

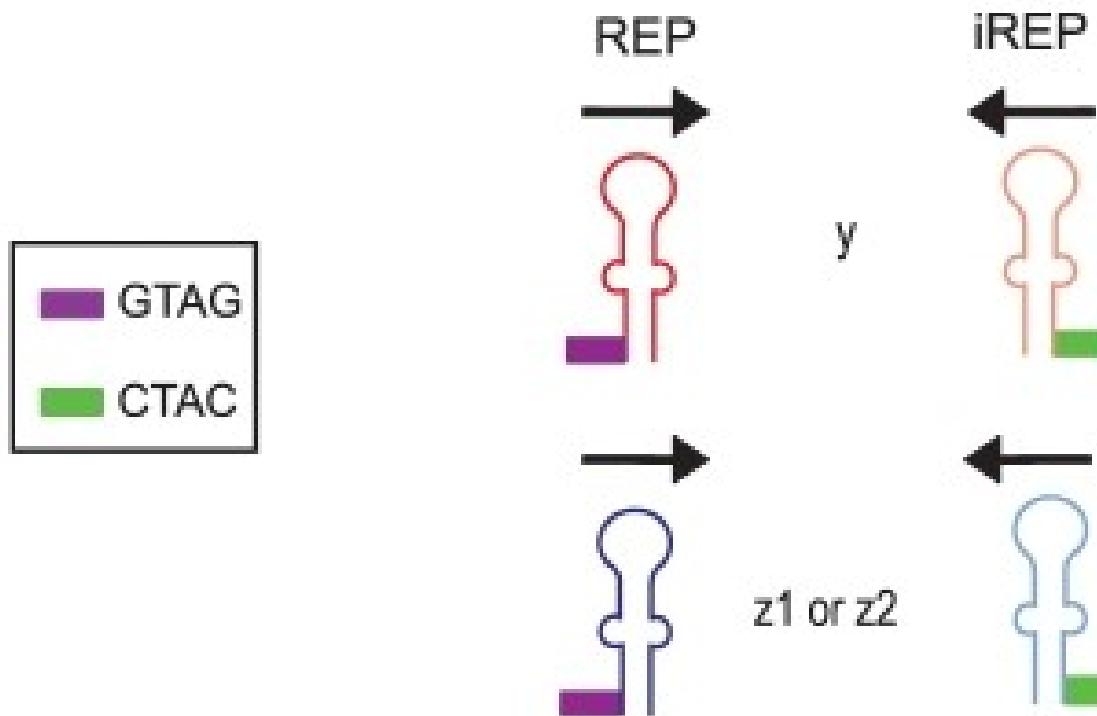
RNA-Seq study of mRNA metabolism in operons: role of BIME in *Escherichia coli*.

Nicolas Jeanne
June 2015

Supervisors: Yves Quentin
Roland Barriot
Reporter : Christine Gaspin

REP & BIME features

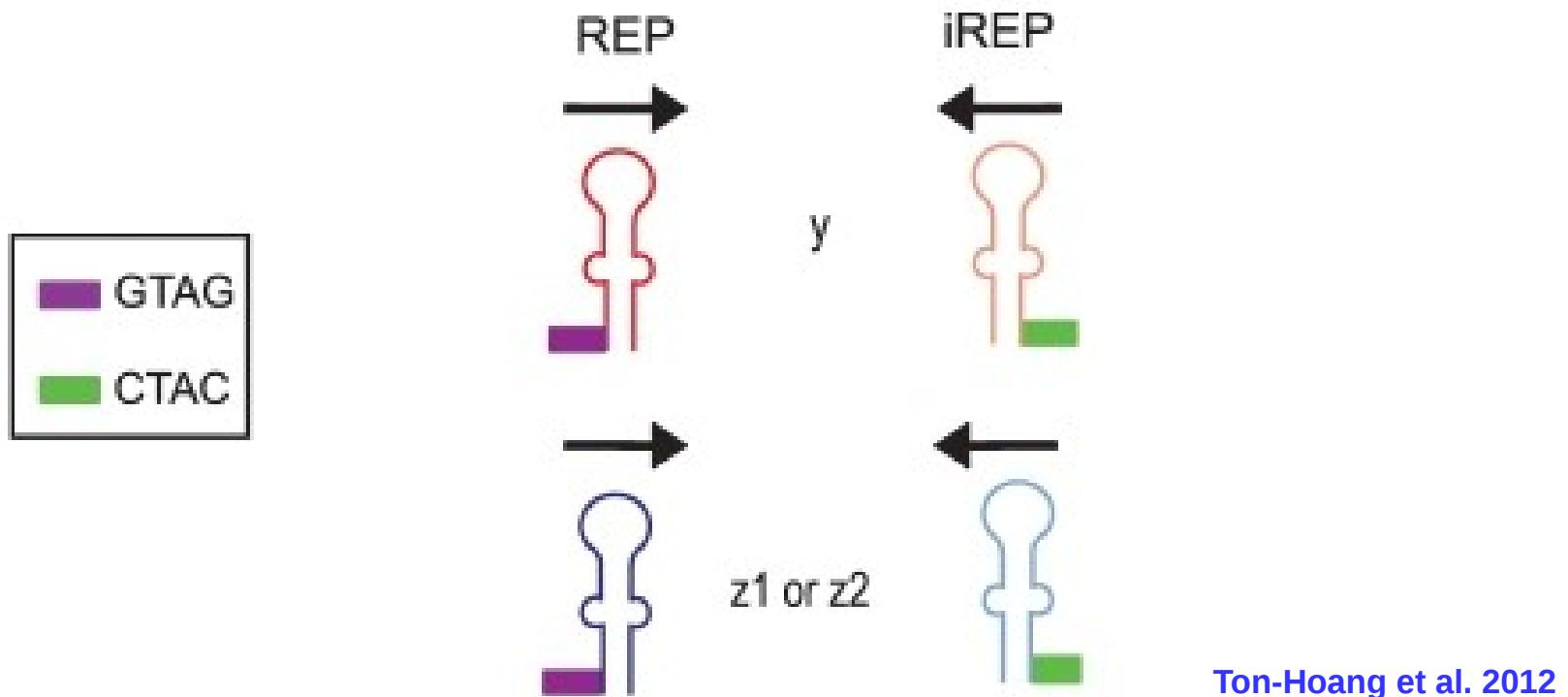
- Repeated Extragenic Palindrome



Ton-Hoang et al. 2012

REP & BIME features

- Repeated Extragenic Palindrome



- BIME set-up (Bacterial Interspersed Mosaic Element)

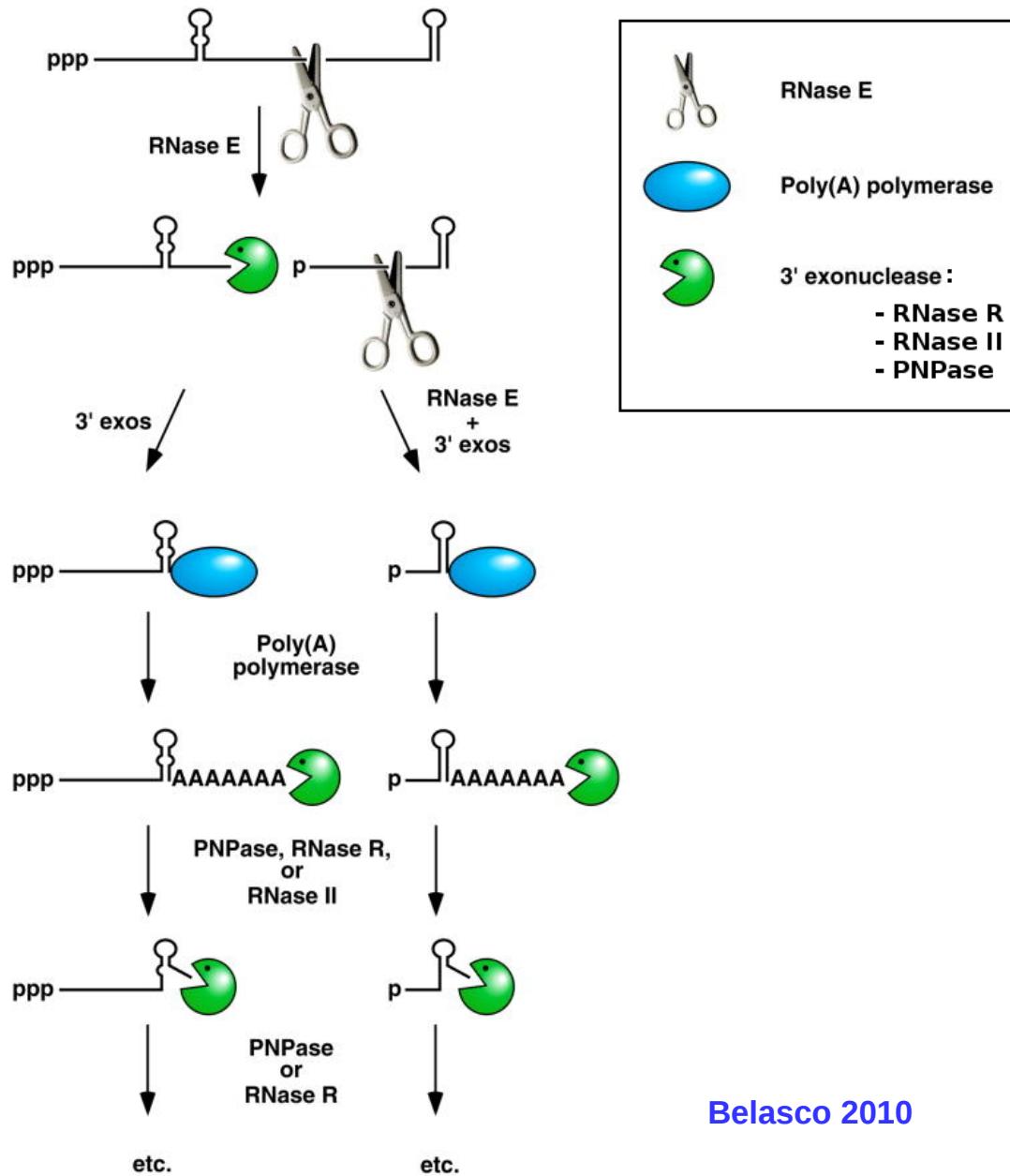
BIME properties?

- Structural roles:
 - ◆ Homologous recombination
 - ◆ Mobile DNA sequence insertion sites
- DNA level:
 - ◆ Protein binding sites
 - ◆ Transcription initiation
 - ◆ Site specific recombination sites
- RNA level:
 - ◆ Control of translation
 - ◆ Transcription termination
 - ◆ 5' end transcript stabilisation

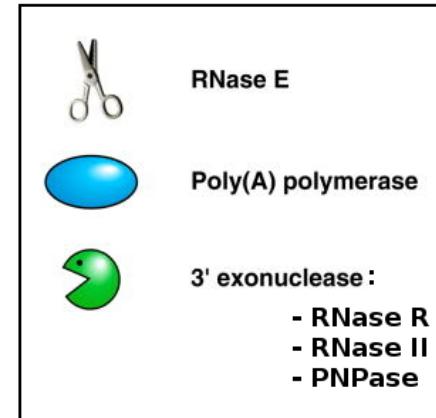
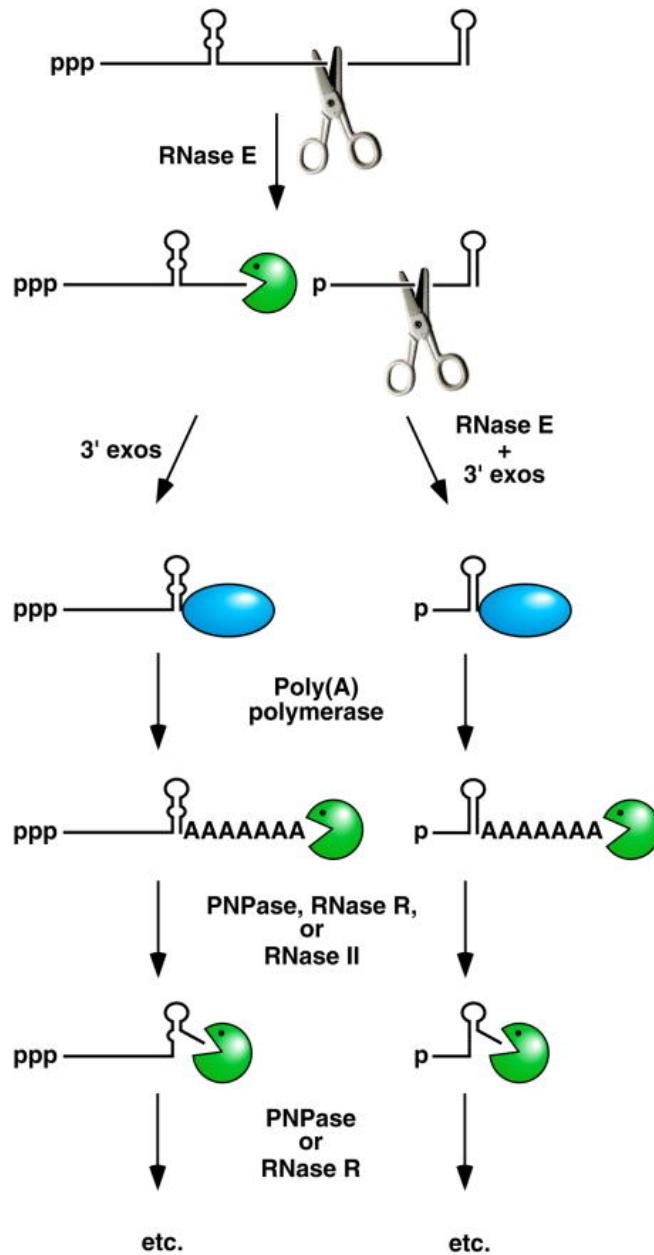
BIME properties?

- Structural roles:
 - ◆ Homologous recombination
 - ◆ Mobile DNA sequence insertion sites
- DNA level:
 - ◆ Protein binding sites
 - ◆ Transcription initiation
 - ◆ Site specific recombination sites
- RNA level:
 - ◆ Control of translation
 - ◆ **Transcription termination?**
 - ◆ **5' end transcript stabilisation?**

mRNA degradosome



mRNA degradosome

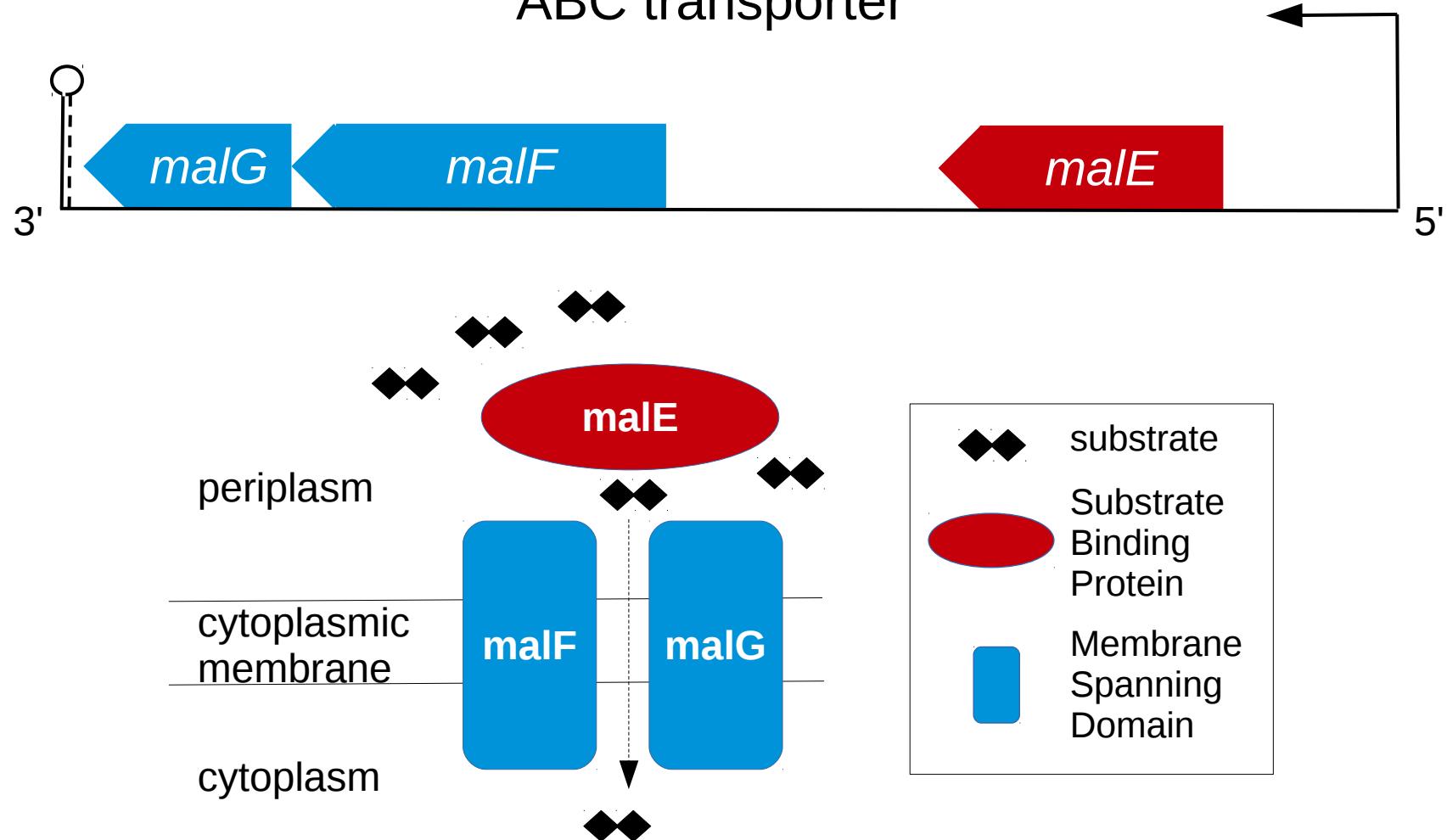


BIME secondary
structures?

