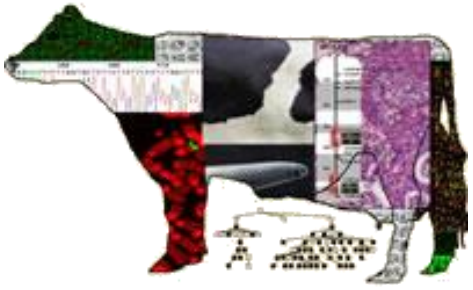
Copy number variation of individual cattle genomes using Next Generation Sequencing

Derek M. Bickhart, PhD
Bovine Functional Genomics Lab
Animal and Natural Resources Institute
USDA ARS BFGL
01/15/2012



Creating CNV maps with NGS data

Background and Introduction

Individualized CNV maps

Detection within populations

Background and Introduction

Genetic Variation

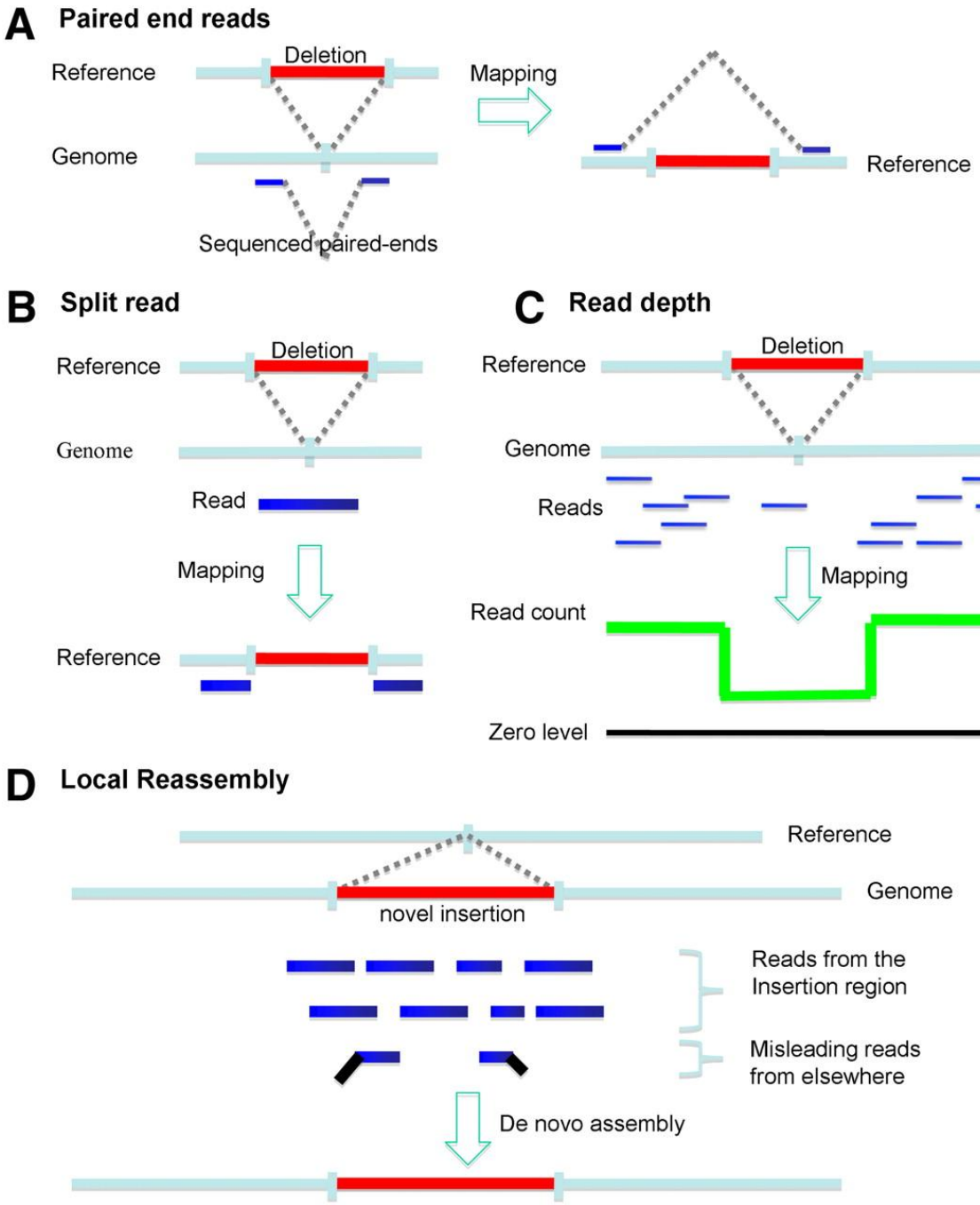
How genomes change over time

Sequence

- Single nucleotide variations – SNP (human 60 million)
- Small insertions/deletions – frameshift, microsatellite, minisatellite
- Mobile Elements – SINE, LINE Transposition (300bp - 6 kb)
- Genomic structural variation (1 kb – 5 Mb)
  - Large-scale Insertions/Deletions (Copy Number Variation: CNV)
  - Segmental Duplications (> 1kb, > 90%)
  - Chromosomal Inversions, Translocations, Fusions.

