CNV impact on gene expression

M. Mielczarek, M. Frąszczak, A. Zielak-Steciwko,
B. Nowak, B. Hofman, J. Pierścińska, J. Szyda

Wroclaw University of Environmental and Life Sciences; Kozuchowska 7, 51-631 Wroclaw, Poland

Background & Objectives

CNV impact on gene expression investigation by combining genomic and transcriptomic data:

- Gene expression quantification   (RNA-seq)
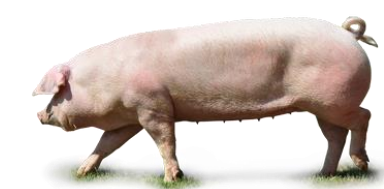- Genome-wide CNVs detection   (DNA-seq)
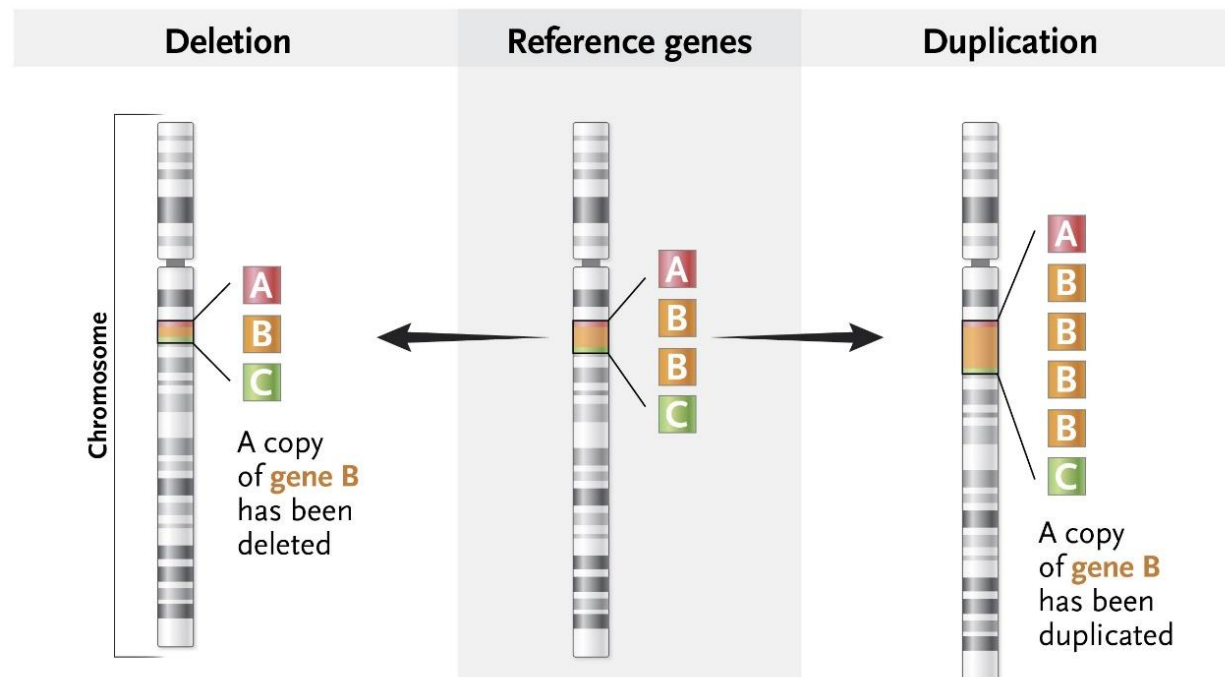
source: illustrated-glossary.nejm.org
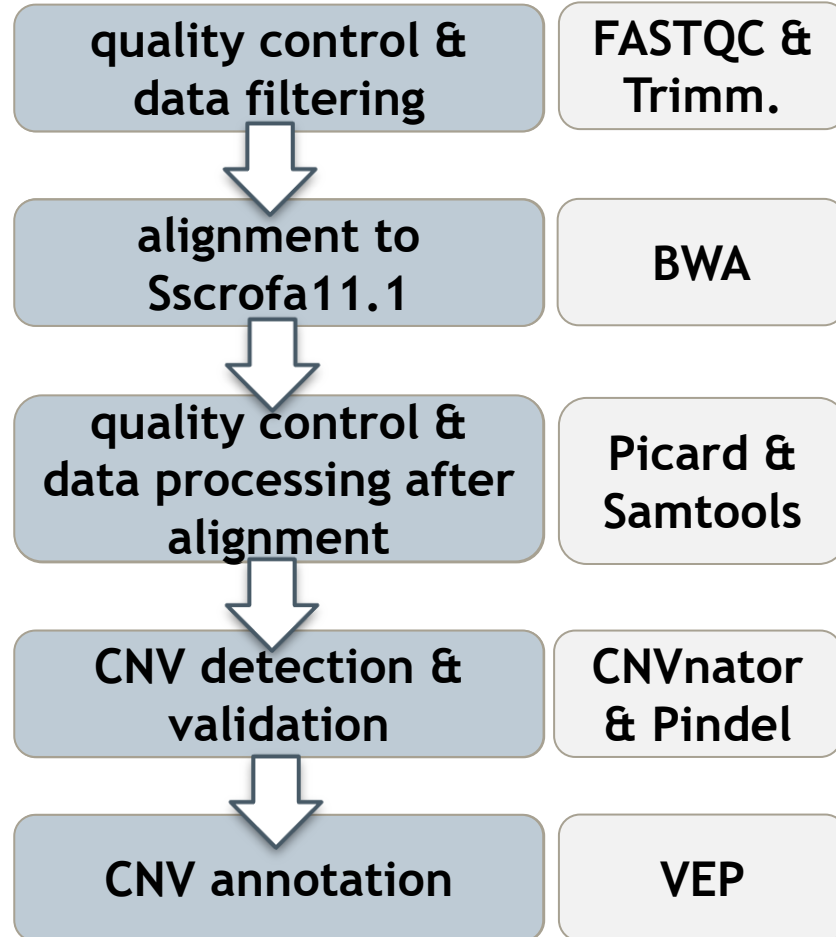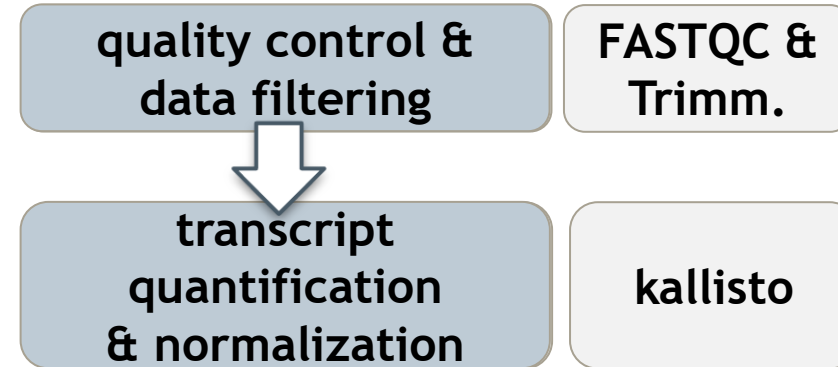
Datasets

