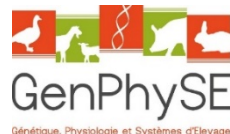


# 3D genome conformation and gene expression in fetal pig muscle at late gestation

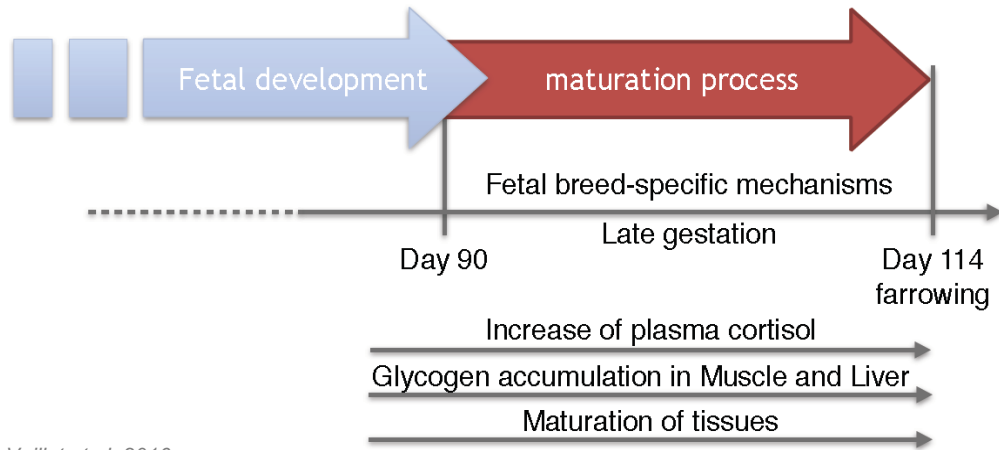
Maria Marti Marimon

4 December 2019



# Agronomic interest

- Factors responsible of piglets mortality: weight, genotype and **maturity**



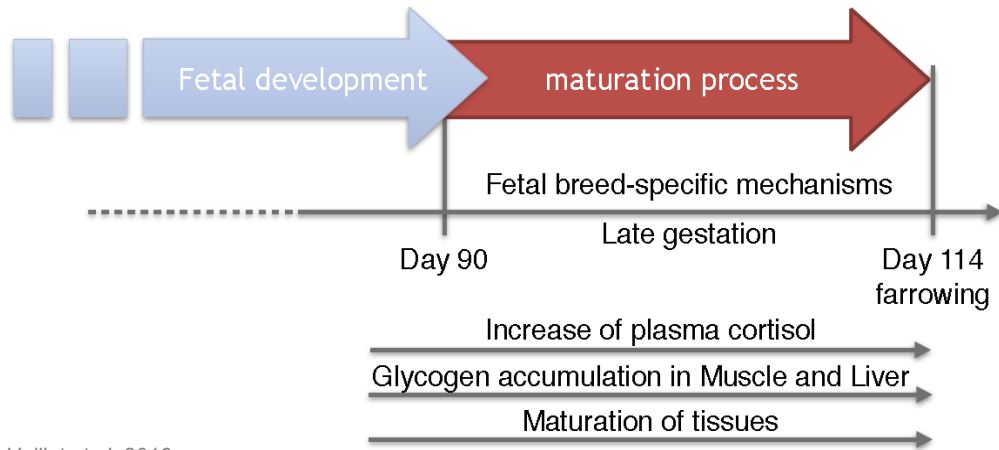
*Voillet et al. 2016*

## Maturity of fetal muscle

- Motor functions
- Thermoregulation

# Agronomic interest

- Factors responsible of piglets mortality: weight, genotype and **maturity**

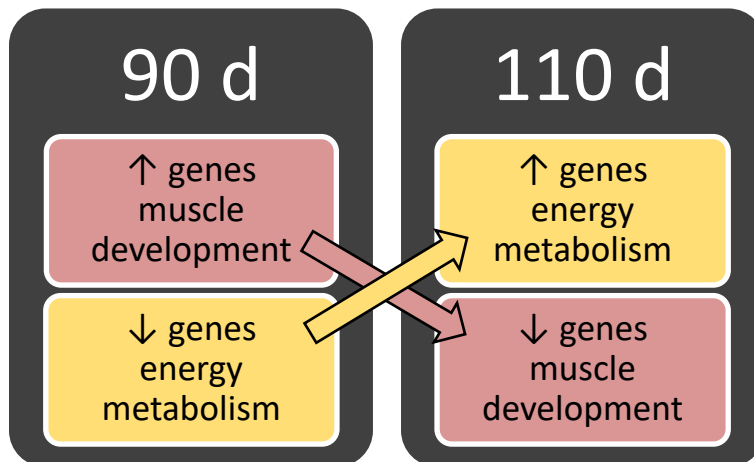


*Voillet et al. 2016*

## Maturity of fetal muscle

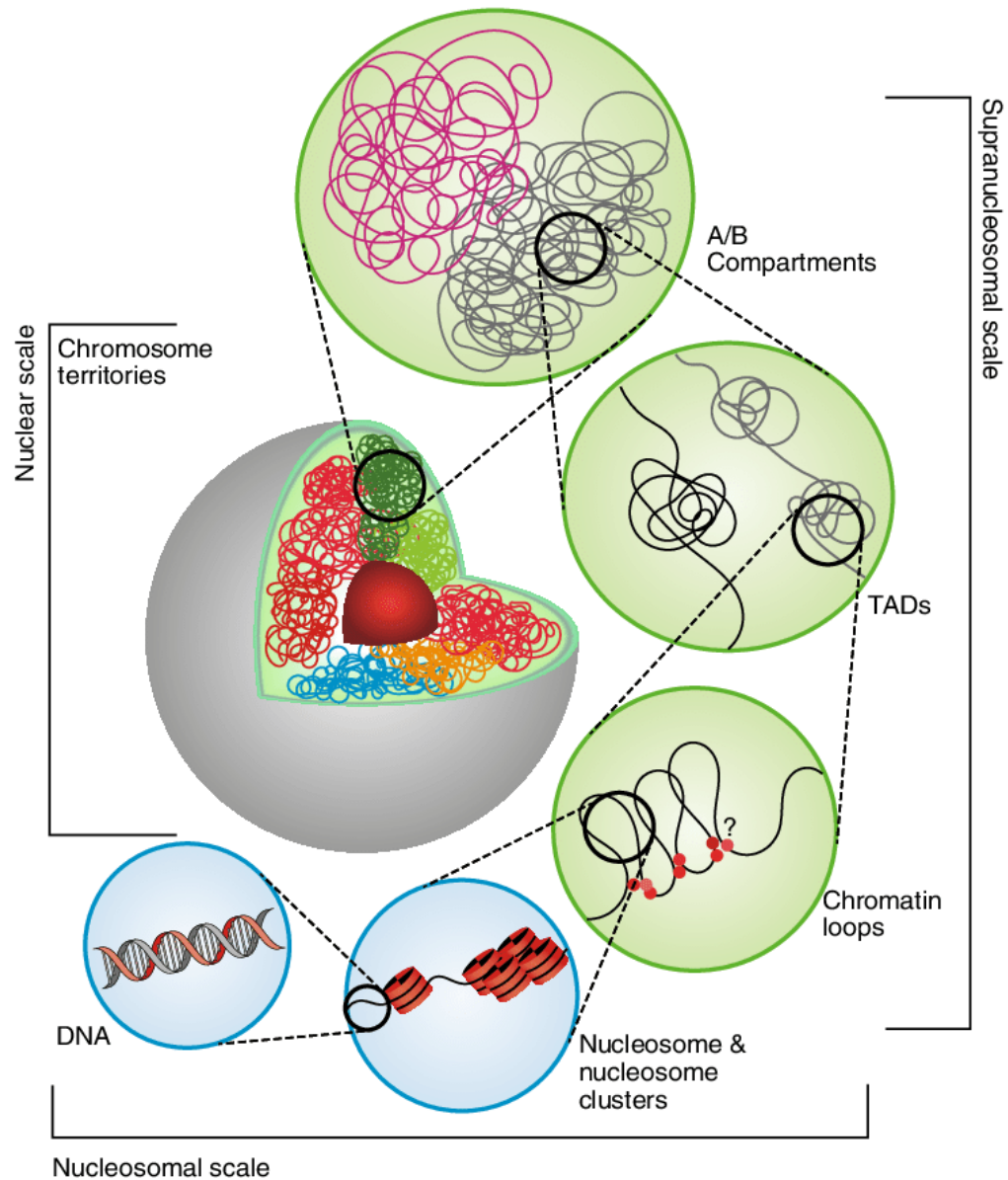
- Motor functions
- Thermoregulation

- Muscle transcriptome study (*Voillet et al. BMC Genomics, 2014*)

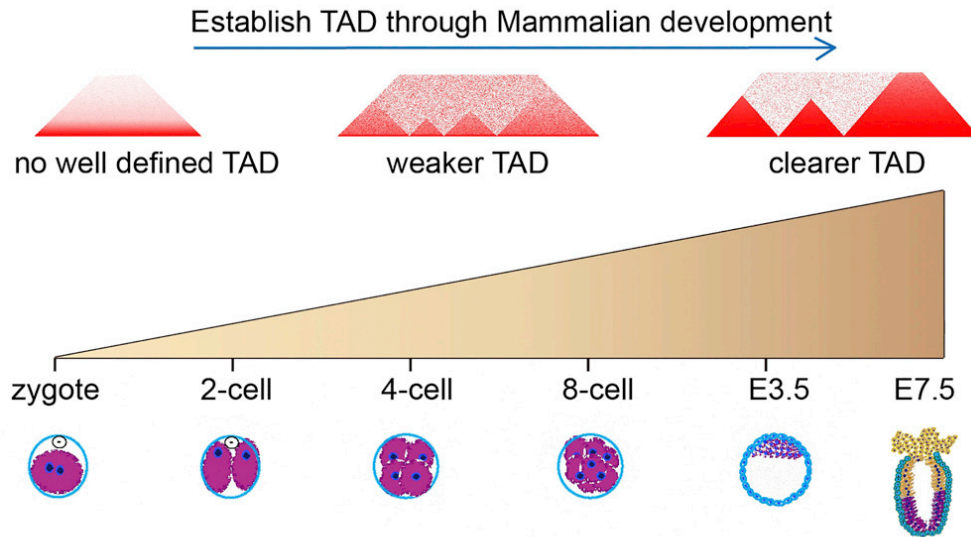


Transcriptional change associated to 3D genome organization?

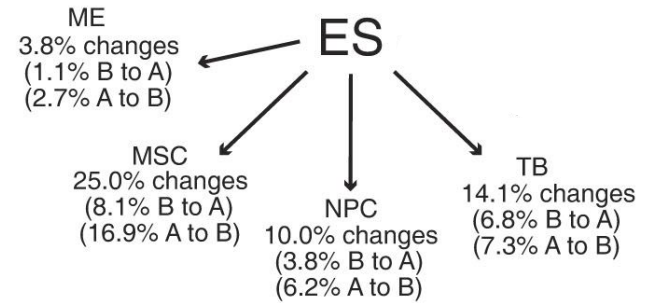
# 3D genome architecture



# 3D Genome dynamics during early development



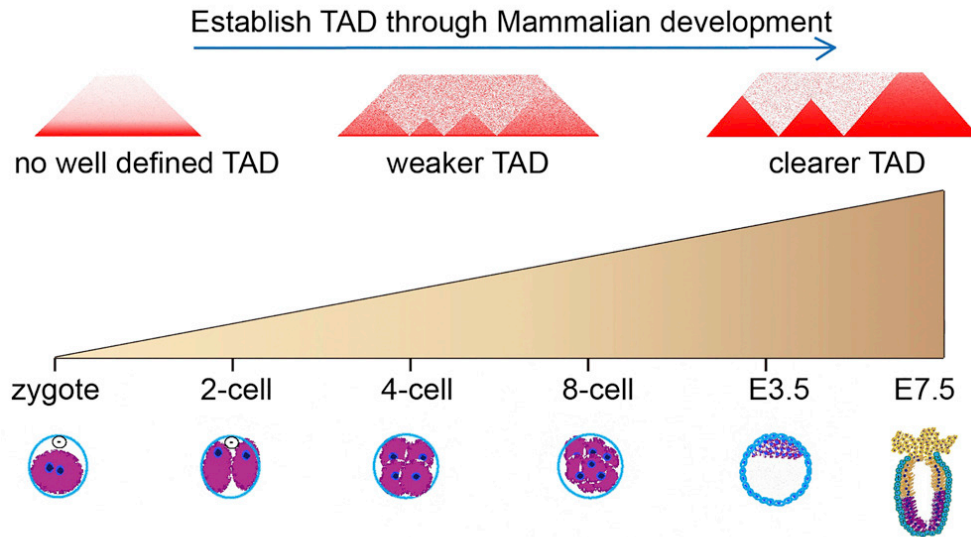
Ke Y., et al., Cell. 2017



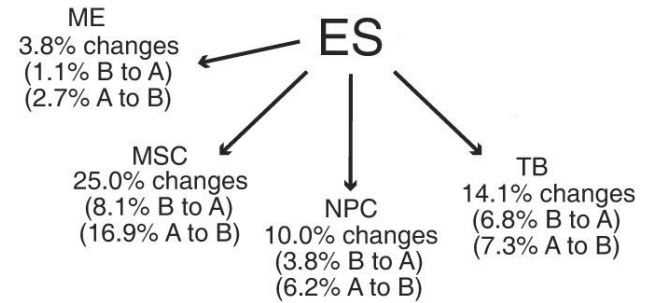
In total, 36% of the genome changes compartment in at least 1 cell type

Dixon JR, et al., Nature 2015

# 3D Genome dynamics during early development



Ke Y., et al., Cell. 2017



In total, 36% of the genome changes compartment in at least 1 cell type

Dixon JR, et al., Nature 2015

Zygote genome activation

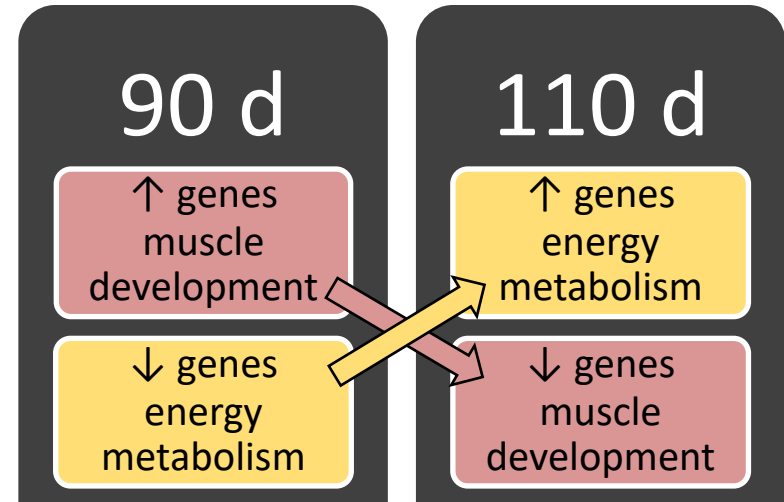
Cell differentiation  
expression programs

# Experimental design

➤ **Gene expression** (*Voillet et al. 2014*)

90 days gestation

110 days gestation



# Experimental design

## ➤ Gene expression (Voillet et al. 2014)

90 days gestation

110 days gestation

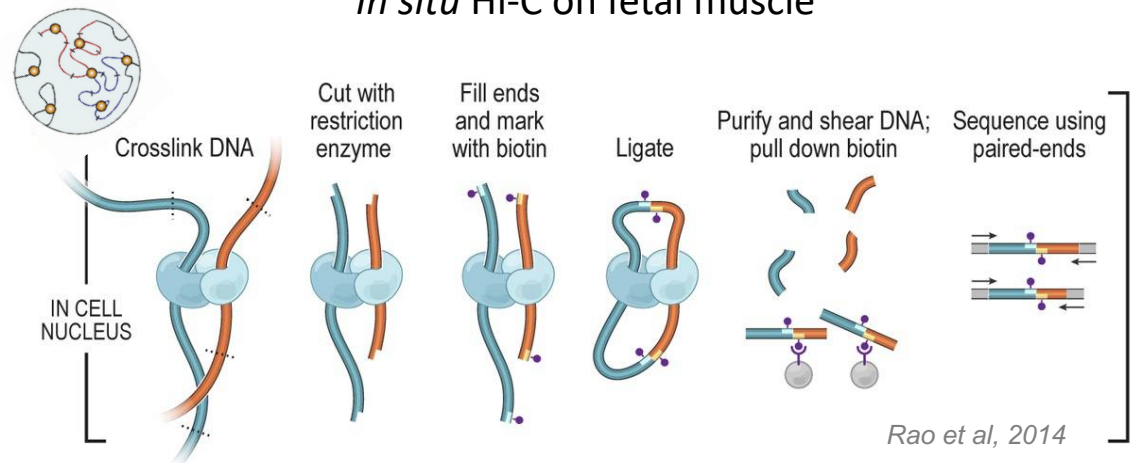
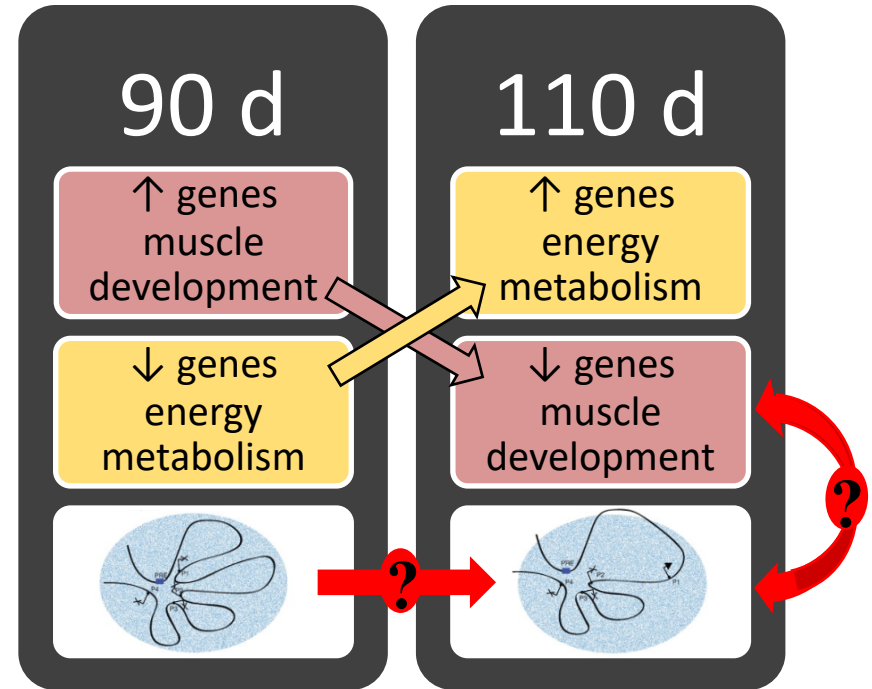
## ➤ 3D Genome organization

3 fetuses (90 days gestation)

- Rep1-90
- Rep2-90
- Rep3-90

3 fetuses (110 days gestation)

