BACTERIAL TRANSCRIPTION

EMBL 2010

Bioconductor Developer Meeting

Leonardo Collado Torres

Winter Genomics

OUTLINE

- Background information
  - Work team
  - Developer team
- Biology
- Goals
- Our work dynamic
- What we've done
- To do list

A DIVERSE WORK TEAM

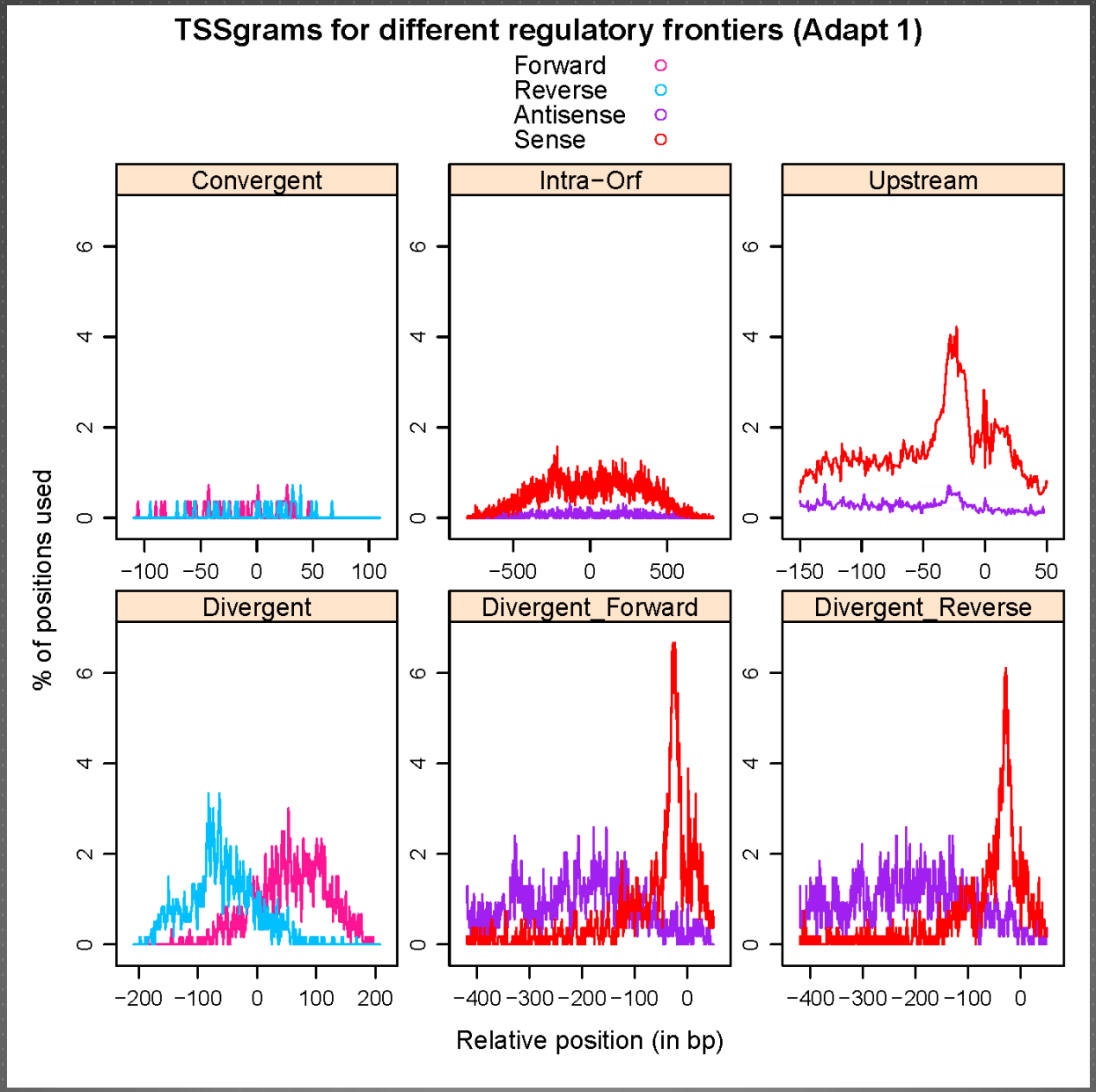
- A benchwork lab (Morett's at iBT UNAM)
  - Developer of new transcription start sites mapping techniques and maintainer of the UUSMD (local seq. facility)
- A bioinformatics lab (Collado-Vides's at CCG UNAM)
  - Transcriptional bacterial regulation and maintainer of RegulonDB
- A new bioinformatics company (Winter Genomics)
  - New high throughput sequencing bioinformatics service company
- Undergraduate Program on Genomic Sciences (LCG) UNAM
  - All of the developers come from this program (graduated and current students)