Dataset 1

Dataset 2

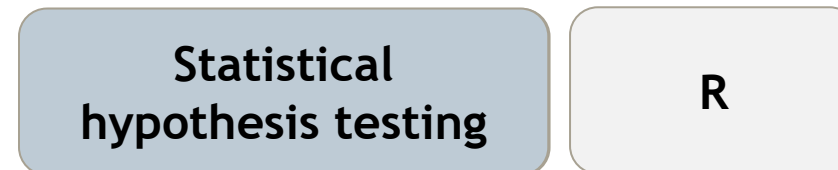| Source	PRJNA354435	PRJNA403969	Self collected data (S1-S6)
Breed	Duroc x Pietrain	Duroc x Pietrain	Polish Landrace
# individuals	1 female	143 females	6 males
DNA-seq	Illumina HiSeq 2500 150bp ×2 (PE) average coverage: 15	✖	NovaSeq 6000, 150bp x2 (PE) average coverage: ~15
RNA-seq	Illumina HiSeq 2500 150bp x2 (PE) Muscle 171,810,298 PE reads	Illumina Hiseq 2500 150bp x2 (PE) Muscle 43,873,302 – 39,455,116 PE reads	NovaSeq 6000 100bp x2 (PE) Muscle 120,802,823 - 168,508,615

