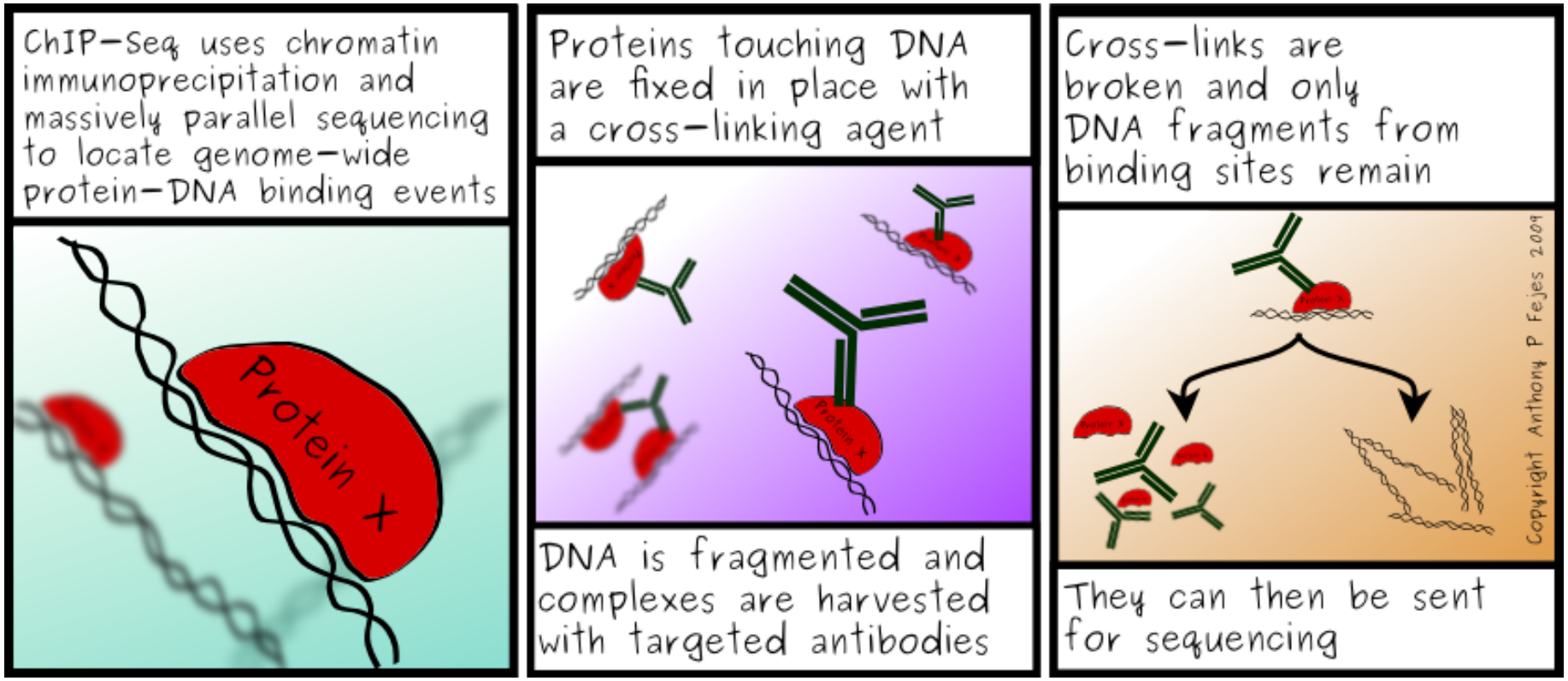
ChIP-seq

Mark Robinson

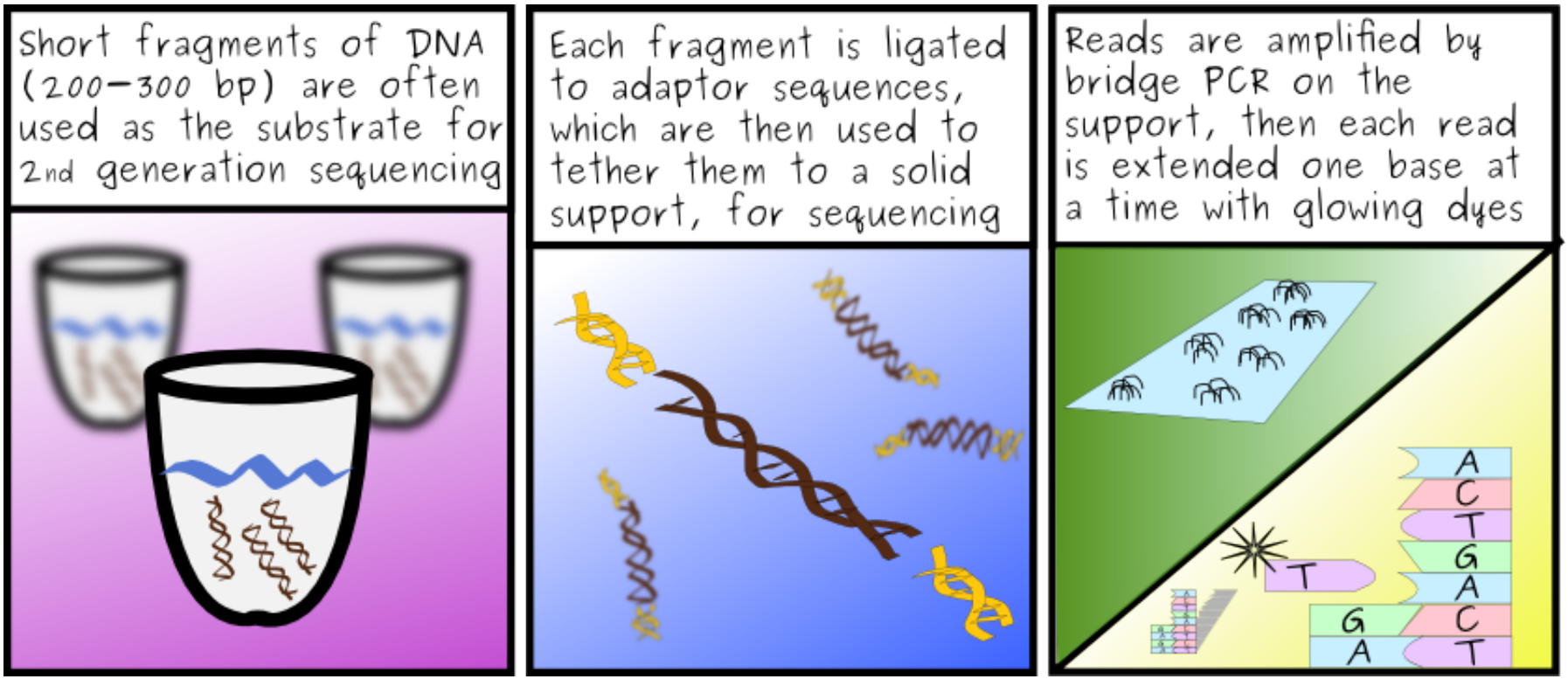
Department of Biochemistry and Molecular Biology,
Michigan State University

ChIP-seq: A method for determining protein-DNA interactions on a genome-wide scale



