

# The analysis of CNV regions in Holstein-Friesian cows



M. Frąszczak<sup>1</sup>, M. Mielczarek<sup>1,2</sup>, E. L. Nicolazzi<sup>3</sup>, G. Minozzi<sup>3</sup>, R. Giannico<sup>3</sup>, J. Williams<sup>3</sup>, K. Wojdak-Maksymiec<sup>4</sup>, J. Szyda<sup>1,2</sup>

<sup>1</sup>Biostatistics group, Wrocław University of Environmental and Life Sciences, <sup>2</sup>National Research Institute of Animal Production, <sup>3</sup>Fondazione Parco Tecnologico Padano, <sup>4</sup>West Pomeranian University of Technology

## CONCLUSIONS

- Duplication regions are longer for healthy cows
- Deletion regions are longer for mastitis cows
- Healthy cows have more common regions than sick cows
- The longest CNV regions are located in intergenic regions

## MATERIAL

- Whole genome NGS with IlluminaHiSeq2000
- Polish Holstein-Friesian cows
- 16 paternal halfsib pairs: 16 healthy (H) & 16 with clinical mastitis (S)
- 5-17 x genome average coverage

## METHODS

### Bioinformatics pipeline

- Alignment to UMD3.1 (BWA-MEM)
- Post alignment filtering (Picard & Samtools)
- CNV calling (CNVnator & Pindel)

### CNV regions analysis

- Descriptive statistics
- Jaccard and Kulzinsky similarity measure
- Mann-Whitney-Wilcoxon test

## RESULTS

The longest CNV regions:  
 • del – 1374000bp (BTA12)  
 • dup – 470700bp (BTA18)

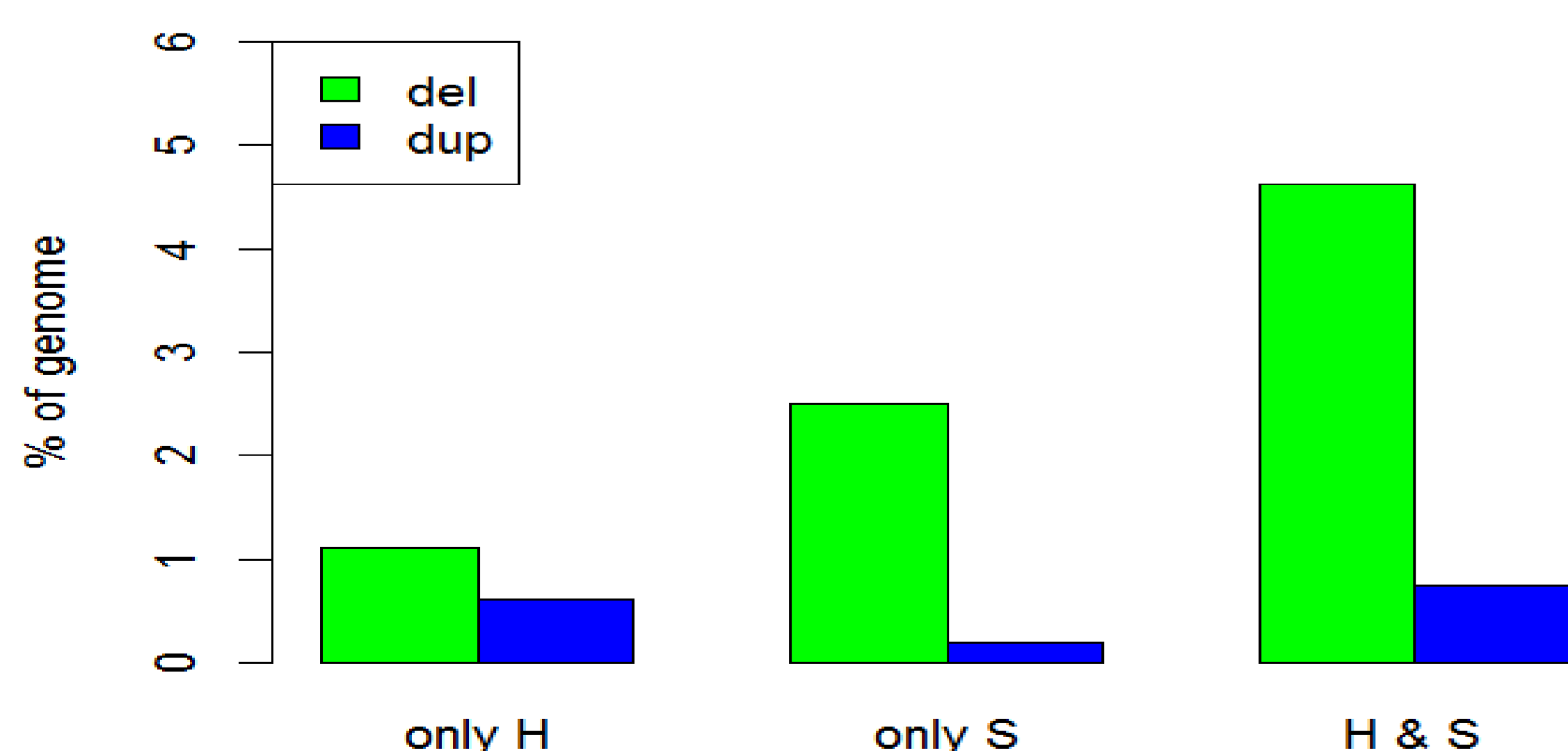
53 554 del → 12 244 del regions  
 16 309 dup → 2 556 dup regions

