrance



Penetrance plots of genome-wide CNAs in four subtypes of canine (c) leukemia including ALL (A), AML (B), B-CLL (C), and T-CLL (D). CFA1-38 and X are plotted across the x-axis, and the percentage of cases that demonstrated either copy number gain (blue, above midline) or loss (red, below midline) within a defined chromosomal region are represented on the y-axis. The horizontal lines above and below the midline indicate the 20% threshold for definition of a recurrent CNA.

Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*
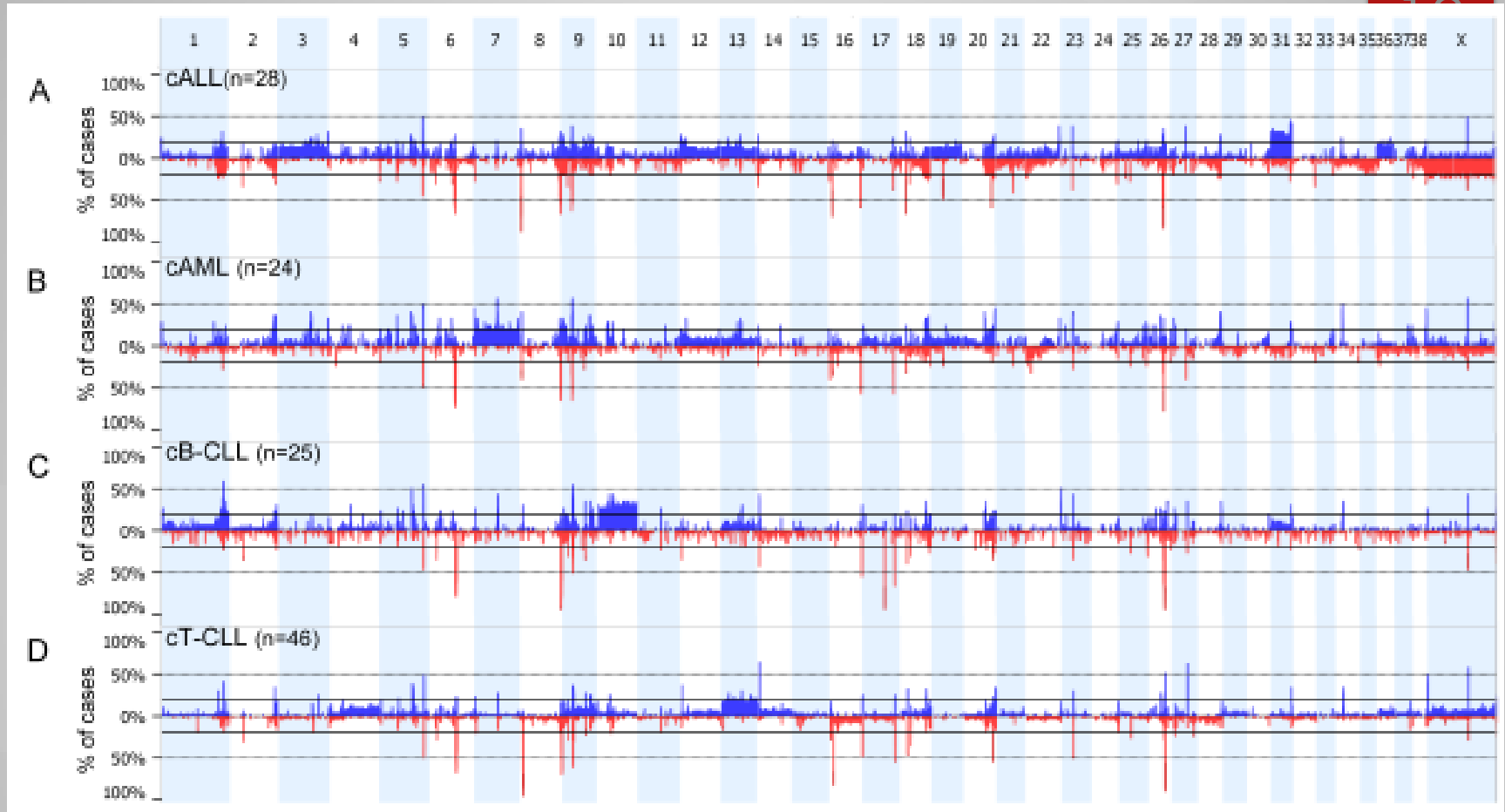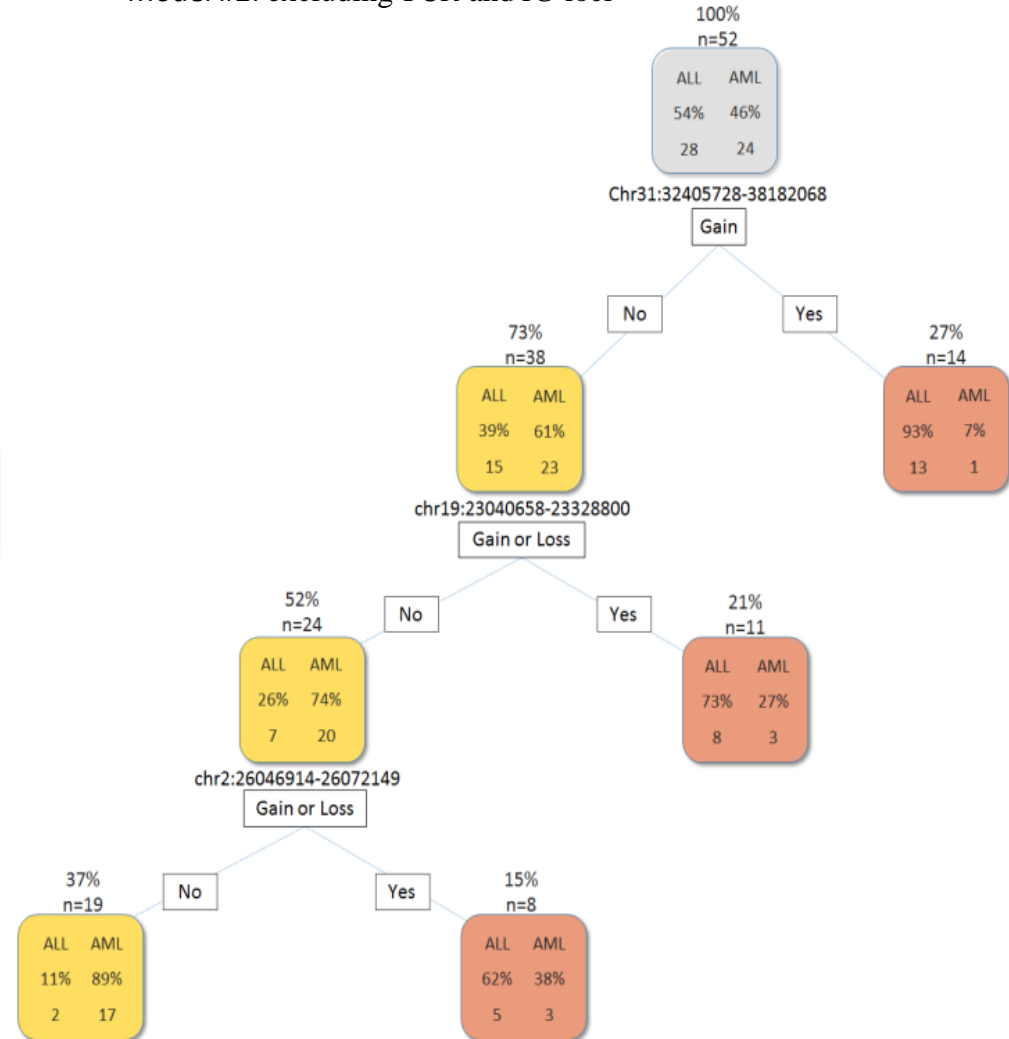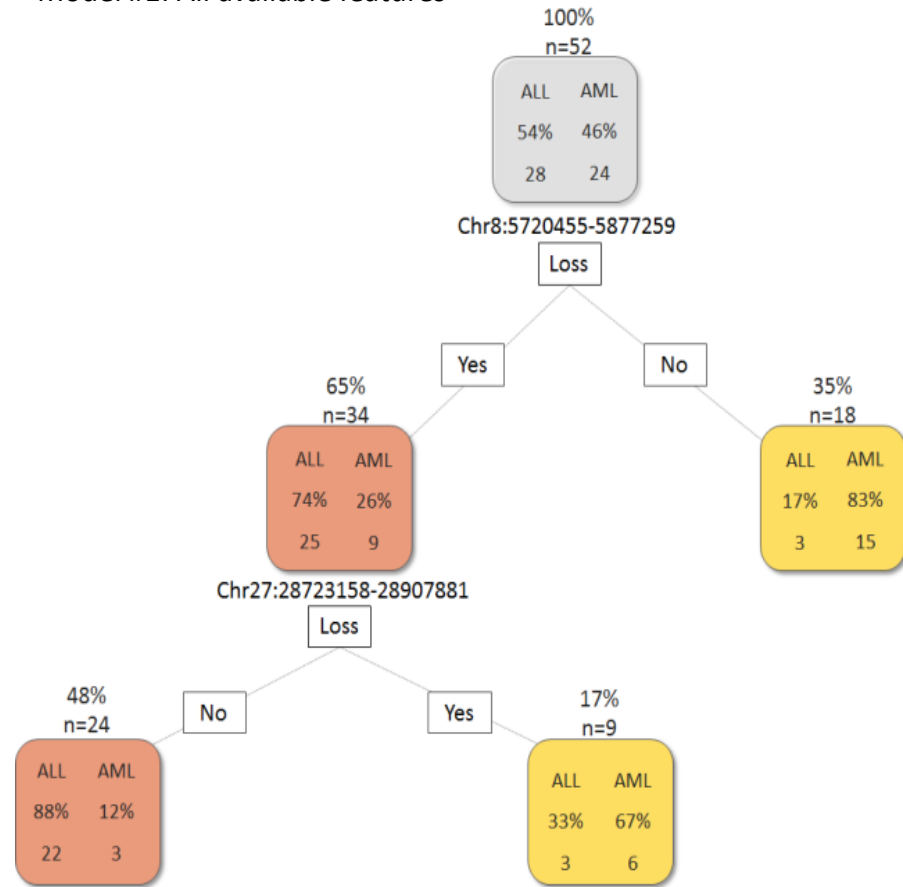
# ALL vs AML