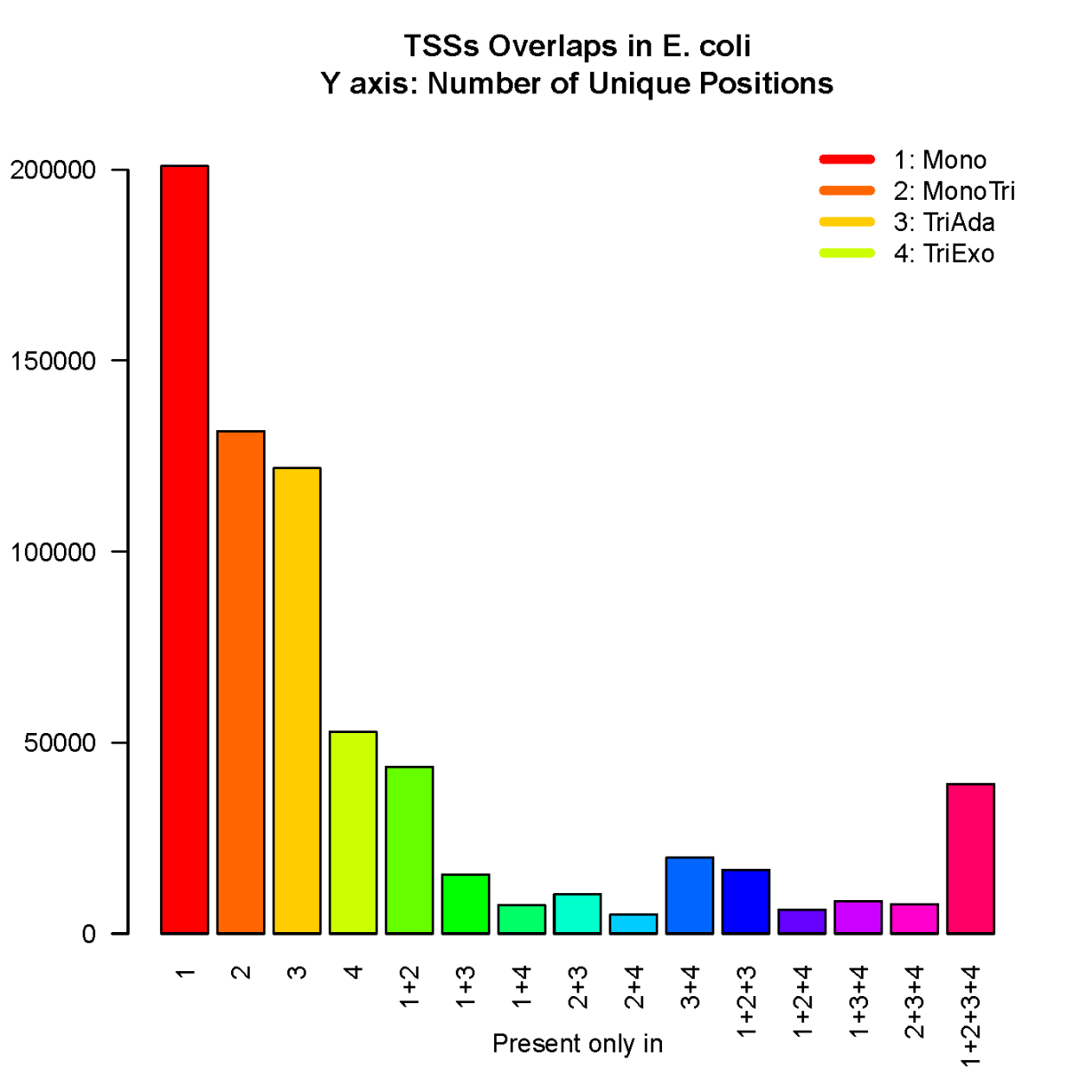
DYNAMIC DEVELOPER TEAM

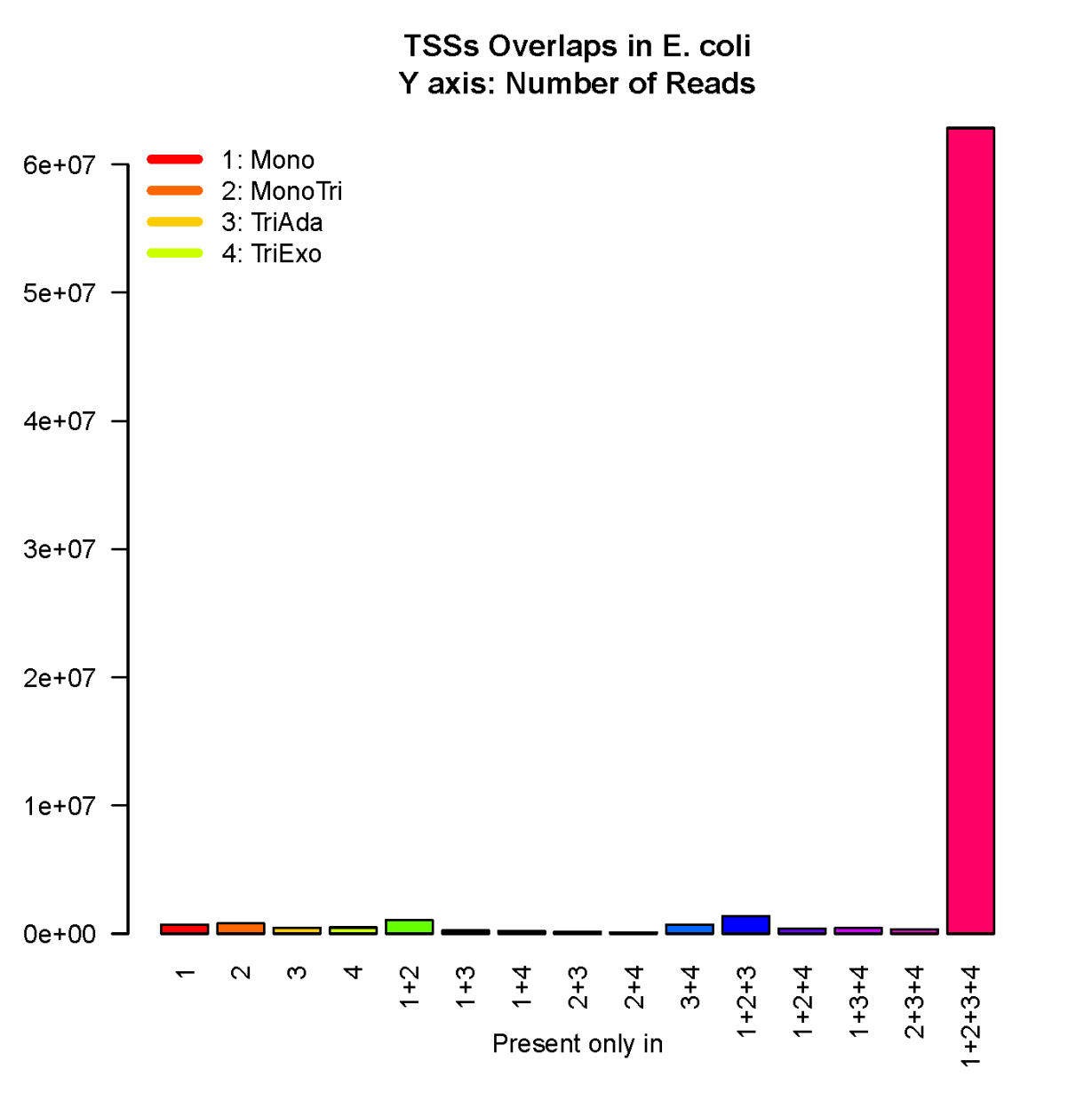
- April – June 2010
  - (5th) Alejandro Reyes Quiroz
  - (5th) Victor Moreno Mayar
  - (5th) Gabriel Cuellar Partida
- Aug 2010 – currently
  - (3rd) Carlos Vargas Chavez
  - (6th) Melvin Noe Gonzalez
  - (6th) Mayela Soto
  - (6th) Daniela Garcia Sorano
- Other programmers
  - Veronica Jimenez Jacinto
  - Leticia Vega Alvarado
  - Blanca Taboada

