ChIP-Seq Transcription Factors Analysis

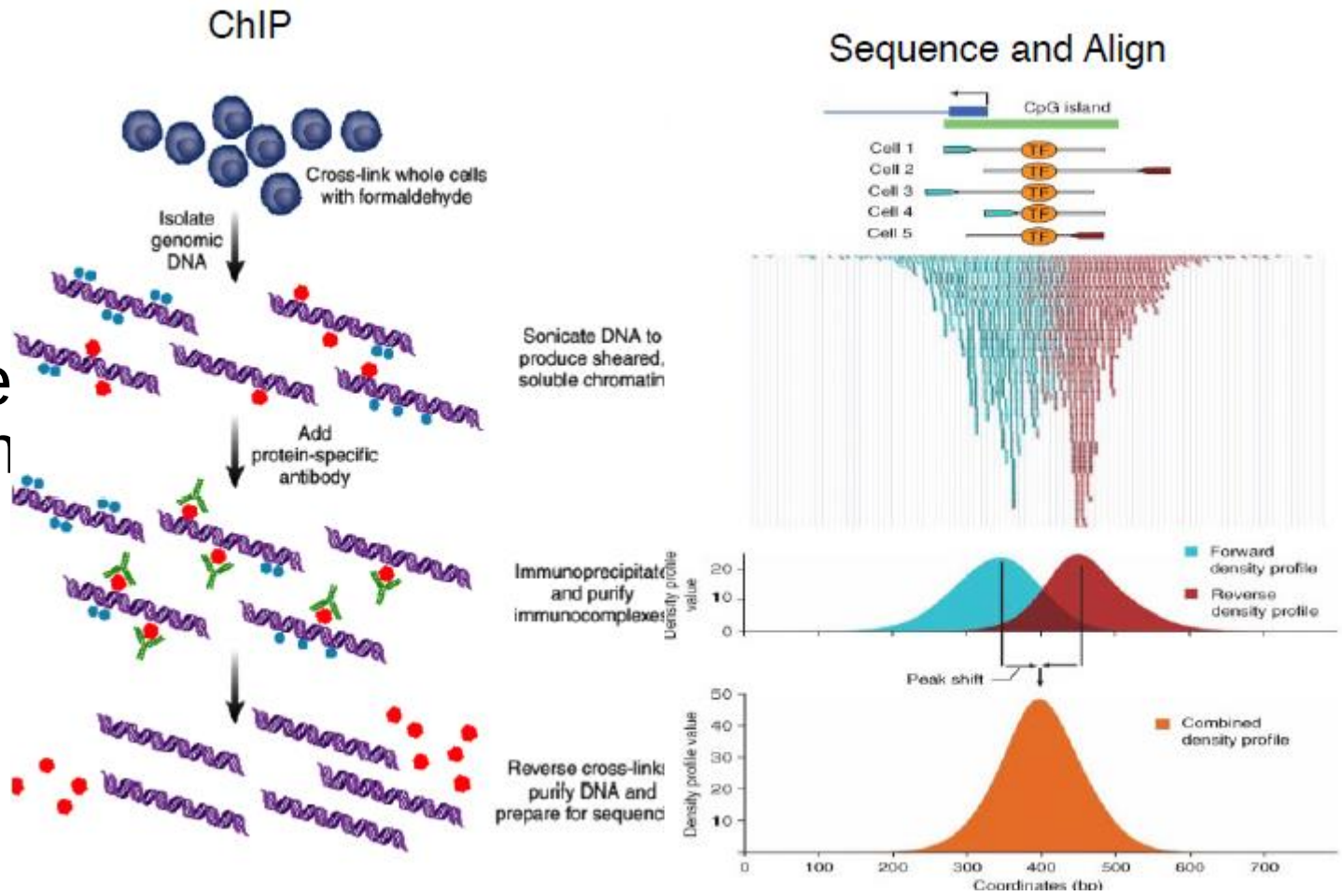
Team D (yciT, perR, allR)

Abdurrahman Addokhi, Abdul Bhuiya, Matthew Ning,

Antonio Ortega, Divya Sha

ChIP-Seq overview

- Identifies possible binding sites for proteins
- Very high throughput
- FLAG-labeled proteins are induced in target organism
- Antibodies used to mark bound DNA
- Sequencing is performed on precipitated DNA fragments



Valouev et al (2008) Nature Methods

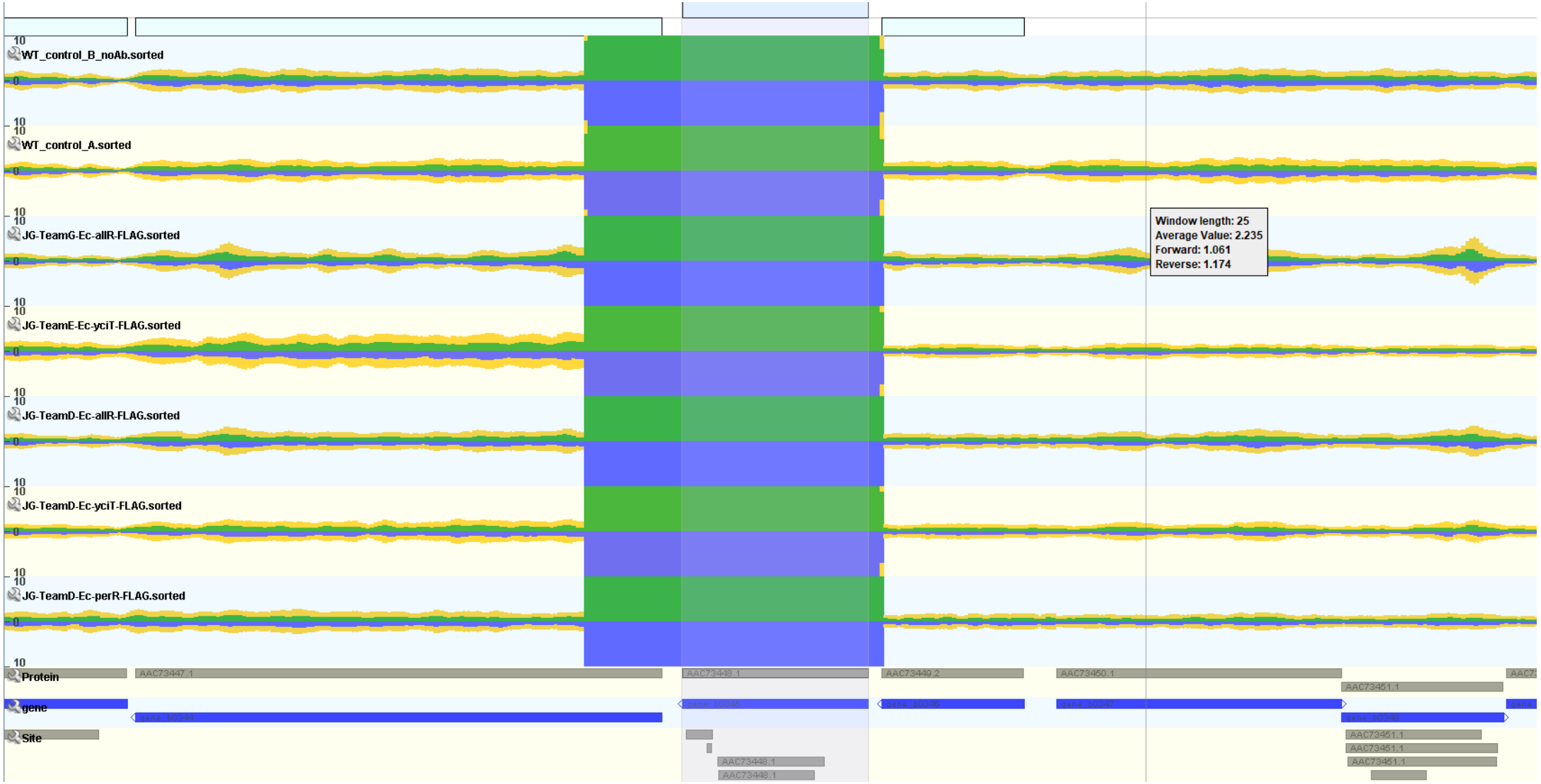
Lab procedure

1. Inoculate strain
2. Re-inoculate to large volume
3. Induce Transcription Factor
4. Crosslink and wash
5. Lyse and shear
6. Incubate with primary antibody
7. Incubate with secondary antibody
8. Wash
9. Elution of DNA-protein
10. Complexes and reverse crosslink
11. Purify DNA
12. Sequence (off-site)

Contents

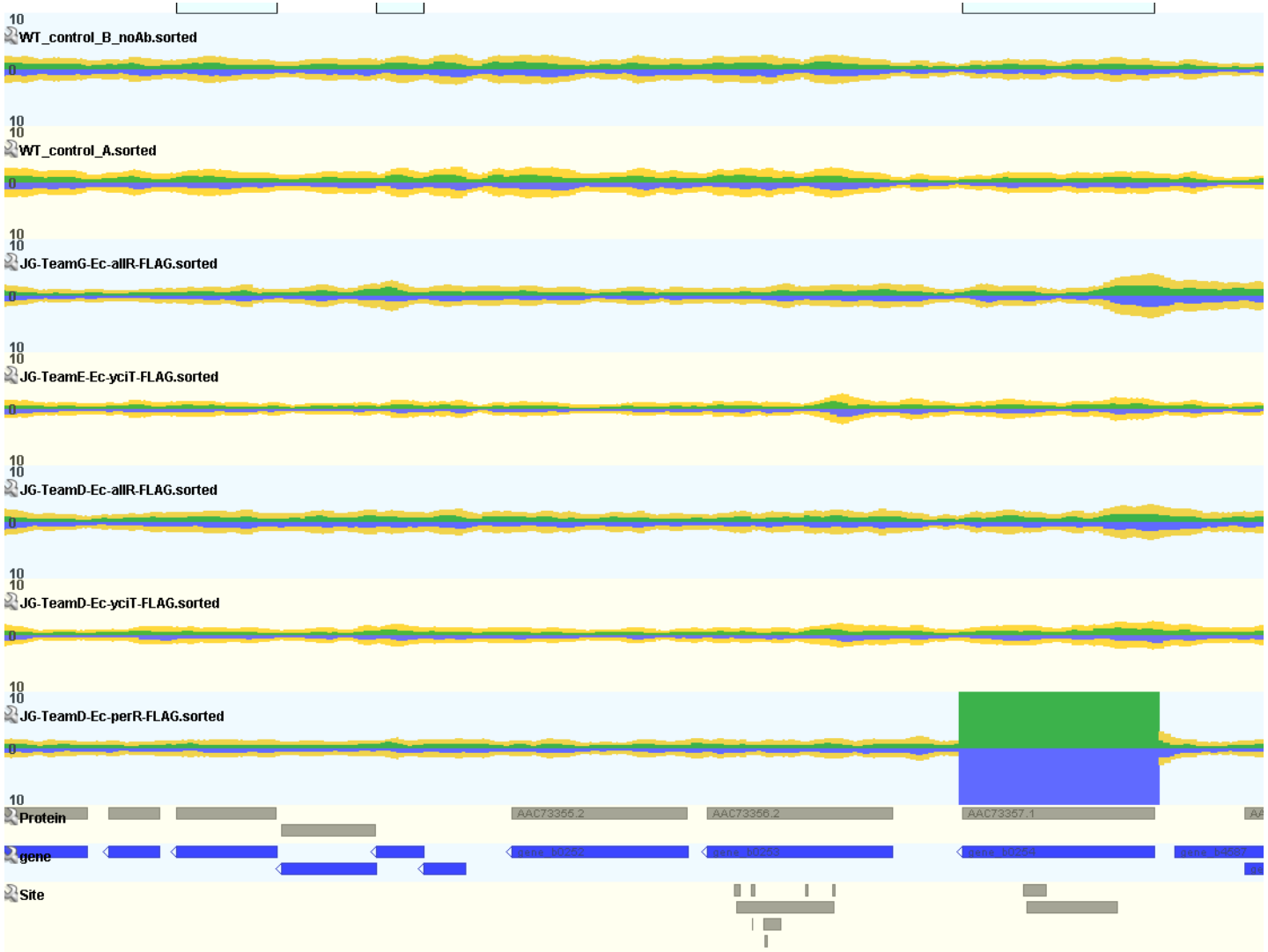
1. Validation
2. Literature review
3. Peak analysis
4. Motif discovery
5. Conclusions

