http://theta.edu.pl/

THETA
Statistical Genetics Group
Institute of Animal Genetics

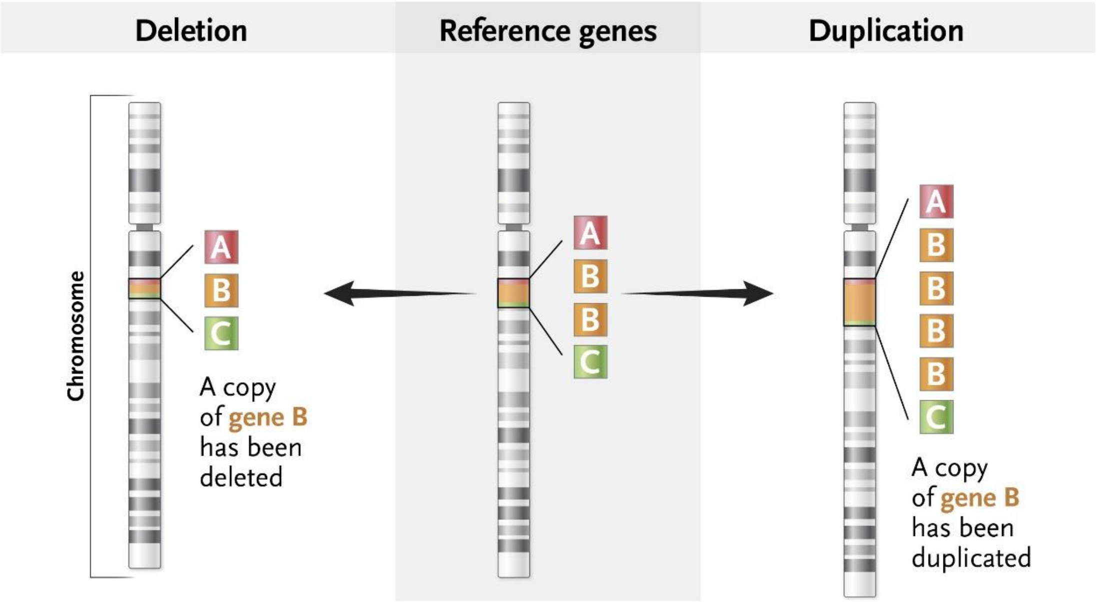
UNIVERSITAS RERUM NATURALIUM WRATISLAVIENSIS

Modulation of gene expression by CNVs

M. Mielczarek, M. Fraszczak, J.Szyda

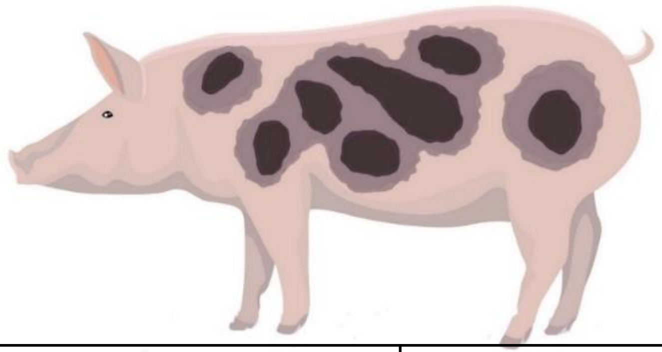
Background & Objectives

- Genome-wide CNVs detection
- **CNV impact on gene expression**



source: https://illustrated-glossary.nejm.org

Dataset



NCBI BioProject	PRJNA354435	PRJNA403969

Breed	Duroc x Pietrain	Duroc x Pietrain
# individuals	**1** female	**168** females
DNA-seq	Illumina HiSeq 2500 150×2 PE average coverage: 15	❌
RNA-seq	Illumina HiSeq 2500 **Muscle** 171,810,298 PE reads	Illumina Hiseq 2500 and 2000 **Muscle** 43,873,302 – 39,455,116 PE reads

