



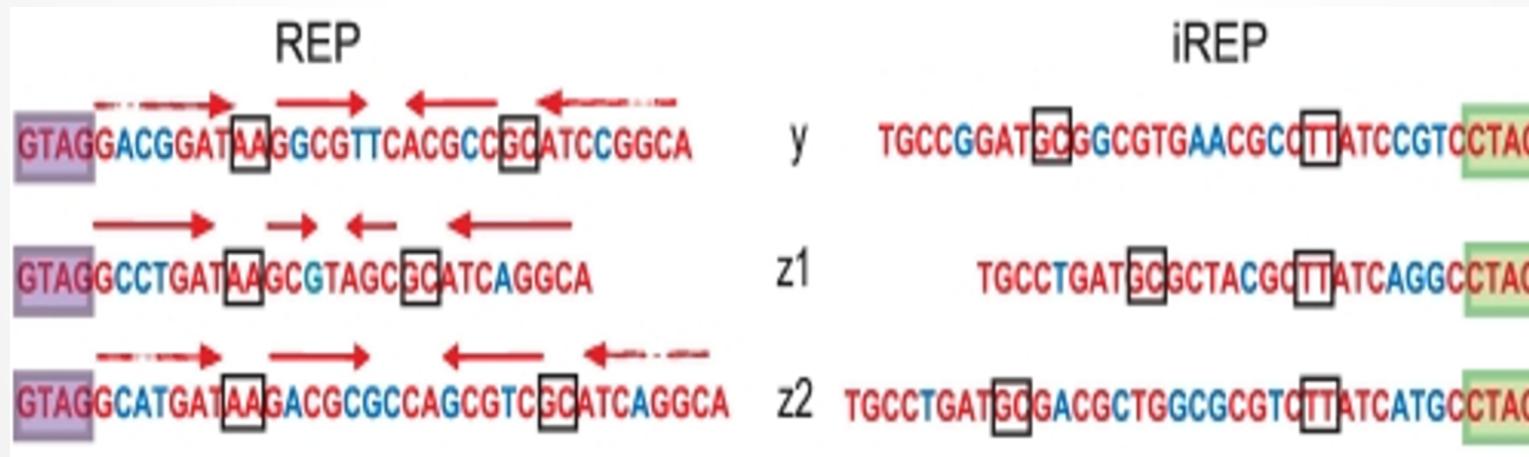
REP involved in transcription regulation?

REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z¹ and Z².

REP features

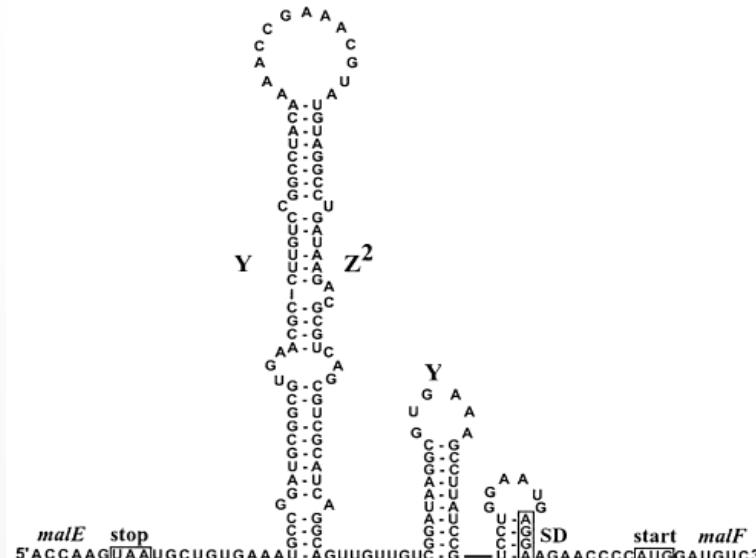
- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z¹ and Z².



Ton-Hoang et al. 2012

REP features

- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z¹ and Z².
- Stem-loop secondary structure with a bulge.



REP features

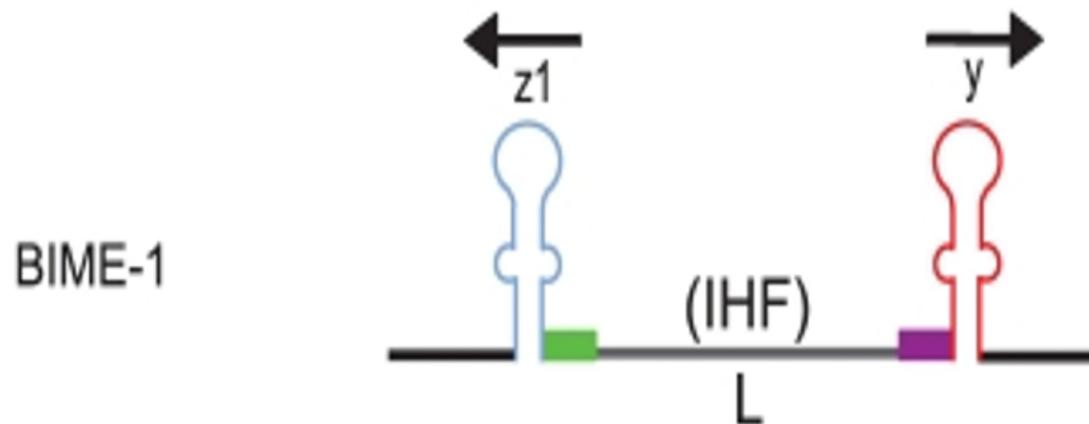
- Repeated Extragenic Palindrome.
- 20 - 40 nucleotides.
- 3 classes depending on length and secondary structure : Y, Z¹ and Z².
- Stem-loop secondary structure with a bulge.
- Numerous bacterial genomes, ≈ 1% for E. coli.
- BIME organization.

BIME features

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

BIME features

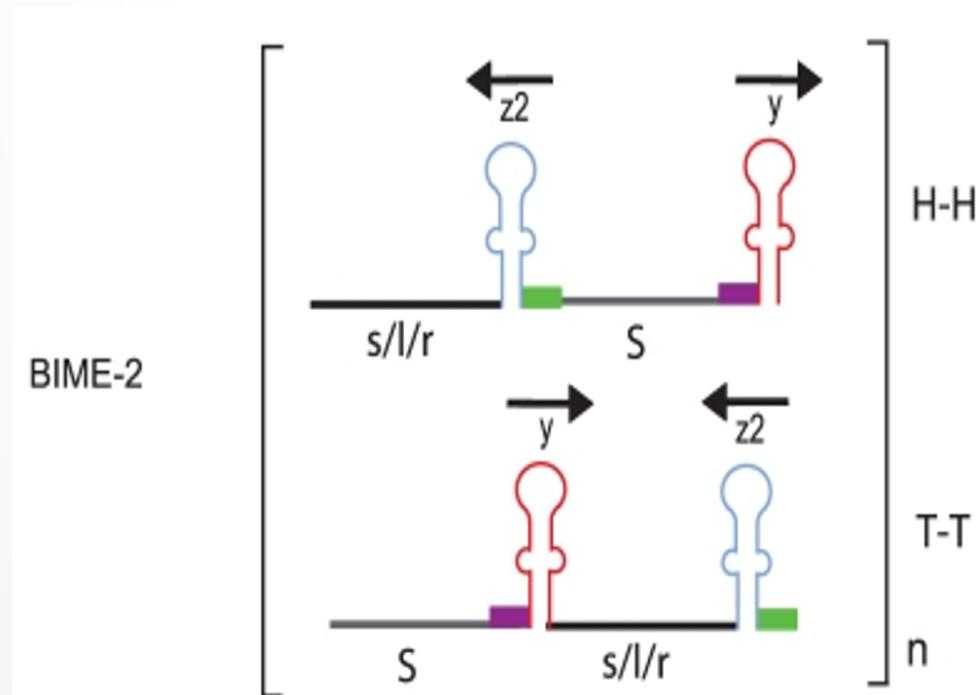
- Bacterial Interspesed Mosaic Element.
- 3 classes depending on REP composition :
Bime-1, Bime-2, atypical.



Ton-Hoang et al. 2012

BIME features

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



Objectives

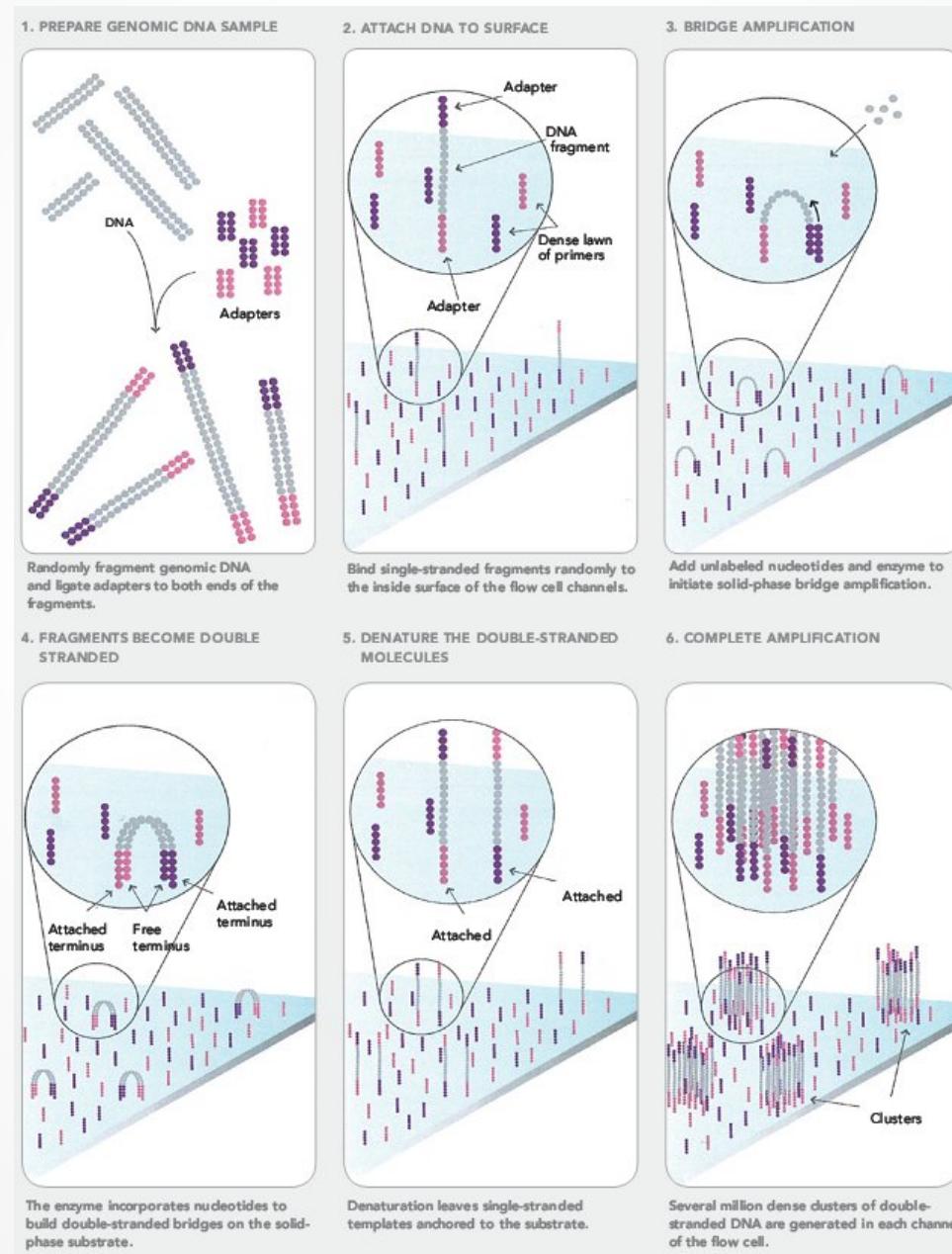
Does REP are involved in the transcription regulation in E. coli genome :