Validation: watermarks (lacI)

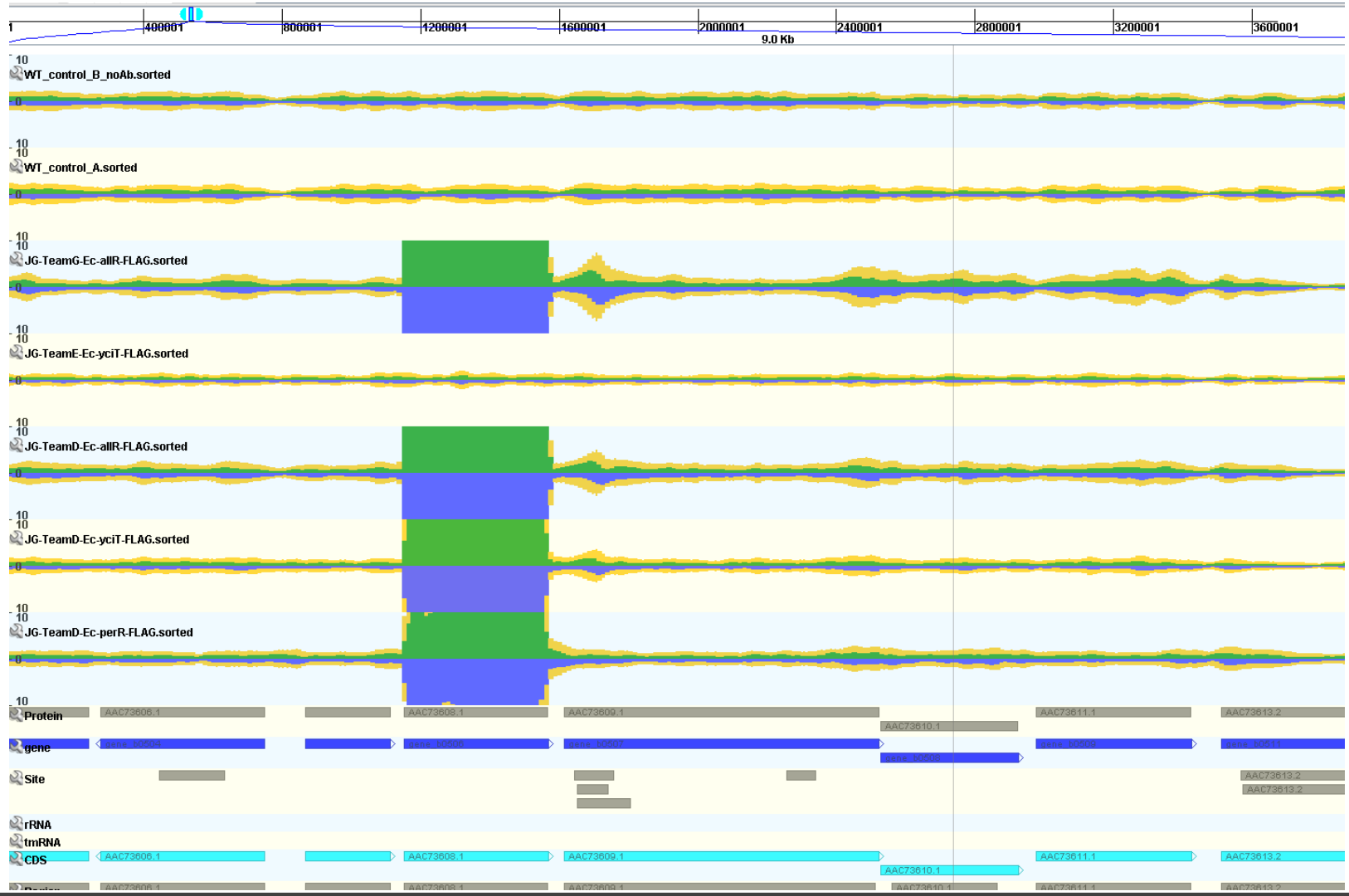


Validation: watermarks (TF)

- Found them for all three TFs but…



Cross contamination with allR



Peak Analysis pipeline

1. Check peaks called by SPAT
2. Start by low p-value, high shift
3. Go to that position in GenomeView
4. Make sure the peak is not in the controls
5. Make sure the peak is in the other team's experiment
6. If TF not allR, check it is not in allR
7. Identify nearest genes
8. For perR, we submitted the genes to DAVID
9. Identify as intragenic or intergenic downstream, intergenic upstream

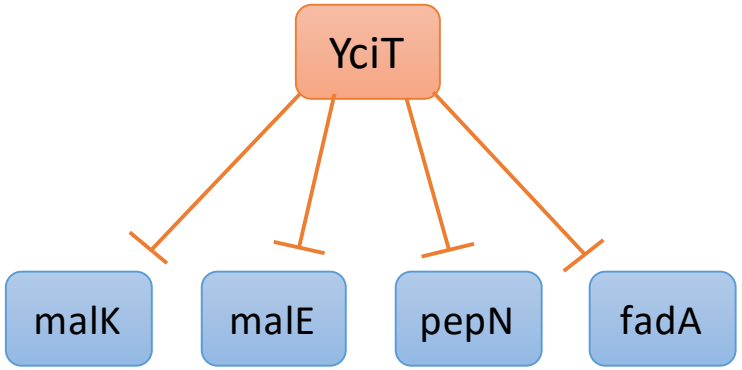
Motif analysis pipeline

- For each transcription factor
- Extract E. Coli sequence from start to end of peak
- Submit to MEME to find motifs
- Searched for motifs around (150 bases) mode
- Searched motifs for the top binding sites (p-values under 0.01)
- Results submitted to TOMTOM to find matching known motifs in E. Coli

yciT

Literature review

- Not well documented
- HTH motif
- Through yciT mutant E. coli, it has been shown that



(perhaps indirectly)

DeoT, a DeoR-type transcriptional regulator of multiple target genes

Maya Elgrably-Weiss*, Eliana Schlosser-Silverman*, Ilan Rosenshine & Shoshy Altuvia

Department of Molecular Genetics and Biotechnology, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

Correspondence: Shoshy Altuvia, Department of Molecular Genetics and Biotechnology, The Hebrew University-Hadassah Medical School, 91120 Jerusalem, Israel. Tel.: +972 2 6757212; fax: +972 2 6757308; e-mail: shoshy@cc.huji.ac.il

Received 19 August 2005; revised 16 October 2005; accepted 16 October 2005.
First published online 15 November 2005.

doi:10.1111/j.1574-6968.2005.00020.x

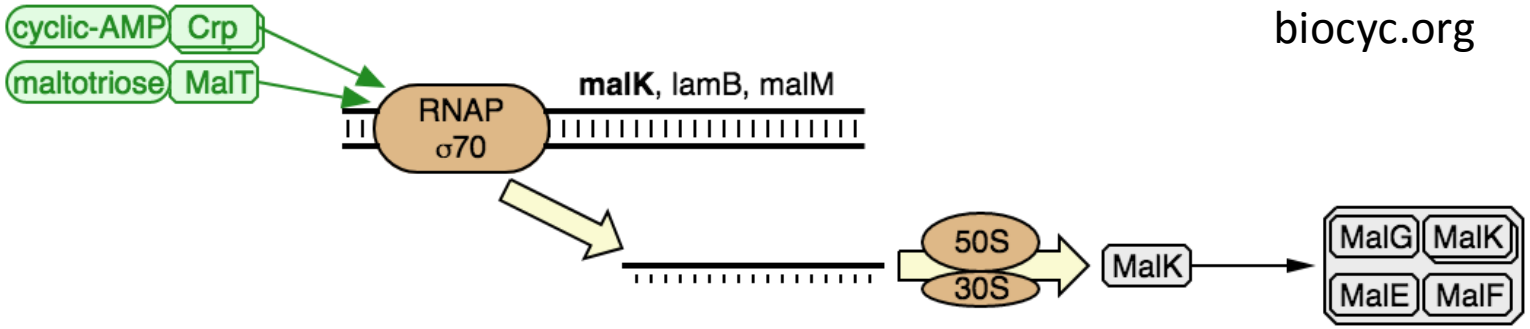
Editor: Marco Soria

Keywords
yciT; DeoR; transcription control; operon; repression; global regulator.

Abstract

In this study, we investigated the genetic organization and function of *Escherichia coli yciT*, a gene predicted by computational methods to belong to the DeoR-type family of transcriptional regulators. We show that transcription of *yciT* (here denoted *deoT* for *deoR*-Type) initiates from a promoter located upstream of a putative open reading frame (denoted *deoL* for *deoT* Leader). We also show that DeoT acts as a global regulator, repressing the expression of a number of genes involved in a variety of metabolic pathways including transport of maltose, fatty acid β-oxidation and peptide degradation.