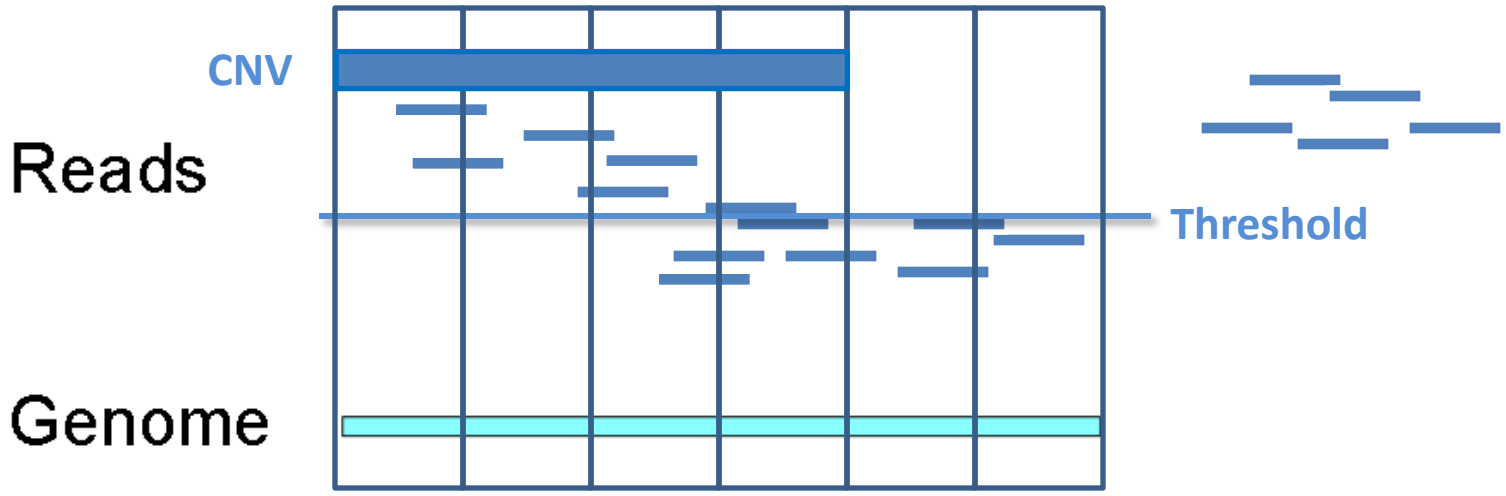
Cytogenetics

CNVs contribute to phenotype



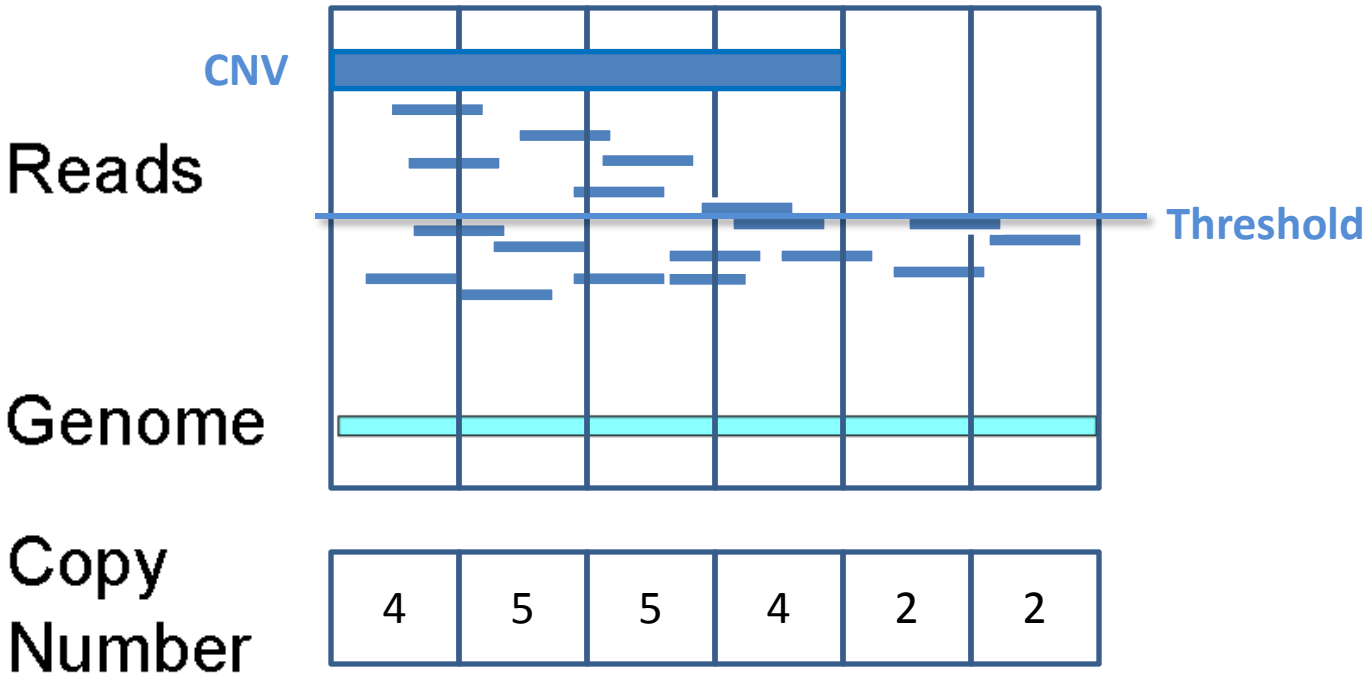
Figures Adapted from
Snyder et al, 2010

Read Depth (RD) Detection



- Align reads to reference genome (used mrsFAST)
- Determine average RD and set Threshold
  - Threshold value: average + 4 Stdev
  - Normalized for GC bias
- Section genome into windows and call CNVs
- Analysis can be done in SD regions

RD provides genome-wide Copy Number (CN)



Advantages of this approach:

- Assign CN to Gene Regions
- CN is not relative to other samples
- Works on all reference assemblies