| Model #1: Decision Tree Classifier Results | | | |
|---|---|---|---|
| | ALL | AML | Precision |
| ALL | 22 | 3 | 0.88 |
| AML | 6 | 21 | 0.78 |

| Model #2: Decision Tree Classifier Results | | | |
|---|---|---|---|
| | ALL | AML | Precision |
| ALL | 26 | 7 | 0.79 |
| AML | 2 | 17 | 0.89 |

Model #1: All available features

Model #2: excluding *TCR* and *IG* loci
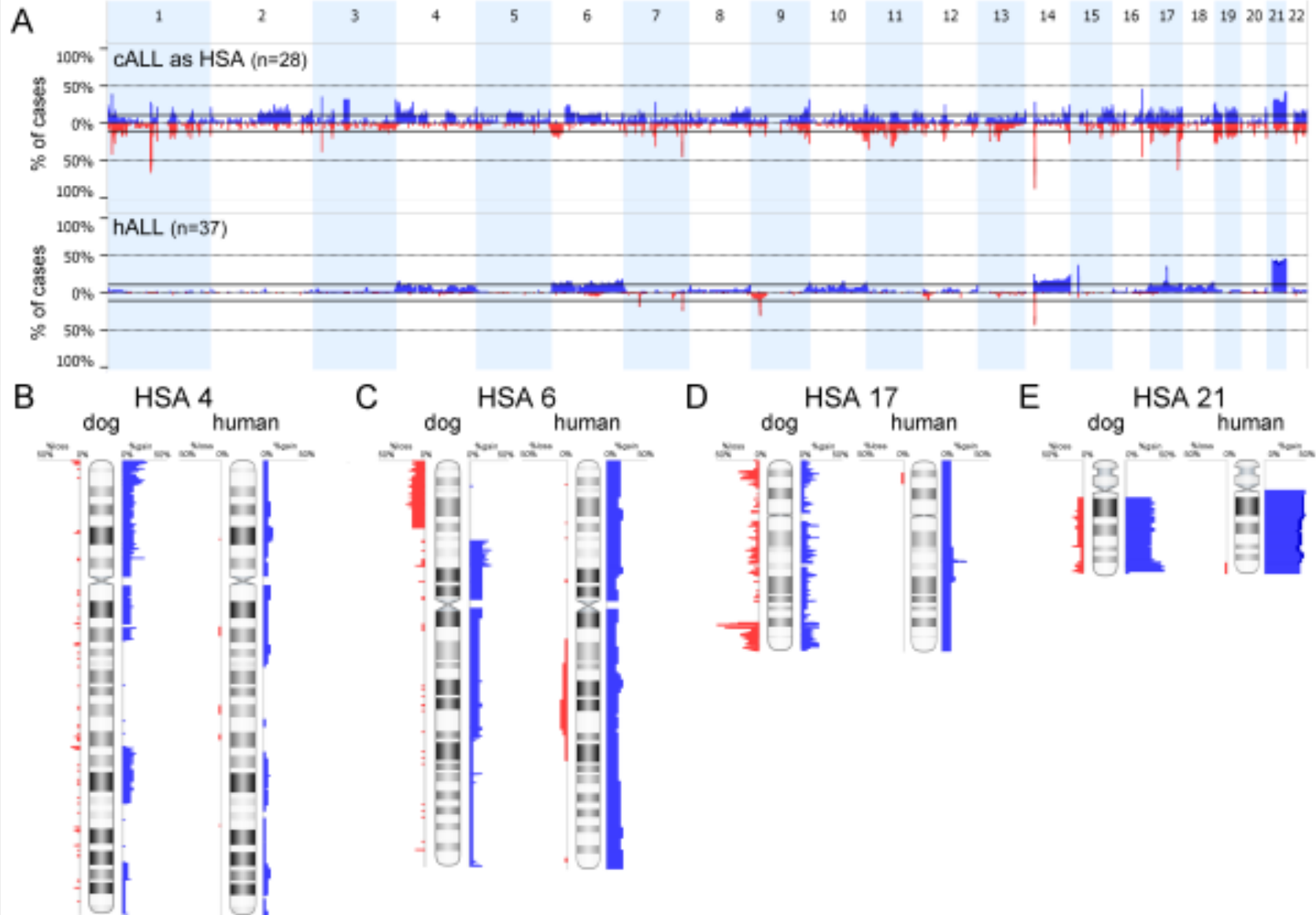


Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*

Roode et al. Genome-wide assessment of recurrent genomic imbalances in canine leukemia identifies evolutionarily conserved copy number changes and regions for subtype differentiation. *In Prep*
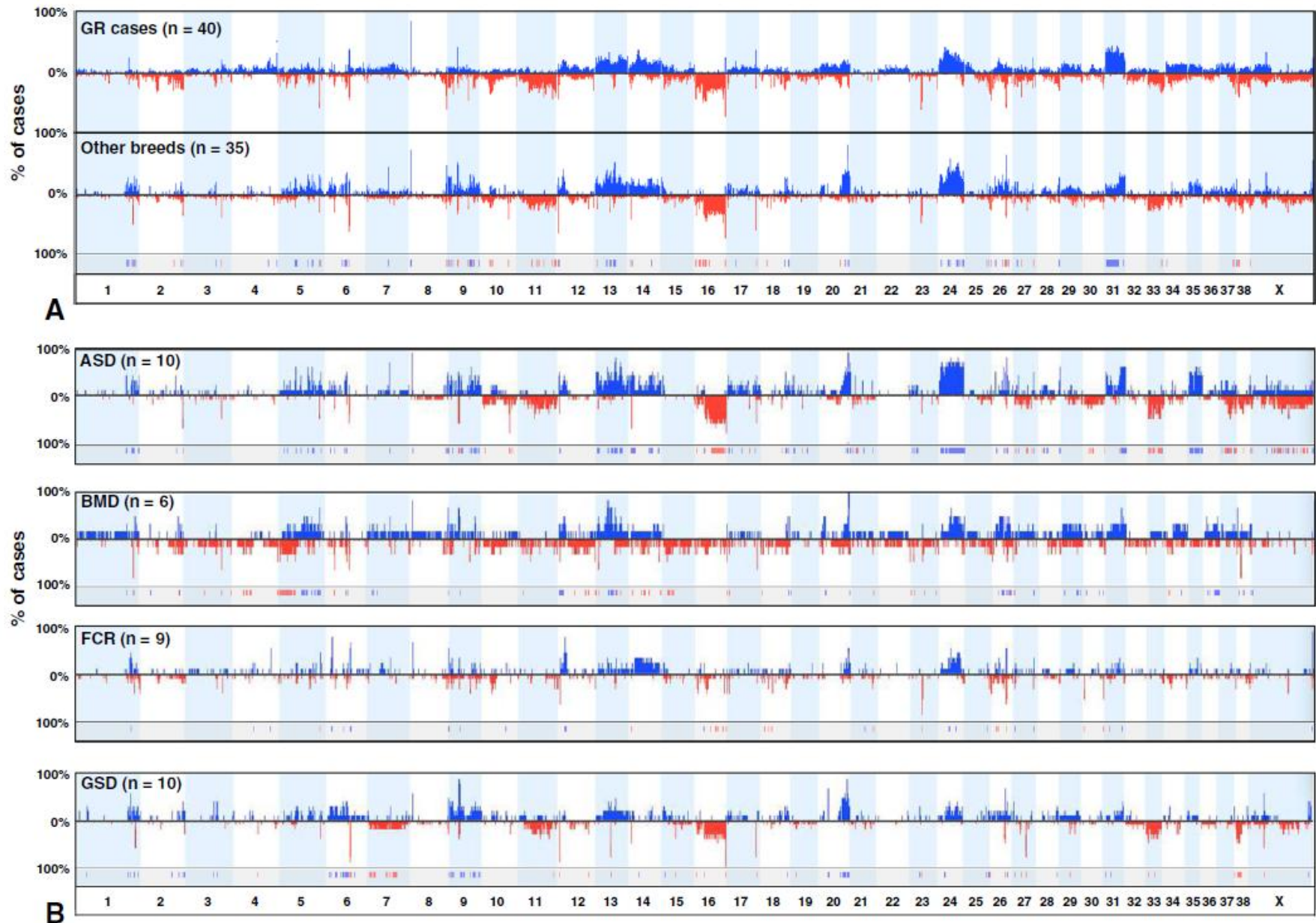
# Hemangiosarcoma



