The Genomic HyperBrowser
Exploring the borders of the galaxy

Geir Kjetil Sandve, University of Oslo, 26.may 2011
The Genomic HyperBrowser
Exploring the borders of the galaxy
Since 2007..
Outline

- The Genomic HyperBrowser
- Life at the borders of the Galaxy
- Why we like Galaxy
Outline

• The Genomic HyperBrowser
• Life at the borders of the Galaxy
• Why we like Galaxy
The Genomic HyperBrowser

Why?
Genomic information is becoming plentiful.
But analysis lags behind
Enter:
The Genomic hyperbrowser

- Do two tracks at a time, but robustly and comprehensively
- Emphasize local analysis
- Each result is in fact a new track
The Genomic HyperBrowser

What?
You asked:

Are 'MEFB1 (BLOC segments)' overlapping 'SINE (Repeating elements)', more than chance?

Simplistic answer:

No support from data for this conclusion in any bin

Precise answer:

0 significant bins out of 19, at 10% FDR*

A collection of FDR-corrected p-values per bin was computed. Not able to compute a global p-value for this analysis.

* False Discovery Rate: The expected proportion of false positive results among the significant bins is no more than 10%.

In each bin, the test of

H0: The segments of track 1 are located independently of the segments of track 2 with respect to overlap

vs

H1: The segments of track 1 tend to overlap the segments of track 2

was performed.

P-values were computed under the null model defined by the following preservation and randomization rules:

Preserve segments of T2, segment and inter-segment lengths of T1, randomize positions (MC)

The test statistic used is:

The number of base pairs that are inside segments of both tracks
Do 'TF binding sites' accumulate more towards the borders of 'Introns'?
Requirements for interactivity

"ls" in file hierarchy

≈ 100 runs on-the-fly

=> Job scheduling
The Genomic HyperBrowser

The power of for-loops
AP2A vs Melanoma

Unmarked points

Marked segments

\textbf{in case} > \textbf{in control}?
All TFs vs Melanoma

AP2ALPHA
E2F
CREBP1
CREB
NFKAPPAB
CREB
ZIC2
CREBP1CJUN
CREL

....
All TFs vs all diseases

[multiply previous slide by 1068]
The disease regulome
The disease regulome
The disease regulome

- Generating hypotheses on the regulation of disease
- ... but also an interactive machine for generating such maps
Outline

- The Genomic HyperBrowser
- Life at the borders of the Galaxy
- Why we like Galaxy
Galaxy selling points (for developers)

- Stop wasting time writing interfaces
- Get your tools used by biologists
Galaxy selling points
(for developers)

- Stop wasting time writing interfaces
- Already had GUI, and still partly external
- Get your tools used by biologists
Galaxy selling points (for developers)

- Stop wasting time writing interfaces
  - Already had GUI, and still partly external
- Get your tools used by biologists
  - Not distributed anyway (have our own server)
Life at the borders of the Galaxy

- Separate, monolithic codebase
- Separate GUI
- Separate data collection
- Separate results files
So, how come we still like Galaxy?

- Web server, user handling, job scheduling
- History is indeed powerful
- With time, we added 20 supporting tools..
- With time, we now consider distribution..
Looking beyond our ego..

- The HyperBrowser can't solve all problems
  - .. but Galaxy can!?
- Working towards active national installation
- Use webtools for internal code sharing
The team

Eivind Tøstesen

Sigve Nakken

Trevor Clancy

Fang Liu

Sveinung Gundersen

Ingrid K. Glad

Arnoldo Frigessi

Halfdan Rydebeck

Geir Kjetil Sandve

Lars Holden

Knut Liestøl

Eivind Hovig

Morten Johansen

Marit Holden

Vegard Nygaard

Egil Ferkingstad

Geir Kjetil Sandve

Lars Holden

Knut Liestøl

Eivind Hovig

Morten Johansen

Marit Holden

Vegard Nygaard

Egil Ferkingstad
Support
Summary

- The Genomic HyperBrowser asks what, and solves how
- Tightly integrated with Galaxy, but expanding borders
- Google -> ‘hyperbrowser’
Conclusion

• Highly modifying a wheel is still better than reinventing it
• Galaxy passed our stress tests, and constantly adds value
• Google -> ‘hyperbrowswer’