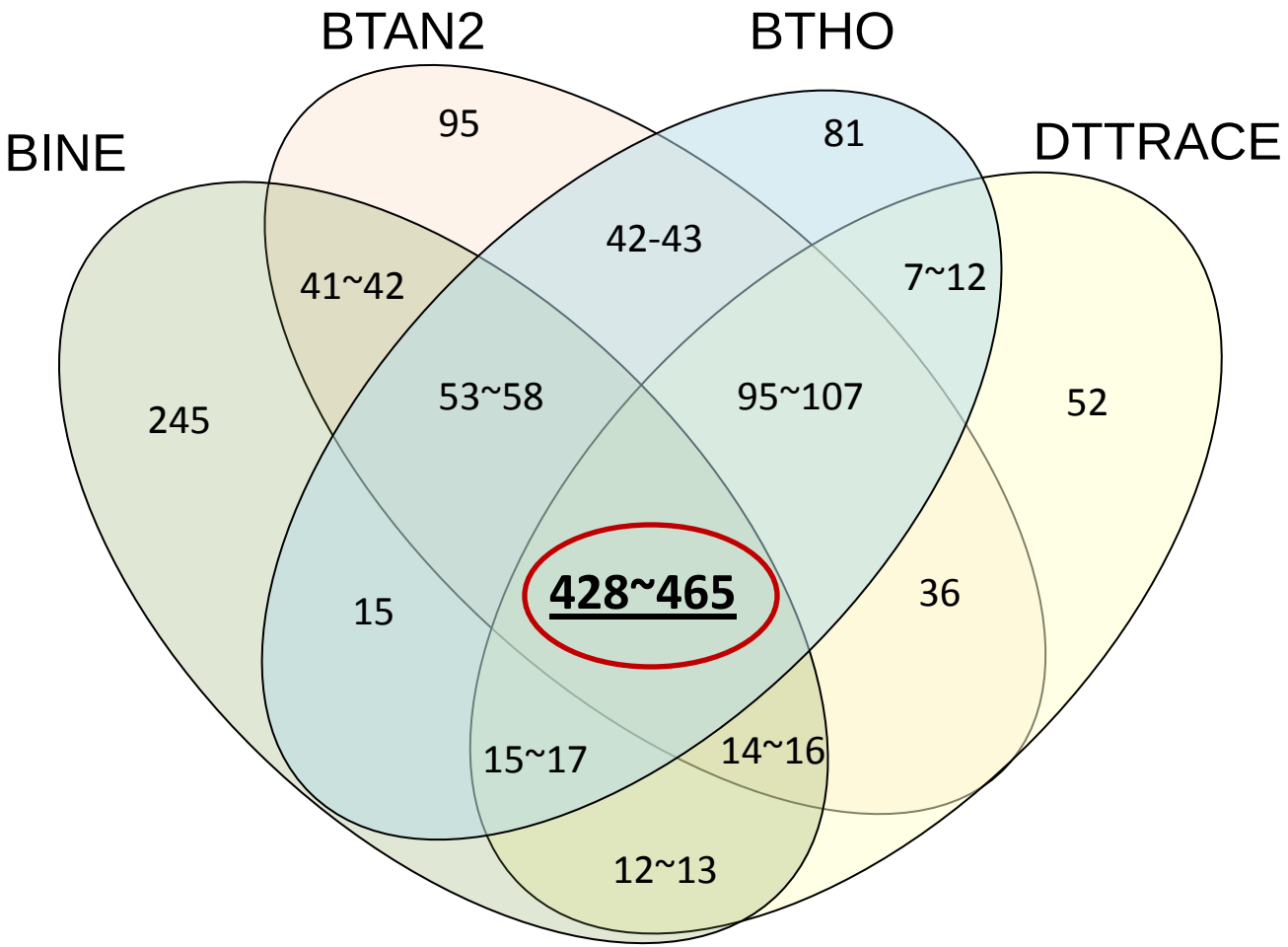
Individualized CNV maps

Animals Selected

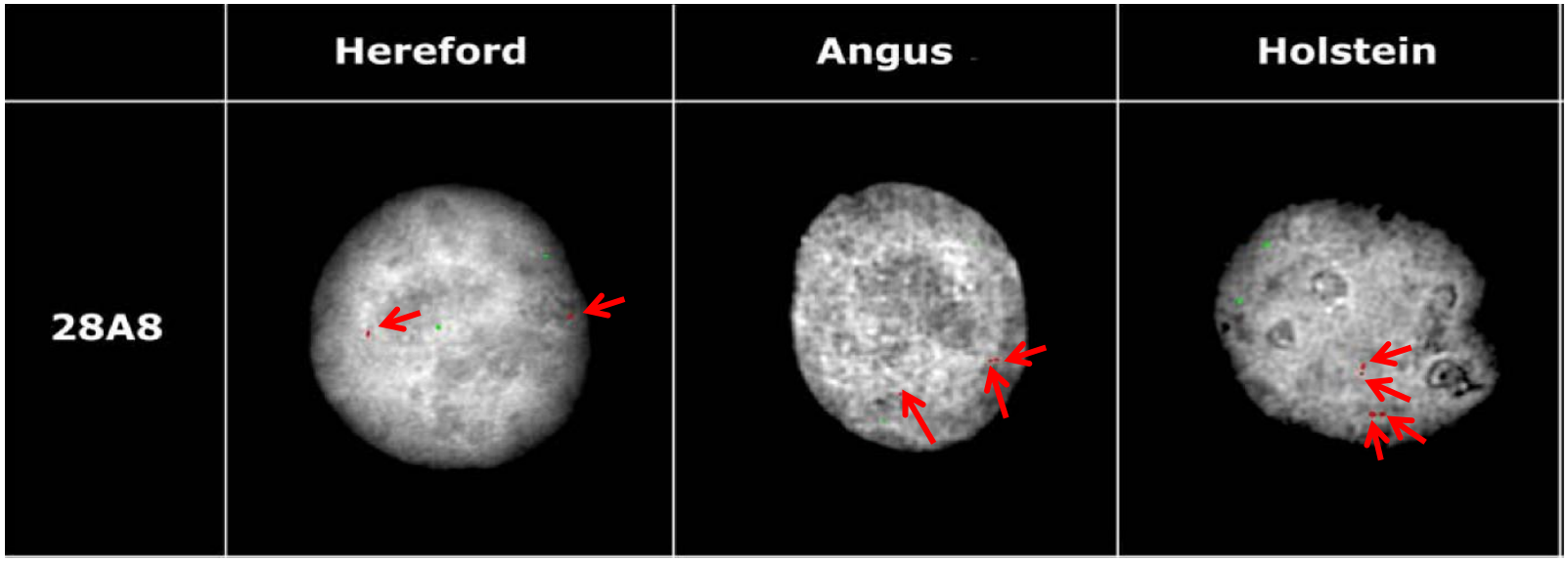
Animal Abbreviation	Breed	Coverage
BINE	Nelore	High
BTAN1	Angus	High
BTAN2	Angus	High
BTAN3	Angus	High
BTHO	Holstein	Low
DTTRACE	Hereford	Low