THETA
Statistical Genetics Group
Institute of Animal Genetics

Methods

1. CNVs detection (DNA-seq)

quality control & data filtering	FASTQC & Trimm.

⬇

alignment to Sscrofa11.1	BWA

⬇

quality control & data processing after alignment	Picard & Samtools

⬇

CNV detection & validation	CNVnator & Pindel

⬇

CNV annotation	VEP

2. Transcript quantification (RNA-seq)

quality control & data filtering	FASTQC & Trimm.

⬇

transcript quantification & normalization	kallisto

3. CNV impact on gene expression investigation

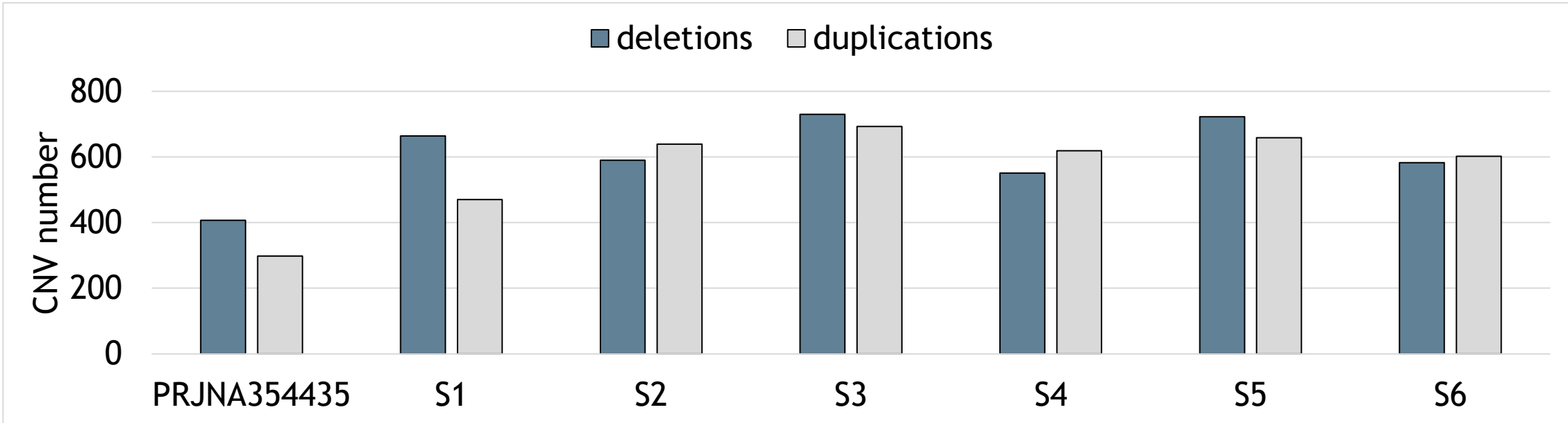|	
Statistical hypothesis testing	R

Results: CNV detection

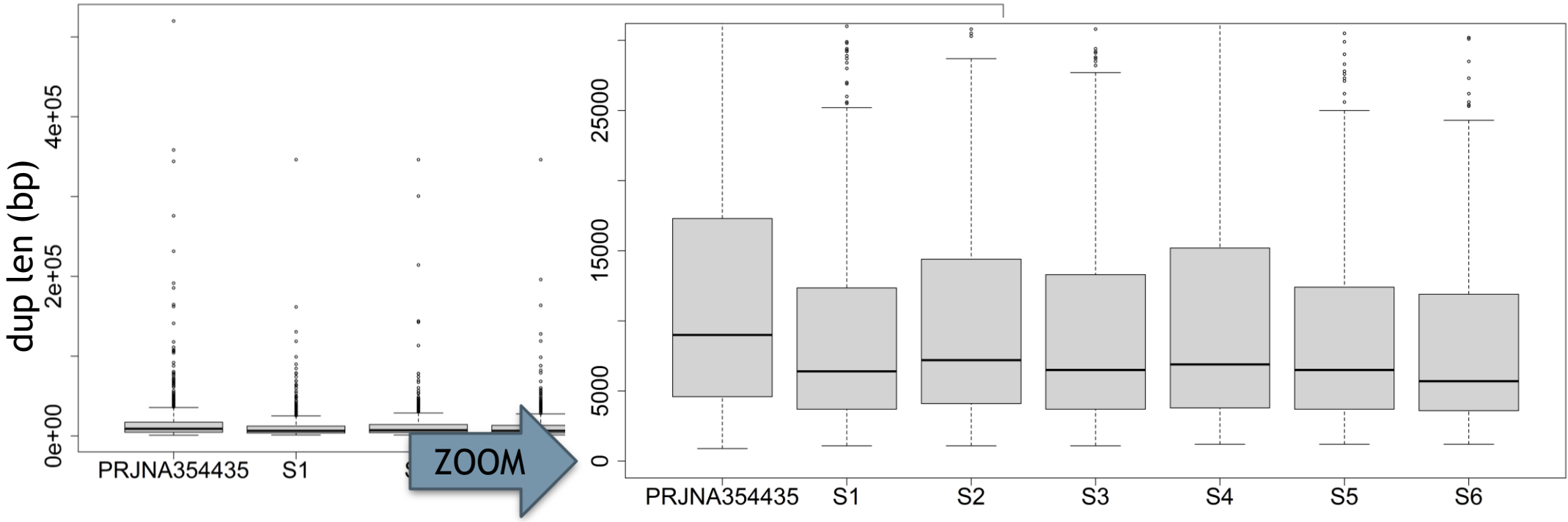**Dataset 1**

#del: 407