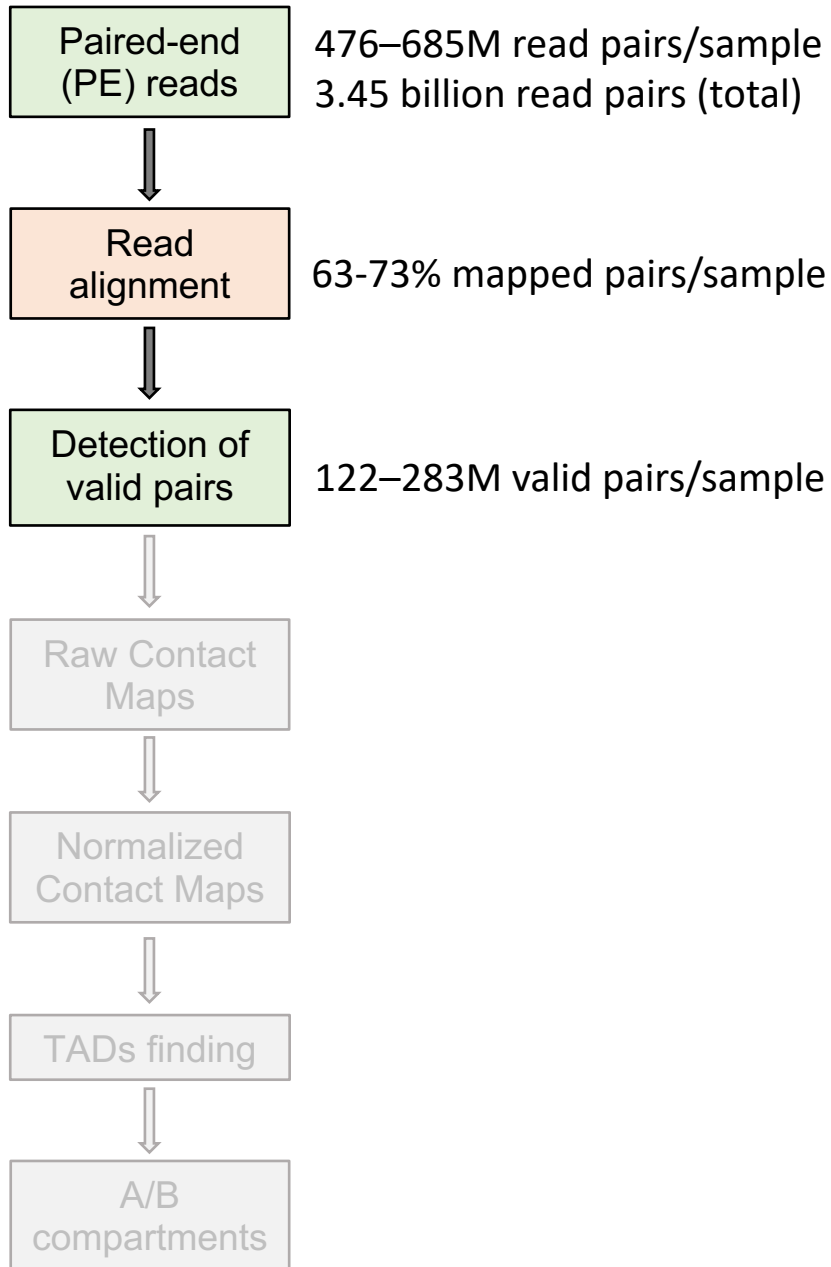
- Rep1-110
- Rep2-110
- Rep3-110





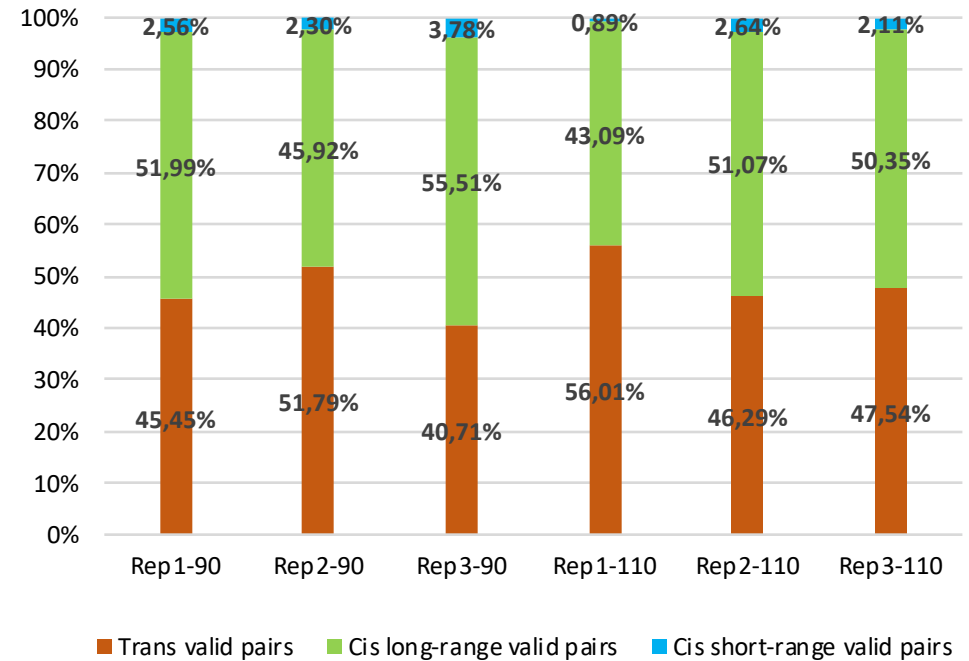
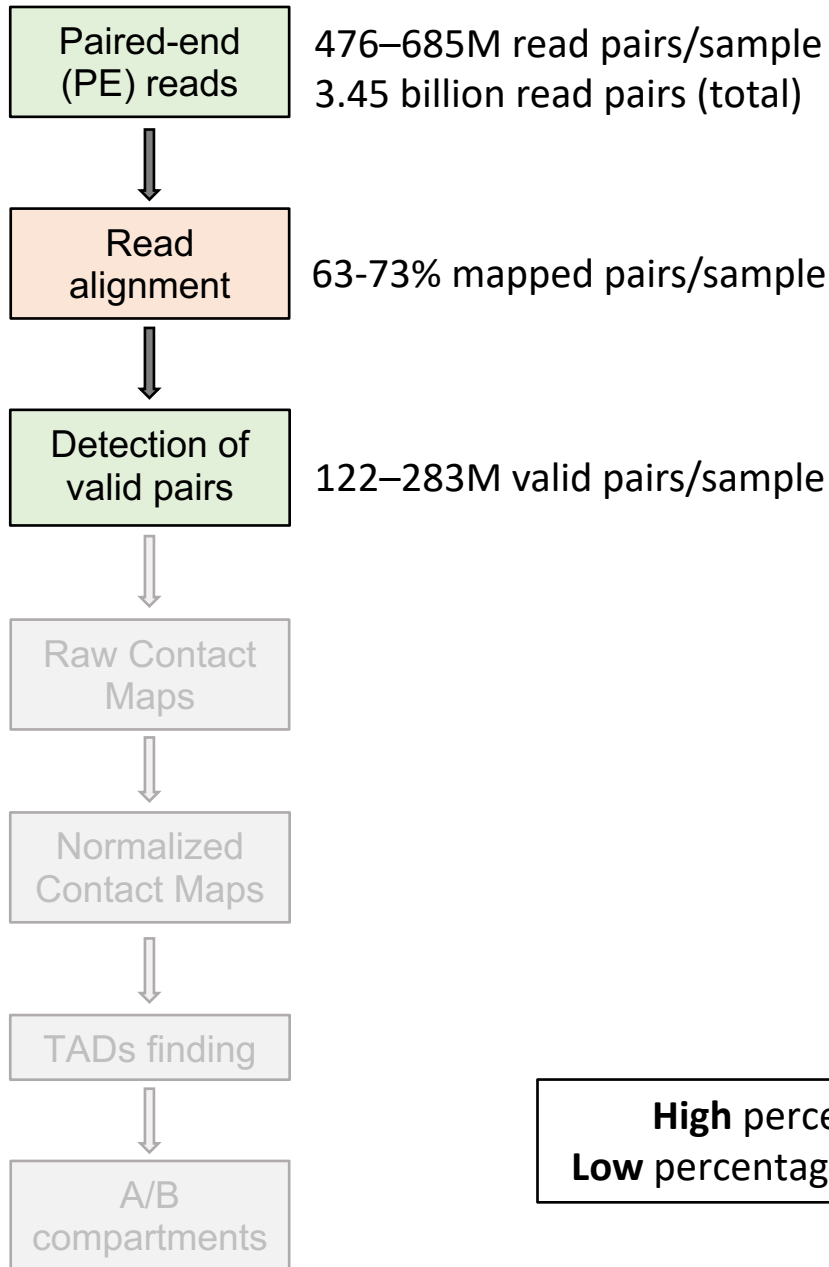
# Hi-C data analysis

HiC-Pro (*Servant et al. 2015*)



# Hi-C data analysis

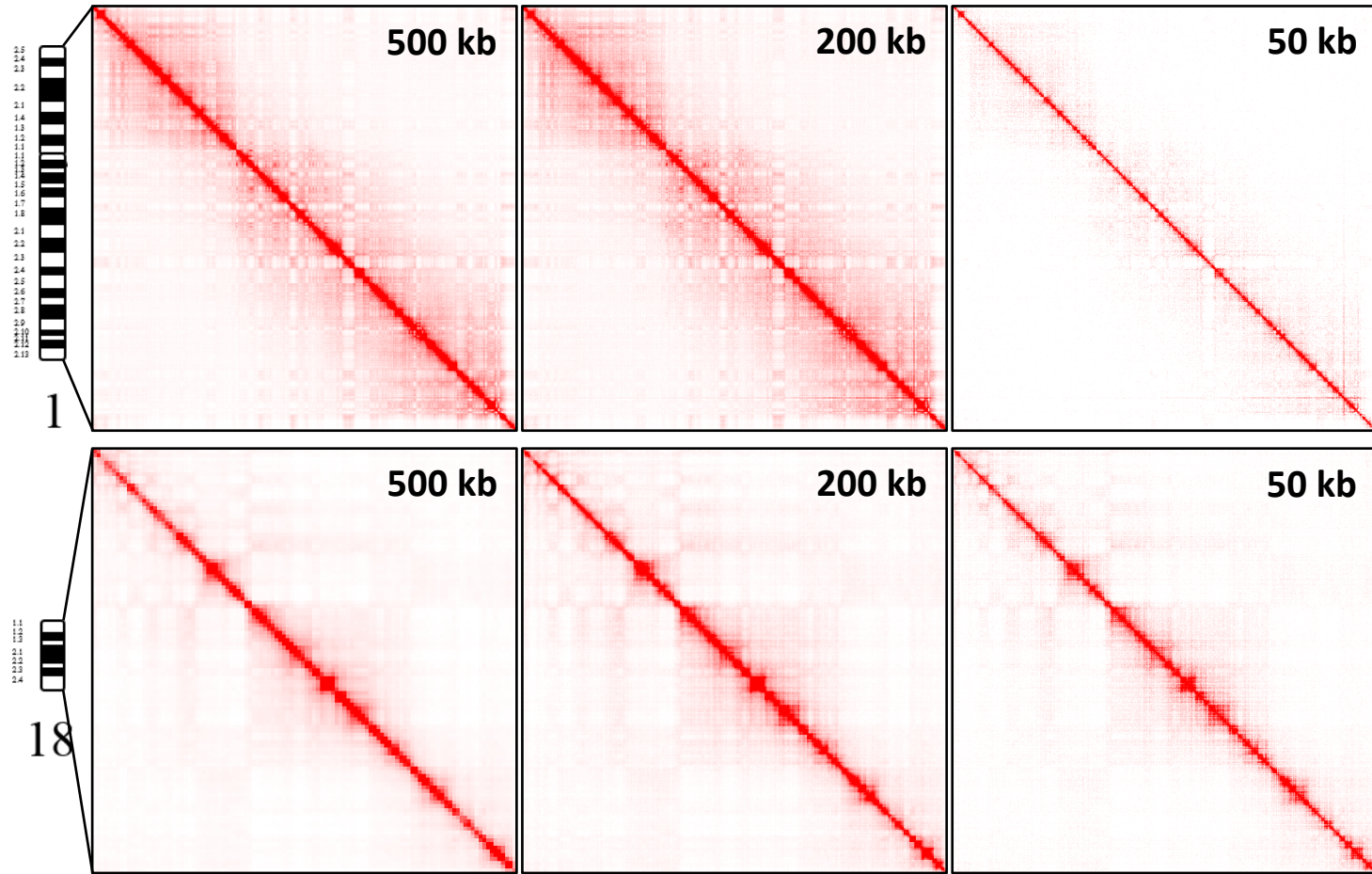
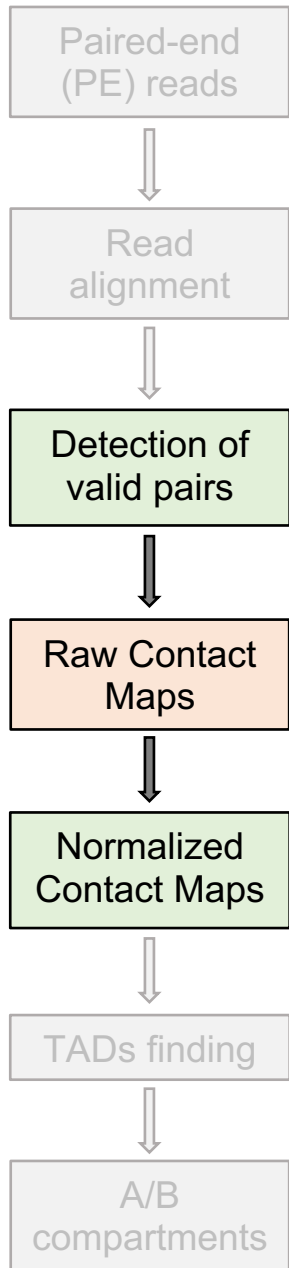
HiC-Pro (Servant et al. 2015)



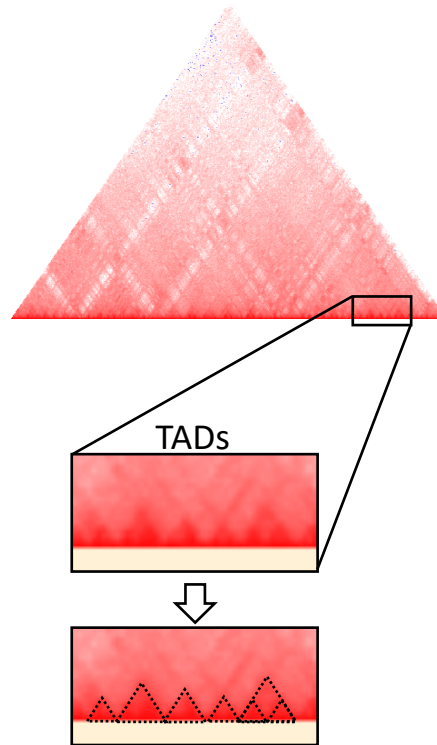
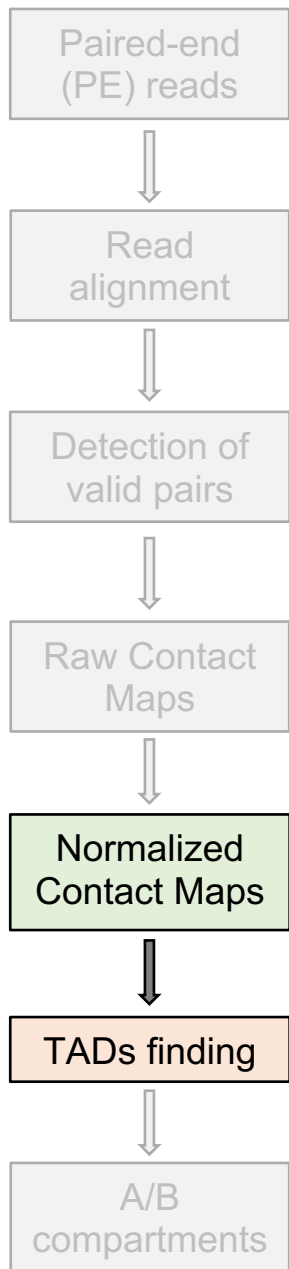
**High percentages of *trans* valid pairs**  
**Low percentages of *cis* short-range valid pairs**

# Hi-C data analysis

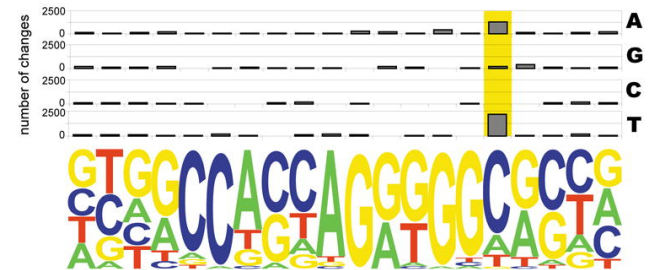
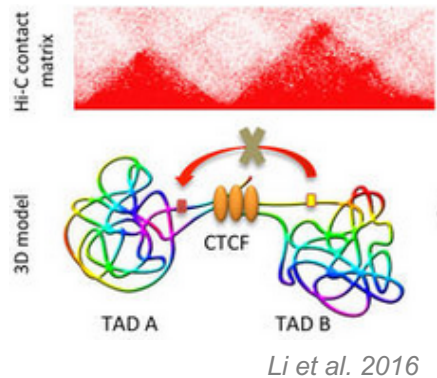
HiC-Pro (Servant et al. 2015)



# Hi-C data analysis (TADs) Juicer: arrowhead (Neva et al., Cell Systems 2016; Rao et al., Cell 2014)



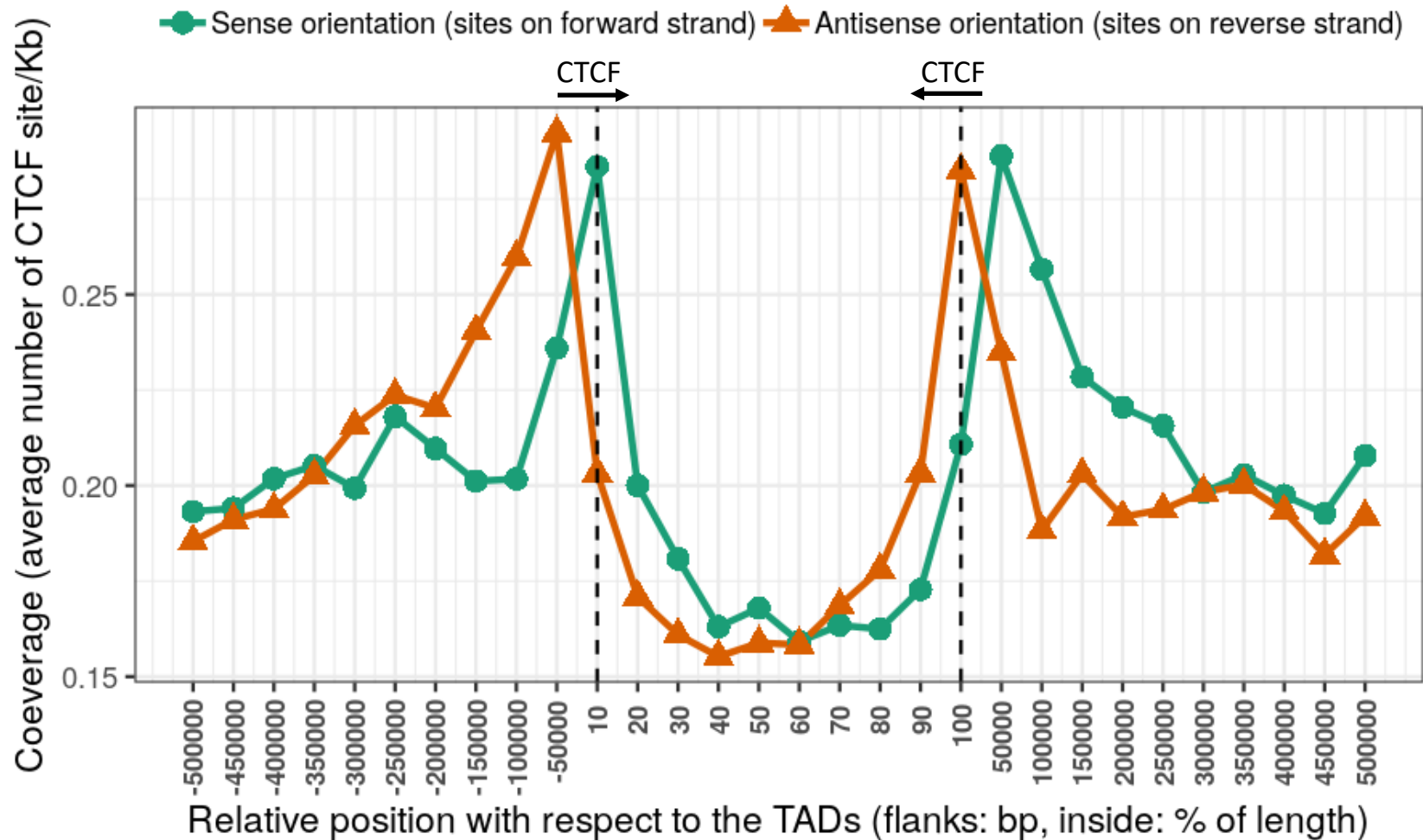
- 50 Kb resolution matrices
- 1312 TADs per replicate on average
- Average mean size: 1480 Kb
- Global conservation of TAD structure (74 – 79% of TAD boundaries from each condition are identical to the other condition)