The shortest CNV regions:  
 • del – 400bp (BTA1-29)  
 • dup – 300bp (BTA12)

### Validation (CNVnator & Pindel)

12% deletions & 23% duplications

### Genome occupation



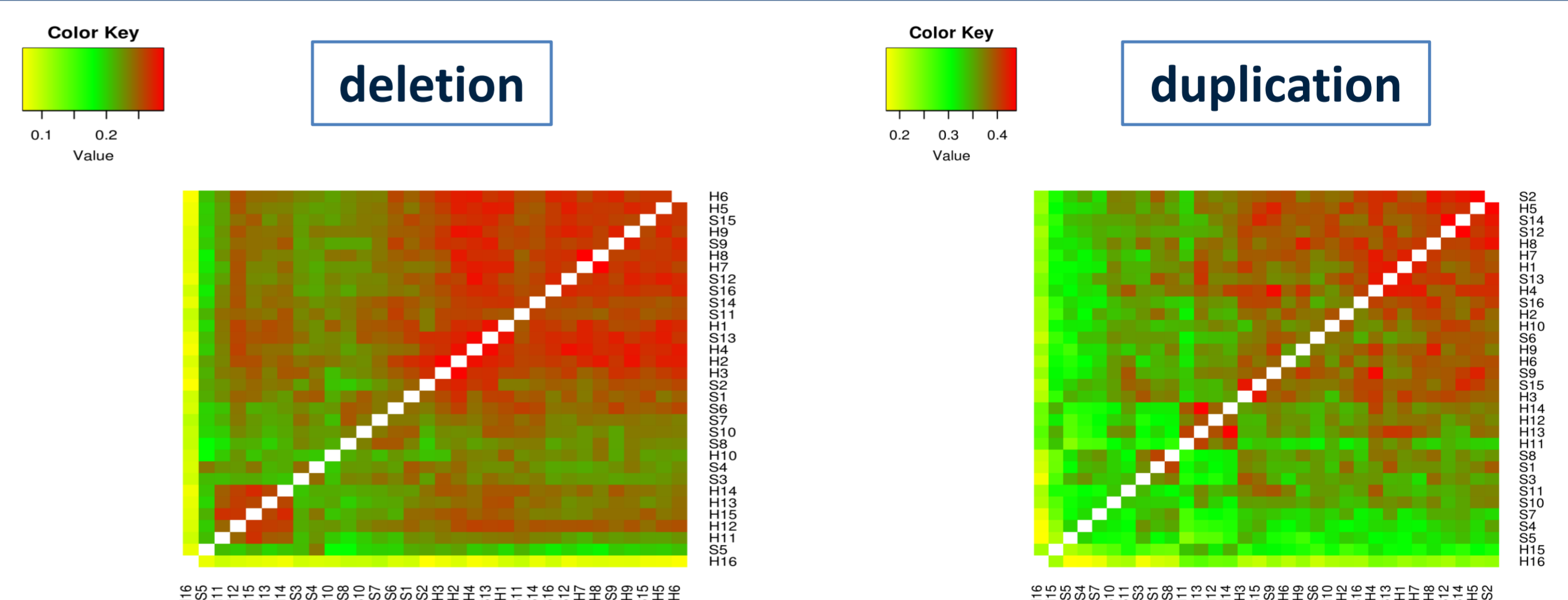
### del regions:

- longer for S cows ( $p=2.6 \cdot 10^{-34}$ )
- very similar numbers of regions (S: 7874, H: 7740)

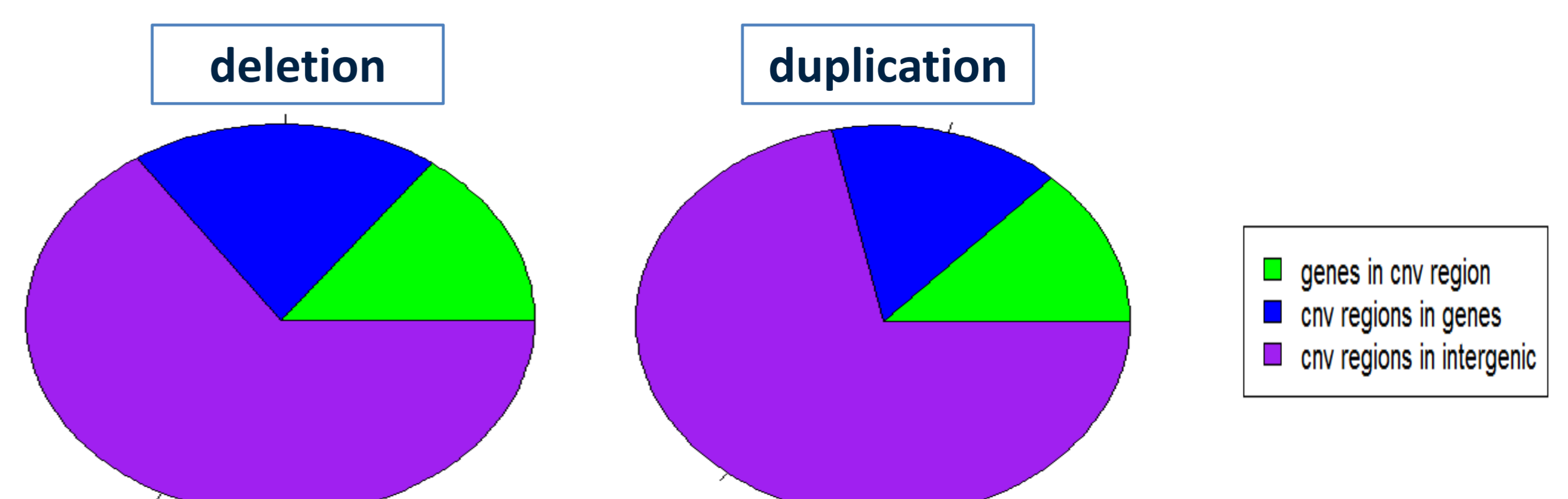
### dup regions:

- longer for H cows ( $p=1.3 \cdot 10^{-23}$ )
- more for H cows regions (S: 2028, H: 1144)

### Jaccard similarity between cows



Longer distances between sick cows than between healthy ( $p=0.007$ )



Acknowledgements: