Copy number variants identified on the new bovine reference genome partly match known variants

A. M. Butty[1], T. C. S. Chud[1], F. Miglior[1], F. S. Schenkel[1], A. Kommadath[2,3], K. Krivushin[2], J. R. Grant[2], I. M. Häfliger[4], C. Drögemüller[4], P. Stothard[2], and C. F. Baes[1,4]

[1]CGIL – University of Guelph, Guelph, Canada; [2]AFNS – University of Alberta, Edmonton, Canada; [3]Lacombe R&D Centre – Agriculture and Agri-Food Canada, Lacombe, Canada; [4]Vetsuisse Faculty – University of Bern, Switzerland

Copy number variants

- Deletions or duplications longer than 50 base pairs
  - Previously 1Kb but technologies are improving at detecting shorter events
  - Maximum length defined to 5 Mb

Copy number variants

- Deletions or duplications longer than 50 base pairs
  - Previously 1Kb but technologies are improving at detecting shorter events
  - Maximum length defined to 5 Mb

- No consensus on definition or quality criteria
  - Lack of definition is not only present in cattle but also in human data

Copy number variants

- Deletions or duplications longer than 50 base pairs
  - Previously 1Kb but technologies are improving at detecting shorter events
  - Maximum length defined to 5 Mb

- No consensus on definition or quality criteria
  - Lack of definition is not only present in cattle but also in human data

- Identification of high confidence CNV possible *in silico* using multiple detection methods
  - Validation done with qPCR

Objectives

- Develop a pipeline for in silico identification of **high confidence CNV regions** (CNVR) using two sources of genomic information

Objectives

- Develop a pipeline for in silico identification of **high confidence CNV regions** (CNVR) using two sources of genomic information

- **Describe** the identified high confidence CNVR and their putative functions

CNV identification

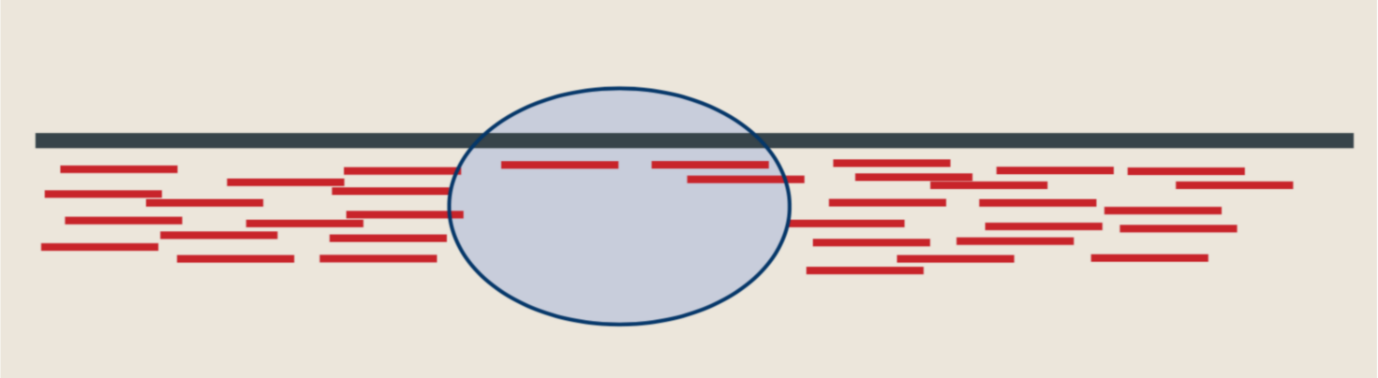
- Samples:
  - 96 Holstein animals

- Dataset:
  - Whole-genome re-sequences (**WGS**)
  - SNP array genotype (**GEN**)

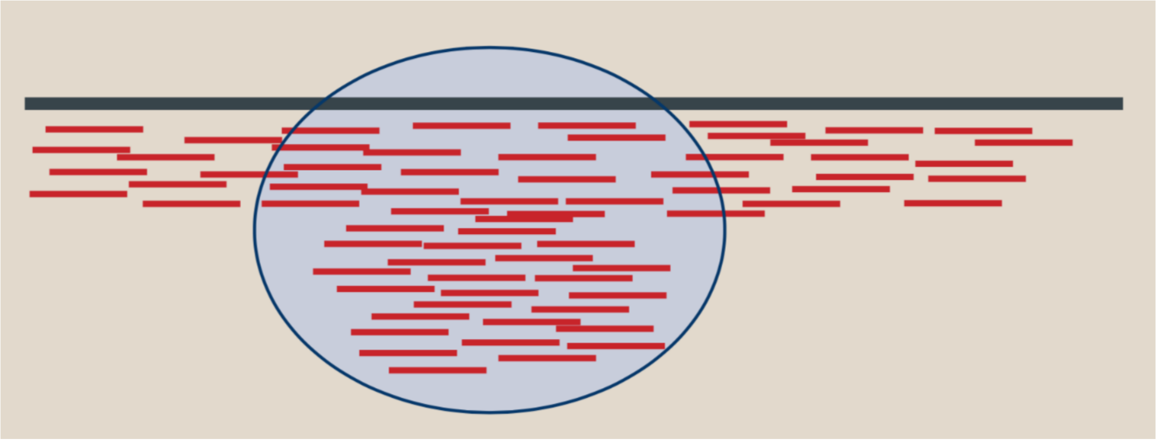
- Algorithms focusing on:
  - Read depth (for WGS)
  - SNP Chip CNV calling (for GEN)