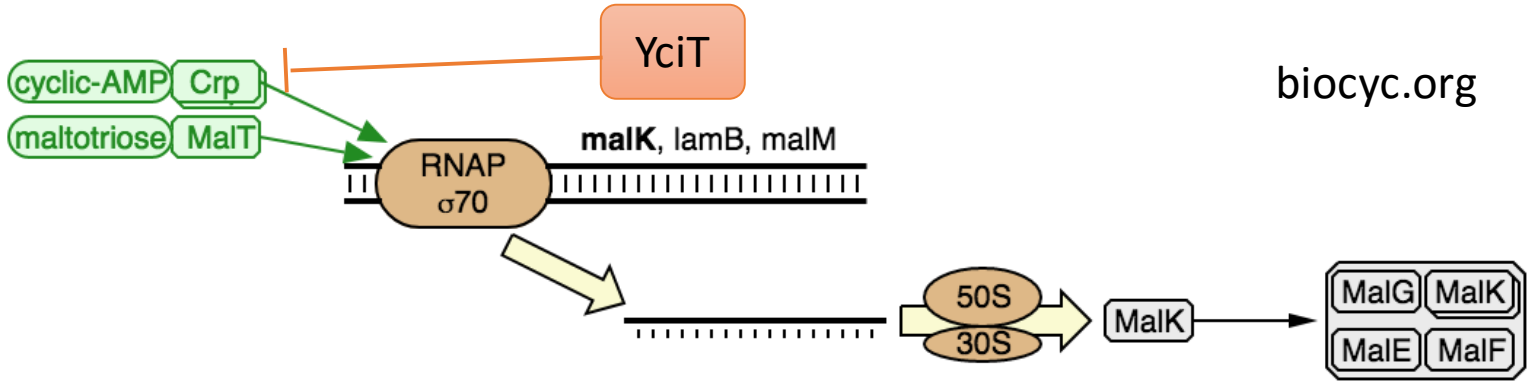
malK



biocyc.org

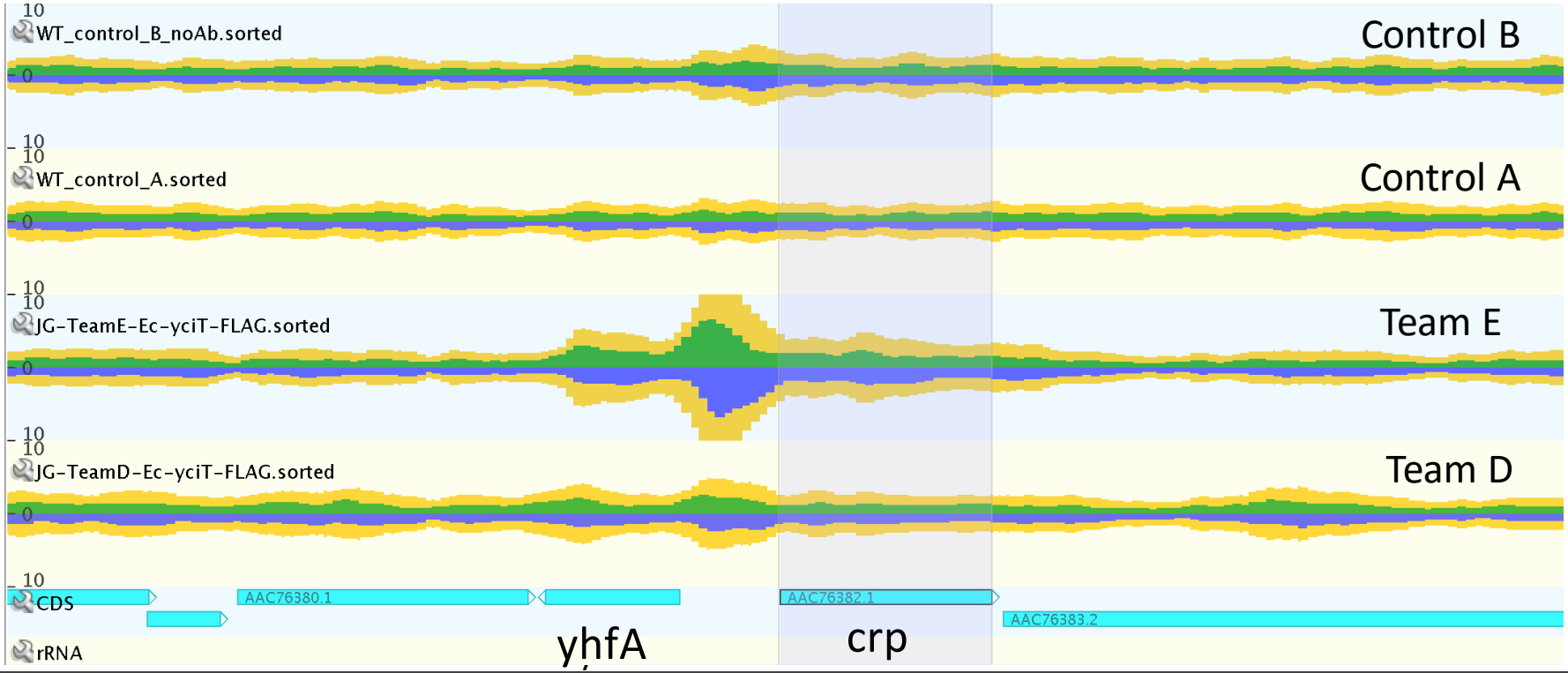
- No significant peaks around malK.

malK



biocyc.org

- No significant peaks around malK.



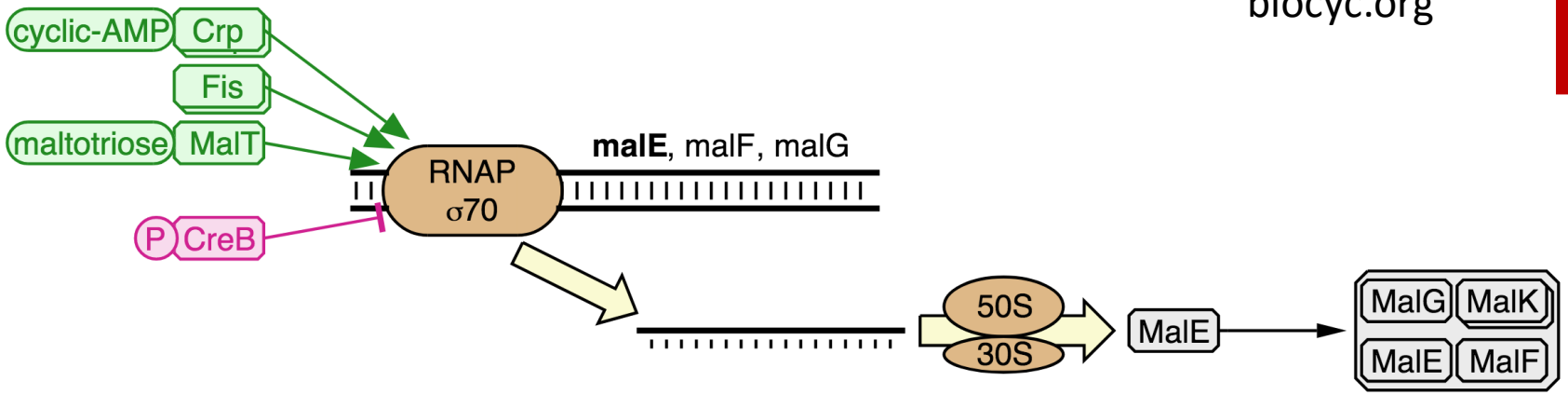
crp

A 2.6 fold peak upstream

malT

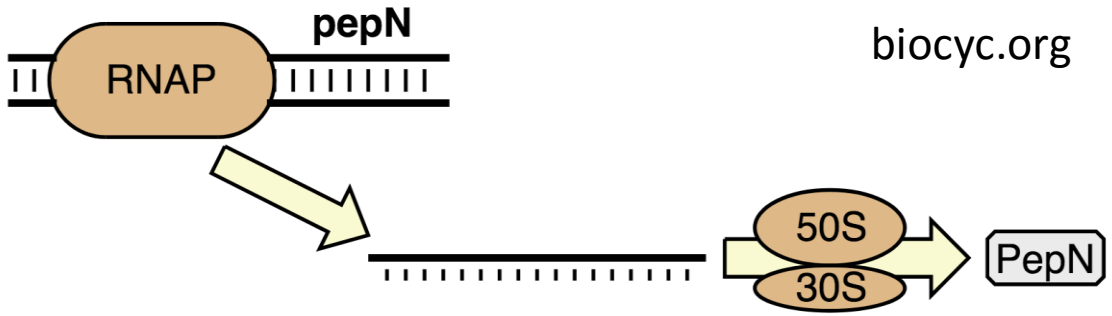
- A 3.5 fold peak upstream of malT, on Team E only.
- malT is also upregulated by Crp.

malE



- No significant peaks around malE.
- malE is regulated by both Crp & MalT.
- No significant peaks around other regulators.

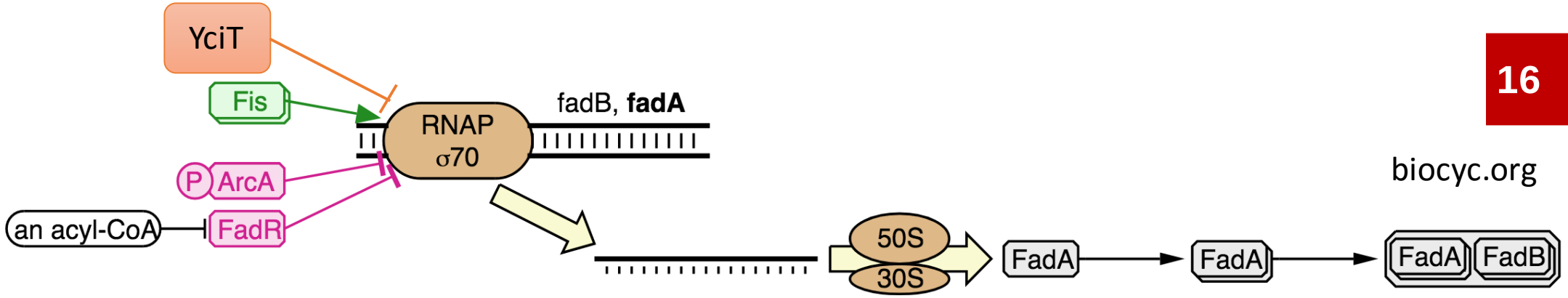
pepN



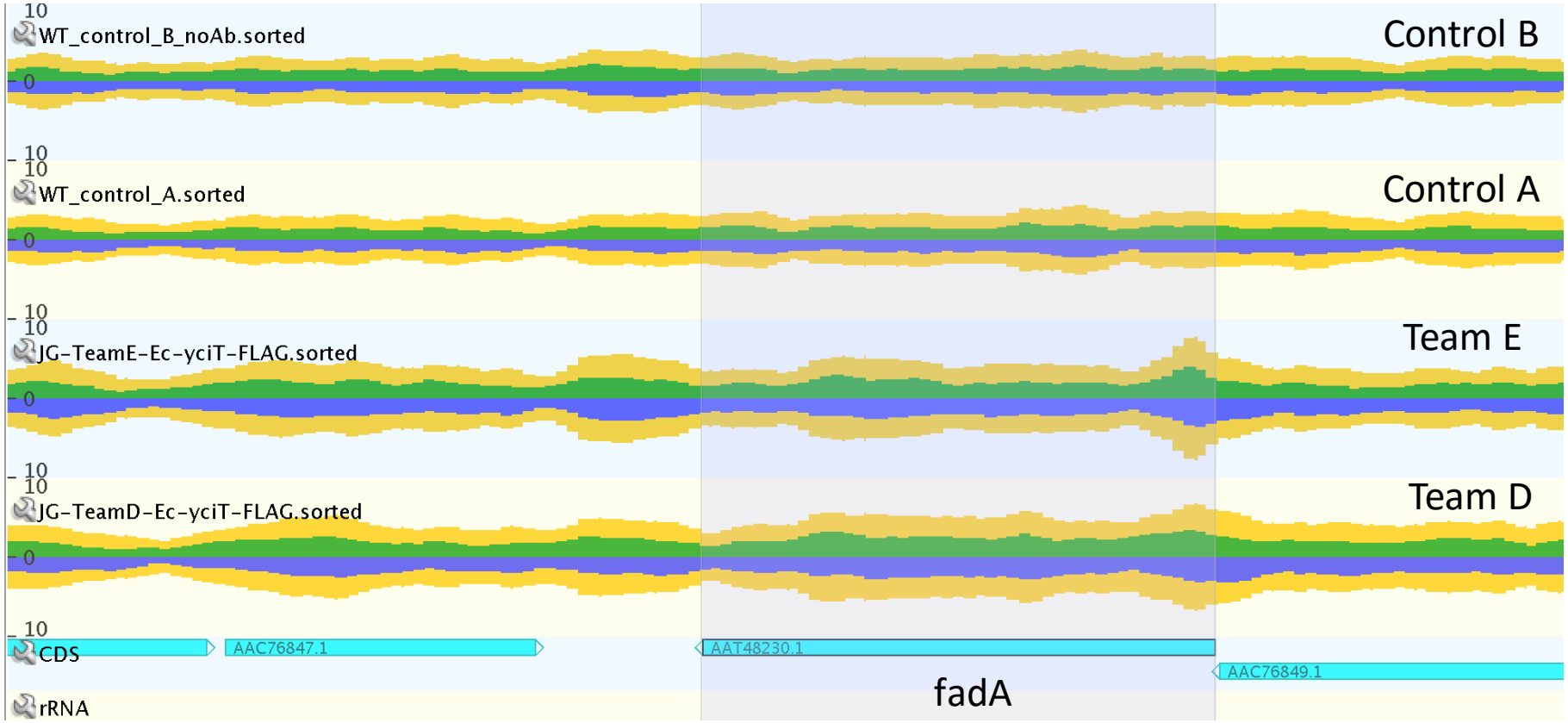
biocyc.org

- No significant peaks around pepN.
- No know regulators.

fadA



biocyc.org

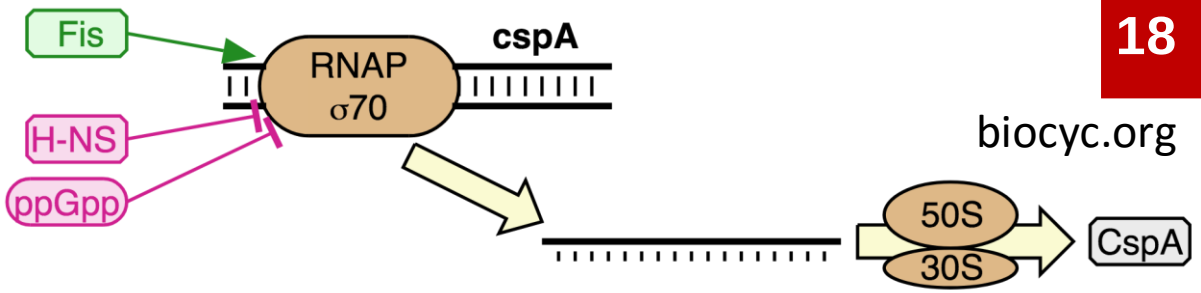


2.1 fold peak that appears in both teams, suggests direct downregulation of fadA by YciT.