#dup: 298

Dataset 2

#del: 551-730
#dup: 619-693
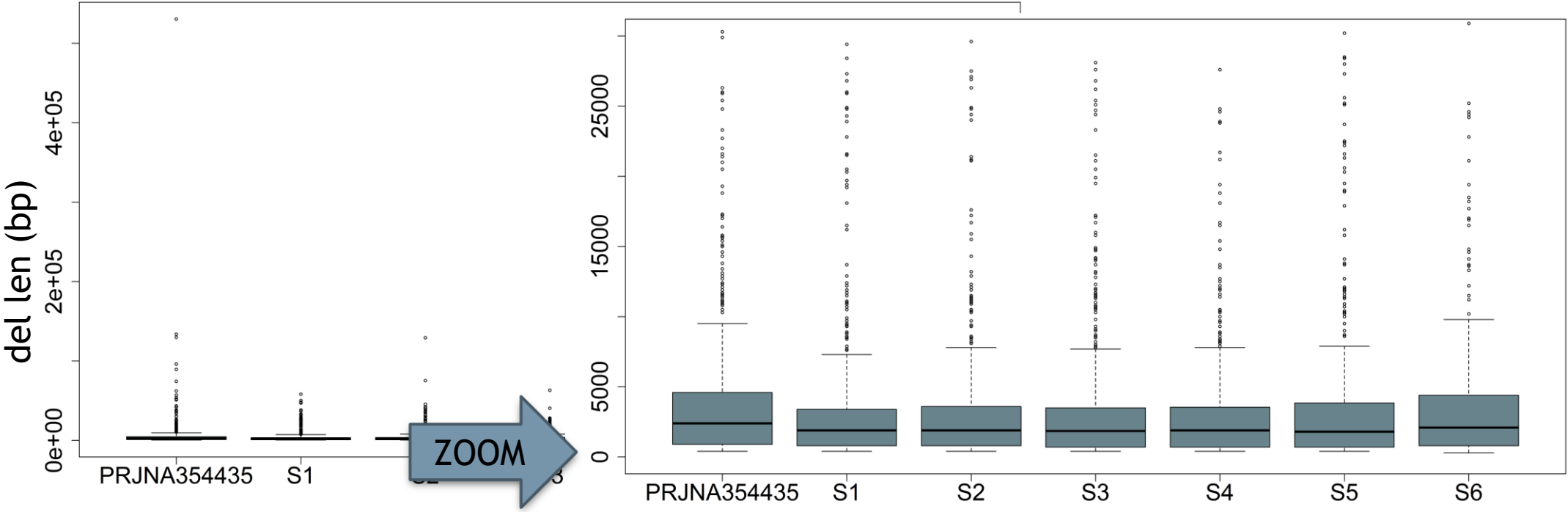
Results: CNV annotation
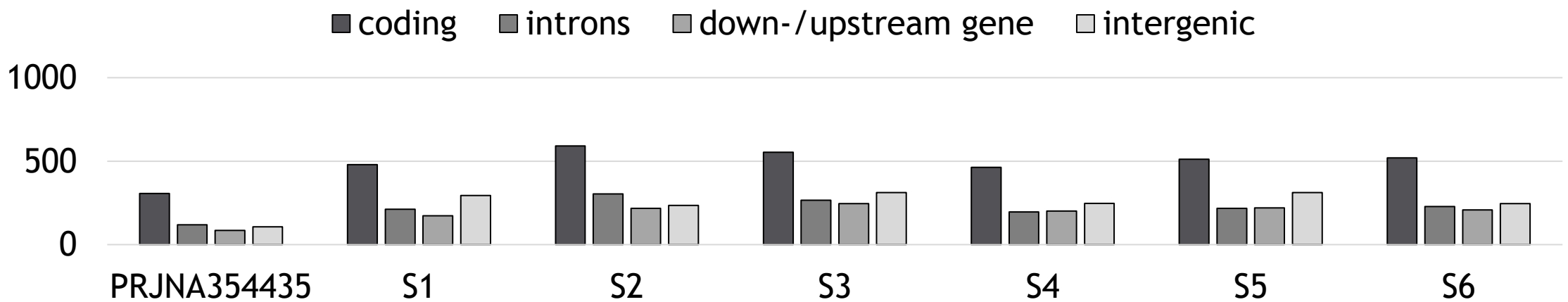
Deleted genomic regions

■ coding ■ introns □ down-/upstream gene □ intergenic

PRJNA354435   S1   S2   S3   S4   S5   S6
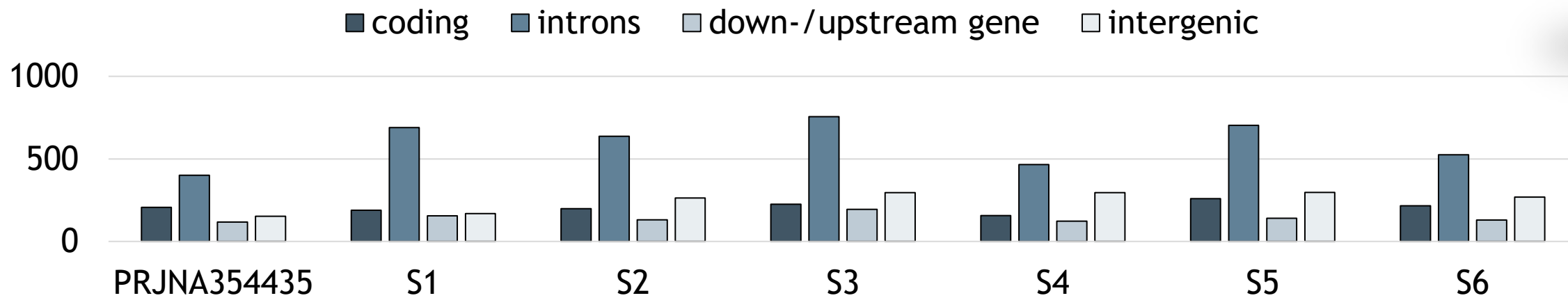
Duplicated genomic regions