Thomas et al. (2014) Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res.*

# Hemangiosarcoma
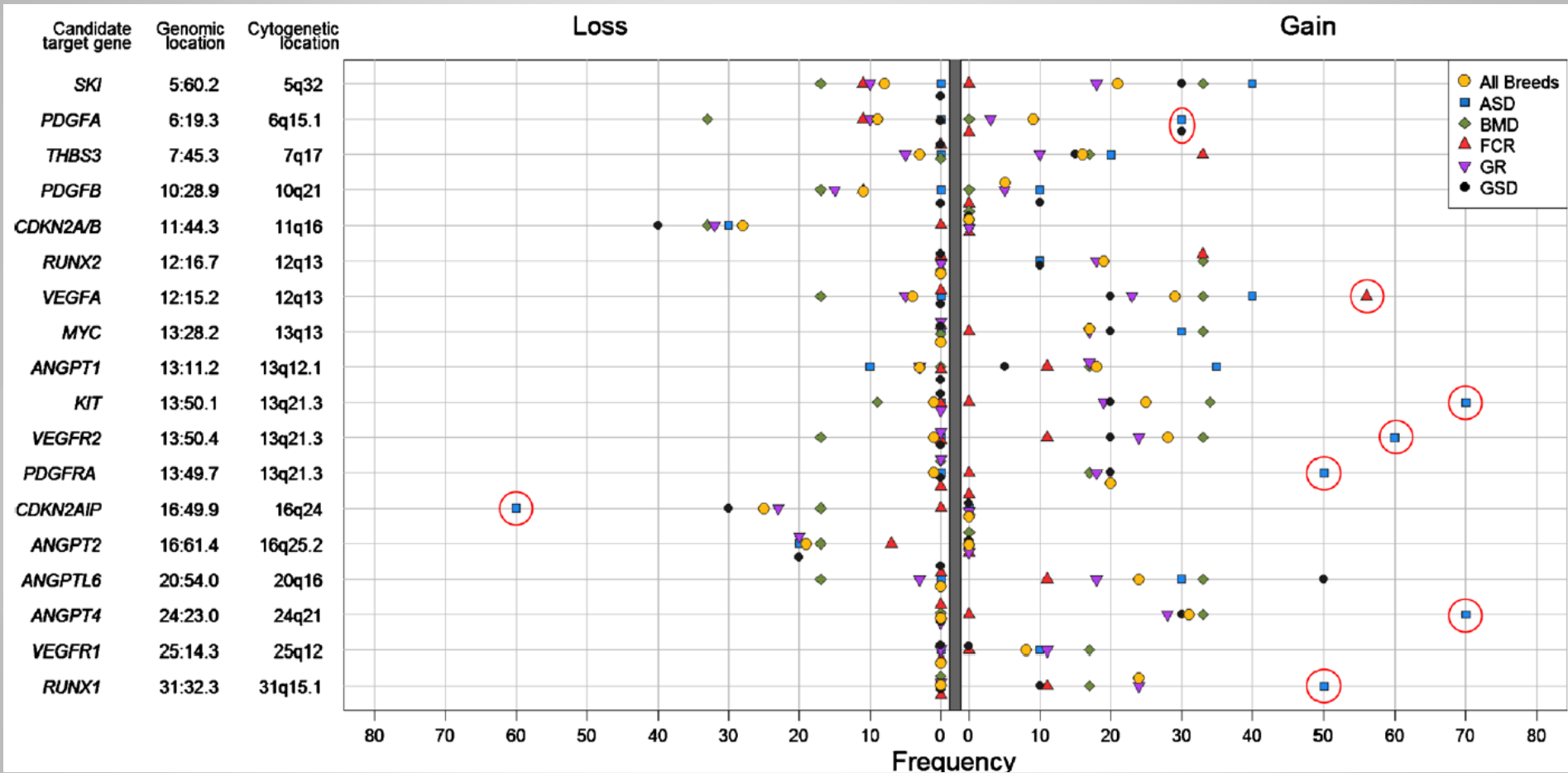
Red circles indicate CNA penetrance values that deviate significantly in one breed compared to all other breeds (p<0.05)

Thomas et al. (2014) Genomic profiling reveals extensive heterogeneity in somatic DNA copy number aberrations of canine hemangiosarcoma. *Chromosome Res.*

# Summary

❑ The oaCGH platform detects CNAs that display distinct differences among cancer classifications and subclassifications

❑ These data can be used to develop predictive models, for example CNAs present in CFA31, CFA19, CFA2 accurately distinguished leukemia ALL and AML subtypes in the present study. Validation of this model is currently underway.

❑ Some cancers (e.g. hemangiosarcoma) show heterogeneity of CNAs across different breeds.

❑ Conserved regions between humans and canines can be mapped and compared to human cancers. These analyses show that for some cancers, common aberrations are observed between species, further highlighting the utility of this model for studying human cancers.

❑ Comparing genes in shared aberrations across breeds and species may help to identify candidate genes that are drivers of carcinogenesis.