Novel Peaks (yciT)

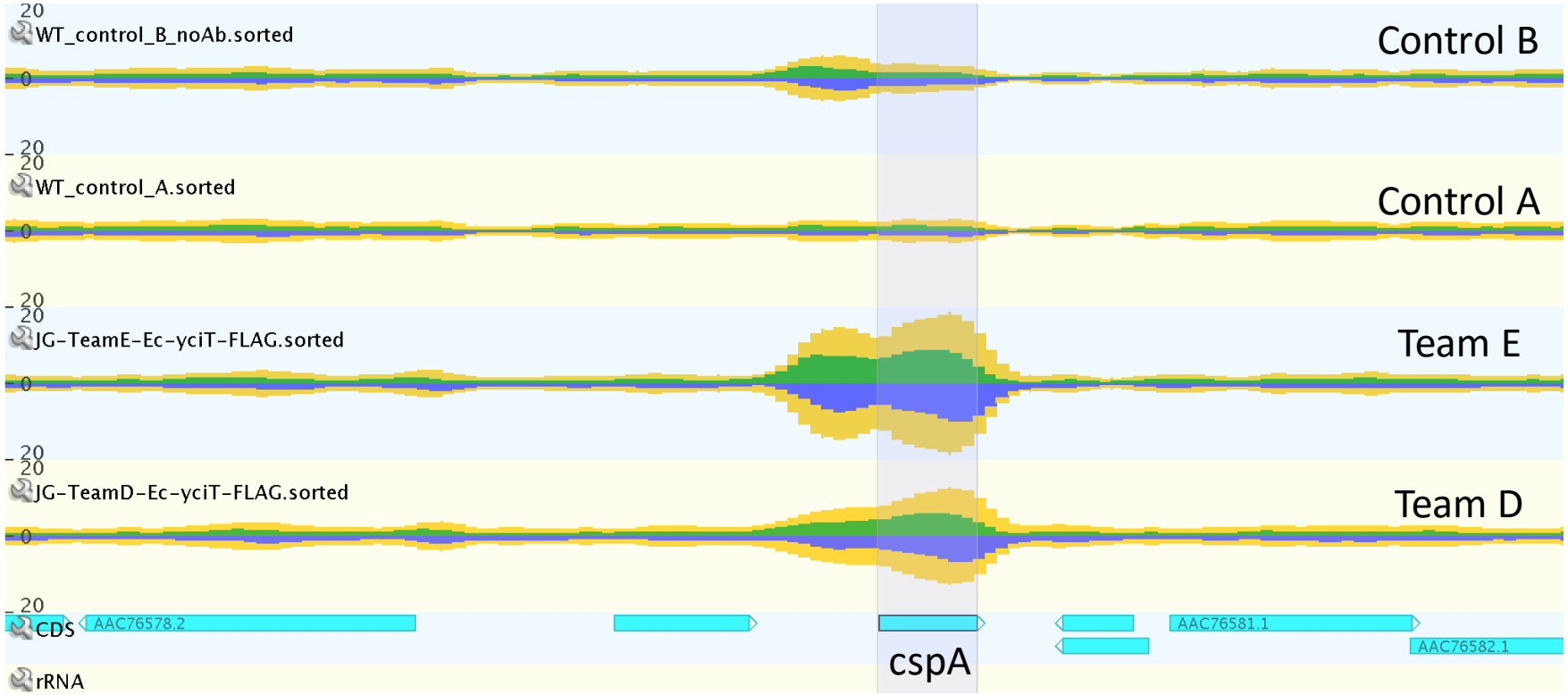
cspA

DNA-binding transcriptional activator CspA



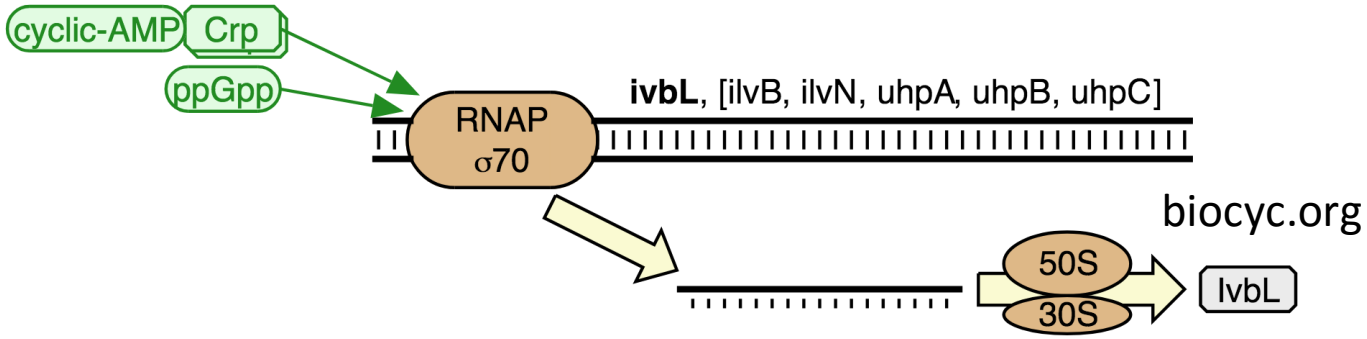
biocyc.org

- "The “cold shock protein A," CspA, is a major cold shock protein and was shown to be detected only during early-log-phase growth at 37°C and during log phase after a shift from 37°C to 10°C" (biocyc.org).

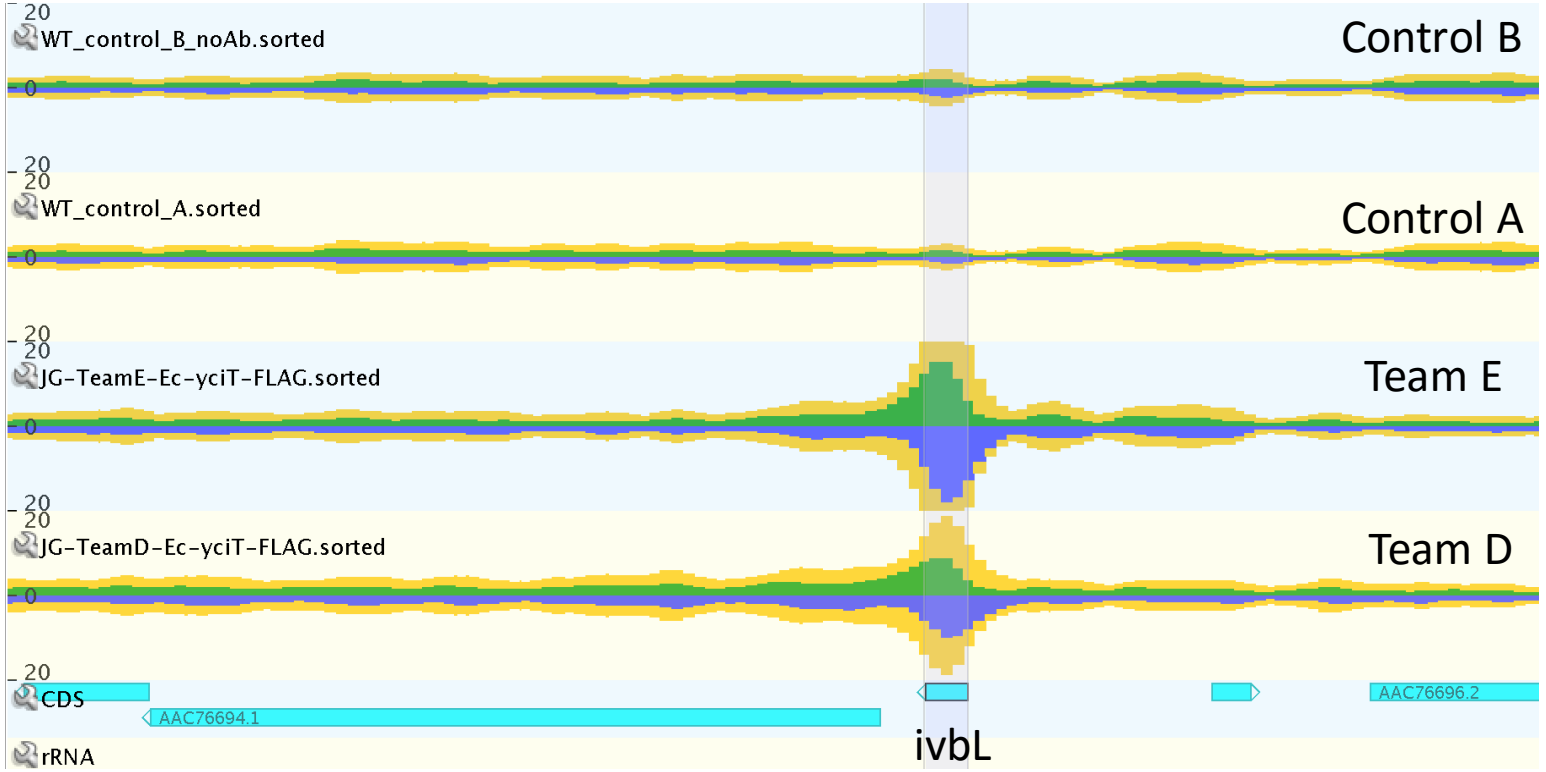


Genic 4.5 fold peak

ivbL
ilvBN operon leader peptide



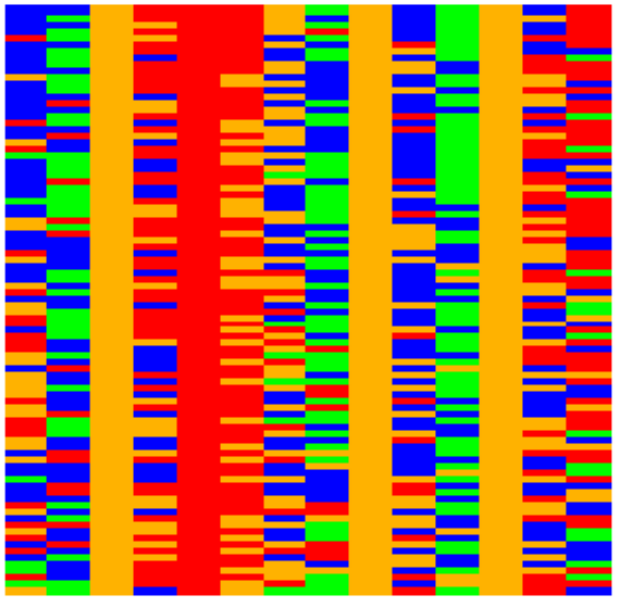
- "The ilvBN operon leader peptide (IvbL) controls by attenuation the expression of the ivbL-ilvBN operon, which codes for an enzyme involved in isoleucine and valine biosynthesis" (biocyc.org).

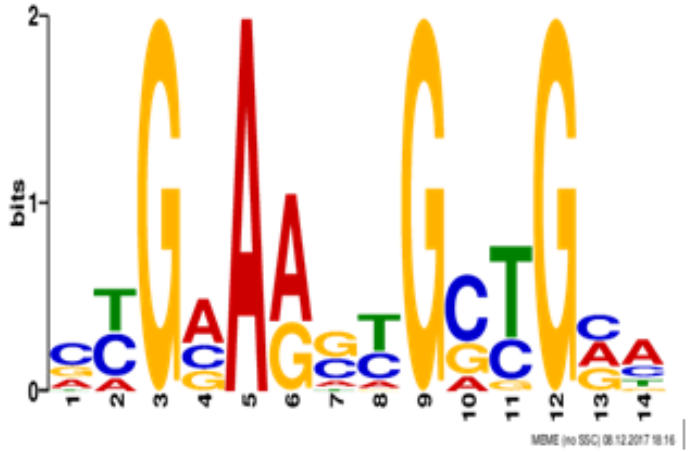


Genic 6.3 fold peak

Motifs

- Found two strong motifs (E-value around 1E-20) around the mode.
- One (shown) is in one third of the SPAT sequences, has some match to NtrC.





allR

Literature review

- Regulates several genes involved in the anaerobic utilization of allantoin (a product of purine degradation) as a nitrogen source as well as aerobic pathway to generate energy.
- AllR and AllS transcription factors work in conjunction
- Known to negatively regulate *allA, allR* and *glc* operons.