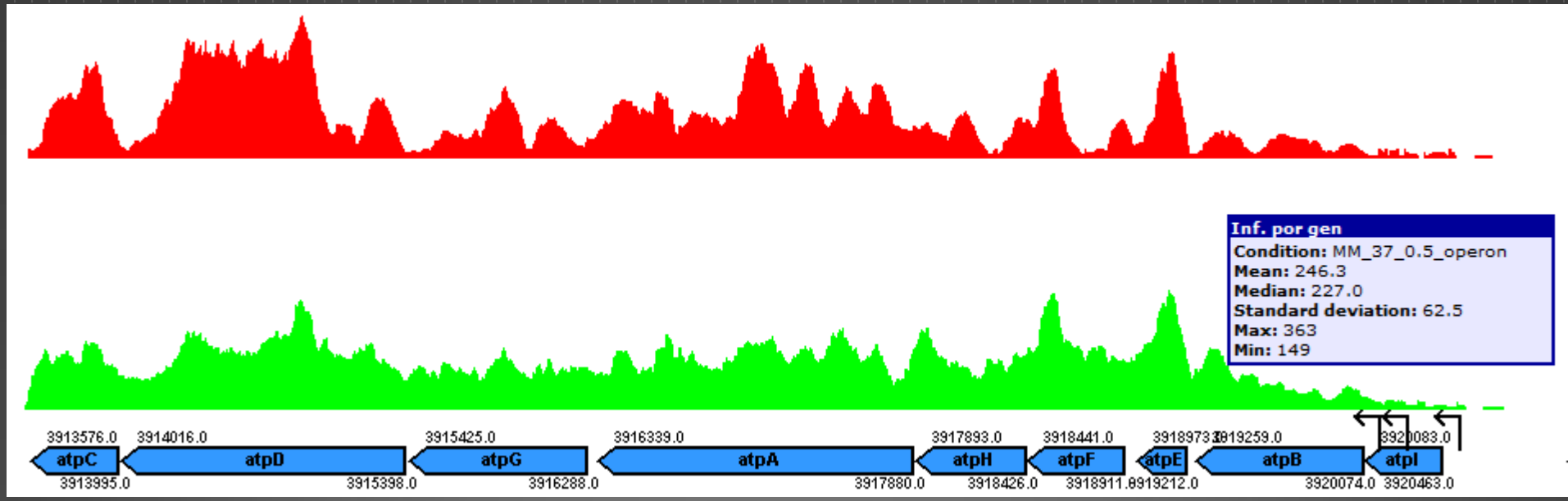
BIOLOGY

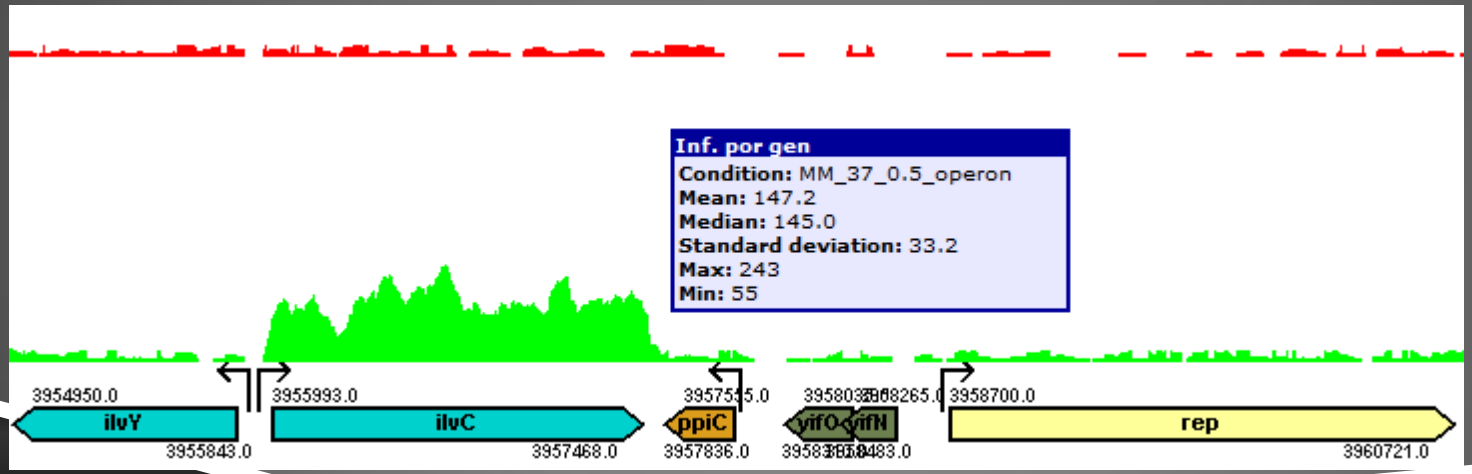
- Understand the transcriptional landscape at the genomic level
- Transcription Start Sites (TSSs)
  - Sites where the mRNA begins its transcription
  - Identify all active TSSs in a given condition
  - Unexpectedly high variability!
- Transcription Units (TUs)
  - One or multiple genes transcribed in the mRNA
  - Overlapping genes lead to complex cases!
- TSSs vs TUs correspondence

