



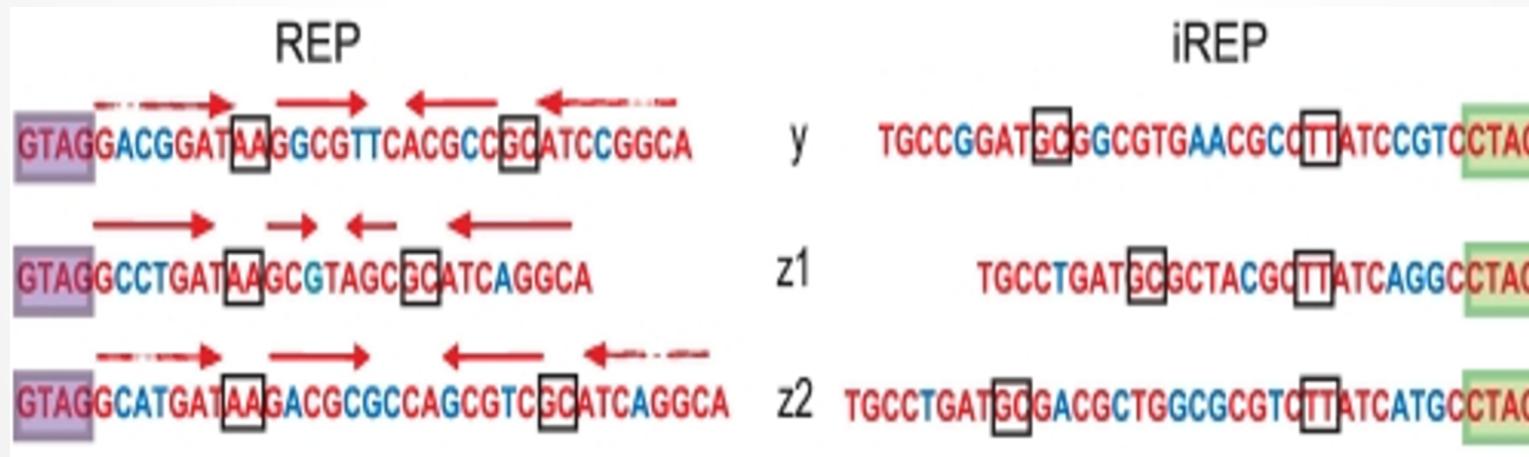
REP involved in  
transcription regulation?

# REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure :  
 $Y$ ,  $Z^1$  and  $Z^2$ .

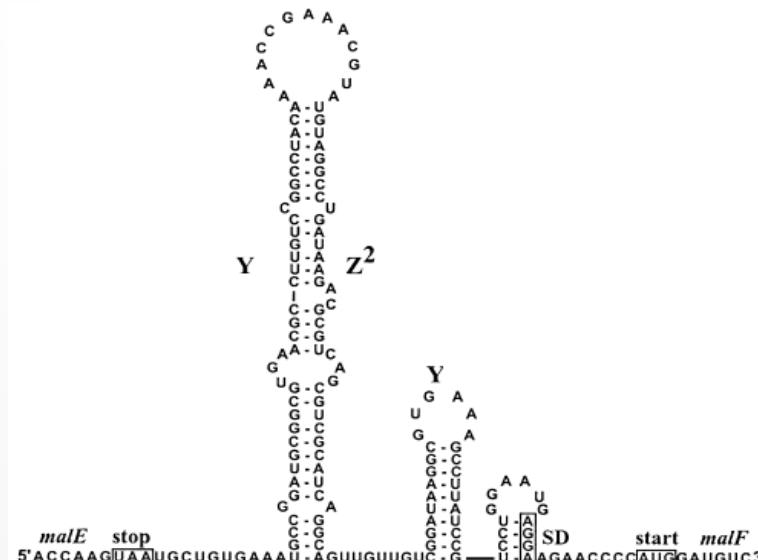
# REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z<sup>1</sup> and Z<sup>2</sup>.



# REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z<sup>1</sup> and Z<sup>2</sup>.
- Stem-loop secondary structure with a bulge.

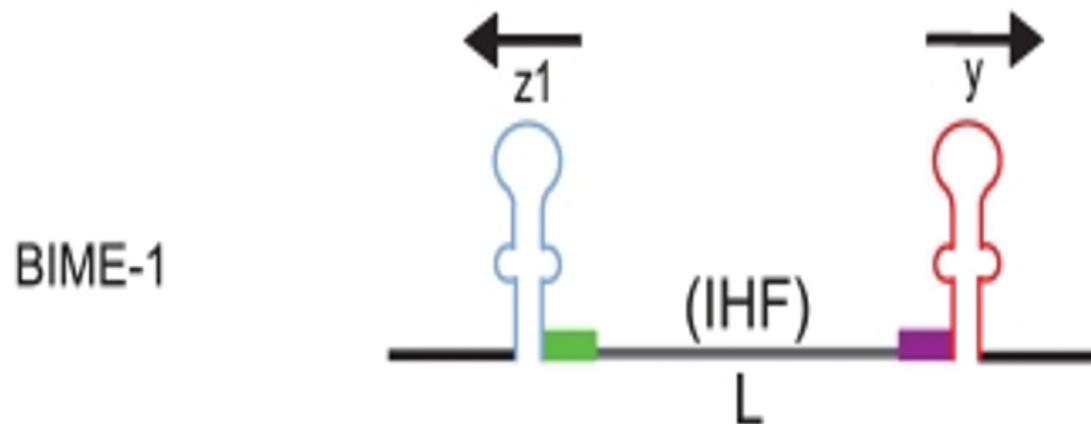


# REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z<sup>1</sup> and Z<sup>2</sup>.
- Stem-loop secondary structure with a bulge.
- Numerous bacterial genomes, ≈ 1% for E. coli.
- BIME organization.

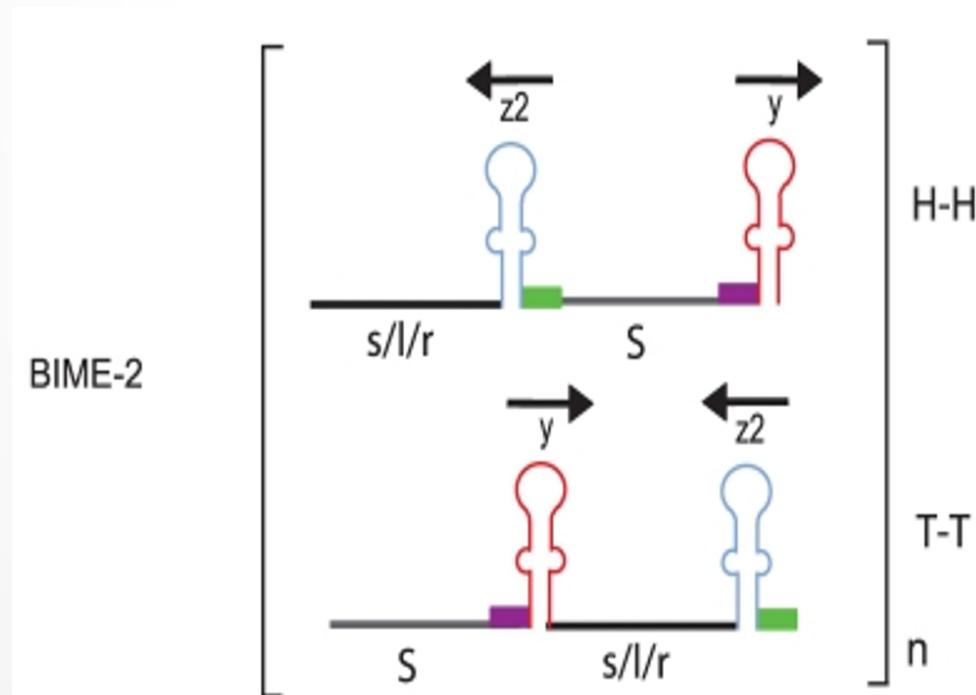
# BIME features

- Bacterial Interspesed Mosaic Element.
- 3 classes depending on REP composition :  
Bime-1, Bime-2, atypical.



# BIME features

- Bacterial Interspesed Mosaic Element.
- 3 classes depending on REP composition :  
Bime-1, Bime-2, atypical.



# Objectives

Does REP are involved in the transcription regulation in E. coli genome :

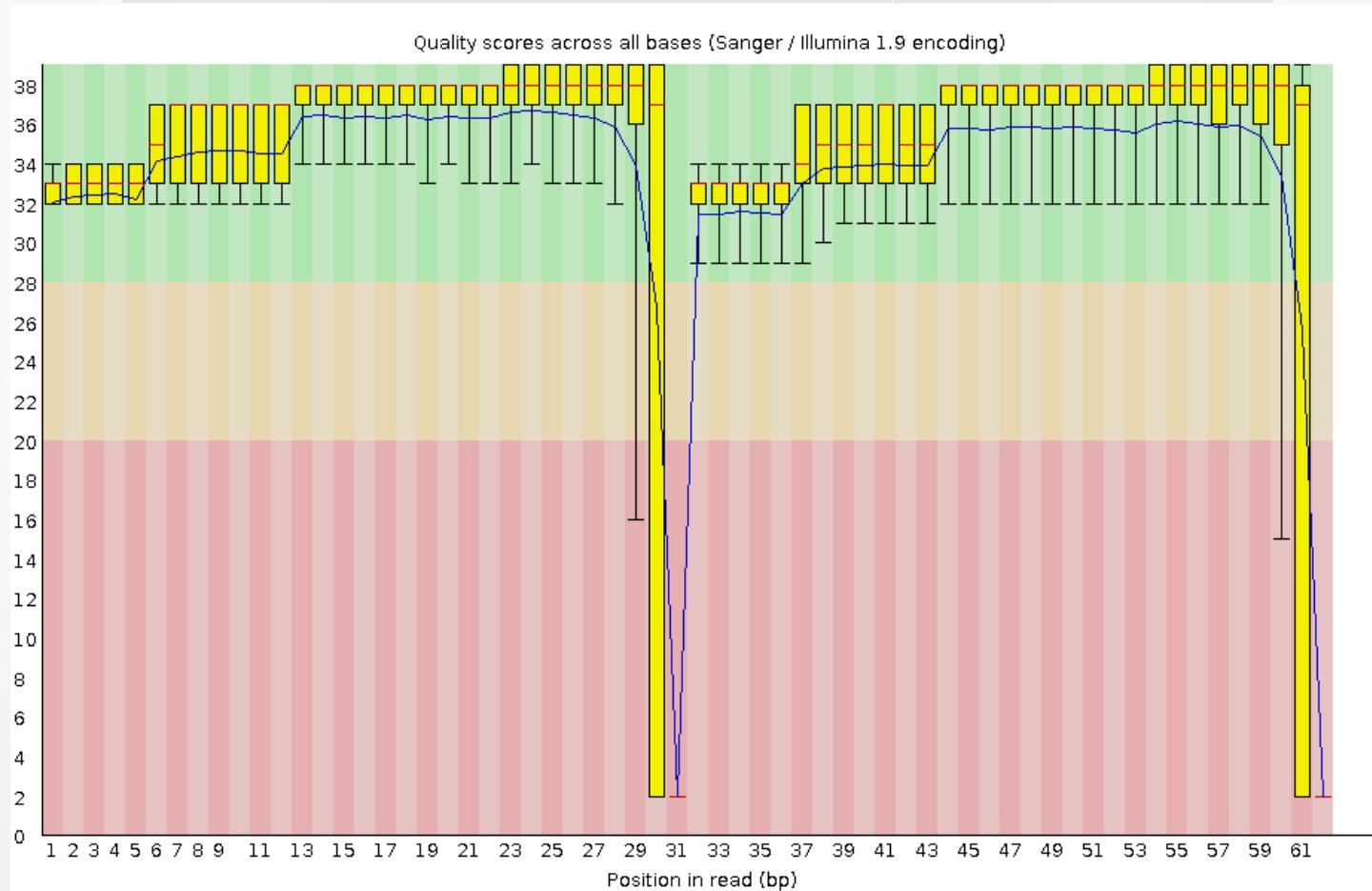
- as protection from RNA degradosome ?
- as transcription terminator ?
- as recognition site for proteic factors ?

# Data choice

- Escherichia coli K-12 MG1655 on Glucose Minimal Medium ([LaCroix et al. 2014](#)).
- Why ?
  - RNA-Seq experiment.
  - Large number of replicates :
    - 9 for Glucose minimal medium (ALE)
    - 2 for the wild type (WT)
  - Paired-end and specific strand library preparation.
- Quality control with `fastqc`.

# Data choice

| Measure                           | Value     |
|-----------------------------------|-----------|
| Total sequences                   | 7 772 244 |
| Sequences flagged as poor quality | 0         |
| Sequence length                   | 62        |
| % GC                              | 53        |



# Data choice

- Escherichia coli K-12 MG1655 on Glucose Minimal Medium ([LaCroix et al. 2014](#)).
- Why ?
  - RNA-Seq experiment.
  - Large number of replicates :
    - 9 for Glucose minimal medium (ALE)
    - 2 for the wild type (WT)
  - Paired-end and specific strand library preparation.
- Quality control with `fastqc`.
- 1 ALE sample rejected.

# Alignments

- Alignment on NC\_000913.2 E.coli genome ([BWA](#)) :
  - Index of the reference genome
  - Alignment (SAM file)

# Alignments

- Alignment on NC\_000913.2 E.coli genome (**BWA**) :
  - Index of the reference genome
  - Alignment (SAM file)
- Post-alignment process (**Samtools**) :
  - Convert to binaries (BAM file)
  - Sort the reads by genomic position
  - Remove PCR duplicates
  - Merge files
  - Index alignment merged file

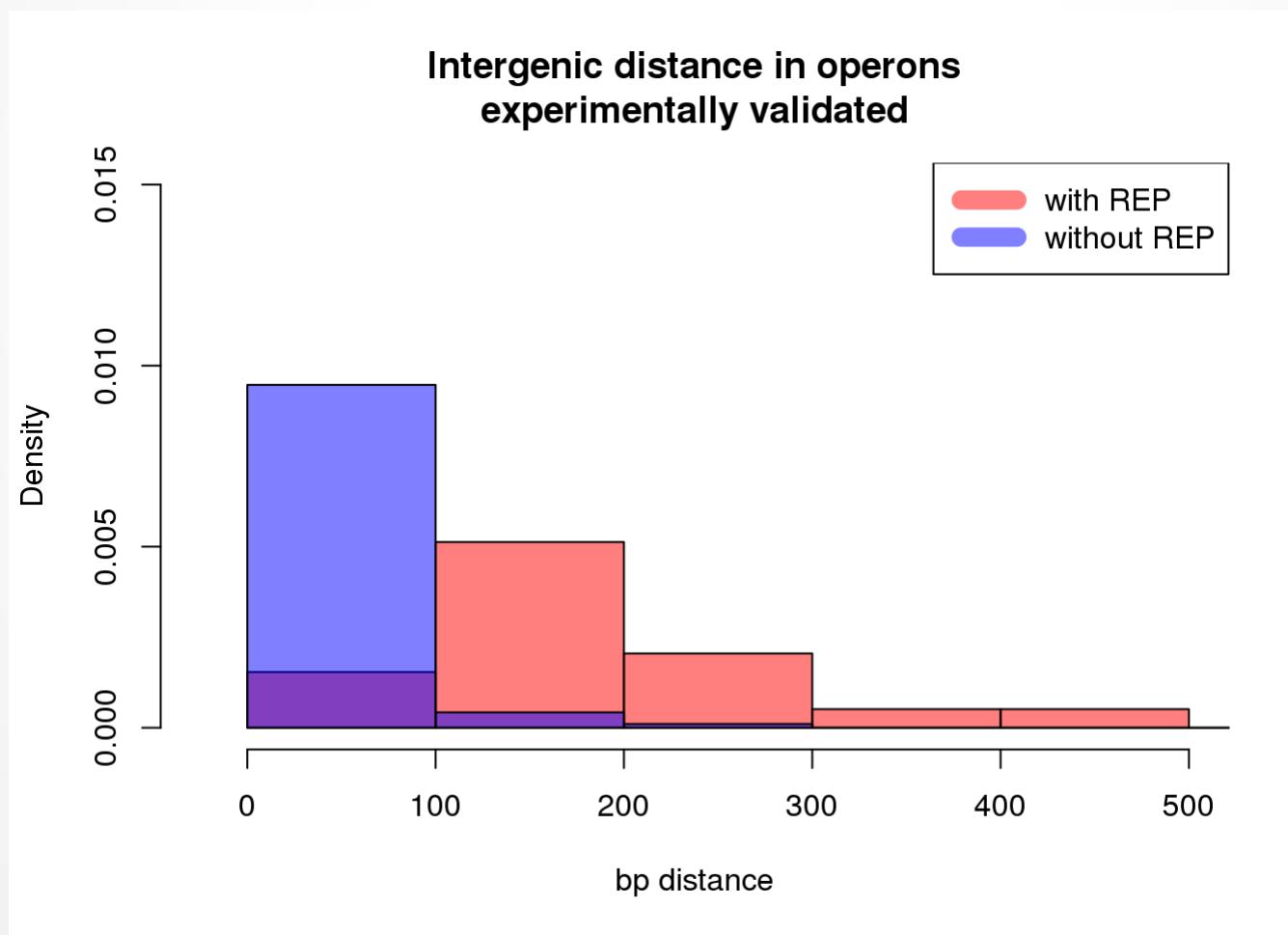
# Genomic informations

From GFF, laboratory database, RegulonDB and DOOR  
design of reference files ([Bedtools](#), [Python scripts](#))

- Operons
- REP in operons
- Genes surrounding REP in operon
- Terminators...

# Genomic informations

For operons from RegulonDB :



# Differential expression in operon with REP

**Detect Differential Expression (DE)  
between genes surrounding a REP in an  
operon.**

# Differential expression in operon with REP

**Detect Differential Expression (DE)  
between genes surrounding a REP in an  
operon.**



Keep in mind :

- Not looking for genes DE in different conditions.
- Focus on minimal Glucose medium condition (ALE).
- Not focus on mRNA, transform the GFF to get the expression of tRNA, rRNA and ncRNA.

# Differential expression in operon with REP

- Reads count table for expressed features (`R package EasyRnaSeq`)

# Differential expression in operon with REP

- Reads count table for expressed features ([R package EasyRnaSeq](#))
- Correct reads count with RPKM (Reads Per Kilobase Per Millions mapped reads) :

$$RPKM = \frac{nb\,mapped\,reads \times \frac{1000\,bases \times 10^6}{length\,of\,transcript \times total\,nb\,of\,mapped\,reads}}{}}$$

# Differential expression in operon with REP

- Reads count table for expressed features ([R package EasyRnaSeq](#))
- Correct reads count with RPKM (Reads Per Kilobase Per Millions mapped reads) :

$$RPKM = \frac{nb\,mapped\,reads \times \frac{1000\,bases \times 10^6}{length\,of\,transcript \times total\,nb\,of\,mapped\,reads}}{}$$

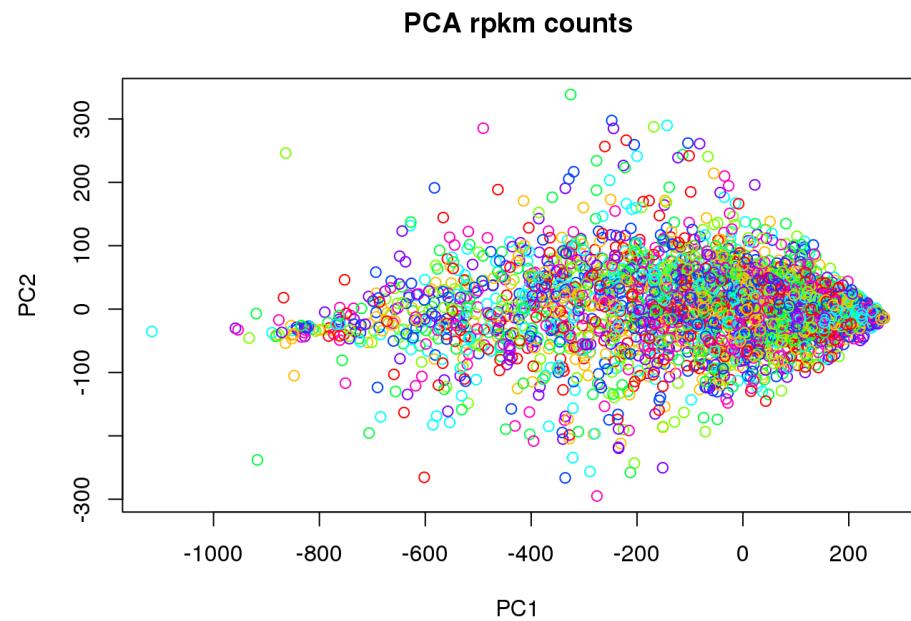
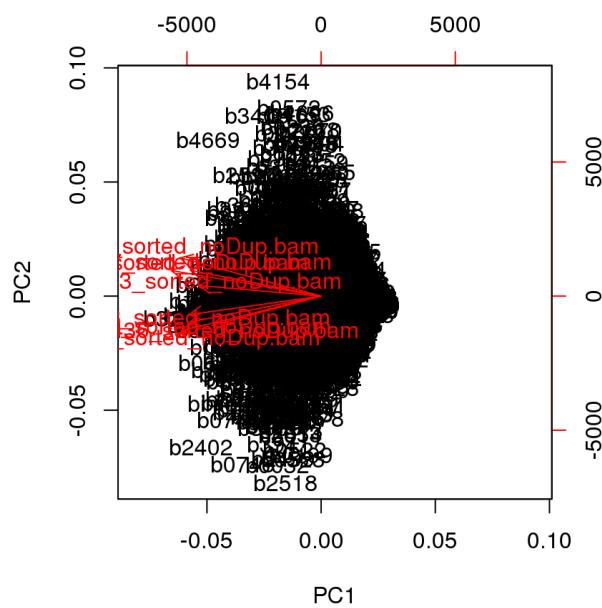
Molar concentration of a transcript :

- RNA length
- Total number of reads in the measurement

Facilitates transparent comparison of transcript levels within and between.