# Acknowledgments

**North Carolina State University**

- ❖ Sarah Roode
- ❖ Rachael Thomas
- ❖ Matthew Breen
- ❖ Alison Motsinger-Reif
- ❖ Steven Suter
- ❖ Luke Borst

**University of Minnesota**

- ❖ Jaime Modiano

**University of Utah**

- ❖ Joshua Schiffman

**University of Guelph**

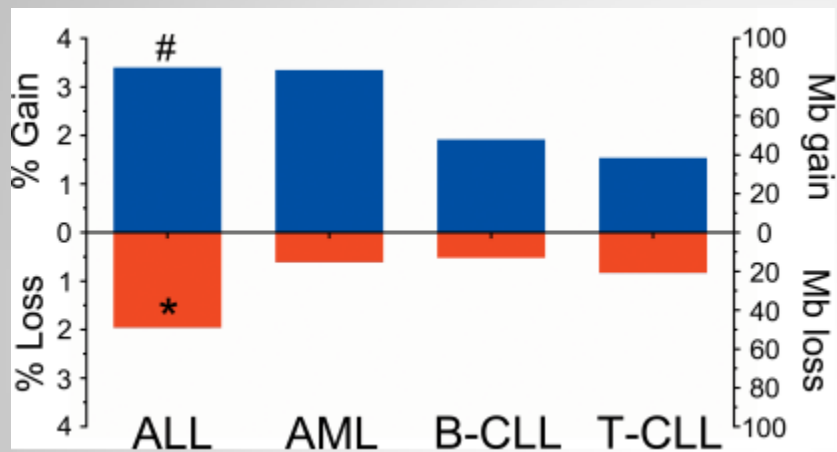- ❖ Dorothee Bienzle

**Colorado State University**

- ❖ Anne Avery

**Broad Institute**

- ❖ Kerstin Lindblad-Toh

# Questions?

Genomic imbalances in each subtype expressed as percent genome changed and the total number of megabases (Mb) within regions of copy number change. The symbol (#) denotes $p<0.05$ for total percent genome changed compared to all other subtypes; and the symbol (*) denotes $p<0.05$ for percent genome loss or gain compared to other subtypes.

Roode et al. *In Prep*
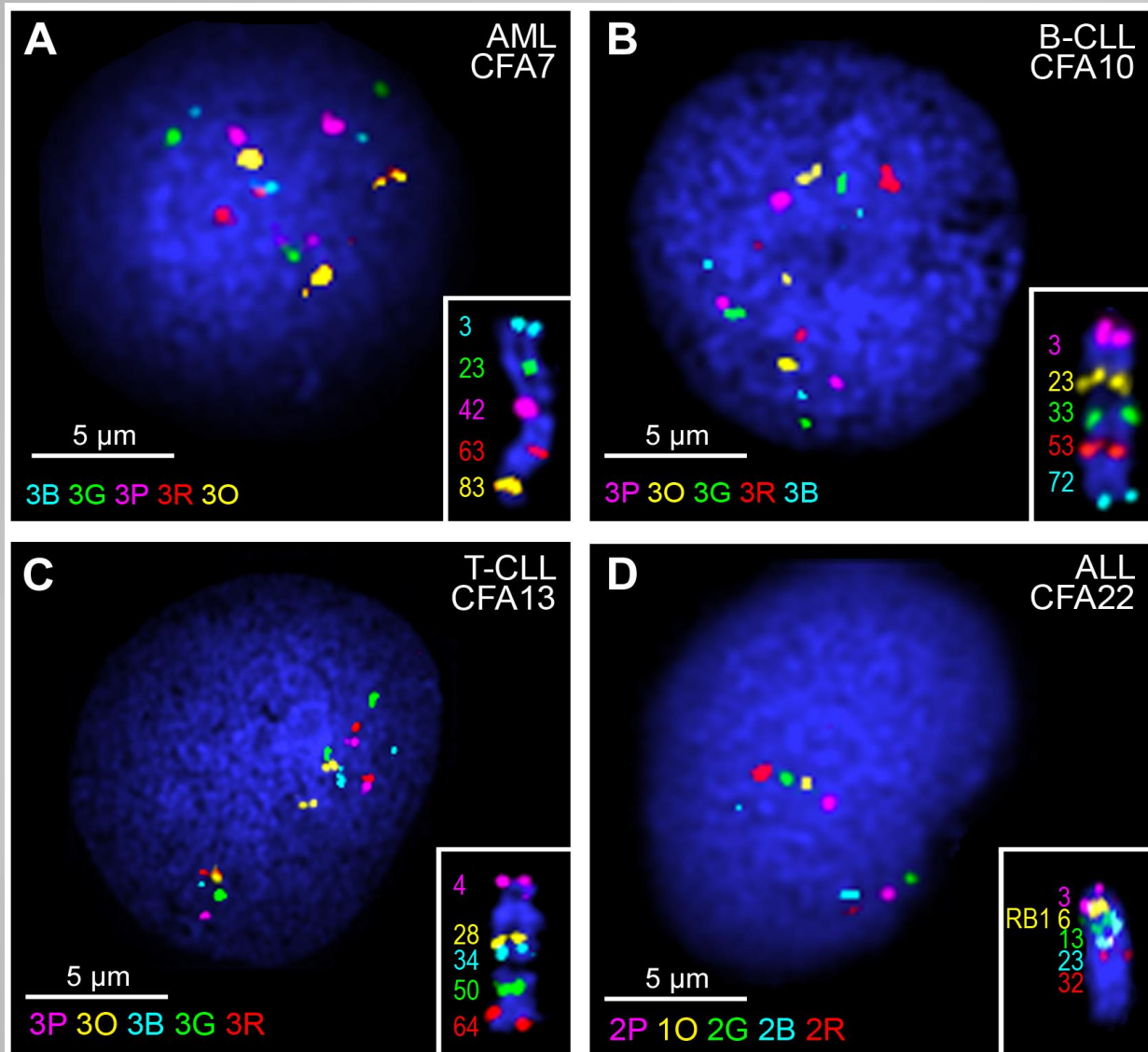
FISH verifies recurrent CNAs identified via oaCGH. Each panel (A-D) includes a representative interphase nuclei harvested from whole blood from a dog with leukemia. The inset shows a control dog chromosome with correct localization of each of the differently labeled BAC clones and the approximate Mb position of each clone. Copy number of each colored probe is also indicated in each panel. (A) Trisomy of CFA 7 in AML. (B) Trisomy of CFA 10 in B-CLL. (C) Trisomy of CFA 13 in T-CLL. (D) Loss of region containing *RB1* in ALL.