Methods

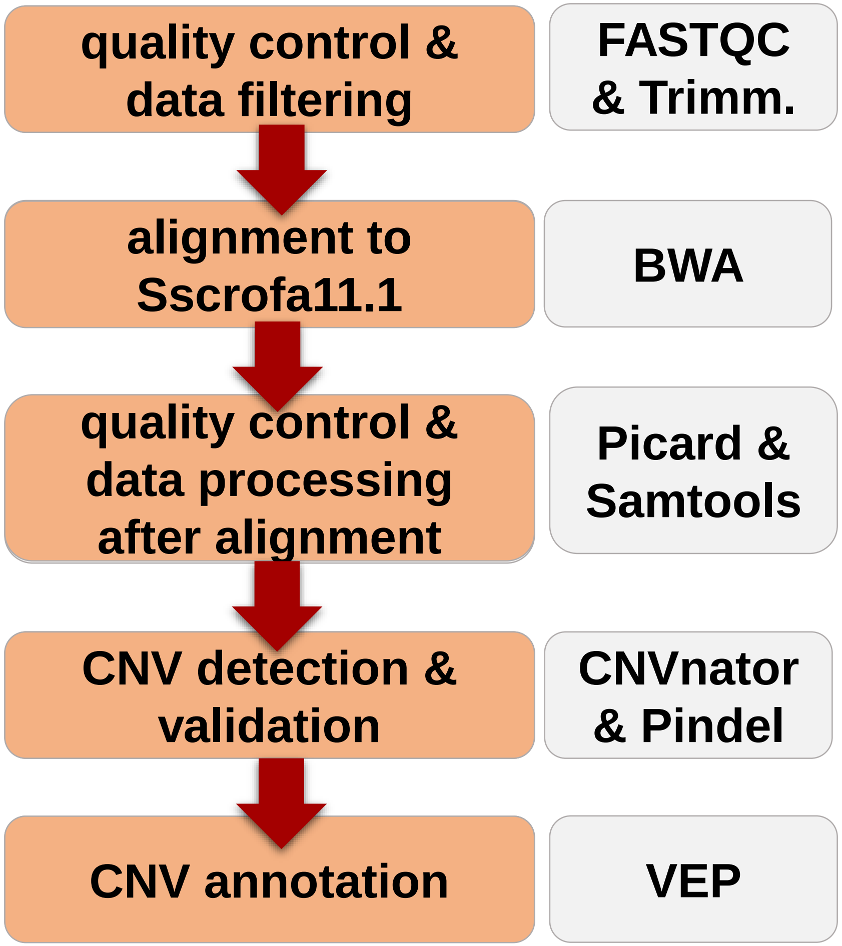
1. CNVs detection (DNA)

2. Transcript quantification (RNA)

3. CNV impact on gene expression investigation

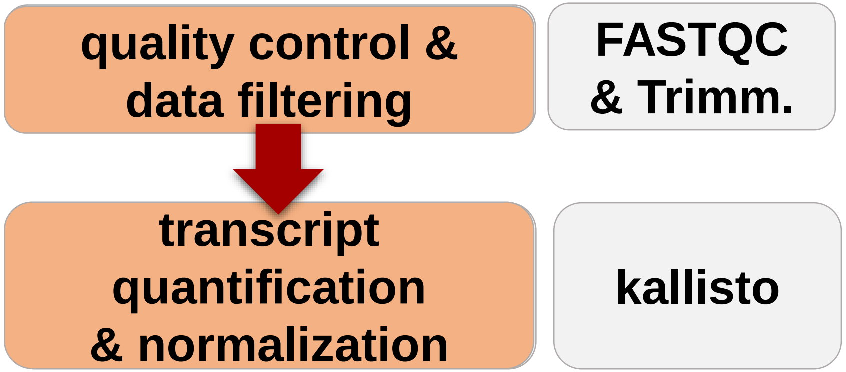
Methods

1. CNVs detection (DNA)



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)



quality control & data filtering	FASTQC & Trimm.
transcript quantification & normalization	kallisto