# Differential expression in operon with REP

- PCA check for RPKM corrected data :

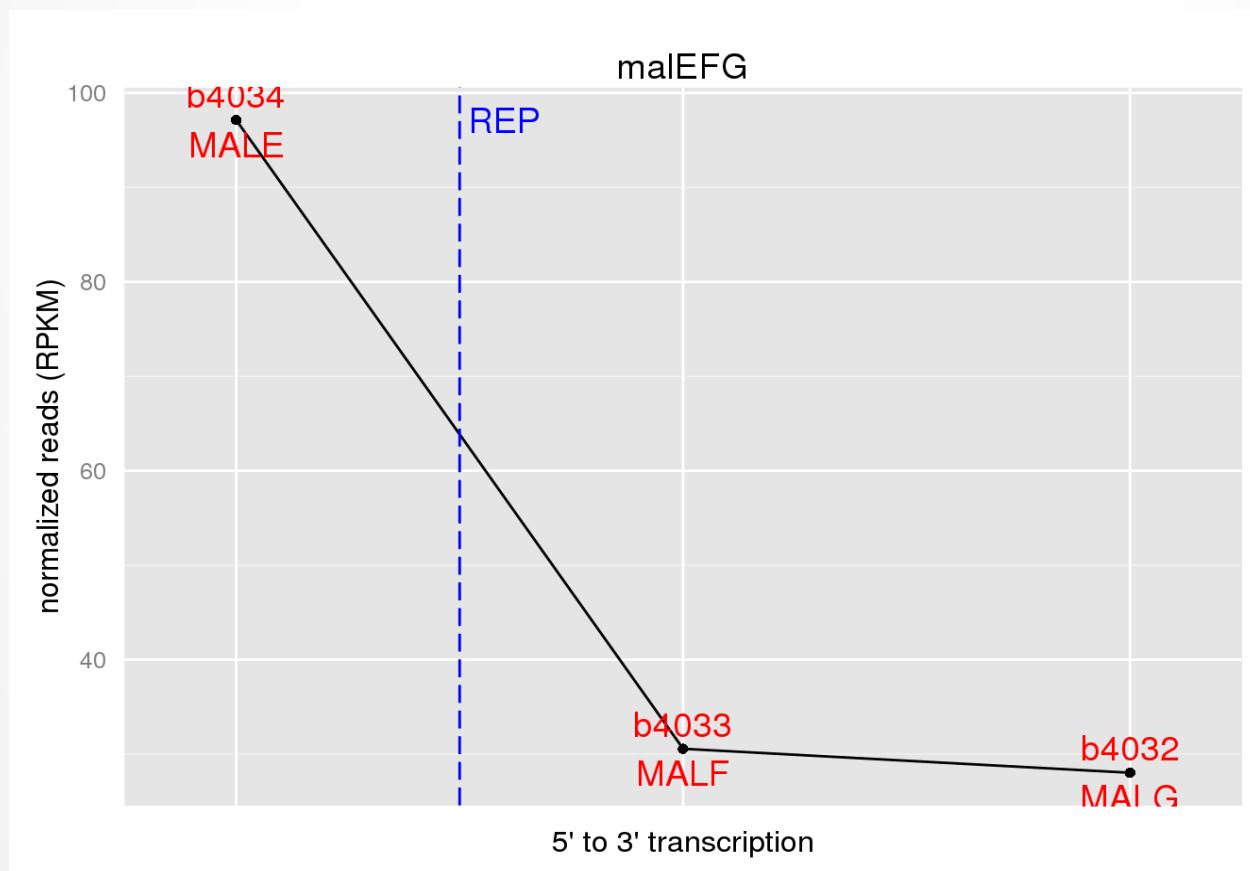


# Differential expression in operon with REP

- For genes surrounding a REP, Wilcoxon test to detect DE :
  - P-value : < 0.01
  - RPKM counts > 10 for at least one gene

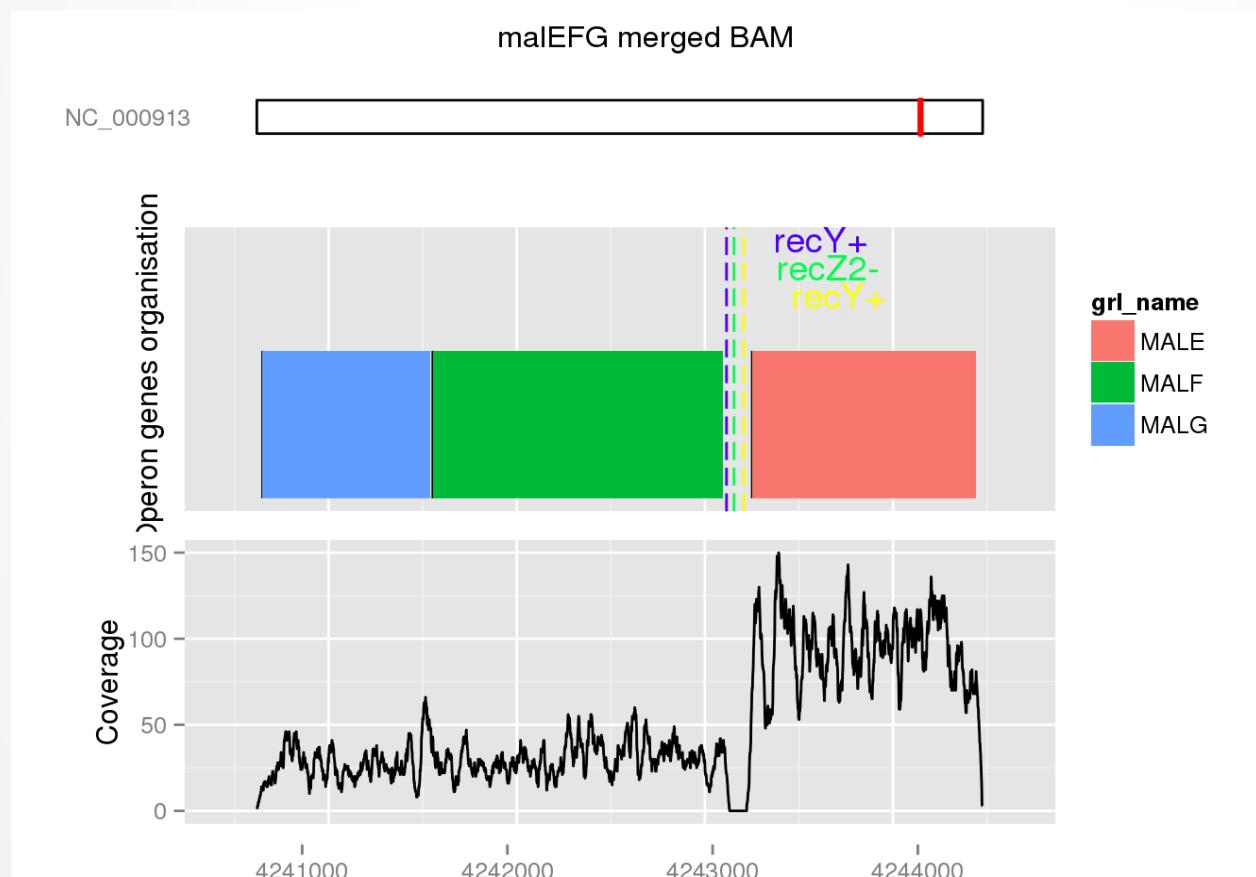
# Differential expression in operon with REP

- For genes surrounding a REP, Wilcoxon test to detect DE :
  - P-value : < 0.01
  - RPKM counts > 10 for at least one gene



# Differential expression in operon with REP

- For genes surrounding a REP, Wilcoxon test to detect DE :
  - P-value : < 0.01
  - RPKM counts > 10 for at least one gene



# Differential expression in operon with REP

- 17 over 36 operons with DE around the REP.

| Function   | Number of operons | DE change feature |
|--|-------------------|-------------------|
| ABC transporter  | 7                 | ↗                 |
|  | 2                 | ↘                 |
| Multipurpose conversions 4-aminobutyrate aminotransferase, PLP-dependent                             | 1                 | ↘                 |
| IF2 : membrane protein/conserved protein   | 1                 | ↗                 |
| Amino acid biosynthesis : Glutamine synthetase   | 1                 | ↘                 |
| Inner membrane protein, DUF485 family  | 1                 | ↗                 |
| Anaerobic respiration carbamoyl phosphate phosphatase and maturation protein for [NiFe] hydrogenases | 1                 | ↗                 |
| Degradation of small molecules: Carbon compounds mannitol-1-phosphate dehydrogenase, NAD-dependent   | 1                 | ↗                 |
| Central intermediary metabolism : Pool, multipurpose conversions glycerol kinase                     | 1                 | ↗                 |

# Differential expression in operon with REP

- 9 ABC transporter operons ([abcDB source](#)) :

| Number of ABC transporter operons | DE change feature | More expressed gene function |
|-----------------------------------|-------------------|------------------------------|
| 7                                 | ↘                 | SBP                          |
| 2                                 | ↗                 | NBD                          |

# Search for transcription change events on BIME

- GFF modifications to convert BIME regions as mRNA.
- Extraction of genomic positions for genes bounded regions with BIME like :  

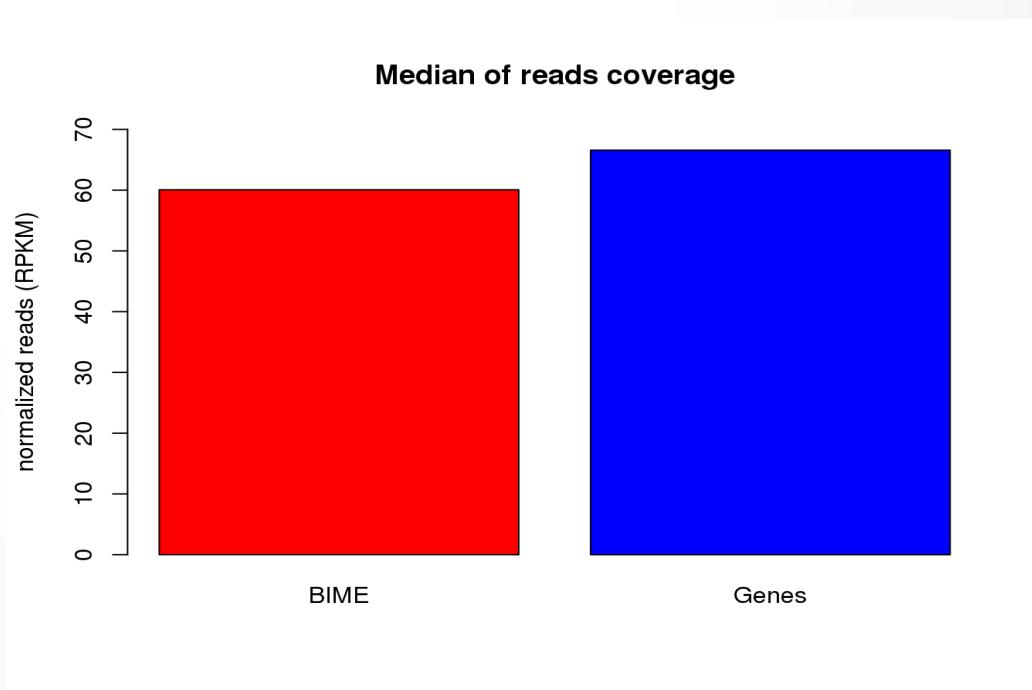

The diagram illustrates a genomic region with two genes, each represented by a blue bracket labeled {Gene}. Between these genes is a central red bracket labeled {IntraGenic-BIME-IntraGenic}, which represents a BIME-like intragenic region. Ellipses indicate additional genes and regions between them.
- Coverage per base calculation on these regions (**Bedtools**)
- RPKM correction.

