

# ENCODE DCC Update 2010

Jim Kent, UC Santa Cruz

# Outline

- Short/Medium Term Logistics
- New and current features and display
  - Saving and sharing sessions
  - Wiggles with whiskers
  - Big Wig and Big Bed
  - BAM
- Future plans
  - Integrating in data from analysis working group
    - Integrated elements tracks
  - Drag and drop track reordering
  - “Live figures” gallery of sessions



# Near term logistical issues

- We'll be reopening submission pipeline March 15
  - New submissions should be on hg19/GRCh37 or mm9
  - Use BAM rather than TagAlign for aligned reads
- Currently DCC data wranglers still focused on getting Jan 15 freeze data polished, through QA and to public.
- DCC will be meeting with new, ARRA funded groups Thursday for lunch to help get things started. Please pick up lunch order form at registration desk if you are attending.

# ENCODE Histone Modifications by Broad Institute ChIP-seq

Maximum display mode:   [Reset to defaults](#)

Select views ([help](#)):

[Peaks](#)  [Signal](#)

Select subtracks by cell line and antibody:

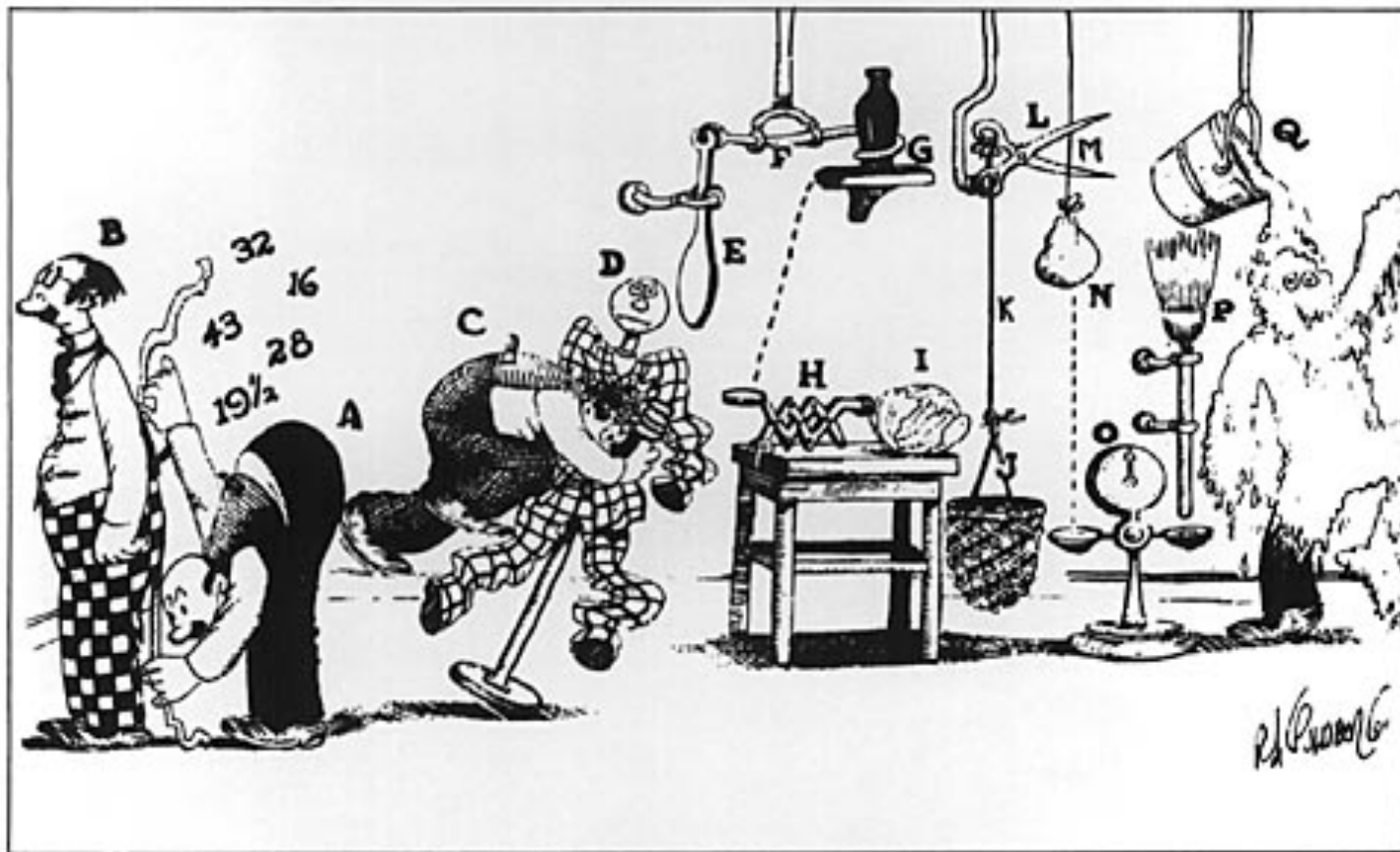
<input type="checkbox"/> <input type="checkbox"/>	All	Cell Line	GM12878	H1-hESC	HepG2	HMEC	HSMM	HUVEC	K562	NHEK	NHLF	Cell Line	All	<input type="checkbox"/> <input type="checkbox"/>
	Antibody		<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	Antibody		
<input checked="" type="checkbox"/>	<a href="#">CTCF</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">CTCF</a>
<input checked="" type="checkbox"/>	<a href="#">H3K4me1</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K4me1</a>
<input checked="" type="checkbox"/>	<a href="#">H3K4me2</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K4me2</a>
<input checked="" type="checkbox"/>	<a href="#">H3K4me3</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K4me3</a>
<input checked="" type="checkbox"/>	<a href="#">H3K9ac</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K9ac</a>
<input checked="" type="checkbox"/>	<a href="#">H3K9me1</a>							<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<a href="#">H3K9me1</a>
<input checked="" type="checkbox"/>	<a href="#">H3K27ac</a>		<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K27ac</a>
<input checked="" type="checkbox"/>	<a href="#">H3K27me3</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K27me3</a>
<input checked="" type="checkbox"/>	<a href="#">H3K36me3</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H3K36me3</a>
<input checked="" type="checkbox"/>	<a href="#">H4K20me1</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">H4K20me1</a>
<input checked="" type="checkbox"/>	<a href="#">Pol2(b)</a>							<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>	<a href="#">Pol2(b)</a>
<input checked="" type="checkbox"/>	<a href="#">Input Control</a>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<a href="#">Input Control</a>