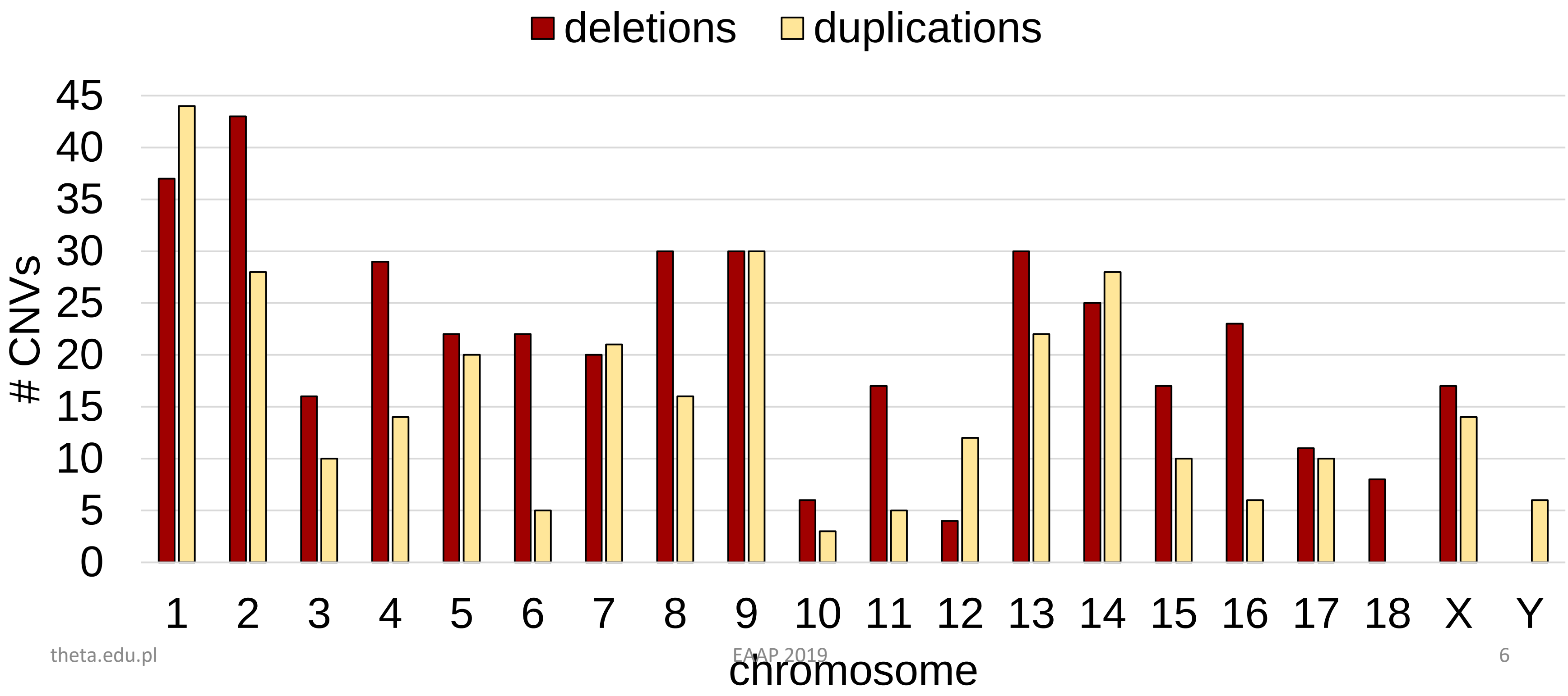
3. CNV impact on gene expression investigation