# Search for transcription change events on BIME

- GFF modifications to convert BIME regions as mRNA.
- Extraction of genomic positions for genes bounded regions with BIME like :

{Gene} {IntraGenic-BIME-IntraGenic} ... {Gene}

- Coverage per base calculation on these regions ([Bedtools](#))
- RPKM correction.



# Search for transcription change events on BIME

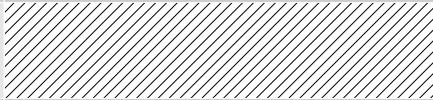
- Looking for correlation between :  
\_\_\_\_|^\_\_ or \_\_|\_\_\_\_ profiles and coverage.
- Profiles simulated with 0 or 1.
- Sliding window (100 bp) on the BIME region.

# Search for transcription change events on BIME

- Looking for correlation between :  
\_\_\_\_|^\_\_ or \_\_|\_\_\_\_ profiles and coverage.
- Profiles simulated with 0 or 1.
- Sliding window (100 bp) on the BIME region.
- Correlation test :
  - same strand : Wilcoxon test for DE
  - one gene with RPKM count > 10
  - 2 fold change
  - Correlation test P-value <  $10^{-7}$
  - Correlation coefficient > 0.7

# Search for transcription change events on BIME

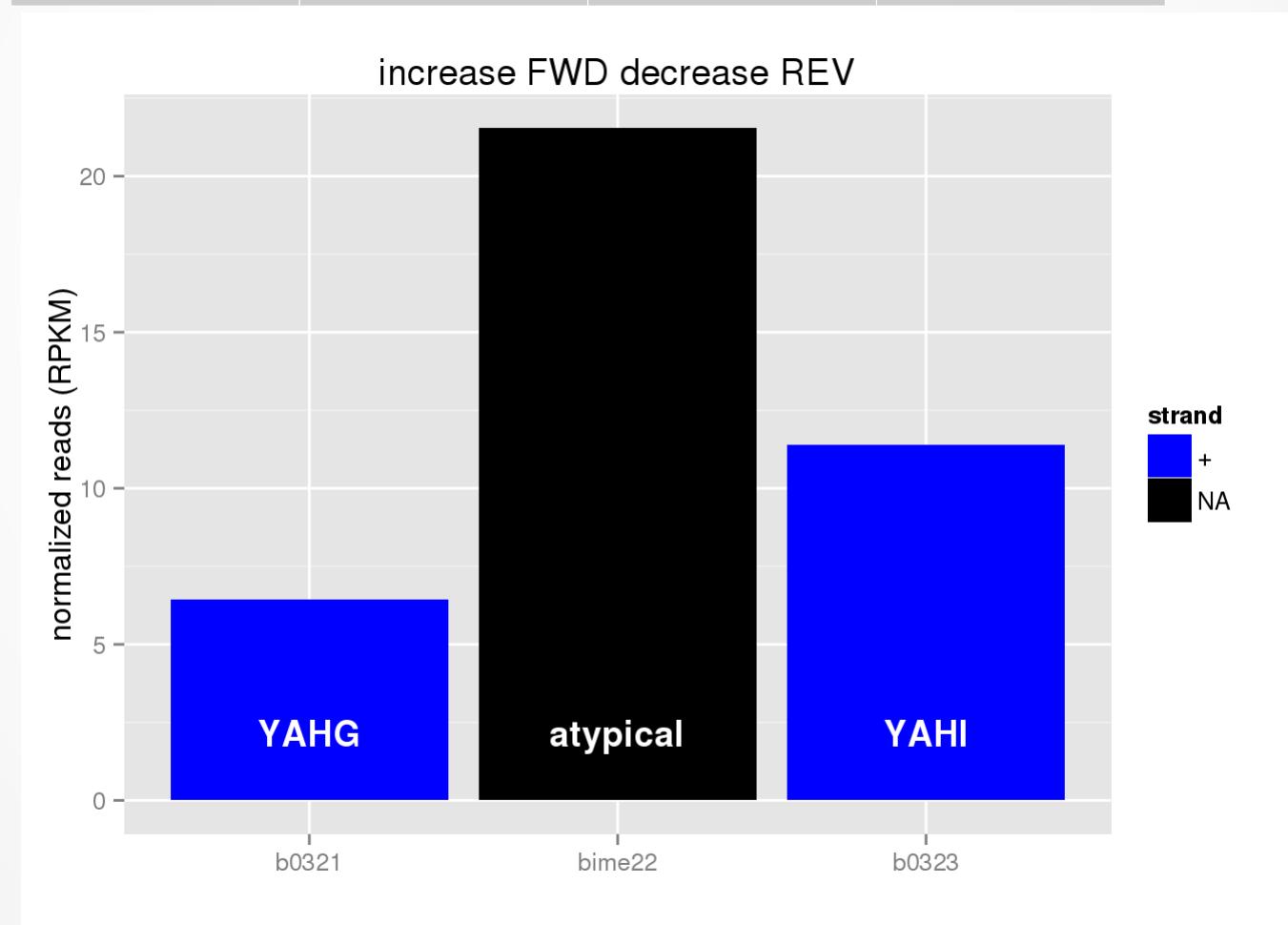
## Results overview :

| Number of regions | Number of regions without unique REP  | Event  | on BIME region |
|-------------------|---|--------|----------------|
| 29                | 25  | __ ^__ | YES            |
| 316               |   | __ ^__ | NO             |
| 43                | 29  | ^__ __ | YES            |
| 308               |  | ^__ __ | NO             |

# On BIME \_\_\_\_|^^^ : stabilisation

1 case :

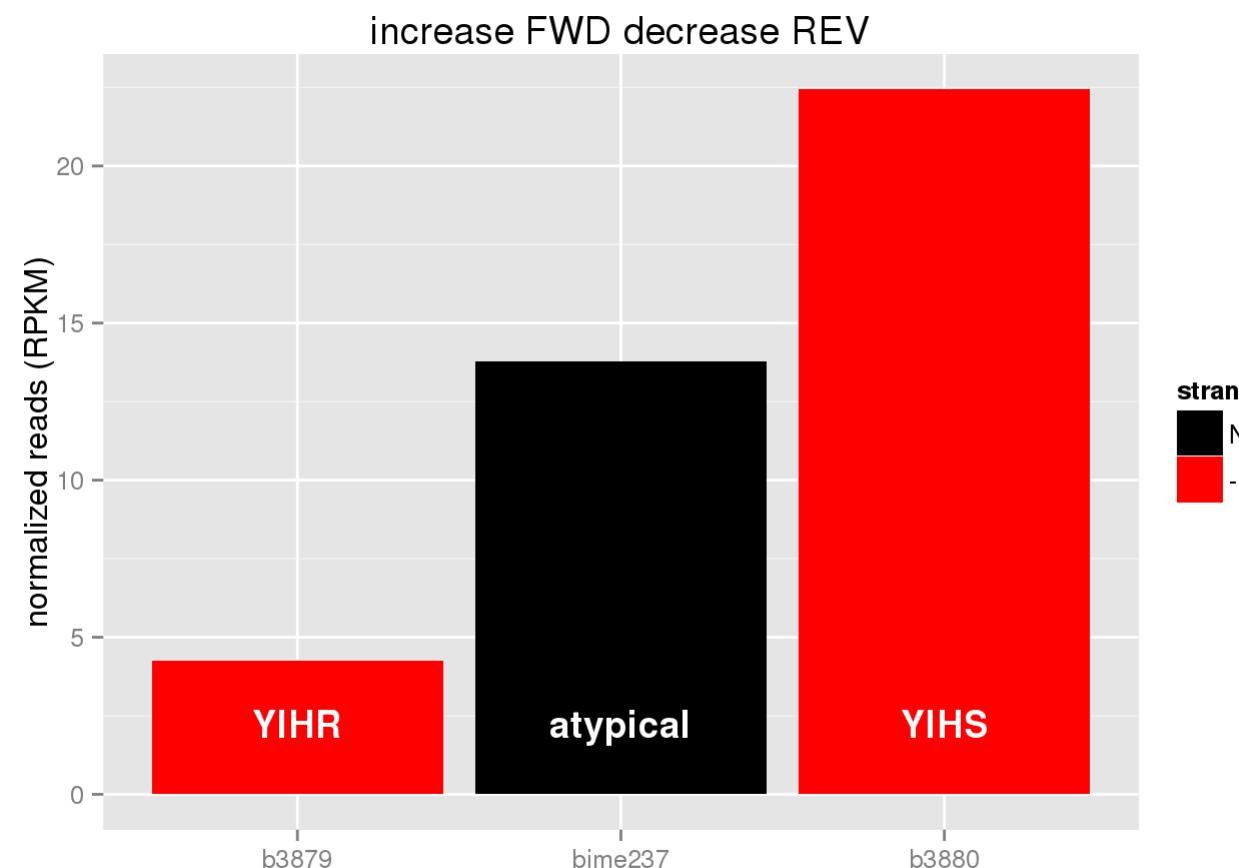
| BIME-1 | BIME-2 | atypical BIME | unique REP |
|--------|--------|---------------|------------|
| 0      | 0      | 1             | 0          |



# On BIME \_\_\_\_|^^\^ : stabilisation

5 cases :