Roode et al. *In Prep*
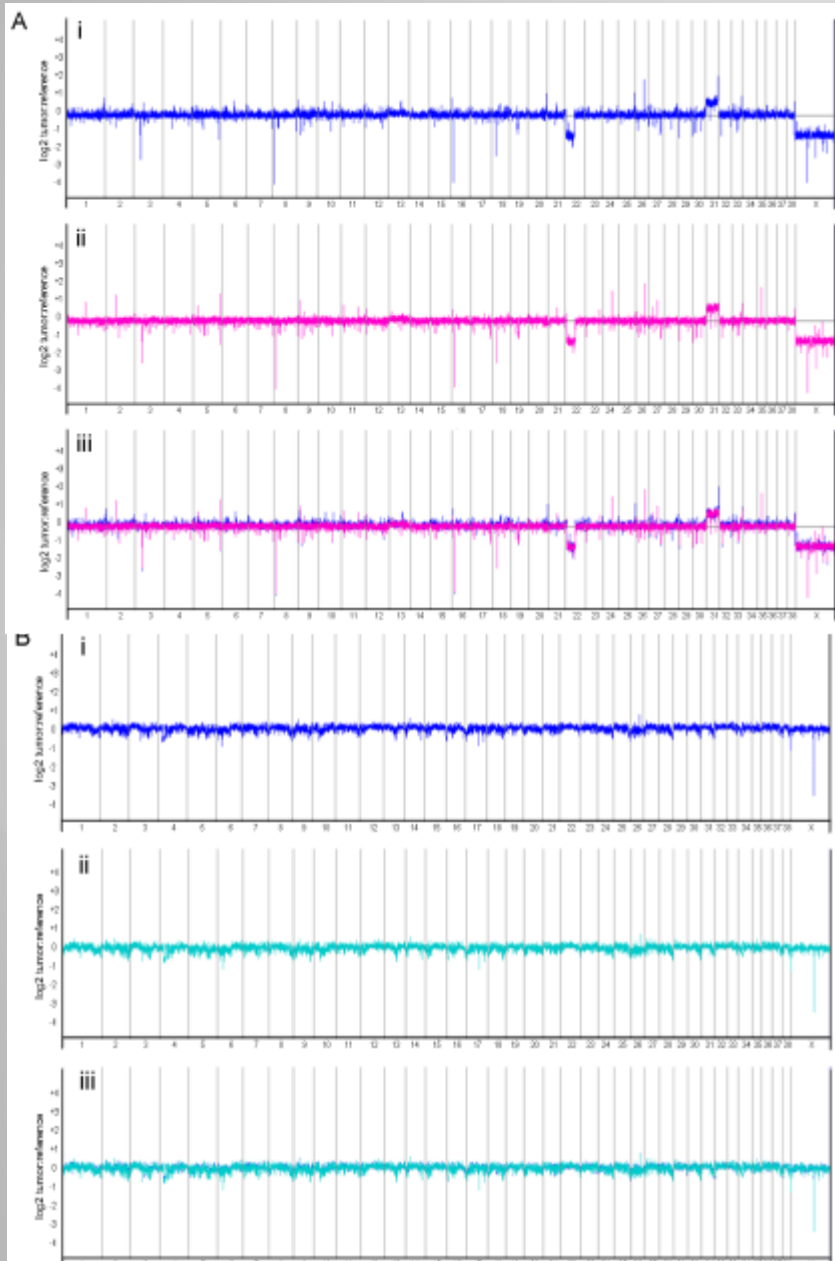
Genome-wide oaCGH profiles comparing DNA isolated from peripheral blood to DNA isolated from flow-sorted neoplastic cells in cases of canine leukemia. Blood samples for representative cases of canine ALL (A) and canine T-CLL (B) were collected and DNA was isolated from both whole blood and a >98% pure population of neoplastic cells derived from fluorescence activated cell sorting. (i) oaCGH profiles of whole blood, (ii) flow-sorted neoplastic cells, and (iii) the stacked overlay of the two profiles, were assessed for differences in aberration detection between sample type due to presumed cell heterogeneity in whole blood. Each oaCGH profile includes the chromosomes (1-38, X) on the x-axis and log2 tumor:reference ratio on the y-axis with gains visible above the midline, and losses below the midline. The case of ALL (A) has a gain of CFA 31 and loss of the proximal half of CFA 22 and CFA X which is equally evident in profiles of both sample types (A,i-iii). The case of T-CLL has few CNAs evident in either sample type (B, i-ii) and the profiles are indistinguishable when overlaid (B, iii).