Statistical analysis	R

Results

CNVs characterization → **the numer of CNVs**

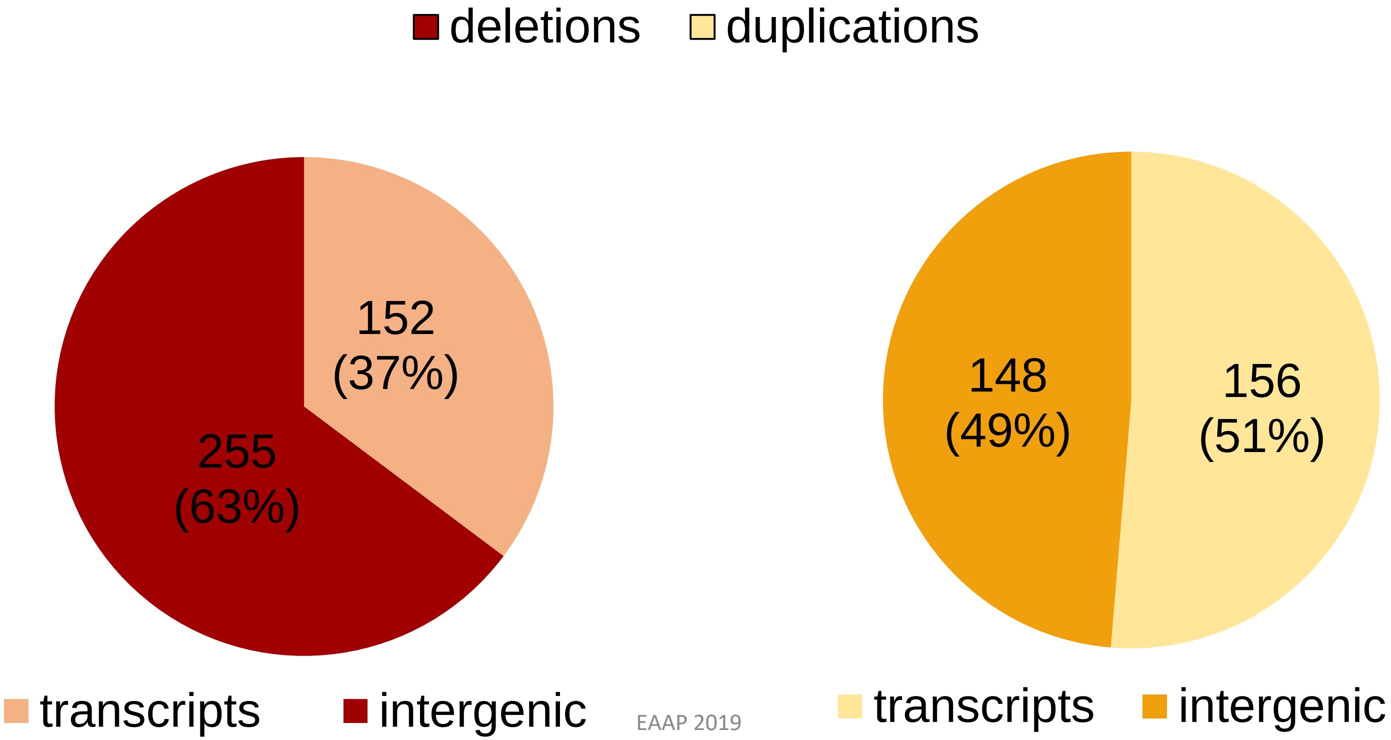