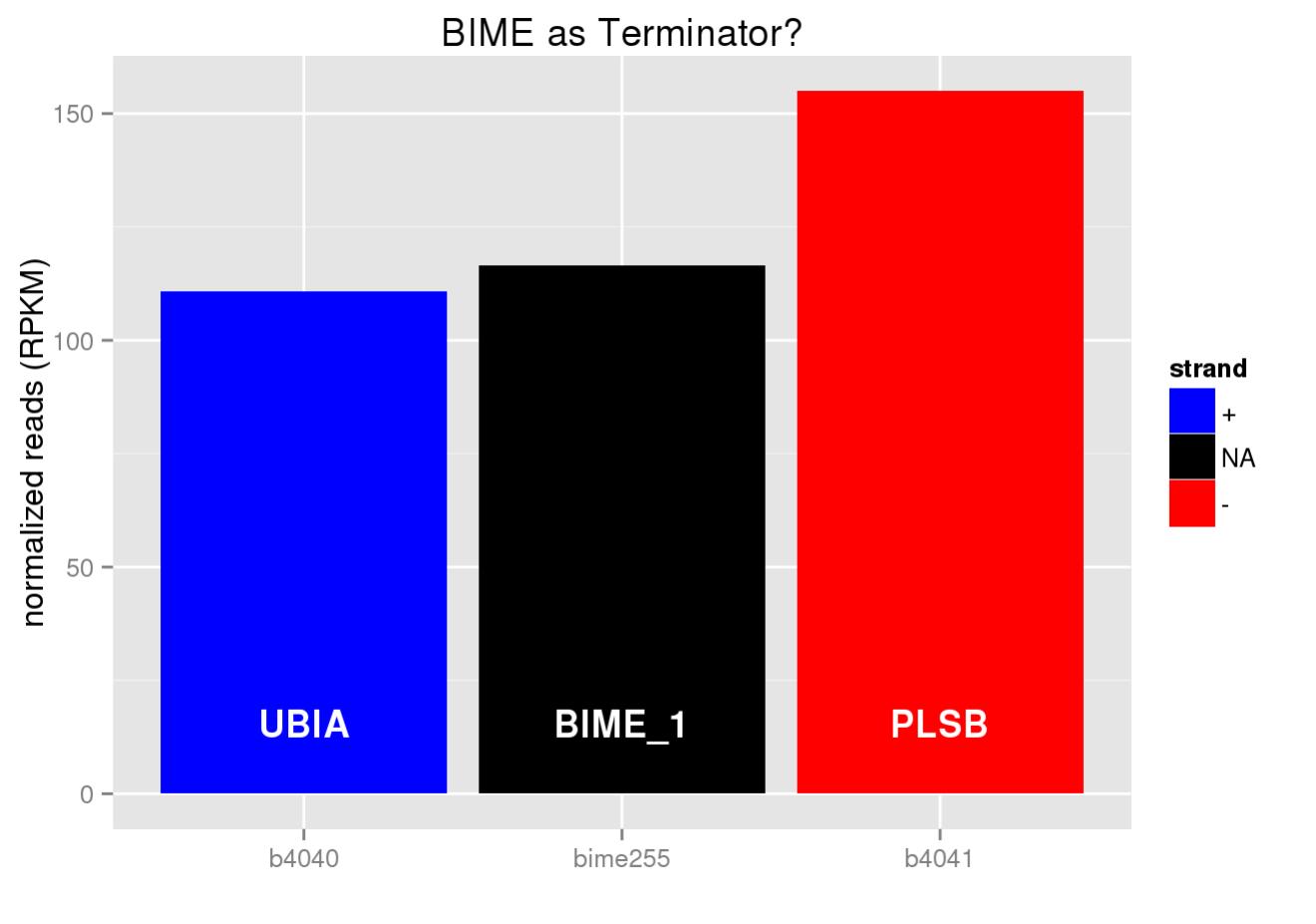
| BIME-1 | BIME-2 | atypical BIME | unique REP |
|--------|--------|---------------|------------|
| 3      | 0      | 1             | 1          |



# On BIME \_\_\_\_|^^^ : terminator

23 cases :

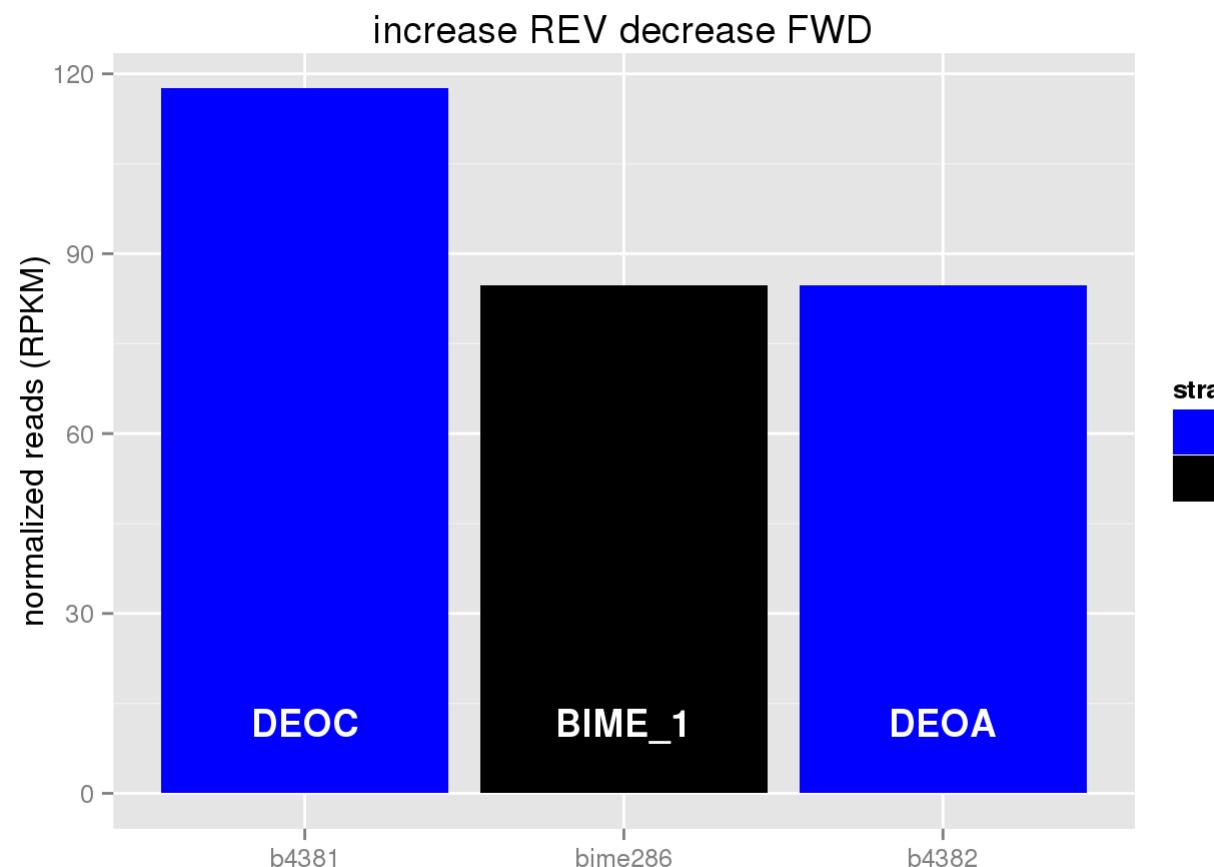
| BIME-1 | BIME-2 | atypical BIME | unique REP |
|--------|--------|---------------|------------|
| 11     | 2      | 3             | 7          |



# On BIME ^^^|\_\_\_\_ : stabilisation

9 cases :

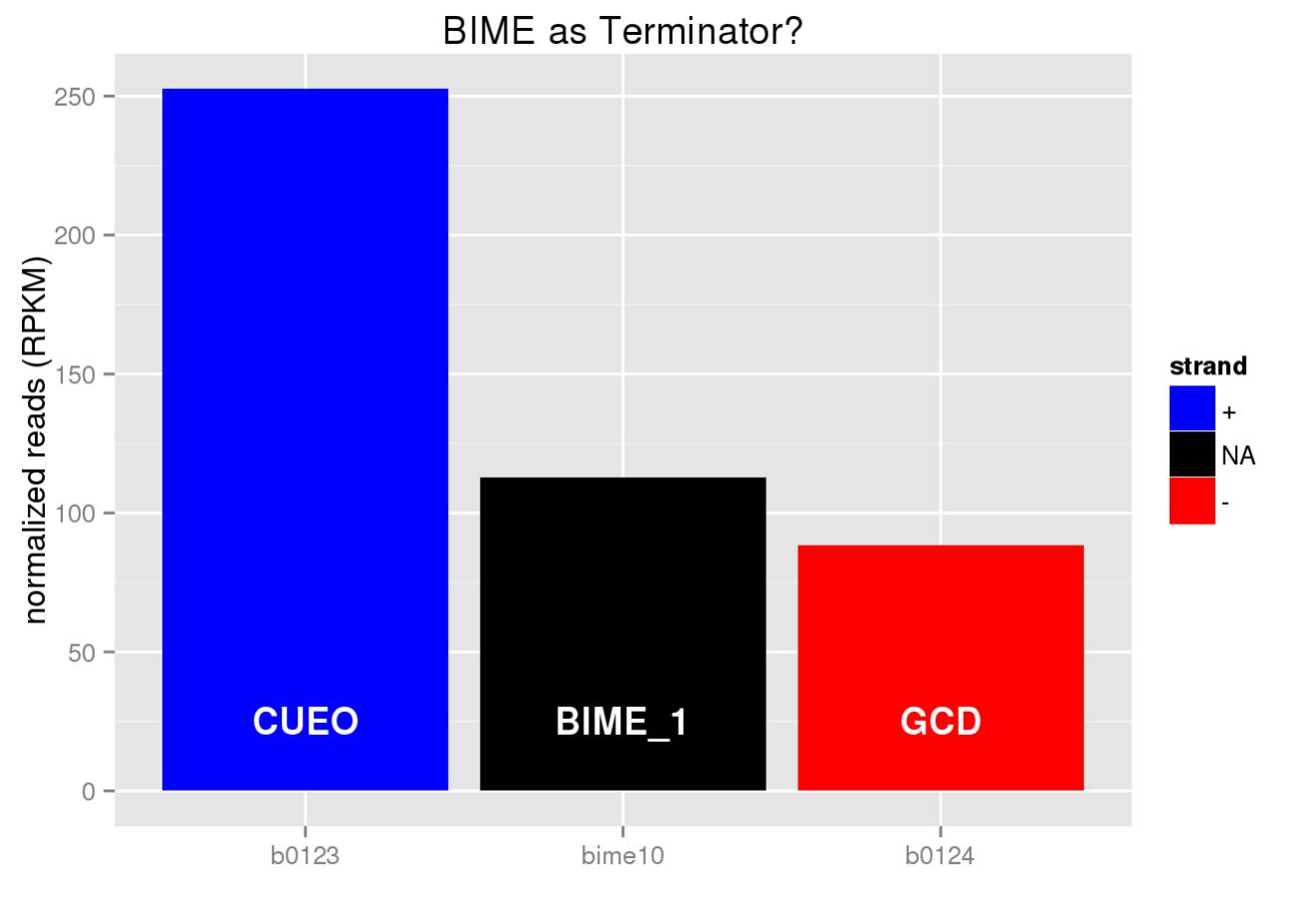
| BIME-1 | BIME-2 | atypical BIME | unique REP |
|--------|--------|---------------|------------|
| 5      | 2      | 1             | 1          |



# On BIME ^^^|\_\_ : terminator

34 cases :