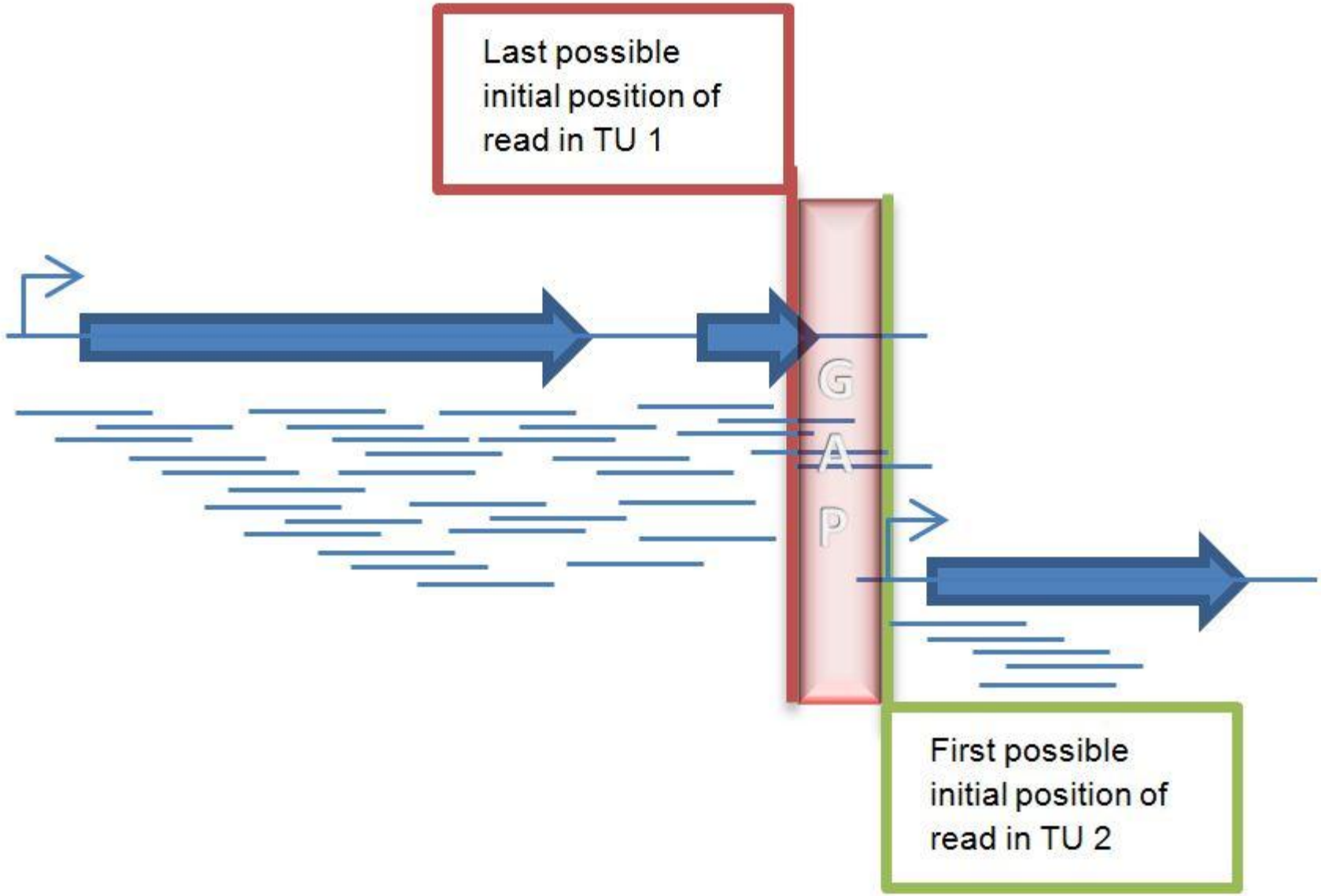

TSSgrams for different regulatory frontiers (Adapt 1)
Forward
Reverse
Antisense
Sense
Convergent
Intra−Orf
Upstream
Divergent
Divergent_Forward
Divergent_Reverse
% of positions used
Relative position (in bp)