■ coding ■ introns ■ down-/upstream gene □ intergenic

PRJNA354435   S1   S2   S3   S4   S5   S6

Results: Wilcoxon signed-rank test

Dataset 1

Erand$_1$ - expression of random transcripts without CNV in 1 animal.
Erand$_{143}$ - average expression of random transcripts across 143 animals.

H$_0$: Erand$_1$ = Erand$_{143}$
H$_1$: Erand$_1$ ≠ Erand$_{143}$

P=0.055

Dataset 2

Erand$_{3\_1}$ - average expression of random transcripts without CNV in 3 out of 6 animals.
Erand$_{3\_2}$ - average expression of random transcripts without CNV in 3 out of 6 animals.

H$_0$: Erand$_{3\_1}$ = Erand$_{3\_2}$
H$_1$: Erand$_{3\_1}$ ≠ Erand$_{3\_2}$

P=0.456

Results: Wilcoxon signed-rank test

Dataset 1

E_1 - expression of transcripts with deletions in 1 animal.

E_{143} - average expression of transcripts across 143 animals.

H_0: $E_1 = E_{143}$
H_1: $E_1 < E_{143}$

	P-value
coding regions	0.002
introns	$1.316 * 10^{-20}$
down-/upstream gene	0.026

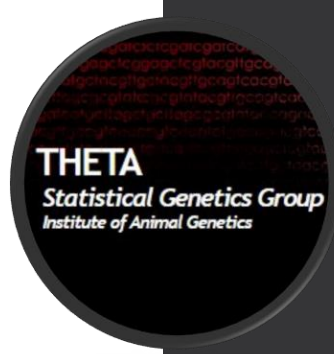## Dataset 2

E - average expression of transcripts without CNV.

E_{CNV} - average expression of transcripts with deletions.

H_0: $E = E_{CNV}$
H_1: $E < E_{CNV}$

	P-value
coding regions	0.008
introns	$1.355 * 10^{-10}$
down-/upstream gene	0.085

Results: Wilcoxon signed-rank test

Dataset 1

E_1 - expression of transcripts with duplications in 1 animal.

E_{143} - average expression of transcripts across 143 animals.