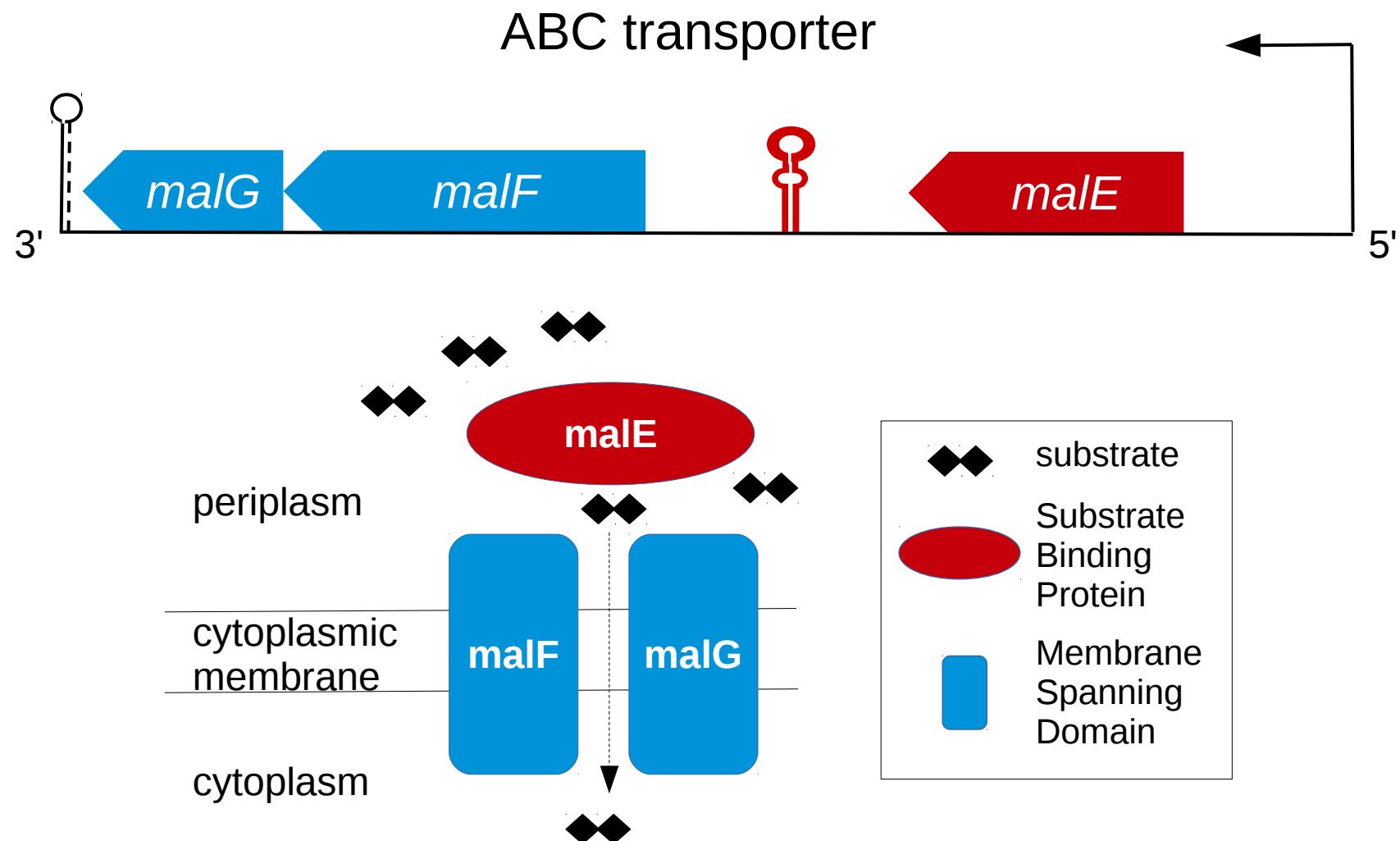
Belasco 2010

malEFG operon

ABC transporter

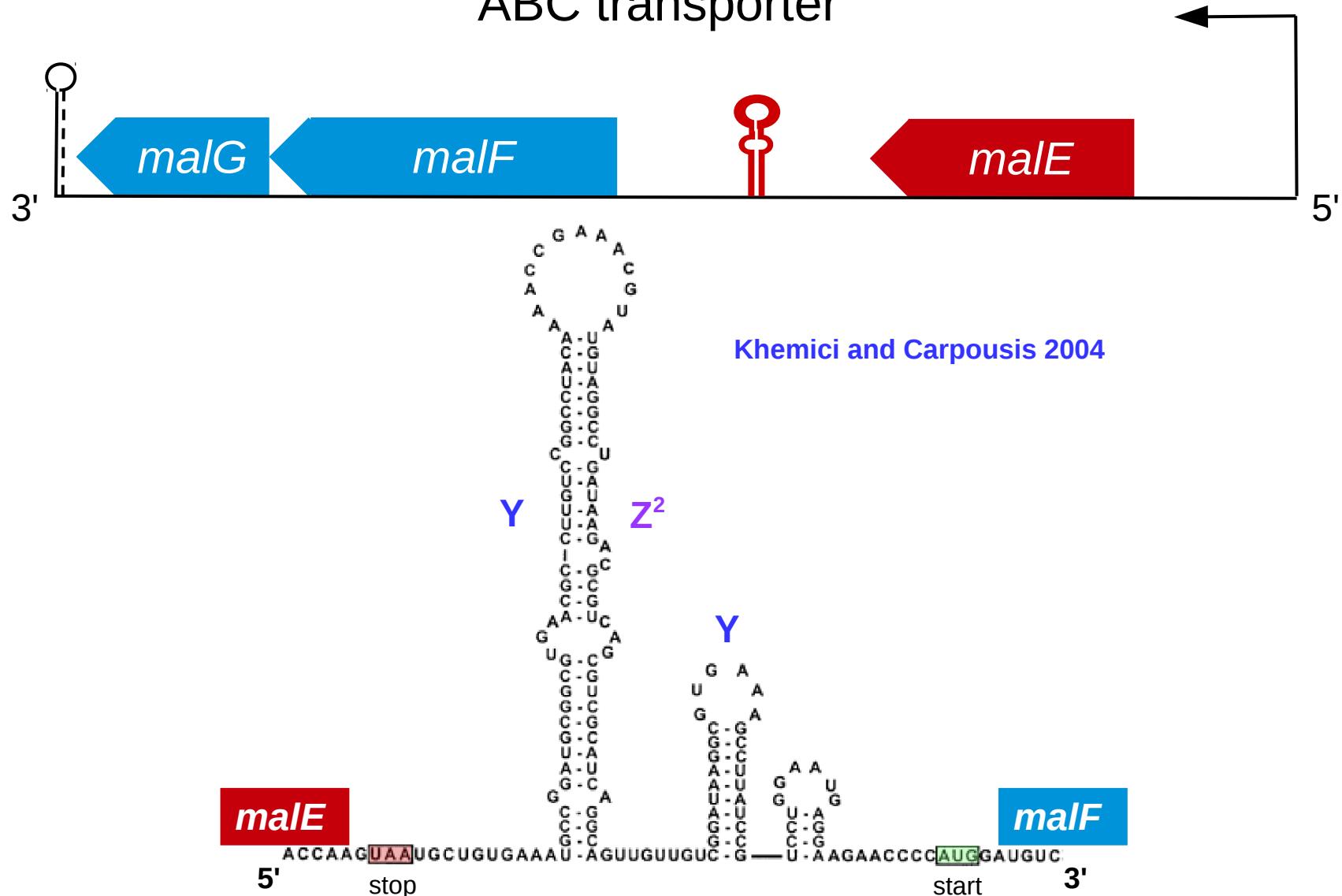


malEFG operon



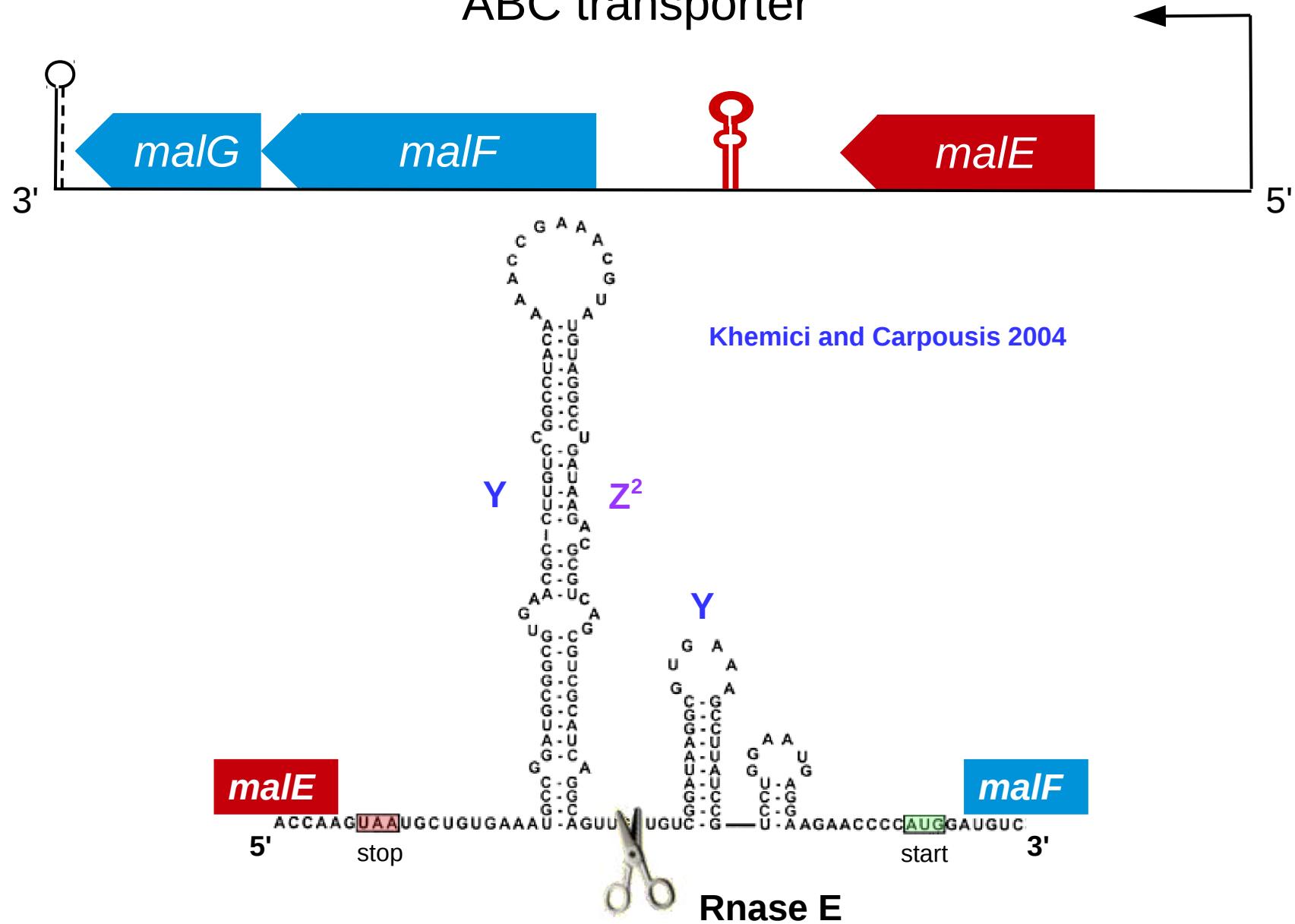
malEFG operon

ABC transporter

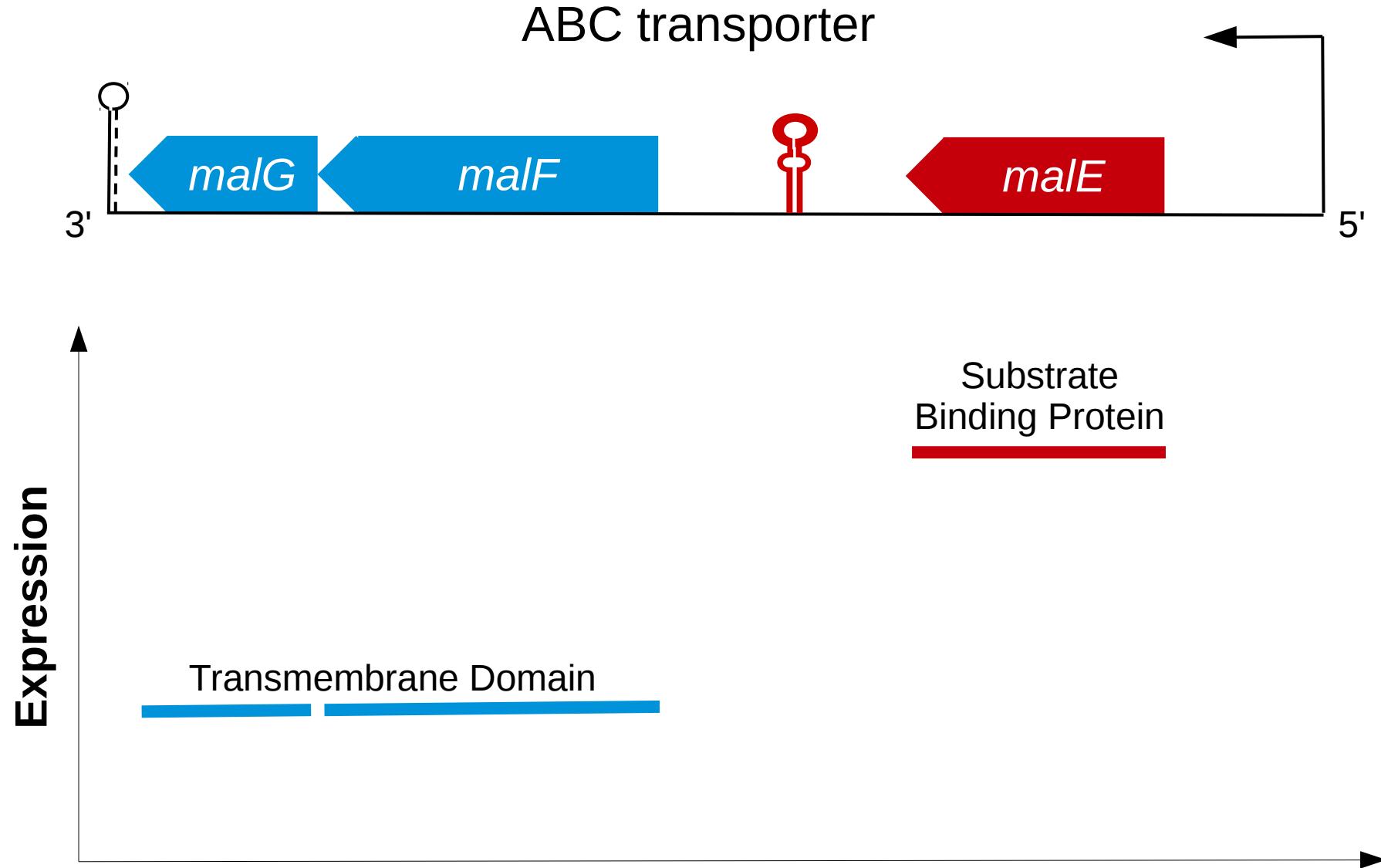


malEFG operon

ABC transporter



malEFG operon



BIME effect study in whole genome

- *E. coli* K12 MG1655 genome:

- 4605 genes ([NCBI : NC_000913_2](#))
- 848 operons > 1 gene ([RegulonDB](#))
- 605 REPs, 287 BIMEs (Team, [Mathias Weyder 2013](#))
- 36 operons with 39 BIMEs
- Promotors ([RegulonDB](#)), Transcription terminators ([Door²DB](#))

BIME effect study in whole genome

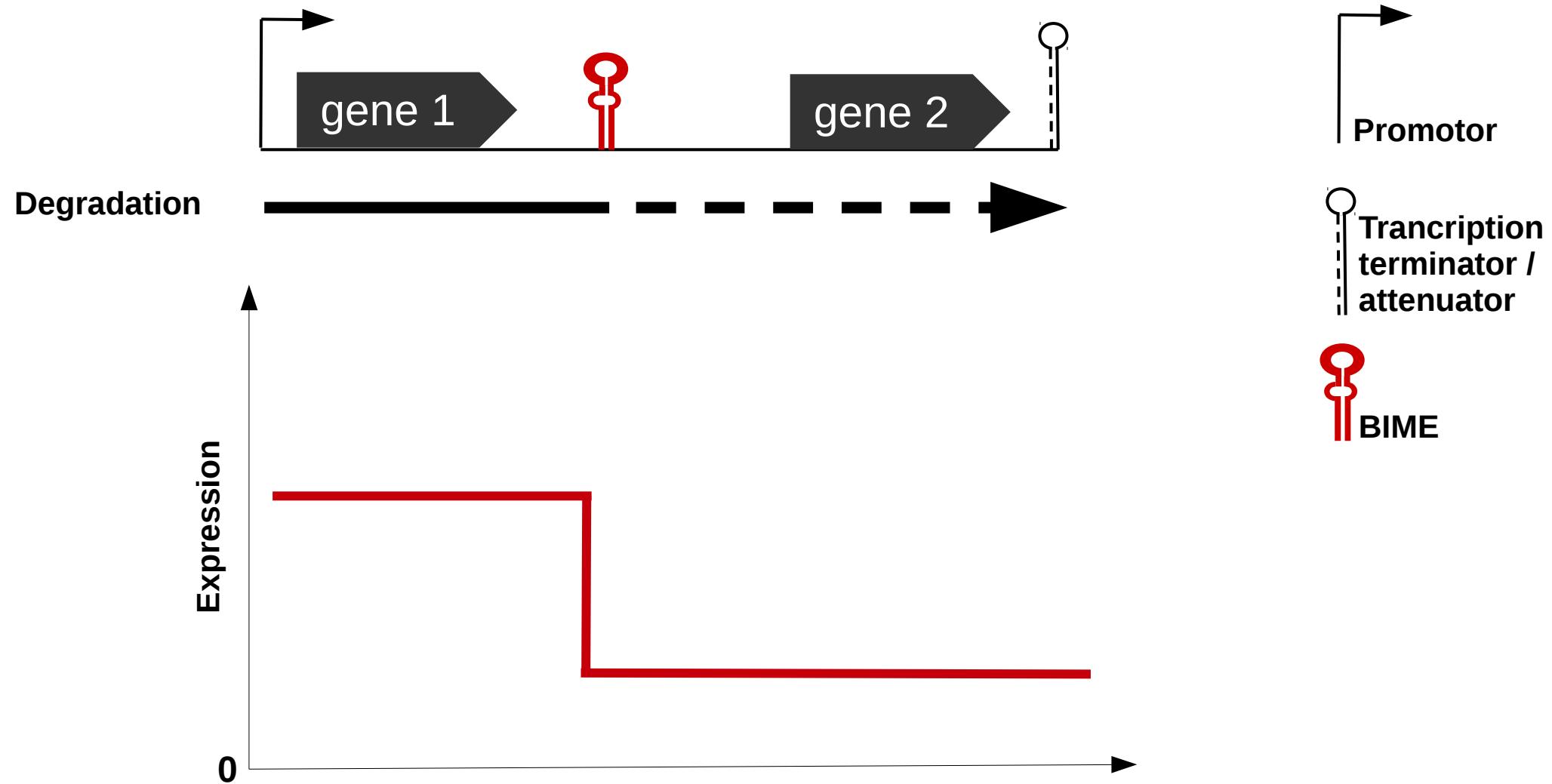
- *E. coli* K12 MG1655 genome:

- 4605 genes ([NCBI : NC_000913_2](#))
- 848 operons > 1 gene ([RegulonDB](#))
- 605 REPs, 287 BIMEs (Team, [Mathias Weyder 2013](#))
- 36 operons with 39 BIMEs
- Promotors ([RegulonDB](#)), Transcription terminators ([Door²DB](#))

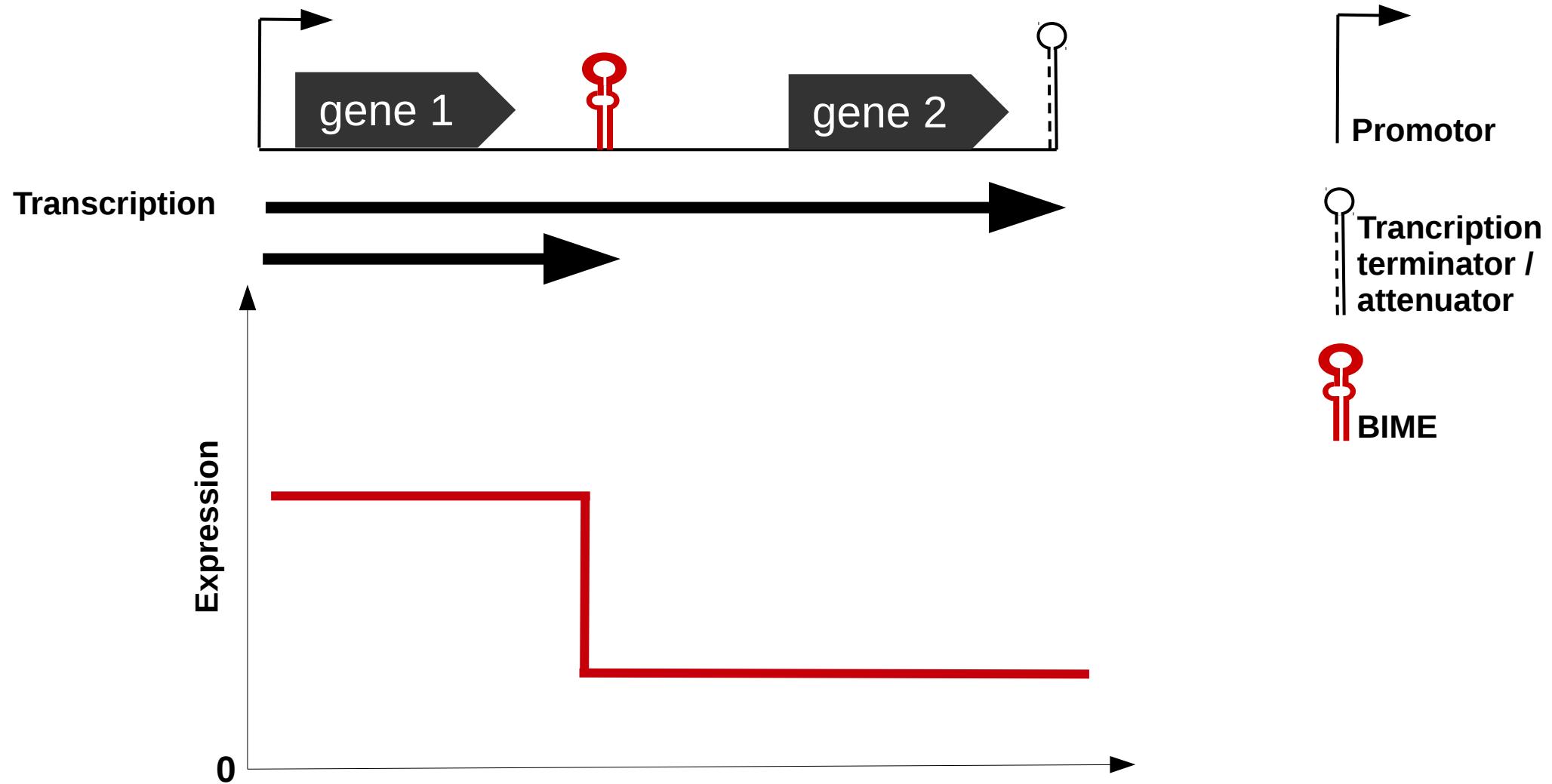
- RNA-Seq:

- Whole transcriptome study
- Snapshot of transcription & degradation
- 3 data sets from 2 experiments ([GEO database](#))
- 8, 3 and 3 replicates