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

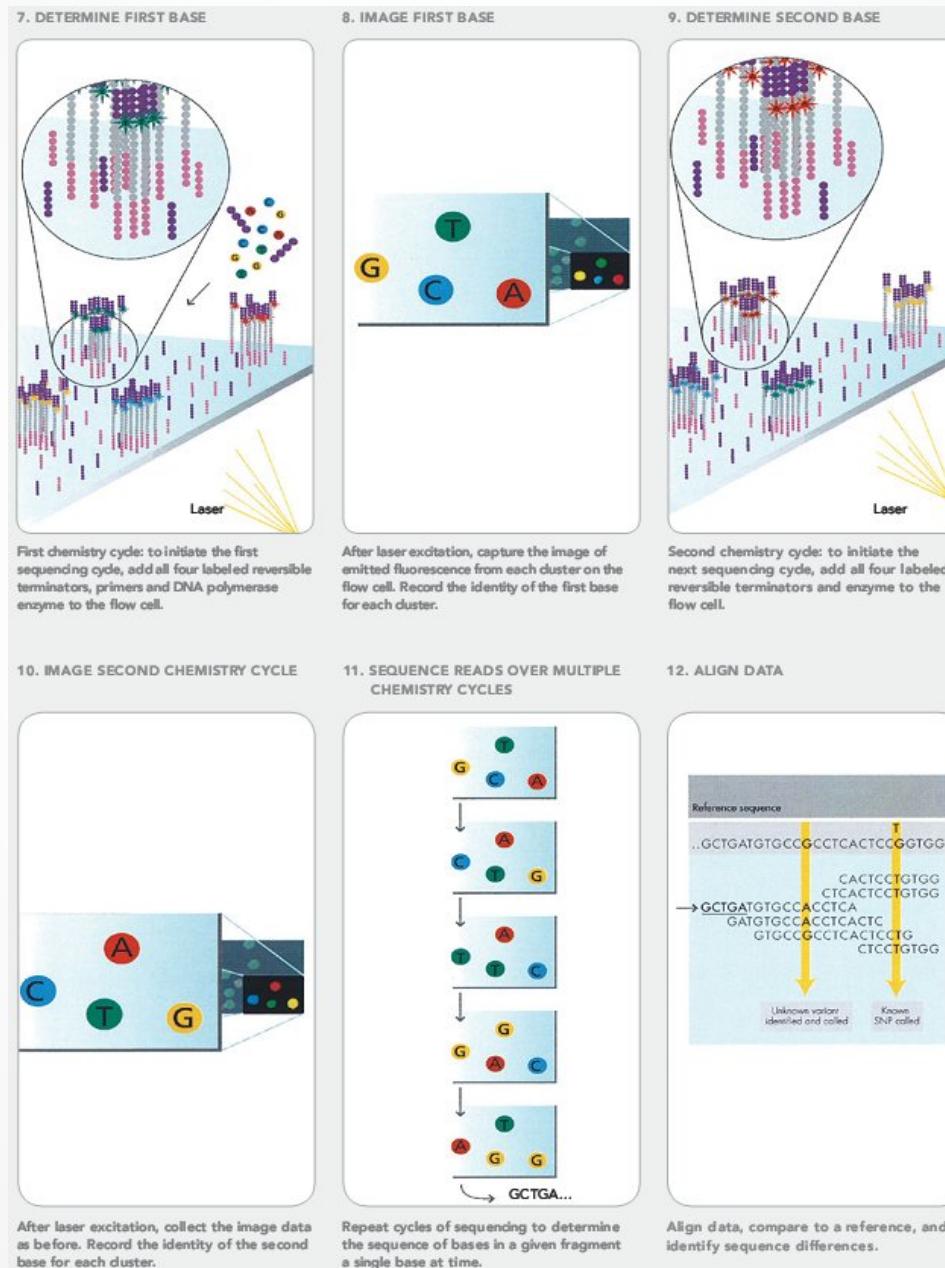
Data choice

- 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.

Illumina technology



Illumina technology

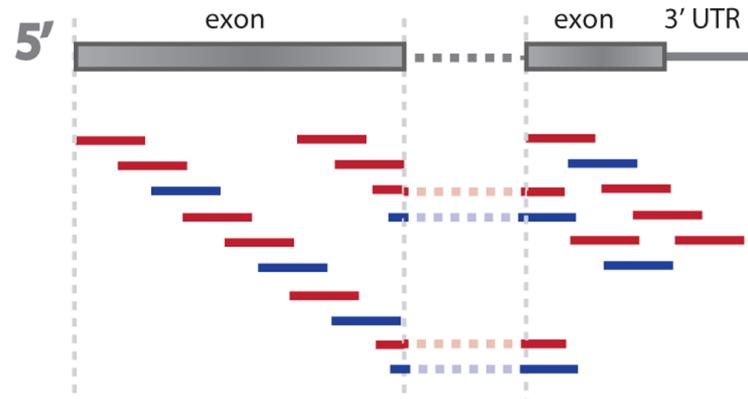


Paired-End vs. Single-End

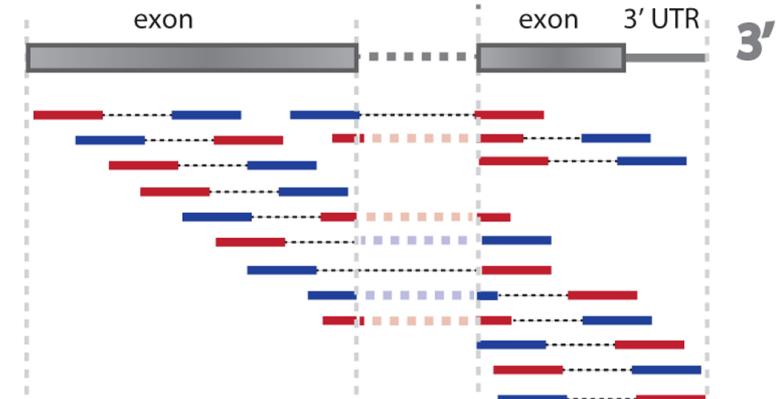
Paired-End :

- Tag1 for 5' end, Tag2 for 3' end
- Distance fixed and known between tags (< 0 to 500 bp)
- Mapping easier even if inner sequence unknown.

Single-end sequencing



Paired-end sequencing

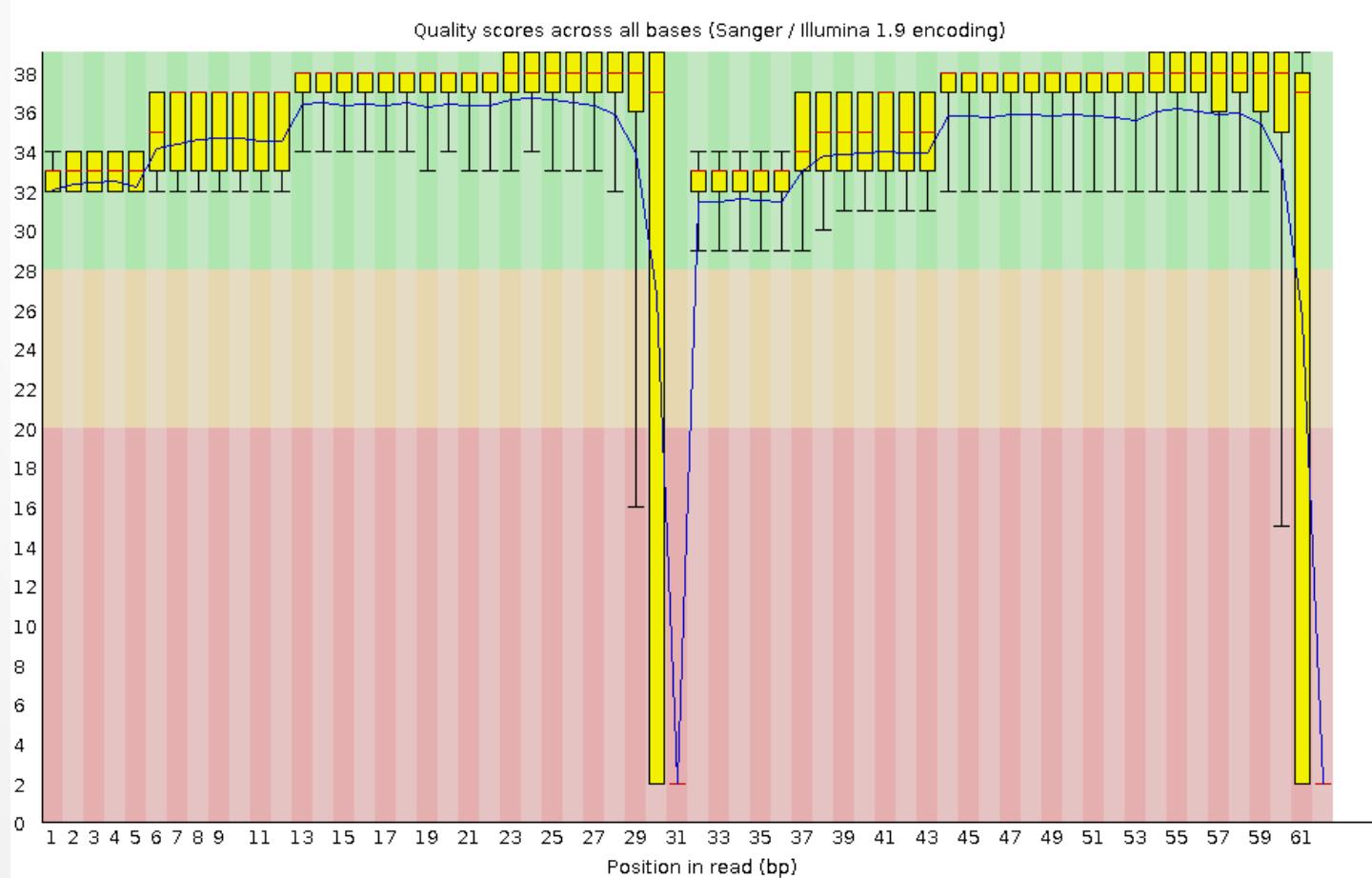


Zhernakova et al. 2013

Quality control

Quality control with
fastqc on ALE
samples :

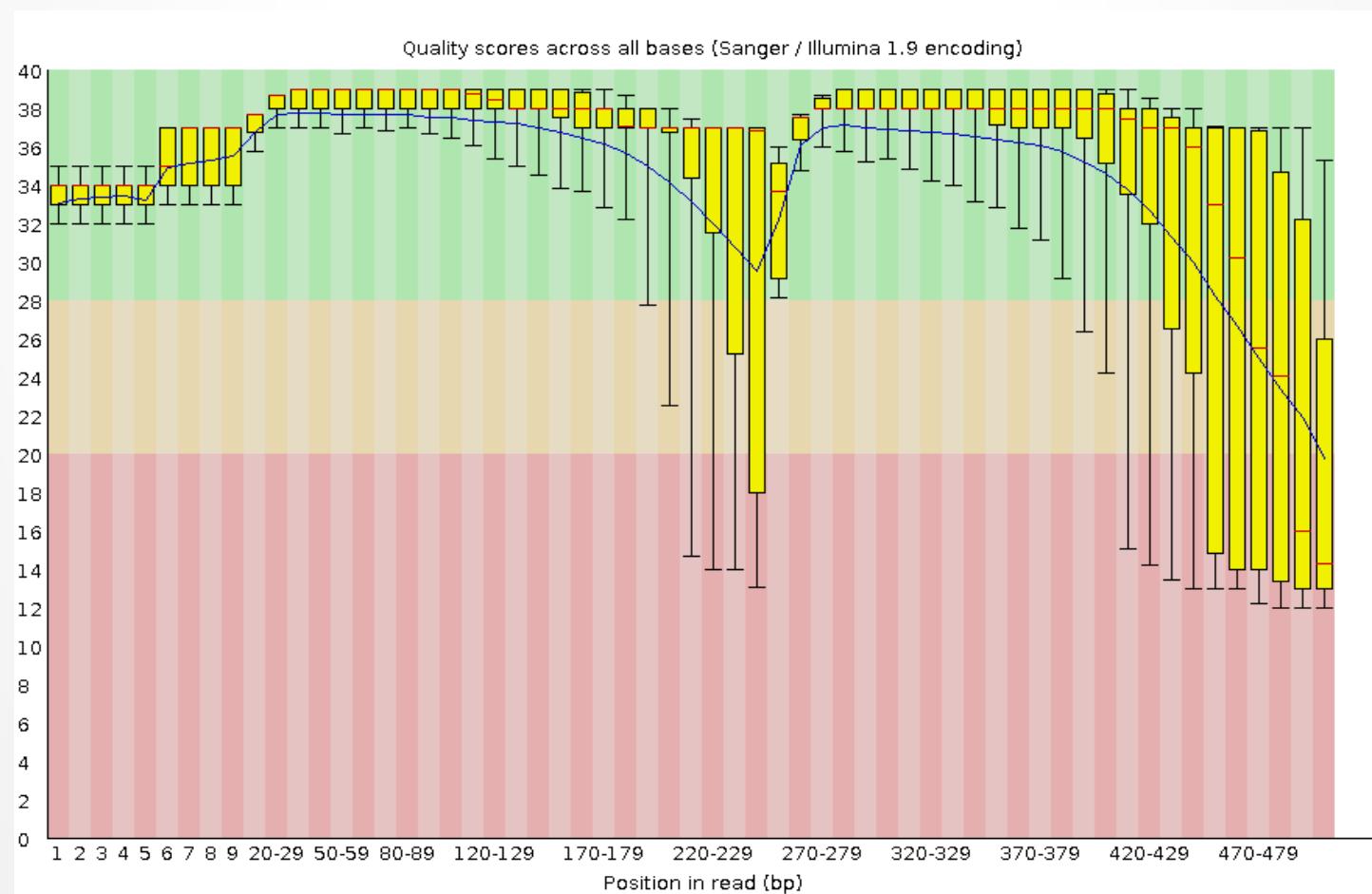
Measure	Value
Total sequences	7 772 244
Sequences flagged as poor quality	0
Sequence length	62
% GC	53