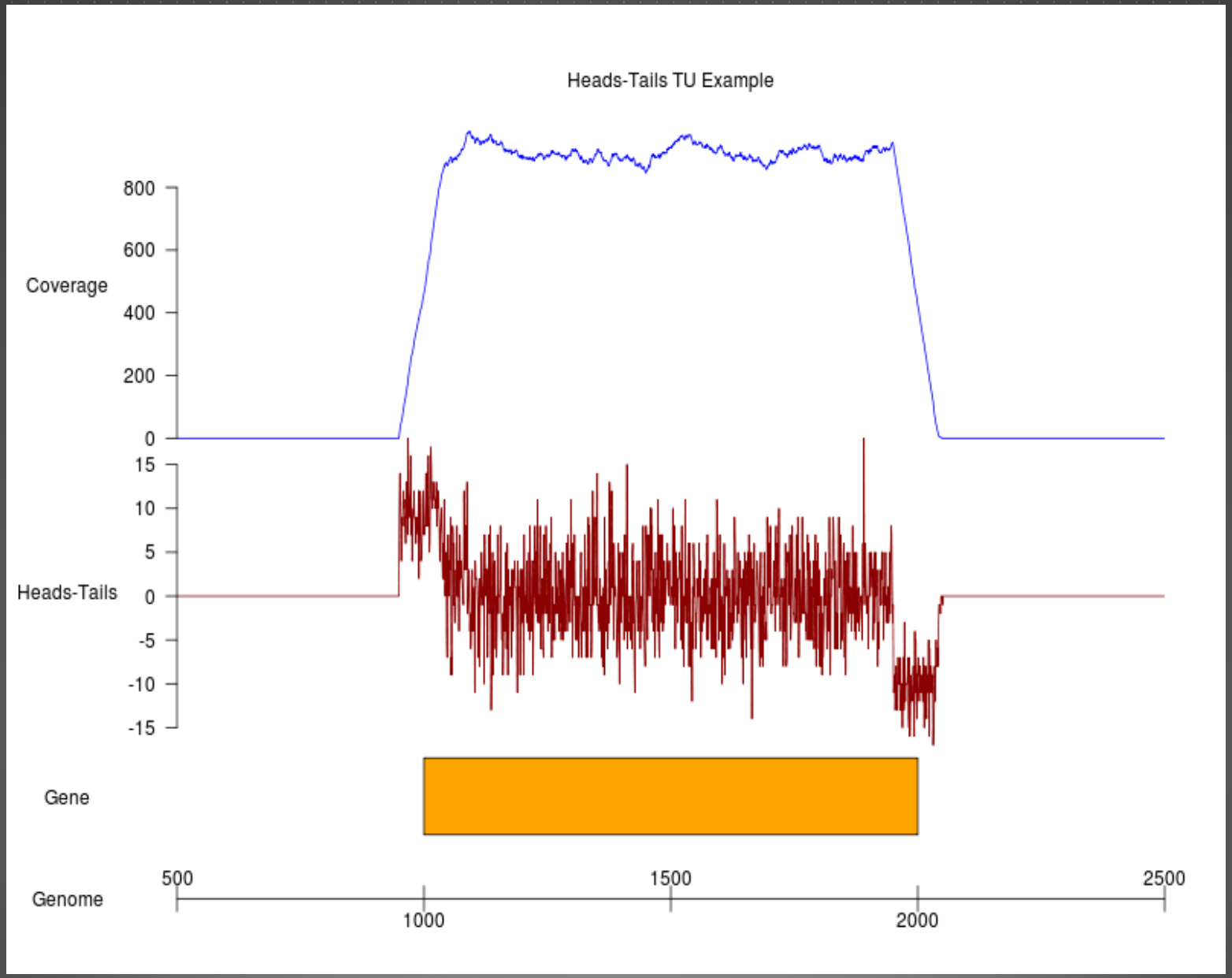

TSSs Overlaps in E. coli
Y axis: Number of Unique Positions
1: Mono
2: MonoTri
3: TriAda
4: TriExo
200000
150000
100000
50000
0
1
2
3
4
1+2
1+3
1+4
2+3
2+4
3+4
1+2+3
1+2+4
1+3+4
2+3+4
1+2+3+4
Present only in