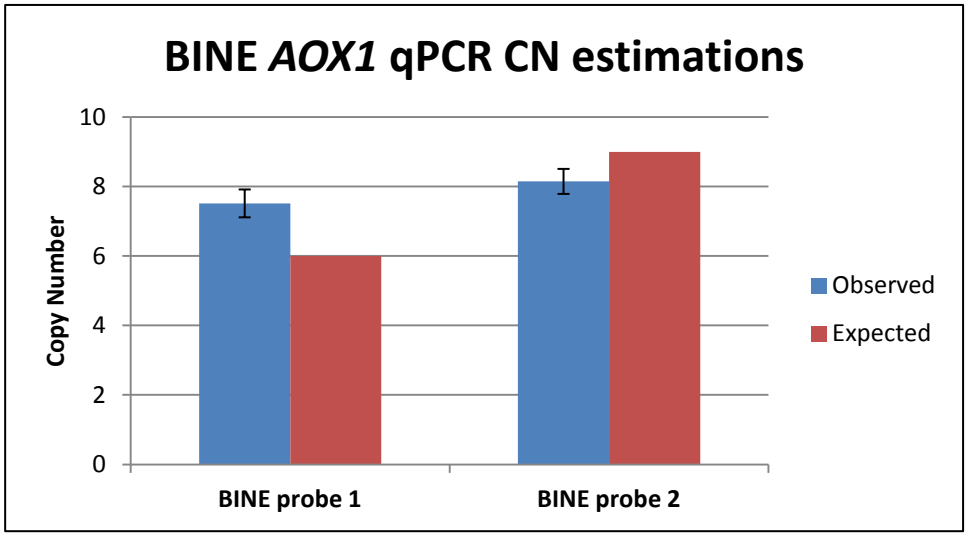
Reference animal

CNV events more diverse than in humans



Experimental validation



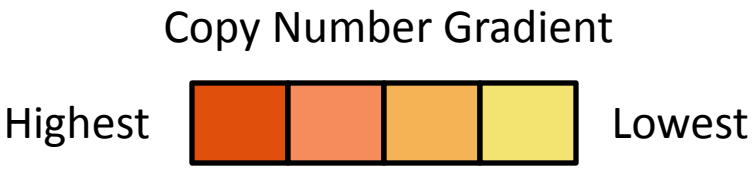


- ~55% agreement with BAC FISH
- 82% agreement with qPCR assays
- 8% false discovery compared to aCGH