Quality control

Quality control with
fastqc on ALE
samples :

1 ALE sample rejected (35-502 bp)
SRR1573441



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

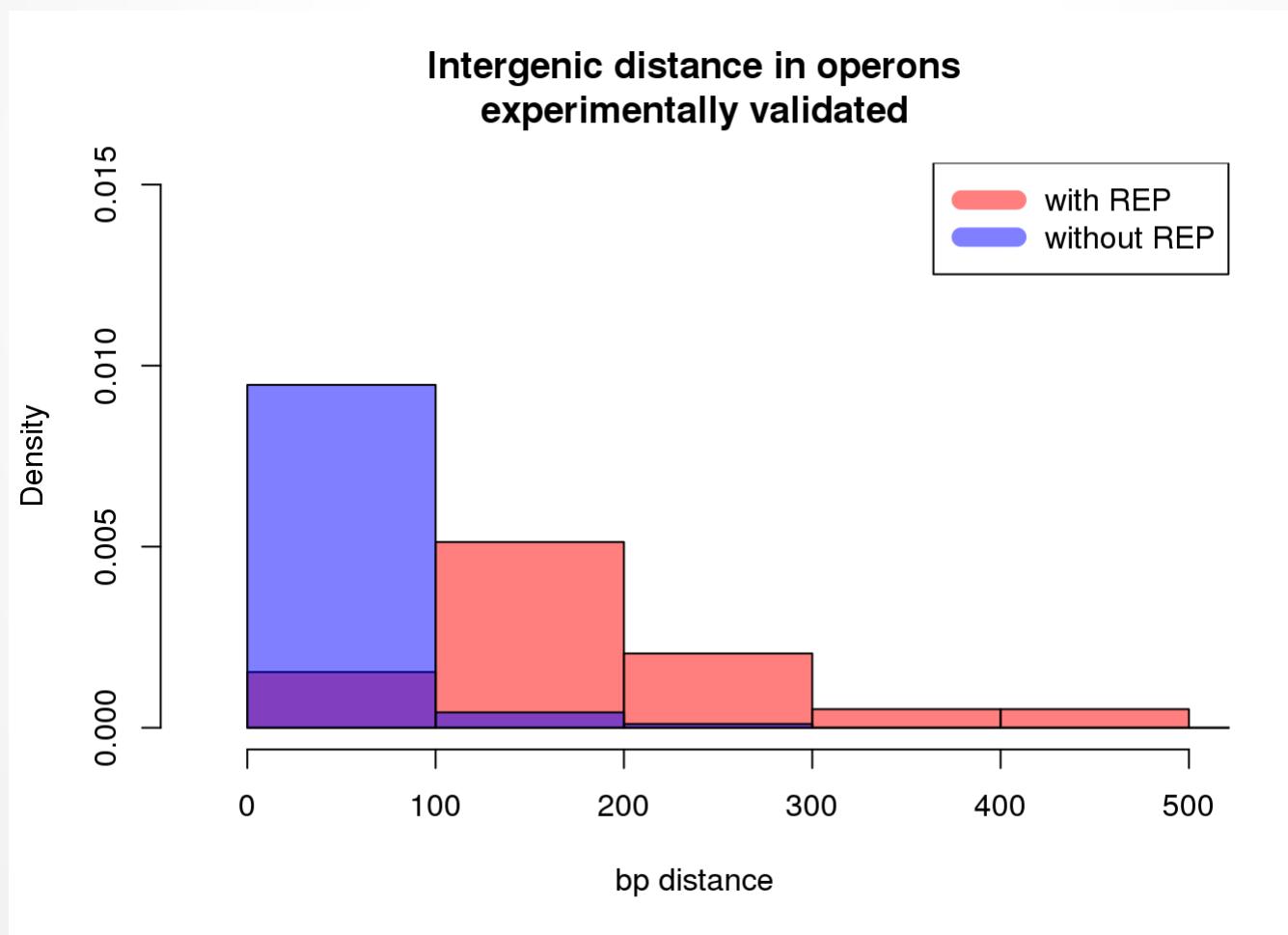
Genomic informations

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 a gene 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
(Bioconductor package easyRnaSeq)

Differential expression in operon with REP

- Reads count table for expressed features
(Bioconductor package easyRnaSeq)
- Correct reads count with RPKM (Reads Per Kilobase Per Millions mapped reads) :

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

Differential expression in operon with REP

- Reads count table for expressed features
(Bioconductor package easyRnaSeq)
- Correct reads count with RPKM (Reads Per Kilobase Per Millions mapped reads) :

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

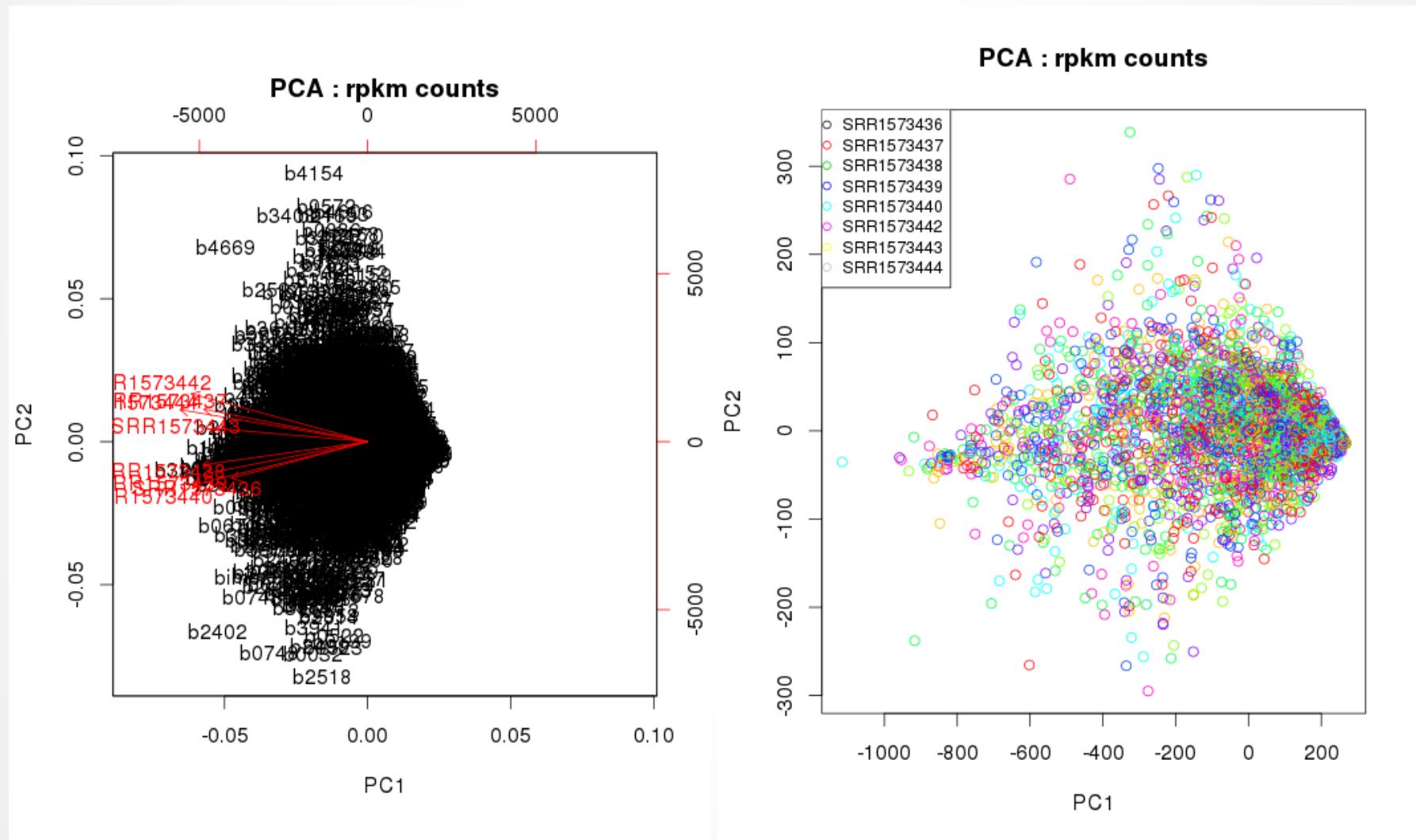
Molar concentration of a transcript :

- 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 on 8 ALE replicates :



Differential expression in operon with REP

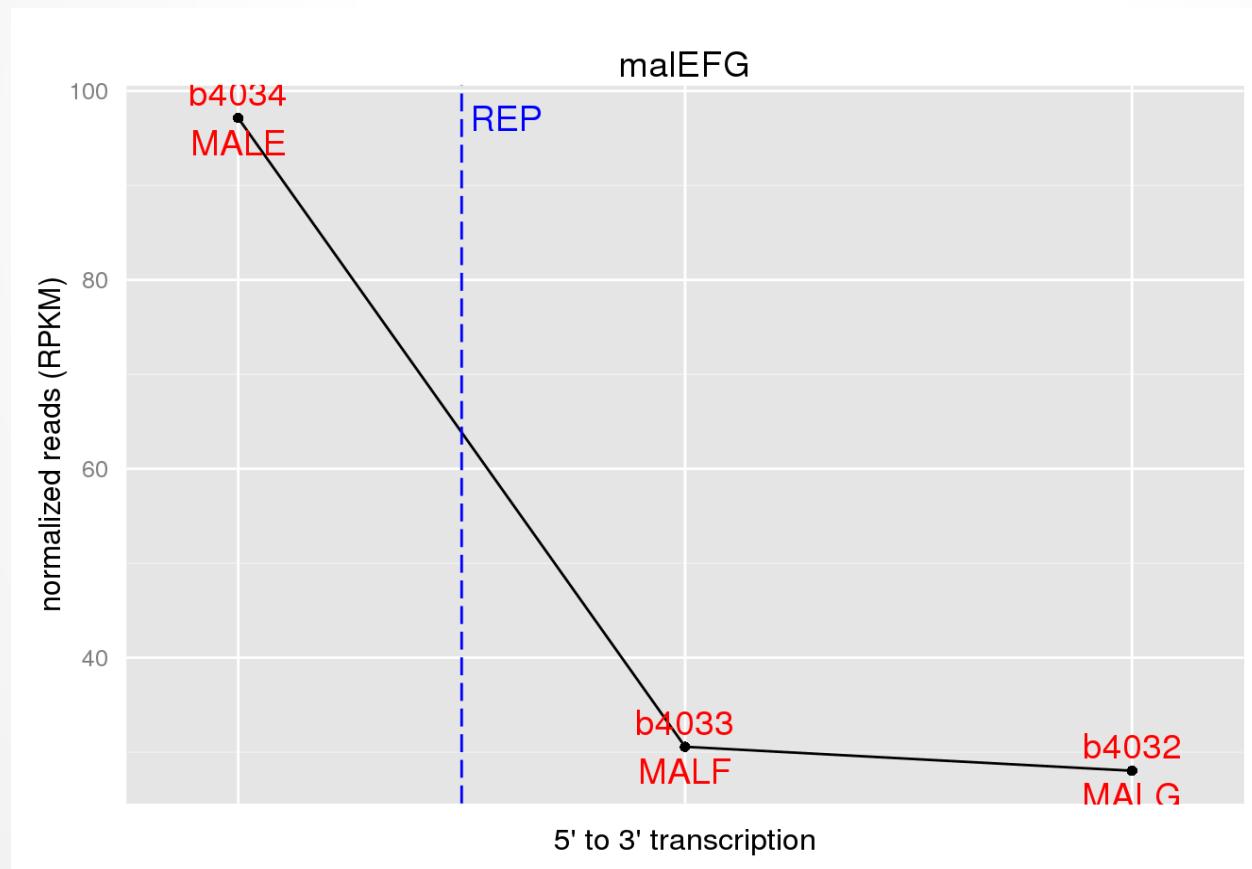
For genes surrounding a REP, Wilcoxon test to detect DE :

- 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 :