Microbiology (2008), **154**, 3366–3378 DOI 10.1099/mic.0.2008/020016-0

The transcription regulator AllR senses both allantoin and glyoxylate and controls a set of genes for degradation and reutilization of purines

Akiko Hasegawa, Hiroshi Ogasawara, Ayako Kori, Jun Teramoto and Akira Ishihama

Department of Frontier Bioscience, Hosei University, Koganei, Tokyo 184-8584, Japan

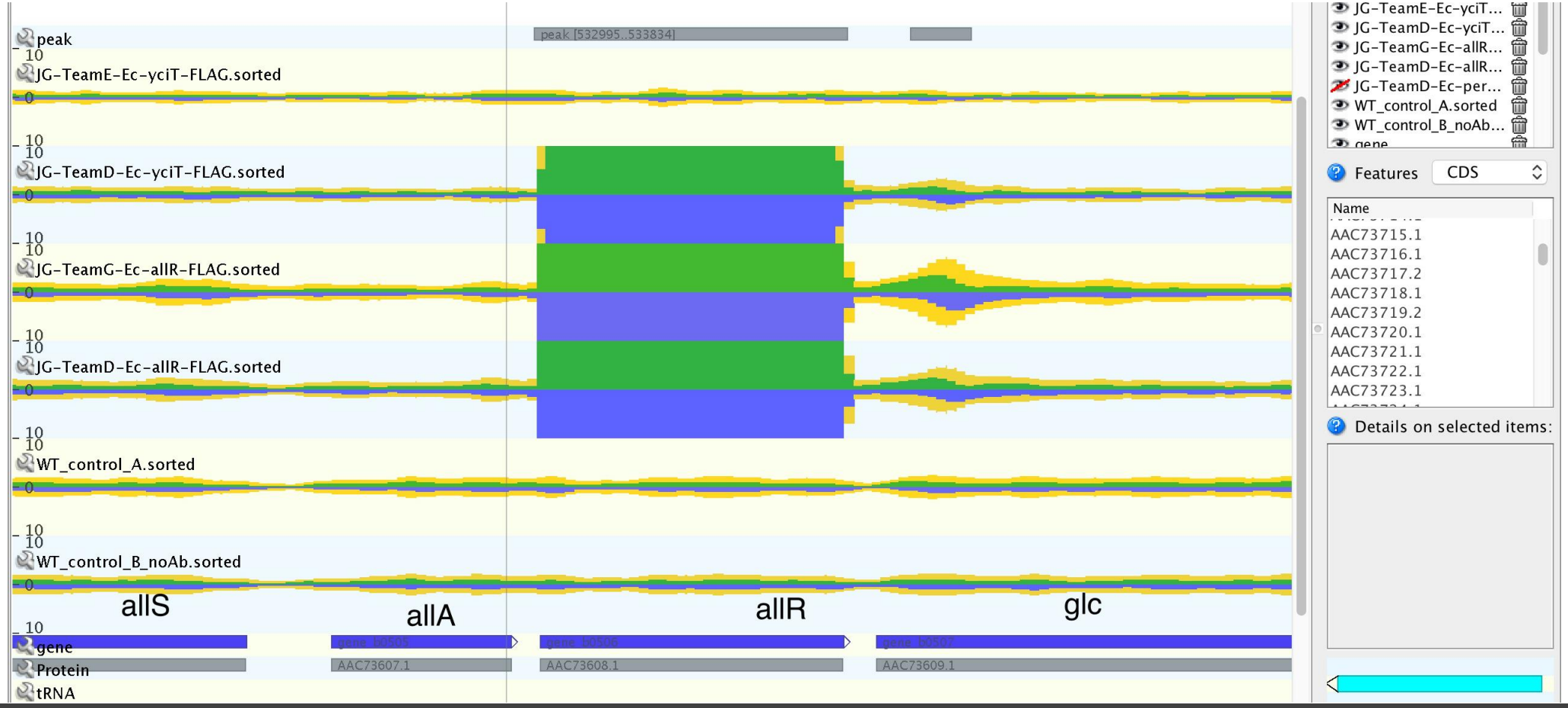
Correspondence
Akira Ishihama
aishiham@hosei.ac.jp

Received 28 April 2008
Revised 14 July 2008

Purines are degraded via uric acid to yield allantoin. Under anaerobic conditions, allantoin is further degraded via carbamoylphosphate to NH_4^+ to provide a nitrogen source and, under aerobic conditions, to 3-phosphoglycerate via glyoxylate for energy production. In this study, we found that a DNA-binding transcription factor AllR, together with AllS, plays a key role in switching control of two pathways, nitrogen assimilation and energy production. The repressor function of AllR is activated in the presence of allantoin, the common substrate for both pathways, leading to repression of the genes for energy production. On the other hand, when glyoxylate is accumulated, AllR is inactivated for derepression of the pathway for energy production. RutR, the

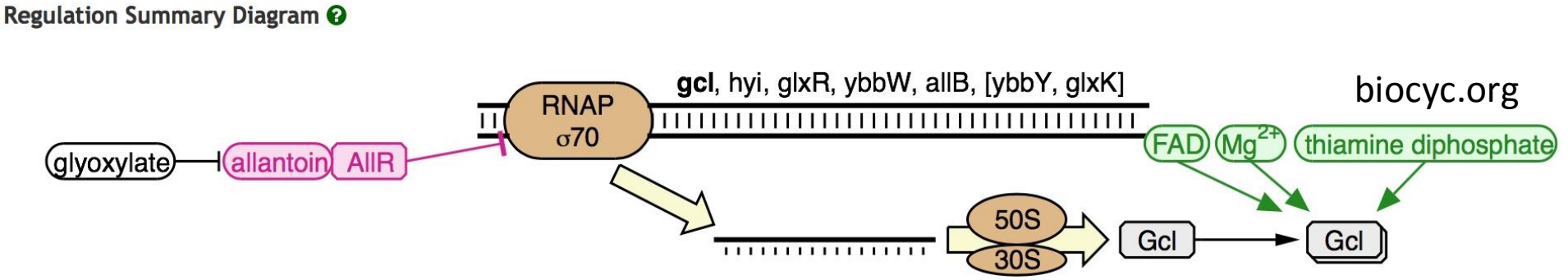
Peaks on literature

- Peak is located within gene b0507 (*glc*) as well as downstream from gene b0505 (*allA*).Literature said it ought to be upstream from *glc*.

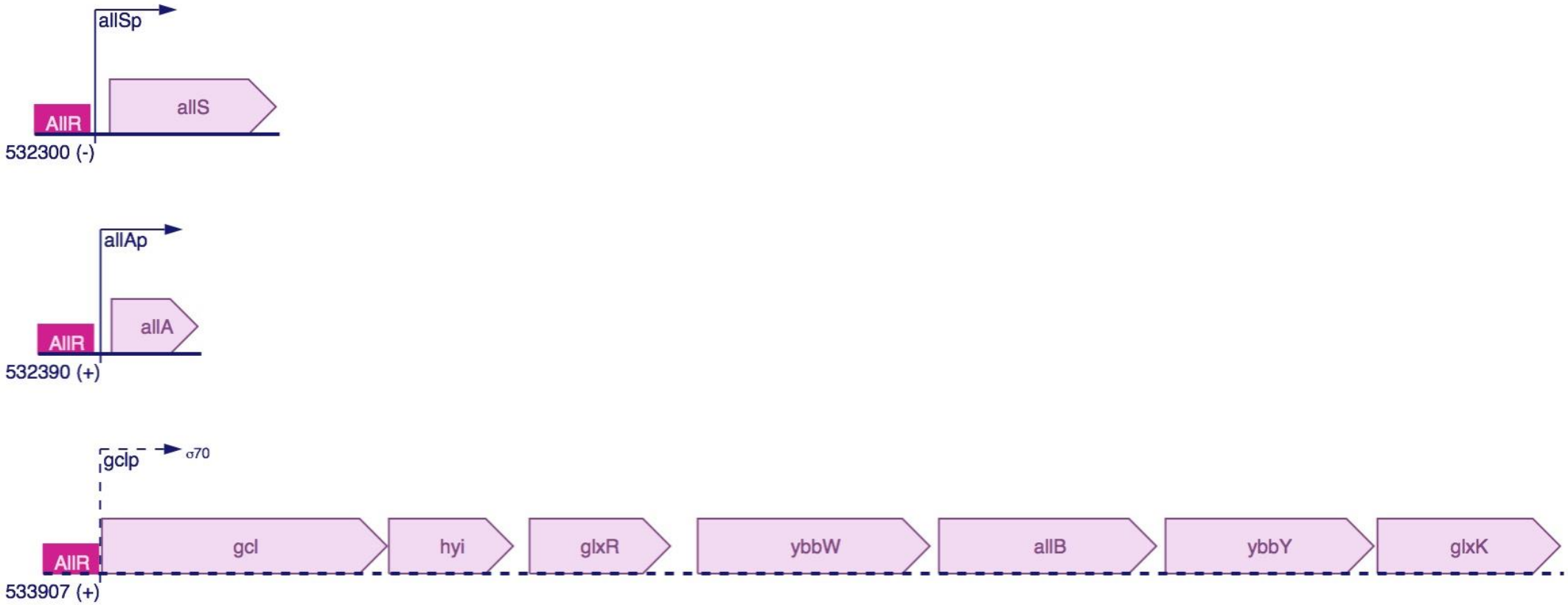


Peaks on literature

- Unfortunately, watermark at b0506 (*allR*) prevents us from seeing (or hoping to see) self-regulation.
- Our experiment did not show peak between *allS* and *allA*
- A corepressor might be necessary



glc, *hyi*, *glxR*, *byyW*, *allB*, *ybbY and glxK* all are co-regulated.