*Kim et al., Cell 2016*

# Hi-C data analysis (TADs)

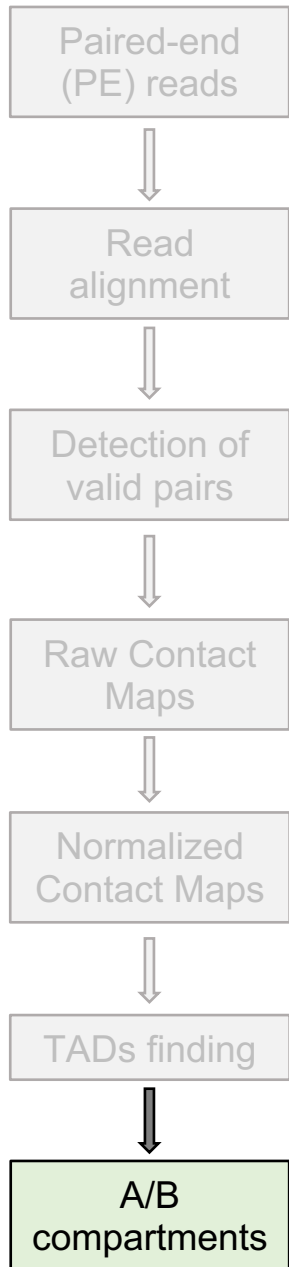
## Merged matrix



TADs validation  
Consistent Hi-C data

# Hi-C data analysis (A/B compartments)

*(Lieberman-Aiden et al., Science 2009)*



500 Kb resolution matrices  
682 compartments/replicate (average)  
Median size 2.6 Mb – 3.5 Mb

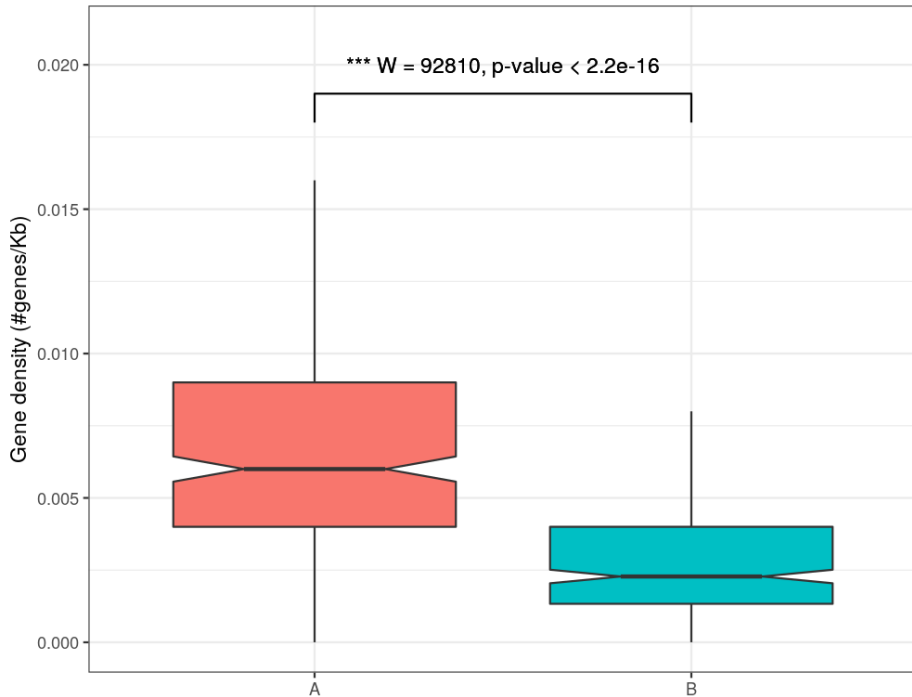
# Hi-C data analysis (A/B compartments)

(Lieberman-Aiden et al., Science 2009)

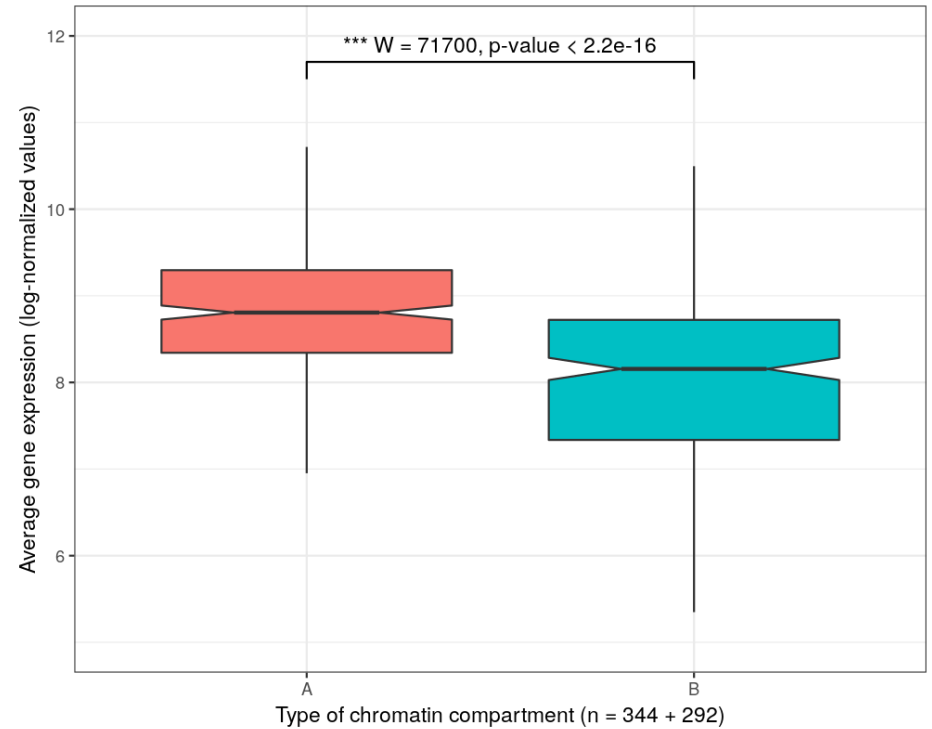
Paired-end  
(PE) reads



Gene density in A/B compartments



Gene expression in A/B compartments



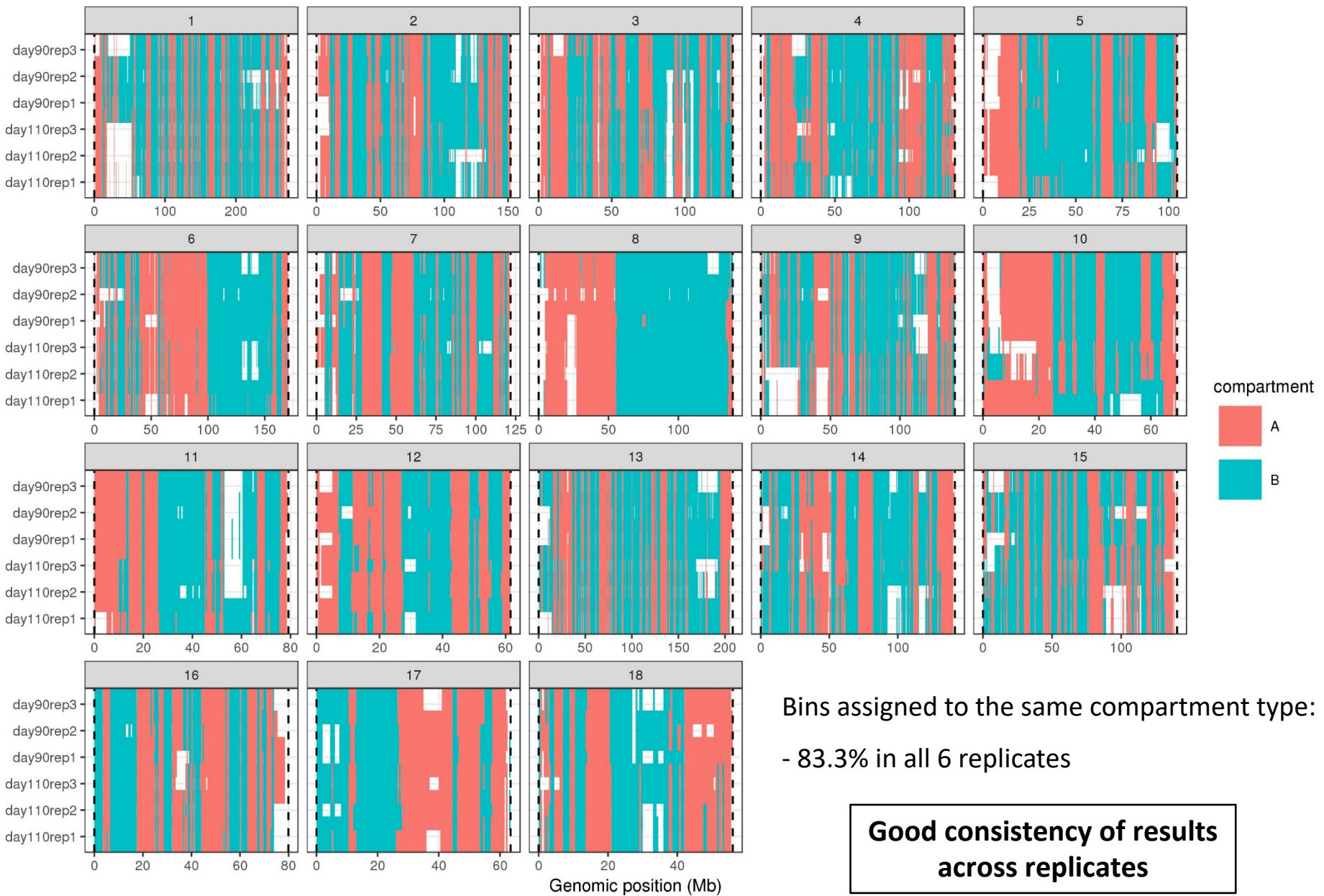
TADs finding



A/B  
compartments

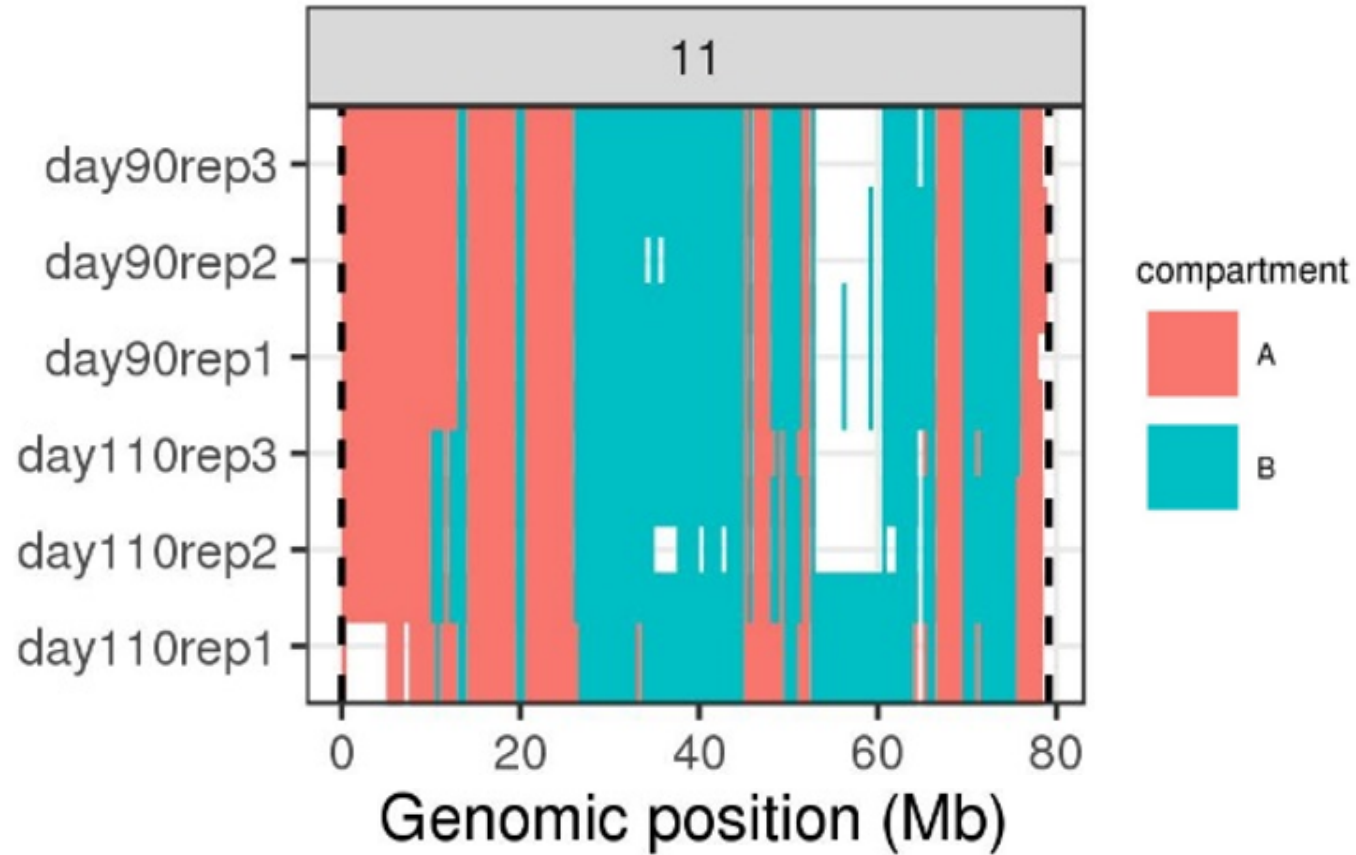
500 Kb resolution matrices  
682 compartments/replicate (average)  
Median size 2.6 Mb – 3.5 Mb

# A/B compartments

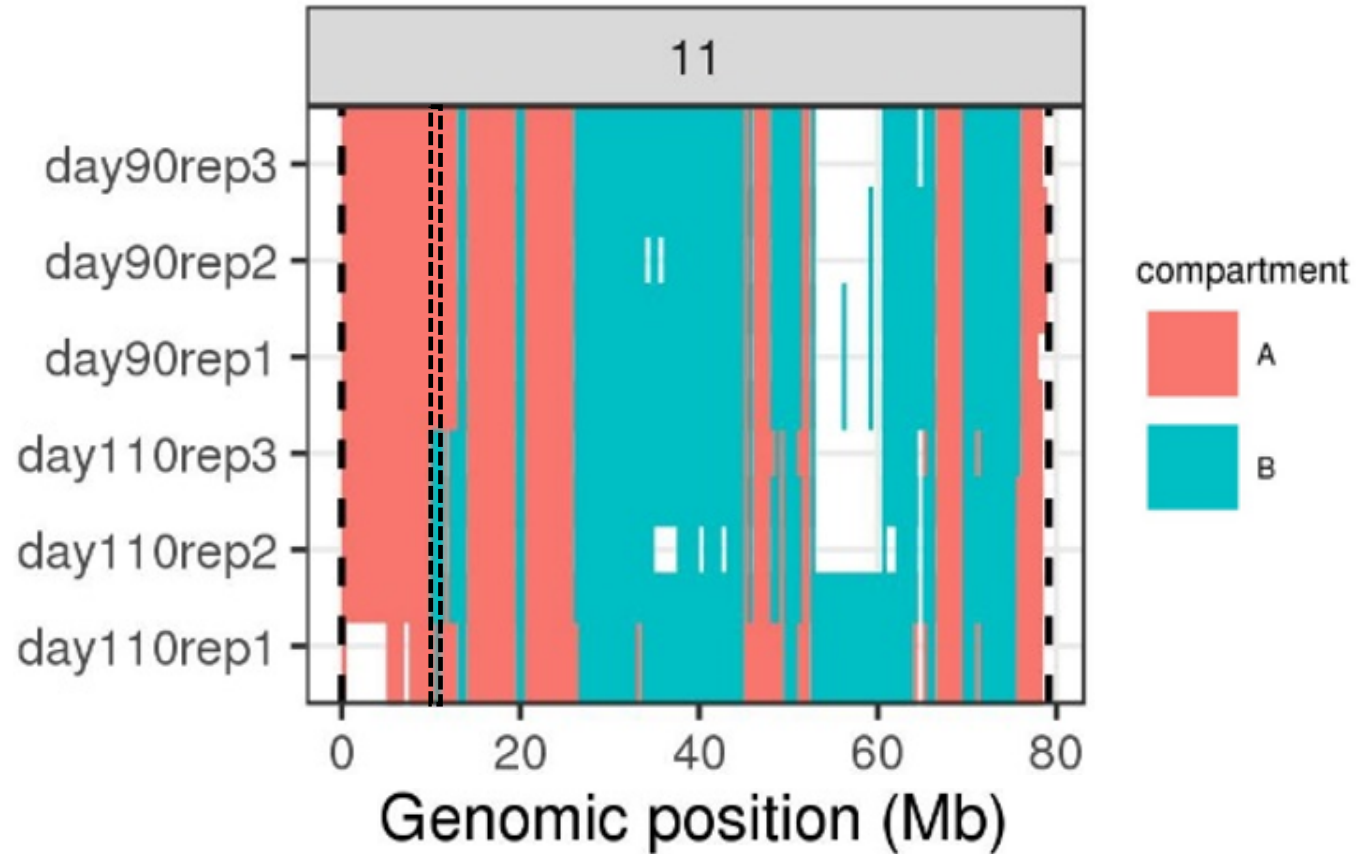




# Genomic regions switching compartments



# Genomic regions switching compartments



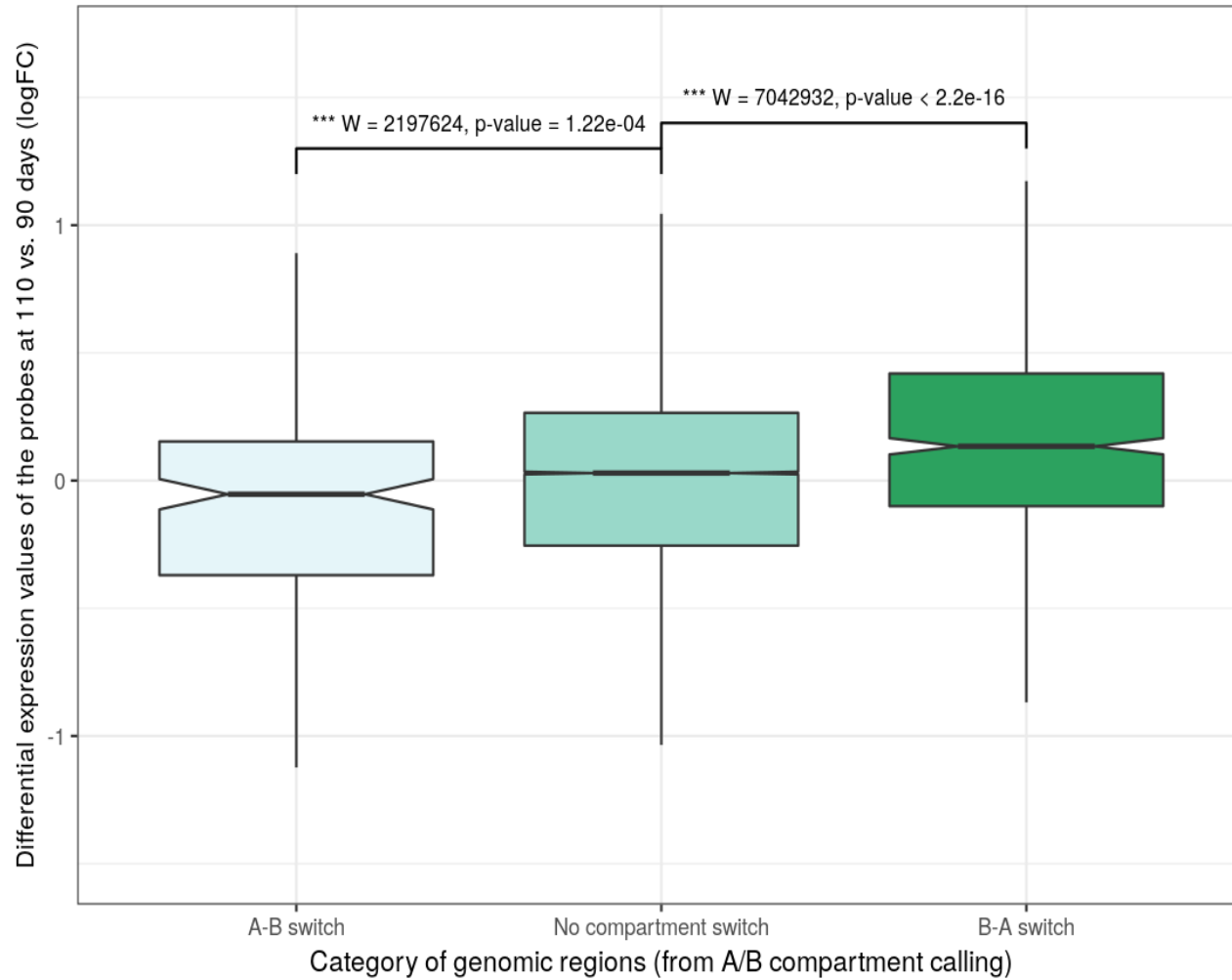
Variability between conditions: 3.3% switching bins (52 Mb)