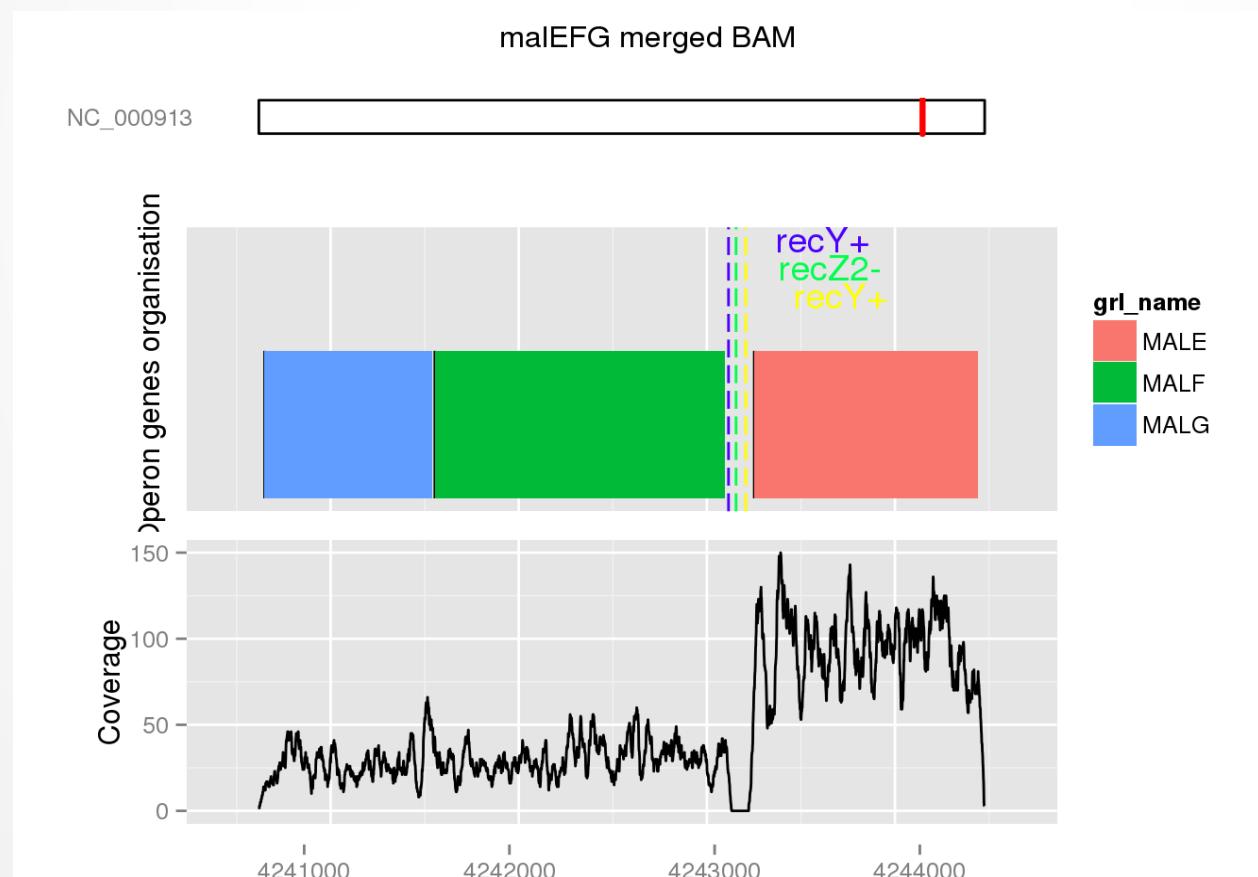
- P-value : < 0.01
- RPKM counts > 10 for at least one gene



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 (abcDB source) :

Function	Number of operons	5' to 3' 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	↙
Glyoxylate bypass isocitrate dehydrogenase kinase/phosphatase	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) :

DE for genes surrounding a REP in ABC transporter :

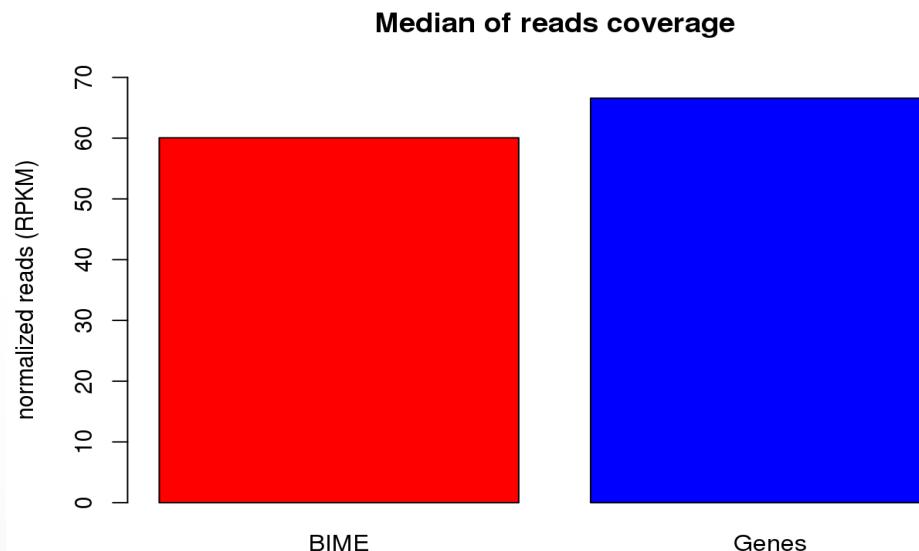
Number of ABC transporter operons	5' to 3' DE change feature	More expressed gene function
7	↙	Solute Binding Protein
2	↗	Nucleotide Binding Domain

Search for transcription change events on BIME

- GFF modifications to convert BIME as mRNA.
- Extraction of genomic positions for genes bounded regions with BIME (**Bedtools**) called BIME regions :
 $\{Gene\} \quad \{IntraGenic-BIME-IntraGenic\} \dots \quad \{Gene\}$
- Coverage per base calculation on these regions (**Bedtools**).
- RPKM correction.

Search for transcription change events on BIME

- GFF modifications to convert BIME as mRNA.
- Extraction of genomic positions for genes bounded regions with BIME (**Bedtools**) called BIME regions :
 $\{Gene\} \quad \{IntraGenic-BIME-IntraGenic\} \dots \{Gene\}$
- Coverage per base calculation on these regions (**Bedtools**).
- RPKM correction.



Search for transcription change events on BiME

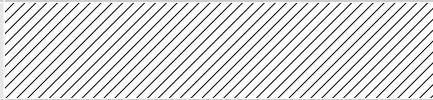
- Looking for correlation between (Fortino et al. 2014) :
____|^__ or __|____ profiles and coverage.
- Profiles simulated with 0 and 1 vectors.
- Sliding window (100 bp) on the BiME region.

Search for transcription change events on BiME

- Looking for correlation between (Fortino et al. 2014) :
____|^__ or __|____ profiles and coverage.
- Profiles simulated with 0 and 1 vectors.
- 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

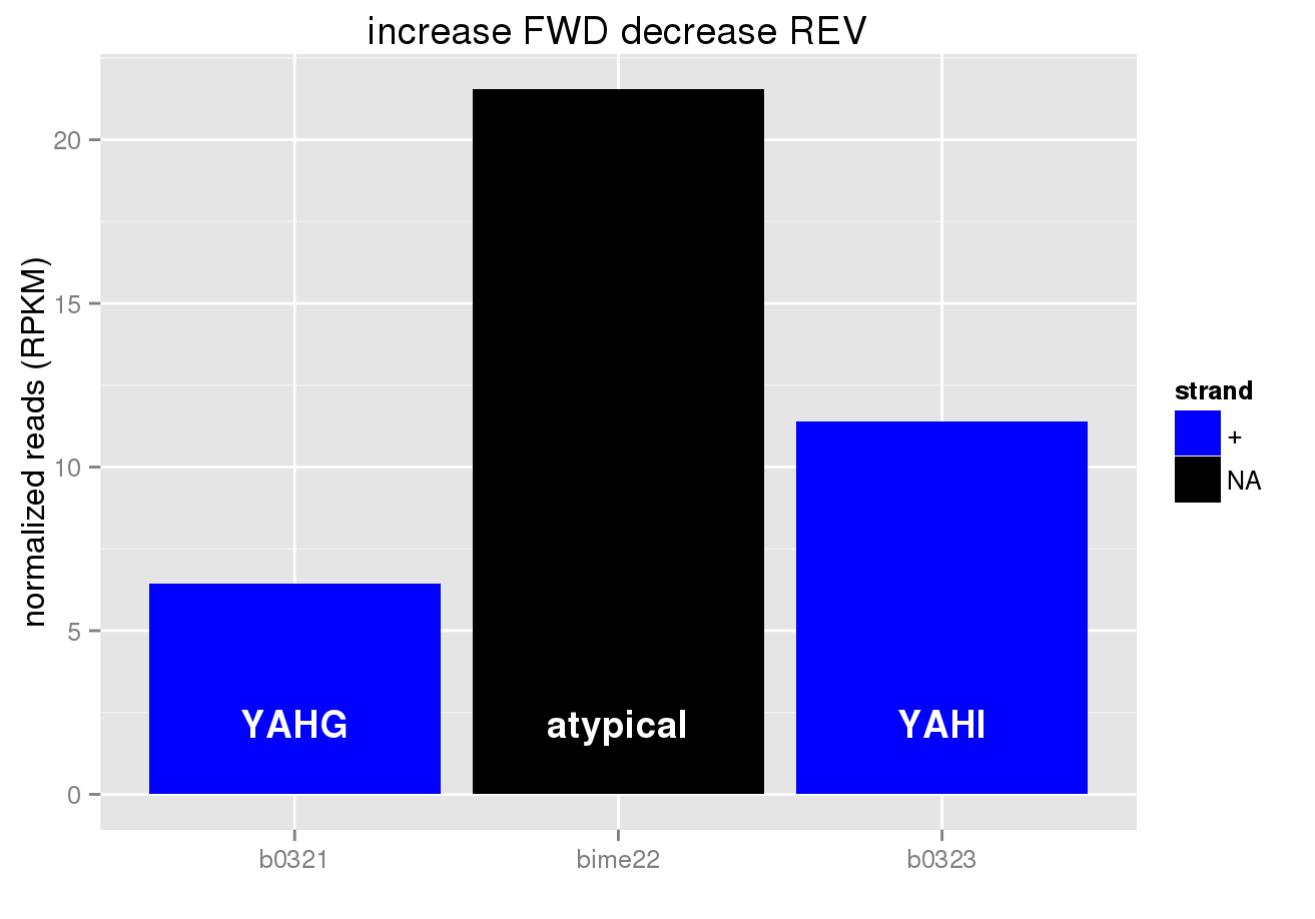
Results overview :

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 :

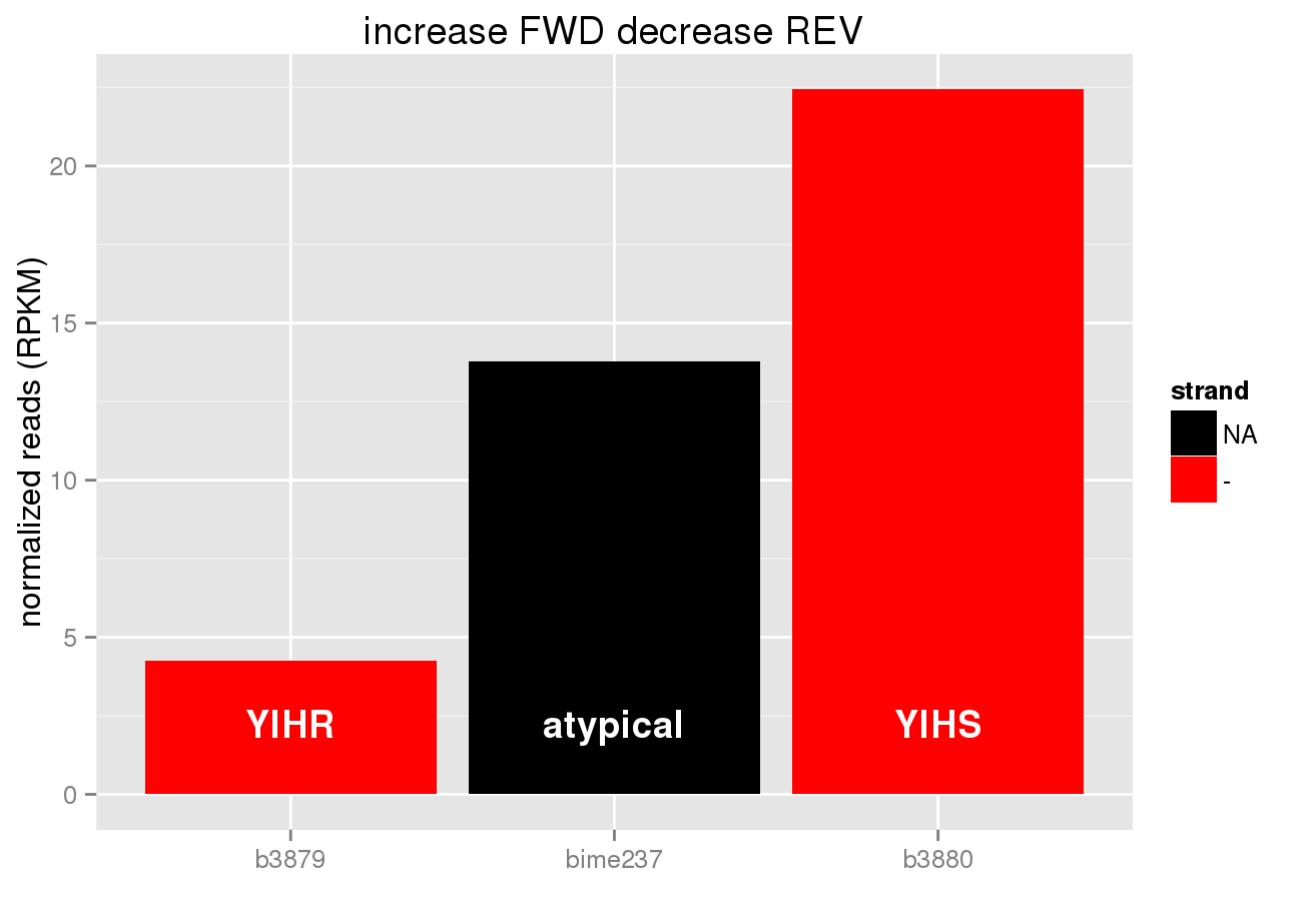
BIME-1	BIME-2	atypical BIME	unique REP
0	0	1	0