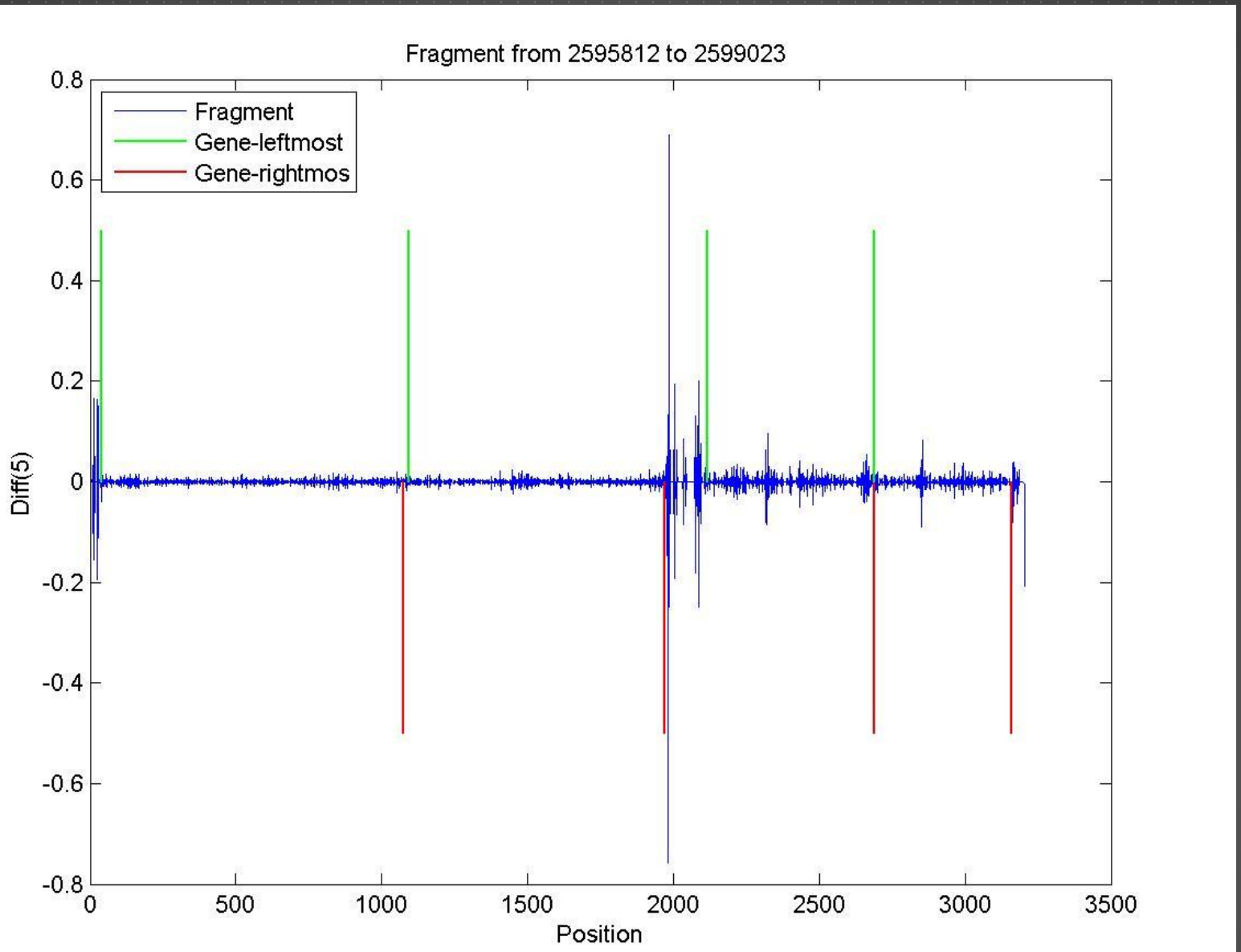

TSSs Overlaps in E. coli
Y axis: Number of Reads
1: Mono
2: MonoTri
3: TriAda
4: TriExo
6e+07
5e+07
4e+07
3e+07
2e+07
1e+07
0e+00
1
2
3
4
1+2
1+3
1+4
2+3
2+4
3+4
1+2+3
1+2+4
1+3+4
2+3+4
1+2+3+4
Present only in


Inf. por gen
Condition: MM_37_0.5_operon
Mean: 246.3
Median: 227.0
Standard deviation: 62.5
Max: 363
Min: 149
3913576.0
3914016.0
3915425.0
3916339.0
3917893.0
3918441.0
3919259.0
3920083.0
atpC
atpD
atpG
atpA
atpH
atpF
atpE
atpB
atpI
3913995.0
3915398.0
3916288.0
3917880.0
3918426.0
3918911.8919212.0
3920074.0
3920463.0


Inf. por gen
Condition: MM_37_0.5_operon
Mean: 147.2
Median: 145.0
Standard deviation: 33.2
Max: 243
Min: 55
3954950.0
3955993.0
3957555.0
3958700.0
ilvY
ilvC
ppiC
rep
3955843.0
3957468.0
3957836.0
3960721.0


Last possible initial position of read in TU 1
G
A
P
First possible initial position of read in TU 2


Heads-Tails TU Example
Coverage
800
600
400
200
0
Heads-Tails
15
10
5
0
-5
-10
-15
Gene
Genome
500
1000
1500
2000
2500


Fragment from 2595812 to 2599023
Fragment
Gene-leftmost
Gene-rightmos
Diff(5)
Position
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