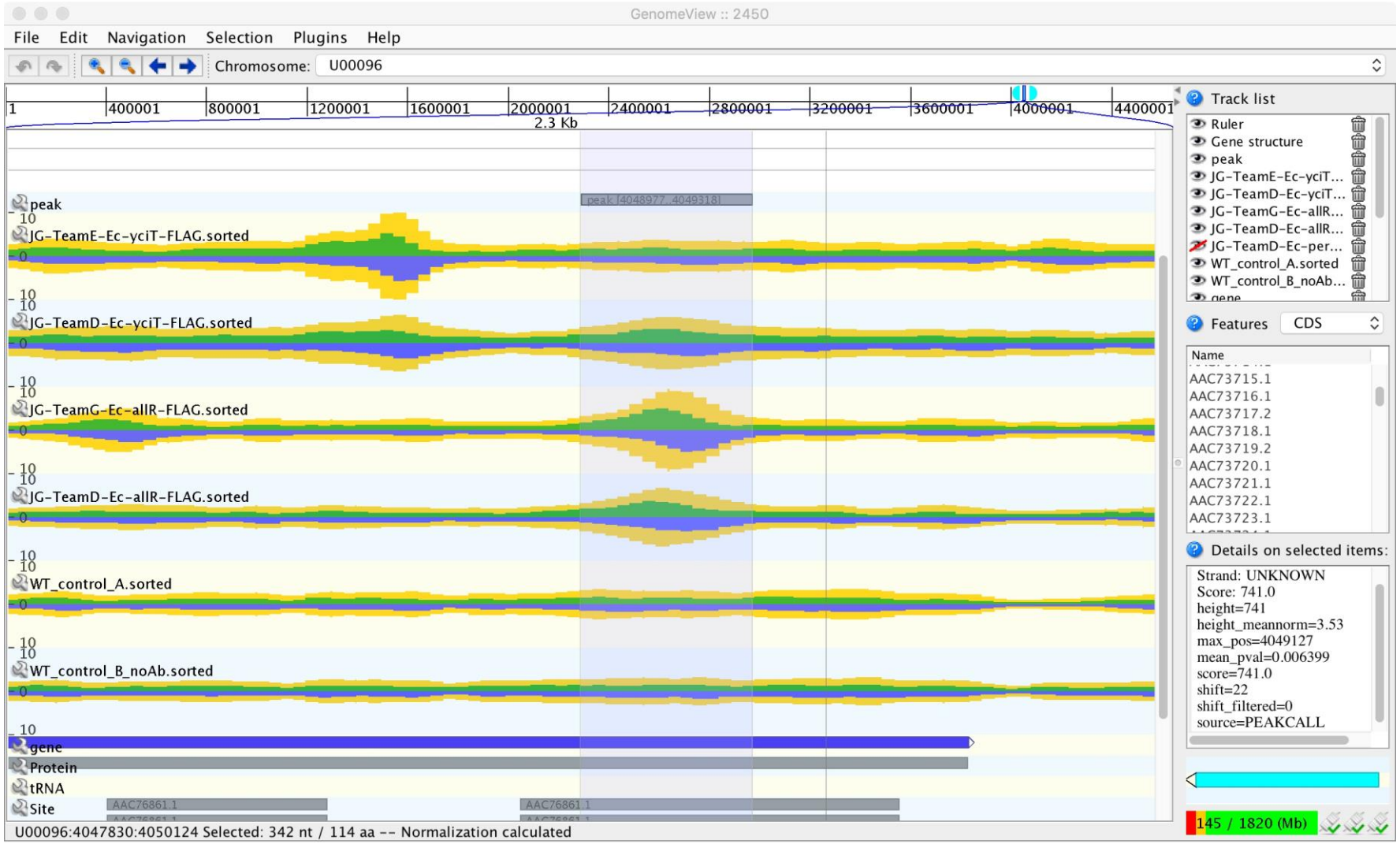
ecocyc.org

Novel peaks (allR)

Novel peaks (allR)

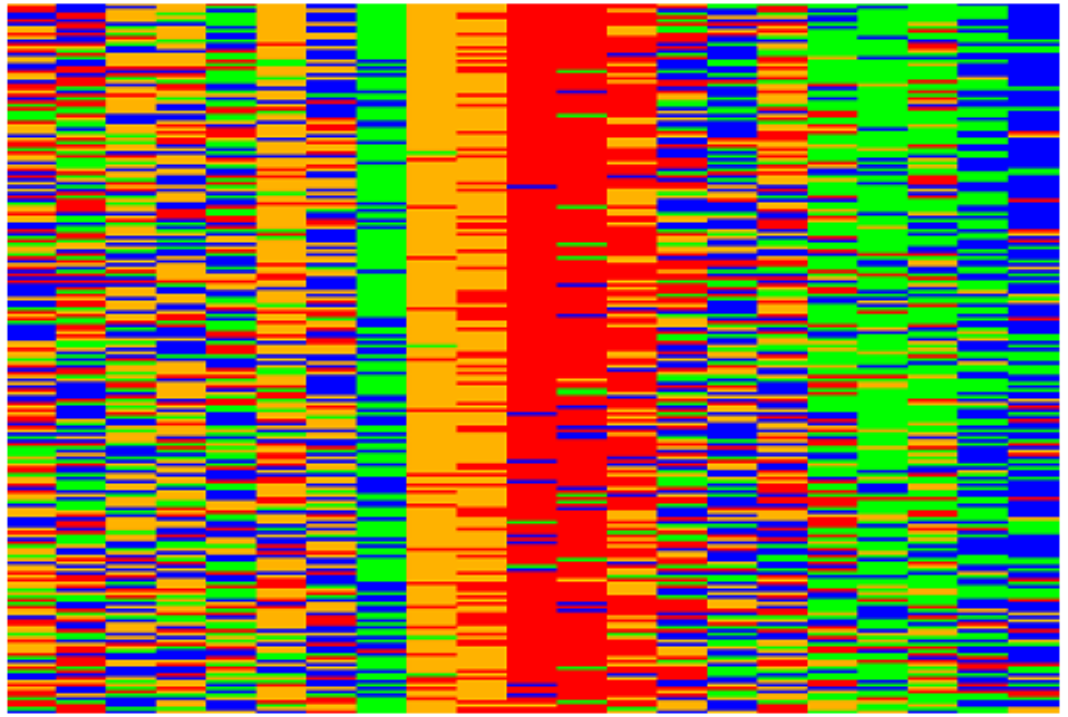
- We found several peaks that do not seem to be functionally related to allR.
- For example, we found peaks within the following genes: citC, polA, bisC, yiaN, mgtA.
- For each peak, we looked at 4 genes nearby, 2 upstream and 2 downstream and they all are not related to the known function of allR.

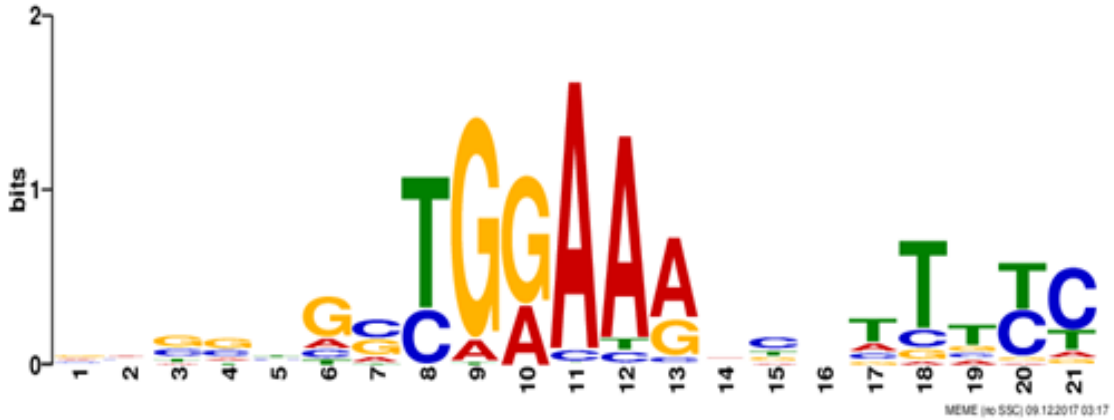
Example: Binding site at *polA*, can possibly negatively regulate transcription of *polA* gene into DNA polymerase I.



Motifs

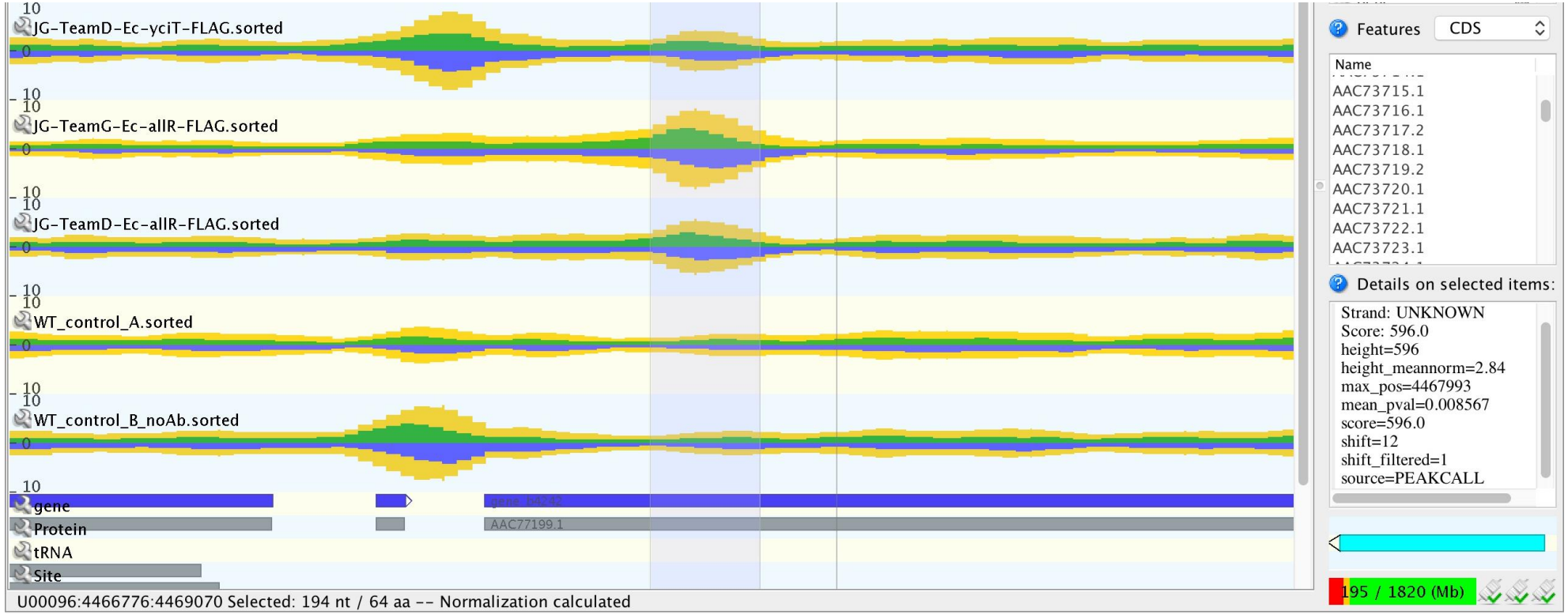
- Found one very strong motif (E-value=2.5eE-126) for all (210) SPAT sequences around mode
- This motif has some match (1.22E-3) with the binding motif of PurR (regulator of purine metabolism)
- Another motif found in top 12 binding sites (by p-value), with some match to PhoP (control of acid resistance genes)





Peak with motif

- Peak at gene b4242 is shown to contain motif for *phoP*.
- Need further experiment to show it.



perR

Literature review

- Not well documented in E. Coli
- More well known in B. Subtilis
  - Involved in sensing peroxides
  - Regulates genes related in iron regulation in the cell
  - Known to repress several genes
- We found no enrichment for the genes suppressed in B. Subtilis that are also in E. Coli (ahpC, ahpF, fur, hemA operon)
- It is self repressed in B. Subtilis, but we did not found enough evidence that this is the case in E. Coli
- 30 potential binding sites found

Novel peaks (perR)

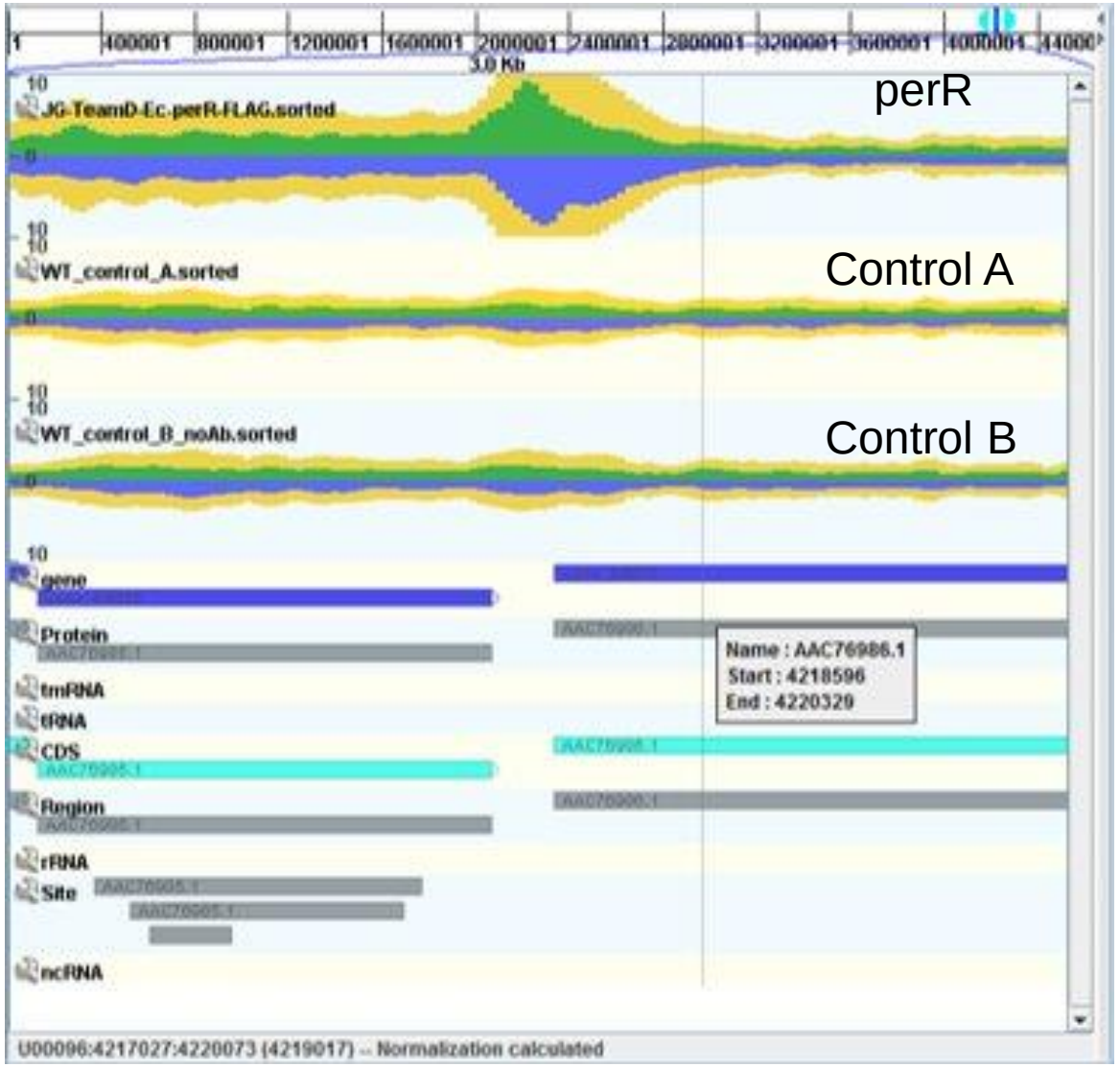
Novel peaks (perR)

- We found 30 peaks with p-value<0.01 and shift higher than zero from SPAT peaks
- The peaks were visually inspected in GenomeView
- Due to the volume of genes, a list was submitted to DAVID to find common themes
- The DAVID analysis showed involvement in iron regulation similar to its role in B. subtilis

Annotation Cluster 1	Enrichment Score: 1.46			Count	P_Value	Benjamini
UP_KEYWORDS	Tricarboxylic acid cycle	RT		4	1.7E-3	1.4E-1
GOTERM_BP_DIRECT	tricarboxylic acid cycle	RT		4	4.6E-3	3.3E-1
KEGG_PATHWAY	Carbon metabolism	RT		4	6.6E-2	5.7E-1
KEGG_PATHWAY	Biosynthesis of secondary metabolites	RT		5	2.6E-1	9.2E-1
KEGG_PATHWAY	Microbial metabolism in diverse environments	RT		4	3.6E-1	9.4E-1
Annotation Cluster 2	**Enrichment Score: 0.64**			**Count**	**P_Value**	**Benjamini**
KEGG_PATHWAY	Metabolic pathways	RT		11	3.7E-2	6.1E-1
KEGG_PATHWAY	Biosynthesis of secondary metabolites	RT		5	2.6E-1	9.2E-1
UP_KEYWORDS	Acetylation	RT		3	2.9E-1	9.9E-1
GOTERM_CC_DIRECT	cytosol	RT		8	9.7E-1	1.0E0
Annotation Cluster 3	**Enrichment Score: 0.48**			**Count**	**P_Value**	**Benjamini**
UP_KEYWORDS	Iron	RT		5	2.5E-1	1.0E0
UP_KEYWORDS	4Fe-4S	RT		3	3.3E-1	9.9E-1
GOTERM_MF_DIRECT	4 iron, 4 sulfur cluster binding	RT		3	3.4E-1	1.0E0
UP_KEYWORDS	Iron-sulfur	RT		3	4.4E-1	1.0E0
Annotation Cluster 4	**Enrichment Score: 0.25**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	topological domain:Periplasmic	RT		9	3.3E-1	1.0E0
UP_SEQ_FEATURE	transmembrane region	RT		12	4.0E-1	1.0E0
UP_KEYWORDS	Transmembrane	RT		12	4.7E-1	9.9E-1
GOTERM_CC_DIRECT	integral component of membrane	RT		9	4.8E-1	9.9E-1
UP_SEQ_FEATURE	topological domain:Cytoplasmic	RT		8	4.8E-1	1.0E0
GOTERM_CC_DIRECT	integral component of plasma membrane	RT		5	5.4E-1	9.8E-1
UP_KEYWORDS	Transmembrane helix	RT		11	5.5E-1	9.9E-1
UP_KEYWORDS	Cell inner membrane	RT		10	6.9E-1	1.0E0
UP_KEYWORDS	Membrane	RT		13	6.9E-1	1.0E0
GOTERM_CC_DIRECT	plasma membrane	RT		11	7.2E-1	9.9E-1
UP_KEYWORDS	Cell membrane	RT		11	7.7E-1	1.0E0
UP_KEYWORDS	Transport	RT		7	8.4E-1	1.0E0
Annotation Cluster 5	**Enrichment Score: 0.14**			**Count**	**P_Value**	**Benjamini**
GOTERM_MF_DIRECT	DNA binding	RT		5	6.1E-1	1.0E0
UP_KEYWORDS	DNA-binding	RT		5	7.0E-1	1.0E0
UP_KEYWORDS	Transcription	RT		3	9.0E-1	1.0E0
Annotation Cluster 6	**Enrichment Score: 0.05**			**Count**	**P_Value**	**Benjamini**
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	RT		3	7.5E-1	1.0E0
UP_KEYWORDS	Nucleotide-binding	RT		4	9.2E-1	1.0E0
UP_KEYWORDS	ATP-binding	RT		3	9.5E-1	1.0E0
GOTERM_MF_DIRECT	ATP binding	RT		3	9.6E-1	1.0E0

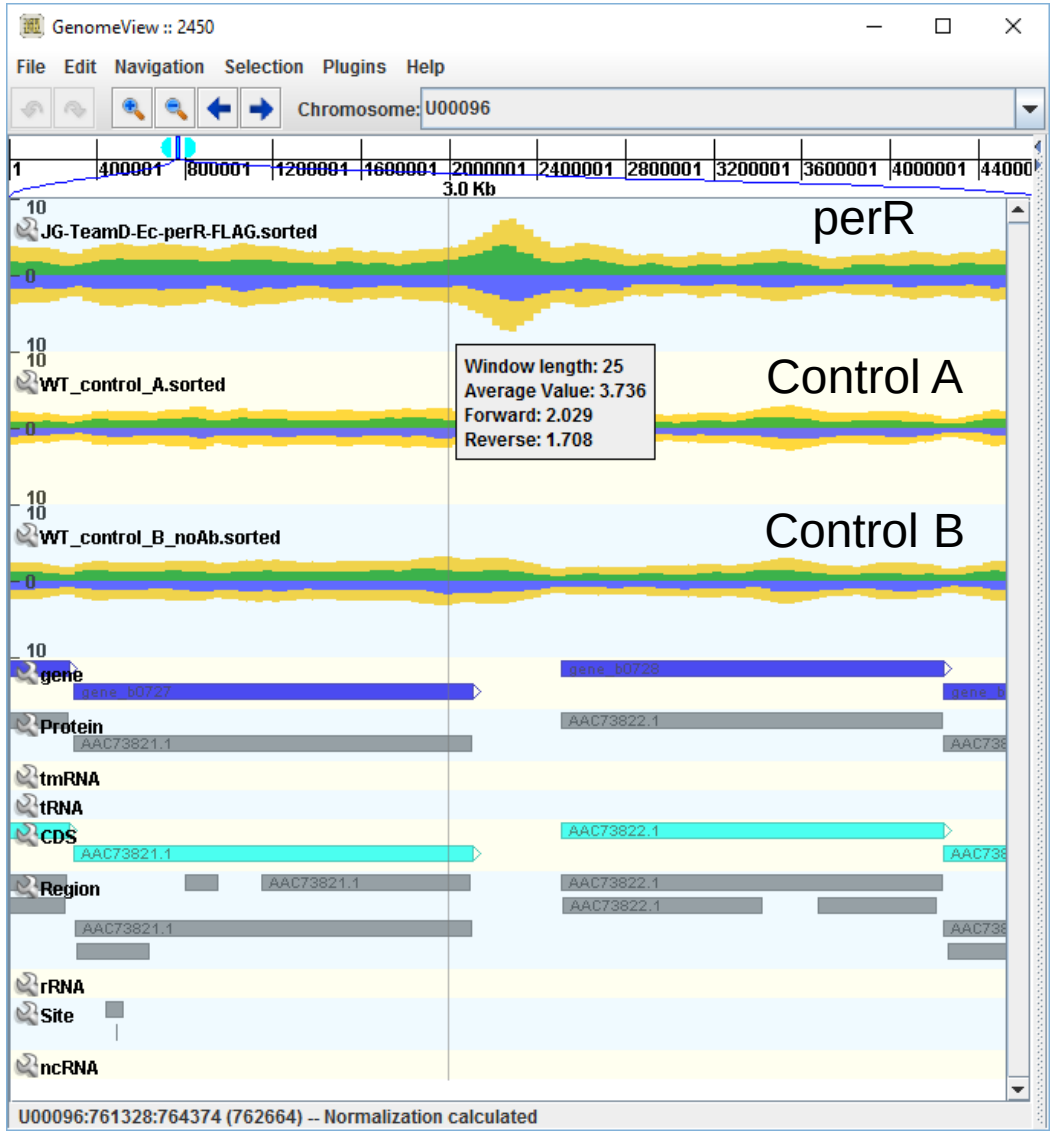
Possible Regulation of aceK and aceA by perR

- aceK is involved in the transition from the Krebs cycle to the glyoxylate cycle
- aceA is a major component of the glyoxylate
- The glyoxylate cycle uses acetates instead of glucose as the source for ATP. This acetate usually comes from the fatty storages of the cell usually when the cell is glucose starved
- This may indicate that perR is involved in controlling the cell's response to glucose starvation
- Possible experiment: observe how perR mutant cells will react to glucose starvation compared to wild type cells