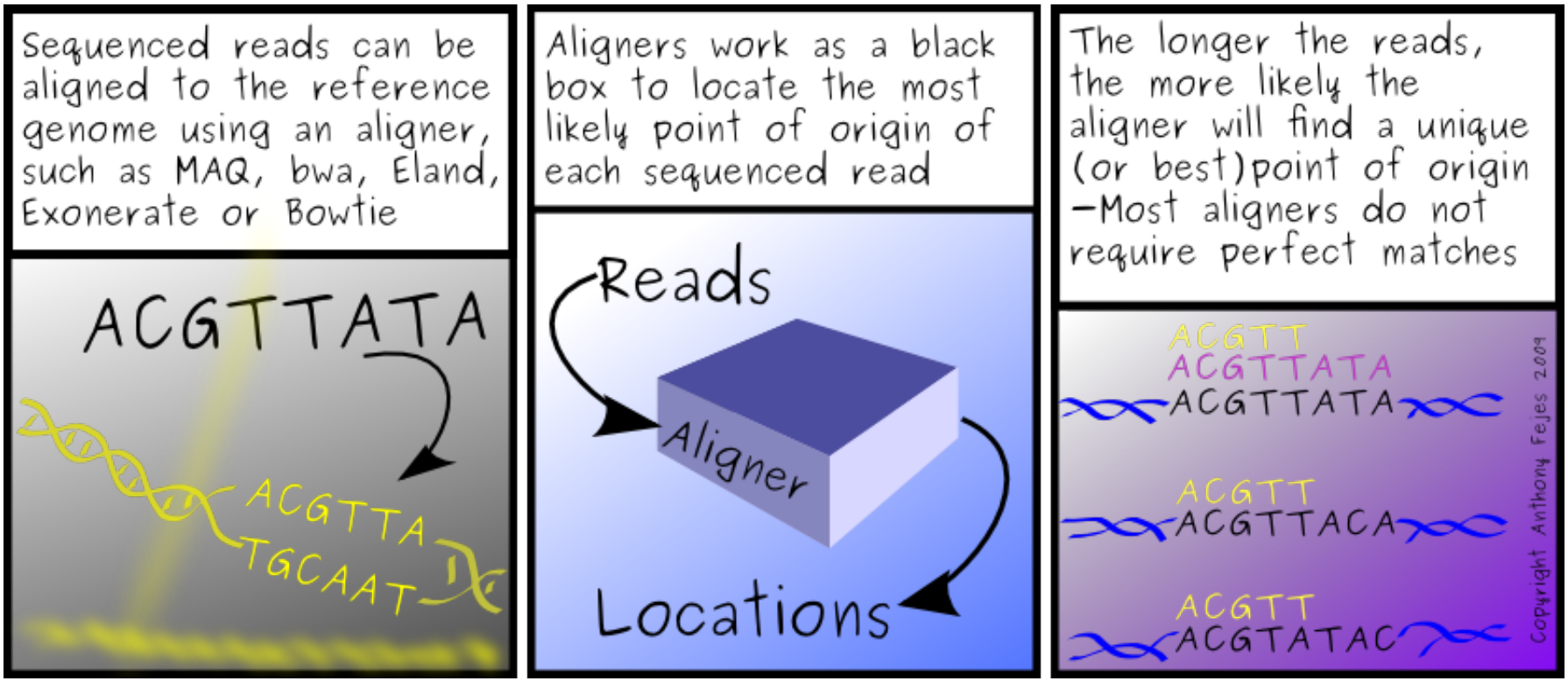
Anthony Fejes 2009

Next generation sequencing: Short read lengths with massive depth of sequencing



Anthony Fejes 2009

Alignment of reads to reference genome allows mapping of binding locations



Peak Detection software

- QuEST: http://mendel.stanford.edu/sidowlab/downloads/quest/

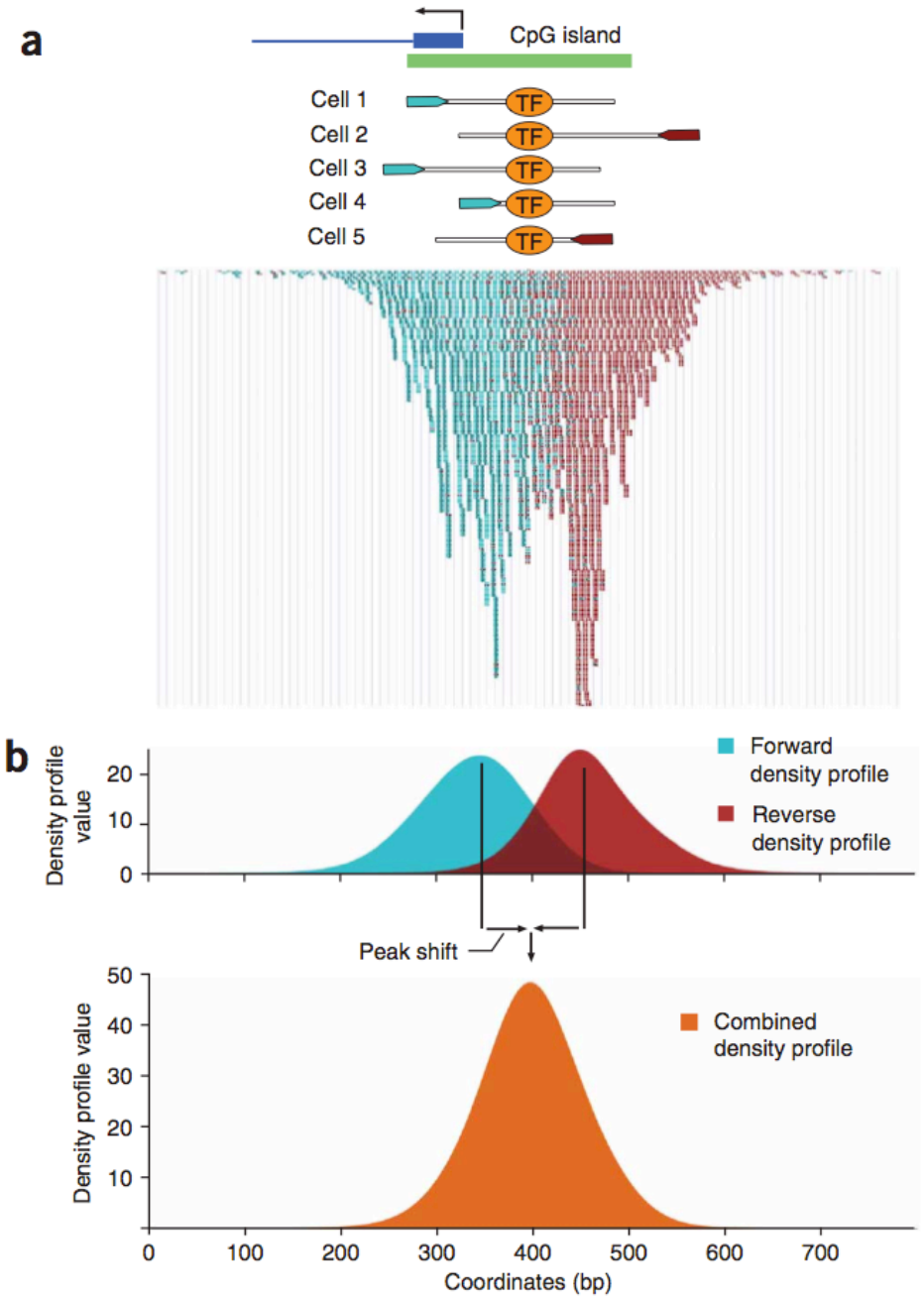
- Findpeaks: http://www.bcgsc.ca/platform/bioinfo/software/findpeaks

- ChIP-seq PeakFinder: http://woldlab.caltech.edu/html/chipseq_peak_finder

Correct handling of short reads

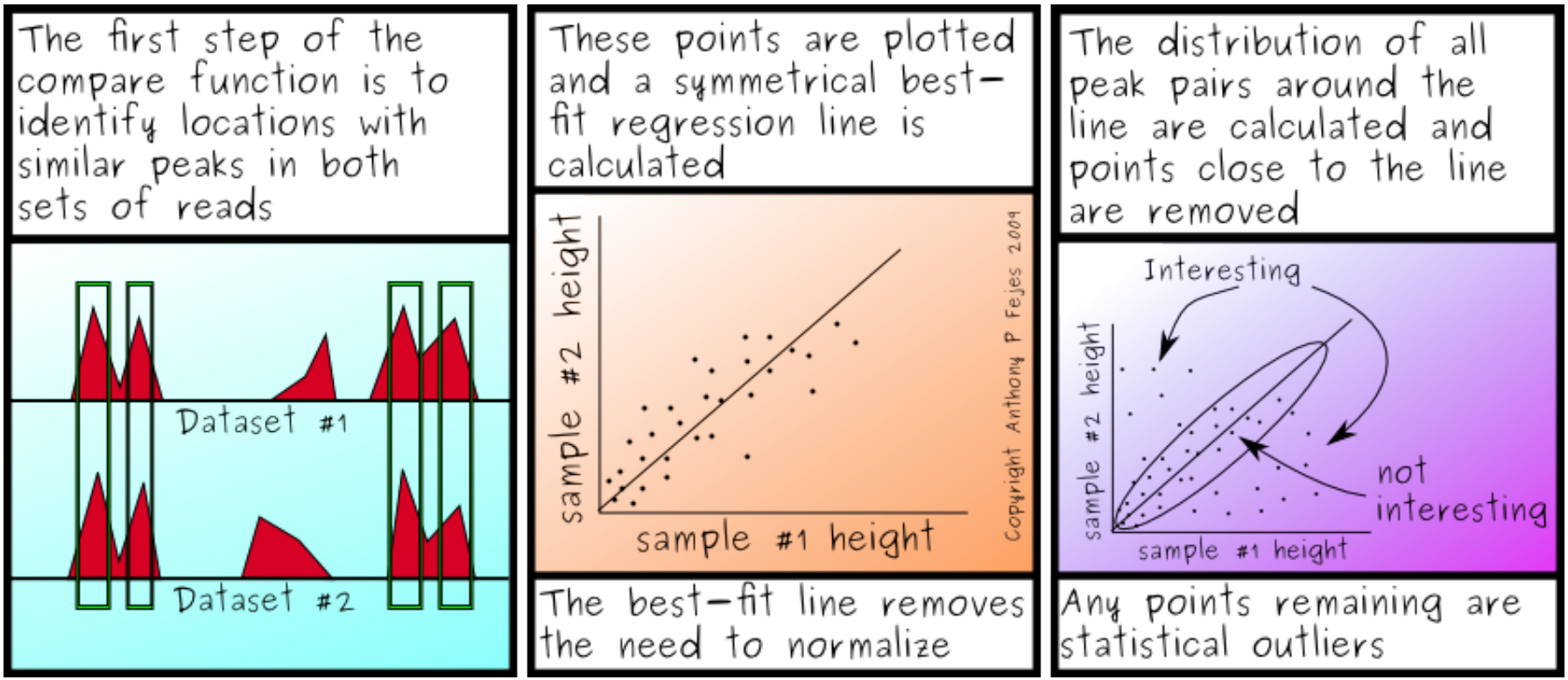
- Transcription factor binding sites are located somewhere on original DNA fragment
  - NOT the sequenced read

- Before attempting to determine binding site location this fact needs to be accounted for.

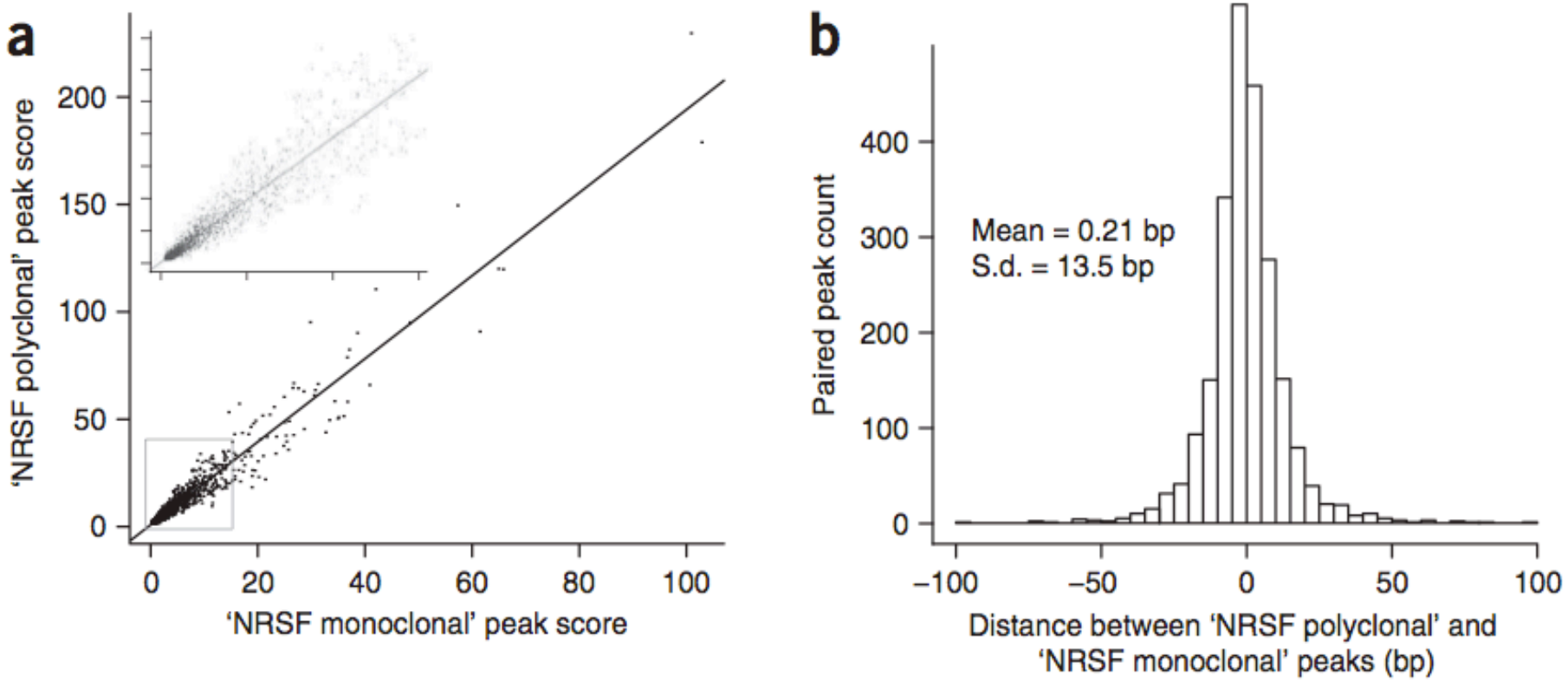


Valouev, A, et al.
Nature Methods, 2008 Sep; 5(9): 829-34

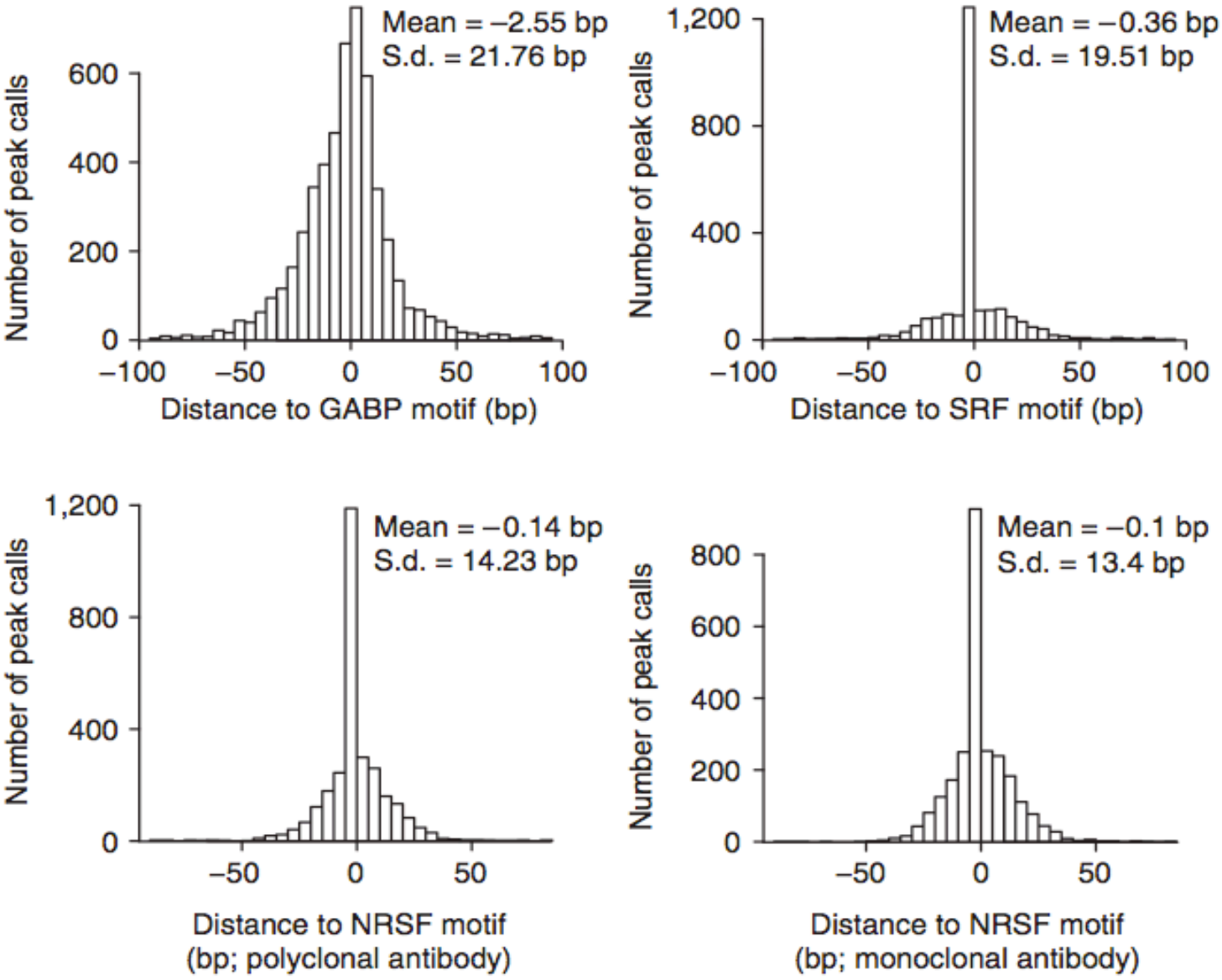
Normalization and Significance detection



Anthony Fejes 2009

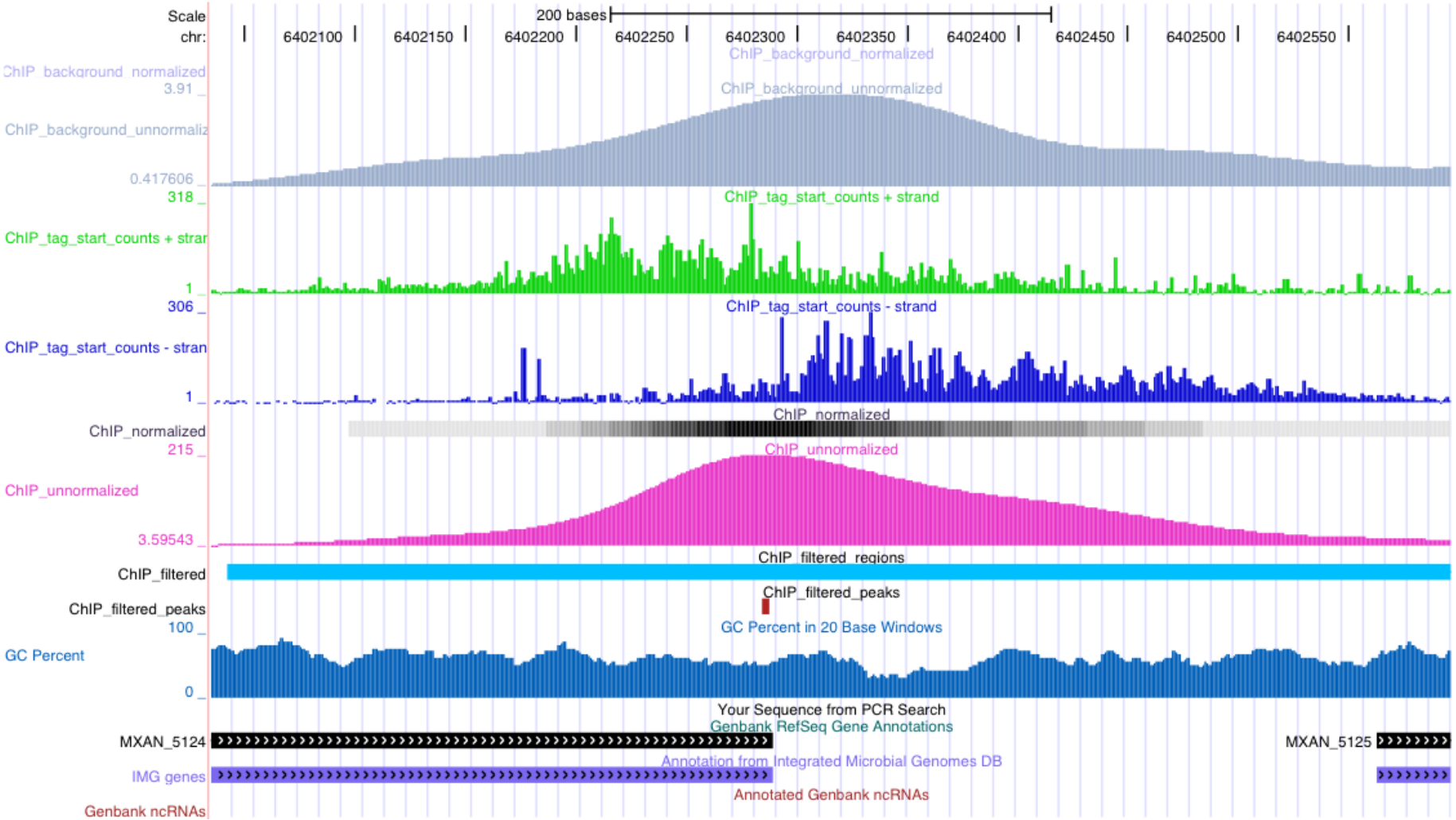


Valouev, A, et al.
Nature Methods, 2008 Sep; 5(9):
829-34

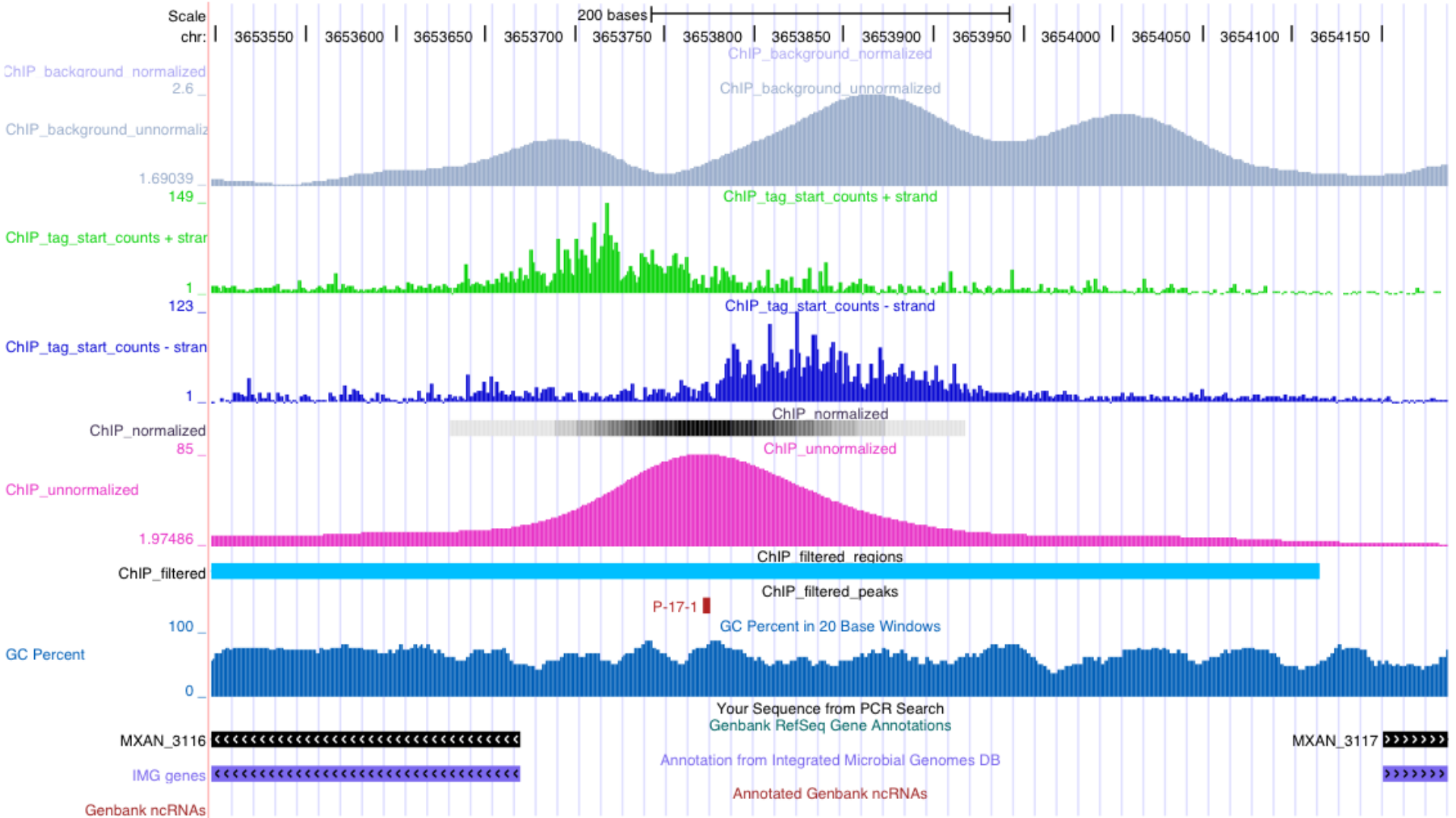


Valouev, A, et al.
Nature Methods, 2008 Sep; 5(9): 829-34

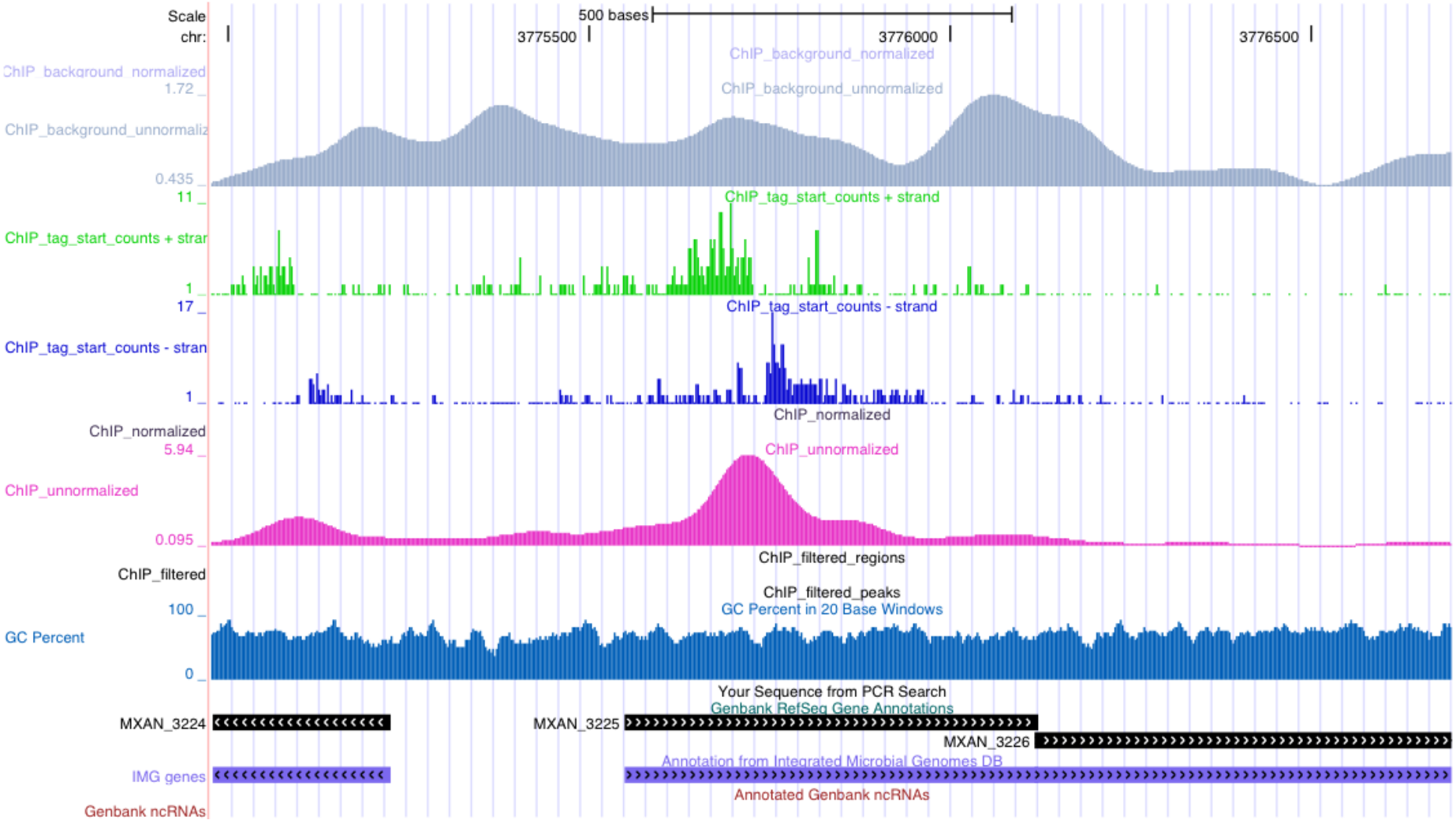
MrpC promoter (positive control)



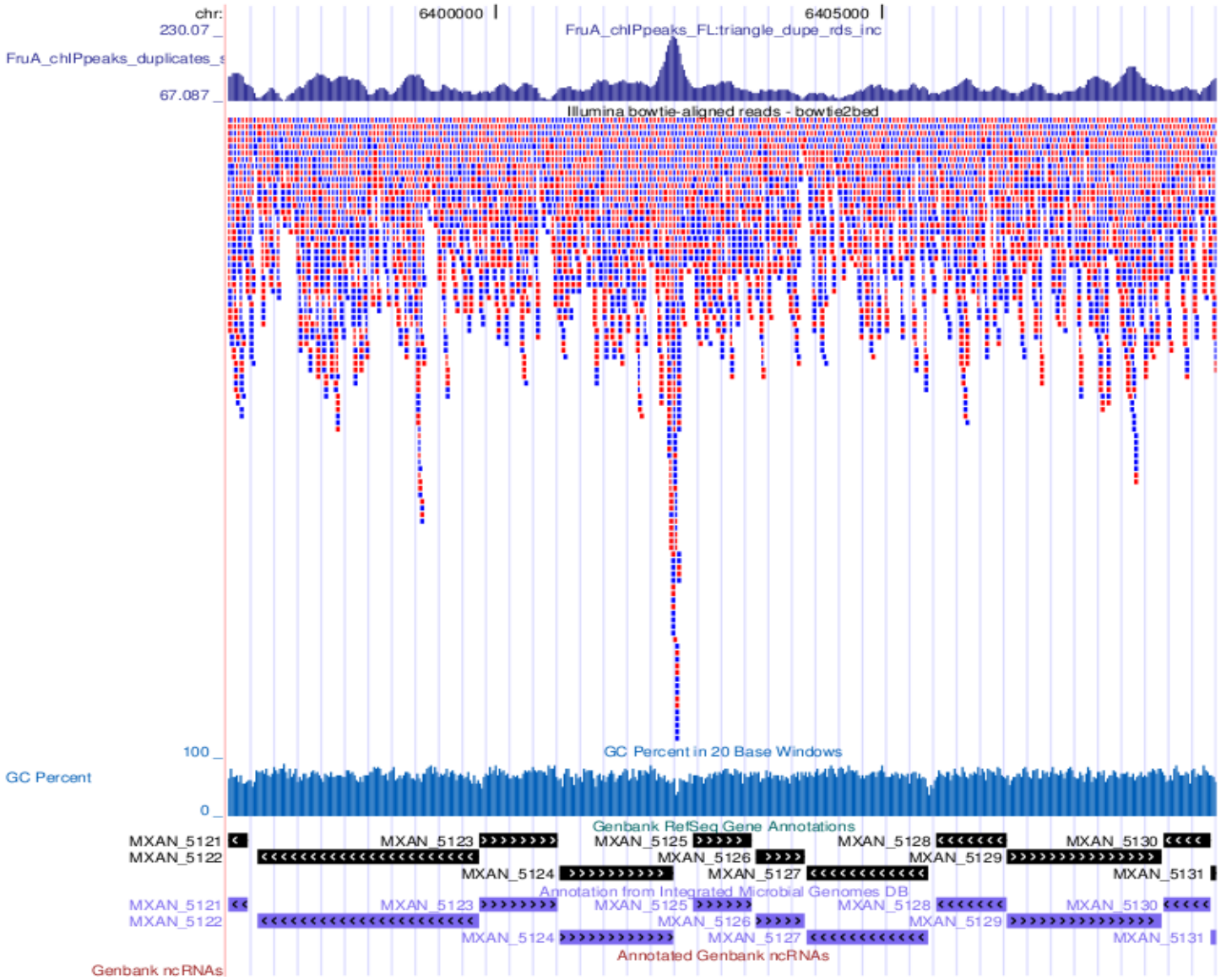
FruA promoter (positive control)



FdgA promoter (negative control)



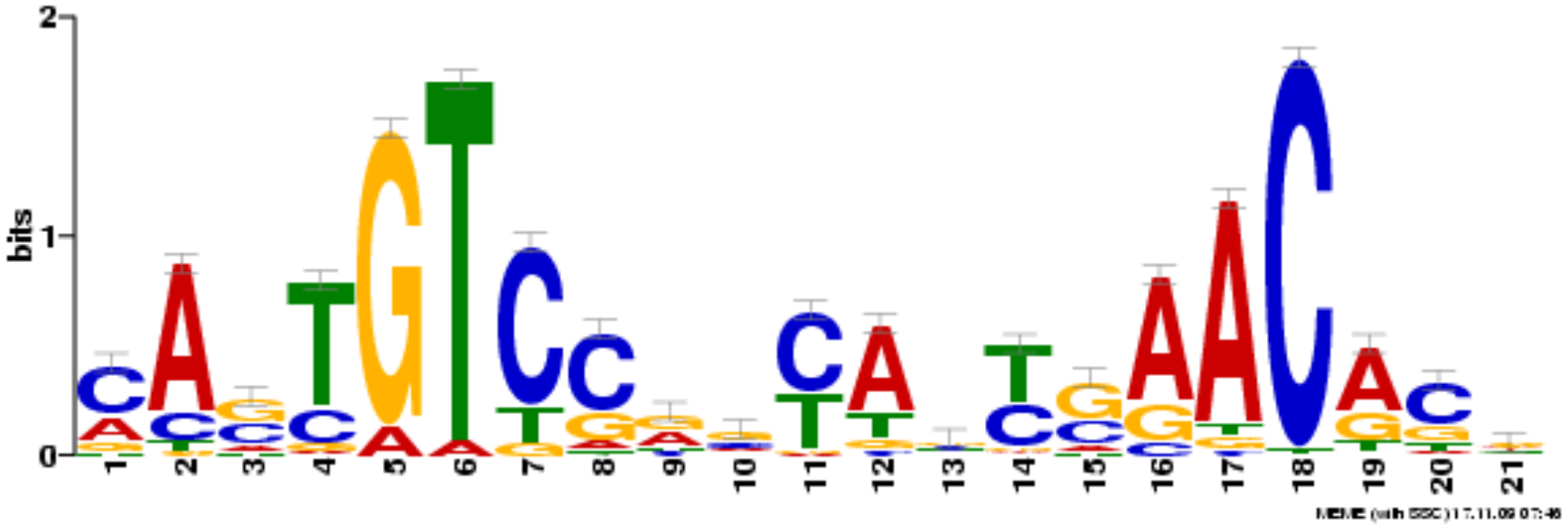
A cautionary tale!



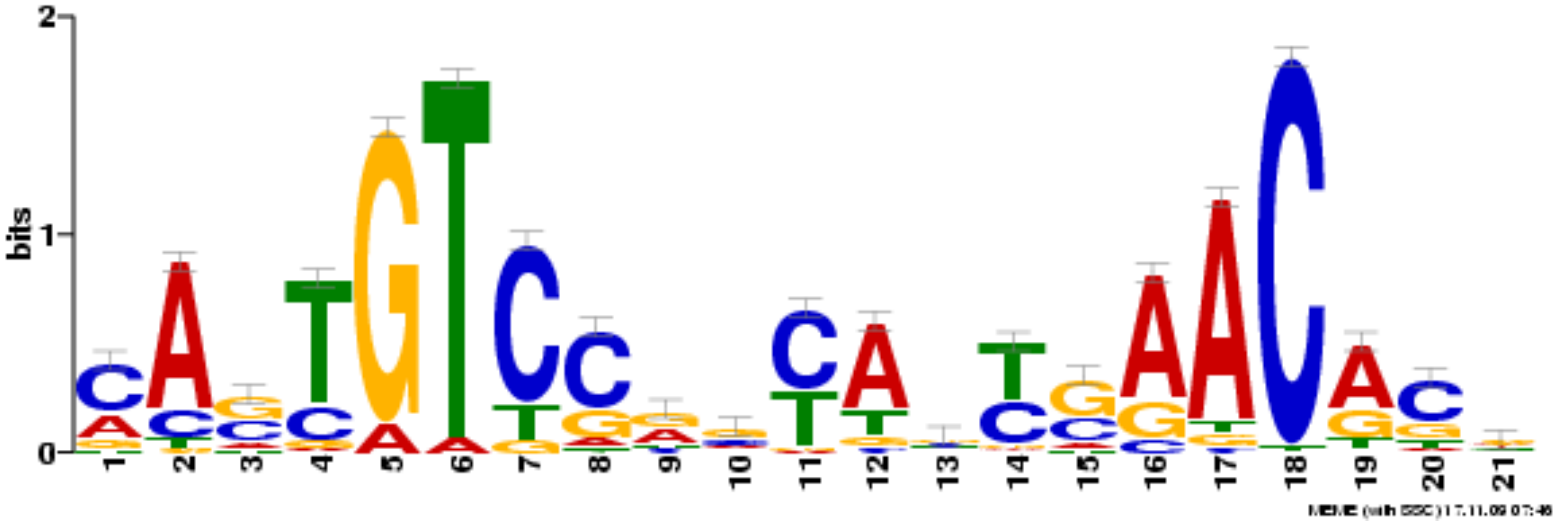
Motif searching

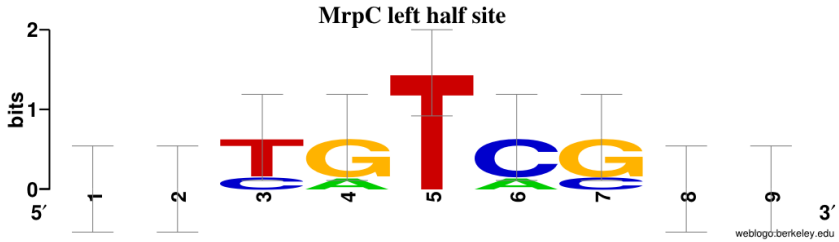
- Can we extract any common motifs from the regions immediately surrounding ChIP-enriched peaks?
  - Do any motifs found resemble MrpC binding sites?
  - Can we use this approach to determine an informative PWM for MrpC binding sites
  - Is such a PWM informative enough for genome wide searching?

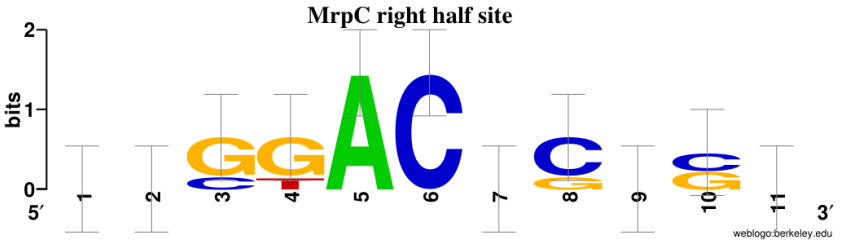
Most informative motif found in all 170 stringently called peaks



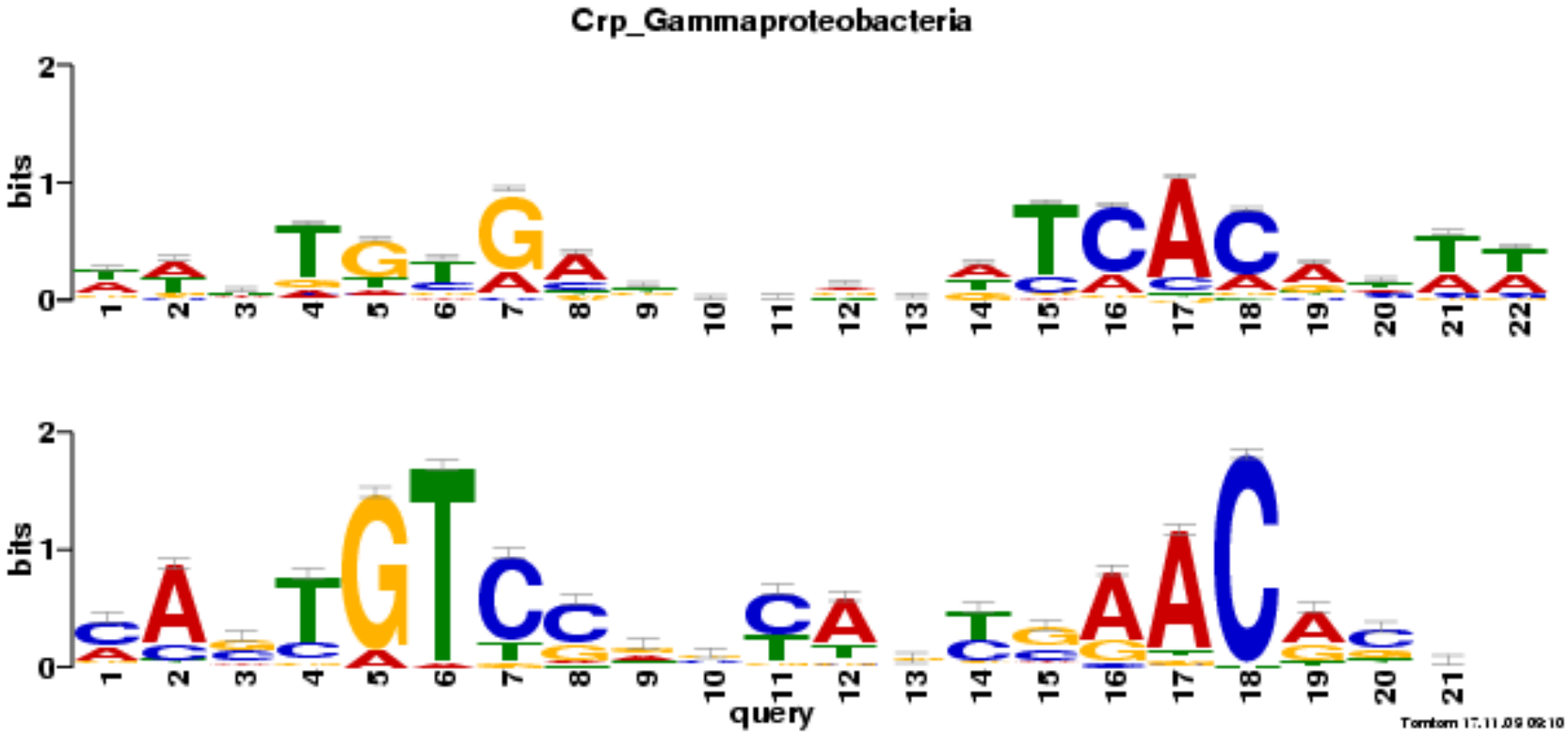
Motif closely resembles PWM constructed from known sites