H_0: $E_1 = E_{143}$
H_1: $E_1 > E_{143}$

	P-value
coding regions	0.020
introns	0.999
down-/upstream gene	0.619

H_1: $E_1 < E_{143}$

coding regions	0.980
introns	$3.84*10^{-5}$
down-/upstream gene	0.389

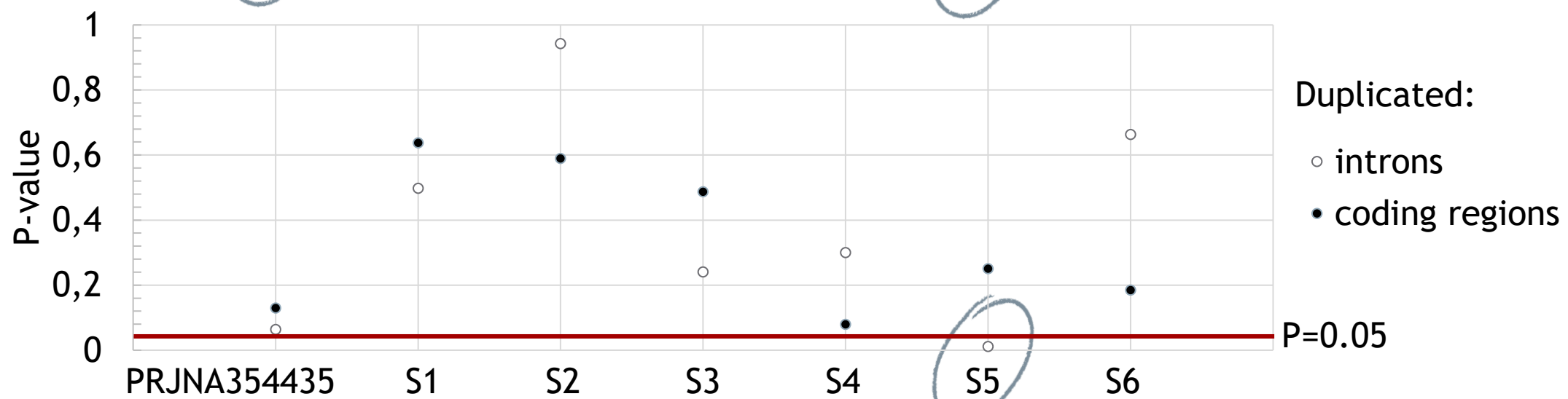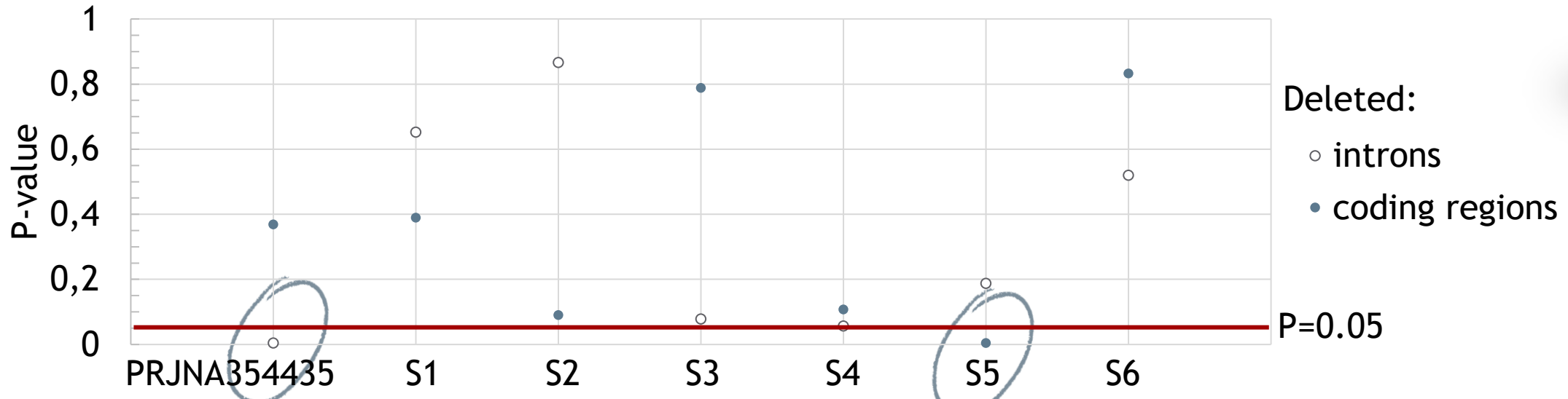## Dataset 2

E - average expression of transcripts without CNV.

E_{CNV} - average expression of transcripts with duplications.

H_0: $E = E_{CNV}$
H_1: $E > E_{CNV}$

	P-value
coding regions	0.999
introns	0.846
down-/upstream gene	0.872

H_1: $E < E_{CNV}$

coding regions	$8.318*10^{-5}$
introns	0.154
down-/upstream gene	0.129

Results: Correlation
Length of a deleted/duplicated transcripts

Conclusions

- Deletions have a potential to reduce gene expression.

- Duplications do not change gene expression.

- There is no correlation between the length of CNV and the expression level.

Acknowledgements

- This study was financed within the frame of the project: "Modulation of gene expression by Copy Number Variation in pigs" (Innovative scientist, ID: N060/0035/20).

- The computational power was provided by the Poznan Supercomputing and Networking Centre.

- Thank you for your attention!