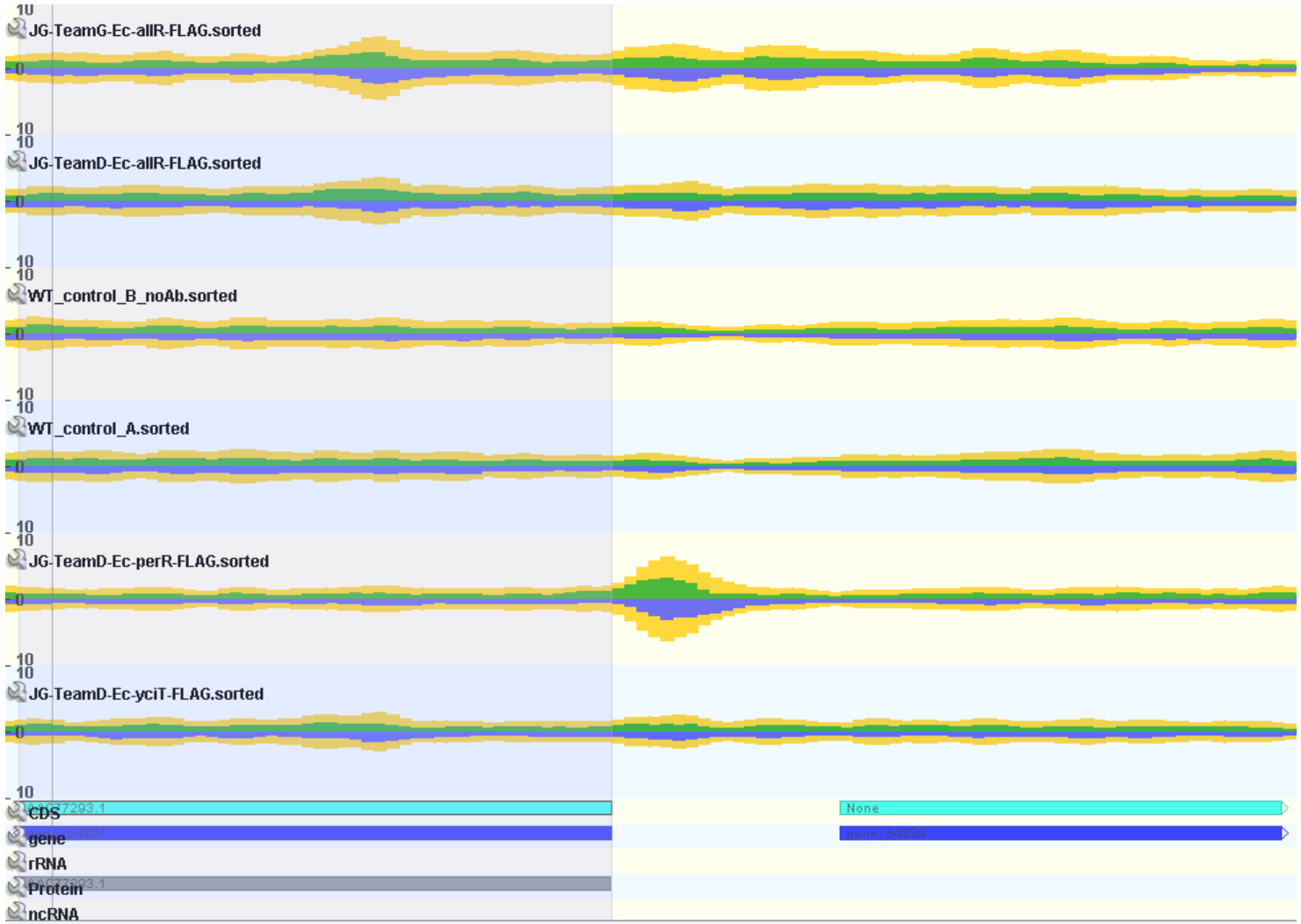
Possible regulation of sucB and sucC by perR

- PerR binds to intergenic region between sucC and sucB, which are part of the OGDH complex (rate-limiting enzyme of the Krebs cycle)
- Might indicate that perR is involved in regulating aerobic respiration
- Future Experiment: Look at product of OGDH catalyzed reaction (succinyl-CoA) in PerR mutant and wild type.



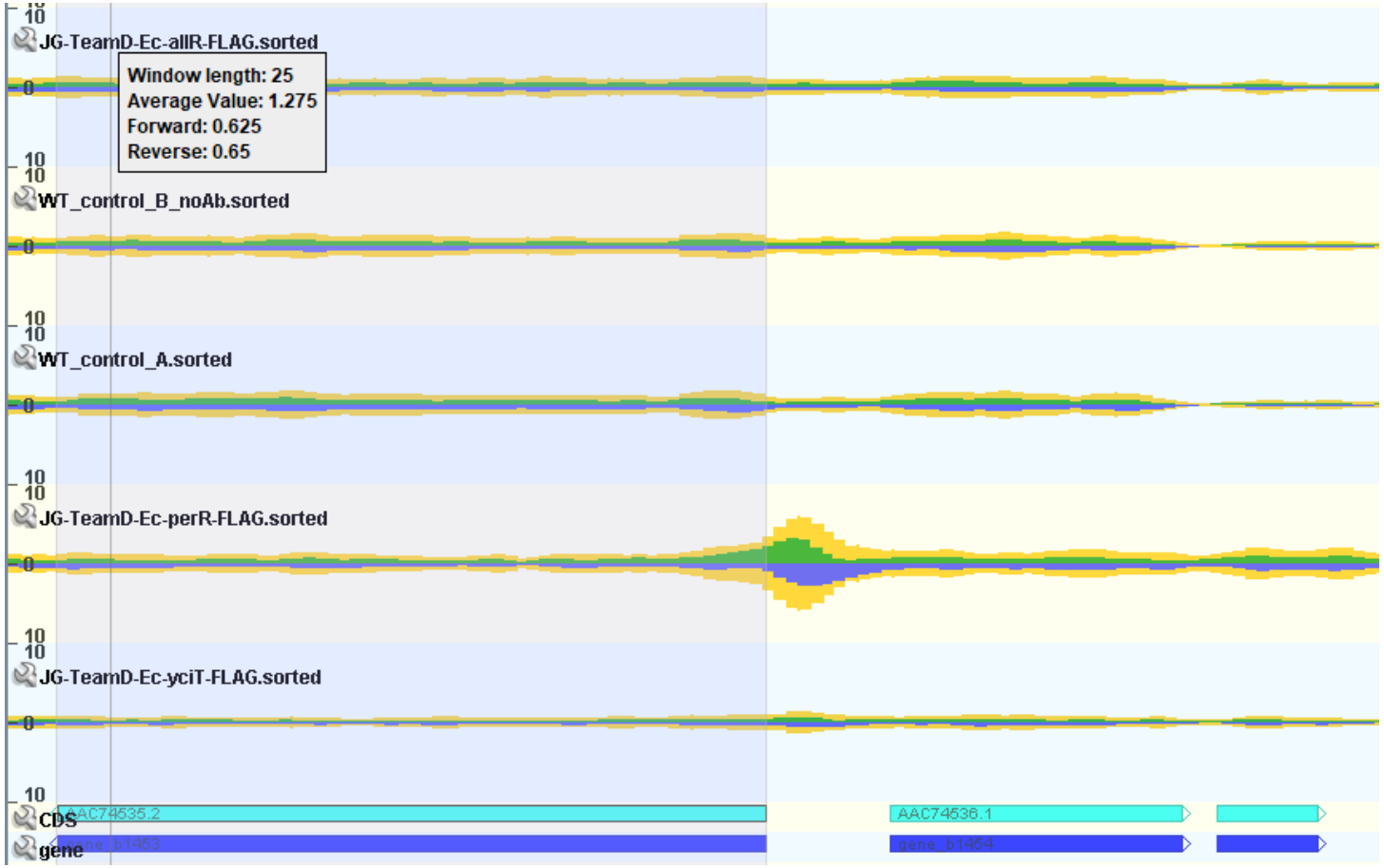
Possible regulation of mdtM by perR conferring antibiotic resistance

- perR binds to many genes that code for transmembrane proteins, like mdtM
- MdtM protein is a multi-drug efflux pump and also both a Na:H+ and K:H+ antiporter
- 4-fold increase in acriflavin resistance and 2-fold increase in chloramphenicol resistance from overexpressed cloned mdtM gene in drug-sensitive strain
- Future experiment- Look at acriflavin and chloramphenicol resistance with perR mutant and wild type perR



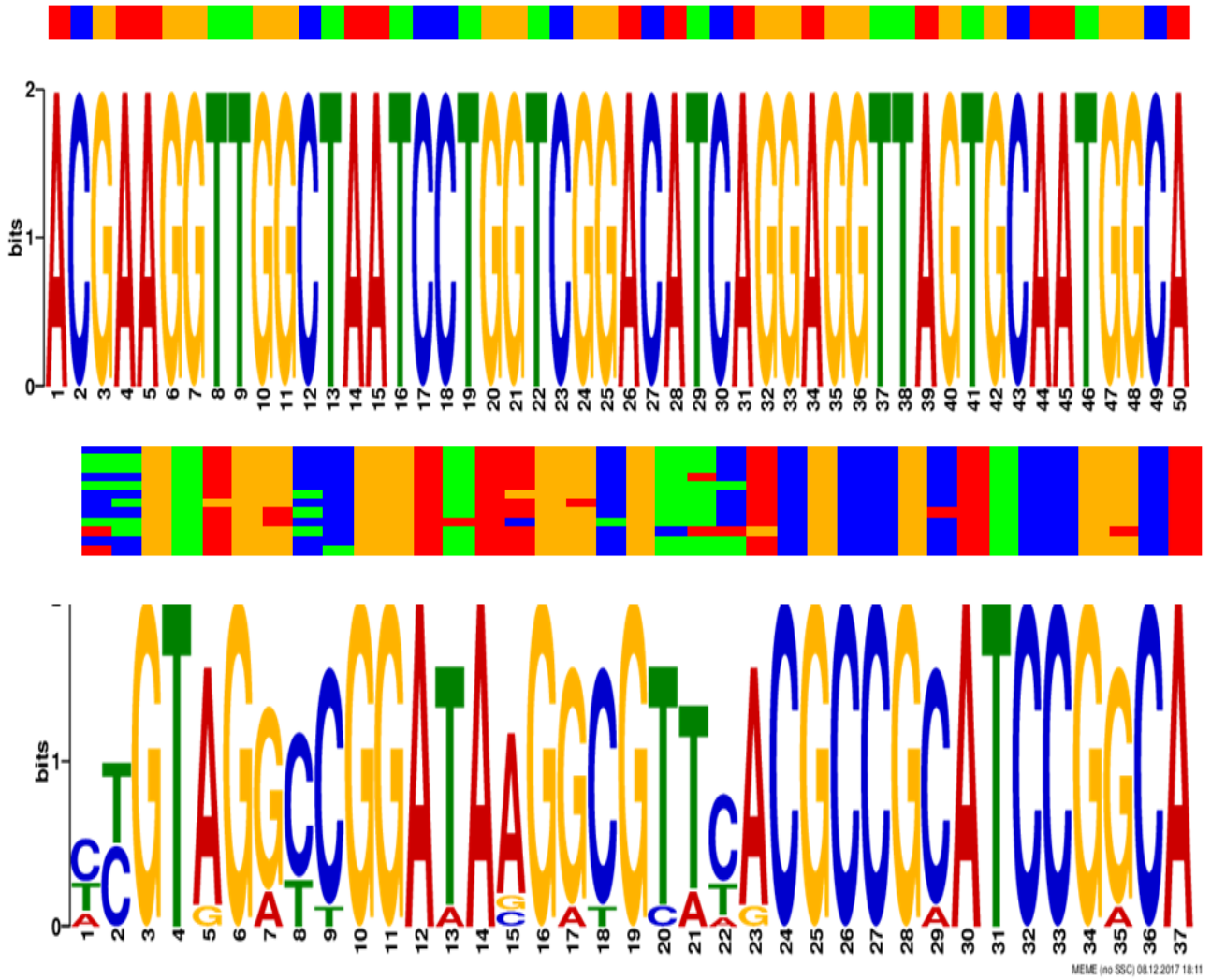
Possible regulation of ansP and mtr by perR

- ansP and mtr code for L-asparagine:H+ symporter and tryptophan:H+ symporter
- PerR may be involved in the growth phase of E.coli, since amino acids such as tryptophan is needed for growth
- PerR may also be involved in the pH balance and ion homeostasis of E.coli
- Future experiment- Look at growth phase E.coli with and without perR mutant, and look at rate of growth



Motifs

- One exact motif for five regions associated with rRNA, similar (7.5E-4) to motif of the Rob TF
- In general, no strong “global” motif found
- Another motif for 12 of the binding sites (E-value 1.6E-113)



Conclusions

- Identified YciT as a possible repressor for crp
- PerR involved in metabolic pathways of E.coli such as in TCA cycle, glyxoxylate cycle, growth by amino acid transporters
- We hypothesize that perR is involved with stress response in E. coli
- Our results for allR were more limited, but found some very strong binding motifs
- Binding motifs were also identified for the other transcription factors
- Our results only show binding sites and hypothesis about possible biological roles, so future experiments need to be performed to validate regulation by transcription factor

Tools Used

- EcoCyc
- MEME
- Tomtom
- SPAT
- GenomeView
- DAVID