500
1000
1500
2000
2500
3000
3500

PROJECT GOALS

- Guarantee reproducibility
- Complete proposal on how to analyze this kind of data
  - Including working software!
- Facilitate future similar analyses from other bacteria
- Create easy (straight forward) to use software

- Learn more about BioC

WHAT WE'VE DONE SO FAR

- A pile of ideas ☺
- Most of the code for the TSSs is ready
  - Granges
  - List (up to 3) of SimpleRleList
- Prototypes for the 3 TU methods
- Summary information
  - GRanges // data.frame
  - Plots mostly using lattice
- Trained undergrads in R / BioC ^_^

TO DO

- Define how to evaluate the TU methods
- Evaluate them

- DOCUMENTATION!

- Feedback on objects that would be less prone to being broken by users
- Check the SummarizedExperiment class

- Aim: getting done prior to the next release

LINKS

- Morett's lab http://www.ibt.unam.mx/server/PRG.base?tipo:doc,dir:PRG.grupo,par:Gem,tit:_Grupo_del__Dr._Juan_Enrique_Morett
- Collado-Vides' lab http://www.ccg.unam.mx/en/ComputationalGenomics
- Winter Genomics http://www.wintergenomics.com/
- UUSMD http://uusmd.unam.mx/
- LCG http://www.lcg.unam.mx/