90 d → 110 d

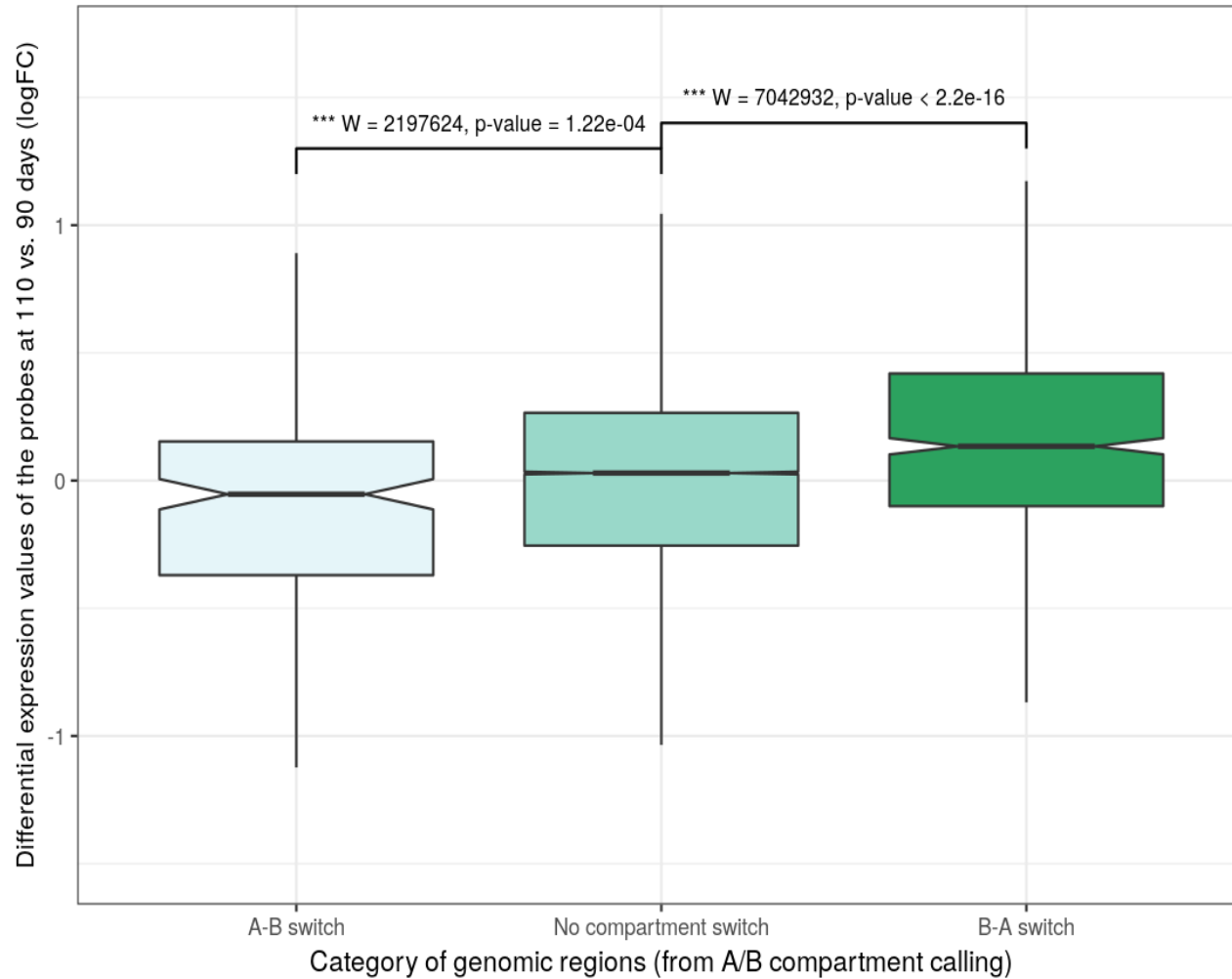
43.3% (AAA → BBB)

56.7% (BBB → AAA)

# A/B compartments and gene expression



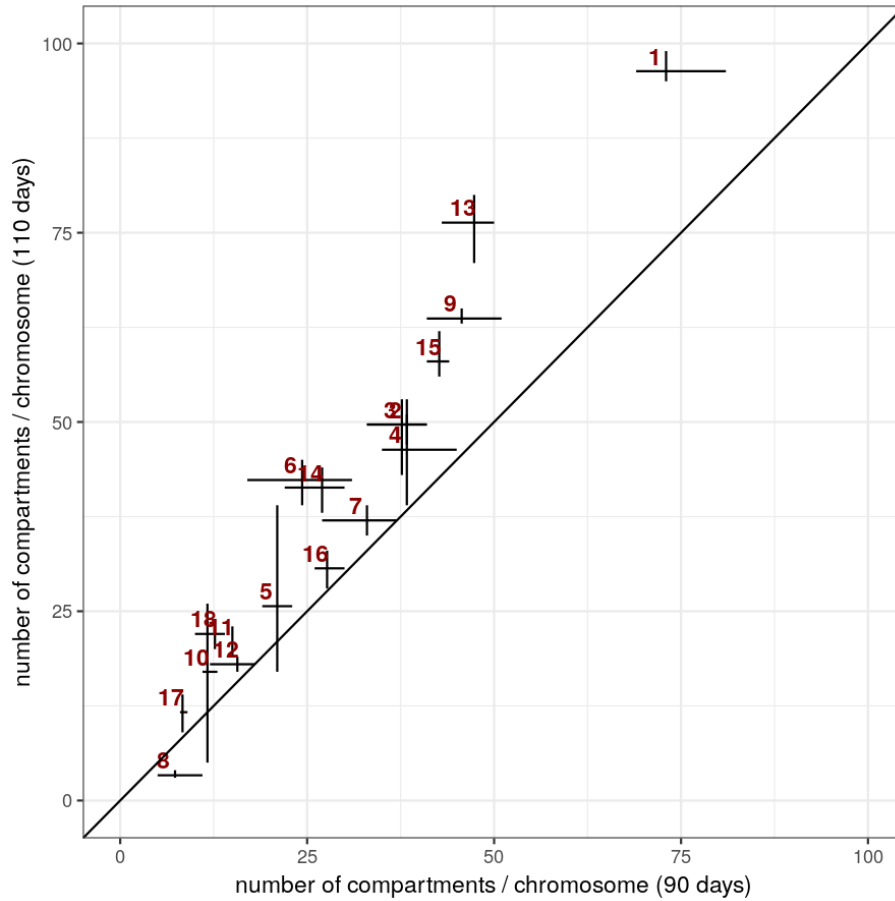
# A/B compartments and gene expression



Switching regions are associated to transcriptional changes

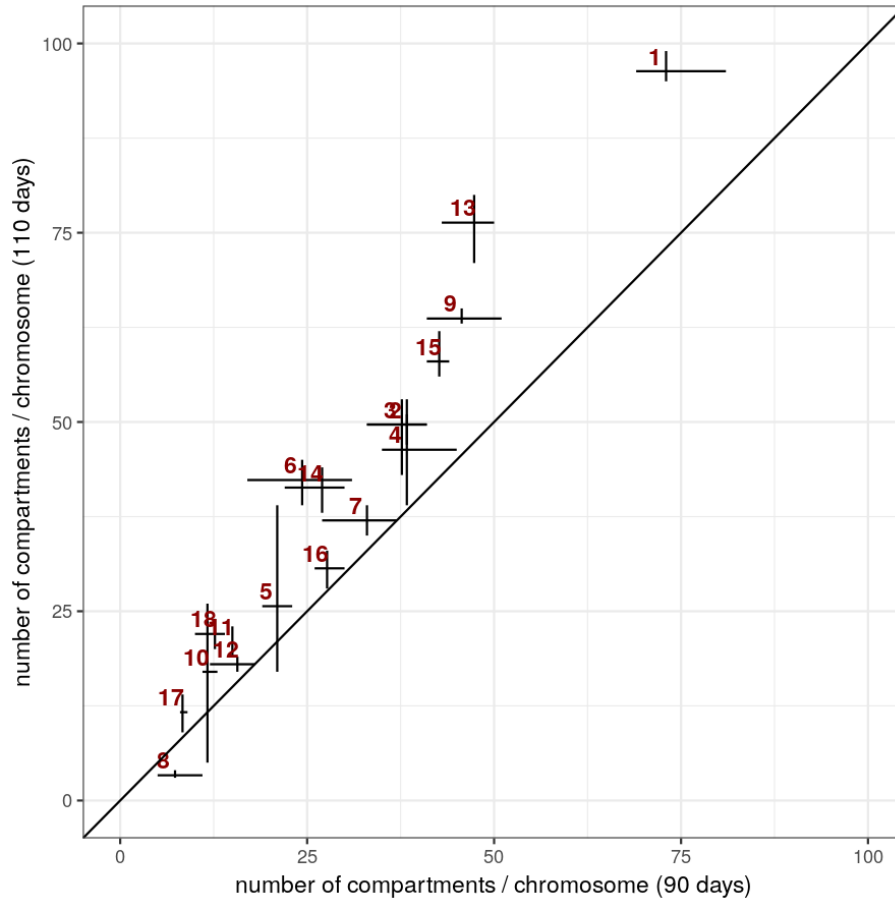
# Genome-wide fragmentation during the muscle maturation process

Number distribution of compartments



# Genome-wide fragmentation during the muscle maturation process

Number distribution of compartments



Fragmentation of genome compartmentalization

# Differentially distal genomic regions

- Filtering, normalization and detection of bin pairs with significant number of contacts (method: Generalized Linear Model “GLM” functionality of edgeR)

	<b>500 Kb</b>	<b>200 Kb</b>
<b>Total bin pairs with any count</b>	9,262,199	3,844,272
<b>Differential bin pairs</b>	<b>10,183 (0.11%)</b>	<b>3,417 (0.09%)</b>

# Differentially distal genomic regions

- Filtering, normalization and detection of bin pairs with significant number of contacts (method: Generalized Linear Model “GLM” functionality of edgeR)

	500 Kb	200 Kb
<b>Total bin pairs with any count</b>	9,262,199	3,844,272
<b>Differential bin pairs</b>	<b>10,183 (0.11%)</b>	<b>3,417 (0.09%)</b>

$$\text{logFC (bin pair)} = \log_2 [ (\text{counts at 110 days}) / (\text{counts at 90 days}) ]$$

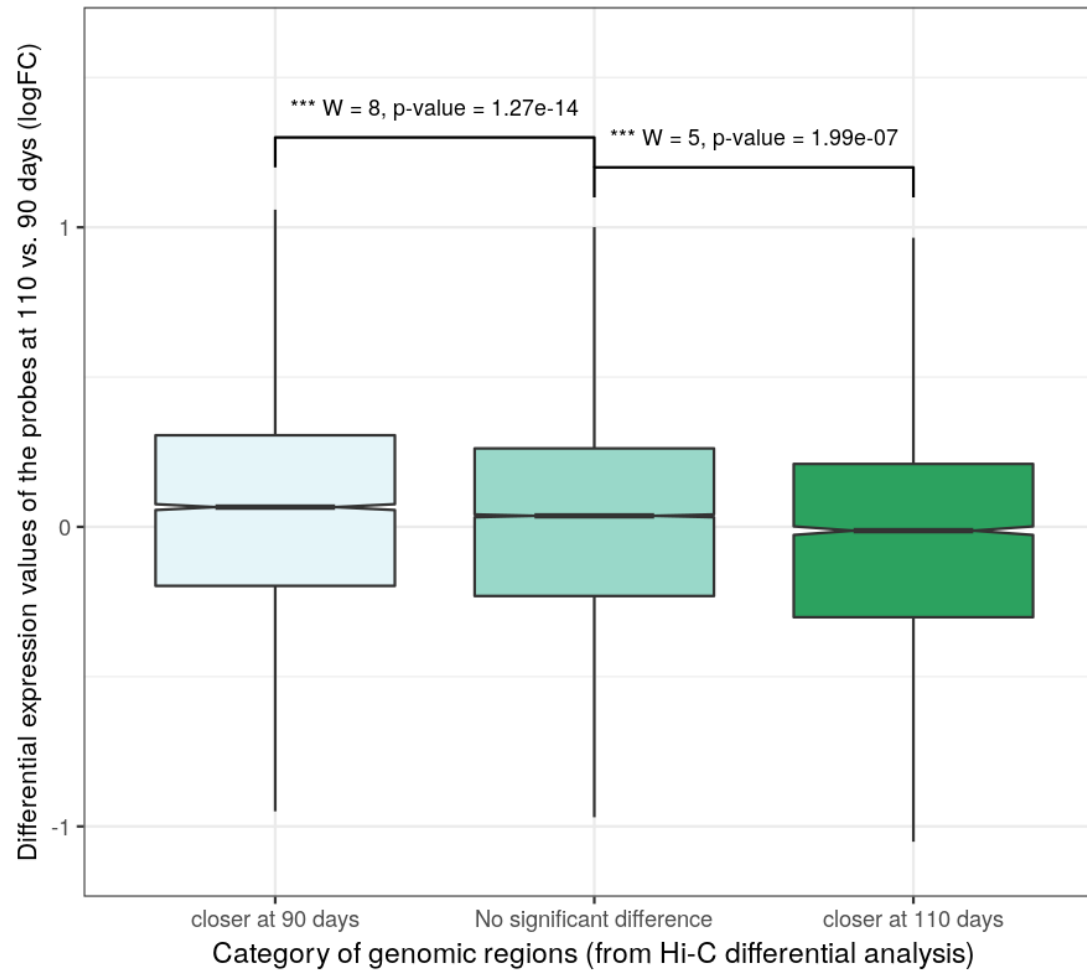
**Positive logFC** = more counts “contacts” **at 110 days** than at 90 days  
= **genomic regions closer to each other**

**Negative logFC** = more counts “contacts” **at 90 days** than at 110 days  
= **genomic regions closer to each other**



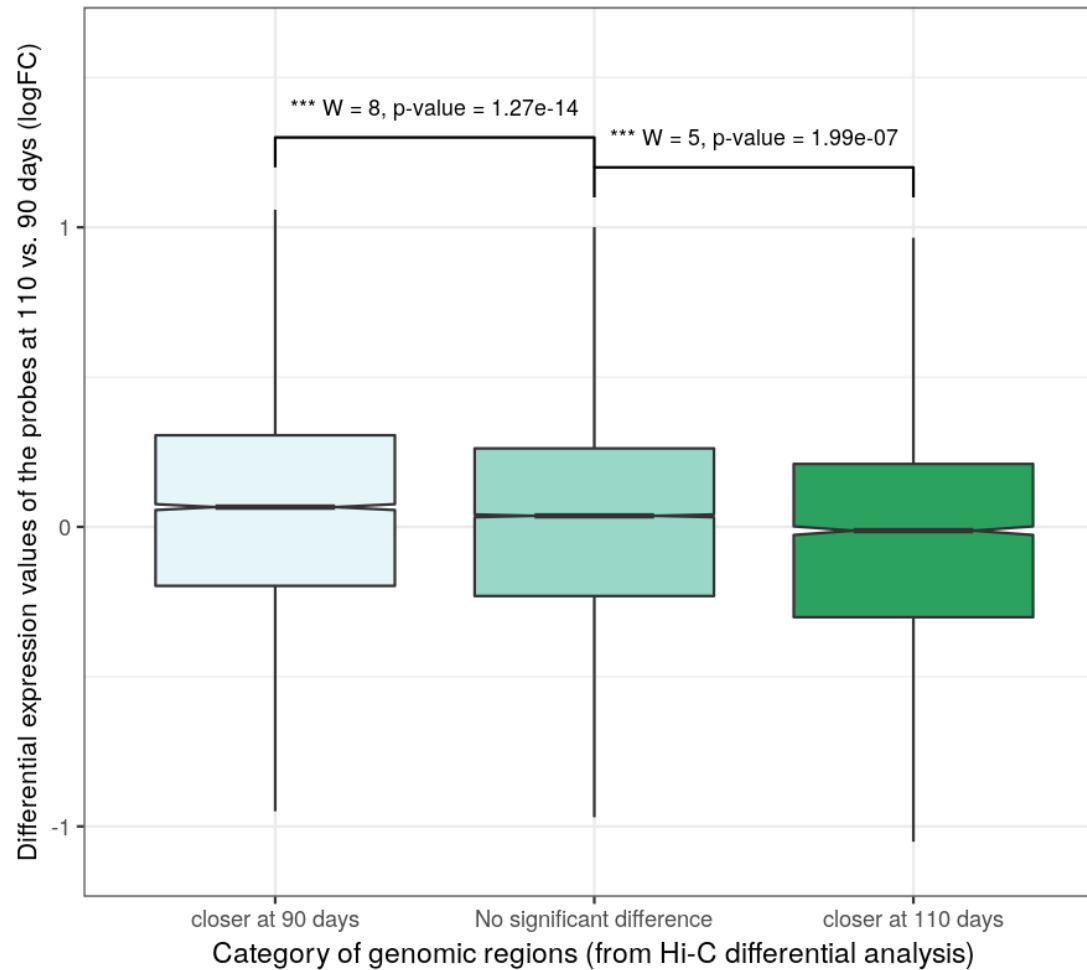
# Gene expression in differentially distal genomic regions

Distributions of logFC expression values of probes mapped to different categories of genomic regions