Roode et al. *In Prep*
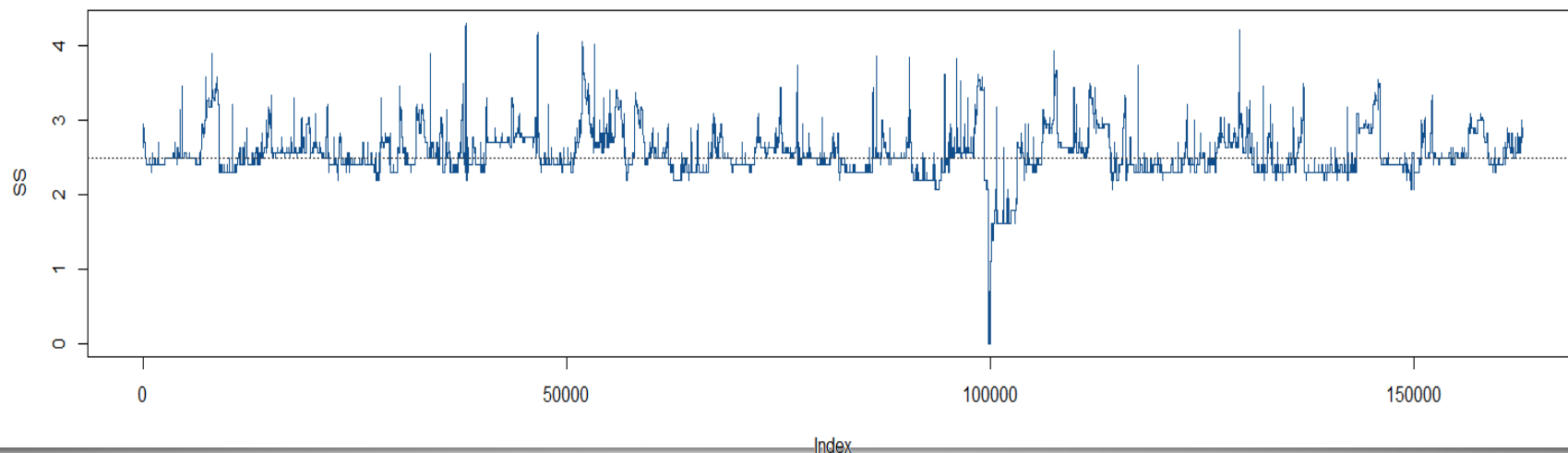
Example of the region calling algorithm for marker at index 100,000. Variance was determined relative to this marker. The more similar a neighboring marker is, the lower the variance value. A value of 0 would indicate an exact match. A threshold hold of 2.5 was used to define regions. Therefore, any contiguous marker with a value of $< 2.5$ was considered to consist of a single region.

# B-CLL vs T-CLL

| | Decision Tree Classifier Results | | |
|---|---|---|---|
| | B-CLL | T-CLL | Precision |
| **B-CLL** | 23 | 1 | 0.96 |
| **T-CLL** | 2 | 45 | 0.96 |

| | Model #2: Decision Tree Classifier Results | | |
|---|---|---|---|
| | B-CLL | T-CLL | Precision |
| **B-CLL** | 22 | 6 | 0.79 |
| **T-CLL** | 3 | 40 | 0.93 |

Model #1: All available features

Model #2: excluding *TCR* and *IG* loci



Roode et al. *In Prep*