On BIME ___|^^\^ : stabilisation

5 cases :

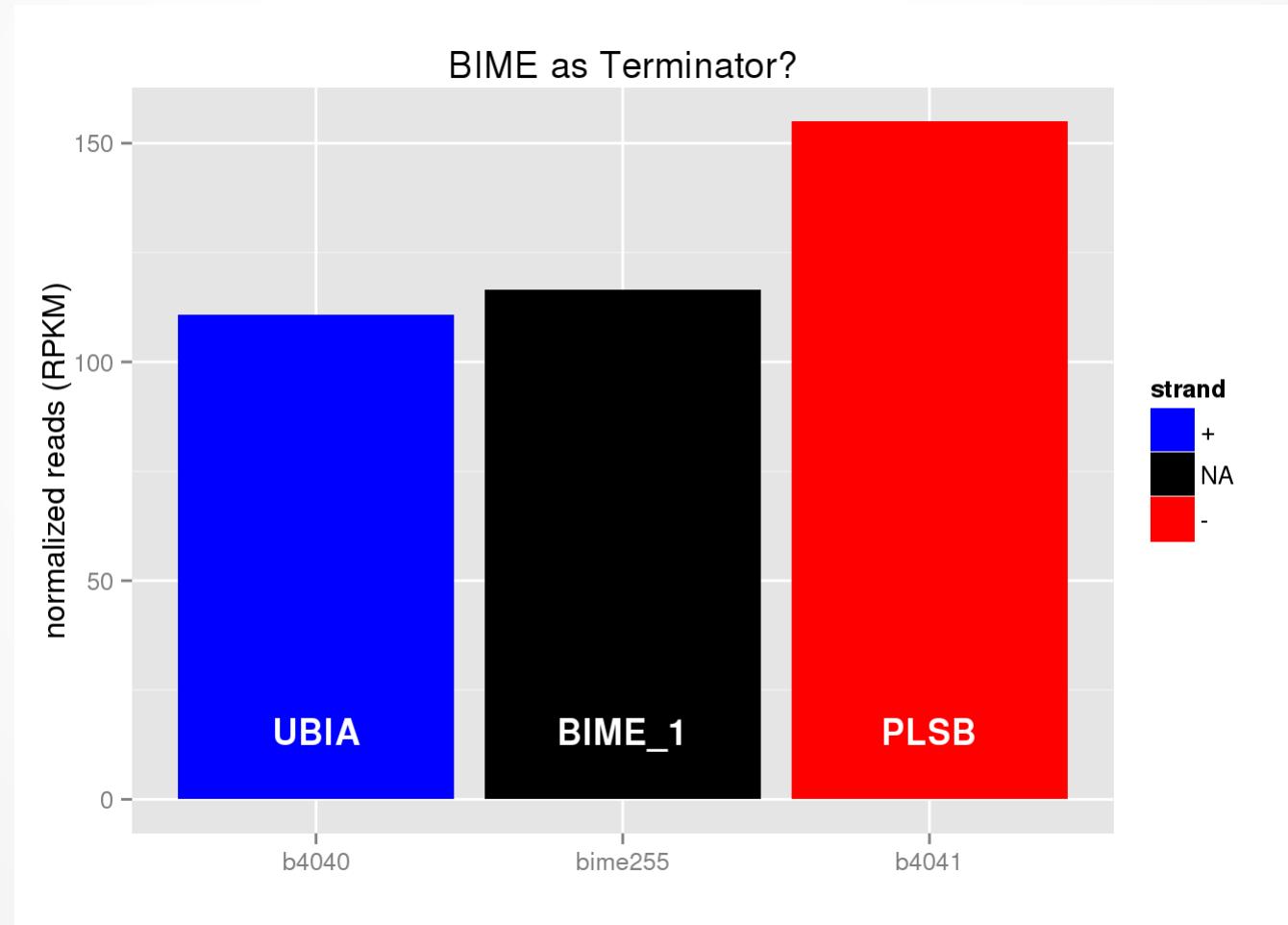
BIME-1	BIME-2	atypical BIME	unique REP
3	0	1	1



On BIME ____|^^\^ : terminator

23 cases :

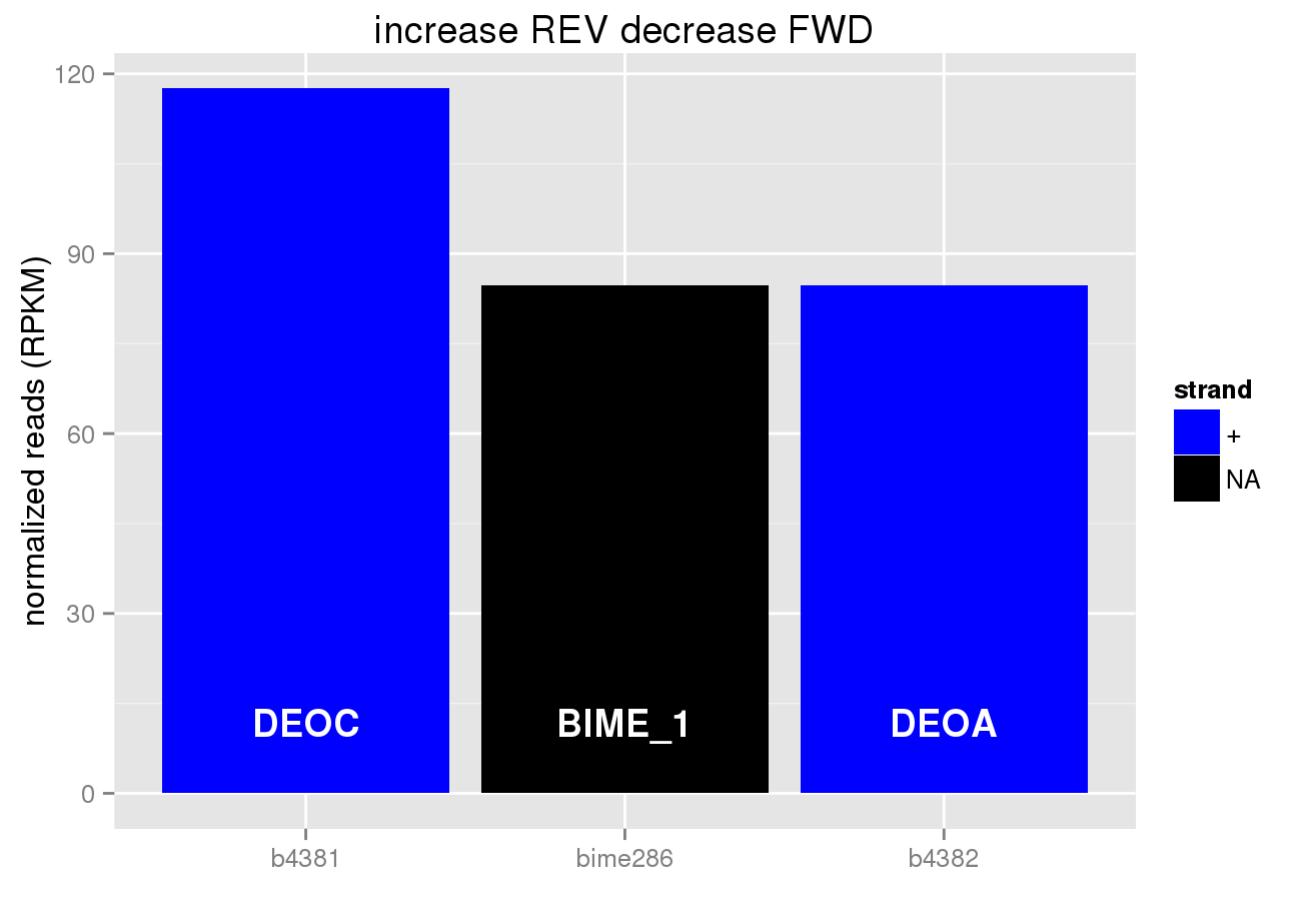
BIME-1	BIME-2	atypical BIME	unique REP
11	2	3	7



On BIME ^^^|_____ : stabilisation

9 cases :

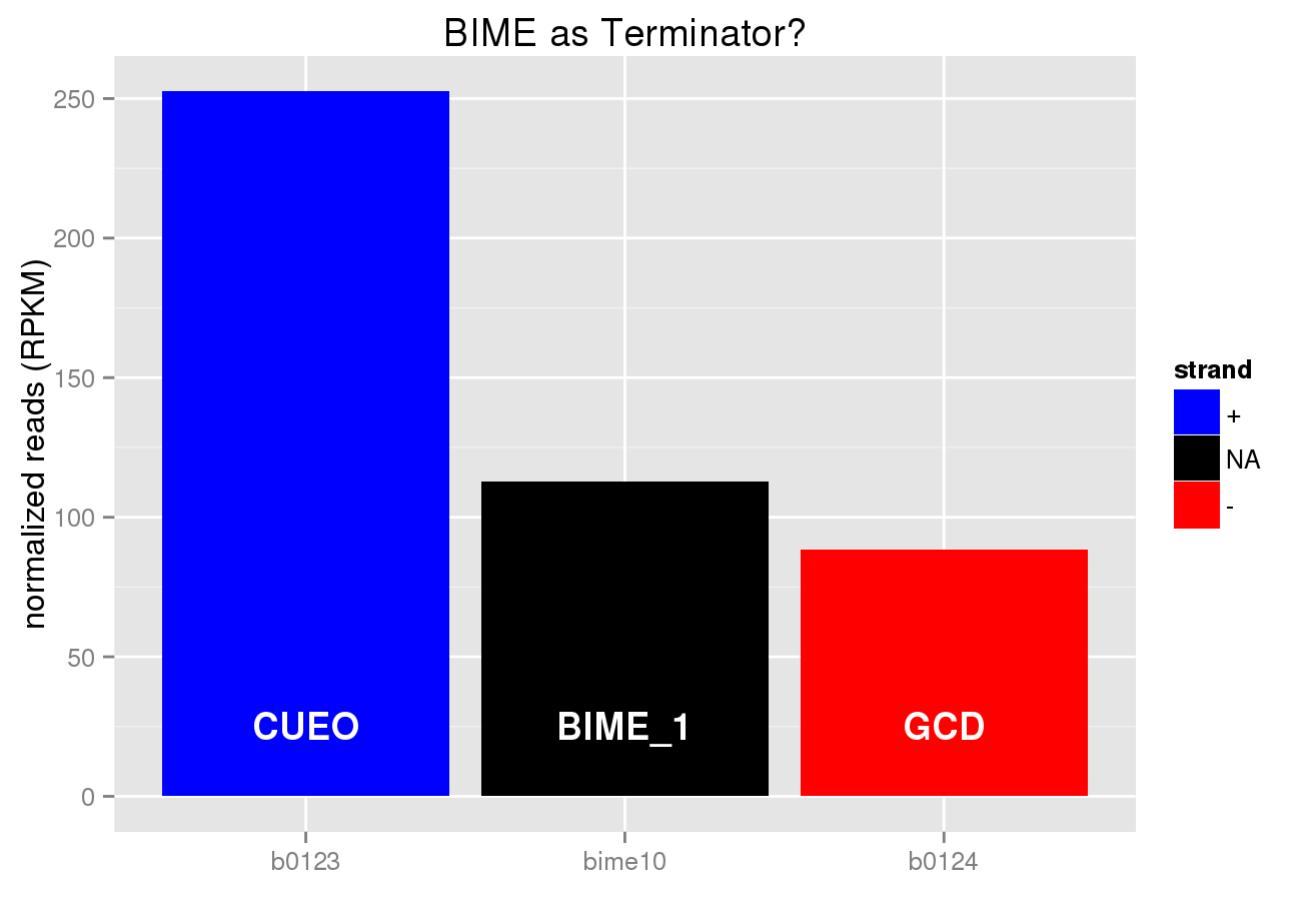
BIME-1	BIME-2	atypical BIME	unique REP
5	2	1	1