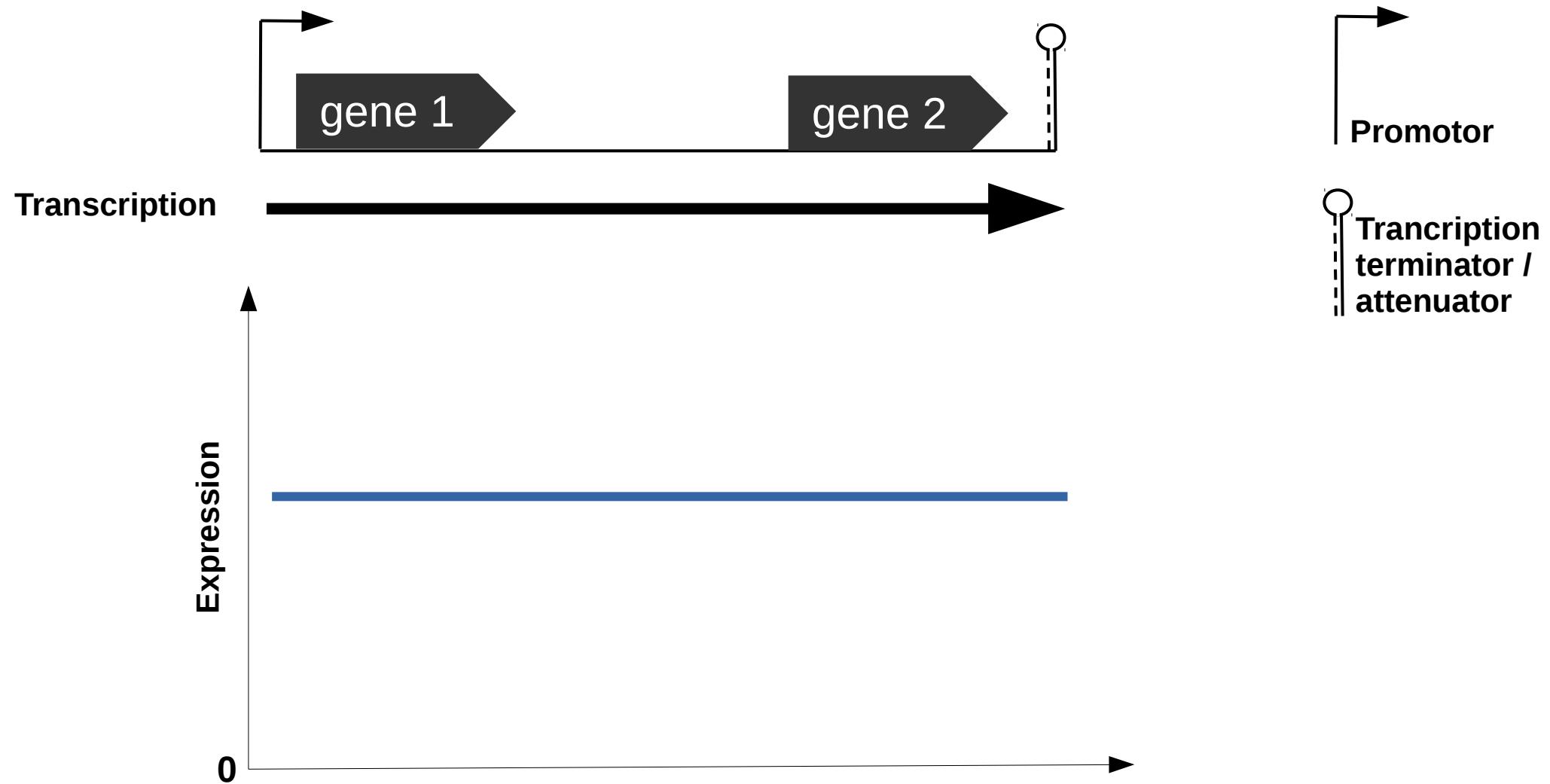
Possible expression patterns



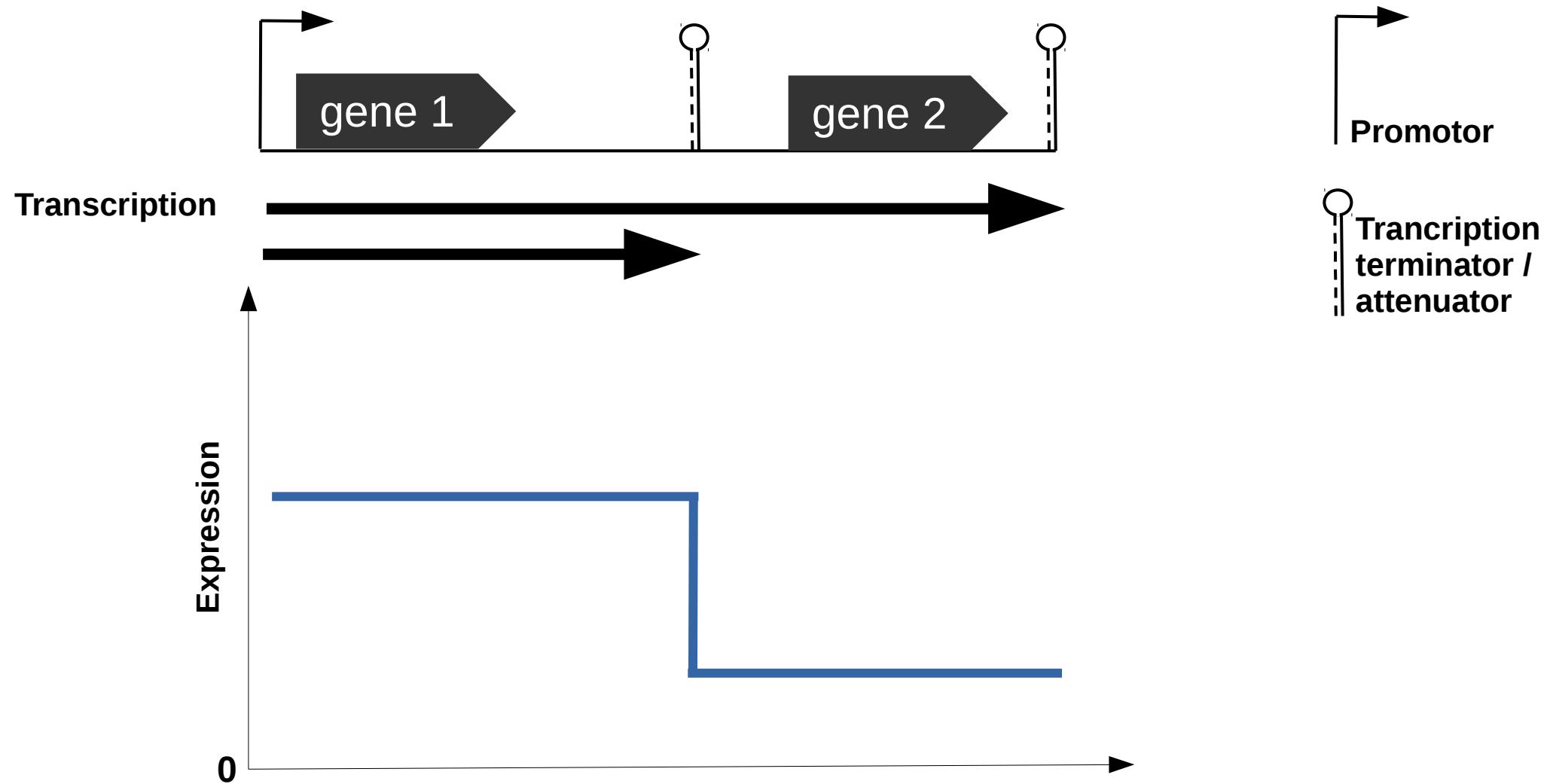
Possible expression patterns



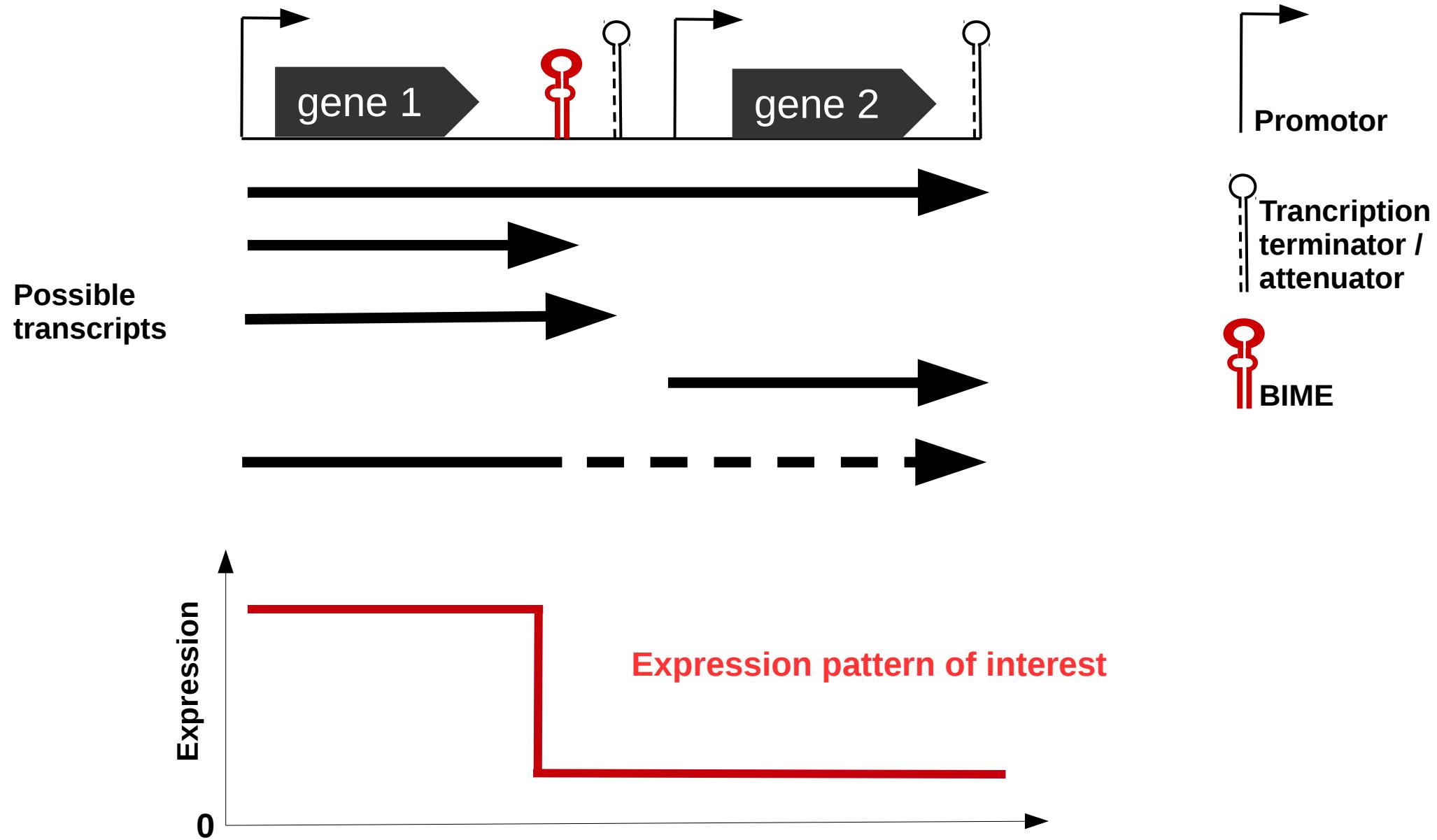
Possible expression patterns



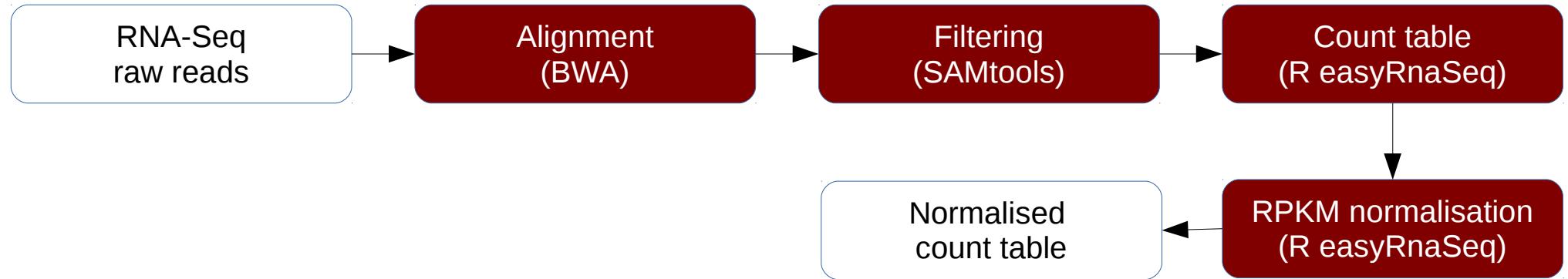
Possible expression patterns



Possible expression patterns



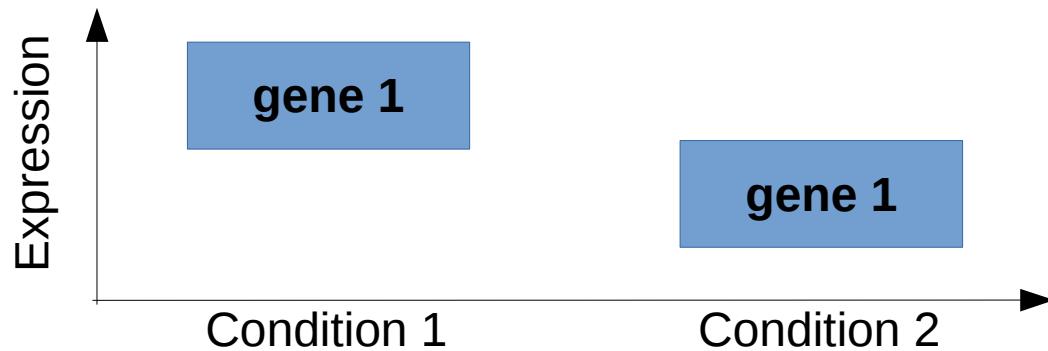
Analysis pipeline



 Data pre-processing

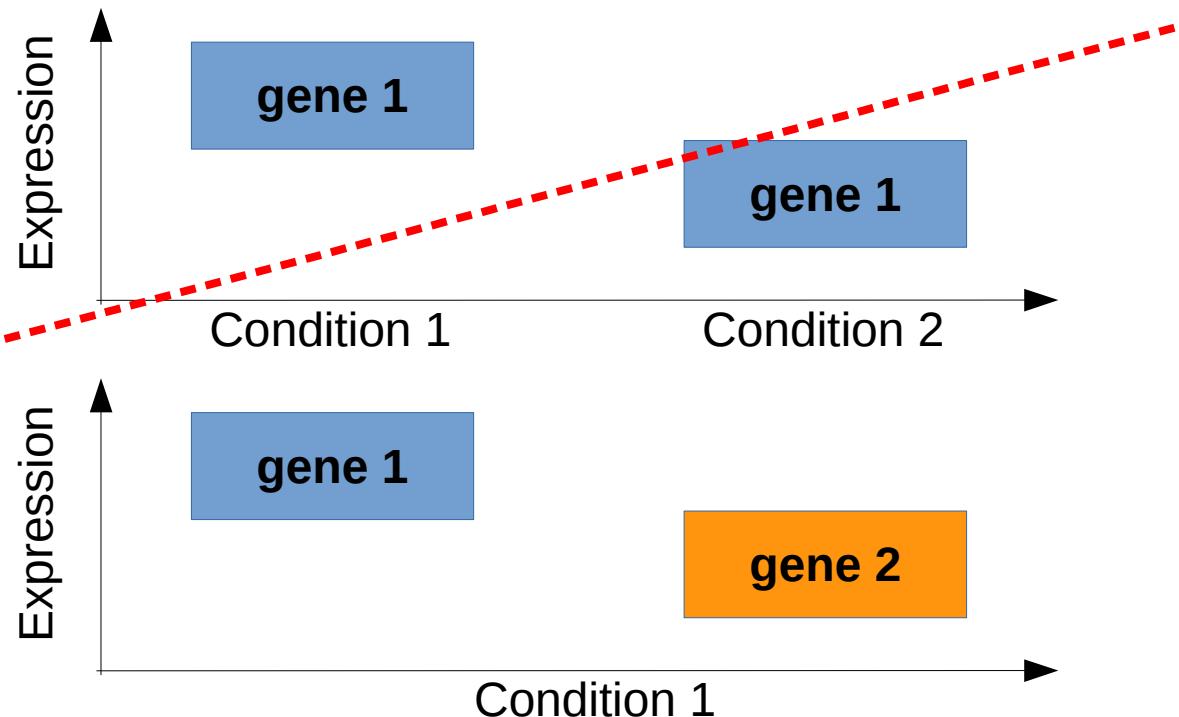
RNA-Seq normalisation

- Keep in mind:



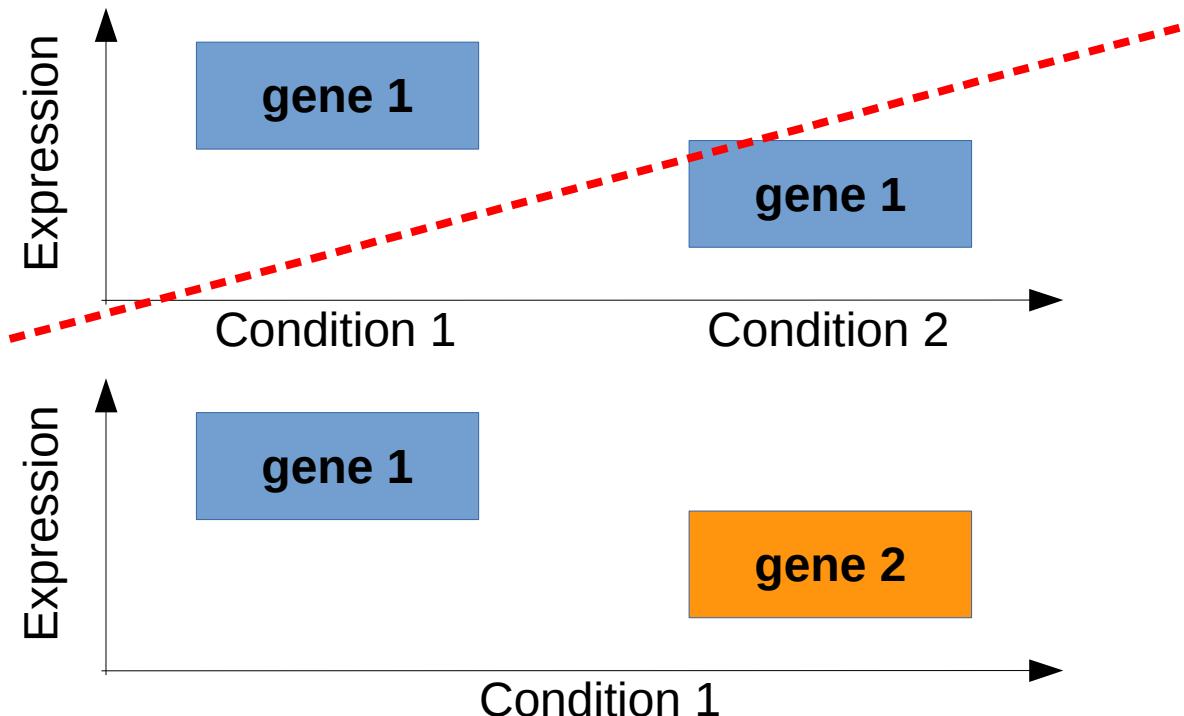
RNA-Seq normalisation

- Keep in mind:



RNA-Seq normalisation

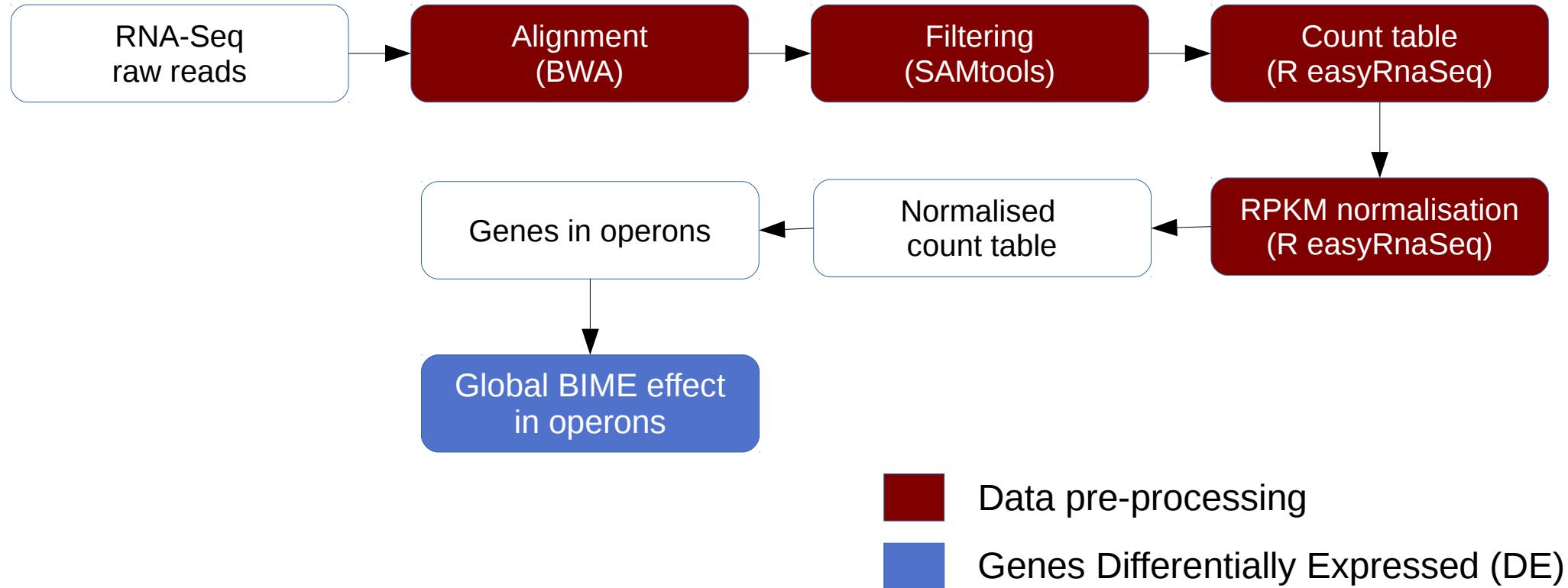
- Keep in mind:



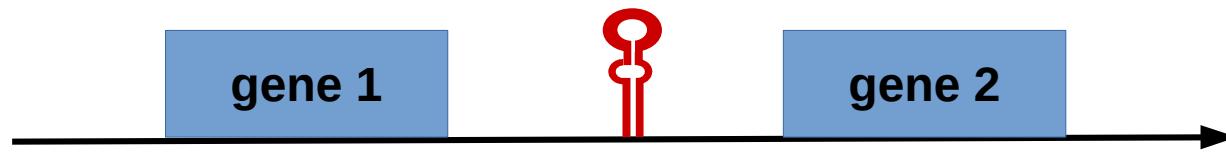
- RPKM normalisation:

- Transcript size
- Library size

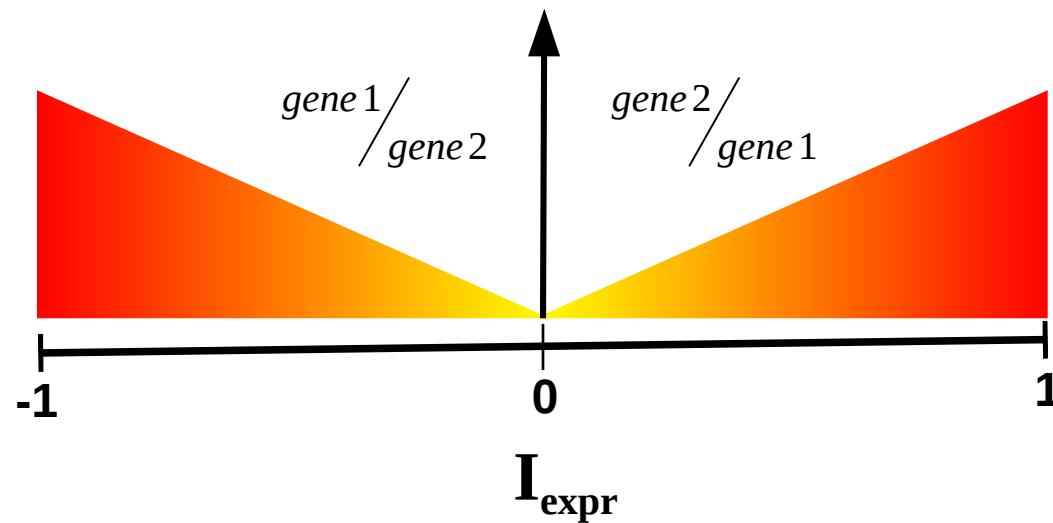
Analysis pipeline



Global BIME effect in operons

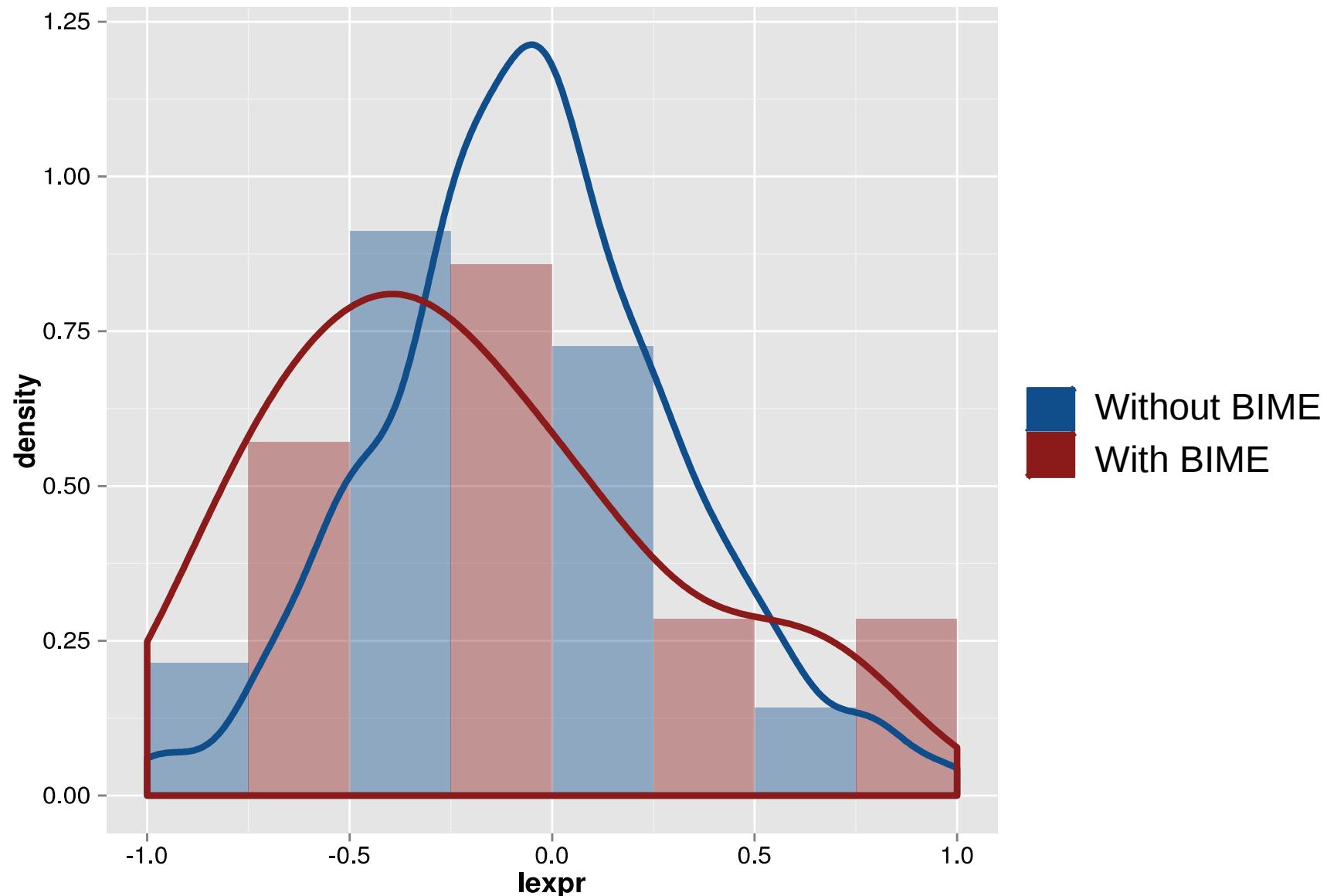


$$I_{expr} = \frac{RPKM_{gene2} - RPKM_{gene1}}{RPKM_{gene2} + RPKM_{gene1}}$$

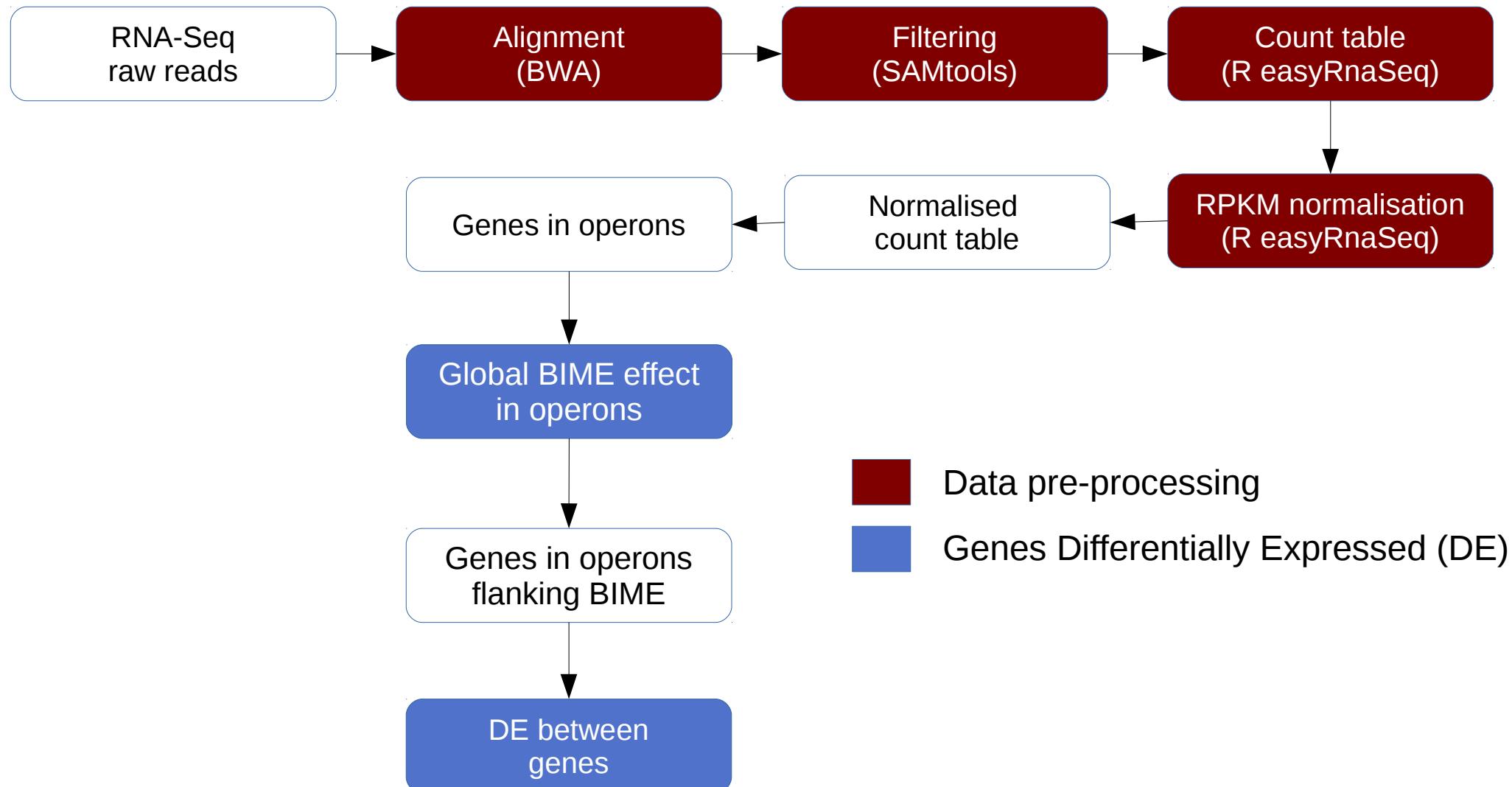


Global BIME effect in operons

Distributions of I_{expr} without annotated promotor and terminator in intergenic region

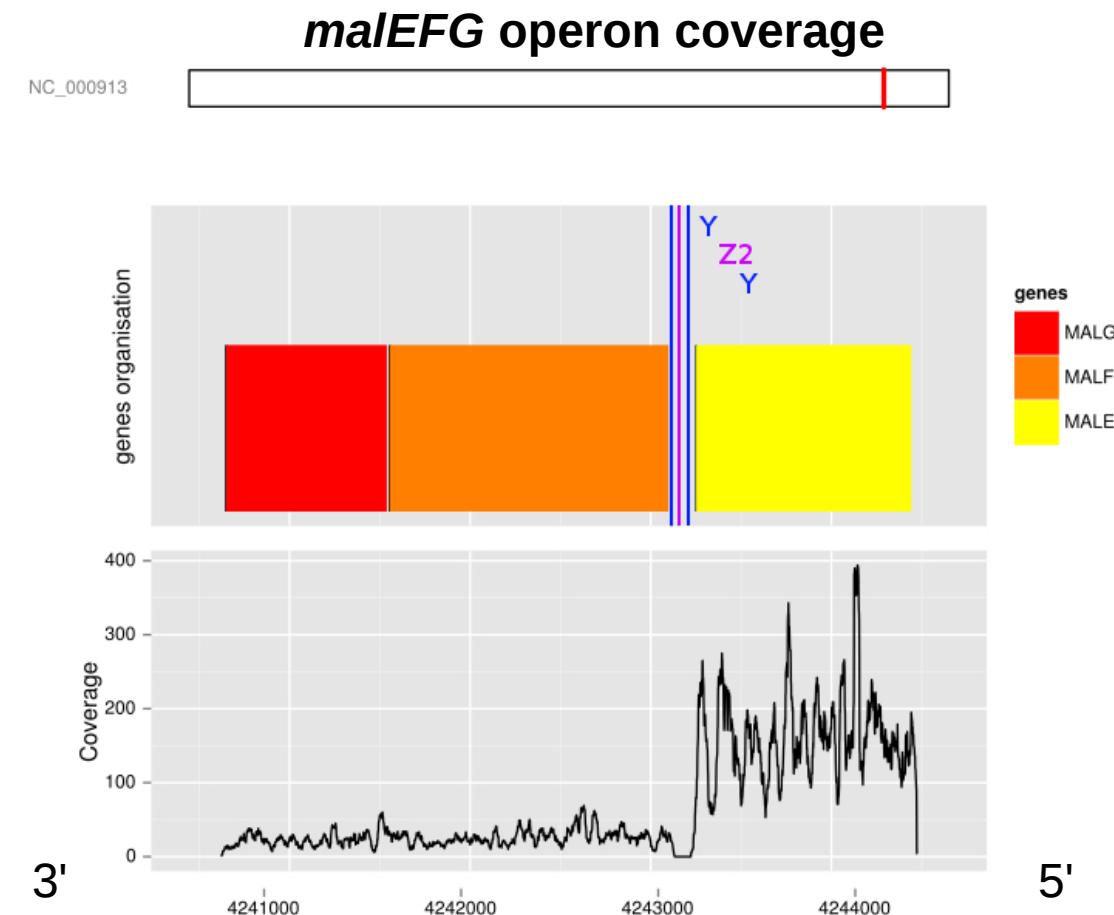


Analysis pipeline



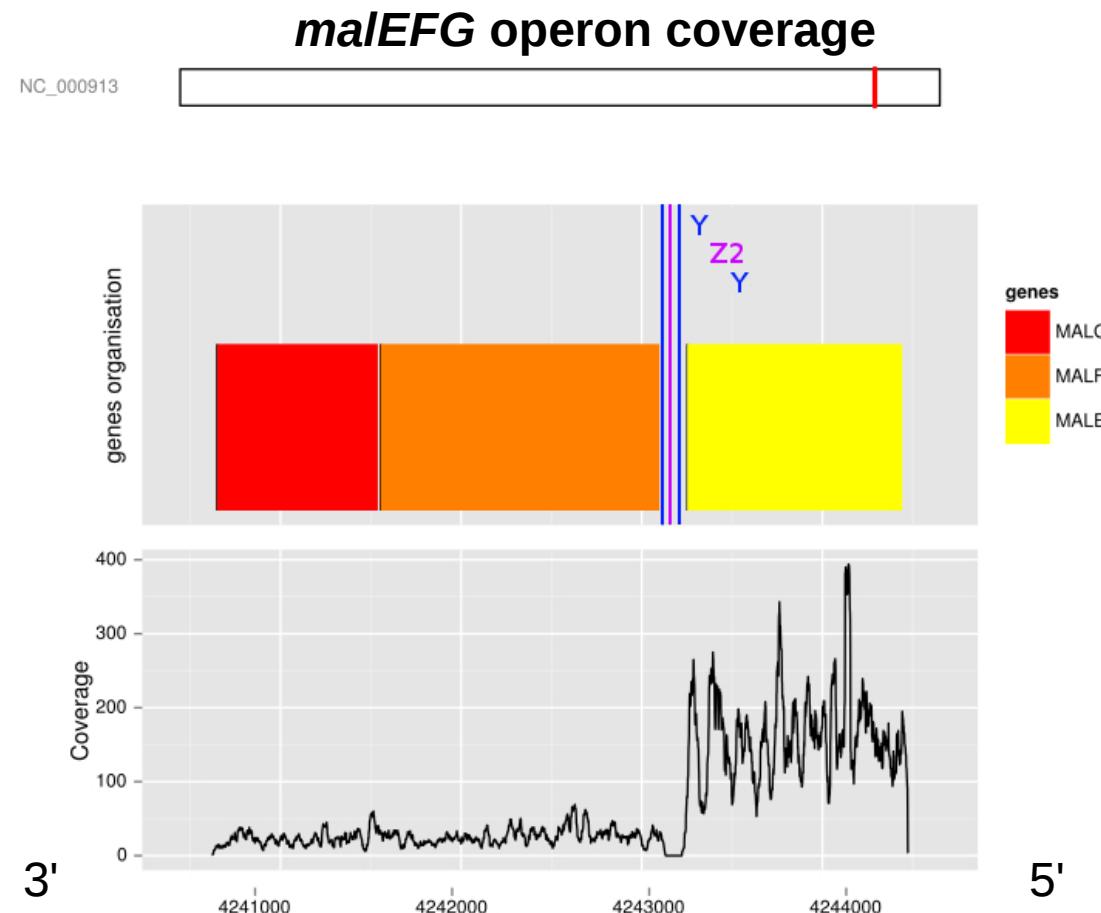
Genes flanking BIME Differential Expression?

- Wilcoxon rank sum test for DE (p-value threshold : 0.01)



Genes flanking BIME Differential Expression?