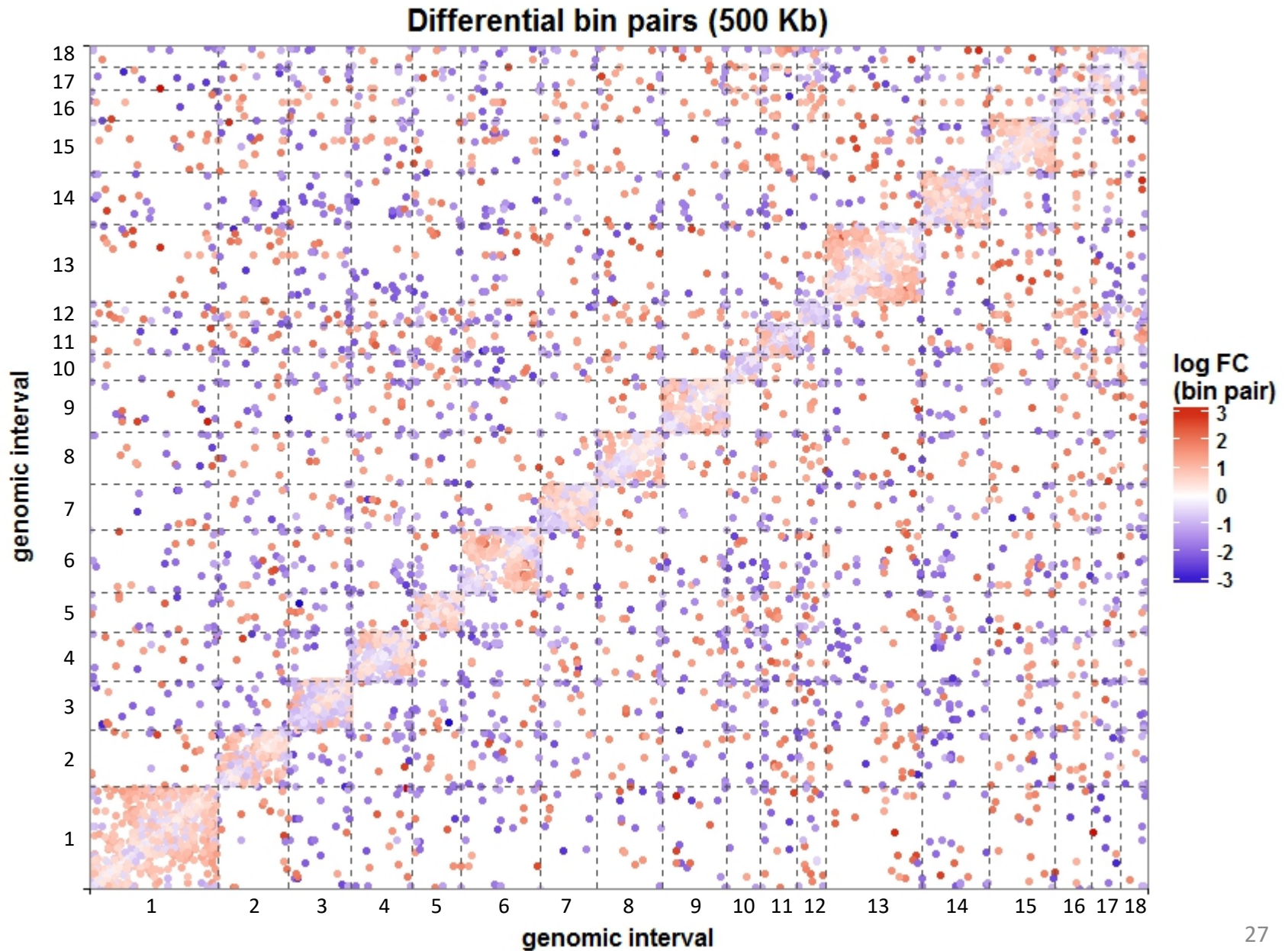
# Gene expression in differentially distal genomic regions

Distributions of logFC expression values of probes mapped to different categories of genomic regions

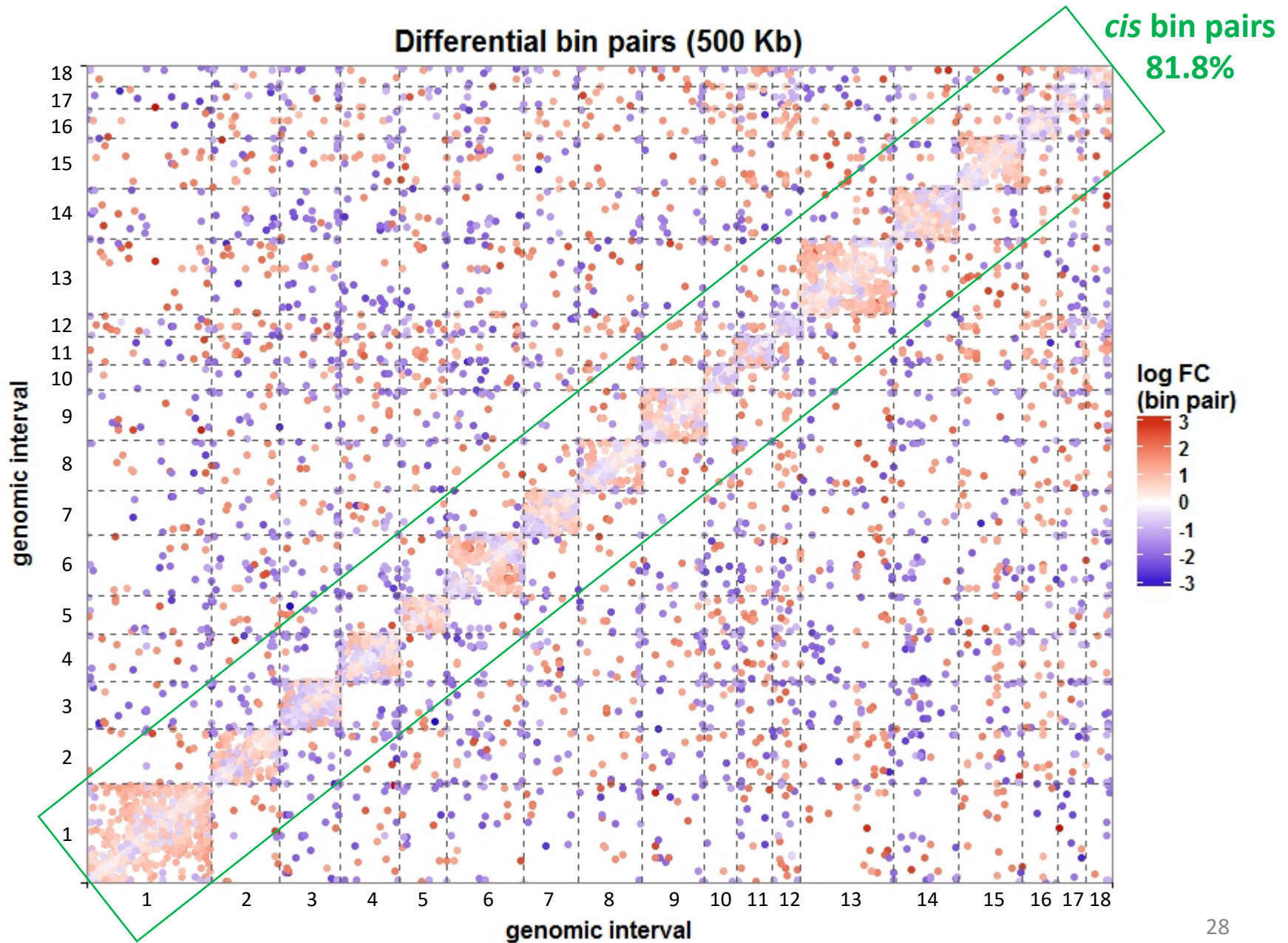


The expression values of probes in genomic regions closer at either 90 or 110 days are significantly lower

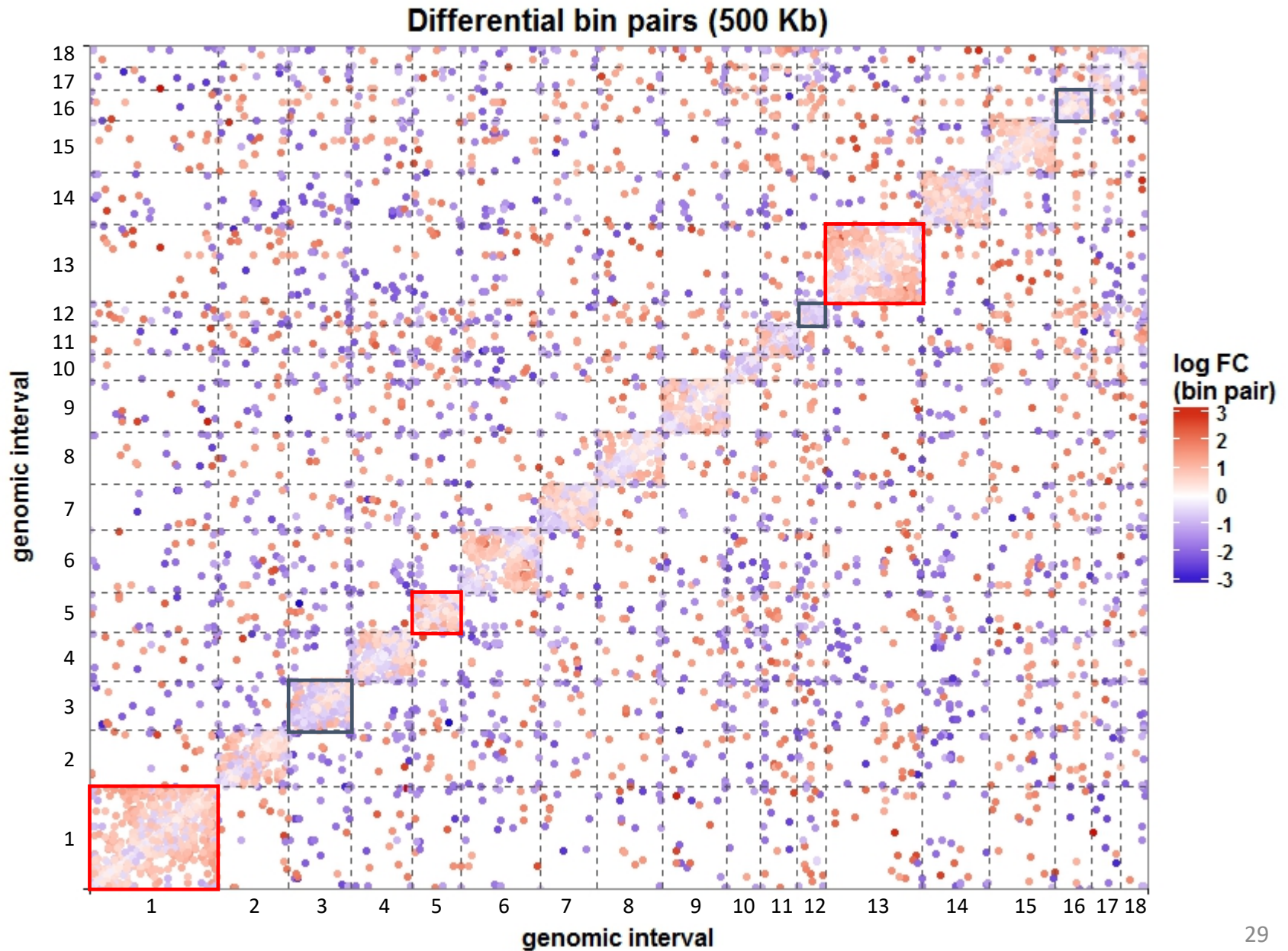
# Differential interacting regions (90-110 days of gestation)



# Differential interacting regions (90-110 days of gestation)

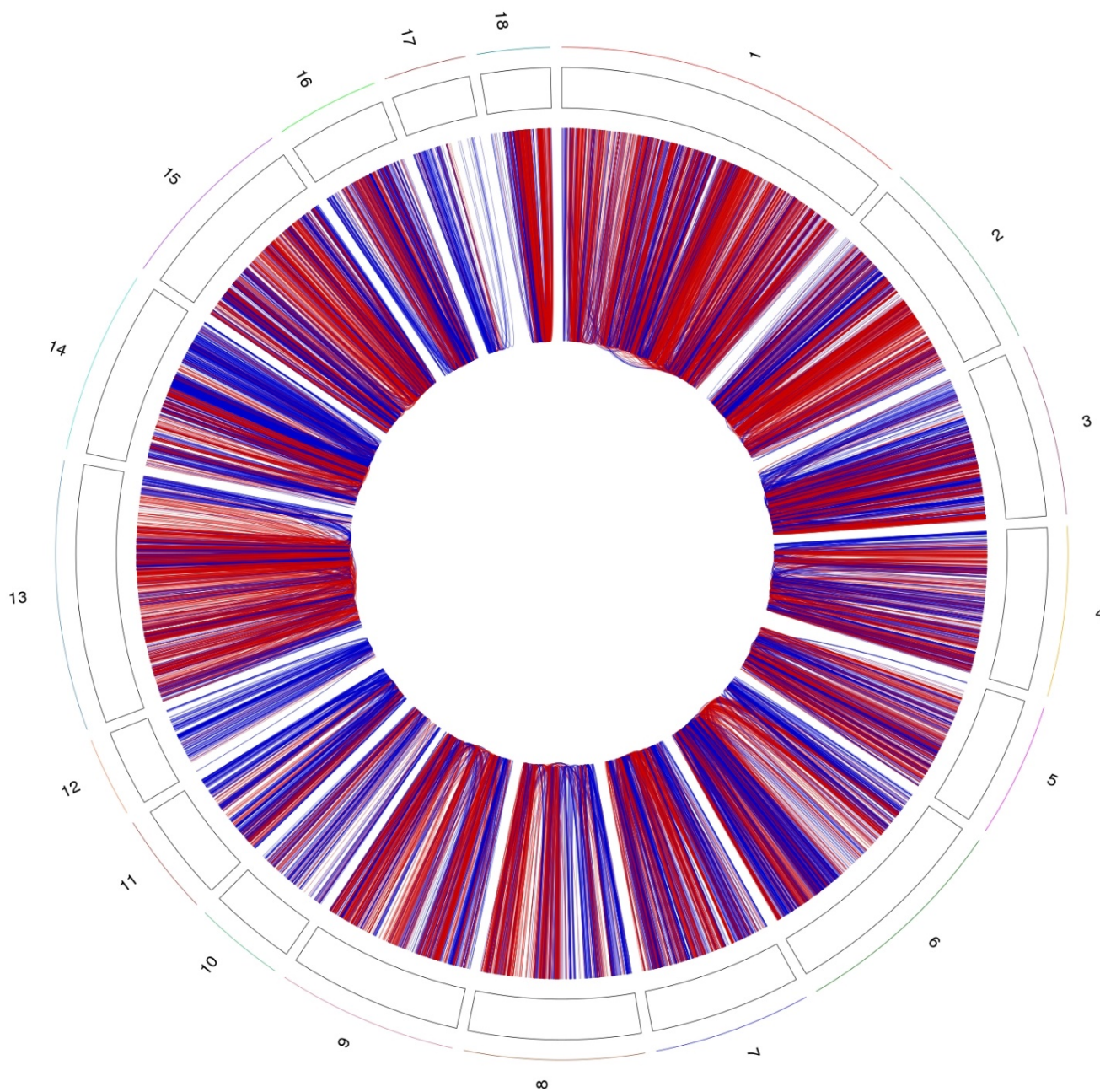


# Differential interacting regions (90-110 days of gestation)



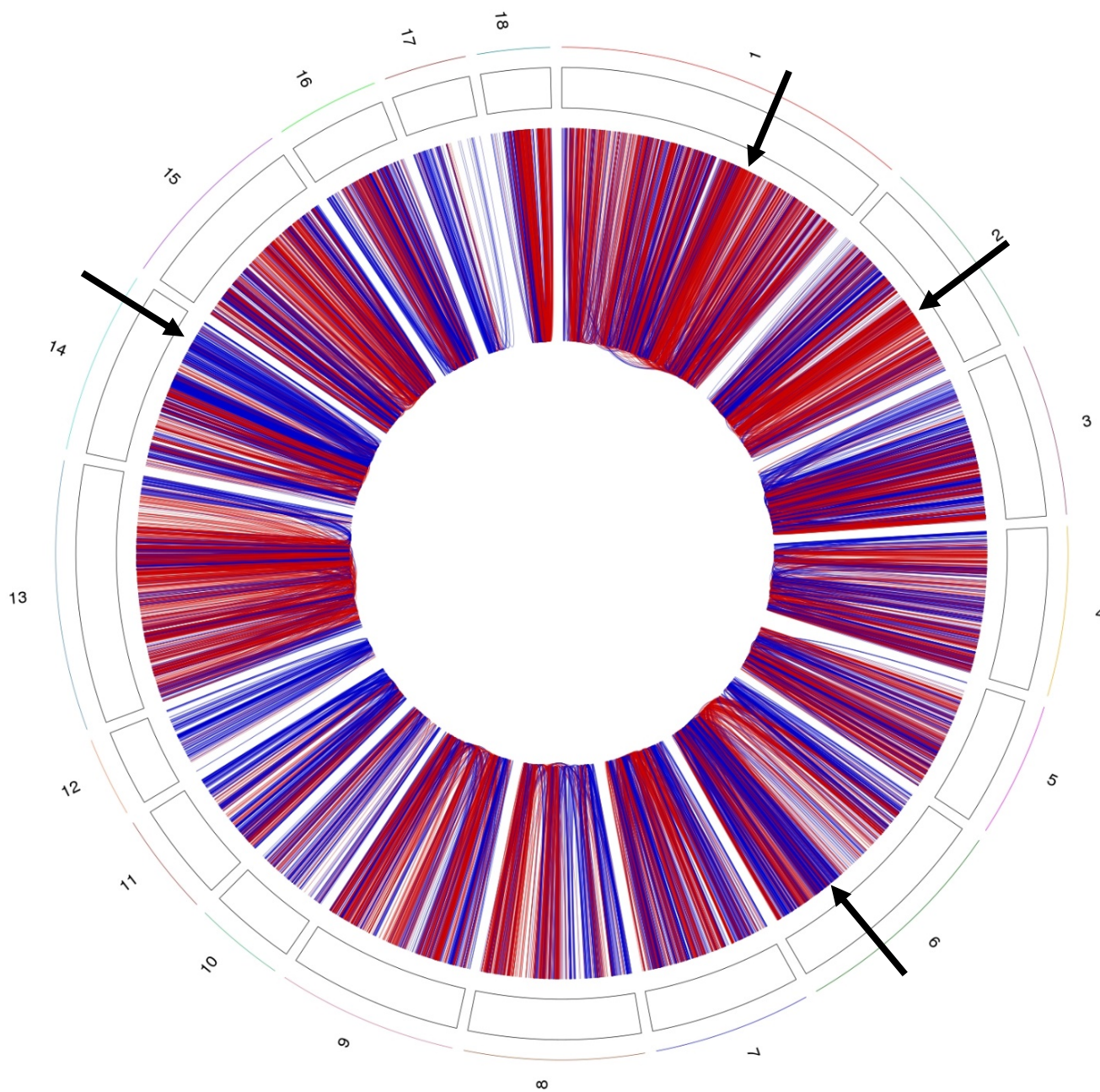
# Differential interacting regions (*cis*)