CNV identification – read depth (WGS)

Deletion



Duplication



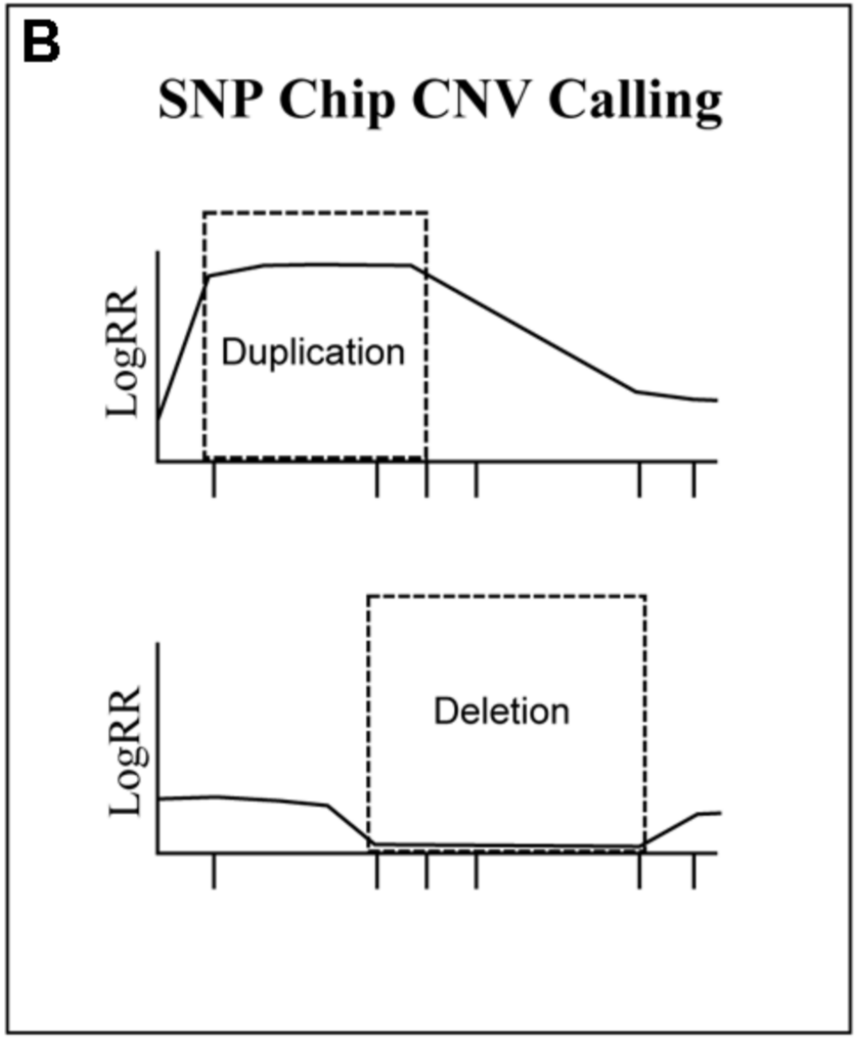
The reference assembly ARS-UCD1.2 was used for alignment of the sequenced reads and CNV identification

Alkan et al. (2011)

CNV identification – signal intensity (GEN)

The formula of the Log R ratio:

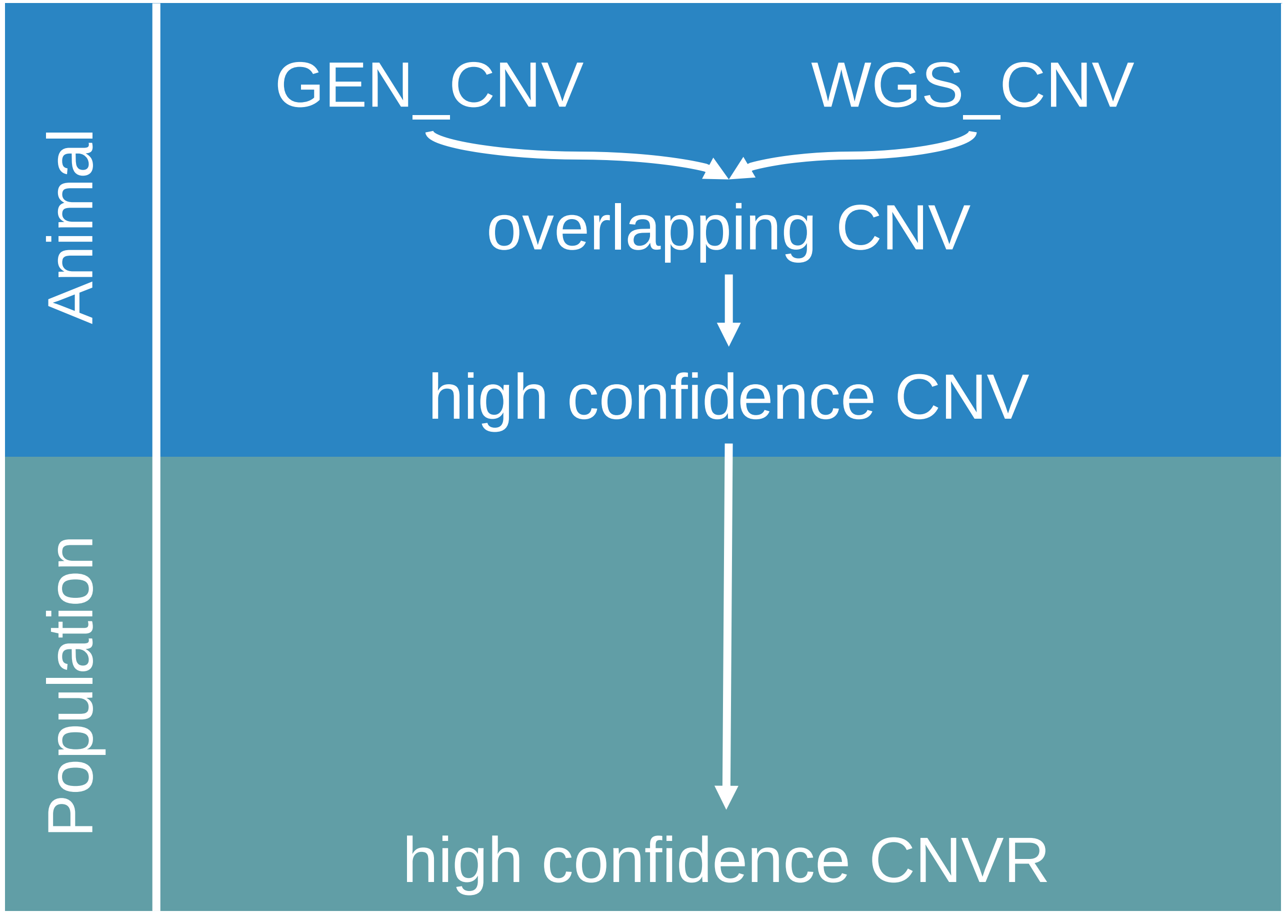
$$log_2(\frac{observed\ intensity}{reference\ intensity})$$



Bickhart and Liu (2014)