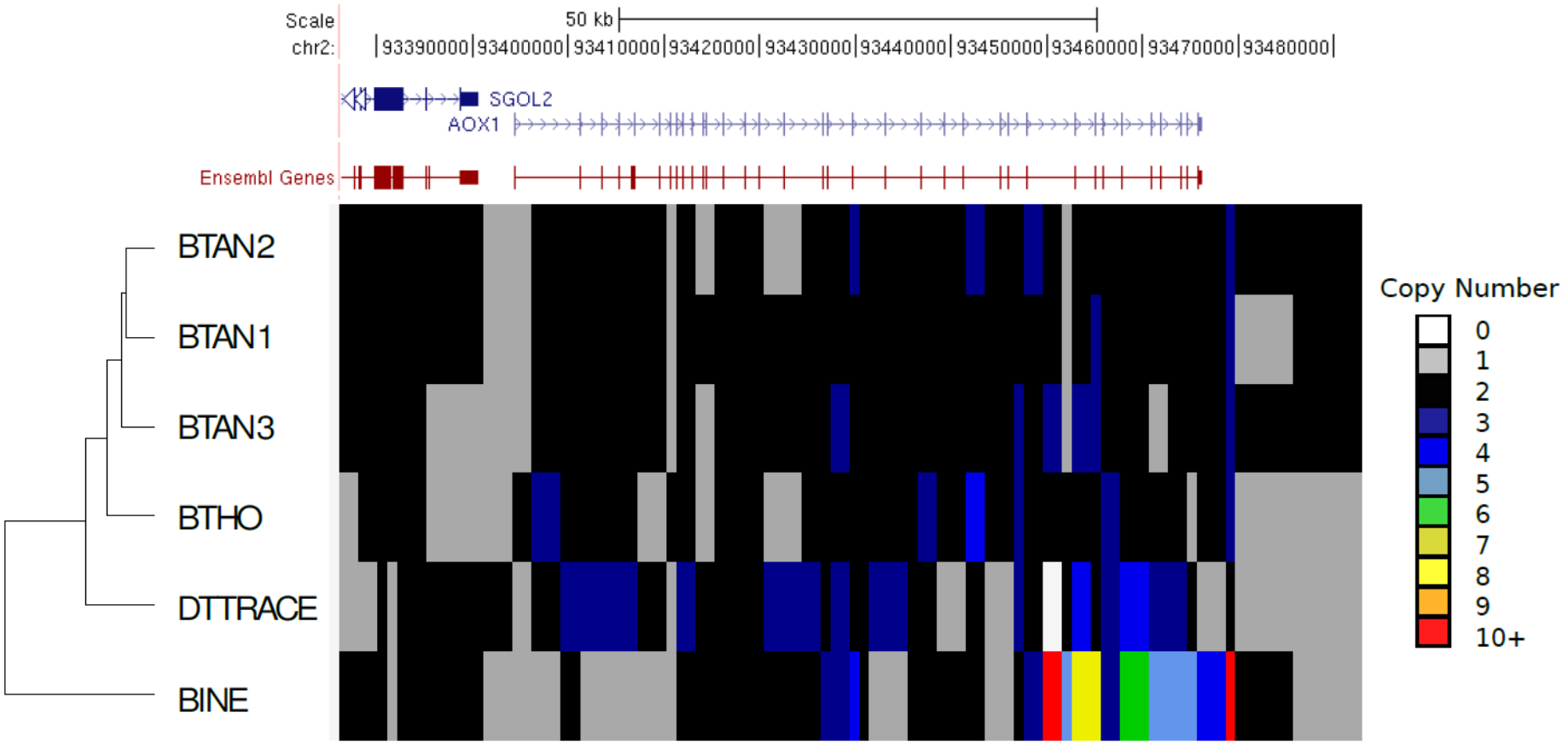
BSP30A is duplicated in all breeds

CN estimates give better resolution

Gene ID	BINE	BTAN1	BTAN2	BTAN3	BTHO	DTTRACE
CATHL4 BSP30A	7.9	9.7	8.7	7.5	11.8	3.6

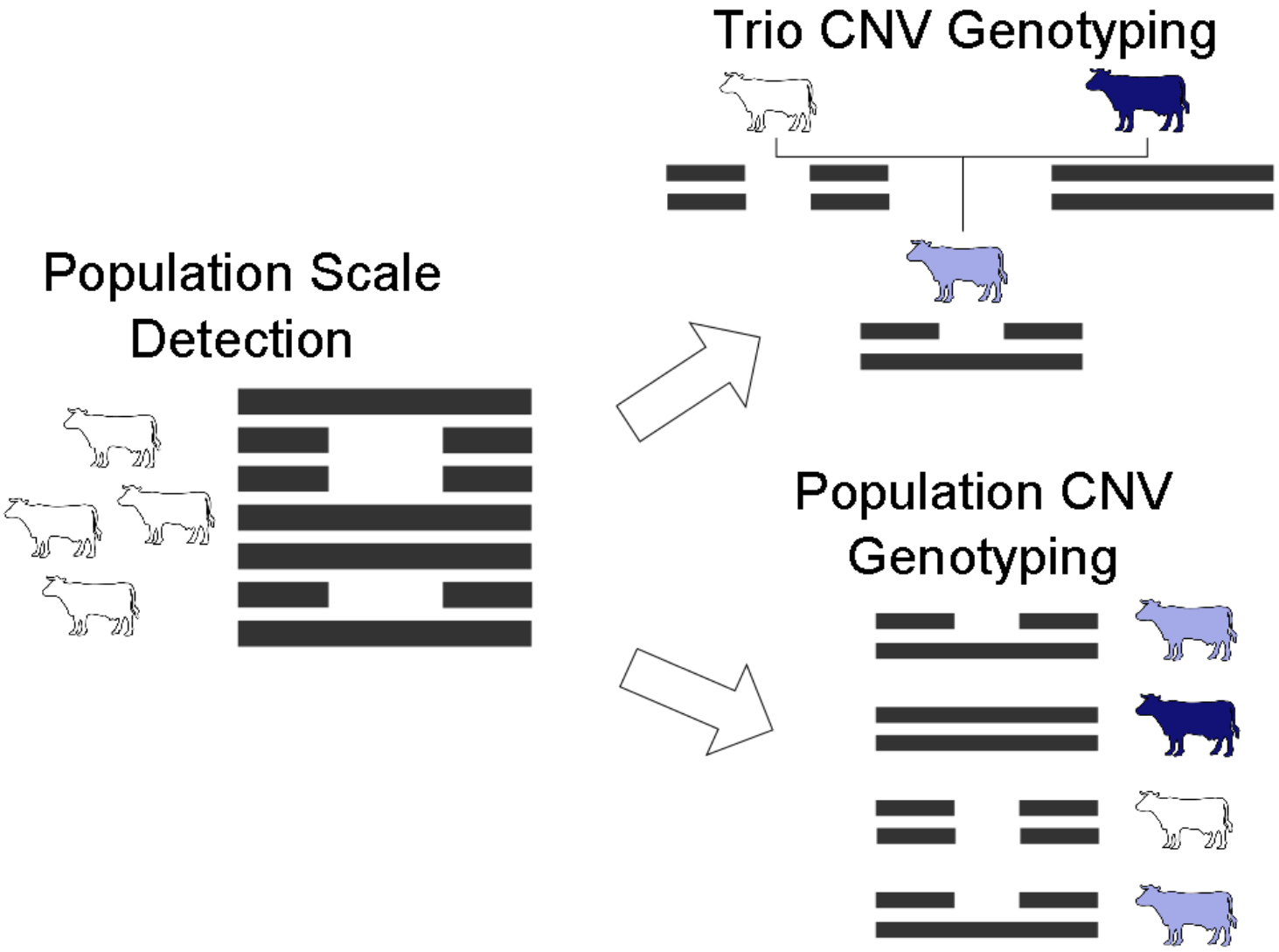


Breed differences can be highlighted

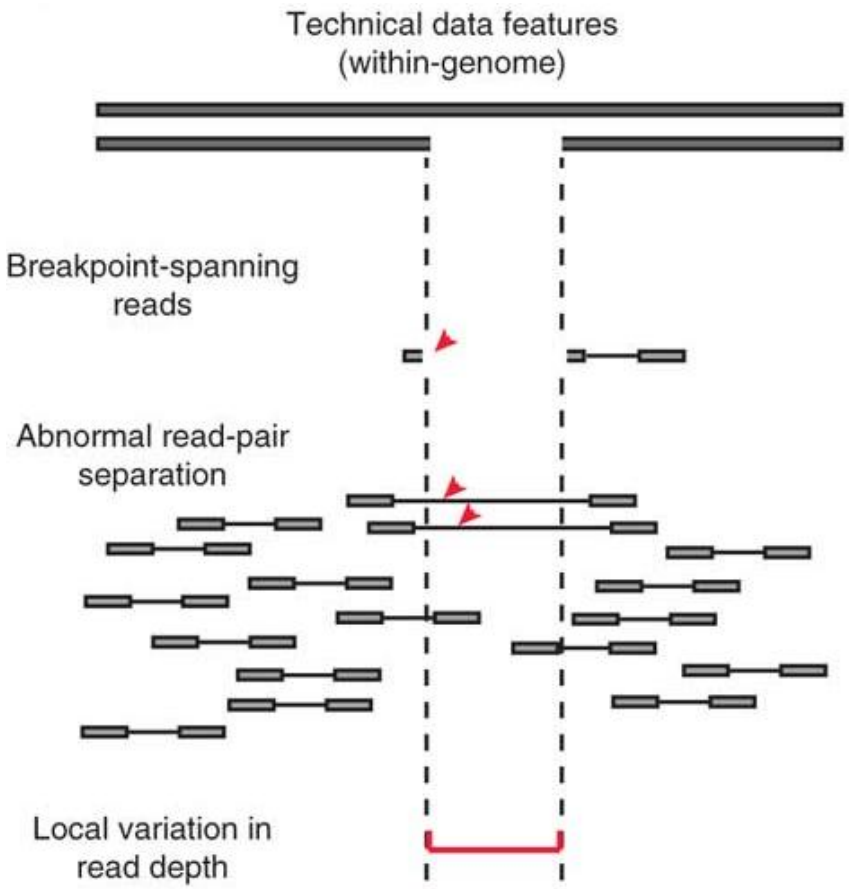


Detection within Populations

Expanding the sample size allows high resolution detection



Improve analysis with multiple methods



Images adapted from Handsaker et al. 2010

- ❑ Multiple techniques increase detection
- ❑ RD disadvantages
  - ○ Only homozygous deletions
  - ○ No mobile element insertion (MEI) detection
- ❑ Genome Strip (pictured left)
  - ○ Combines three techniques
  - ○ Limited to Deletions
- ❑ Include methods:
  - ○ VariationHunter v2 (MEI detection)
  - ○ RD method (duplication detection)
- ❑ Also include:
  - ○ SNP array data
  - ○ CGH array data

Summary

Selected suitable detection strategy

Created high resolution individual maps

Transitioning to population scale analysis

Acknowledgements

- Dr. George Liu's lab
  - Dr. Yali Hou
  - Reuben Anderson
  - Alexandre Dimtchev
- USDA BFGL
  - Dr. Tad Sonstegard
  - Dr. Curtis Van Tassell
  - Dr. Steven Schroeder
- Dr. Evan Eichler's lab
- Dr. Jeremy Taylor's and Dr. Robert Schnabel's lab
- Dr. Jose Fernando Garcia

- Funded by National Research Initiative (NRI) Grant No. 2007-35205-17869 and 2011-67015-30183 from USDA-NIFA

This presentation's Poster Number: P0549

Buffalo CNV discovery Poster Number: P0574