Positive logFC  
Negative logFC



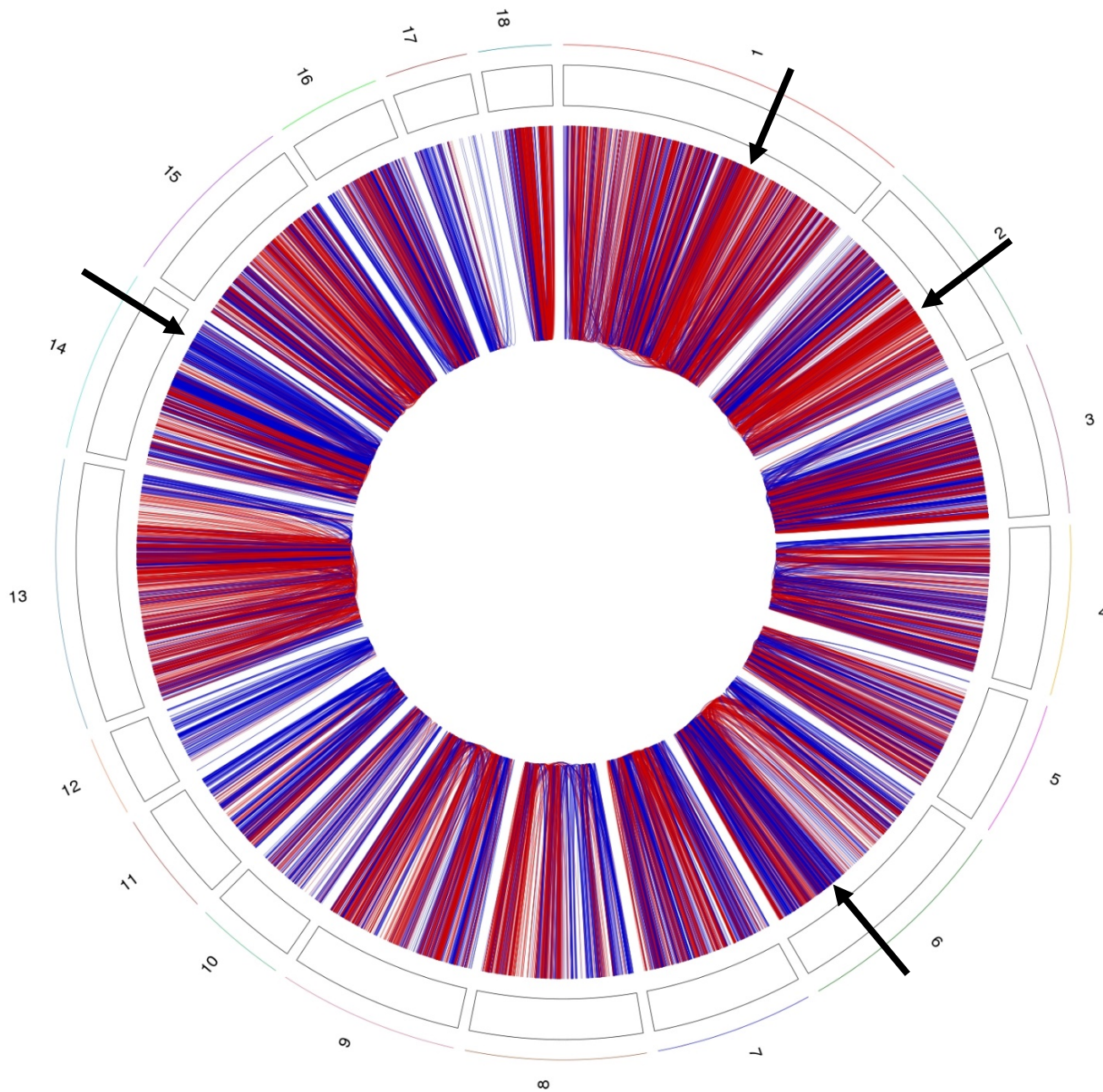
# Differential interacting regions (*cis*)

Positive logFC  
Negative logFC



# Differential interacting regions (*cis*)

Positive logFC  
Negative logFC

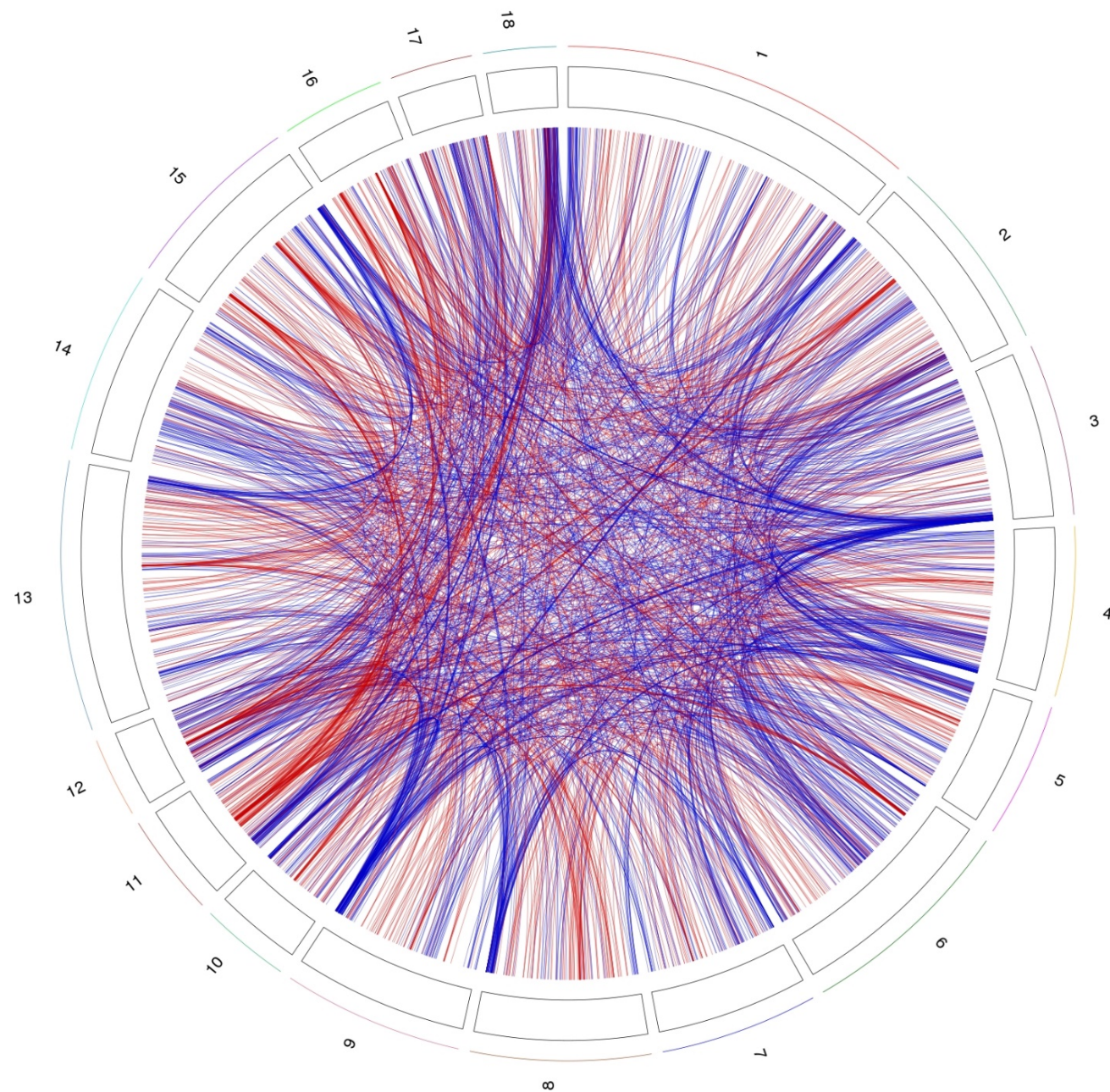


Large dynamic differential regions (90-110 days gestation)