- 407 deletions
- 304 duplications



Results

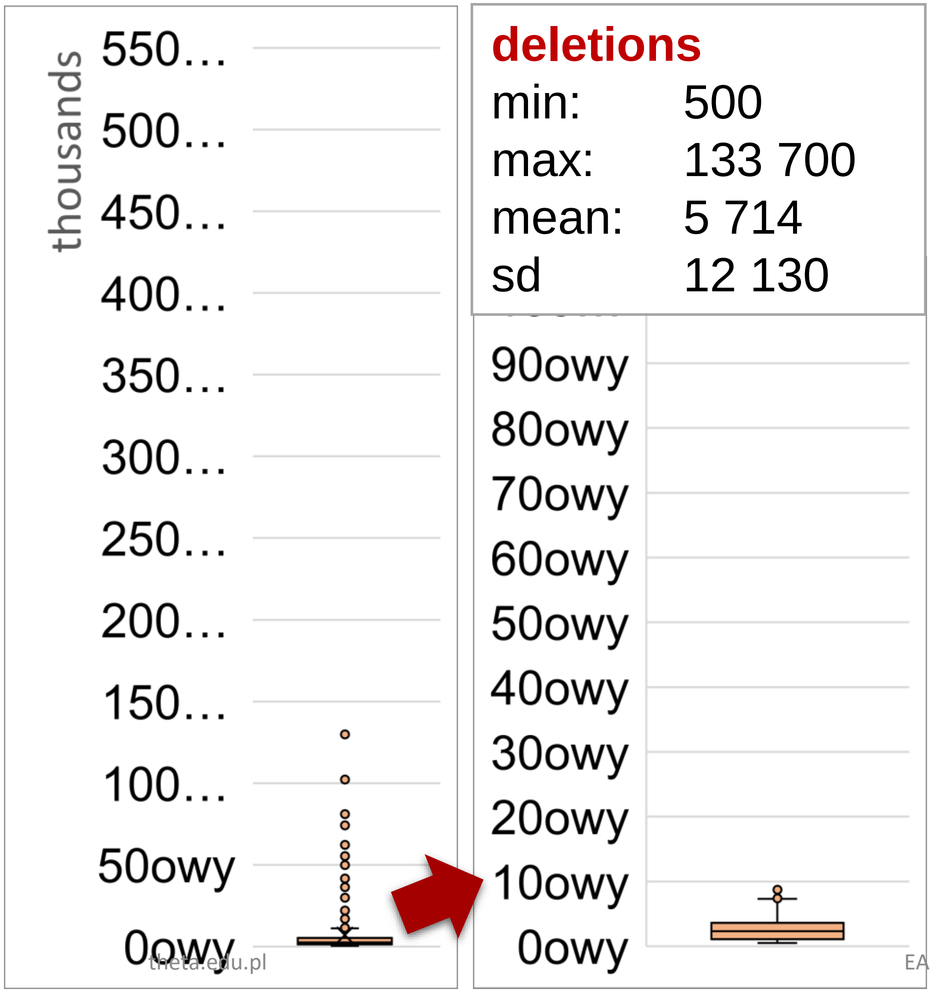
CNVs characterization → **the numer of CNVs**