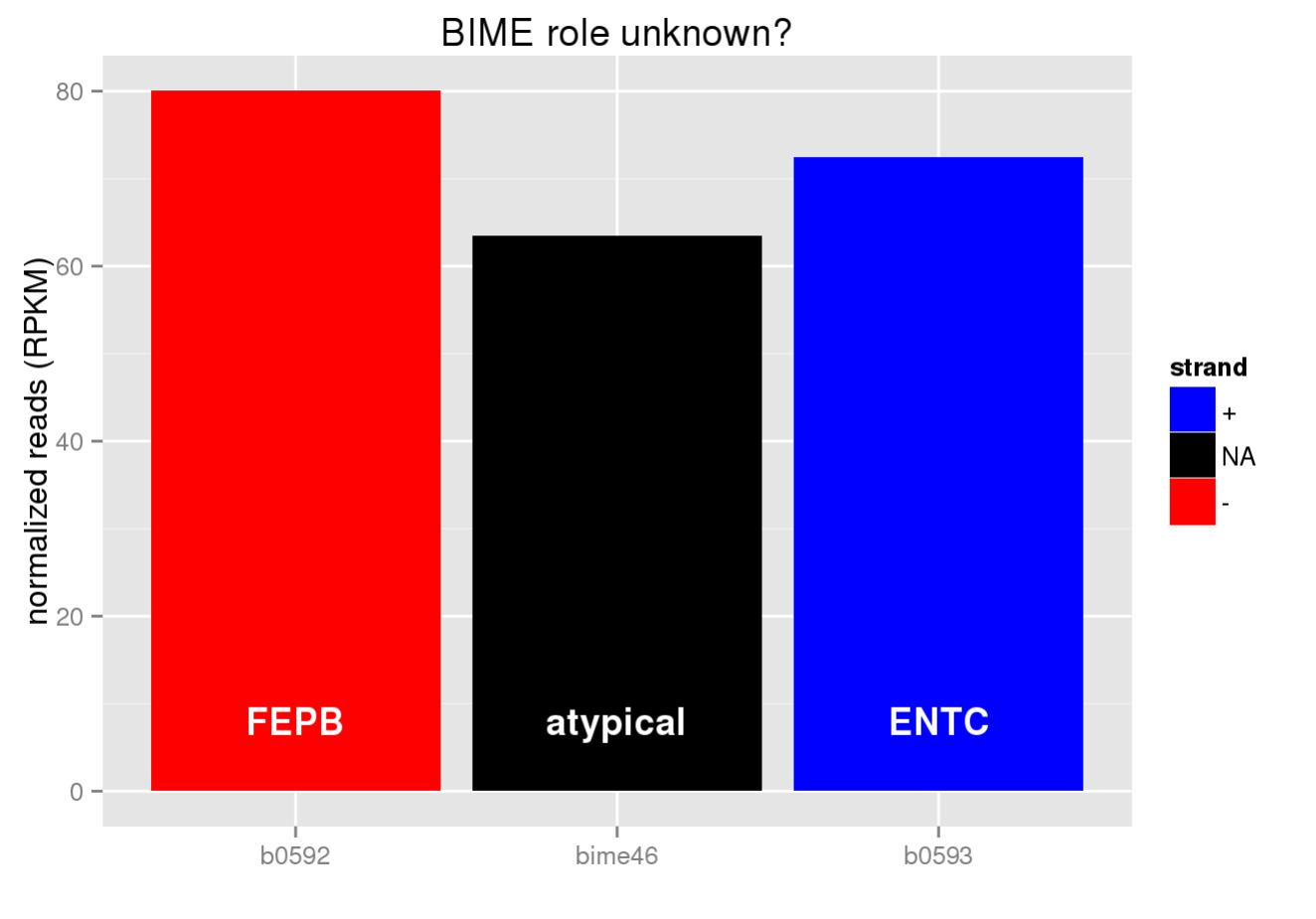
| BIME-1 | BIME-2 | atypical BIME | unique REP |
|--------|--------|---------------|------------|
| 10     | 8      | 3             | 13         |



# On BIME ^^^|\_\_ : enhancing

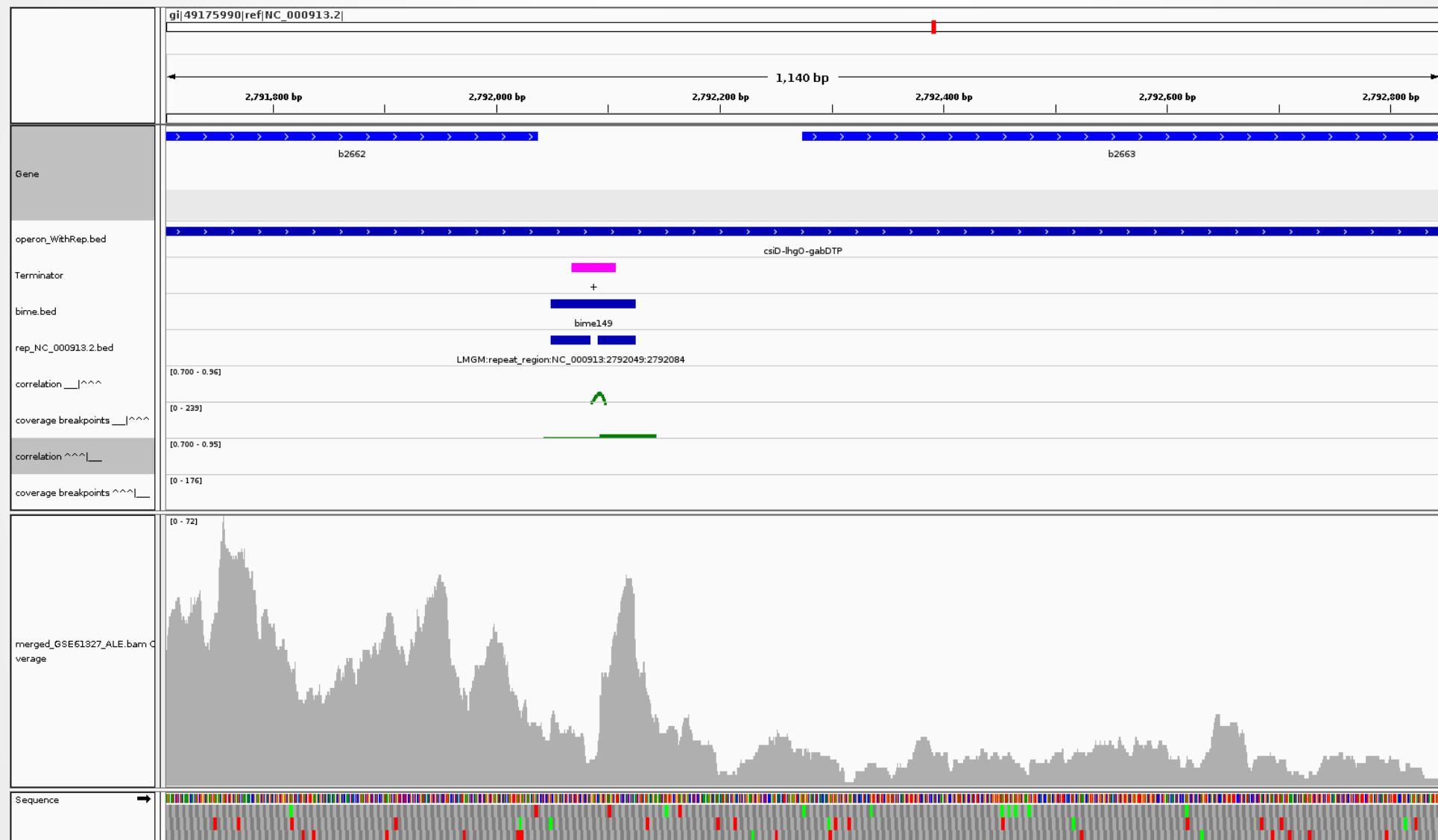
1 case :

| BIME-1 | BIME-2 | atypical<br>BIME | unique REP |
|--------|--------|------------------|------------|
| 0      | 0      | 1                | 0          |



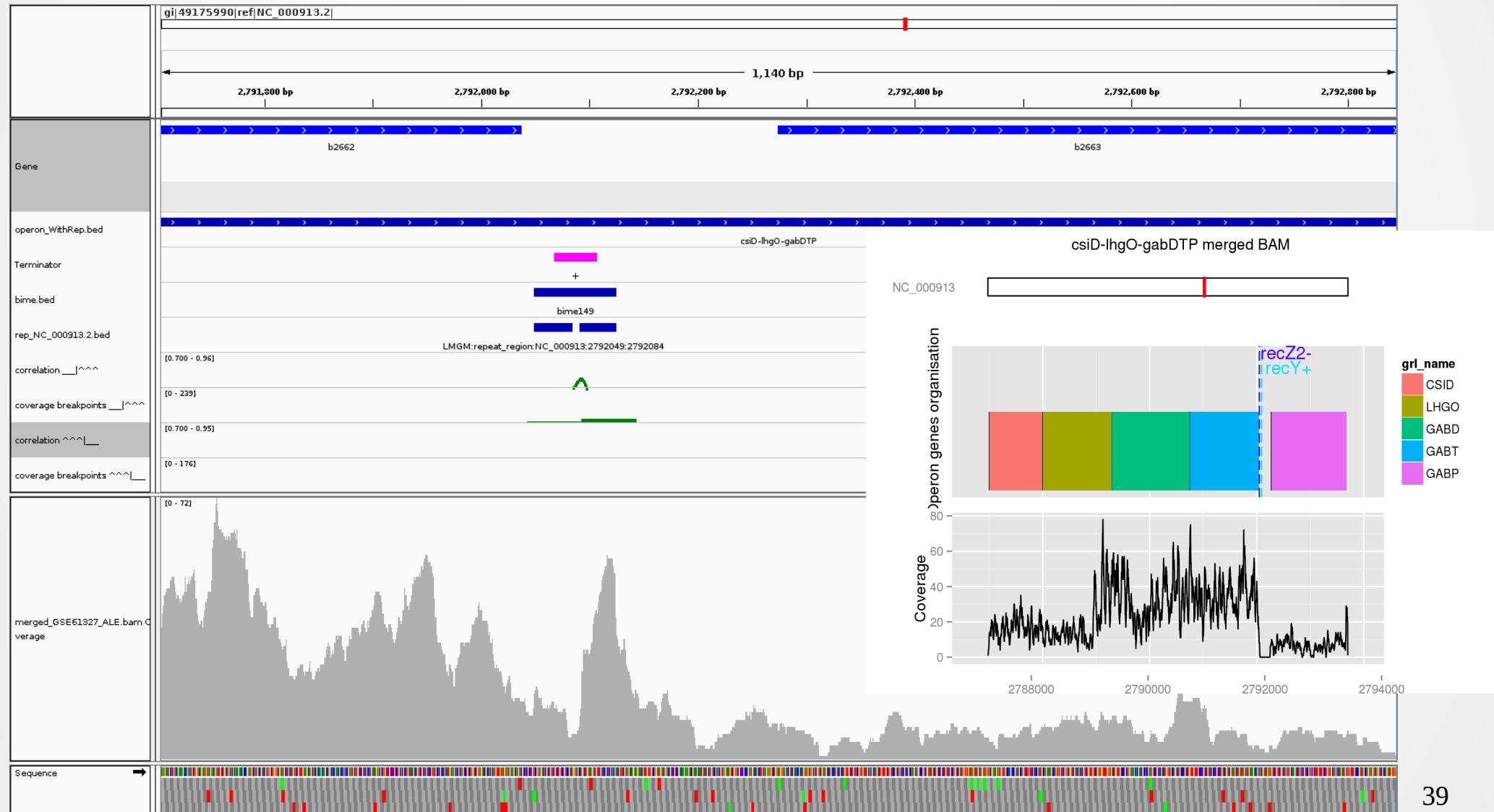
# Links with REP results

Operon [csiD-lhgO-gabDTP](#) (Multipurpose conversions 4-aminobutyrate aminotransferase, PLP-dependent)