# Differential genomic regions (*trans*)

Positive logFC  
Negative logFC



# Differential genomic regions (*trans*)

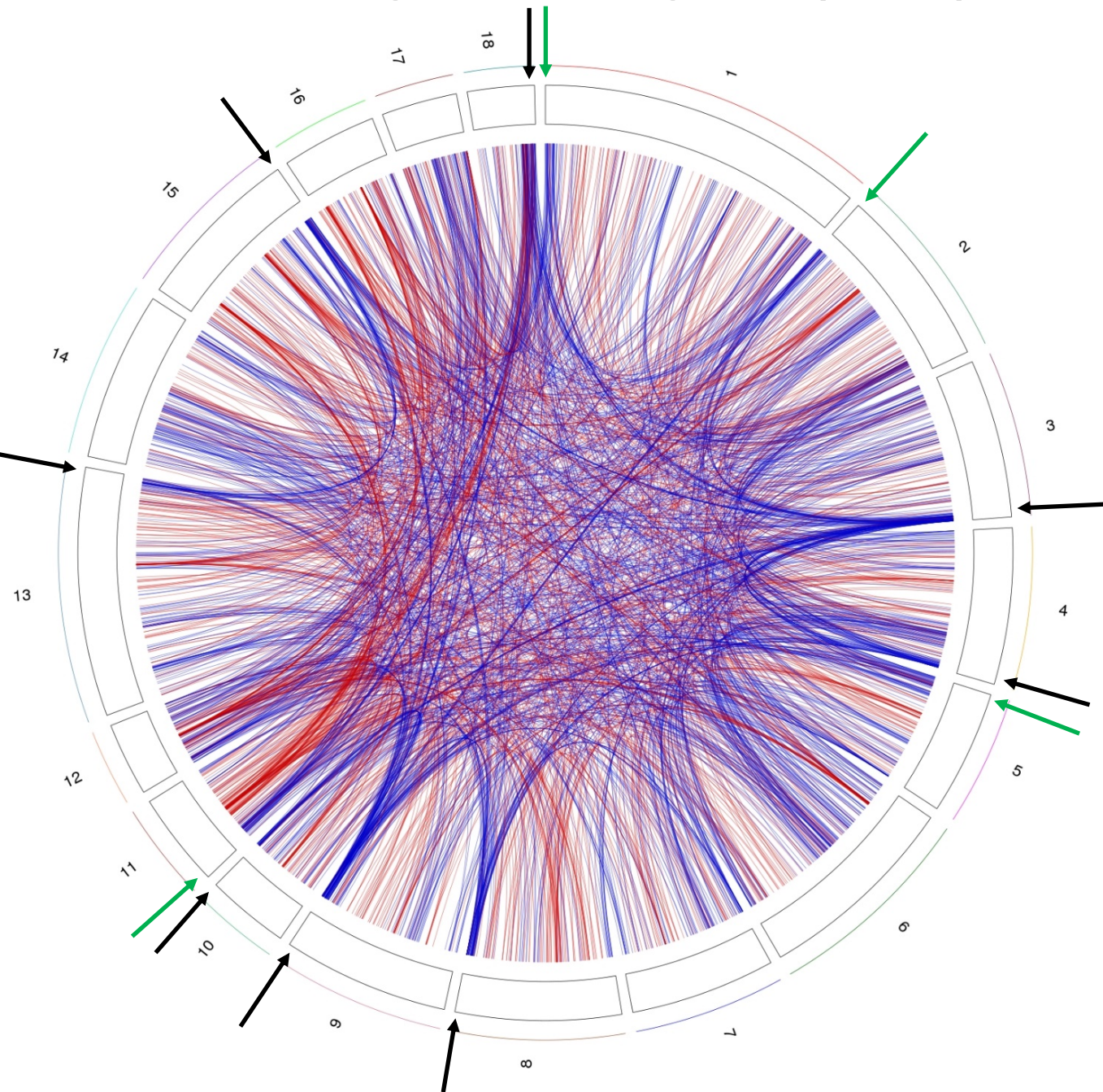
Positive logFC  
Negative logFC

Telomeric regions

Negative logFC

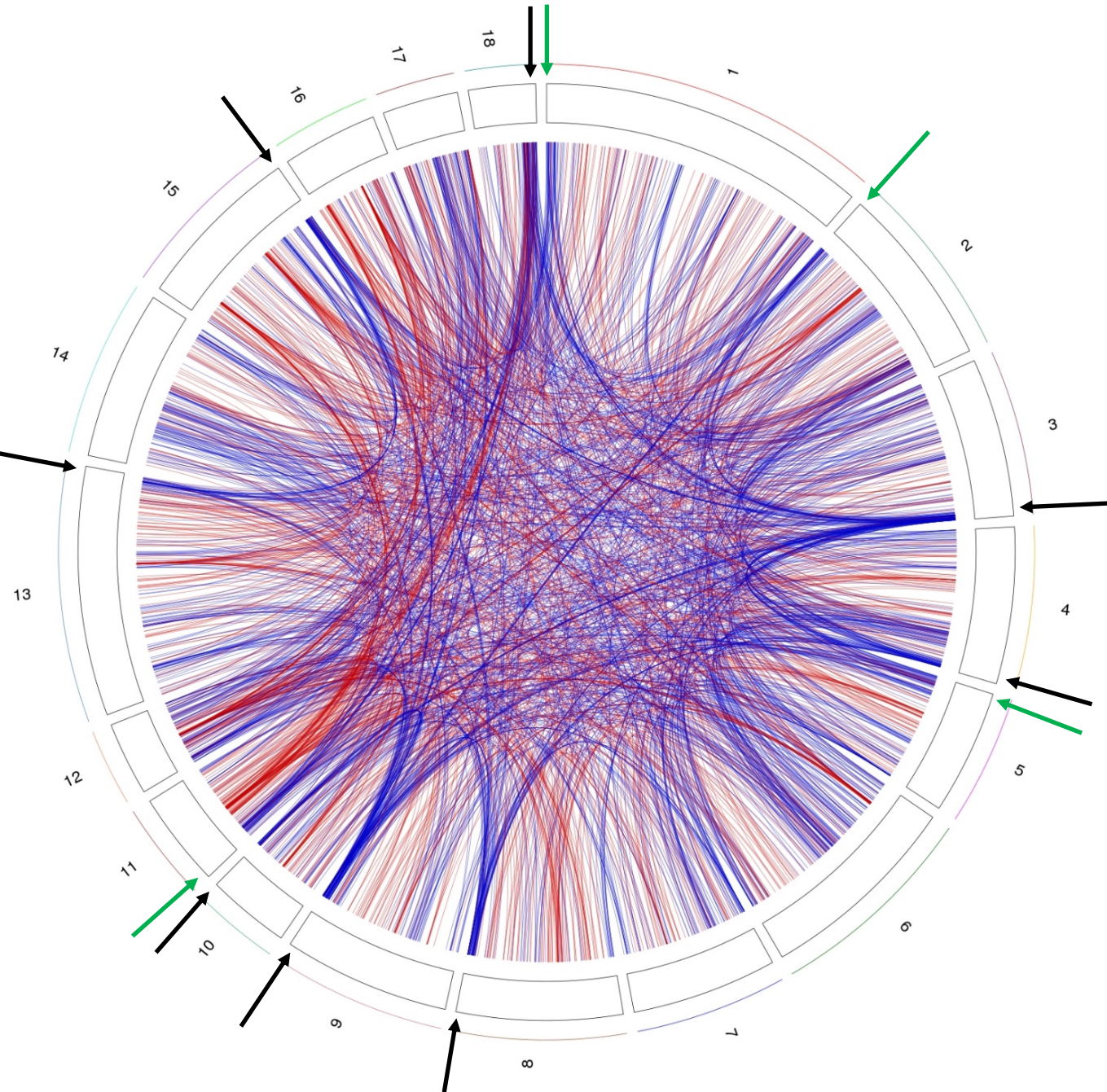
ptel

qtel



# Differential genomic regions (*trans*)

Positive logFC  
Negative logFC



Telomeric regions

Negative logFC

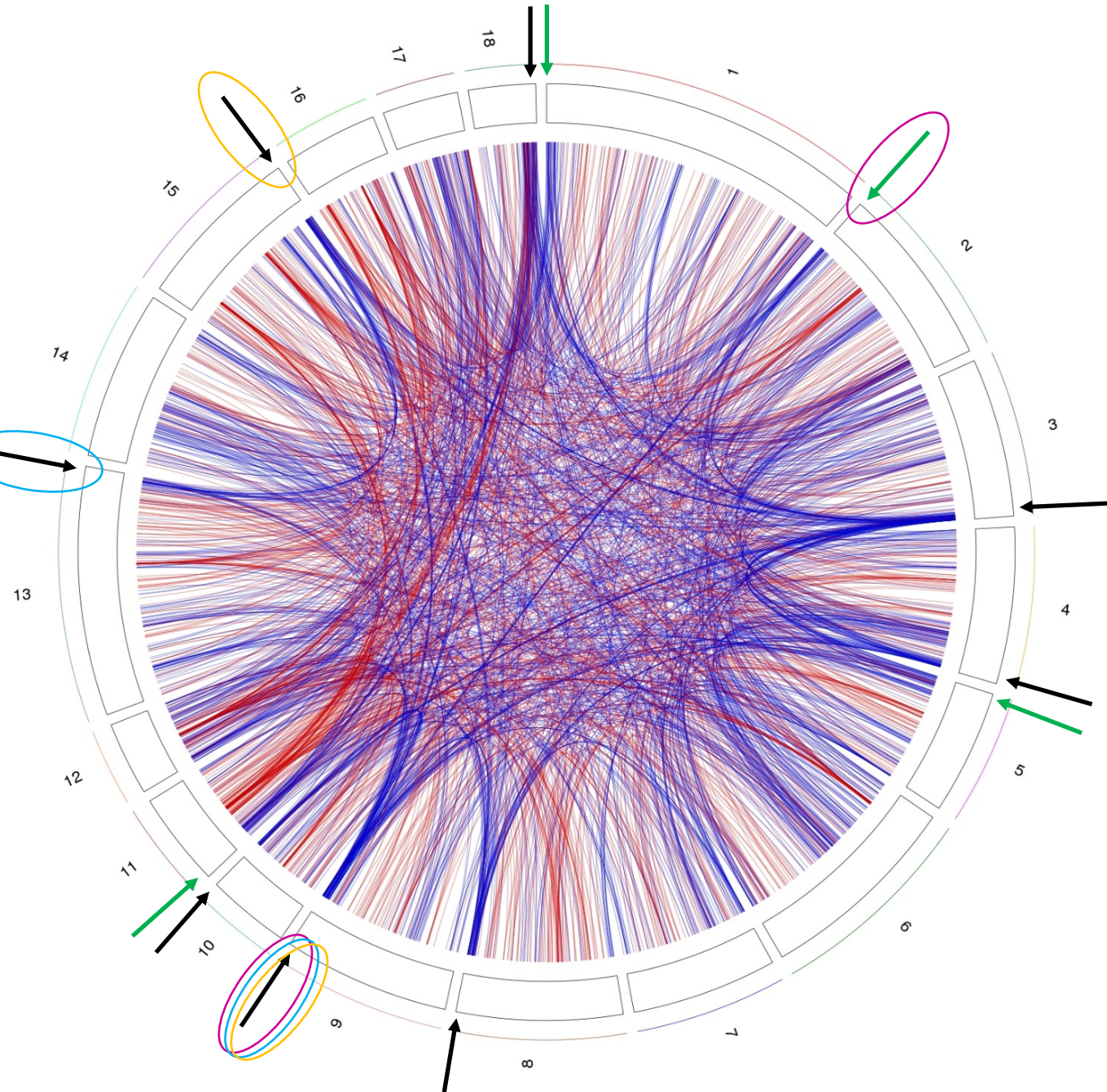
← ptel

← qtel

Preferential clustering  
of telomeres at 90 days

# Differential genomic regions (*trans*)

Positive logFC  
Negative logFC



Telomeric regions

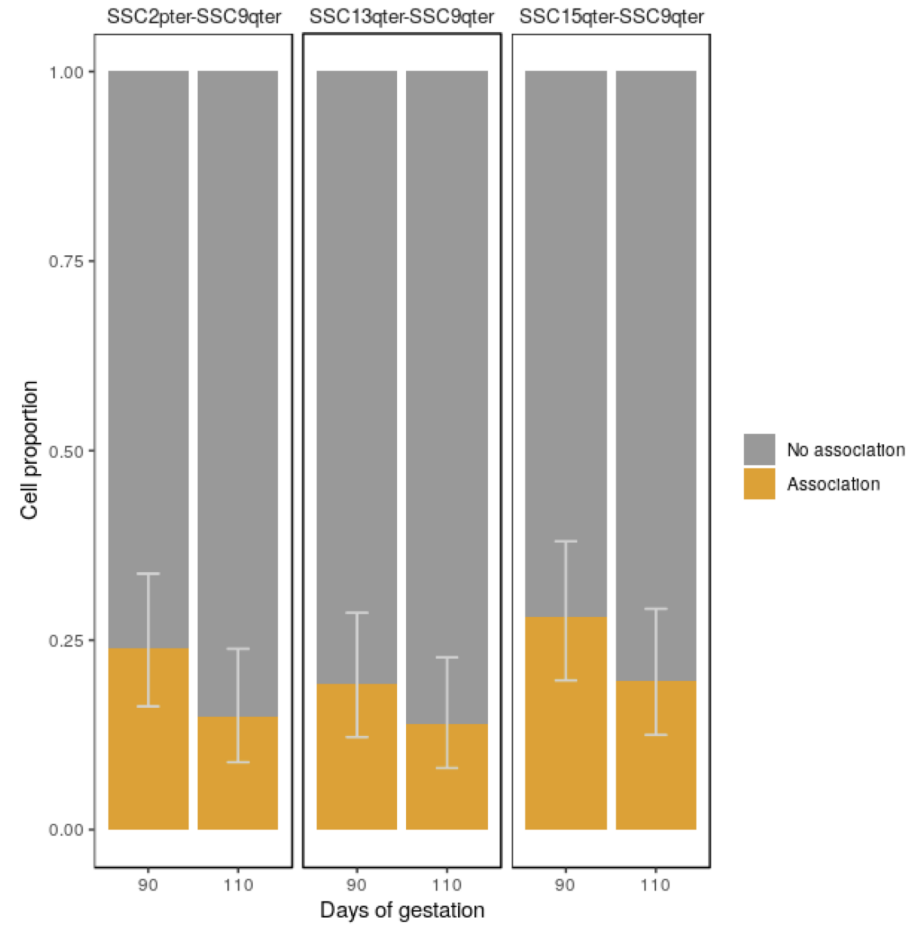
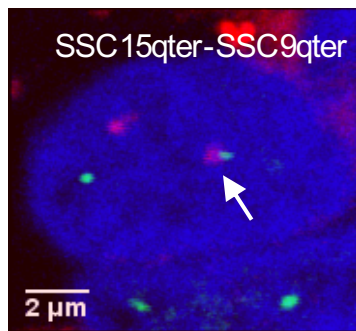
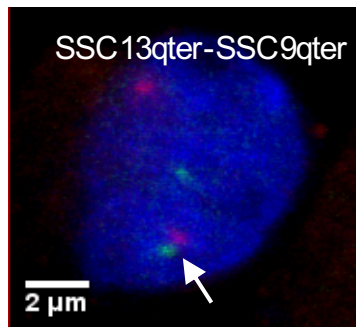
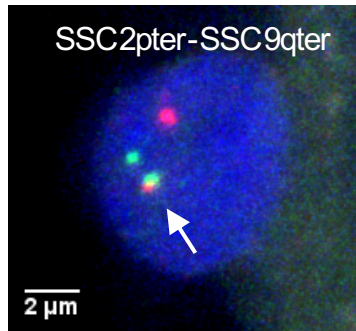
Negative logFC

← ptel

← qtel

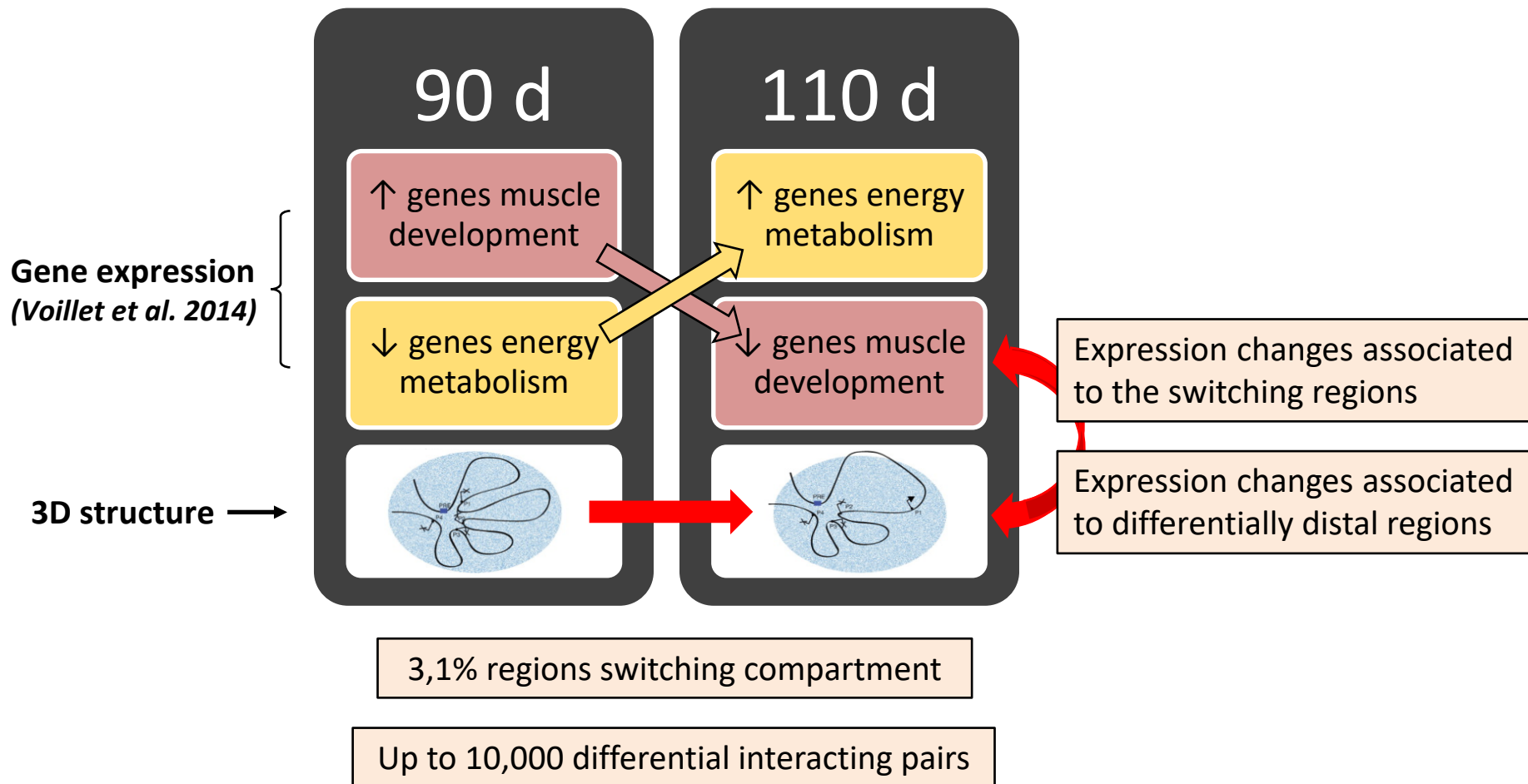
Preferential clustering  
of telomeres at 90 days

# Preferential associations of telomeres (90 days gestation)



# General output

- Changes in genome structure at late gestation → switching A/B compartments
  - genome-wide fragmentation
  - differentially interacting regions (telomeres)
- These changes are associated with variations in gene expression



**Hi-C working team:**

Experiments: **Hervé Acloque & Florence Mompert**

Sequencing: **Diane Esquerré**

Data analysis: **Sylvain Foissac, Sarah Djebali, Matthias Zytnicki & David Robelin**

Statistic analysis : **Nathalie Vialaneix**

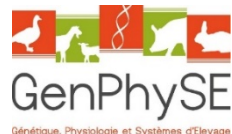
**Cytogenetic team:**

**Yvette Lahbib-Mansais**

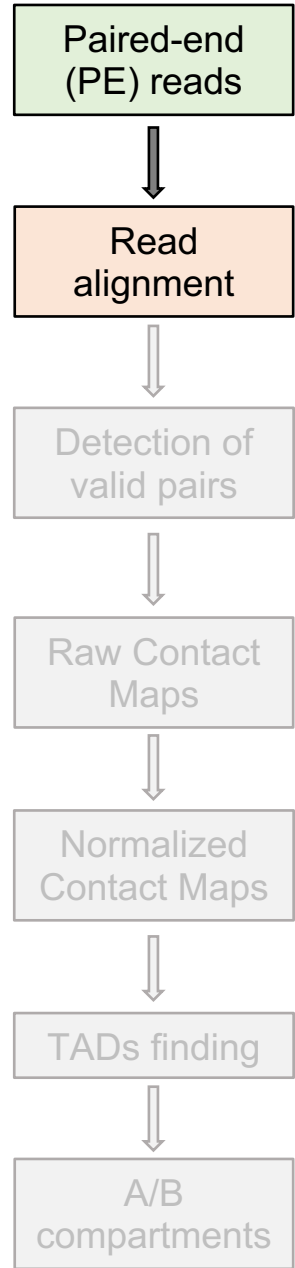
**Martine Bouissou-Matet**

**Funding:**

SCALES projet (CNRS): **Pierre Neuvial & Nathalie Vialaneix**



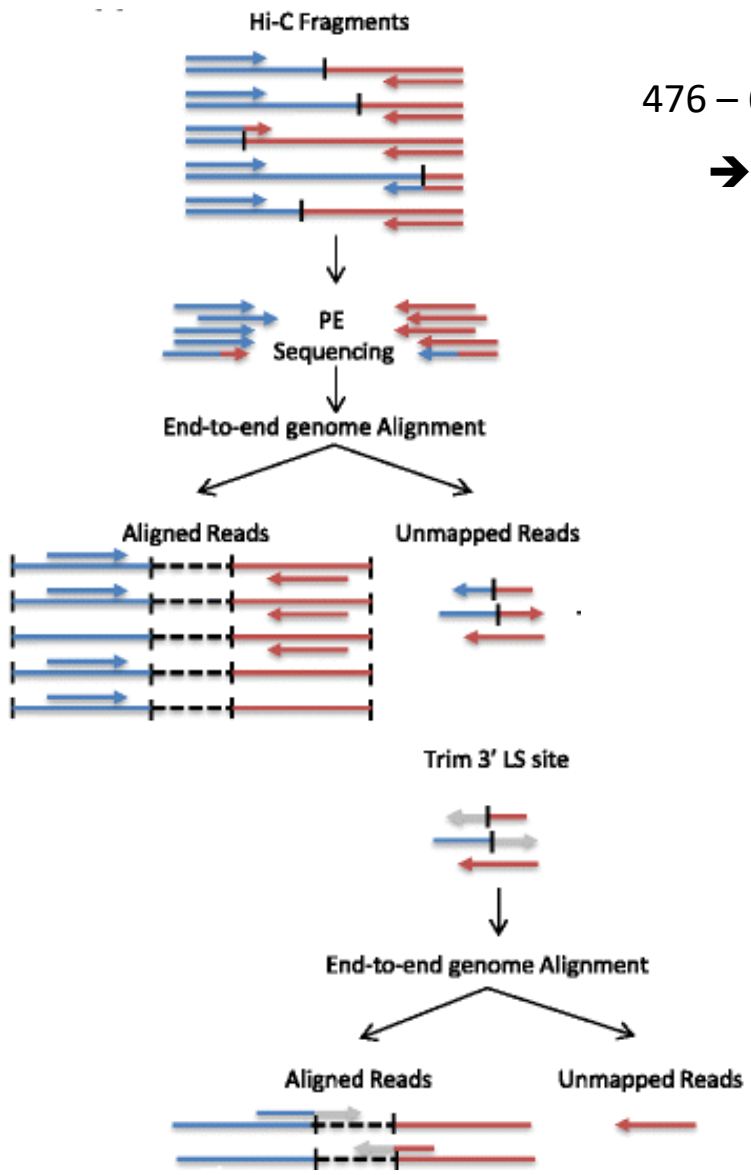
# Hi-C bioinformatics workflow → Read alignment



HiC-Pro (Servant et al. 2015)