- 407 deletions
- 304 duplications



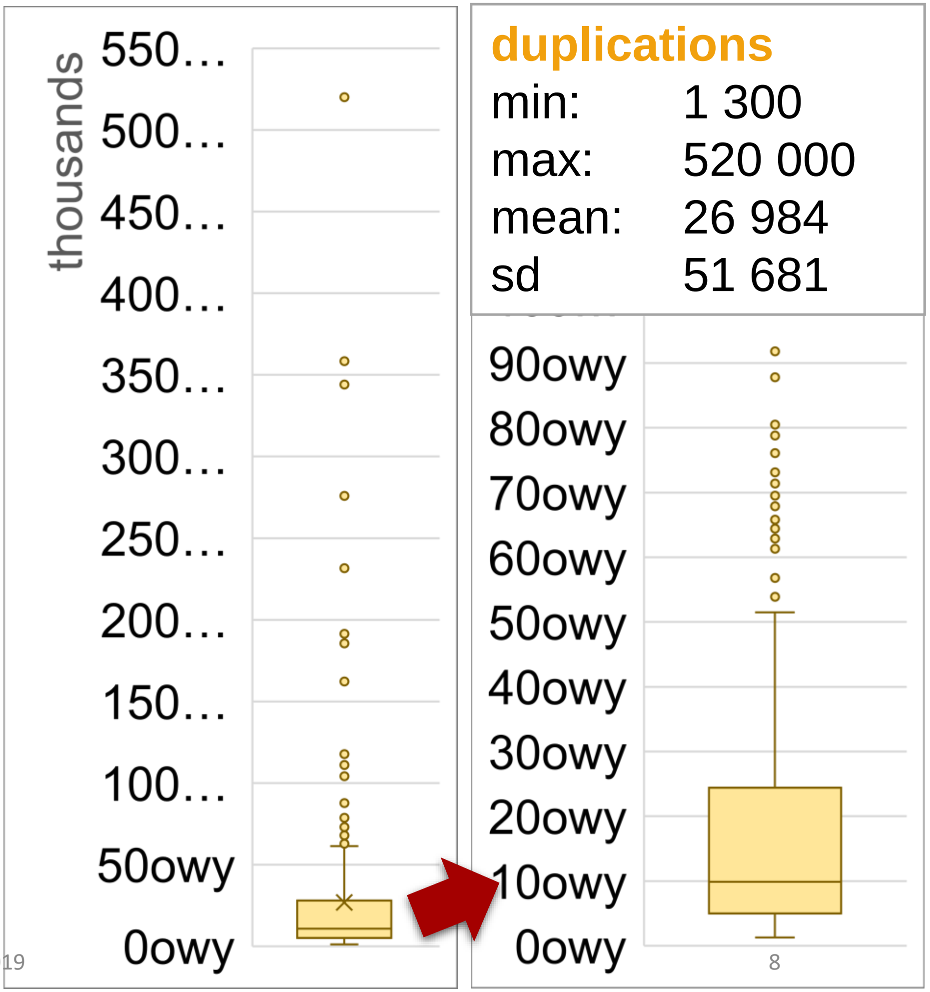
Results

CNVs characterization → **CNVs length (bp)**



deletions

min:	500
max:	133 700
mean:	5 714
sd	12 130



duplications

min:	1 300
max:	520 000
mean:	26 984
sd	51 681

Results

Deletions impact → **Wilcoxon signed-rank test**

$H_0\colon Y_{CNV} \geq \bar{Y}$
$H_1\colon Y_{CNV} < \bar{Y}$

Y_{CNV} - expression of a transcript overlaping with CNV
$\bar{Y}$ - population averaged expression

→*Transcript expression is lower if deletion appears (p = 0.0088)*

$H_0\colon Y_R \geq \bar{Y}_R$
$H_1\colon Y_R < \bar{Y}_R$

Y_R - expression of a random transcript with no CNV
$\bar{Y}_R$ - population averaged expression of a random transcript with no CNV

→ *Transcript expression is not lower in the investigated sow (p = 0.217)*

Results

Duplications impact → **Wilcoxon signed-rank test**

H_0: $Y_{CNV} \leq \bar{Y}$
H_1: $Y_{CNV} > \bar{Y}$

Y_{CNV} - expression of a transcript overlaping with CNV
$\bar{Y}$ - population averaged expression

→ Transcript expression is not higher if duplication appears (p = 0.9885)

