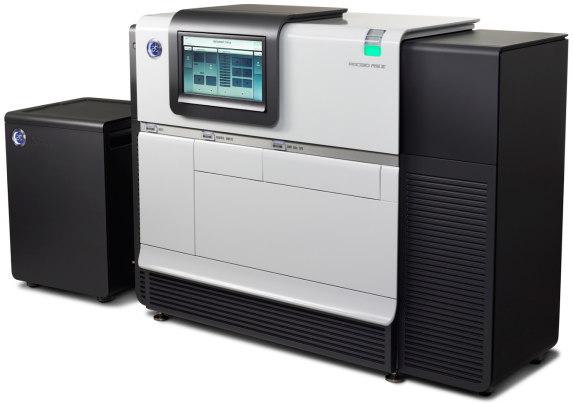
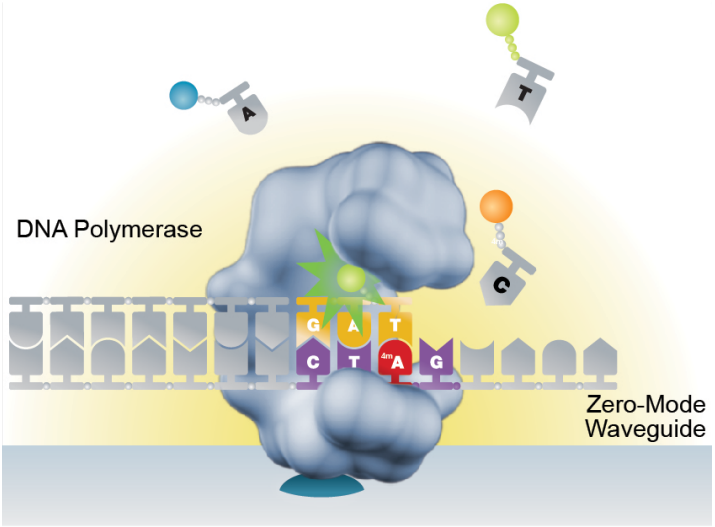
Methylation and eQTL analyses

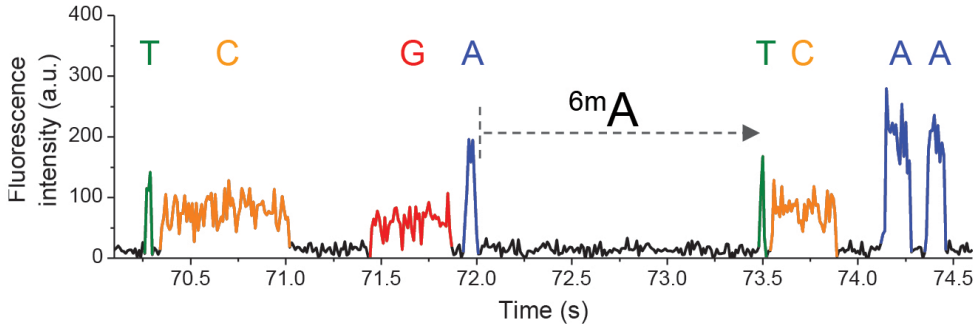
London School of Hygiene and Tropical Medicine

Pacific Biosciences SMRT sequencing

- Polymerase incorporates template bases
- Light signals are detected
- Florescence intensities are converted into base calls

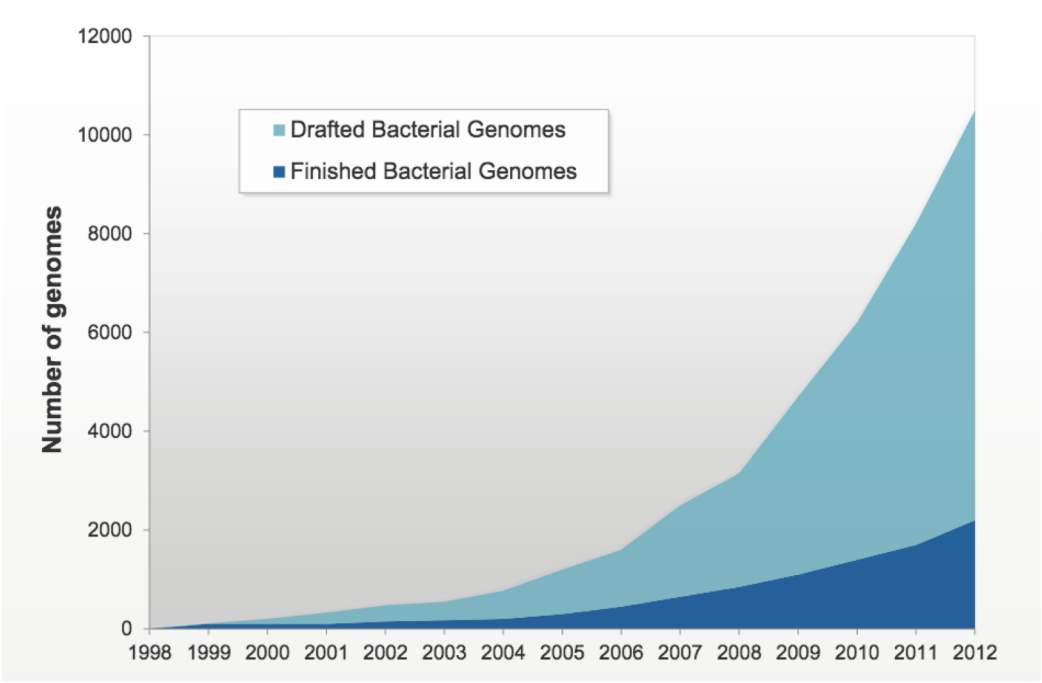


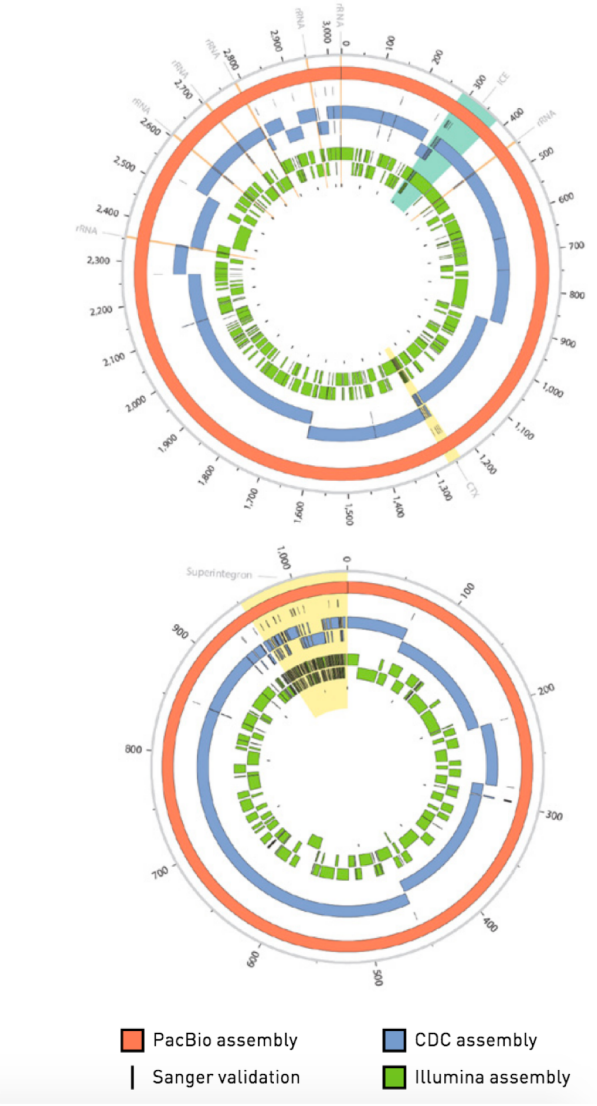




Finishing genomes

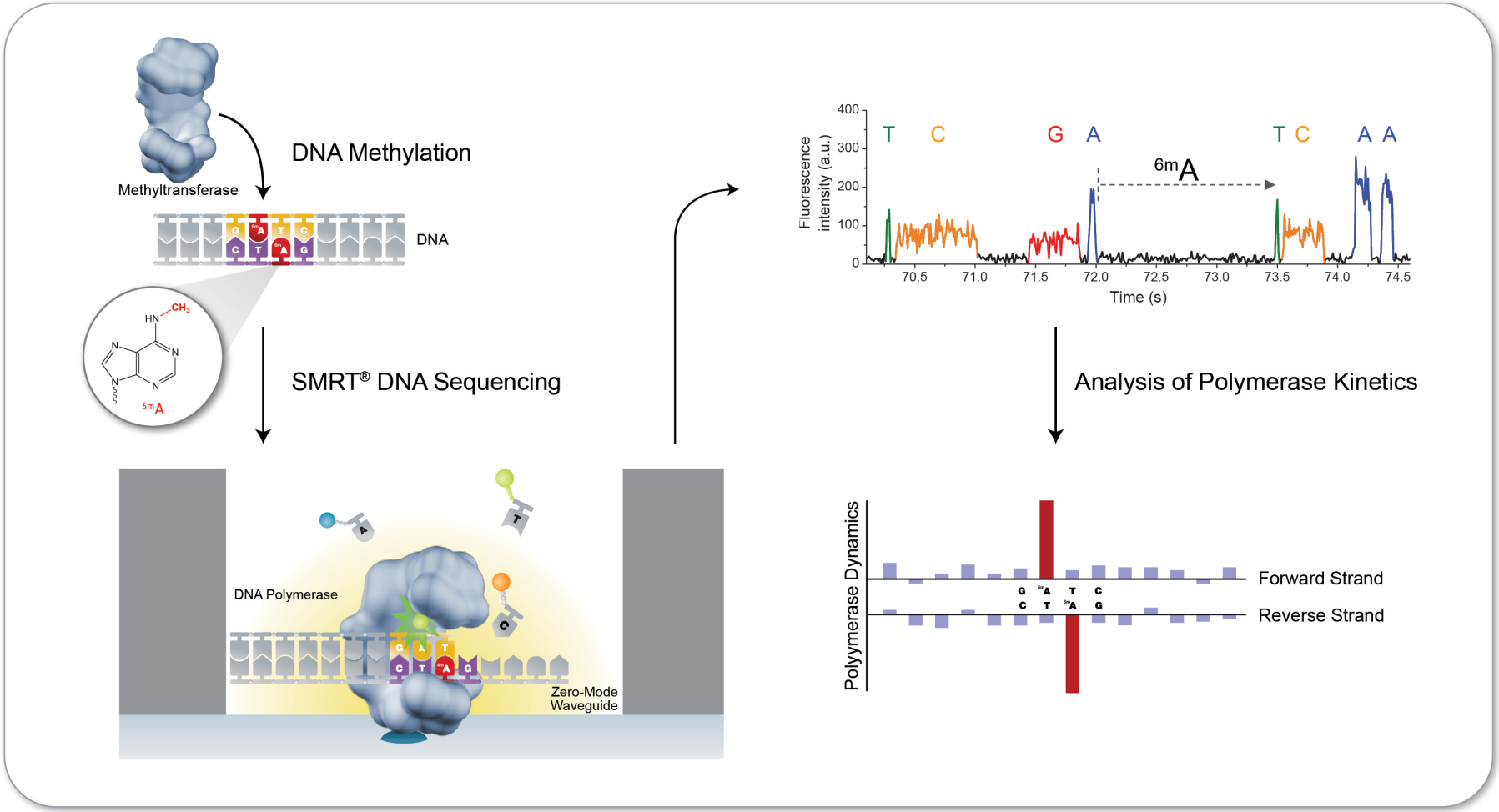
- Long reads allow for finishing of genomes
- Better resolution than short read assemblies
- Followed by whole genome alignment



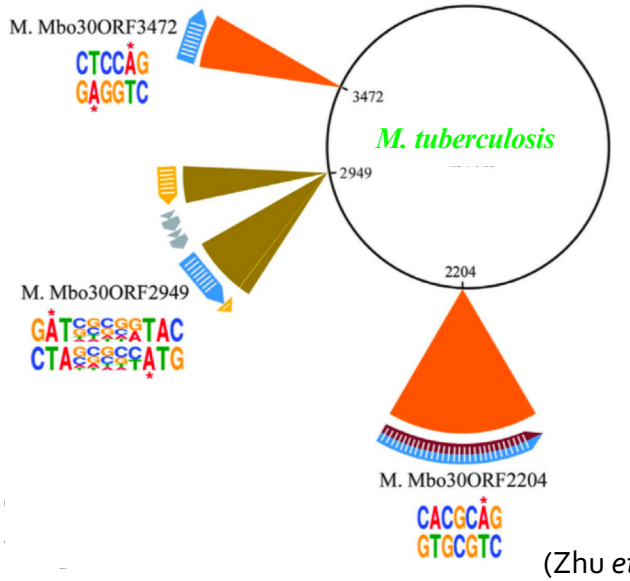


A. Bashir et al., Nat Biotechnol, 2012

Methylation



Methylation



(Zhu *et al.*, 2016)

- Virulent *M. tuberculosis* has been reported to contain 6mA.
- 3 methyltransferases (MTases) identified:

  mamA (Rv3263): CTCC**A**G, CTGG**A**G

  mamB (Rv2756c): CACGC**A**G

  hsdM (Rv2024c): G**A**TN$_4$RTAC, GT**A**YN$_4$ATC

- Orphan enzymes: no cognate restriction endonuclease with the same target site in the proximity of their genes.
- *mamA* and *mamB* are type IIG Mtases.
- *hsdM (hsdS1, hsdM* and *hsdR*) is a type I Mtase.
- Gene expressioDifferent mechanisms proposed (methylation in coding regions/promoter regions)
- n regulated by methylation
- Disruption of *mamA* decreased gene expression

(Shell *et al.*, 2013)

Gene	Symbol	Δ*mamA*/wildtype	Δ*mamA::mamA*/wildtype
Rv3263	*mamA*	−4.05	2.51
Rv0142		−1.32	−0.16
Rv1239c	*corA*	−0.80	−0.22
Rv3197A	*whiB7*	−0.75	0.02
Rv3083		−0.72	−0.19
Rv0102		−0.72	0.03
Rv3085		−0.68	−0.12
Rv3084	*lipR*	−0.62	−0.06
Rv3378c		−0.59	−0.07
tRNA-pro	*proU*	−0.15	−1.91
Rv2463	*lipP*	0.016	−2.62
tRNA-gly^f	*glyV*	−0.0089	−2.66

(Shell *et al.*, 2013)

Methylation Analysis

- Single-molecule real-time (SMRT) sequencing was performed over the 22 samples, based in the kinetic variation of single base.
- Modification was found through the Modification and Motif Analysis pipeline in SMRT Portal (PacBio).
- Three different motifs previously reported were identified, two of them with partner motif (methylated in both strands) and one of them hemi-methylated (Zhu *et al*., 2016; Phelan *et al*., 2018).

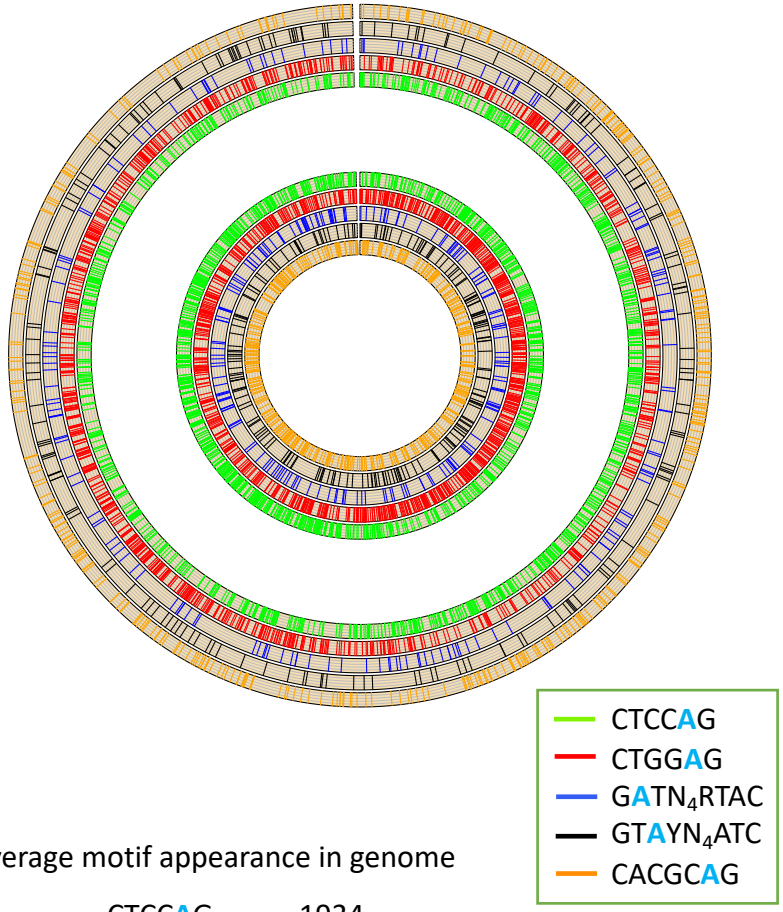
  mamA (Rv3263) - CTCC**A**G, CTGG**A**G

  mamB (Rv2024c) - CACGC**A**G

  hsdM (Rv2756c) - G**A**TN$_4$RTAC, GT**A**YN$_4$ATC

- m6A methylation.
- Other modifications were found but not in motifs (demonstrated as mainly false positives by Zhu *et a*l. by WGS analysis).

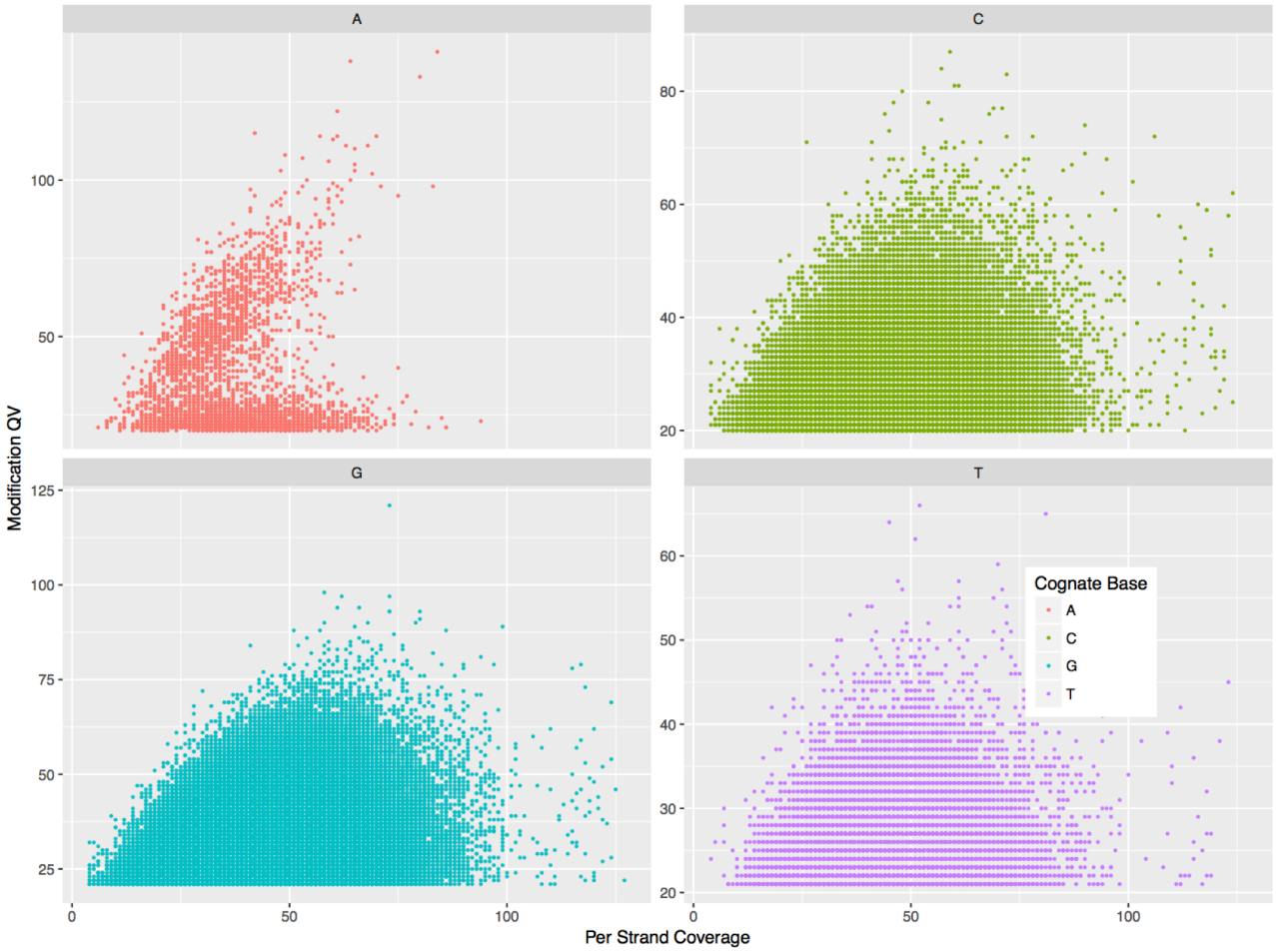
Distribution of motifs along the genome



Average motif appearance in genome

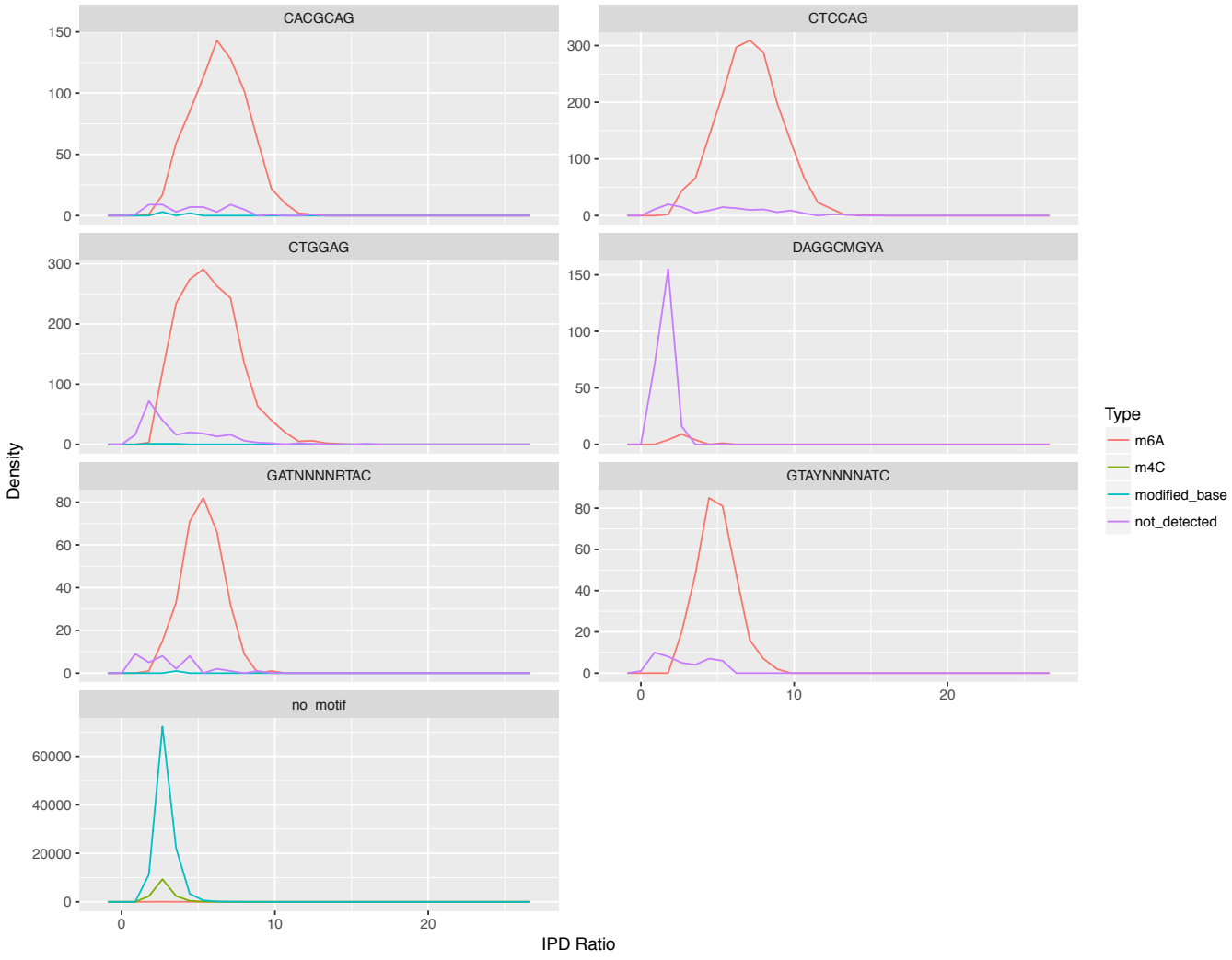
CTCC**A**G	1934
CTGG**A**G	1935
G**A**TN$_4$RTAC	357
GT**A**YN$_4$ATC	357
CACGC**A**G	813

Modification coverage vs Score



- Modification QV: -10 log (p value) score for the detection of the event.
- Min. Modification QV = 30 (p value = 0.001)
- Min. Strand Coverage = 20

IPD Ratio Distribution of the found motifs

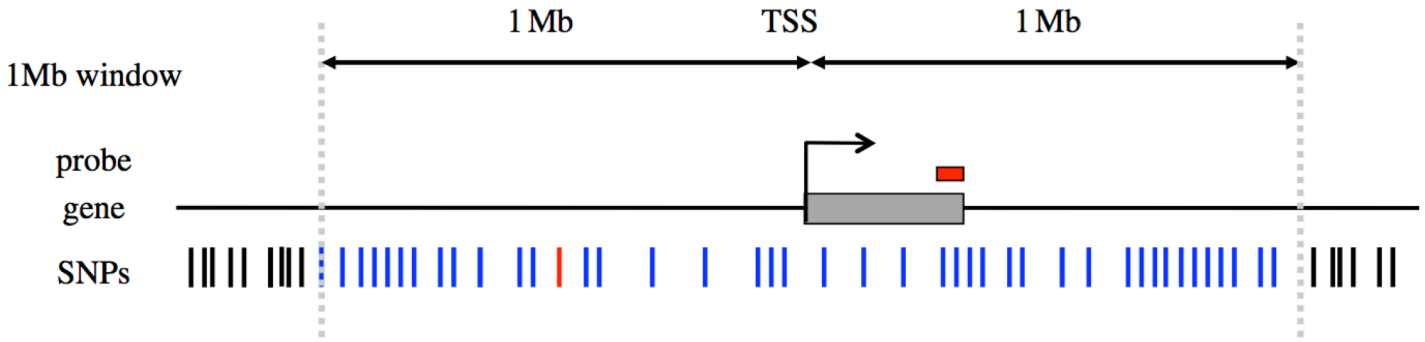


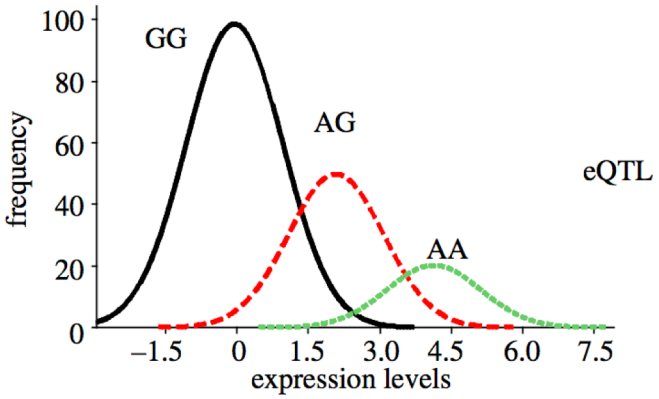
- IPD ratios of motifs found are between 3-10
- High quality motifs, high density of methylated m6A.
- High density of other modified bases but not in motifs found.

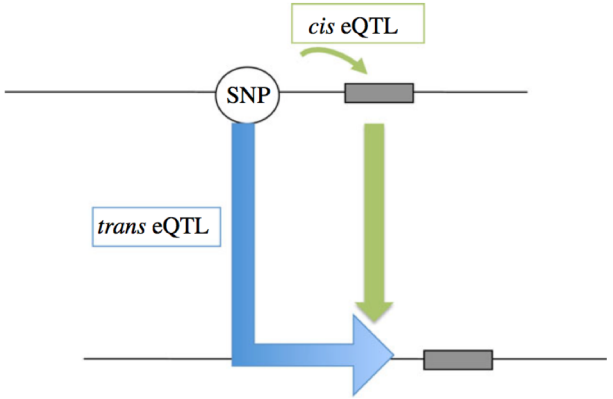
eQTL

- Analysis of genome function
- Discovery of candidate regulators
- **eQTL = expression Quantitative Trait Loci**: genomic loci that contributes to variation in expression levels of mRNAs
- Statistical associations
  - Genetic markers (SNPs)
  - Gene expression levels
  - Modification?
- cis/trans-eQTLs

eQTL



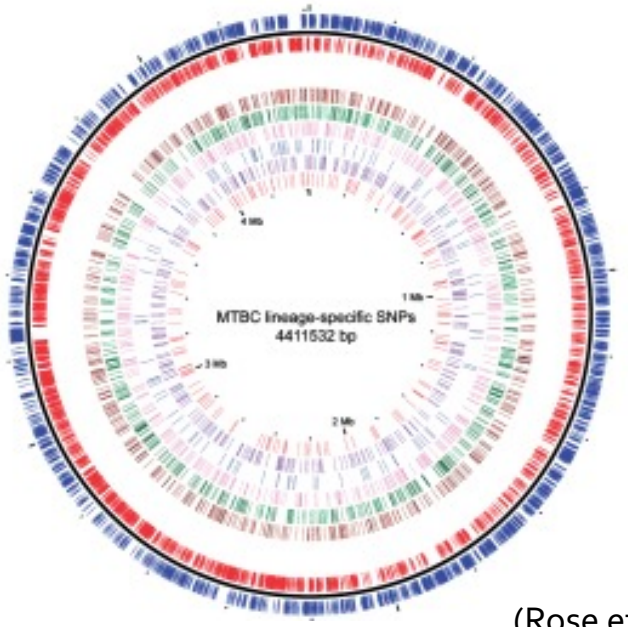




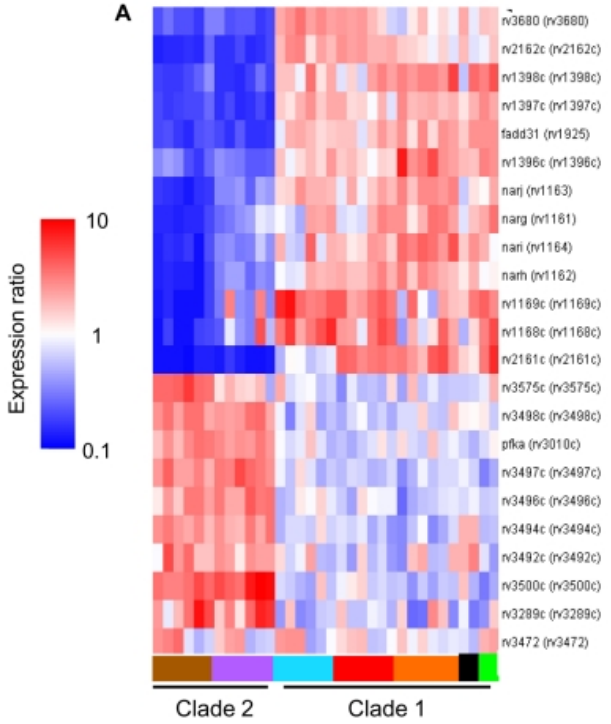
(Nika & Dermitzakis, 2013)

eQTL

- ***Why are eQTL studies interesting?***
- Mapping genotype-phenotype diversity
- i.e. Find associations in *Mtb*:
  - Lineage specific SNPs
  - Lineage specific transcriptomes

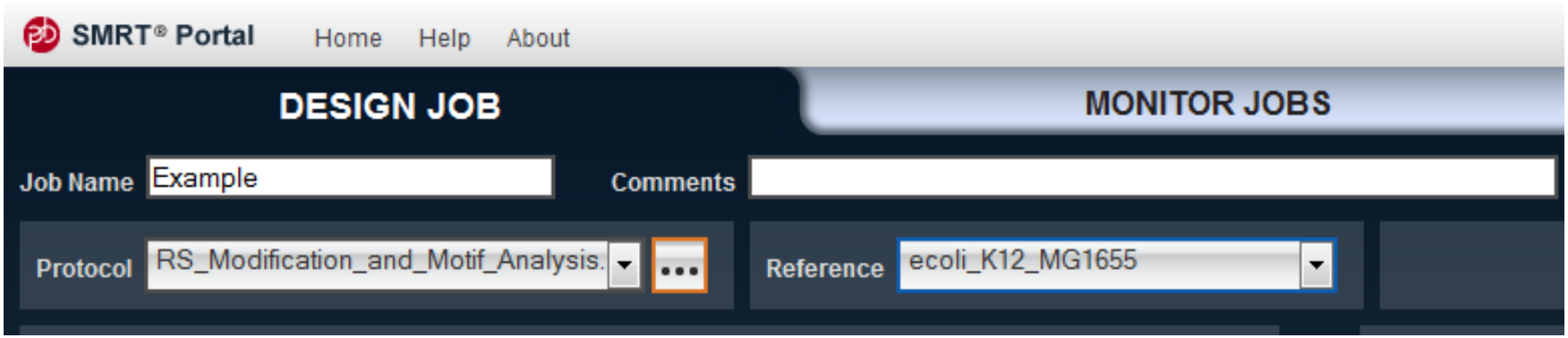


(Rose et al., 2013)

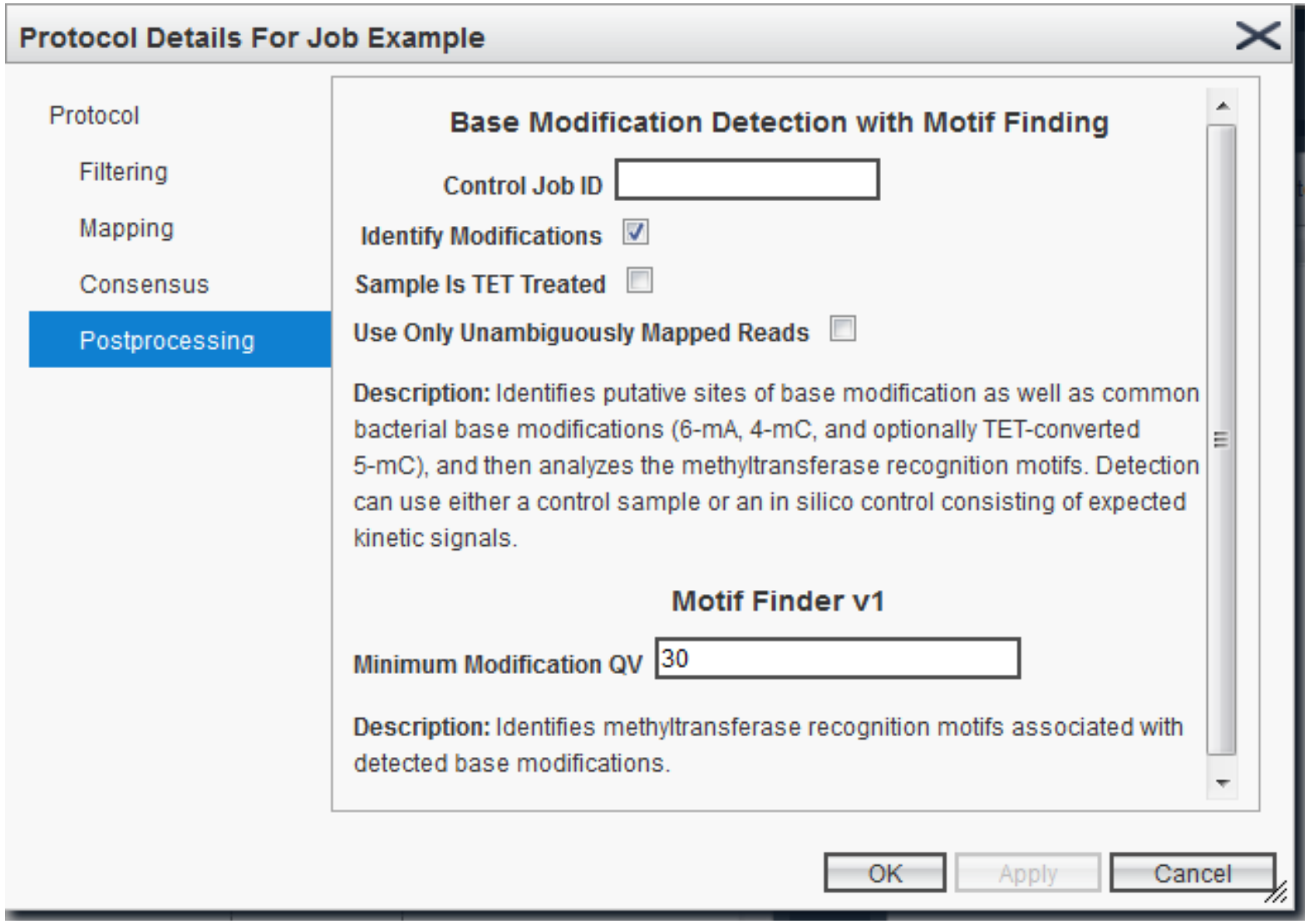


(Homolka et al., 2010)

SMRT Portal

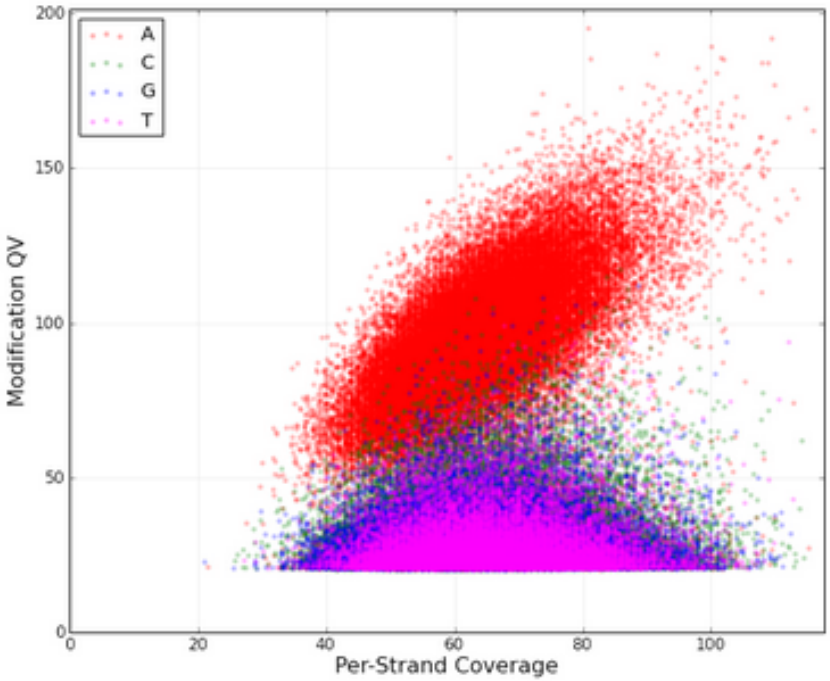

SMRT® Portal
Home
Help
About
DESIGN JOB
MONITOR JOBS
Job Name
Example
Comments
Protocol
RS_Modification_and_Motif_Analysis.
Reference
ecoli_K12_MG1655

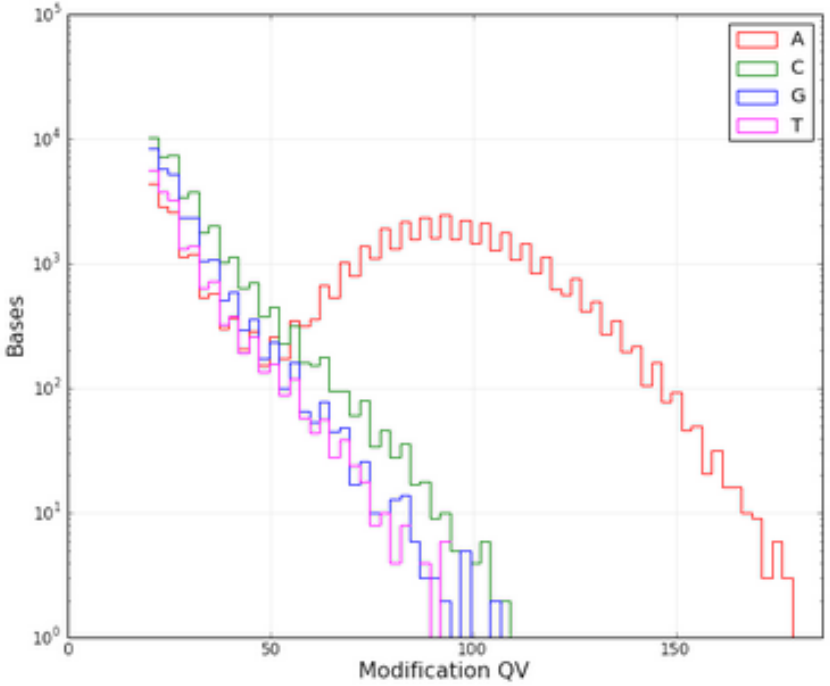
SMRT Portal


Protocol Details For Job Example
Protocol
Filtering
Mapping
Consensus
Postprocessing
Base Modification Detection with Motif Finding
Control Job ID
Identify Modifications
Sample Is TET Treated
Use Only Unambiguously Mapped Reads
Description: Identifies putative sites of base modification as well as common bacterial base modifications (6-mA, 4-mC, and optionally TET-converted 5-mC), and then analyzes the methyltransferase recognition motifs. Detection can use either a control sample or an in silico control consisting of expected kinetic signals.
Motif Finder v1
Minimum Modification QV
30
Description: Identifies methyltransferase recognition motifs associated with detected base modifications.
OK
Apply
Cancel

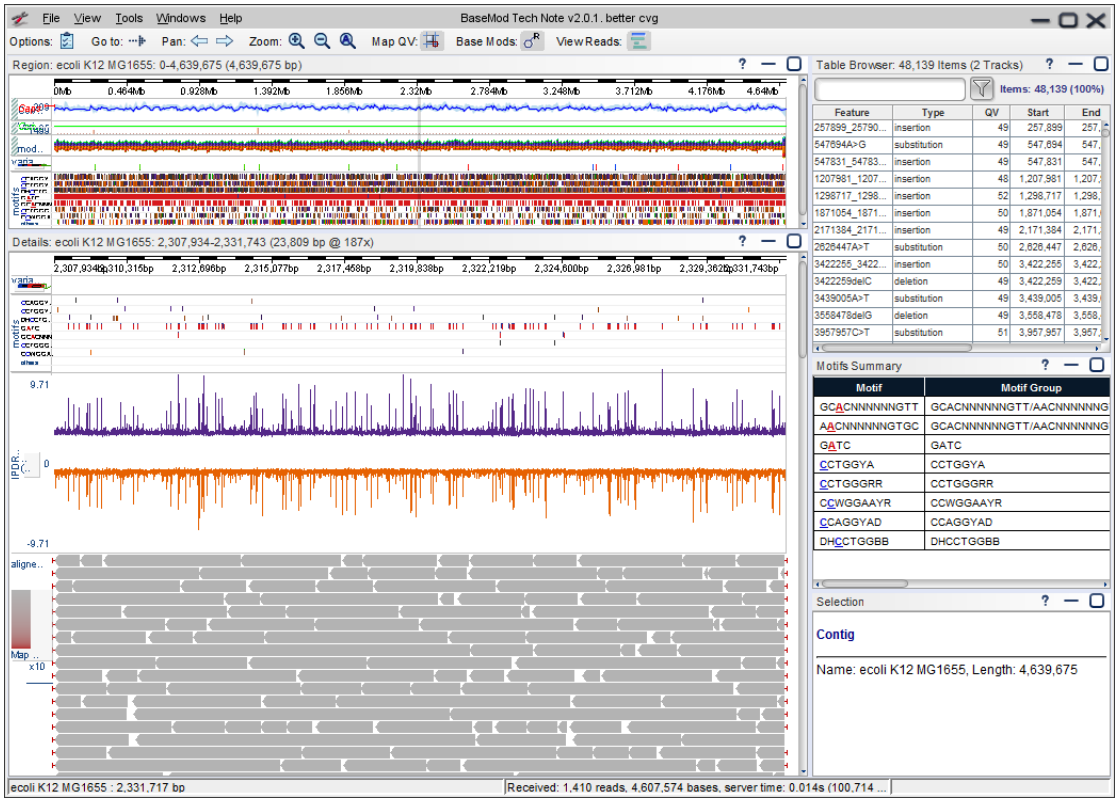
SMRT Portal

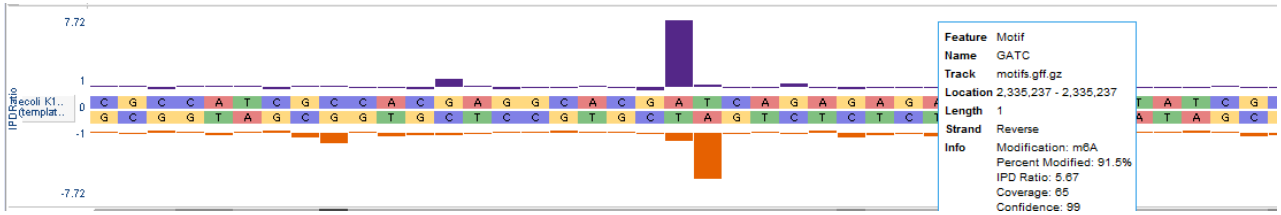
Kinetic Detections





SMRT Portal


File View Tools Windows Help
BaseMod Tech Note v2.0.1. better cvg
Options: Go to: Pan: Zoom: Map QV: Base Mods: View Reads:
Region: ecoli K12 MG1655: 0-4,639,675 (4,639,675 bp)
Details: ecoli K12 MG1655: 2,307,934-2,331,743 (23,809 bp @ 187x)
Table Browser: 48,139 Items (2 Tracks)
Items: 48,139 (100%)
Feature Type QV Start End
257899_25790... insertion 49 257,899 257,
547694A>G substitution 49 547,694 547,
547831_54783... insertion 49 547,831 547,
1207981_1207... insertion 48 1,207,981 1,207,
1298717_1298... insertion 52 1,298,717 1,298,
1871054_1871... insertion 50 1,871,054 1,871,
2171384_2171... insertion 49 2,171,384 2,171,
2626447A>T substitution 50 2,626,447 2,626,
3422255_3422... insertion 50 3,422,255 3,422,
3422259delC deletion 49 3,422,259 3,422,
3439005A>T substitution 49 3,439,005 3,439,
3558478delG deletion 49 3,558,478 3,558,
3957957C>T substitution 51 3,957,957 3,957,
Motifs Summary
Motif Motif Group
GCACNNNNNNGTT GCACNNNNNNGTT/AACNNNNNNG
AACNNNNNNGTGC GCACNNNNNNGTT/AACNNNNNNG
GATC GATC
CCTGGYA CCTGGYA
CCTGGGRR CCTGGGRR
CCWGGAAYR CCWGGAAYR
CCAGGYAD CCAGGYAD
DHCCTGGBB DHCCTGGBB
Selection
Contig
Name: ecoli K12 MG1655, Length: 4,639,675
ecoli K12 MG1655 : 2,331,717 bp
Received: 1,410 reads, 4,607,574 bases, server time: 0.014s (100,714 ...


7.72
1
0
-1
-7.72
IPDRatio
ecoli K1..
(templat..
Feature Motif
Name GATC
Track motifs.gff.gz
Location 2,335,237 - 2,335,237
Length 1
Strand Reverse
Info Modification: m6A
Percent Modified: 91.5%
IPD Ratio: 5.67
Coverage: 65
Confidence: 99

SMRT Portal

Motif	Modified Position	Modification Type	% Motifs Detected	# Of Motifs Detected	# Of Motifs In Genome	Mean Modification QV	Mean Motif Coverage	Partner Motif
GCACNNNNNNGTT	3	m6A	100.00	595	595	95.4	63.7	AACNNNNNNGTGC
AACNNNNNNGTGC	2	m6A	100.00	595	595	96.2	62.2	GCACNNNNNNGTT
GATC	2	m6A	99.91	38,205	38,240	102.6	64.1	GATC
CCTGGYA	1	unknown	54.32	936	1,723	45.5	65.6	
CCTGGGRR	1	unknown	39.39	169	429	43.5	67.0	
CCWGGAAYR	2	unknown	36.63	152	415	41.5	62.8	
CCAGGYAD	1	unknown	21.48	304	1,415	39.5	67.3	
DHCCTGGBB	3	unknown	19.61	747	3,809	40.1	66.4	
Not Clustered	0		0.15	14,150	9,232,129	37.3	66.2	

Modification QV Histogram By Motif