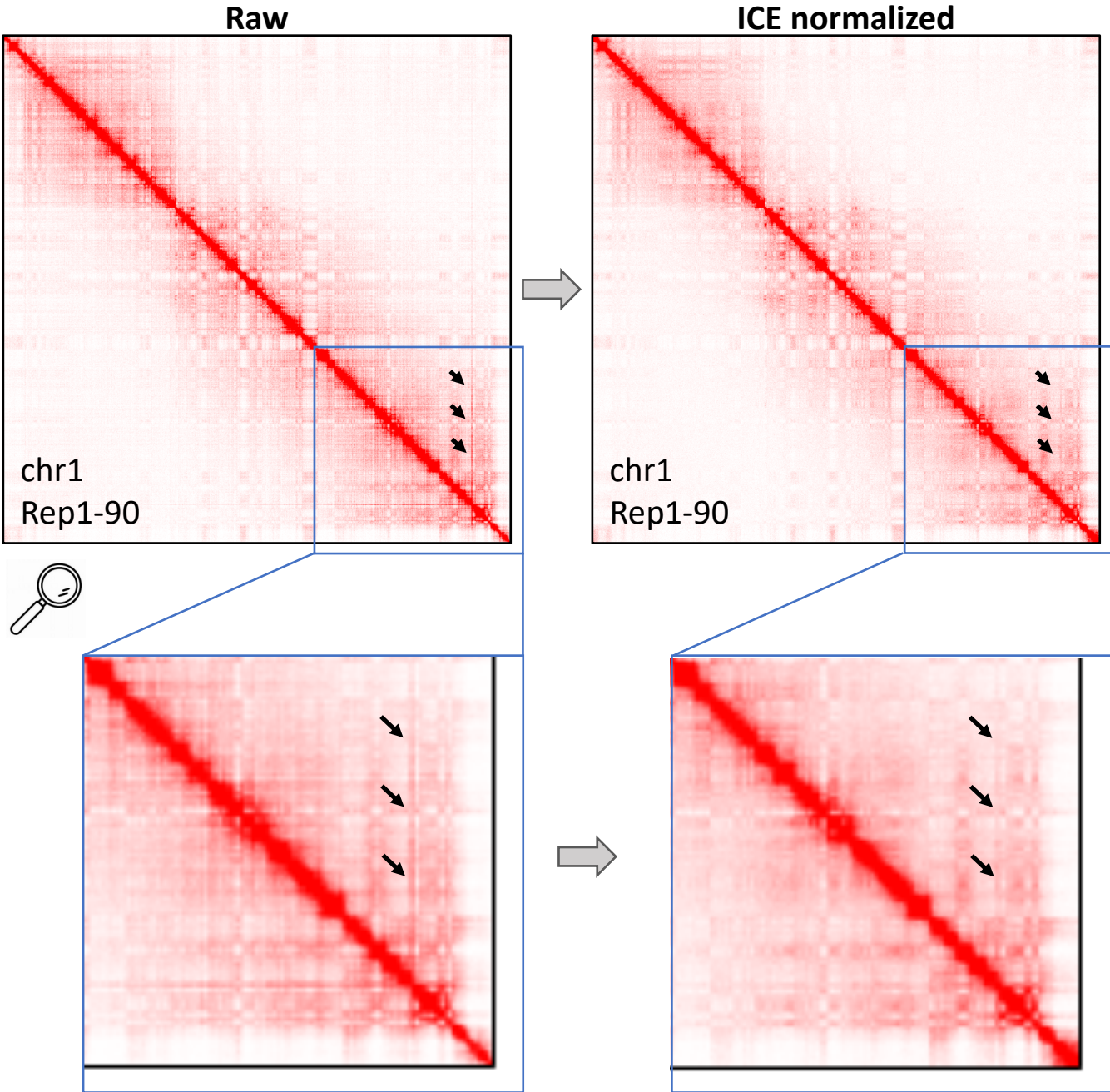
476 – 685 M read pairs / sample

→ 3.45 billion read pairs

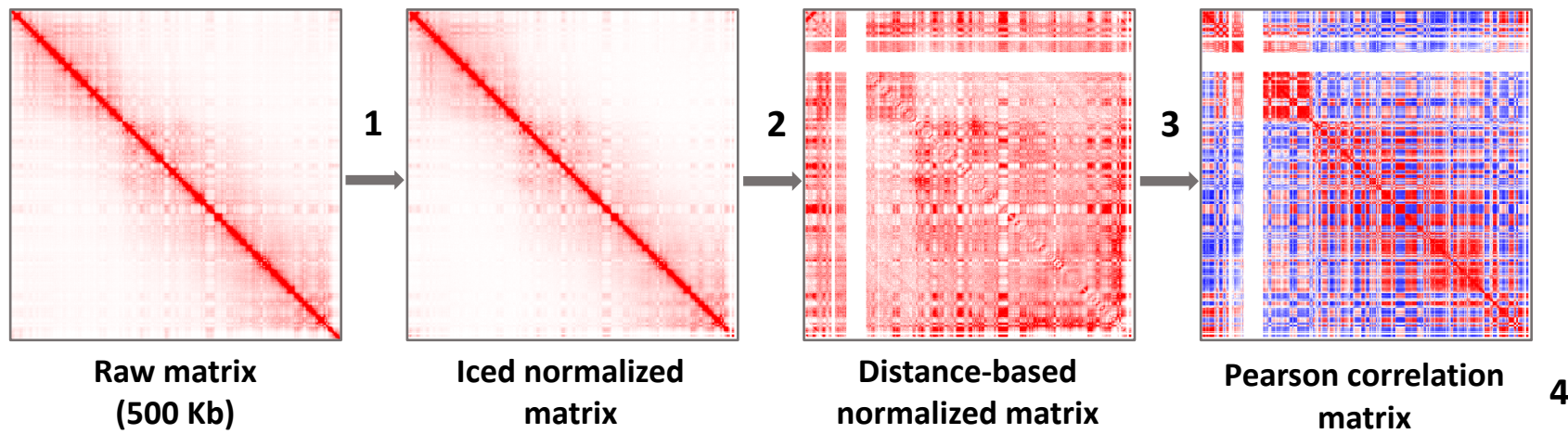




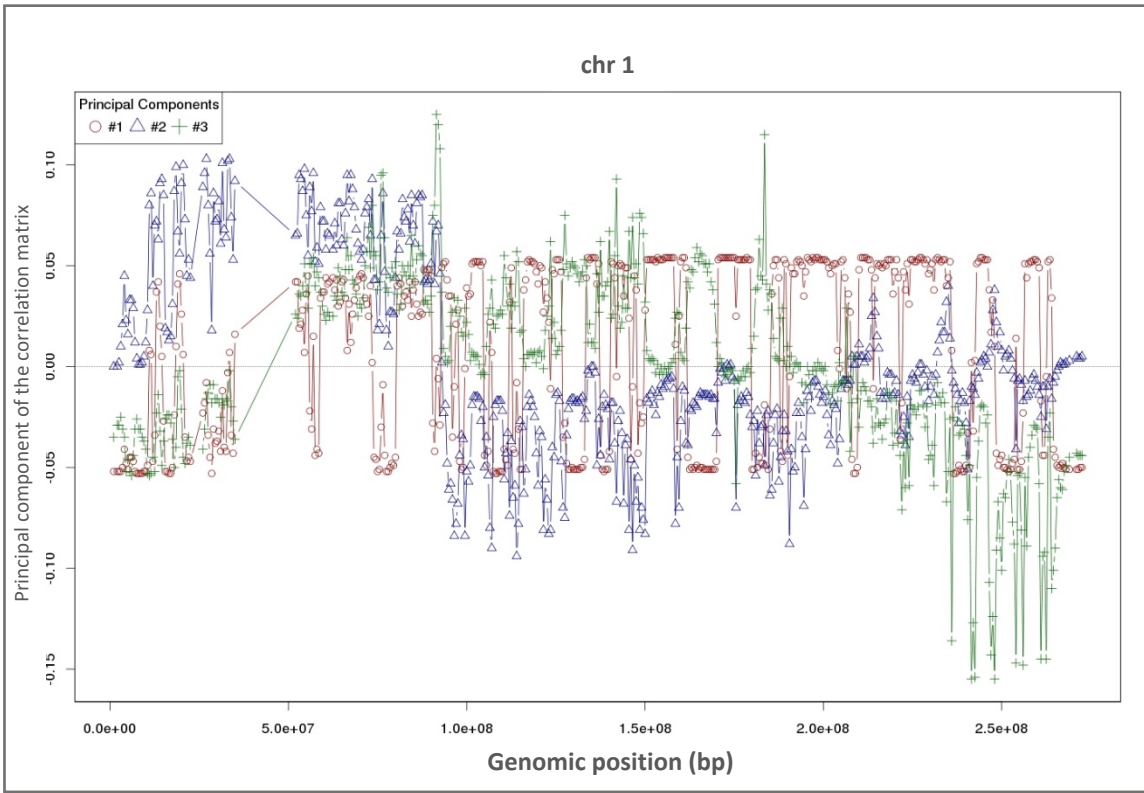
# Normalized matrices



# Identification of A/B compartments

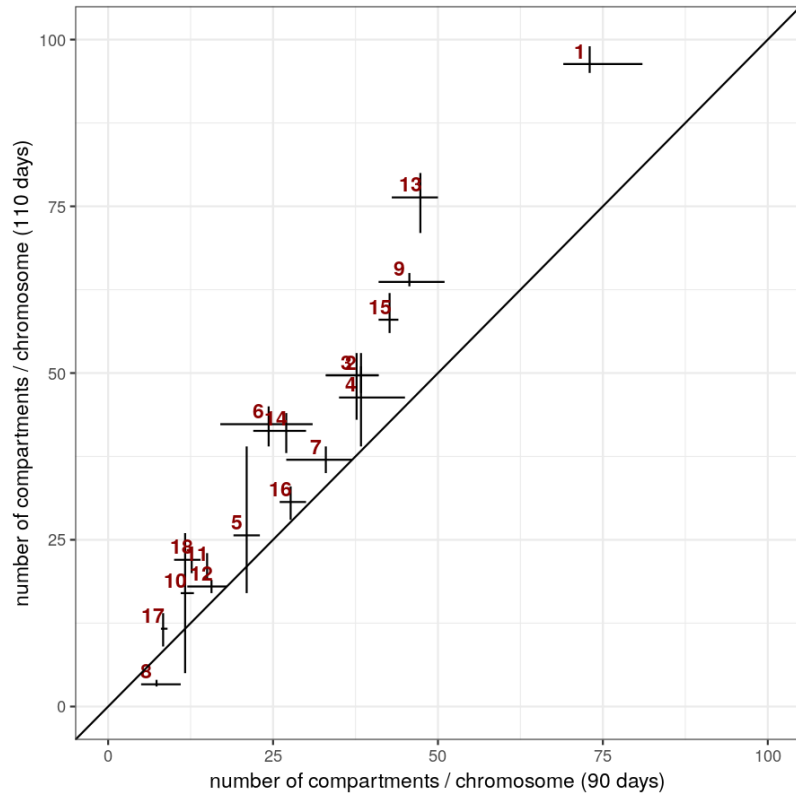


- 1- ICE normalization (matrix balancing)
- 2- « Distance normalization » (observed/expected)
- 3- Pearson correlation matrix
- 4- Principal Component Analysis on the bins

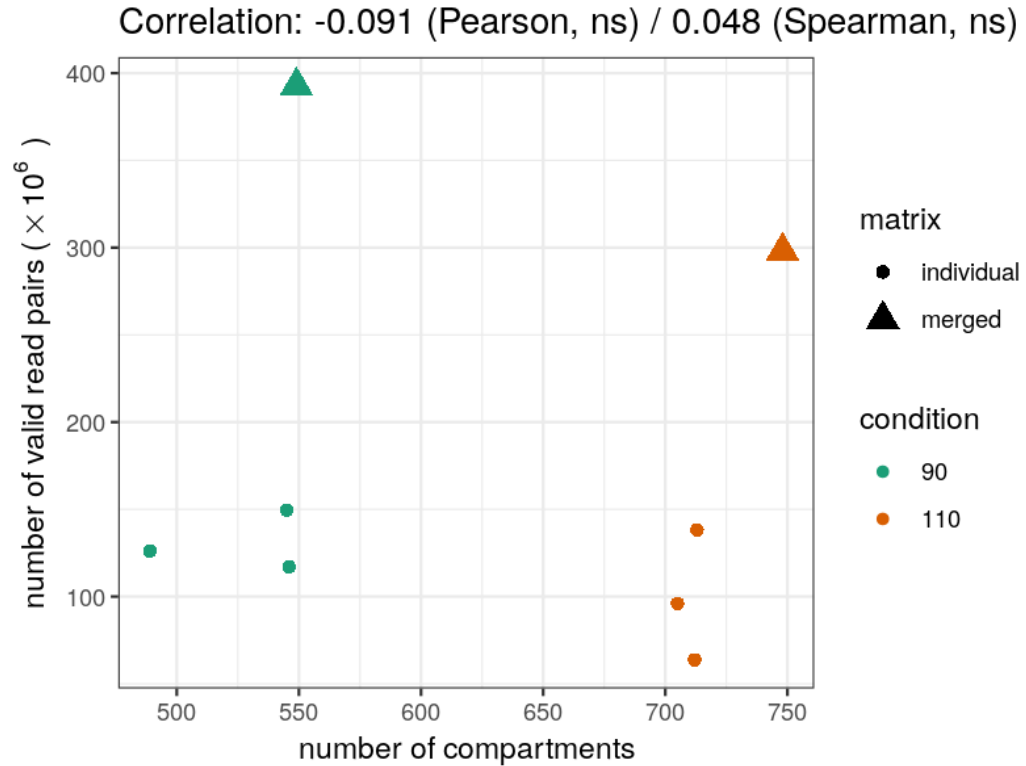


# Genome-wide fragmentation during development

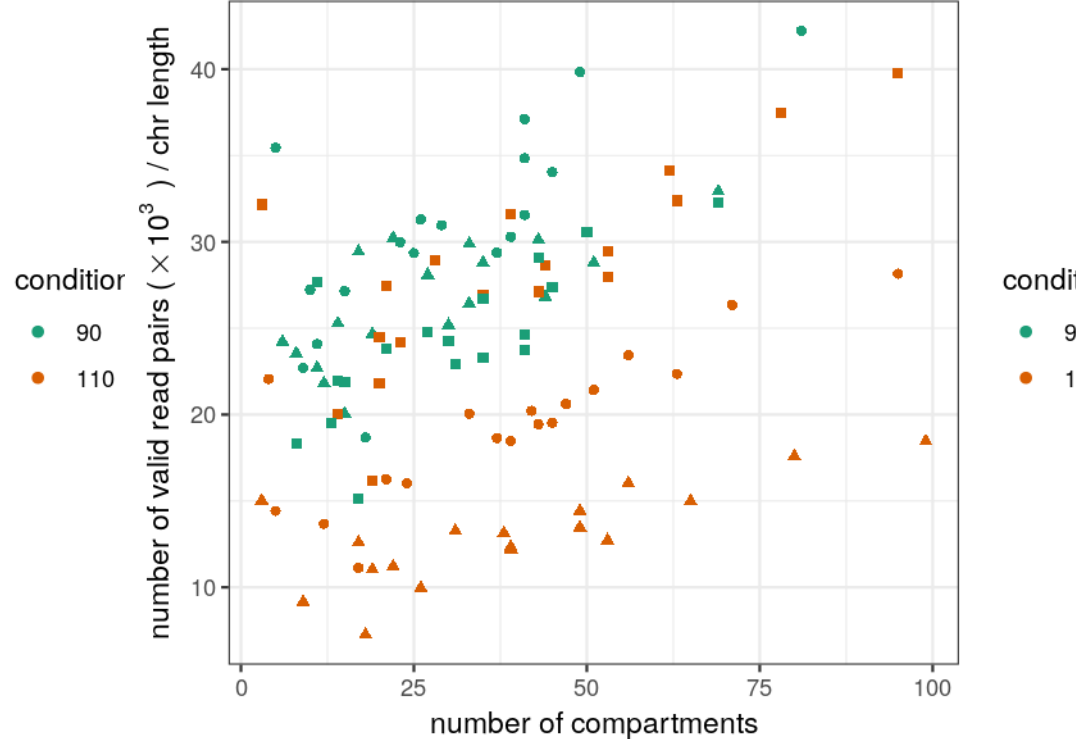
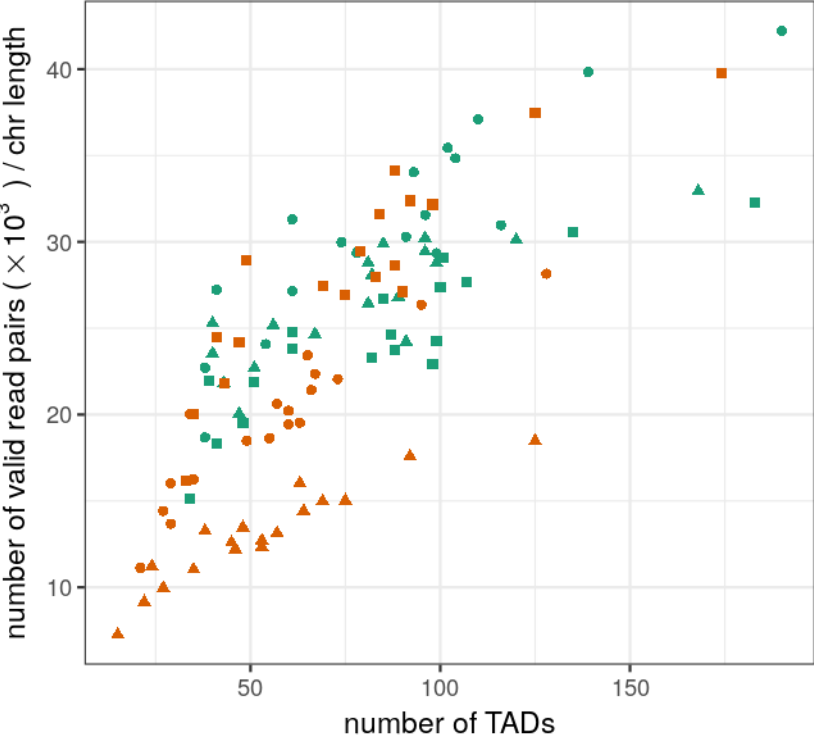
## Number distribution of compartments



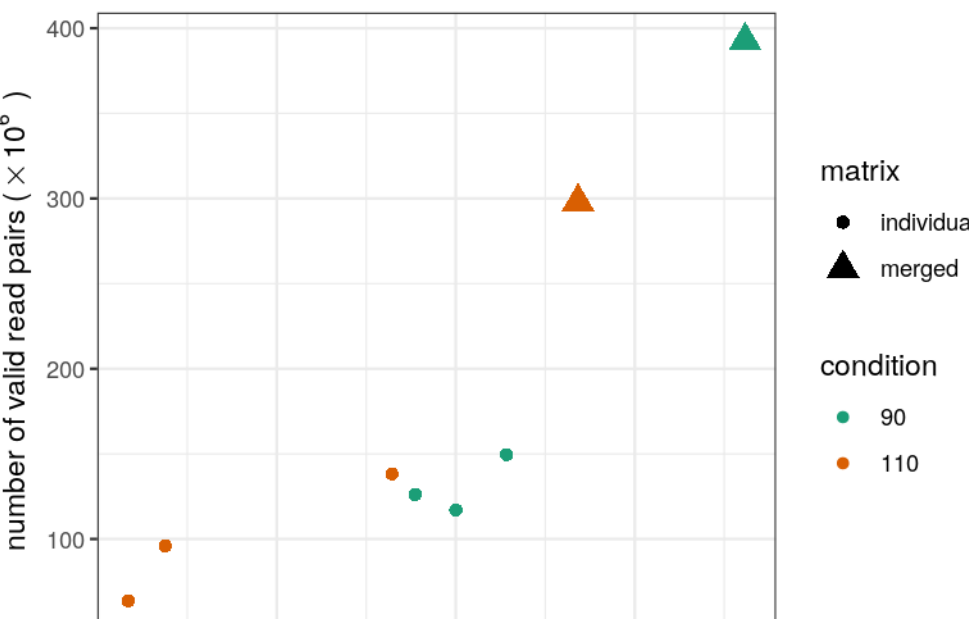
## Number compartments vs. coverage



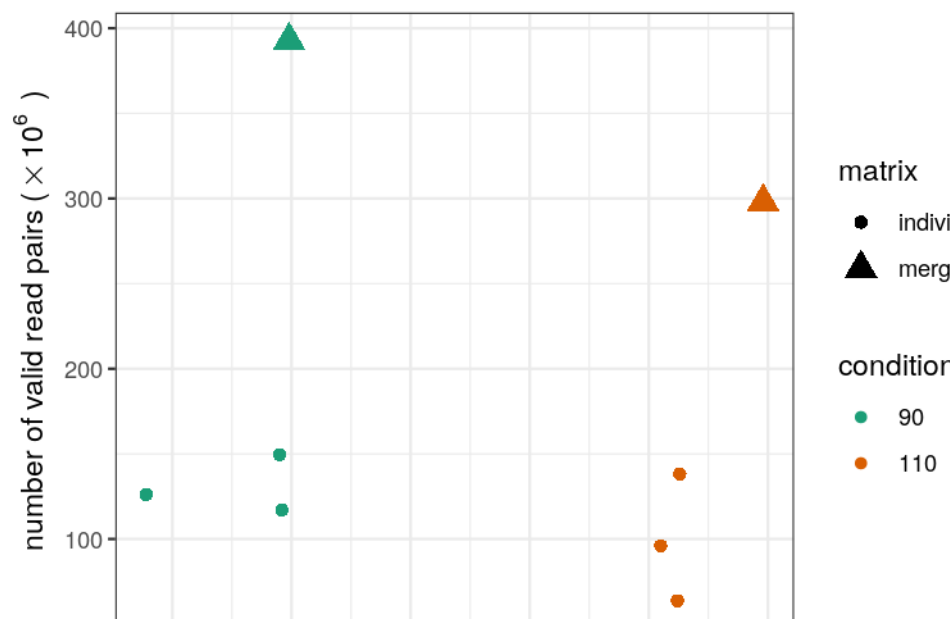
Fragmentation of genome  
compartmentalization



Correlation: 0.871 (Pearson, \*\*\*) / 0.905 (Spearman, \*\*\*)



Correlation: -0.091 (Pearson, ns) / 0.048 (Spearman, ns)



# Differentially distal genomic regions

- Filtering, normalization and detection of bin pairs with significant number of contacts (method: Generalized Linear Model “GLM” functionality of edgeR)

	500 Kb	200 Kb
Total bin pairs with any count	9,262,199	3,844,272
Differential bin pairs	10,183 (0.11%)	3,417 (0.09%)
% differential bin pairs with logFC(+)	56.9	50.7
% differential bin pairs with logFC(-)	43.1	49.3

$$\text{logFC (bin pair)} = \log_2 [ (\text{counts at 110 days}) / (\text{counts at 90 days}) ]$$

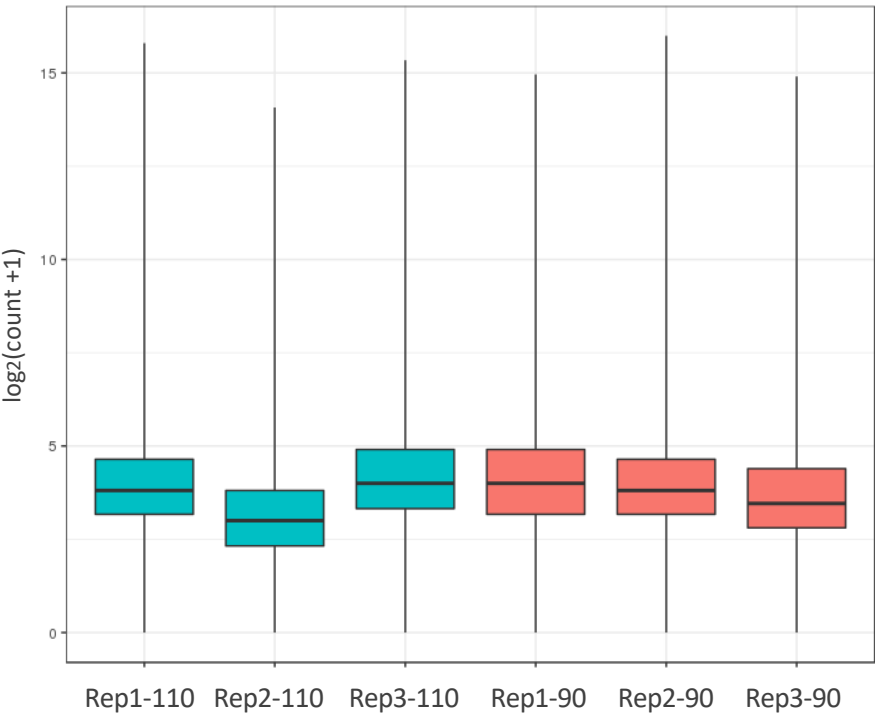
**Positive logFC** = more counts “contacts” at **110 days** than at 90 days  
= **genomic regions closer to each other**

**Negative logFC** = more counts “contacts” at **90 days** than at 110 days  
= **genomic regions closer to each other**

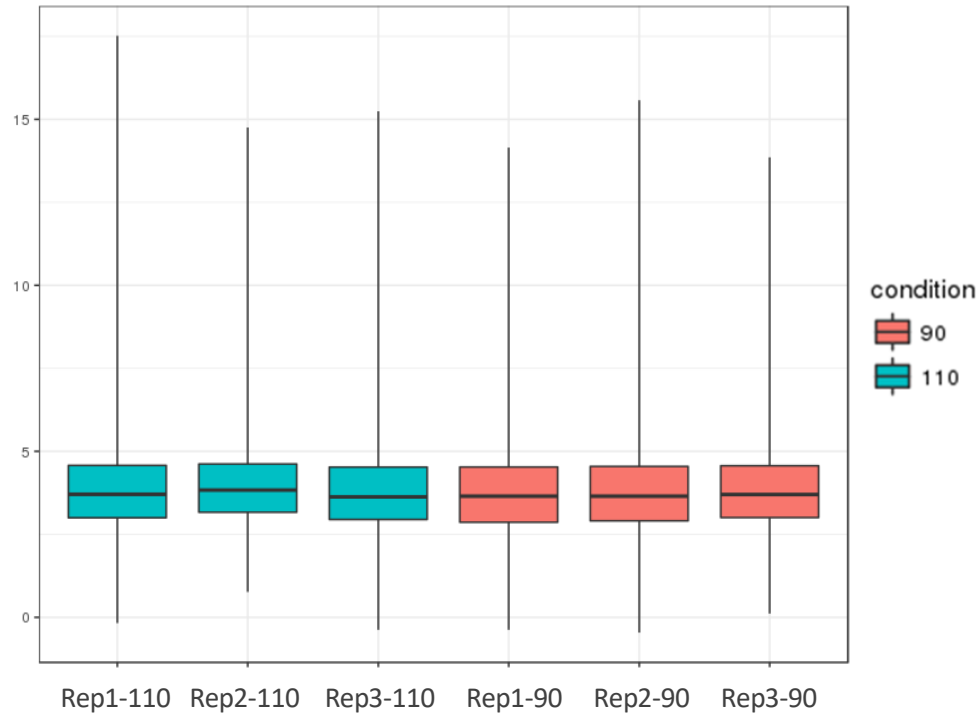
# Differential analysis (90 – 110 days of gestation)

- Raw matrices of the 18 autosomes (500, 200 and 40 kb)
- Inter-matrix normalization

Pseudo counts per simple (before normalization)



Pseudo counts per simple (after normalization)



- Detecting pairs of bins with a significant difference in the number of counts

Generalized Linear Model based on the negative binomial distribution (edgeR)