On BIME ^^^|____ : terminator

34 cases :

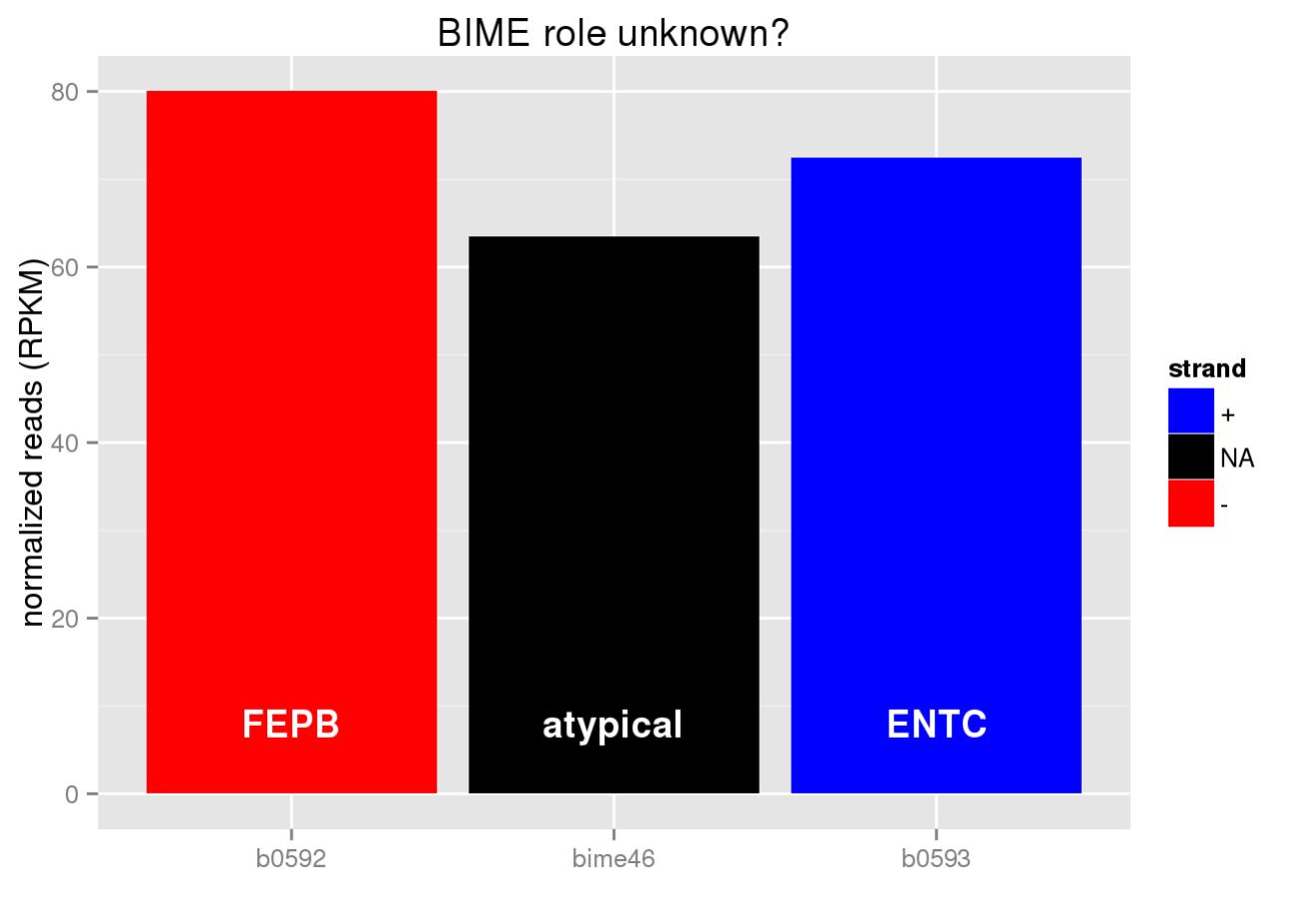
BIME-1	BIME-2	atypical BIME	unique REP
10	8	3	13



On BIME ^^^|____ : enhancing

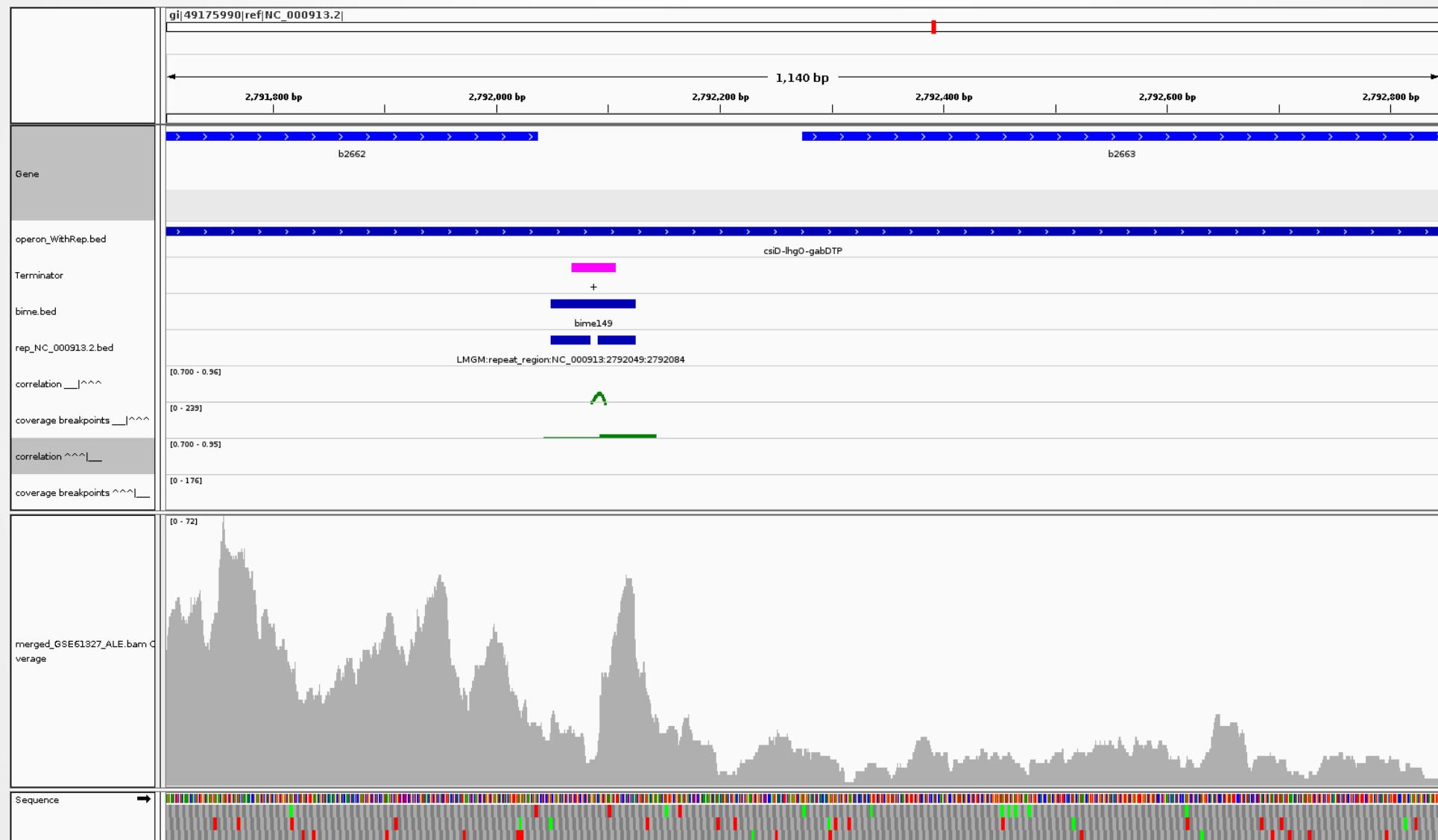
1 case :