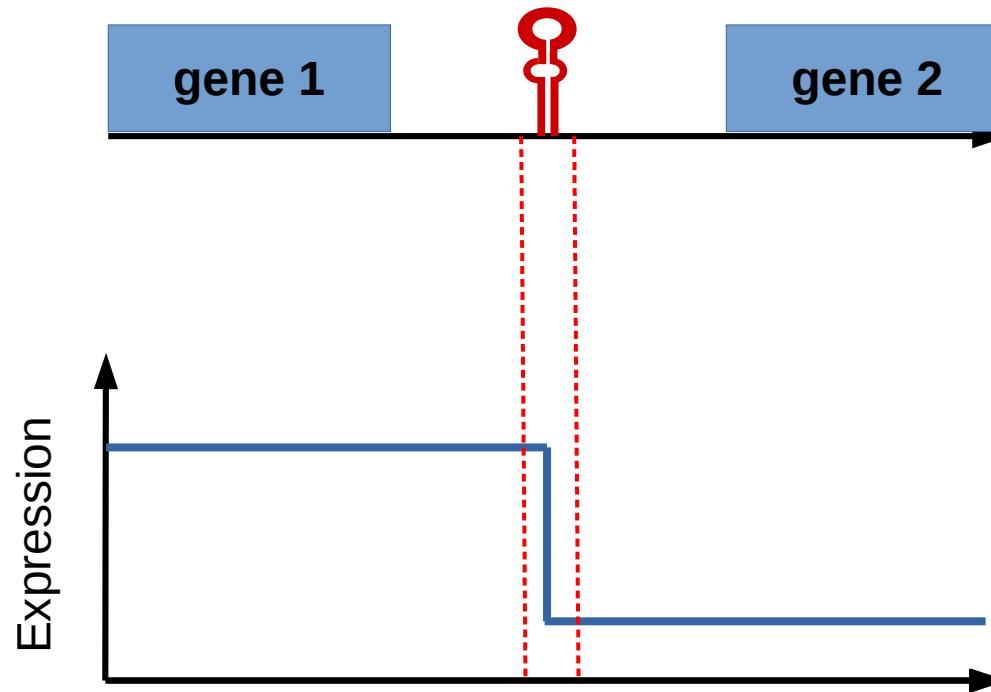
- Wilcoxon rank sum test for DE (p-value threshold : 0.01)



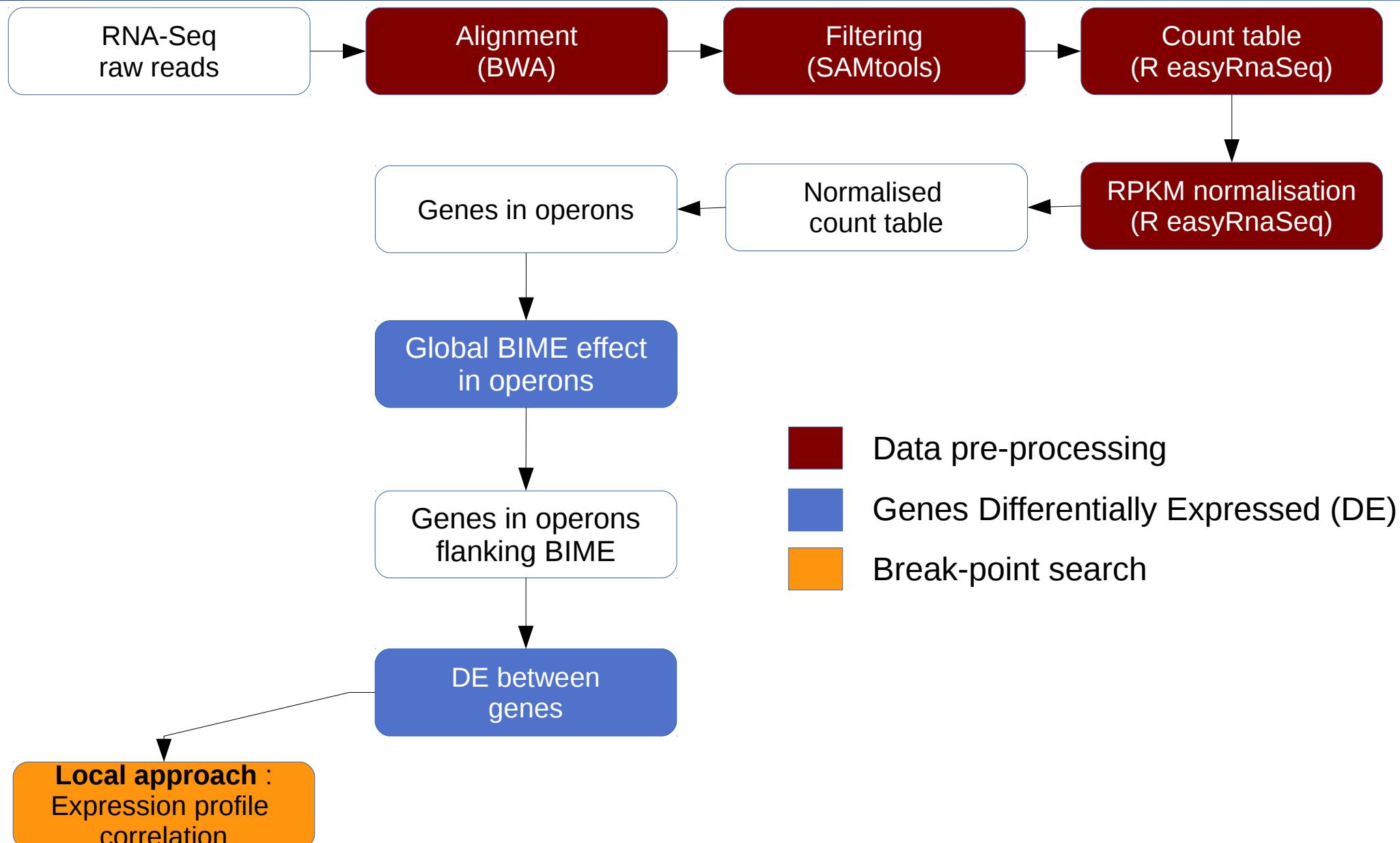
- 25 DE on 39 pairs of gene flanking a BIME (3 data sets)

Genes difference of expression linked to the BIME?

- Coverage break-point on BIME?

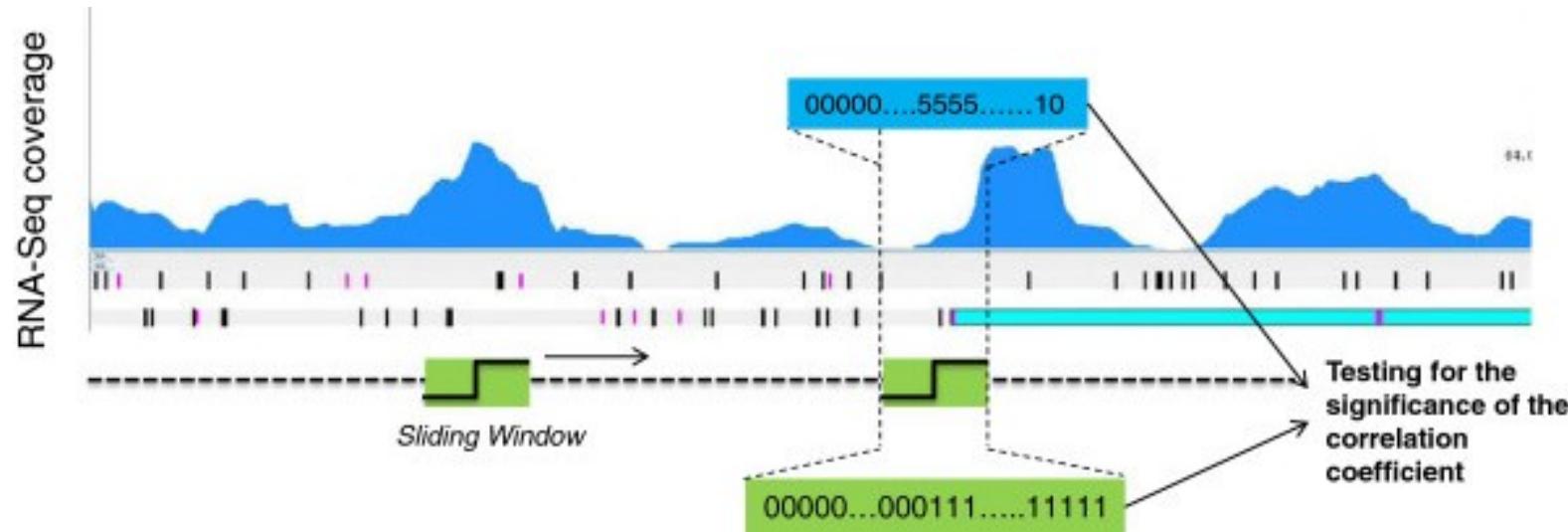


Analysis pipeline



Adaptation of operon detection method (Local approach)

Expression profile correlation

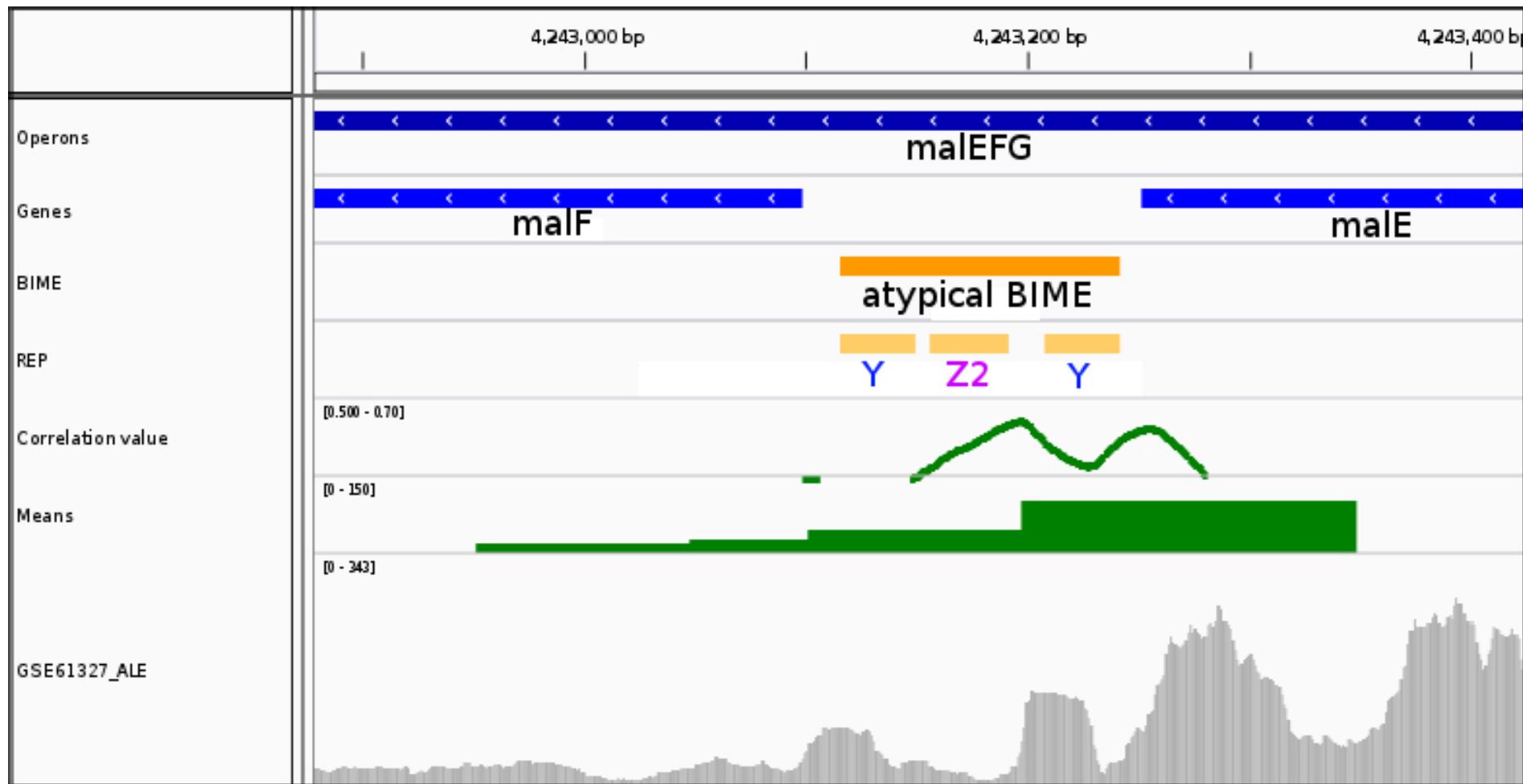


Fortino et al. 2014

- Window size: 300 bp
- 2 fold change of expression
- Correlation > 0.5 & p-value threshold < 10^{-7}

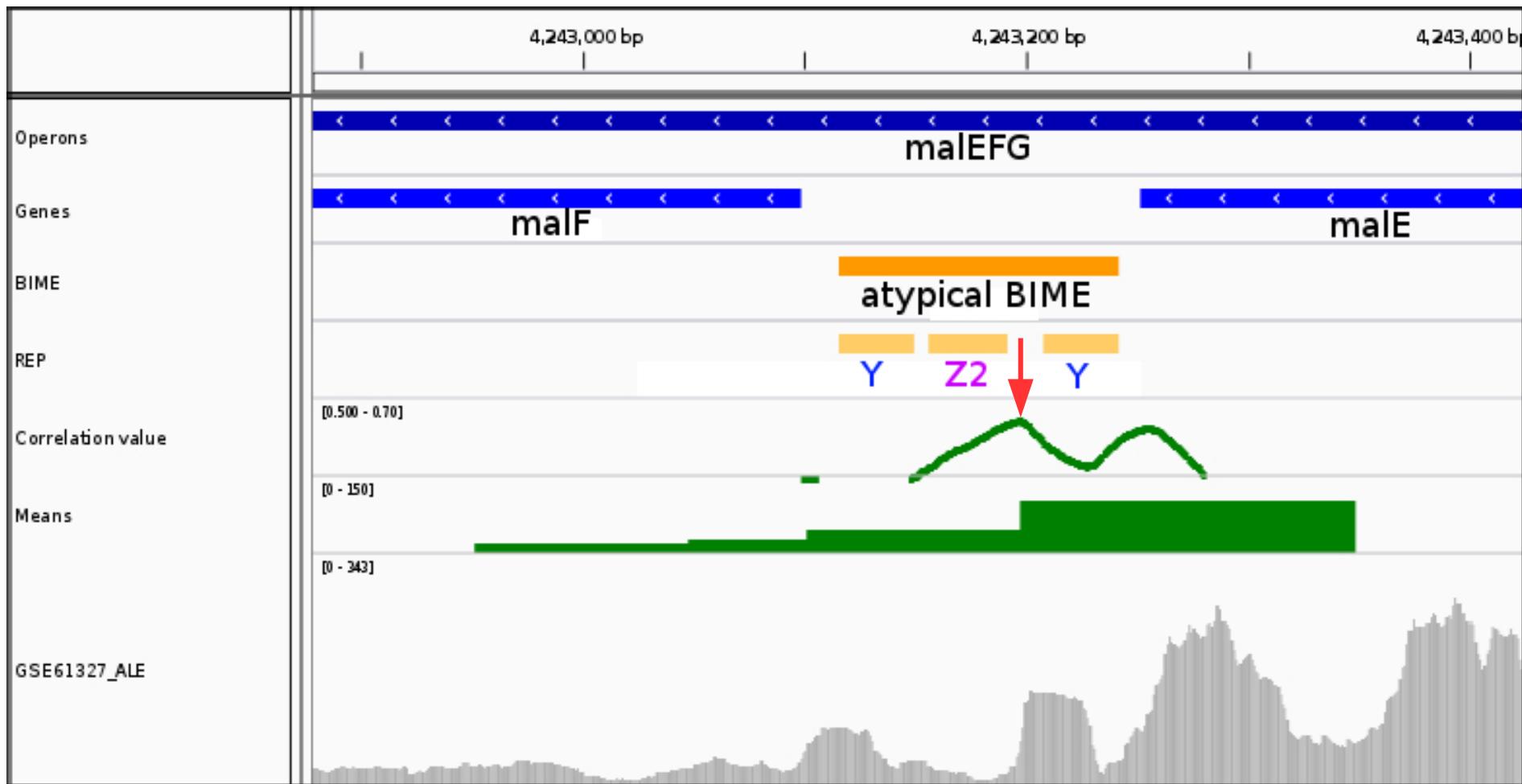
Adaptation of operon detection method (Local approach)

Expression profile correlation on *malE-malF* BIME



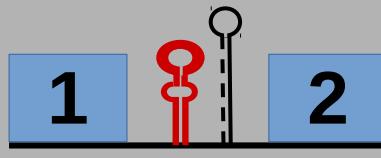
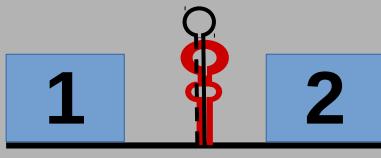
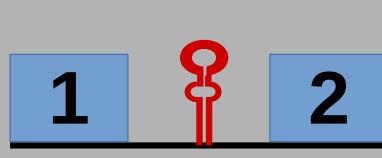
Adaptation of operon detection method (Local approach)

Expression profile correlation on *malE-malF* BIME



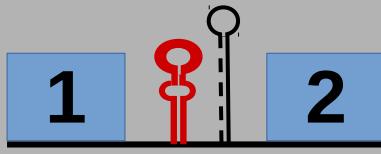
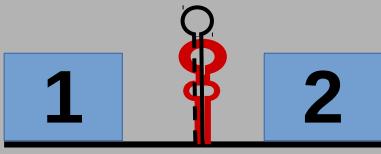
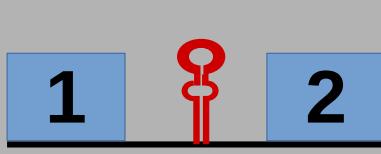
Local approach results

- 18 break-points on BIMEs for 25 DE

			
Coverage break-points on BIME	3/18	4/18	13/18

Local approach results

- 18 break-points on BIMEs for 25 DE

			
Coverage break-points on BIME	3/18	4/18	13/18

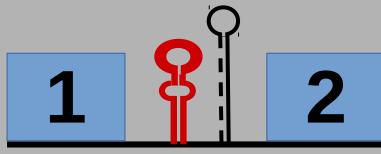
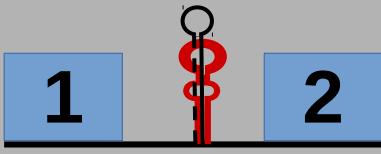
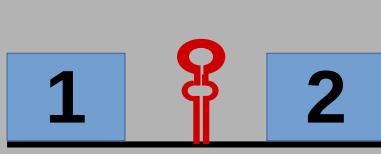
- Weakness:



- RNA-Seq coverage not uniform
- False positive detection

Local approach results

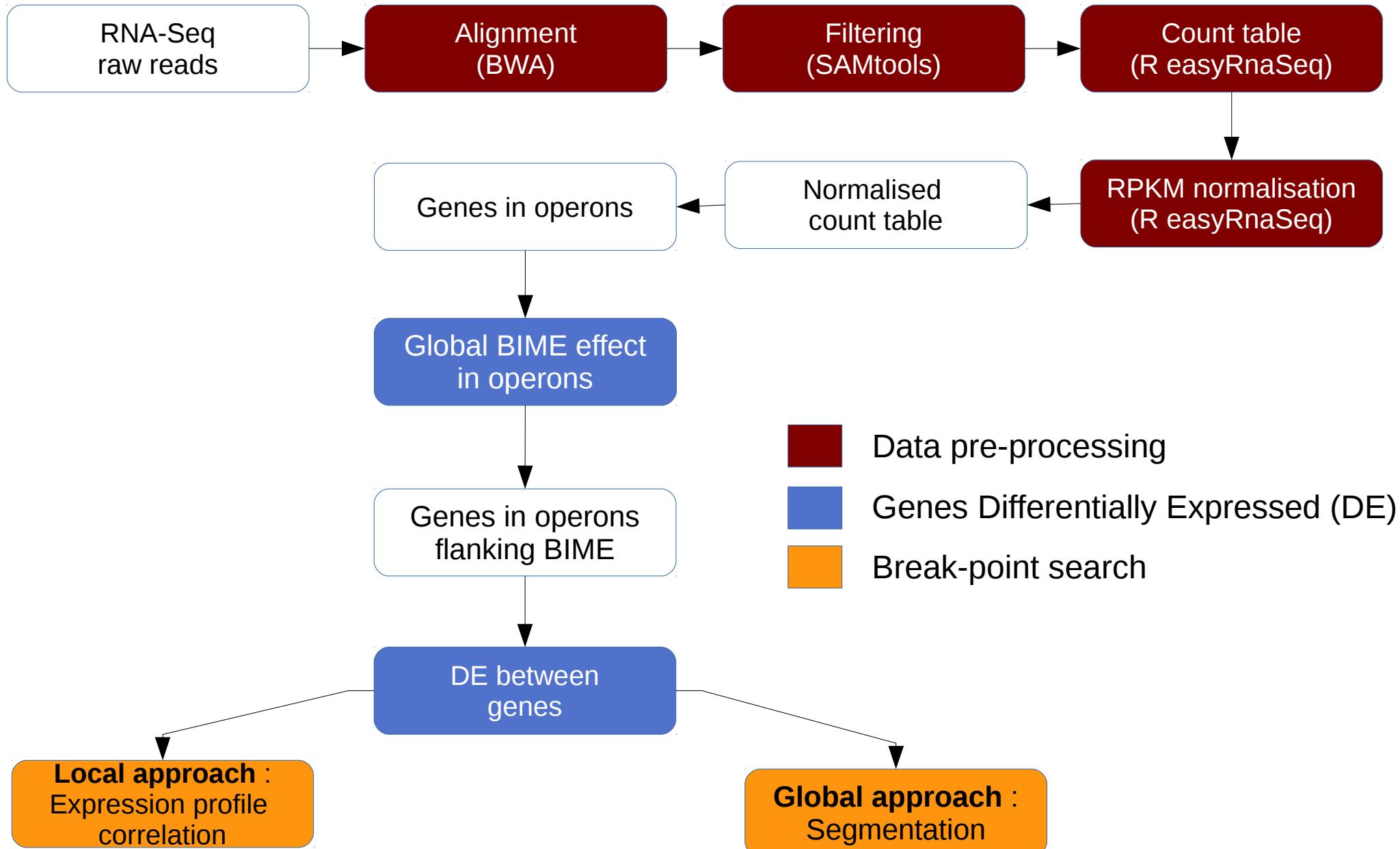
- 18 break-points on BIMEs for 25 DE

			
Coverage break-points on BIME	3/18	4/18	13/18

- Weakness:
 - RNA-Seq coverage not uniform
 - False positive detection
- Other approach?