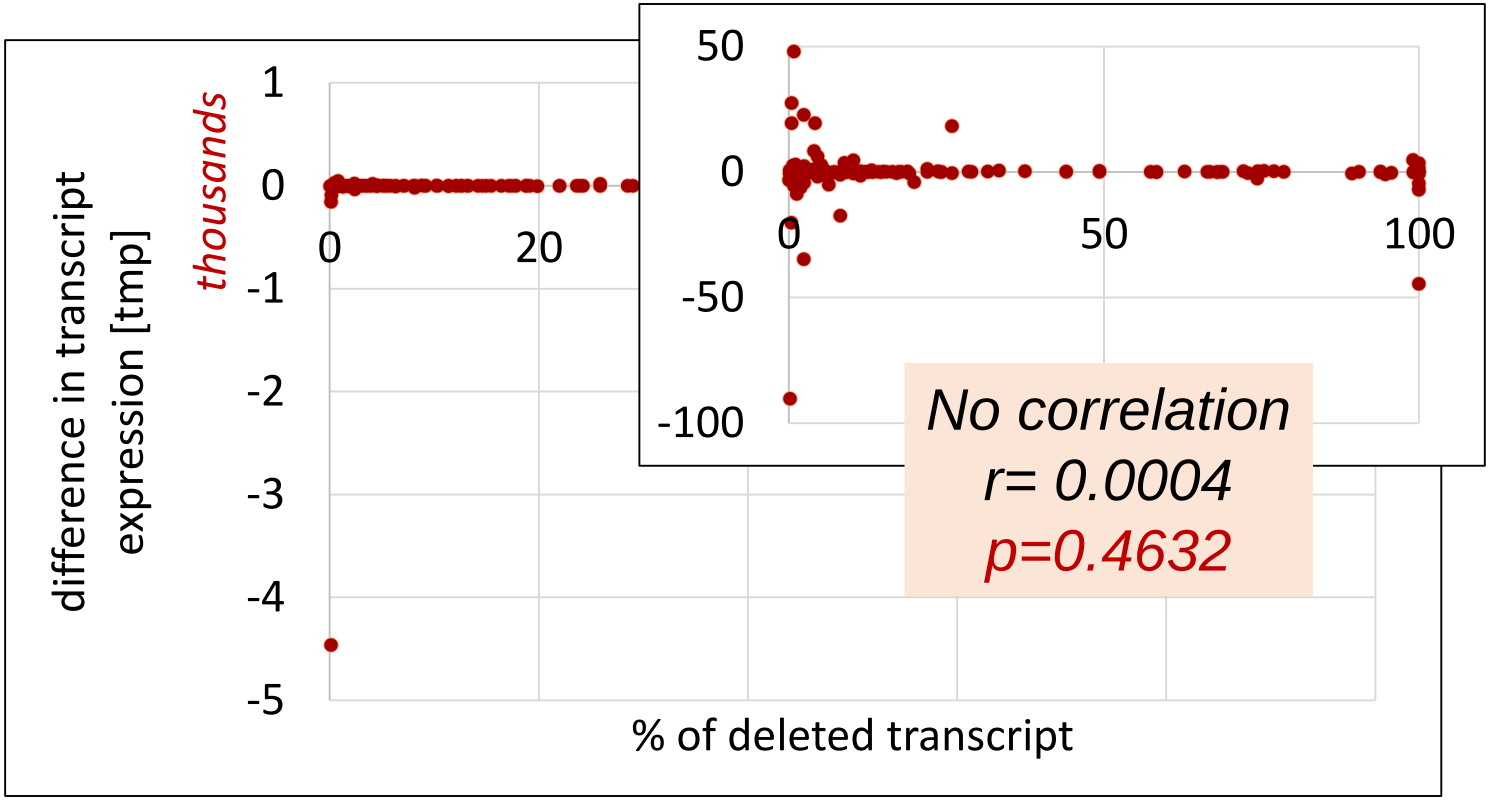
H_0: $Y_R \leq \bar{Y}_R$
H_1: $Y_R > \bar{Y}_R$

Y_R - expression of a random transcript with no CNV
$\bar{Y}_R$ - population averaged expression of a random transcript with no CNV

→ Transcript expression is not higher in the investigated sow (p = 0.7556)