Motif shows homology with Crp motif, a known homolog of MrpC

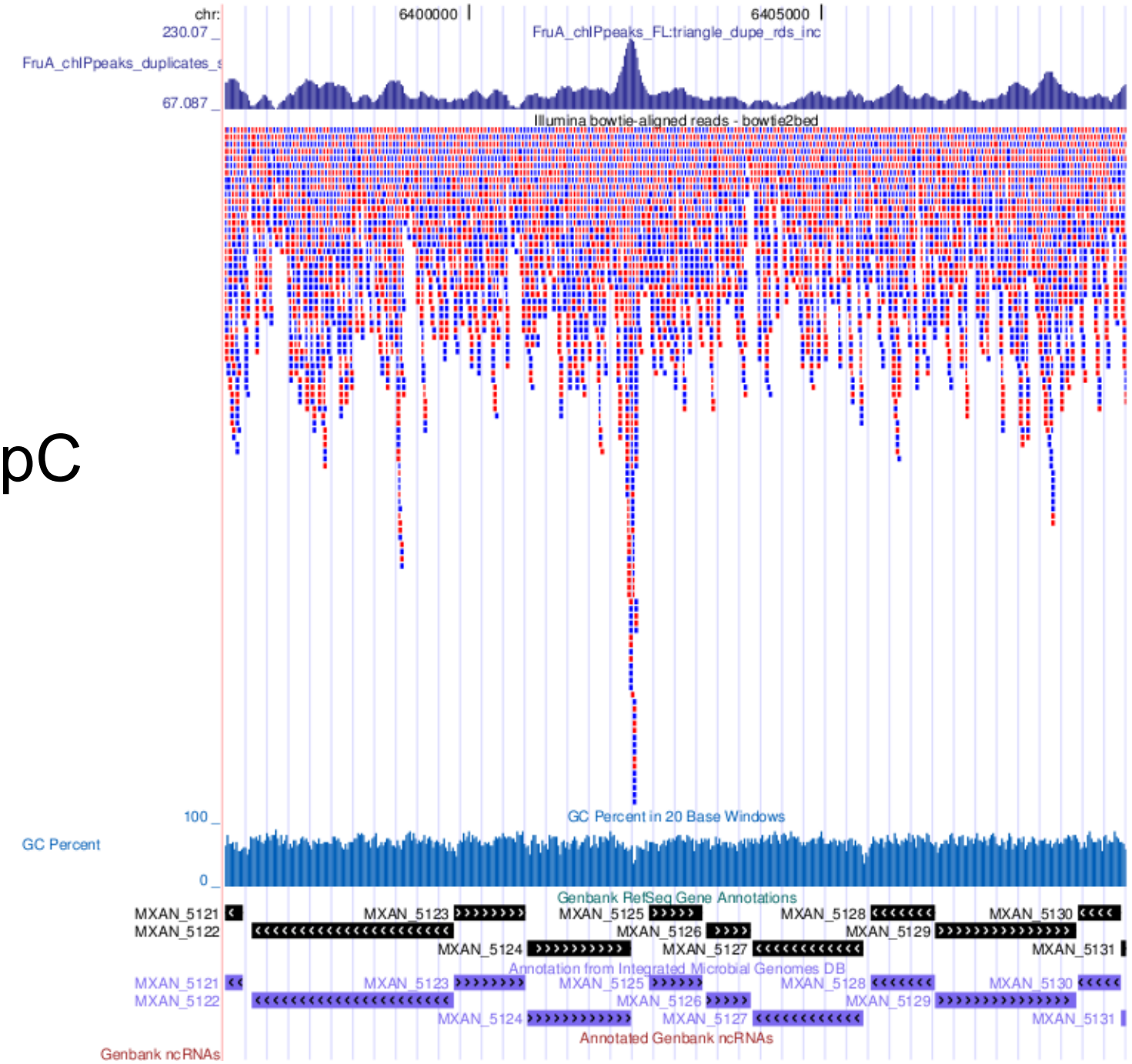


Genome-wide computational predictions

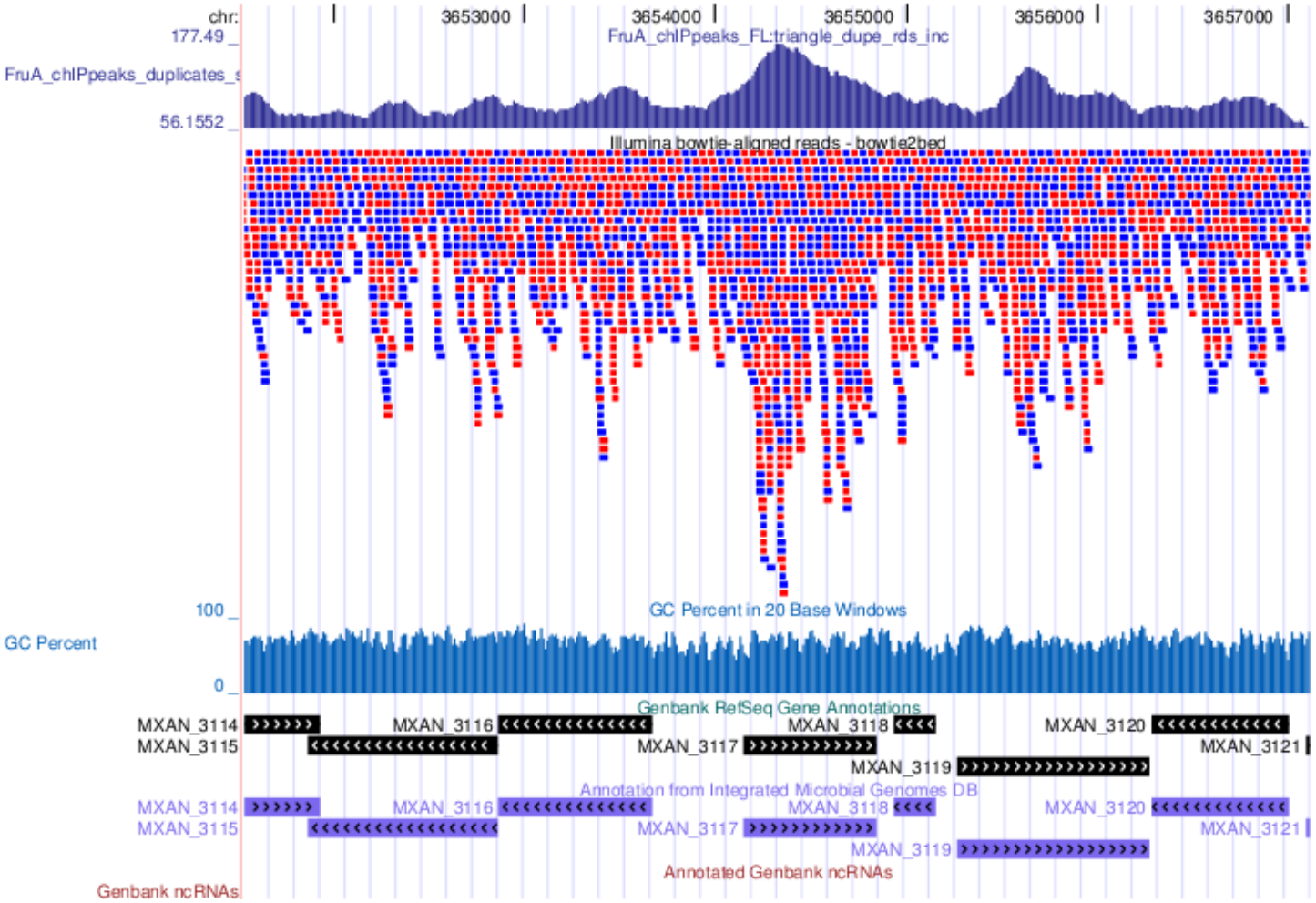
- FruA: ~ 26,000 hits genome-wide (exp: 22,000)
- MrpC: ~ 11,000 hits genome-wide (exp: 12,000)

- Co-located motifs in intergenic regions:
  - Expection genome wide: 120
  - Expection restricted to intergenic regions: 12
  - 38 motifs found genome-wide
  - MrpC promoter predicted to be bound by both FruA and MrpC2

MrpC


chr:
6400000
6405000
230.07
FruA_chIPpeaks_FL:triangle_dupe_rds_inc
FruA_chIPpeaks_duplicates_s
67.087
Illumina bowtie-aligned reads - bowtie2bed
100
GC Percent in 20 Base Windows
GC Percent
0
Genbank RefSeq Gene Annotations
MXAN_5121
MXAN_5122
MXAN_5123
MXAN_5124
MXAN_5125
MXAN_5126
MXAN_5127
MXAN_5128
MXAN_5129
MXAN_5130
MXAN_5131
Annotation from Integrated Microbial Genomes DB
Annotated Genbank ncRNAs
Genbank ncRNAs

FruA


chr:
3653000
3654000
3655000
3656000
3657000
177.49
FruA_chIPpeaks_FL:triangle_dupe_rds_inc
FruA_chIPpeaks_duplicates_s
56.1552
Illumina bowtie-aligned reads - bowtie2bed
100
GC Percent in 20 Base Windows
GC Percent
0
Genbank RefSeq Gene Annotations
MXAN_3114
MXAN_3115
MXAN_3116
MXAN_3117
MXAN_3118
MXAN_3119
MXAN_3120
MXAN_3121
Annotation from Integrated Microbial Genomes DB
MXAN_3114
MXAN_3115
MXAN_3116
MXAN_3117
MXAN_3118
MXAN_3119
MXAN_3120
MXAN_3121
Annotated Genbank ncRNAs
Genbank ncRNAs