Comparison GEN vs. WGS

ANIMAL_CNVR



GEN_CNV

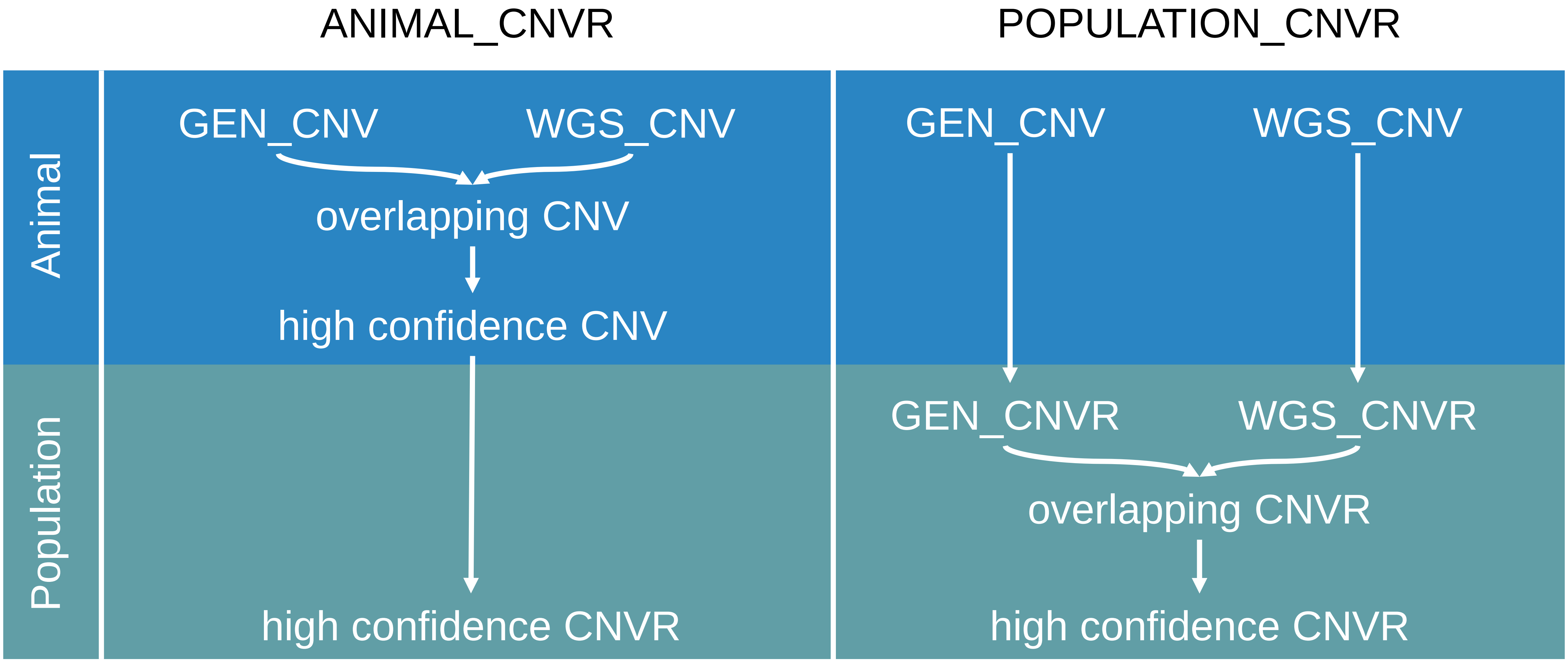
WGS_CNV

overlapping CNV

high confidence CNV

high confidence CNVR

Comparison GEN vs. WGS



ANIMAL_CNVR

POPULATION_CNVR

Animal

Population

GEN_CNV

WGS_CNV

overlapping CNV

high confidence CNV

high confidence CNVR

GEN_CNV

WGS_CNV

GEN_CNVR

WGS_CNVR

overlapping CNVR

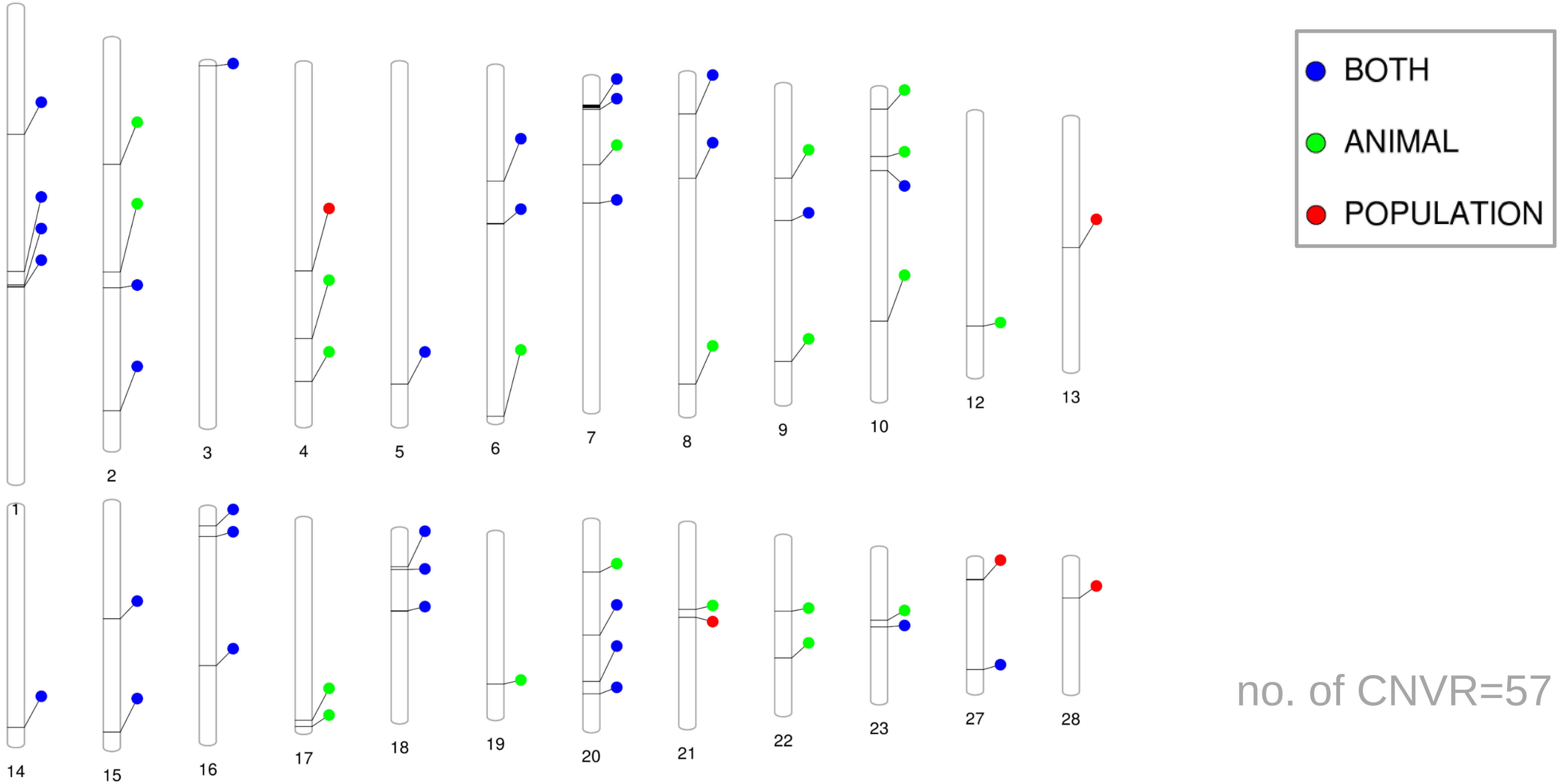