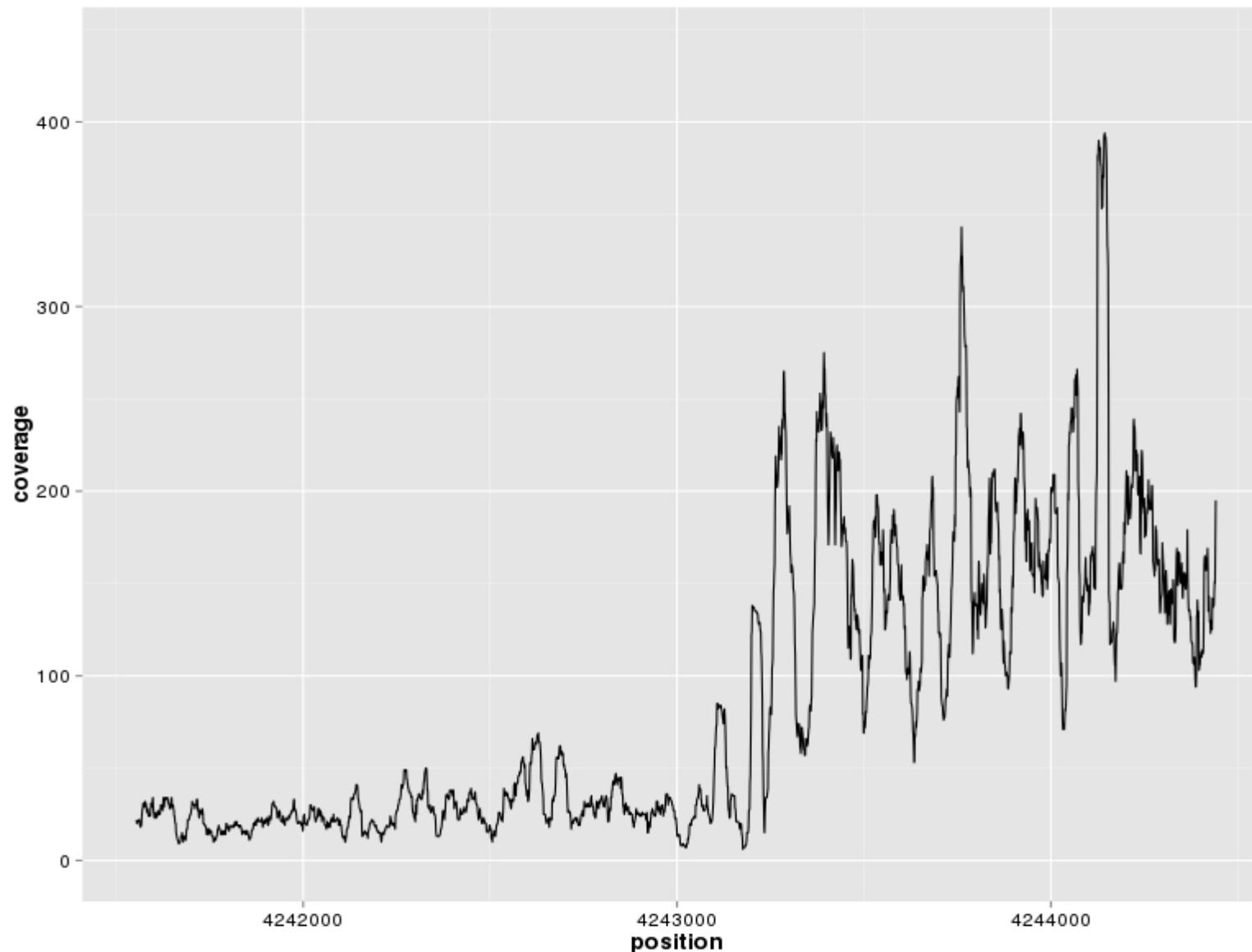


Analysis pipeline



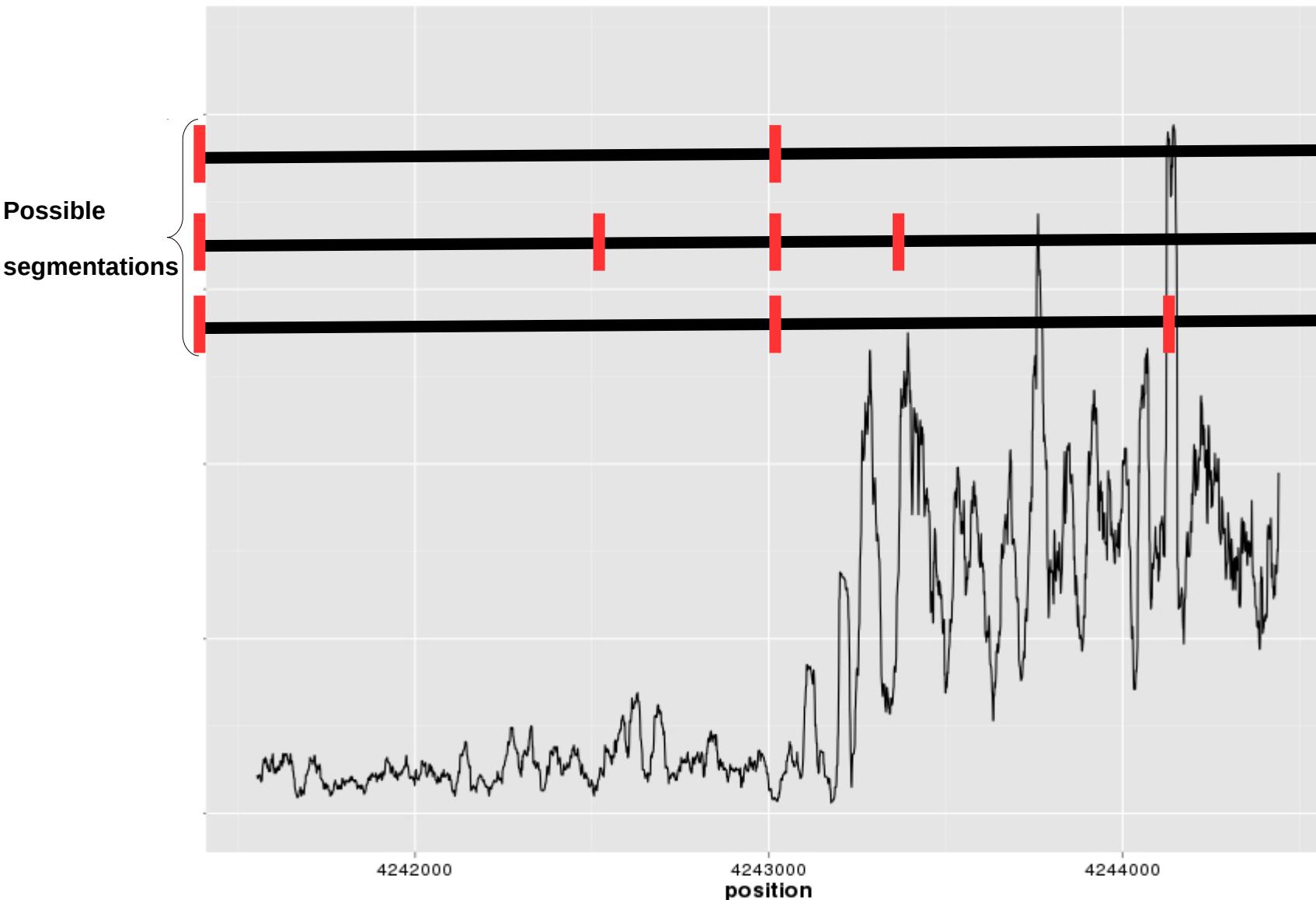
Segmentation (Global approach)

Segmentation on *malE-malF* BIME



Segmentation (Global approach)

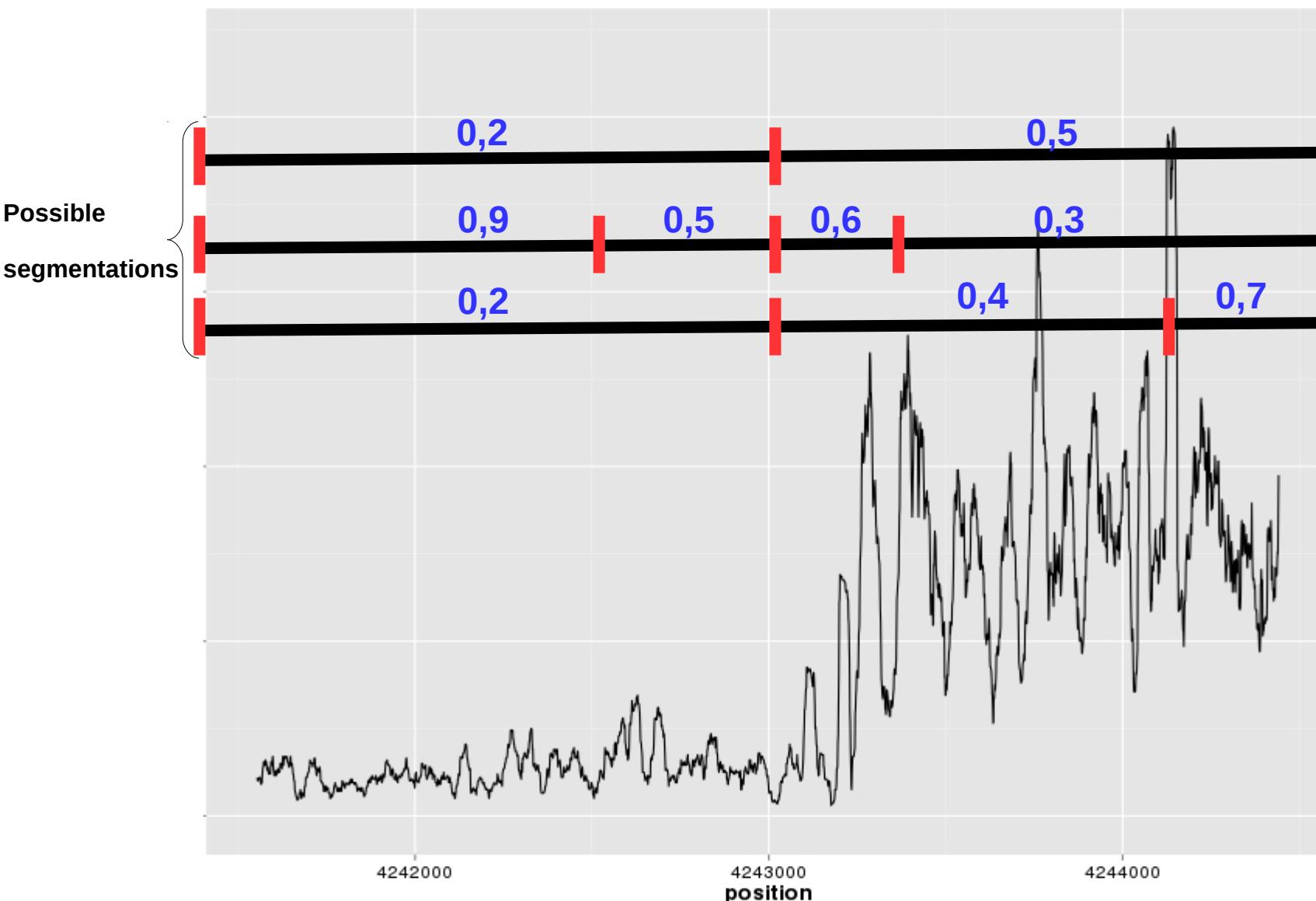
Statistical model : $Y_t \sim G(\theta_r)$



Cleynen et al. 2014

Segmentation (Global approach)

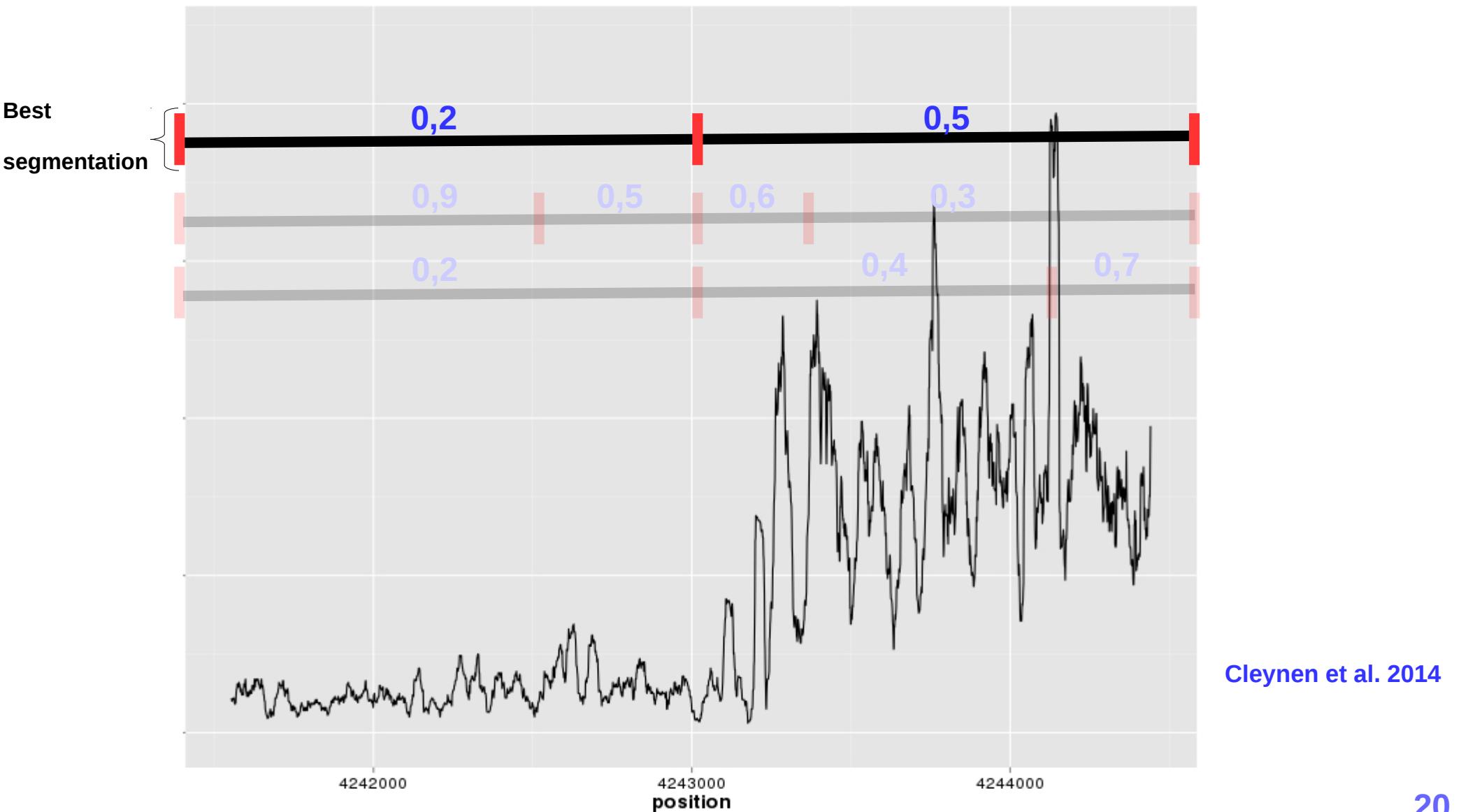
Optimal cost : $c(r) = \min_{\theta} (\sum_{i \in r} y(y_i, \theta))$



Cleynen et al. 2014

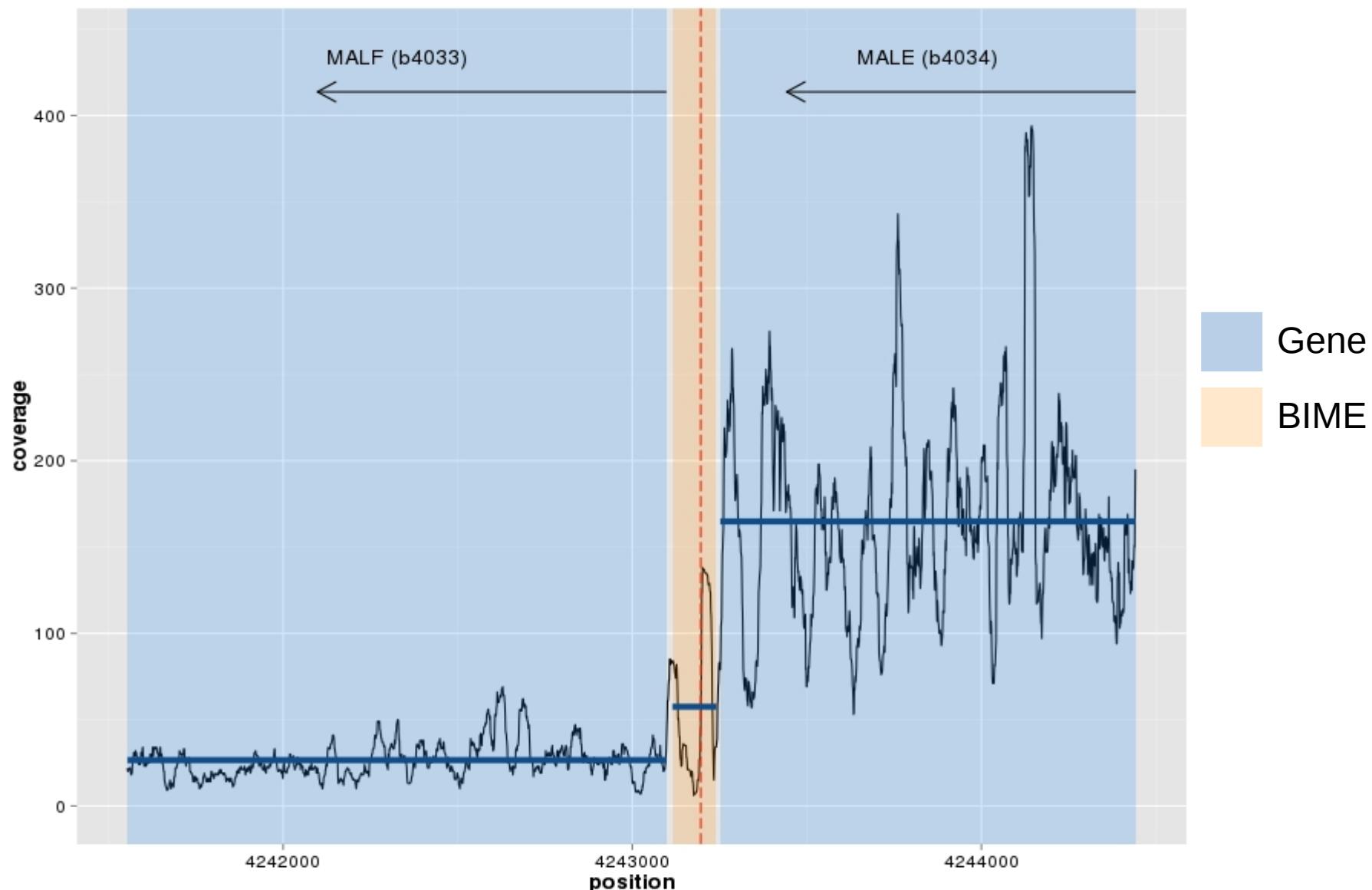
Segmentation (Global approach)

Pruned Dynamic Programming :



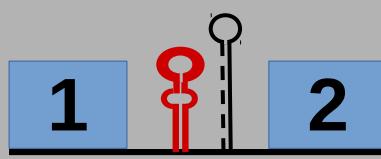
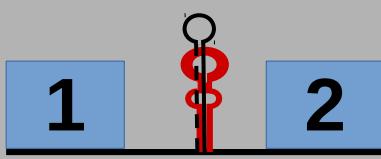
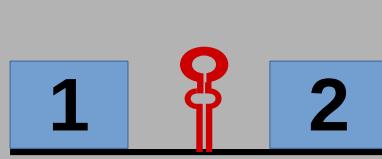
Segmentation (Global approach)

Segmentation on *malE-malF* BIME



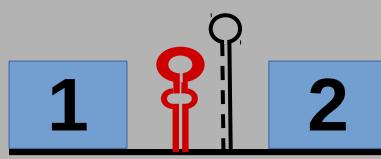
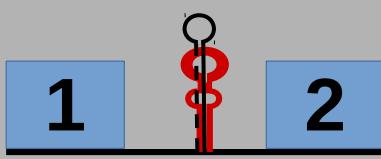
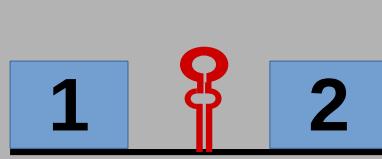
Global approach results

- 17 break-points on BIMEs for 25 DE

			
Coverage break-points on BIME	3/17	3/17	12/17

Global approach results

- 17 break-points on BIMEs for 25 DE

			
Coverage break-points on BIME	3/17	3/17	12/17

- Weakness?
Sensitive to the number of replicates?