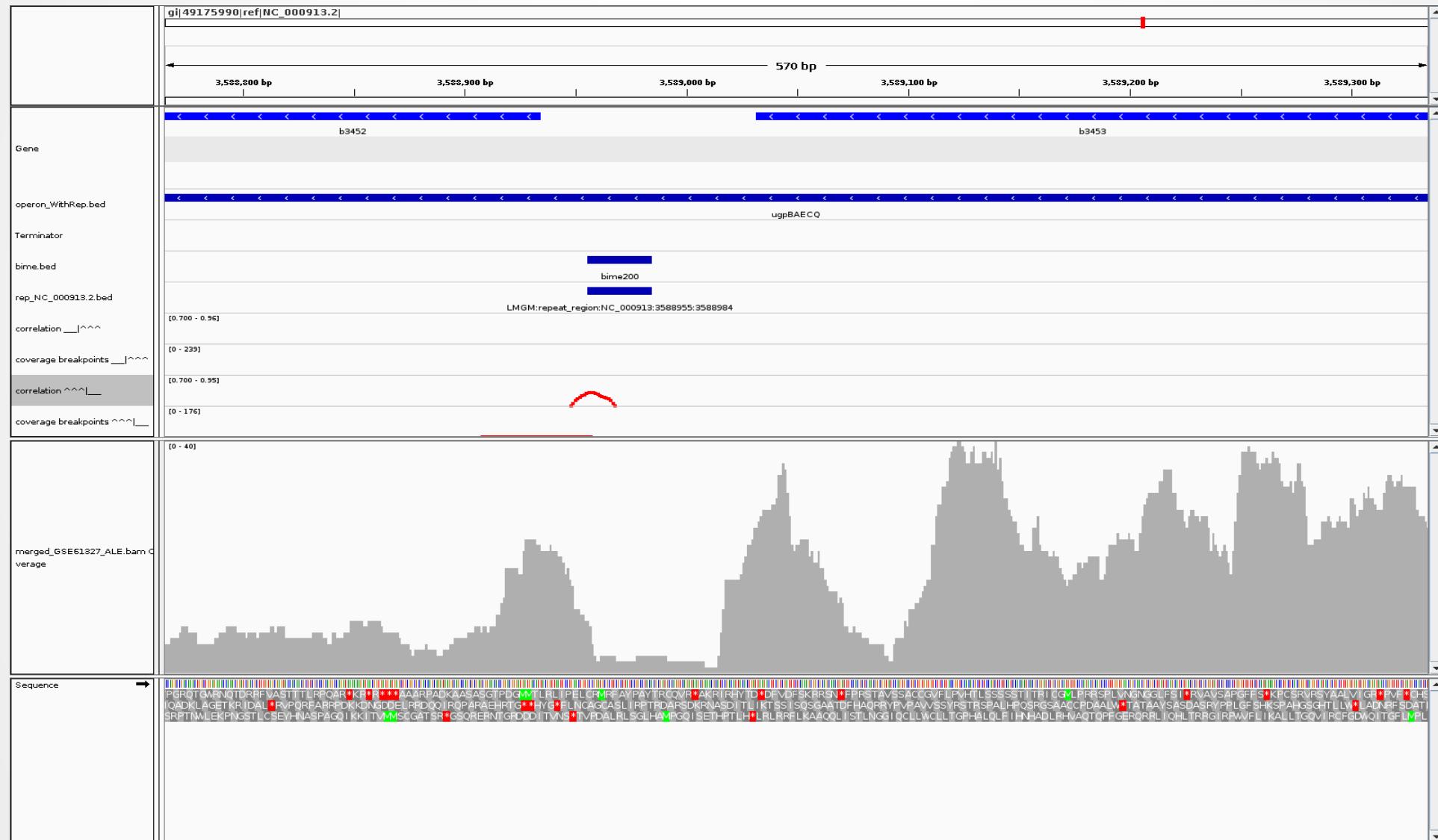
# Links with REP results

Operon [csiD-lhgO-gabDTP](#) (Multipurpose conversions 4-aminobutyrate aminotransferase, PLP-dependent)



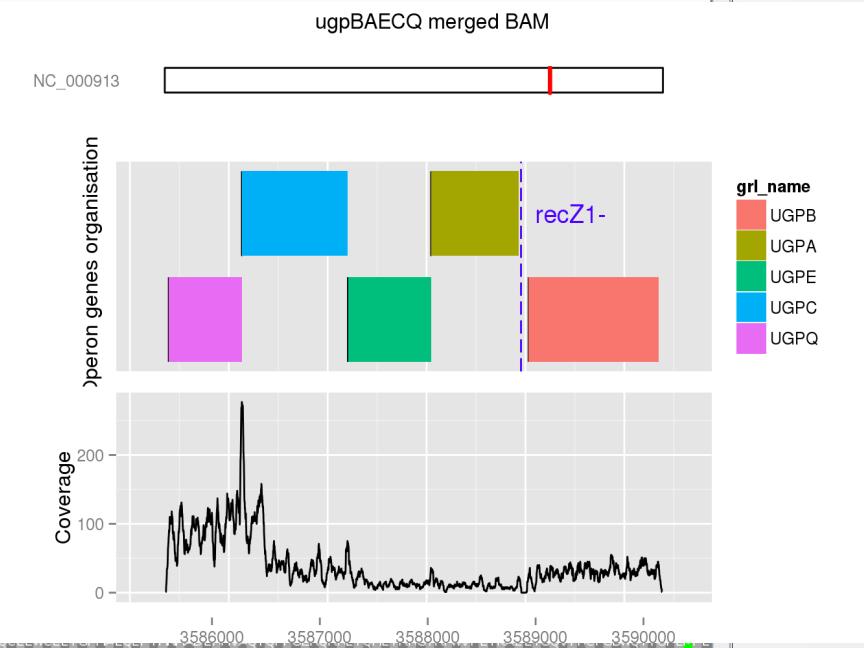
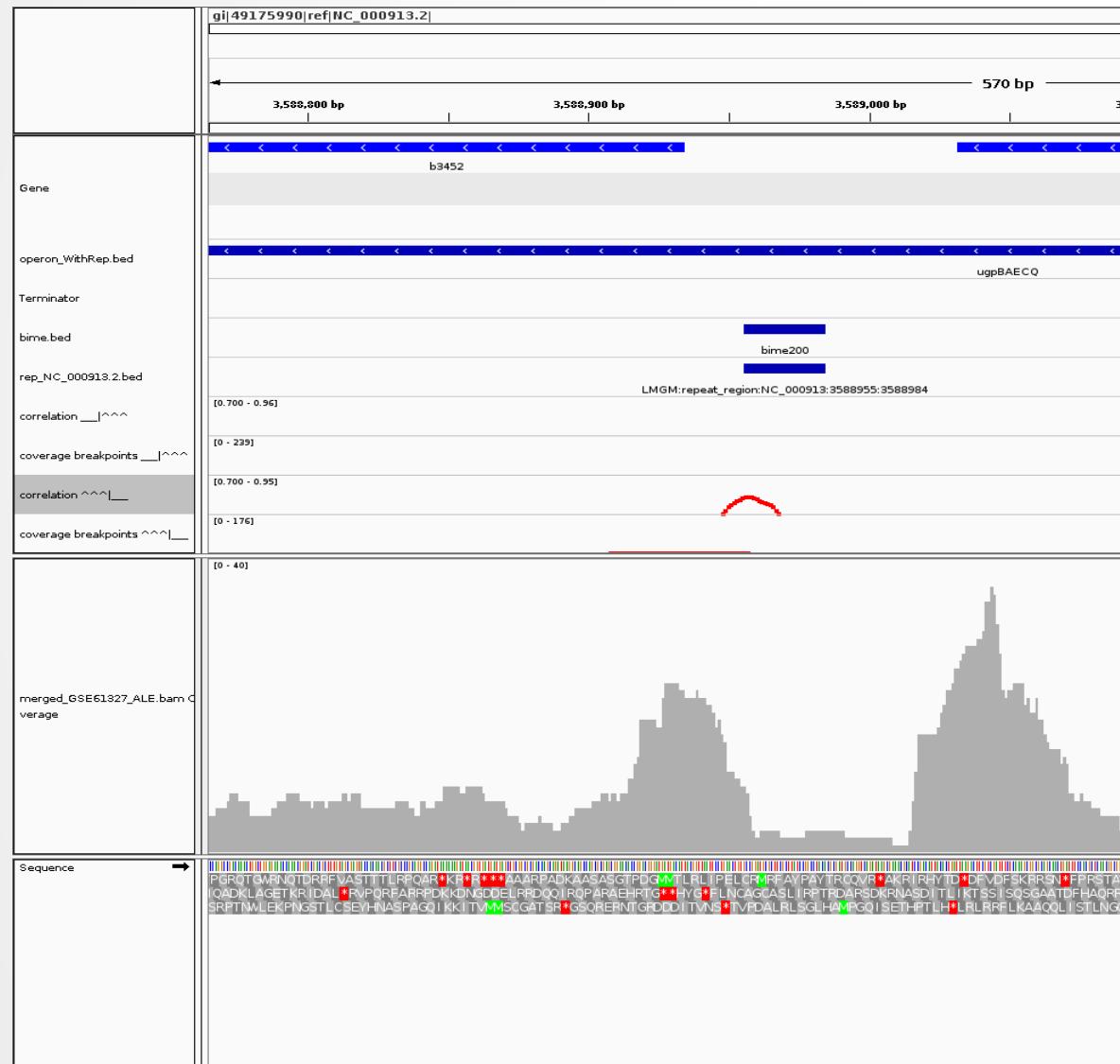
# Links with REP results