high confidence CNVR

Identified high confidence CNVR

Parameter		ANIMAL_CNVR	POPULATION_CNVR
Number of CNVR		52	36
Number of chromosomes		22	20
Type	Deletions	30	15
	Duplications	21	7
	Deletions and duplications	1	14

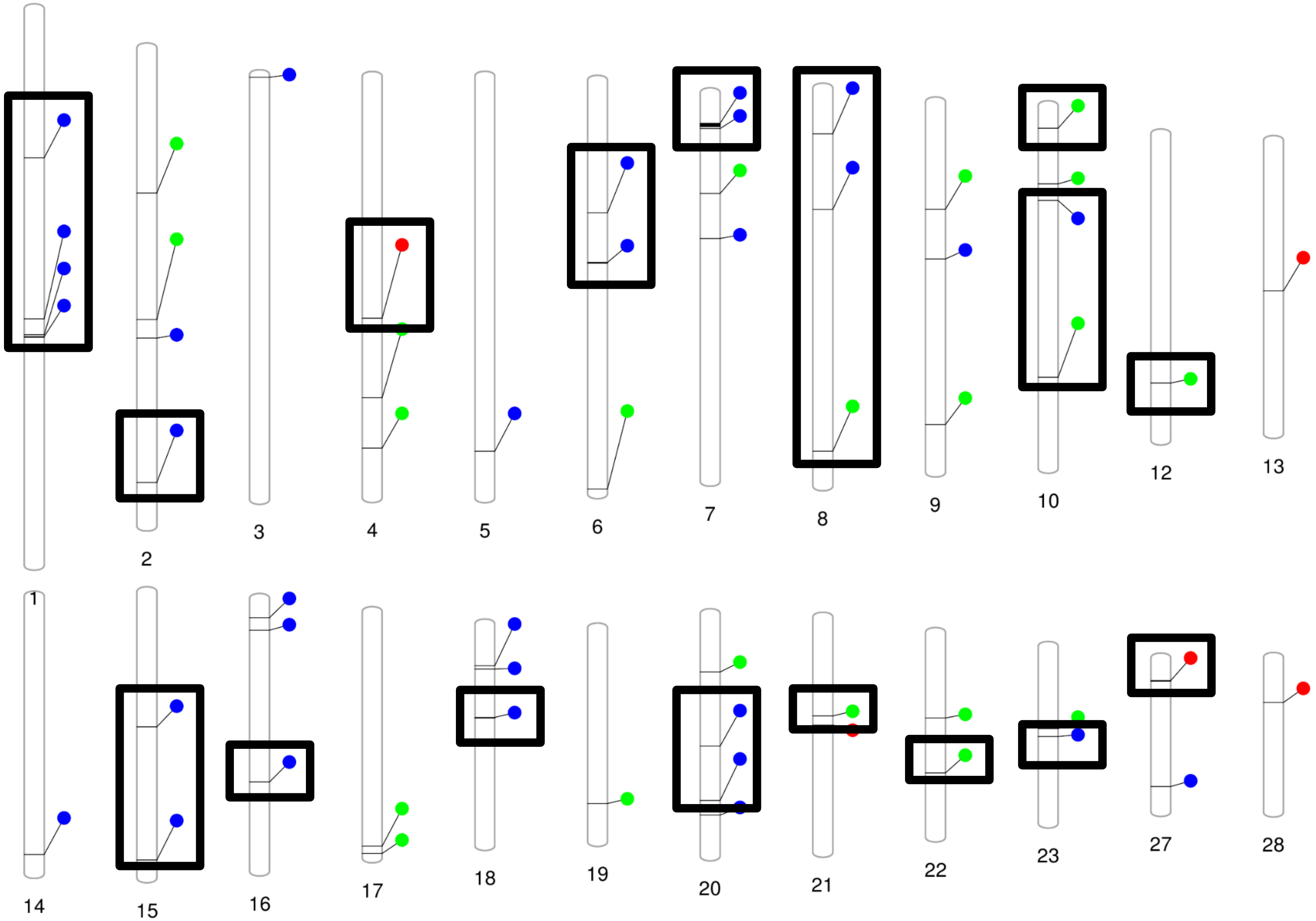
- Identified CNVR sets were similar in regard to:
  - their average lengths, and
  - the proportion of genome they covered