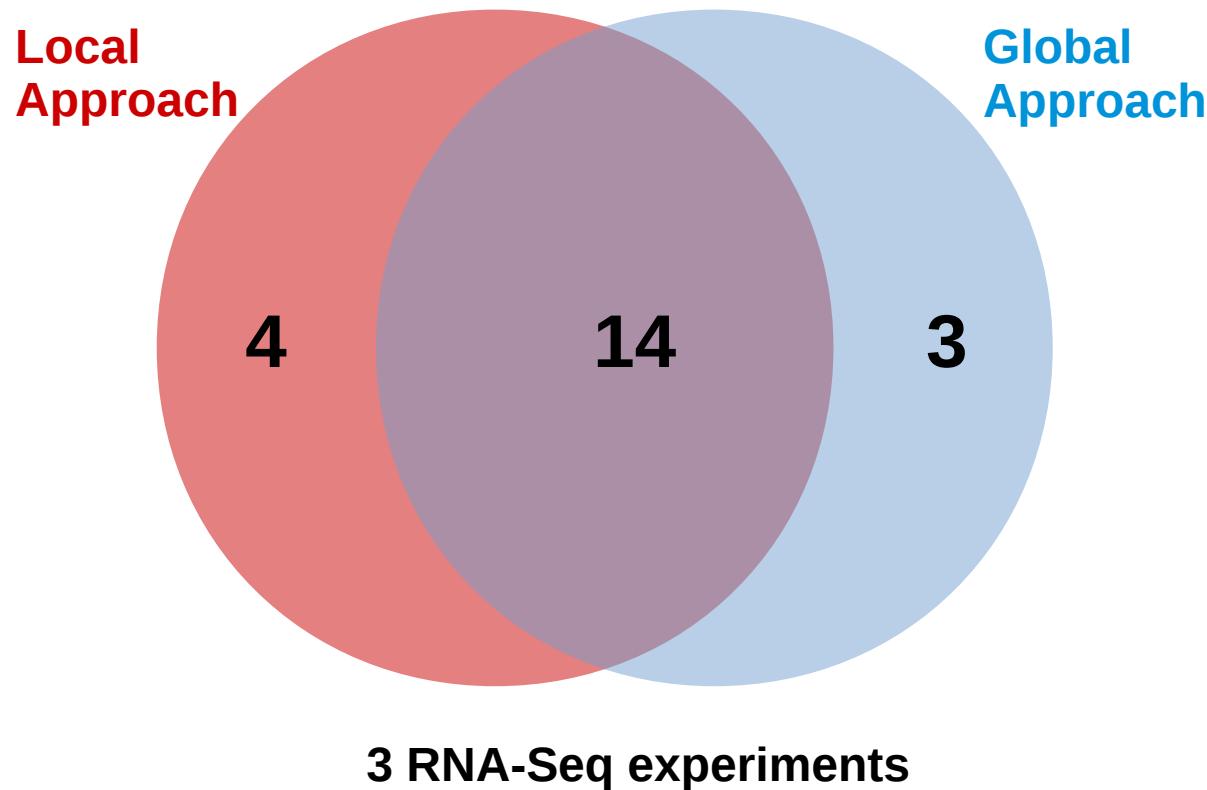
Question

Are BIMEs involved in the mRNA metabolism in operons for *E. coli* genome:

- as transcription terminator?
- as stabiliser of 5' end of transcript?

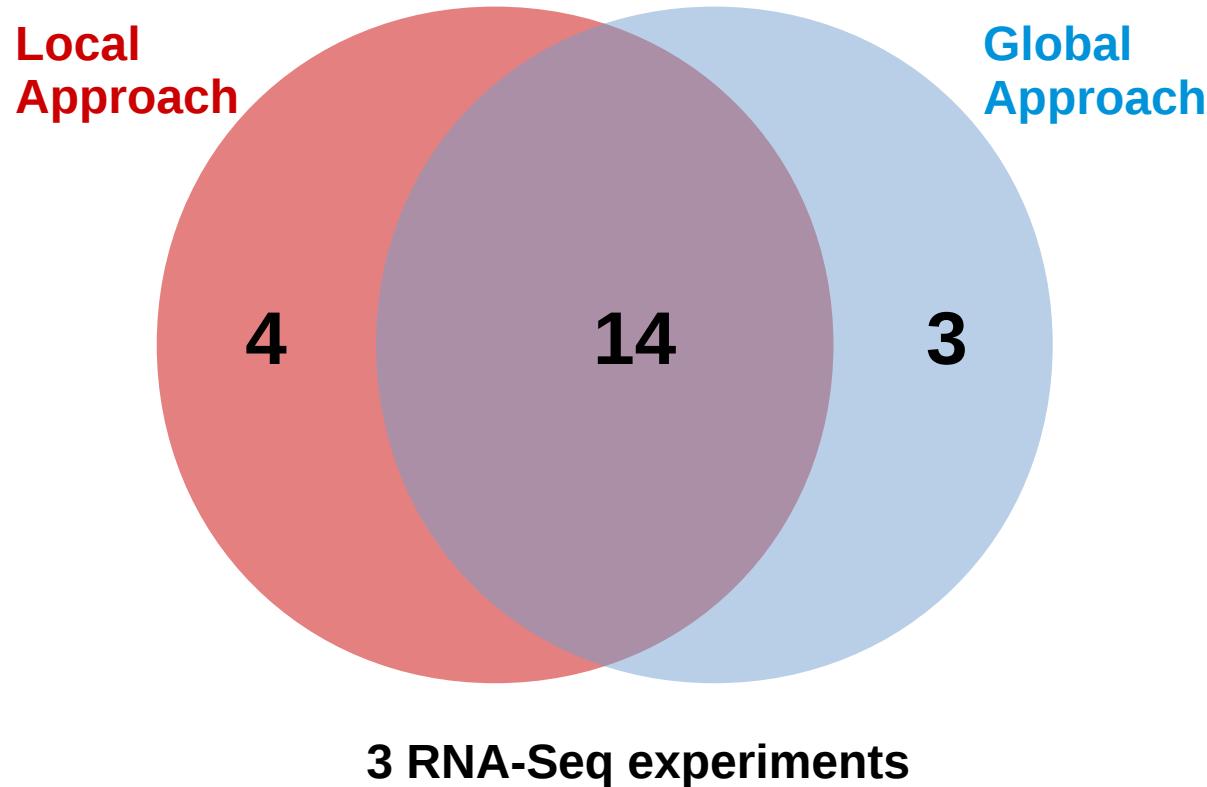
Break-point search result synthesis

- Two methods: 21 break-points



Break-point search result synthesis

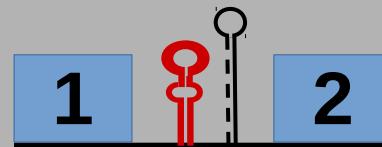
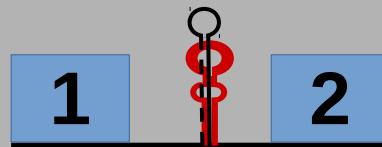
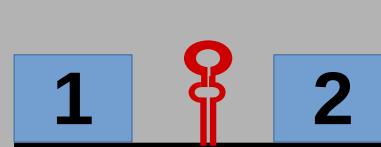
- Two methods: 21 break-points



- 14 common detections

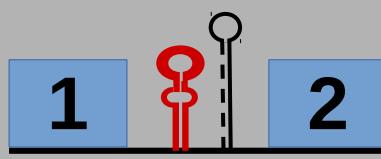
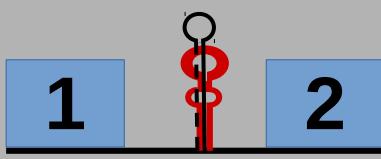
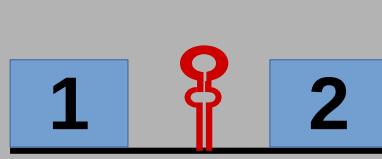
Break-point search result synthesis

- Intergenic region structure with respect to BIME coverage break-point:

			
Coverage break-points on BIME	4/21	4/21	15/21

Break-point search result synthesis

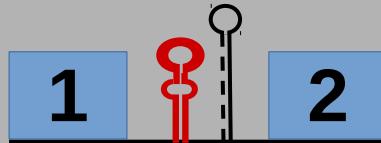
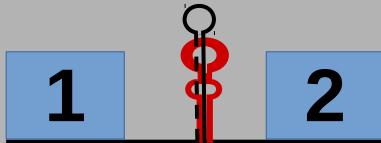
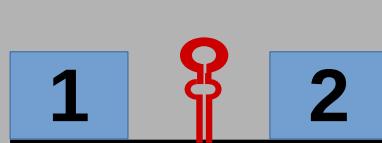
- Intergenic region structure with respect to BIME coverage break-point:

			
Coverage break-points on BIME	4/21	4/21	15/21

- Transcription terminator function

Break-point search result synthesis

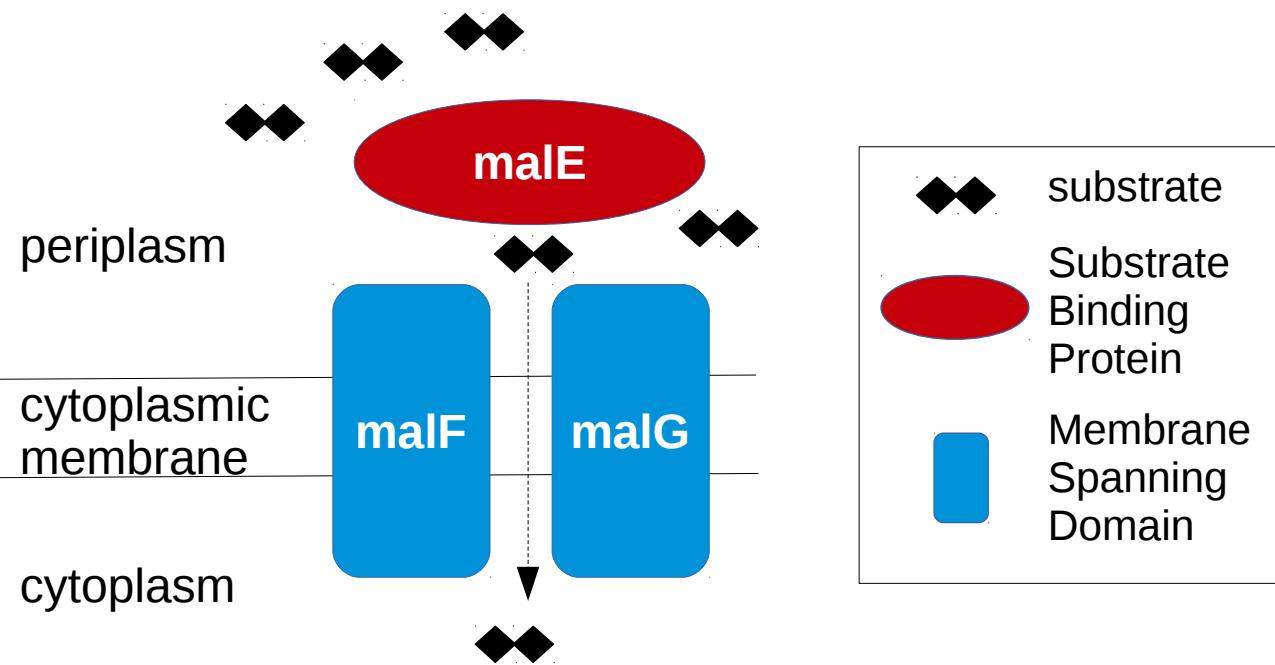
- Intergenic region structure with respect to BIME coverage break-point:

			
Coverage break-points on BIME	4/21	4/21	15/21

- Transcription terminator function
- Stabiliser of 5' end transcript

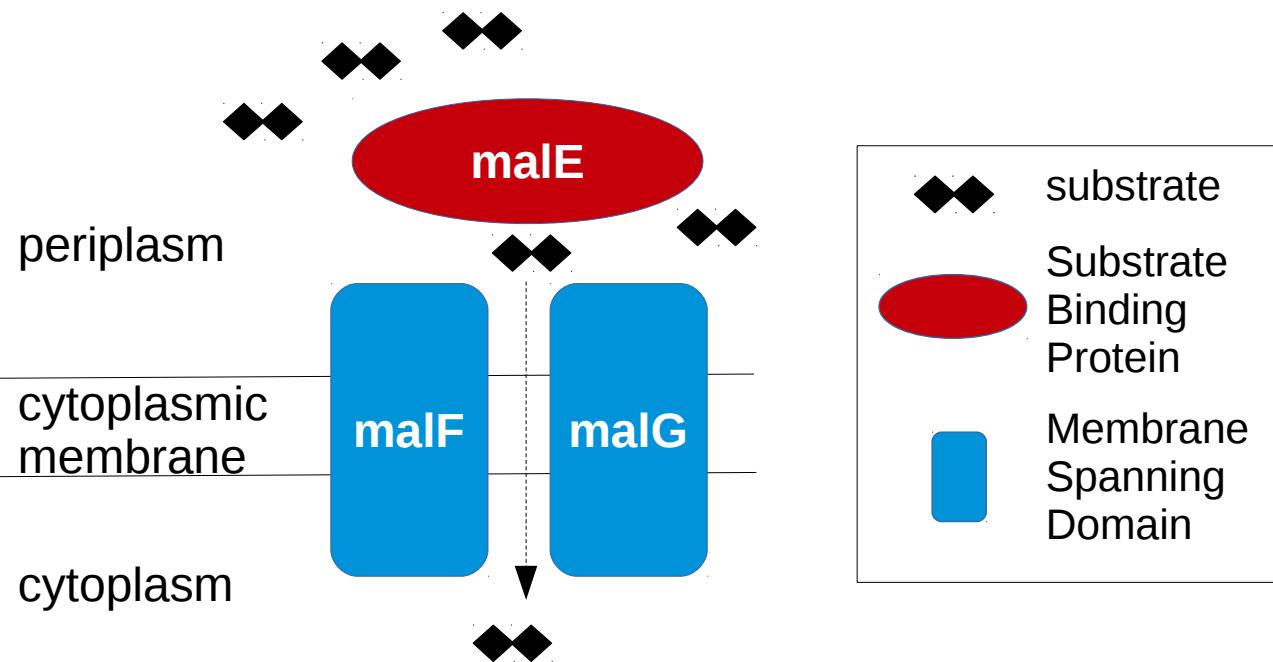
Biological synthesis

- *malEFG* model validated



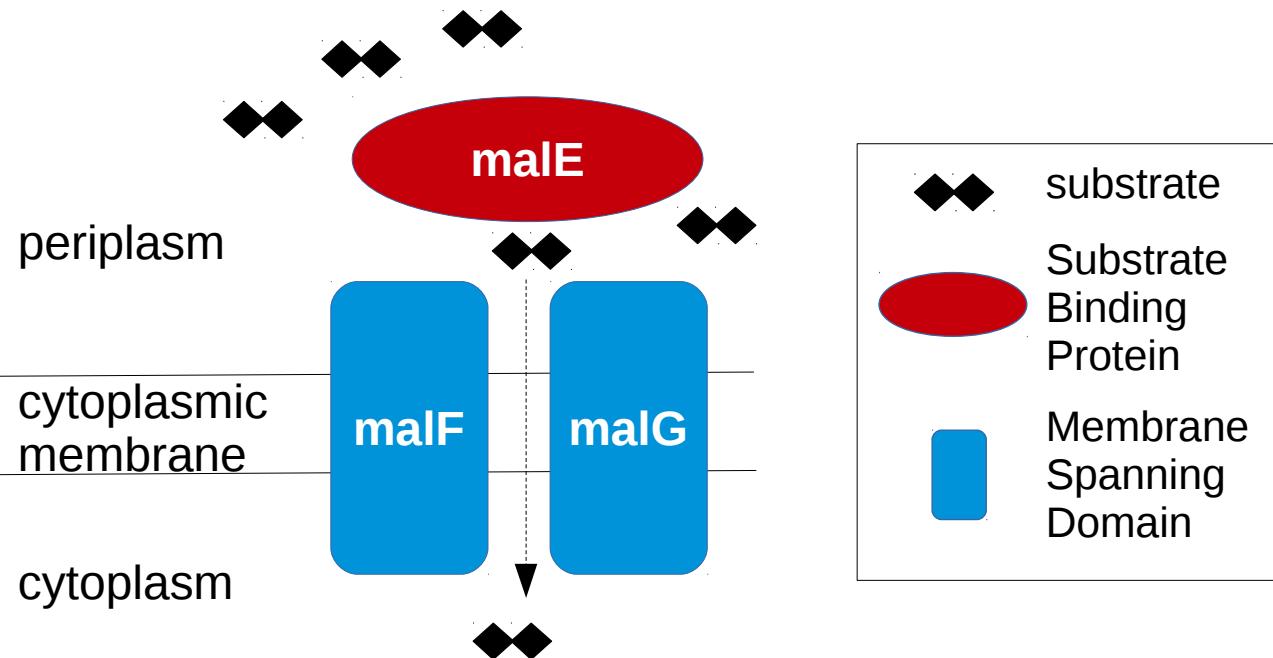
Biological synthesis

- *malEFG* model validated
- 28 ABC transporters in *E.coli* K12



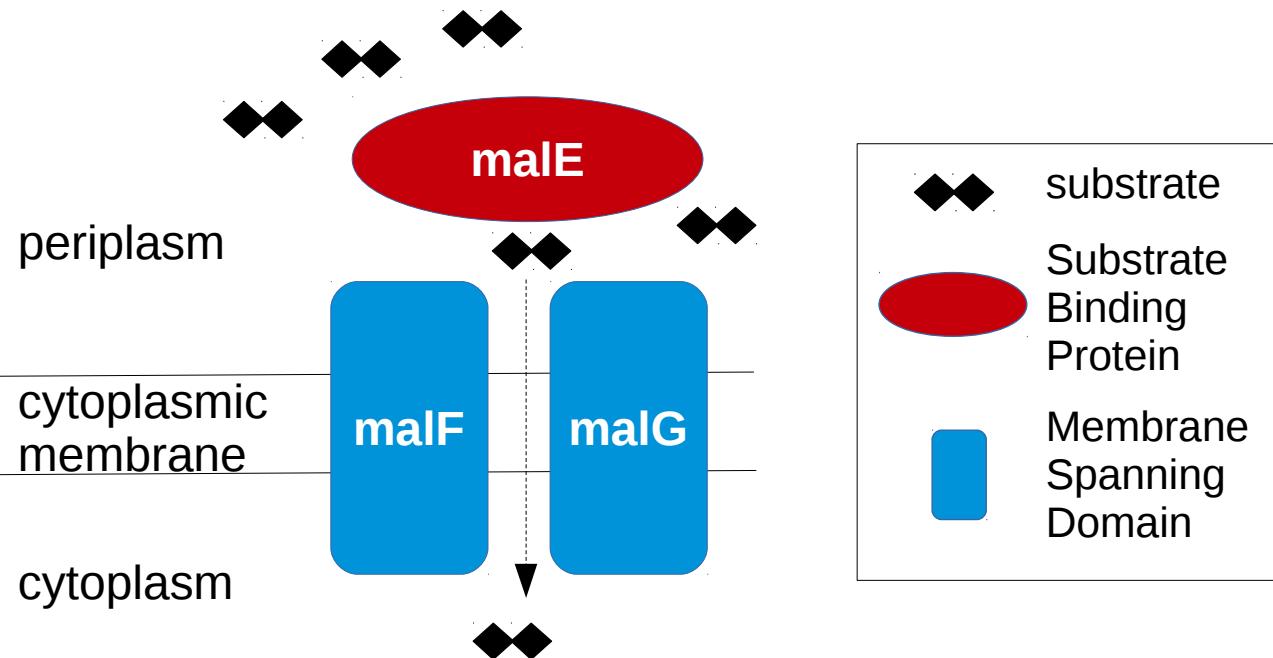
Biological synthesis

- *malEFG* model validated
- 28 ABC transporters in *E.coli* K12
- 7 ABC transporters with 8 BIMEs on 21 detections



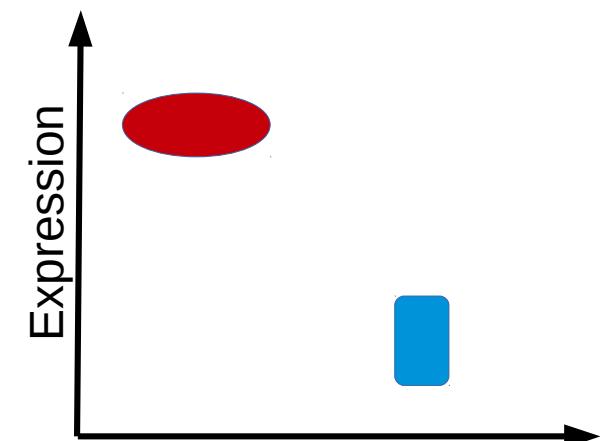
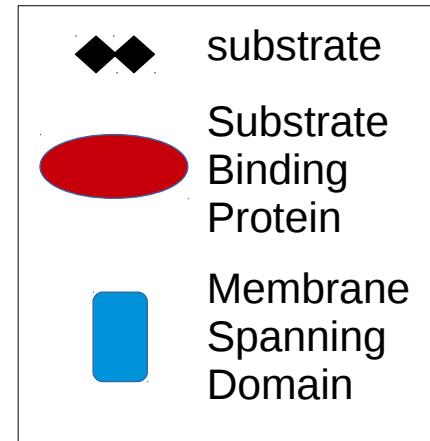
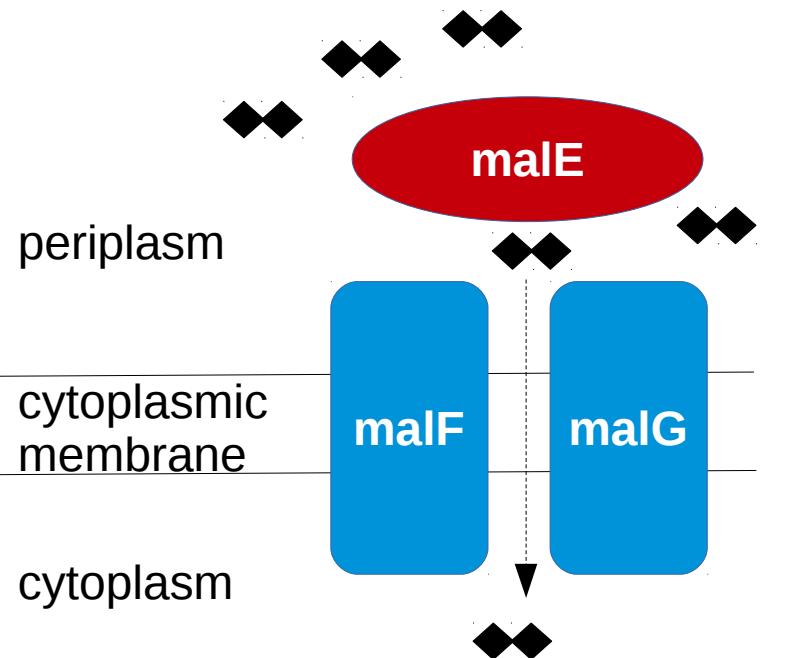
Biological synthesis