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



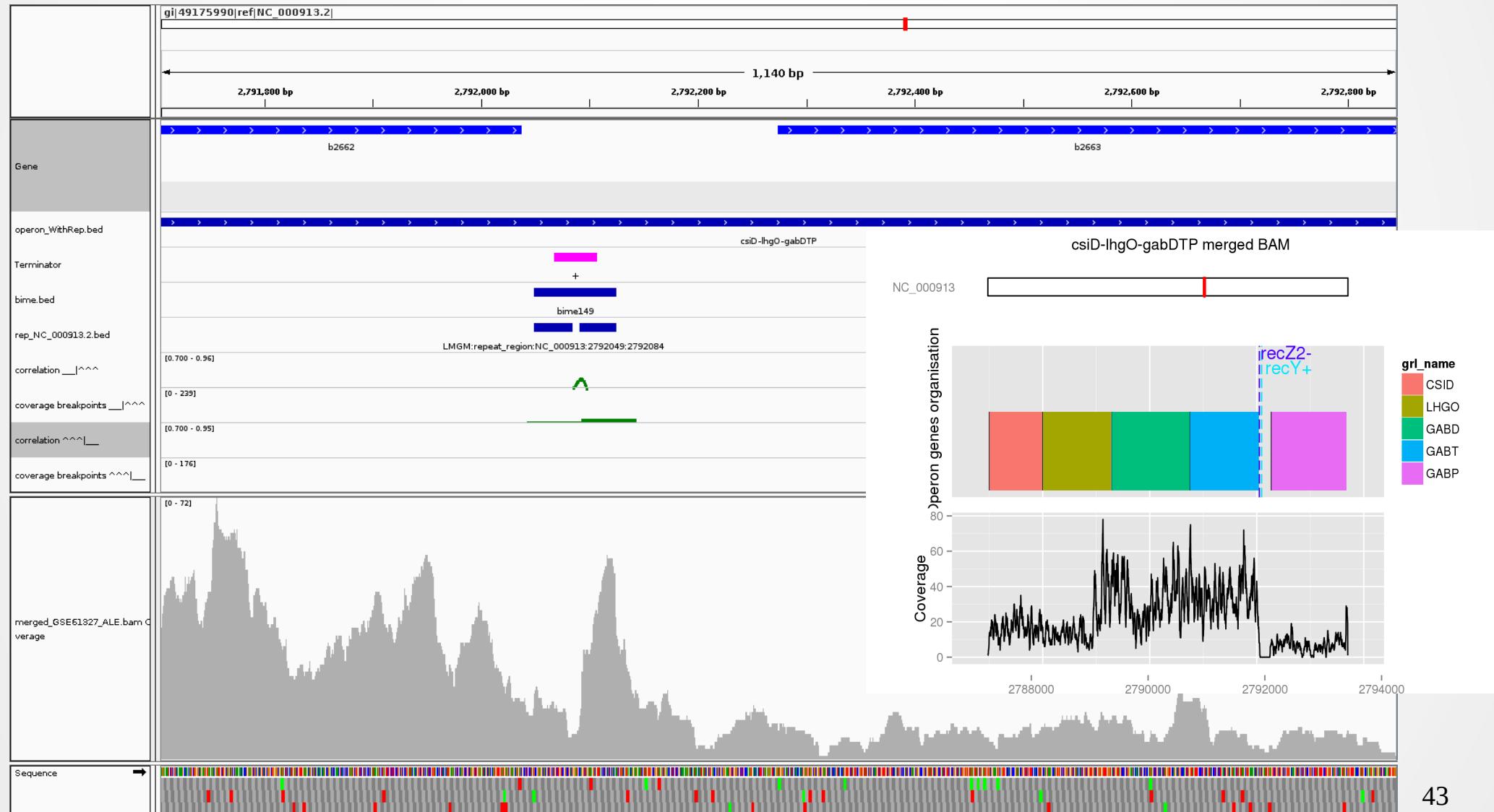
Links with REP results

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



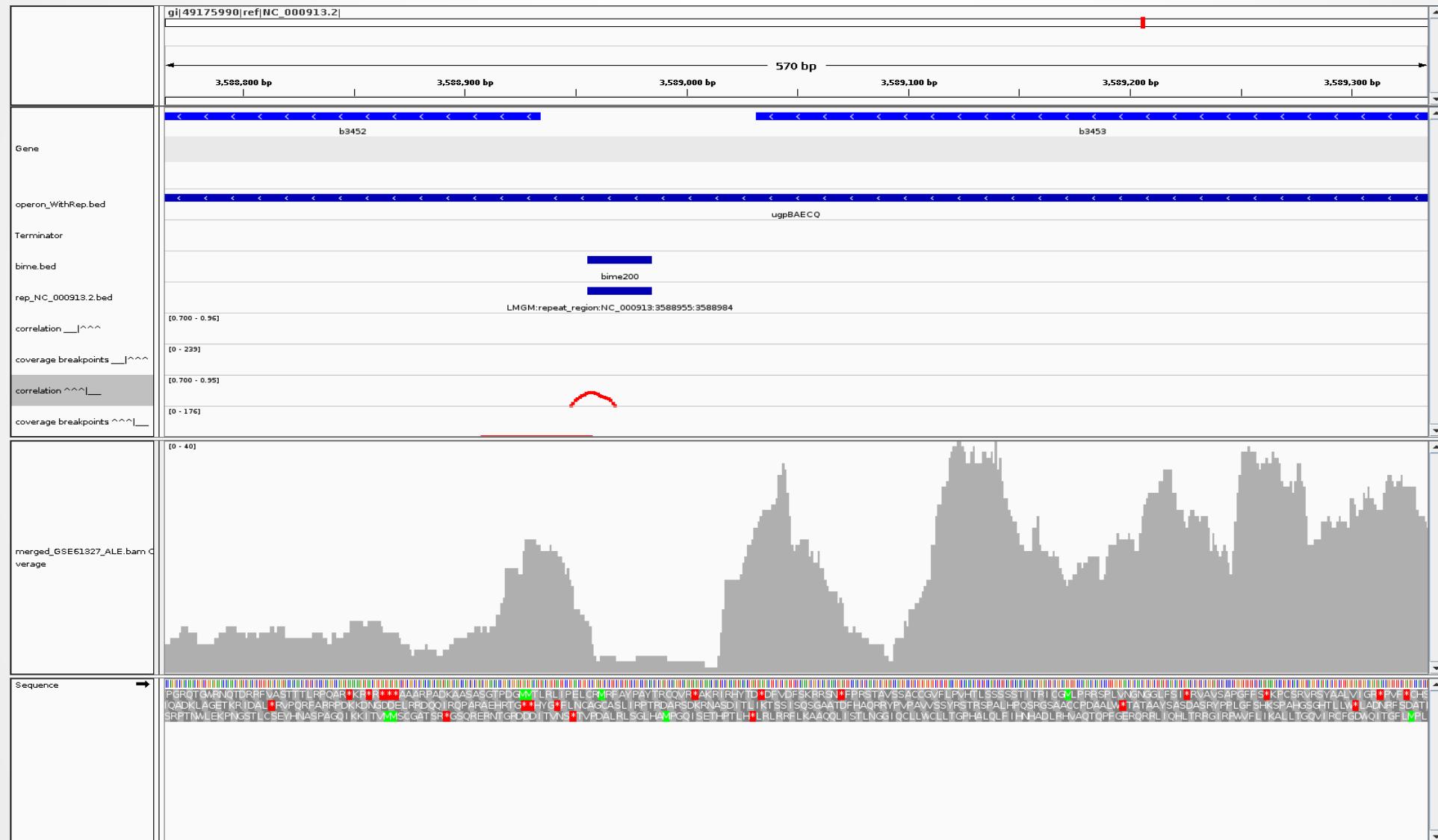
Links with REP results

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



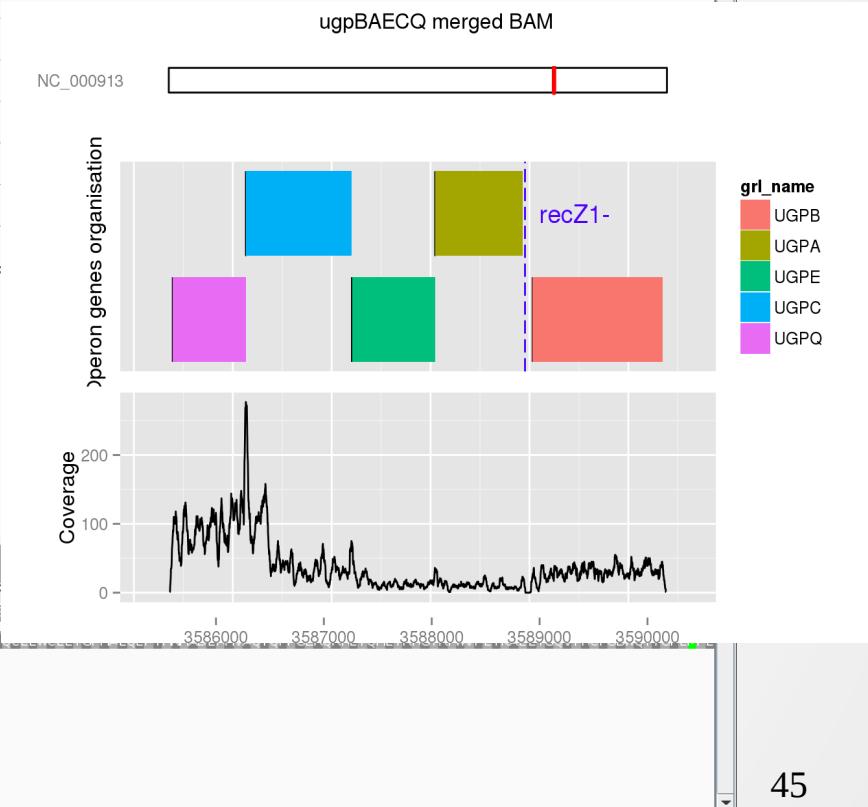
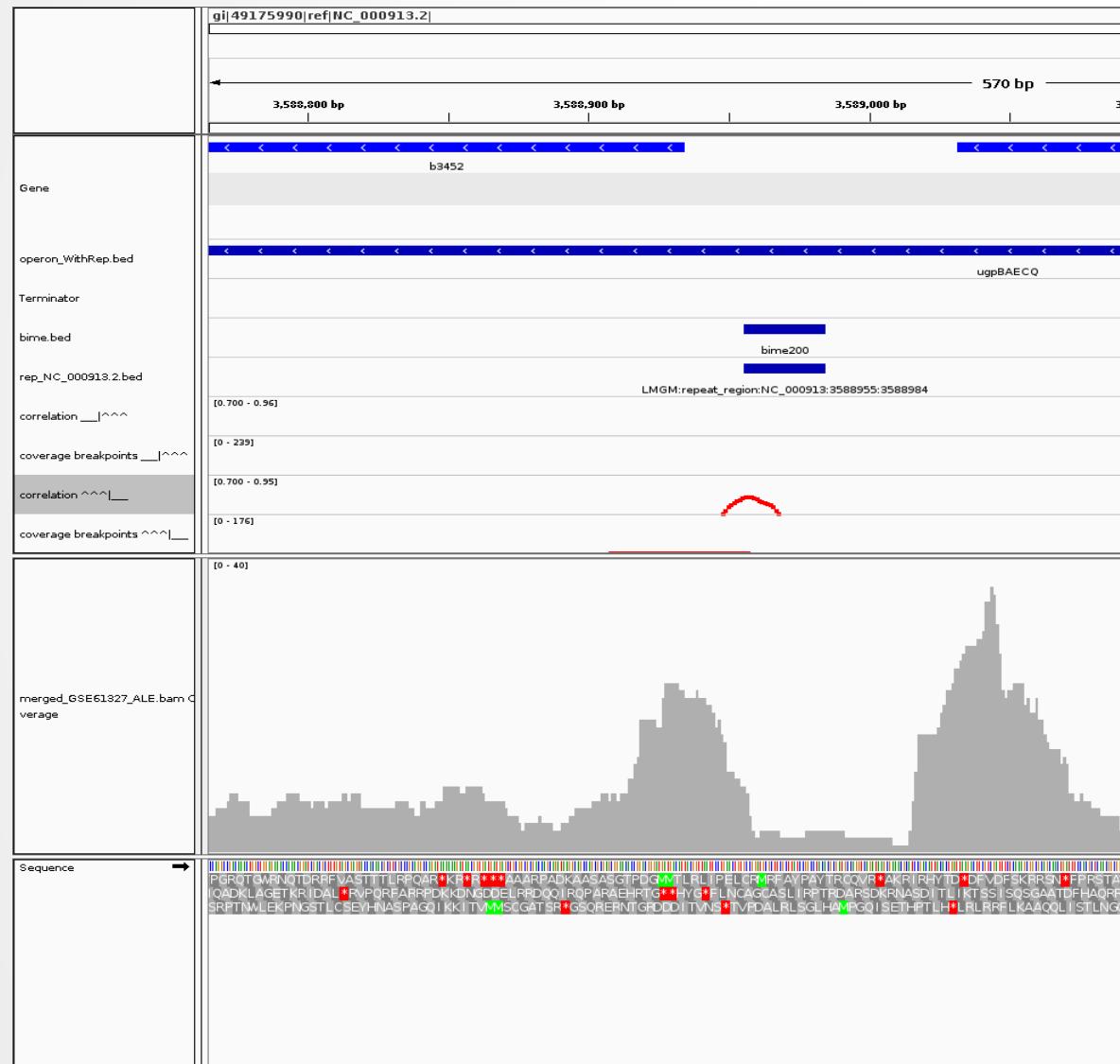
Links with REP results

Operon upgBAECQ (ABC transporter)



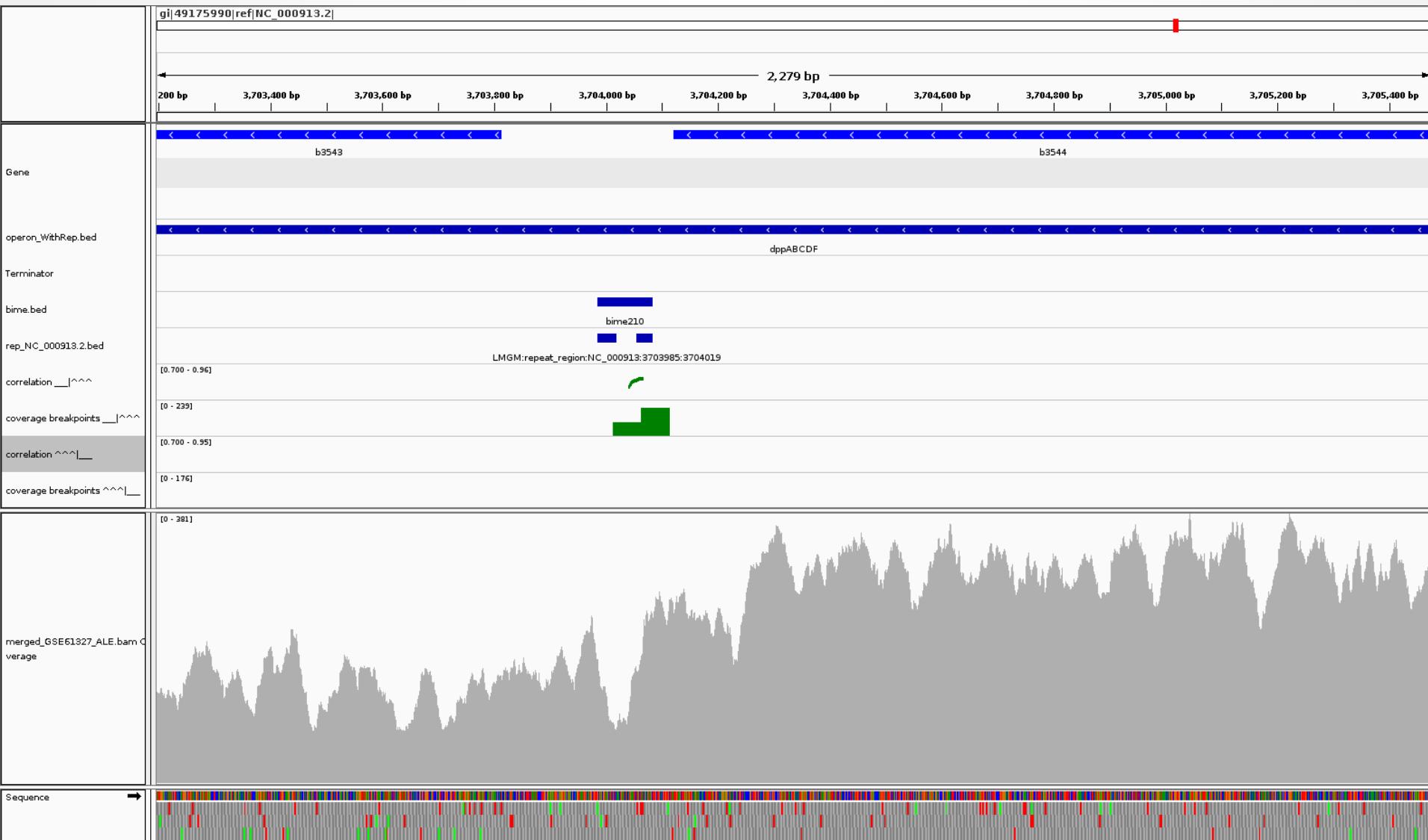
Links with REP results

Operon upgBAECQ (ABC transporter)