Unique high confidence CNVR



no. of CNVR=57

Unique high confidence CNVR



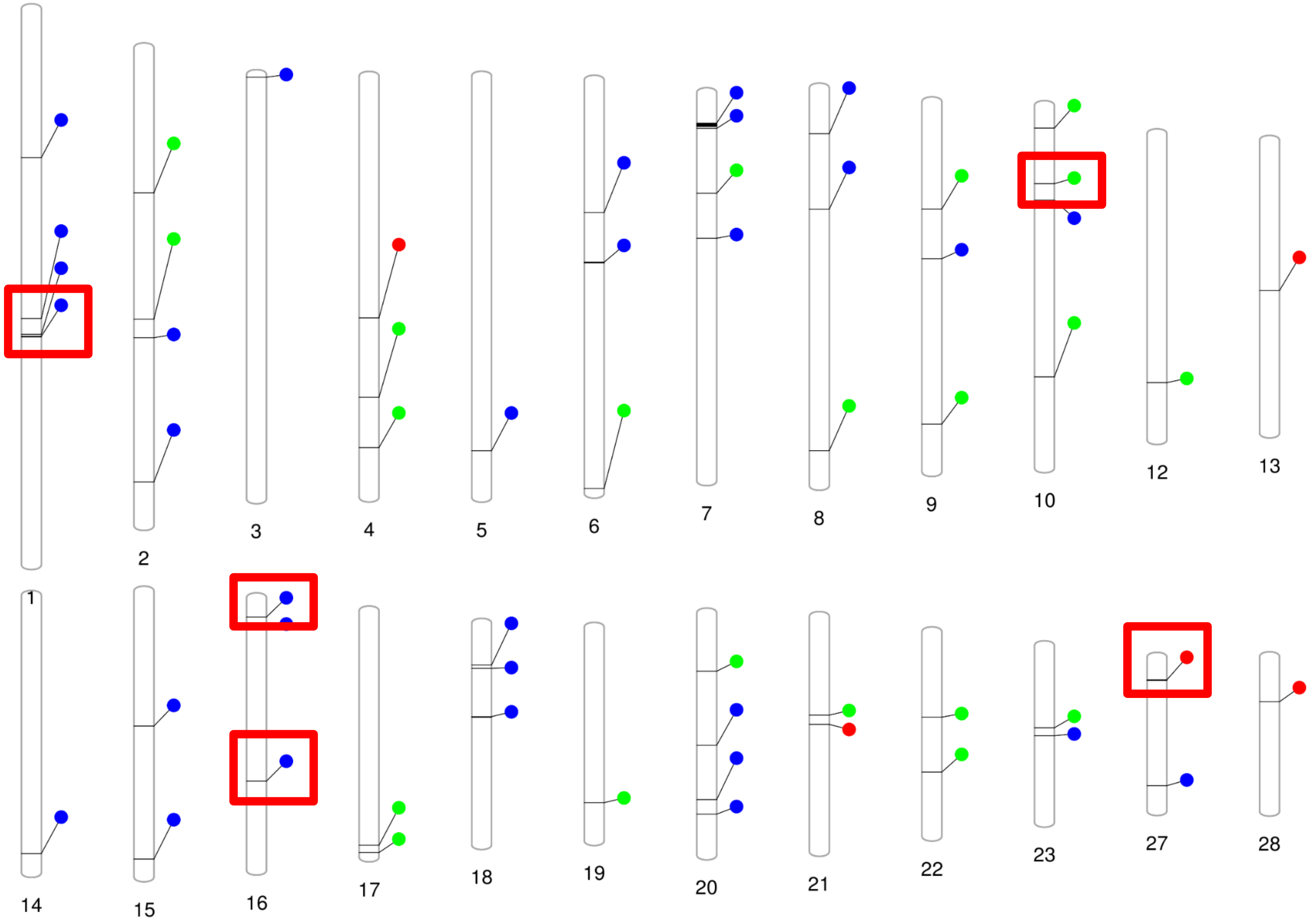
- ● BOTH
- ● ANIMAL
- ● POPULATION

Found in the Database of Genomic Variants archive

DGVarchive

no. of CNVR=57

Unique high confidence CNVR

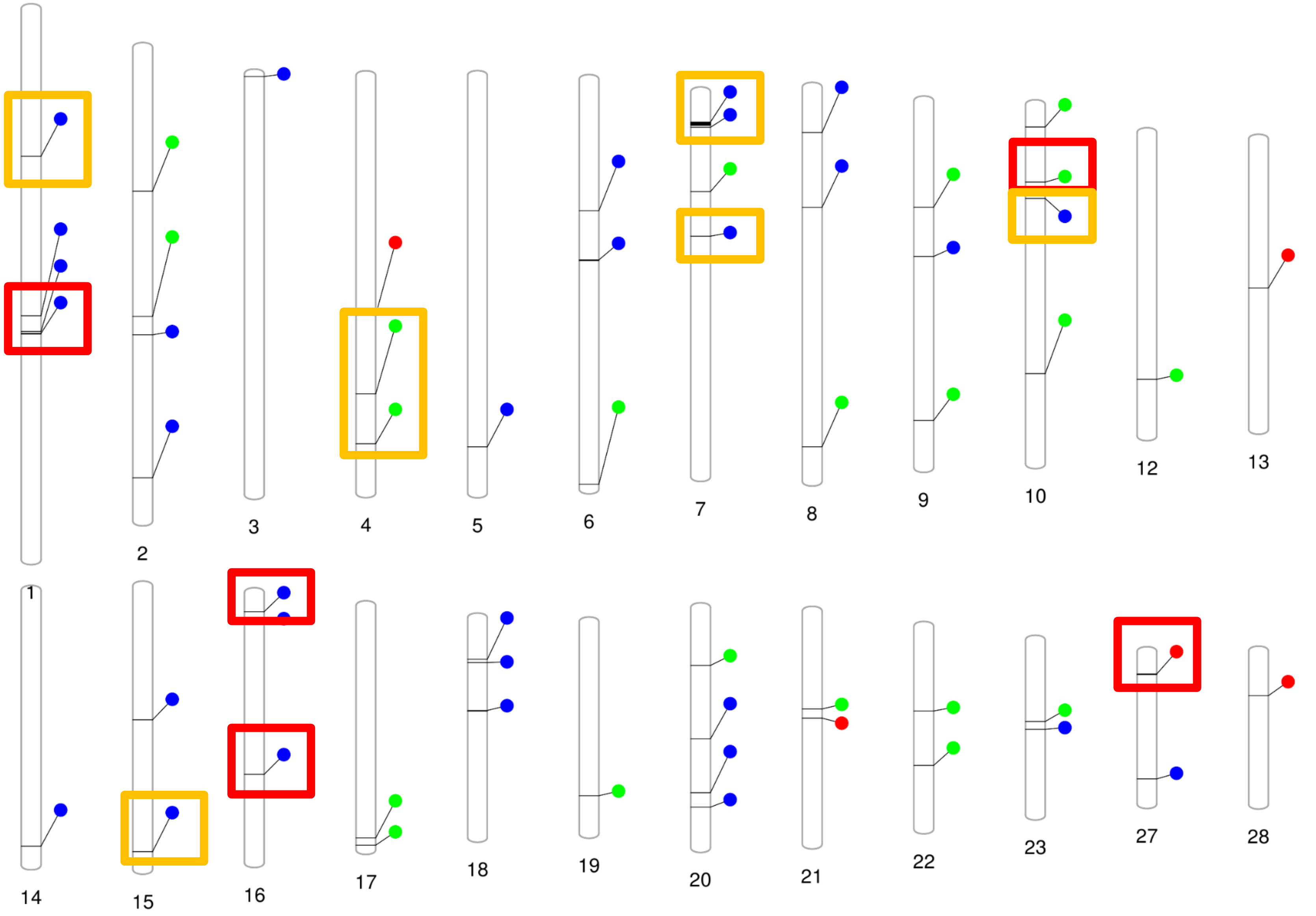


● BOTH
● ANIMAL
● POPULATION

Immunity

no. of CNVR=57

Unique high confidence CNVR



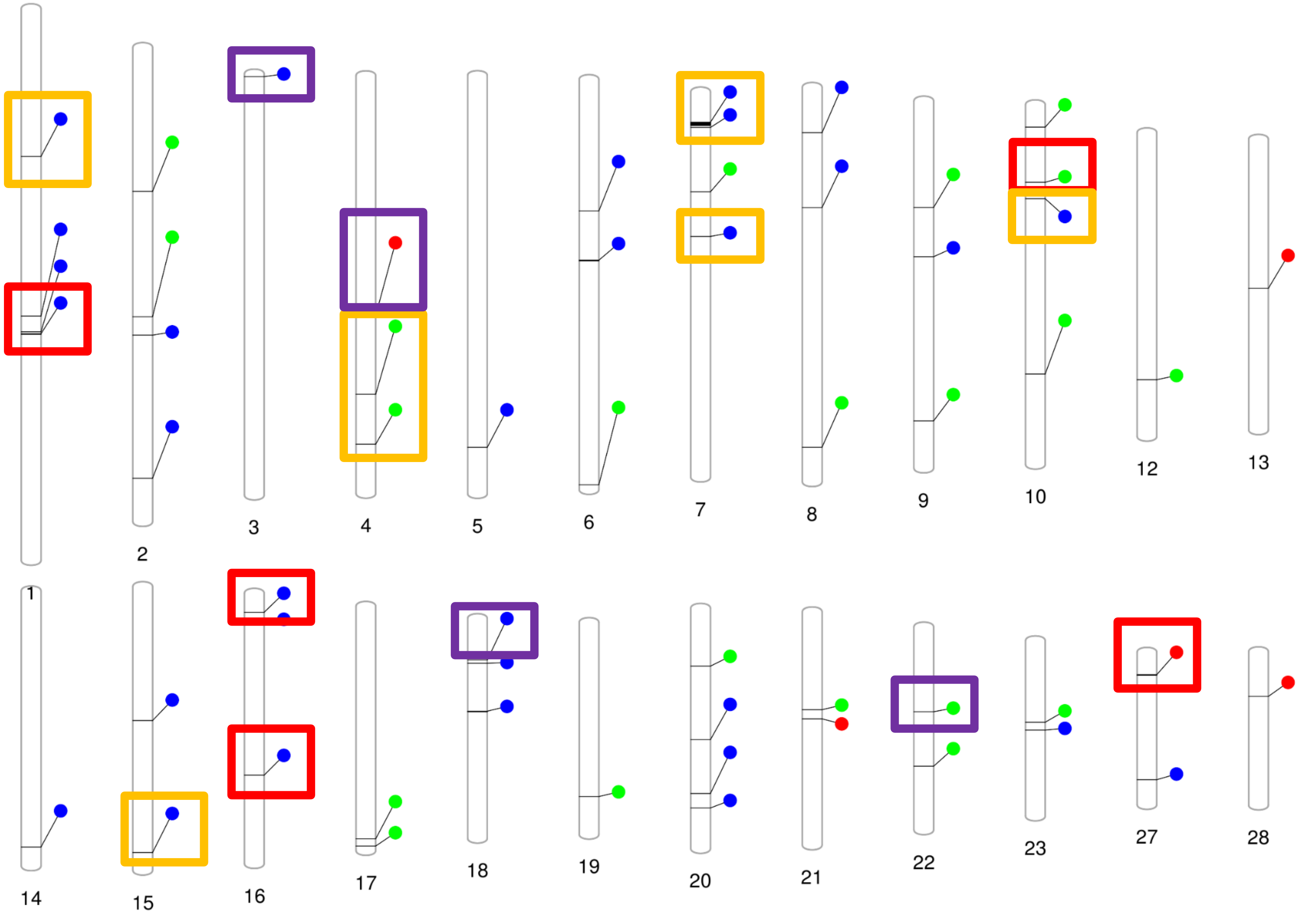
● BOTH
● ANIMAL
● POPULATION

Immunity

Olfactory sense

no. of CNVR=57

Unique high confidence CNVR



● BOTH

● ANIMAL

● POPULATION

Immunity

Olfactory sense

Transcription

no. of CNVR=57

Conclusions

- High confidence CNVR can be identified *in silico* with multiple methods and data sources

  ➔ **Although another reference assembly was used, high confidence CNVR could be matched with previously described variants**

Conclusions

- High confidence CNVR can be identified *in silico* with multiple methods and data sources

  ➔ **Although another reference assembly was used, high confidence CNVR could be matched with previously described variants**

- The identified high confidence CNVR overlap with protein coding sequences

  ➔ **Identified regions often have expected CNV characteristics**

Funders & participating organizations

THE
Efficient Dairy
GENOME PROJECT

GenomeCanada

GenomeAlberta

Ontario Genomics

USDA

UNIVERSITY of GUELPH

UNIVERSITY OF ALBERTA

AARHUS UNIVERSITY

AGRICULTURE VICTORIA

SRUC

QUALITAS

Alberta
Agriculture
and Forestry

Ontario
Ministry of Agriculture,
Food and Rural Affairs

Ontario
MINISTRY OF RESEARCH & INNOVATION

Lactanet

SEMEX®
Genetics for Life®

GrowSafe
SYSTEMS®

ALBERTA milk