- *malEFG* model validated
- 28 ABC transporters in *E.coli* K12
- 7 ABC transporters with 8 BIMEs on 21 detections
- 6 without terminator



Biological synthesis

- *malEFG* model validated
- 28 ABC transporters in *E.coli* K12
- 7 ABC transporters with 8 BIMEs on 21 detections
- 6 without terminator → Substrate Binding Proteins



Conclusions

- DE → Segmentation method suited (Global search for coverage break-points)
- Ready to use with other data sets
- Annotation quality?
- For now RNA-Seq data most appropriate for those studies (coverage not uniform)
- BIME effect on upstream gene
- Predominance of 5' end transcript stabiliser function

Perspectives

- RNA-Seq experiment designs
 - ◆ Test transcription termination function:
WT & alteration of degradosome genes
 - ◆ Test transcript 5' end stabiliser function:
Transcription blocking & time sampling for monitoring of the transcript degradation
- Benchmark with segmentation
 - ◆ reduce the number of replicates (same experiment)
- Generalisation to other species

Acknowledgments

LMGM: Genomics of integrated systems

Gwennaele Fichant

Yves Quentin

Roland Barriot

Petra Langendijk-Genevaux

Mathias Weyder

Arnaud Frèche

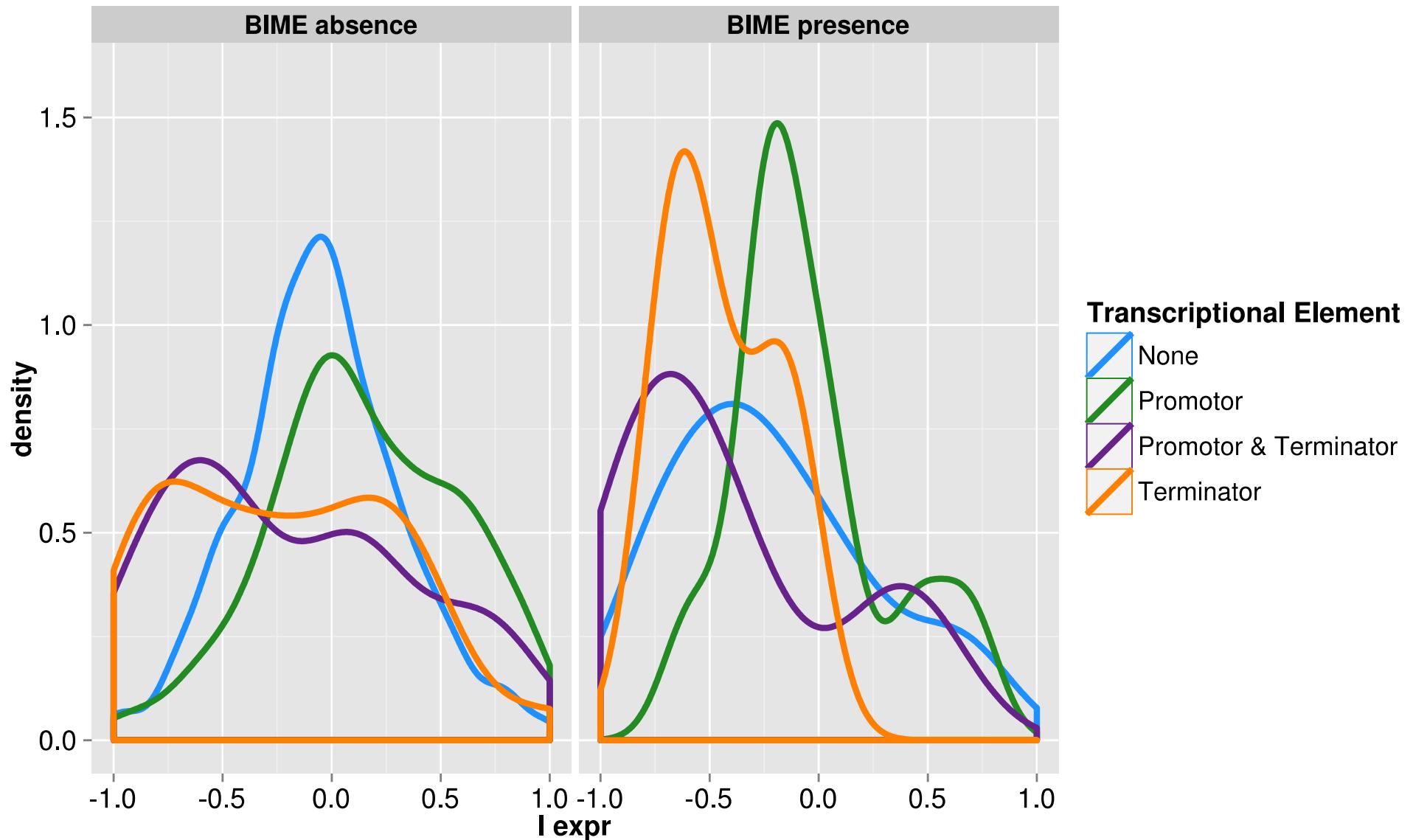
INRA: Applied Mathematics and Informatics

Christine Gaspin



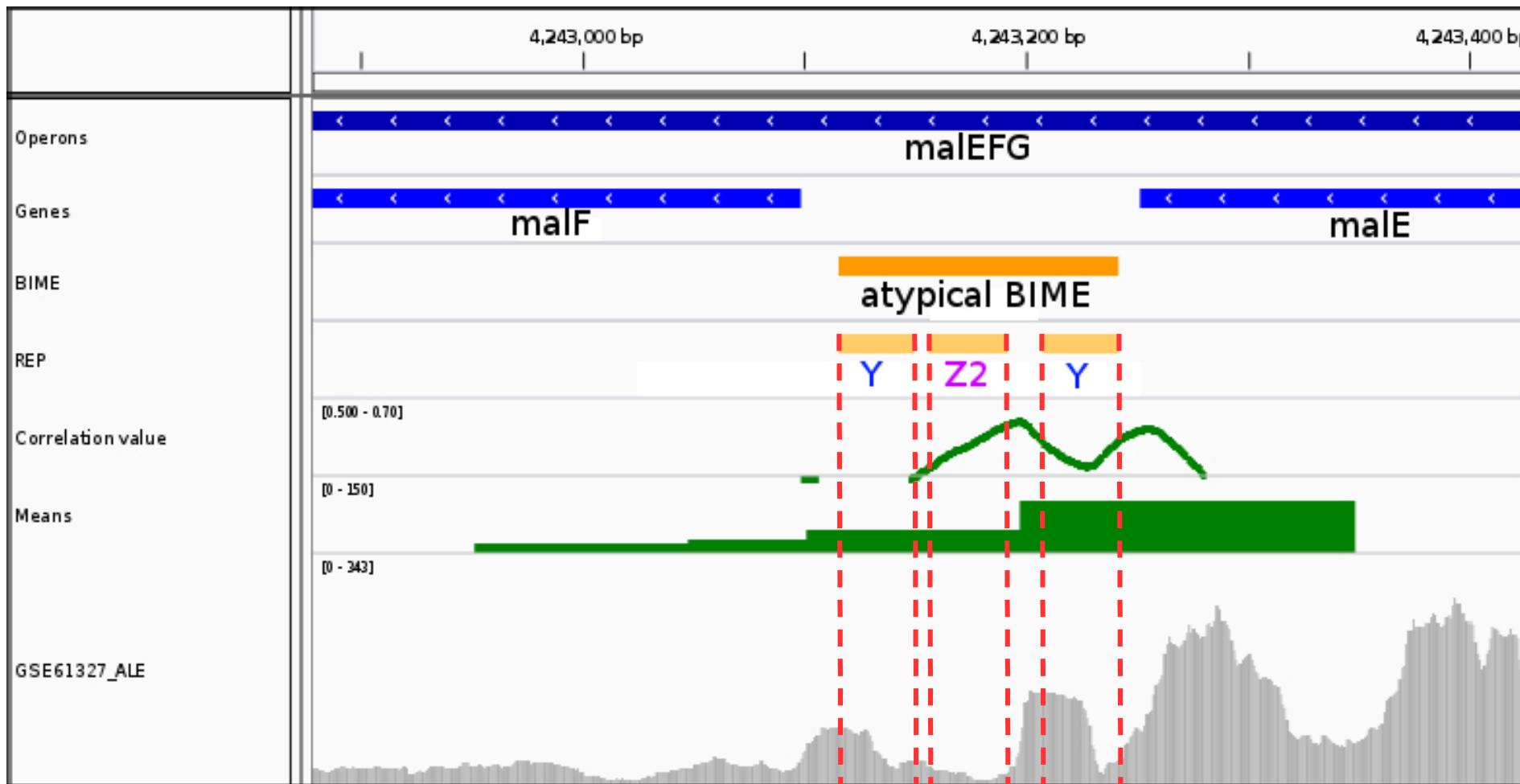
Global BIME effect in operons

Distributions of I_{expr} depending on Transcriptional Elements



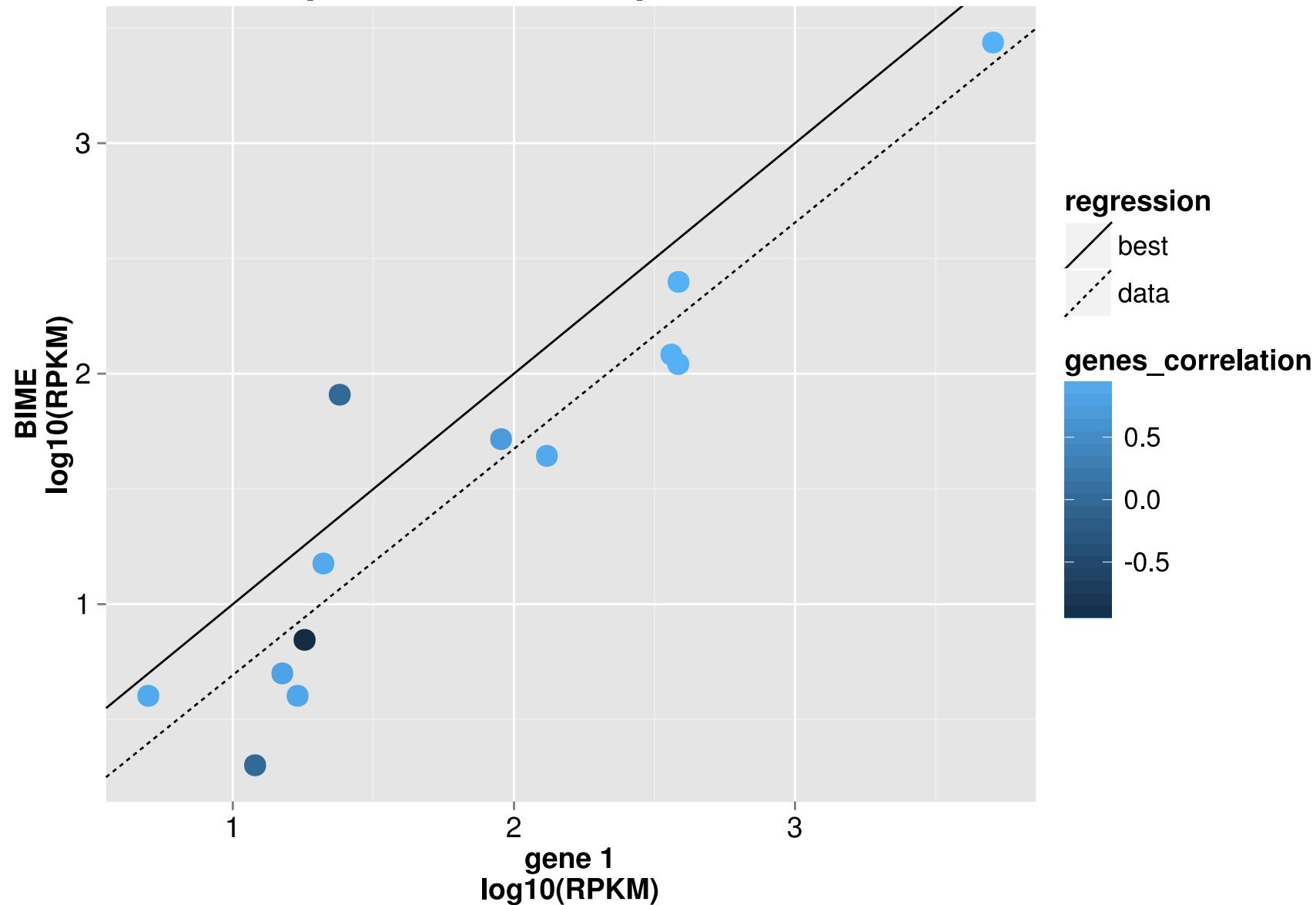
Adaptation of operon detection method (Local approach)

Expression profile correlation on *malE-malF* BIME



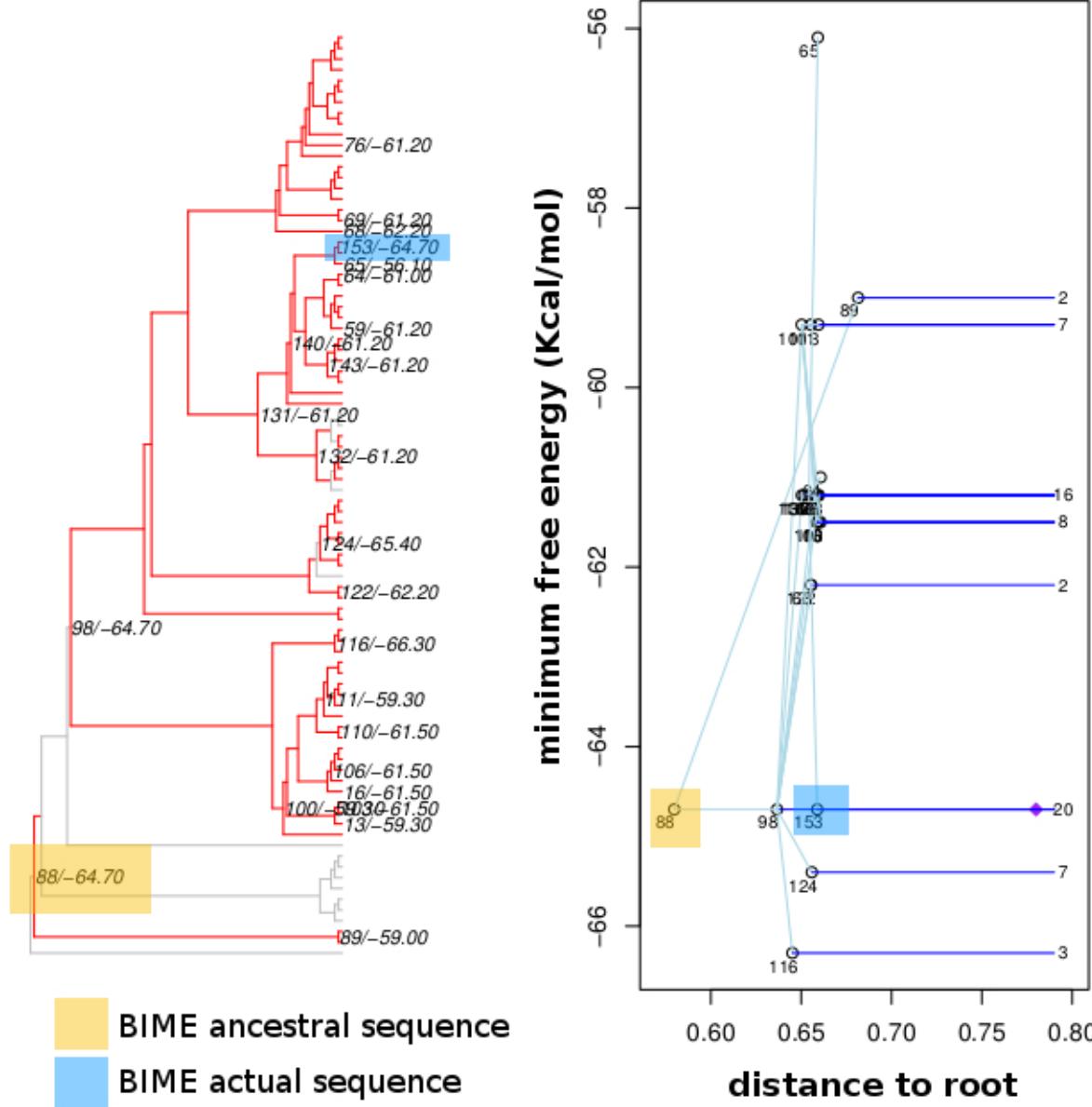
BIME sequencing bias

No DE in operons, no promotor or terminator



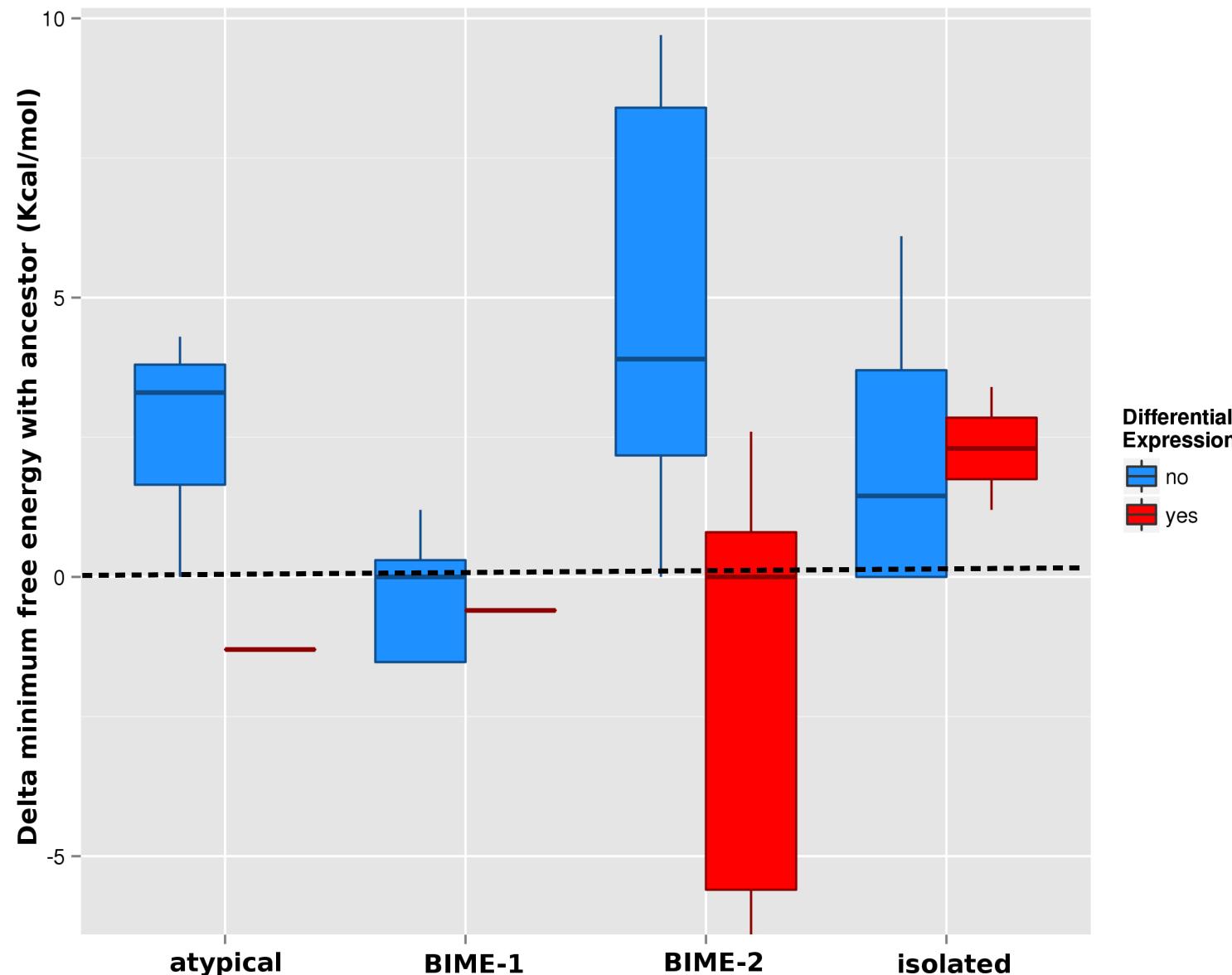
Ancestral sequences and secondary structures

malE-malF BIME 2nd structure energy evolution



Ancestral sequences and secondary structures

Delta of energy between ancestor & actual BIME



Segmentation efficiency & replicates