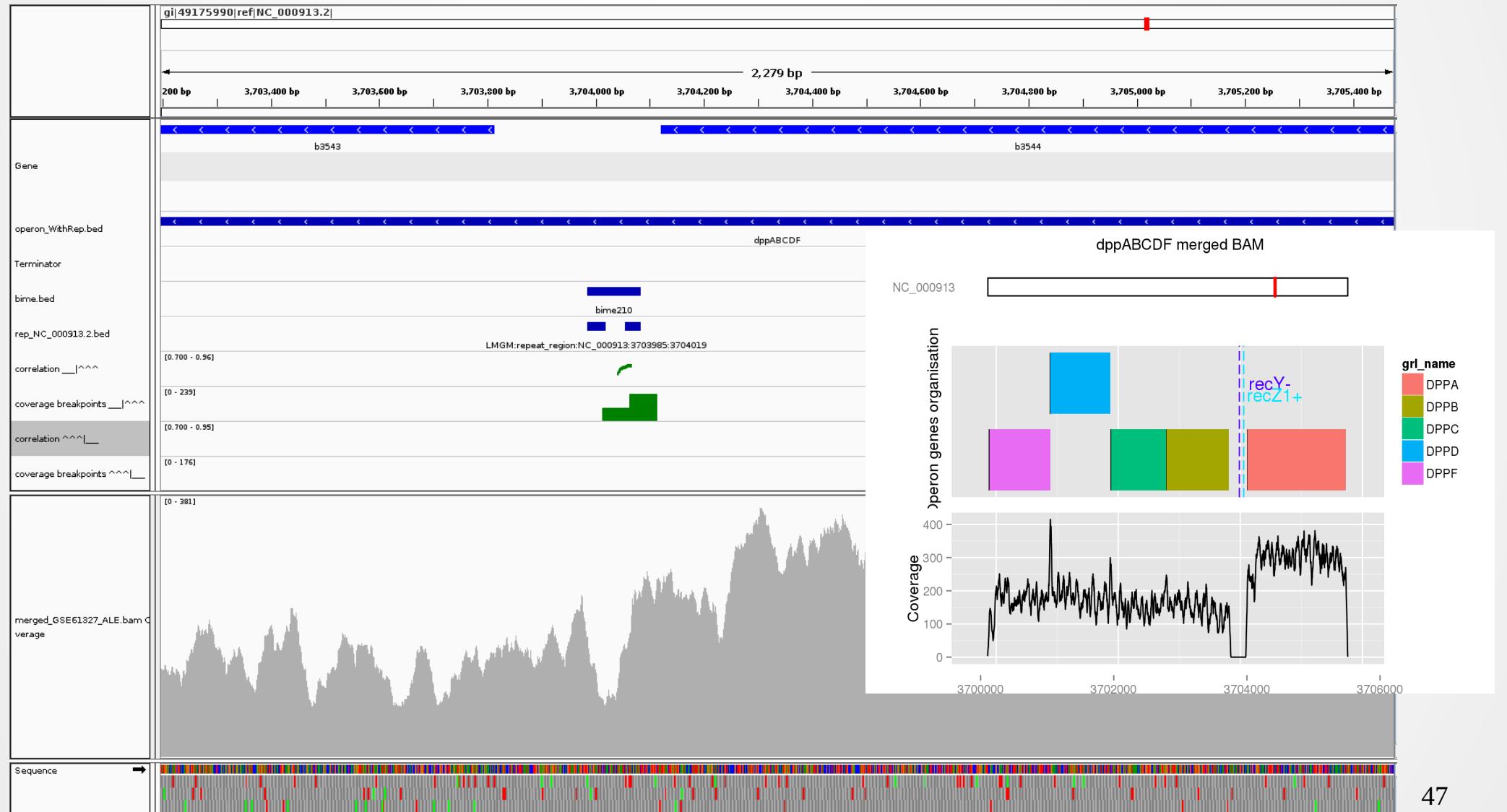
Links with REP results

Operon dppABCDF (ABC transporter)



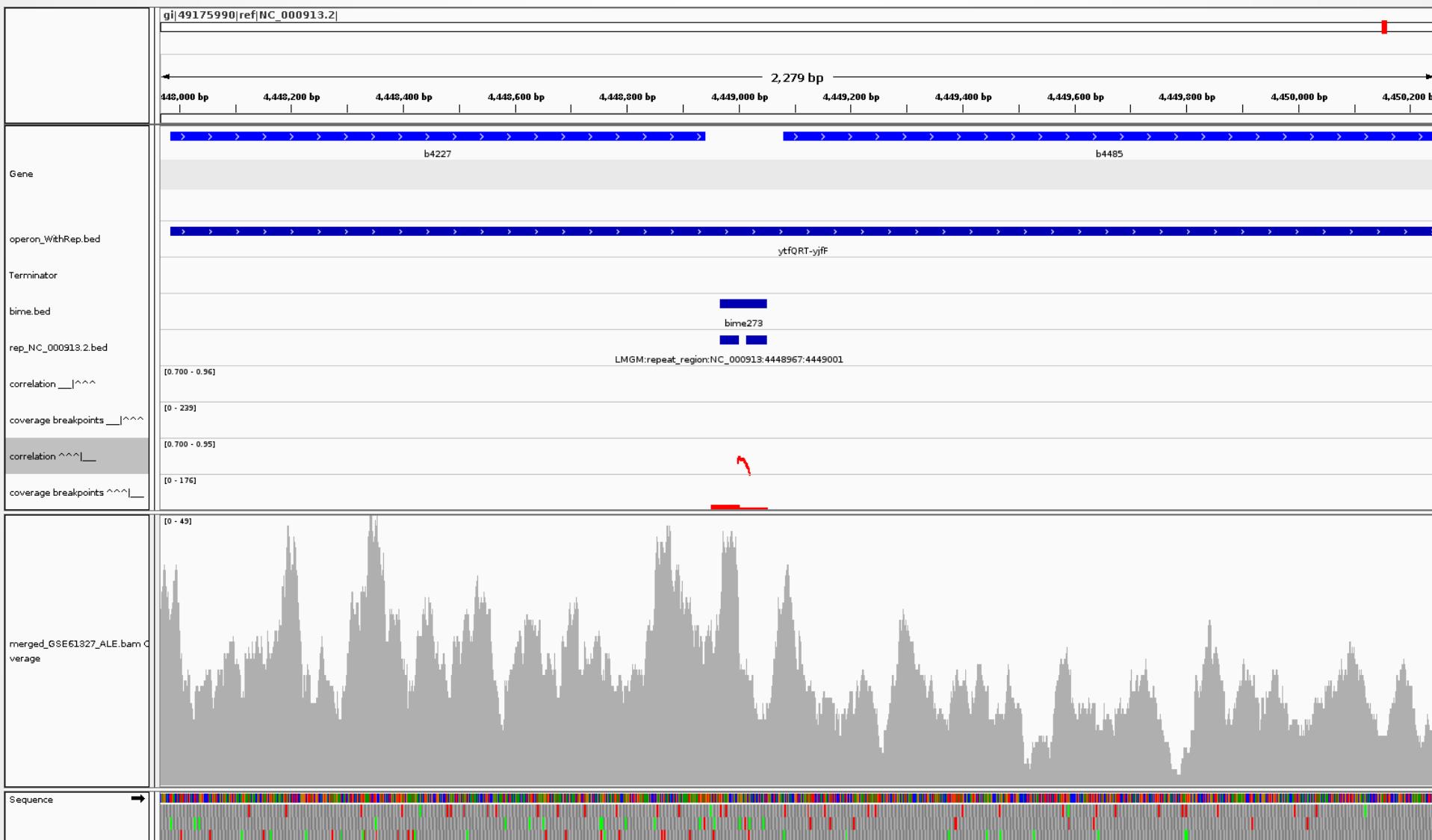
Links with REP results

Operon dppABCDF (ABC transporter)



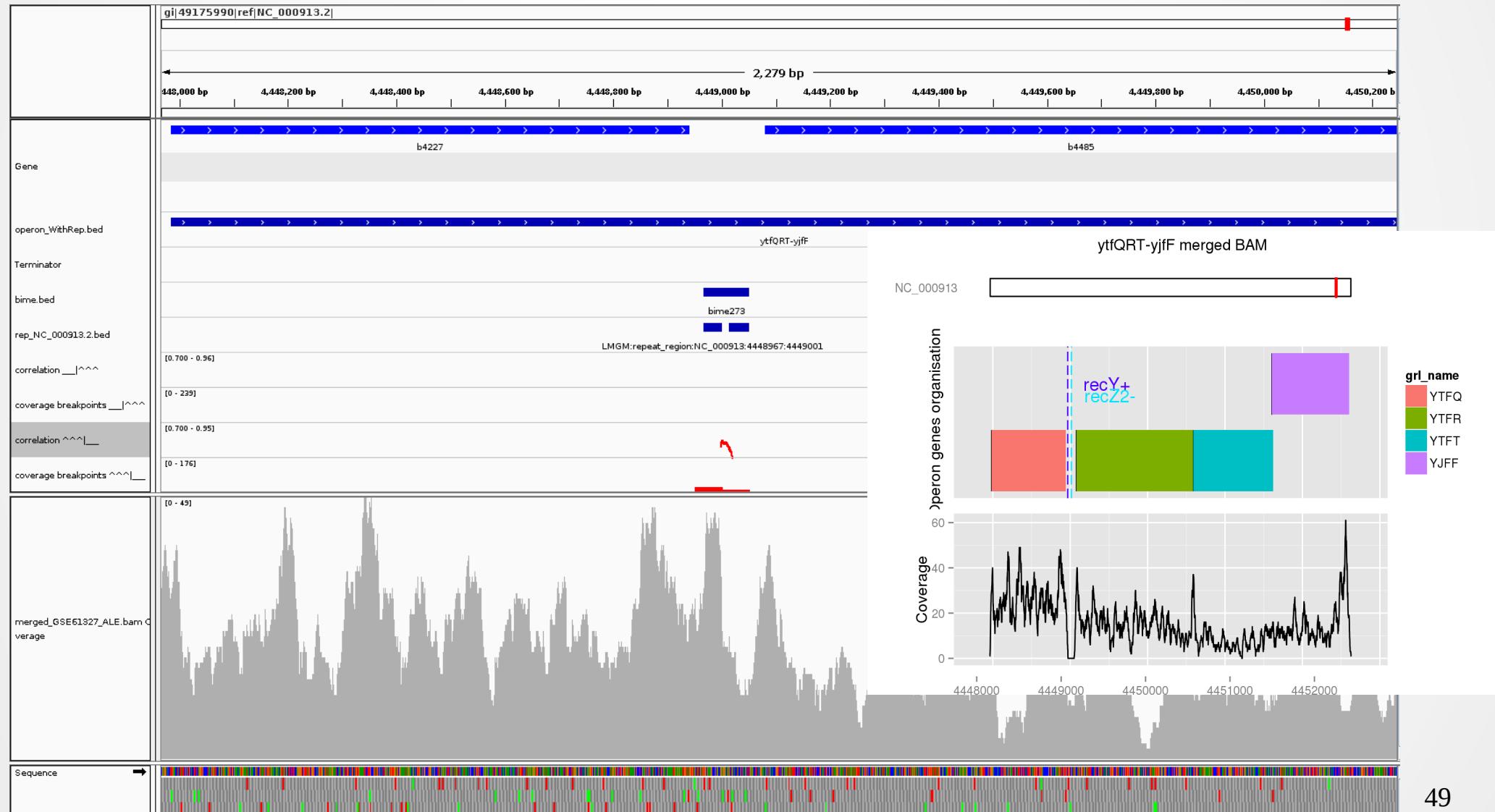
Links with REP results

Operon *ytfQRT-yjfF* (ABC transporter)



Links with REP results

Operon *ytfQRT-yjfF* (ABC transporter)



Perspectives

- Improve transcription change discovery ?
- Alignment of REP or BIME expression profiles to look for a pattern.
- Kolmogorov test on expression profiles for BIME classes to search differences.
- RIP-Seq or RIP-Chip analysis to check interactions RNA-proteins (data?, methodology?)