## Operon upgBAECQ (ABC transporter)



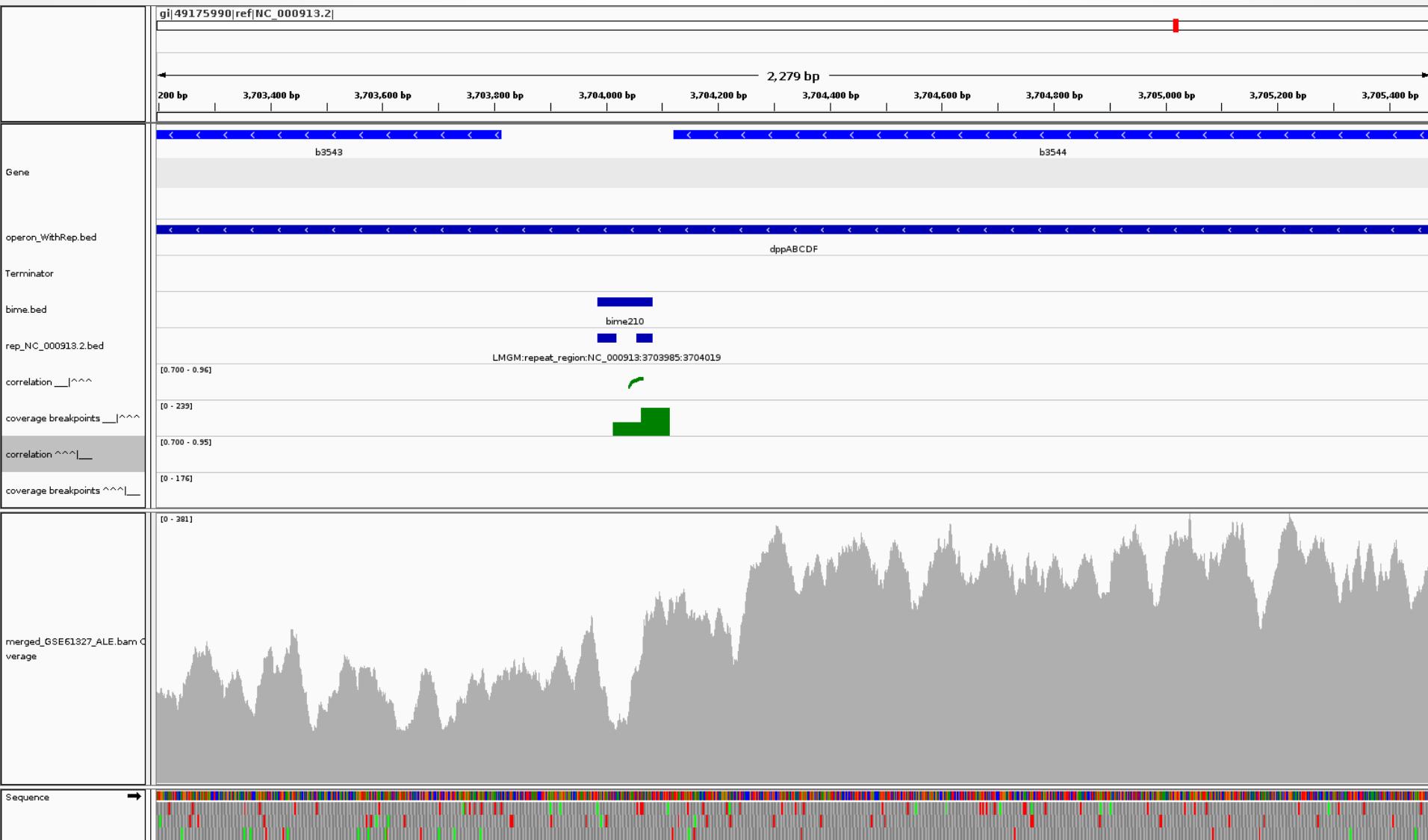
# Links with REP results

## Operon upgBAECQ (ABC transporter)



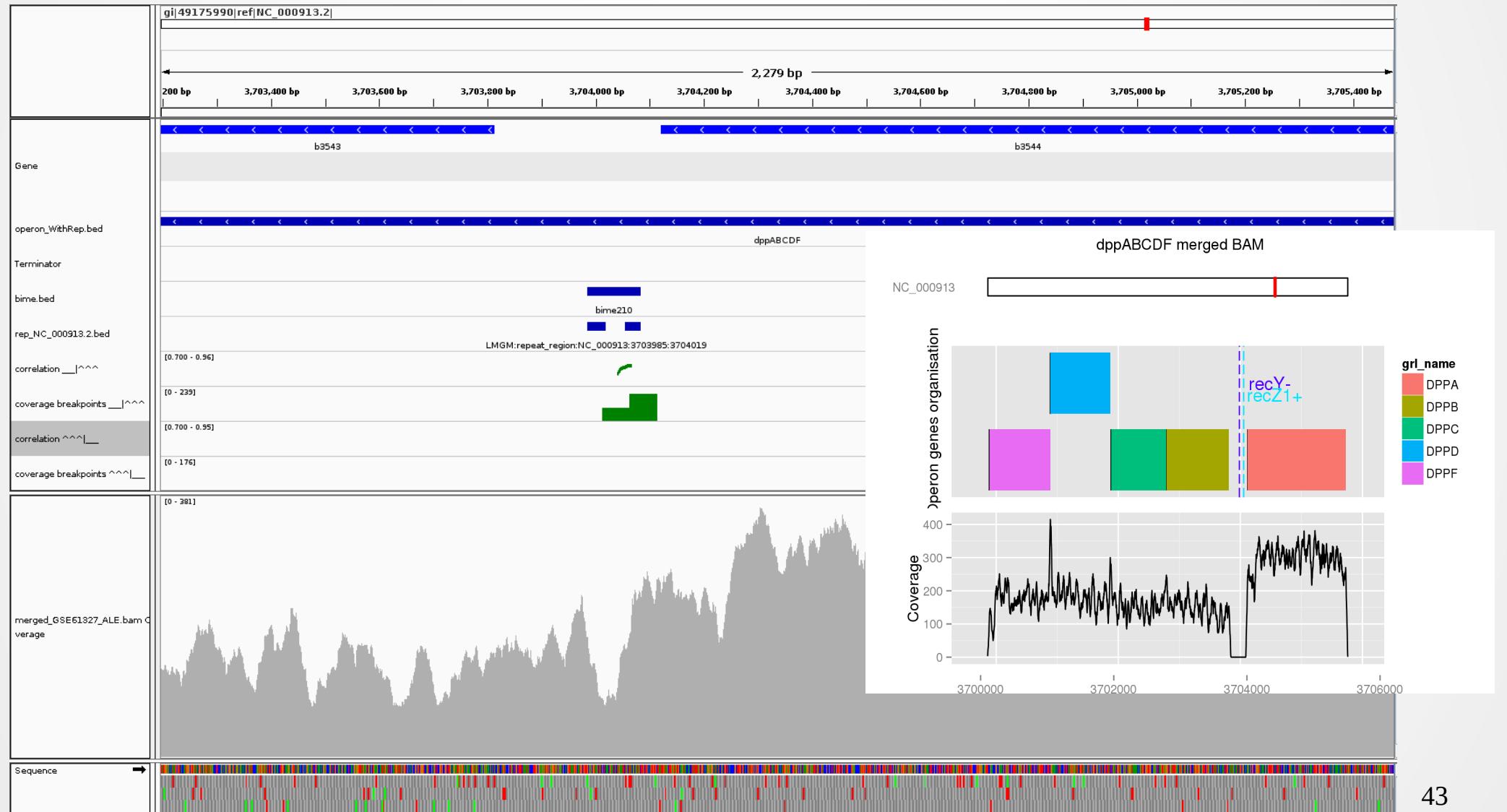
# Links with REP results

## Operon dppABCDF (ABC transporter)



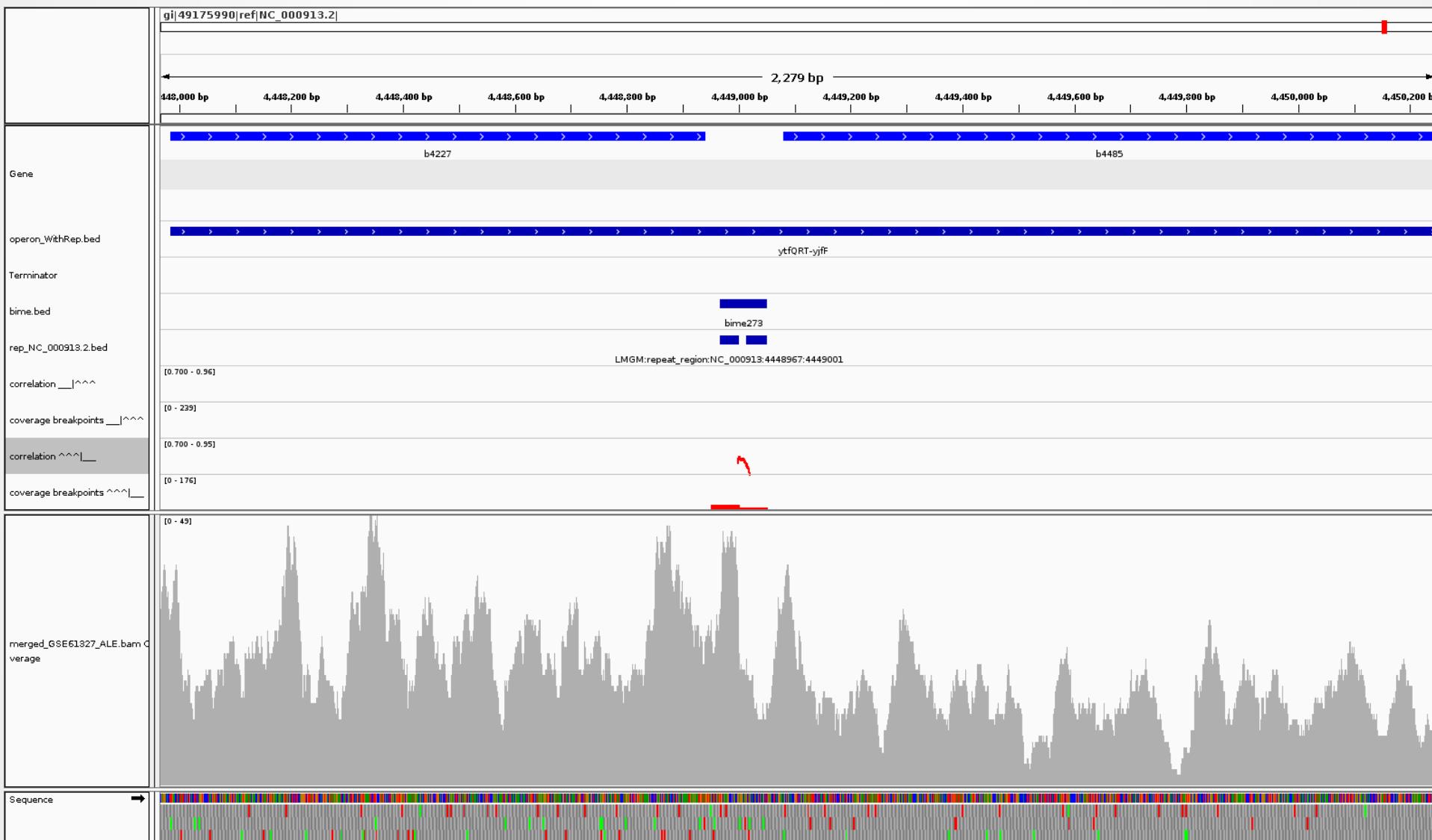
# Links with REP results

## Operon dppABCDF (ABC transporter)



# Links with REP results

## Operon *ytfQRT-yjfF* (ABC transporter)



# Links with REP results

## Operon *ytfQRT-yjfF* (ABC transporter)