Results

CNVs impact → **size of a deleted transcript**



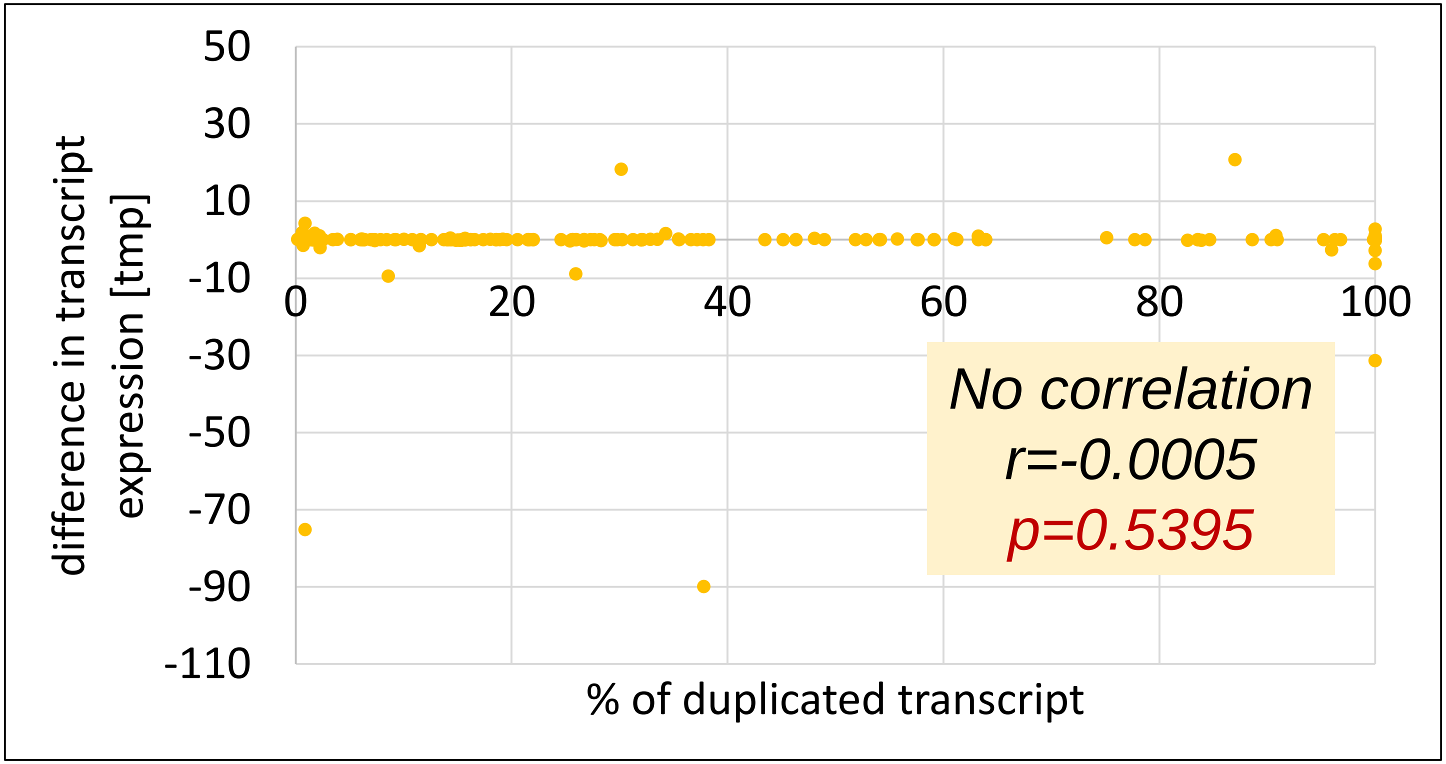
No correlation

r= 0.0004

p=0.4632

Results

CNVs impact → **size of duplicated transcript**



No correlation
r=-0.0005
p=0.5395

Conclusions

- Deletions have a potential to reduce gene expression
- Duplications does not change gene expression

- There is no correlation between the size of deletions/duplications and gene expression level

Acknowledgements

Leading National Research Centre in Poland

Poznan Supercomputing and Networking Center

Polish National Science Centre

Thank you!

supplement

CNVs impact

Wilcoxon signed-rank test

Let **D=** $\mathbf{E_1 - E_{168}}$,where ($\mathbf{E_1}$, $\mathbf{E_{168}}$) is a vector of paired observations of transcript expression in a single swine ($\mathbf{E_1}$) where CNV was detected and average expression of a transcript for 168 swines ($\mathbf{E_{168}}$)

H_0: D = -D

H_1: D < -D

Wilcoxon signed-rank test statistics

$$W = \sum_{j=1}^{k} \mathrm{r}(D_j^{+})$$

Where ***k*** denoted the total number of considered transcripts, ***D*** denoted the vector of differences of expression values between transcripts overlapping with del in the investigated pig and the average expression calculated based on 168 pigs.

r(D+) denote the rank of possitive values of difference $\mathbf{E_1 - E_{168}}$