List subtracks:  only selected/visible  all

<input checked="" type="checkbox"/>	Cell Line <sup>1</sup>	Antibody <sup>2</sup>	Views <sup>3</sup>	Restricted Until
<input checked="" type="checkbox"/>	GM12878	CTCF	Peaks ENCODE Histone Mods, Broad ChIP-seq Peaks (CTCF, GM12878) ... <a href="#">schema</a>	2009-10-05
<input checked="" type="checkbox"/>	GM12878	CTCF	Signal ENCODE Histone Mods, Broad ChIP-seq Signal (CTCF, GM12878) ... <a href="#">schema</a>	2009-10-05
<input checked="" type="checkbox"/>	GM12878	H3K4me1	Peaks ENCODE Histone Mods, Broad ChIP-seq Peaks (H3K4me1, GM12878) ... <a href="#">schema</a>	2009-10-05
<input checked="" type="checkbox"/>	GM12878	H3K4me1	Signal ENCODE Histone Mods, Broad ChIP-seq Signal (H3K4me1, GM12878) ... <a href="#">schema</a>	2009-10-05

Meanwhile the matrices are filling in, though not fast enough for Peter Good!

# New(ish) Features



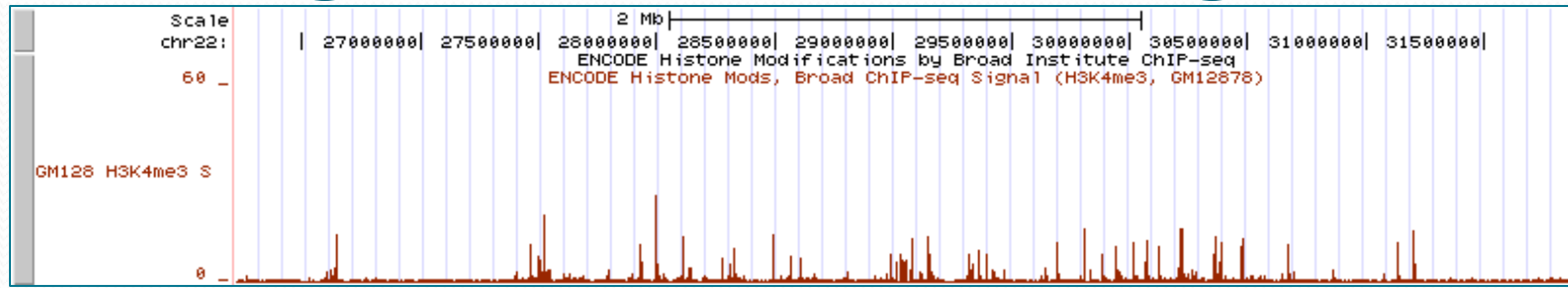
Idea For Dodging Bill Collectors RUBE GOLDBERG (tm) RGI 046



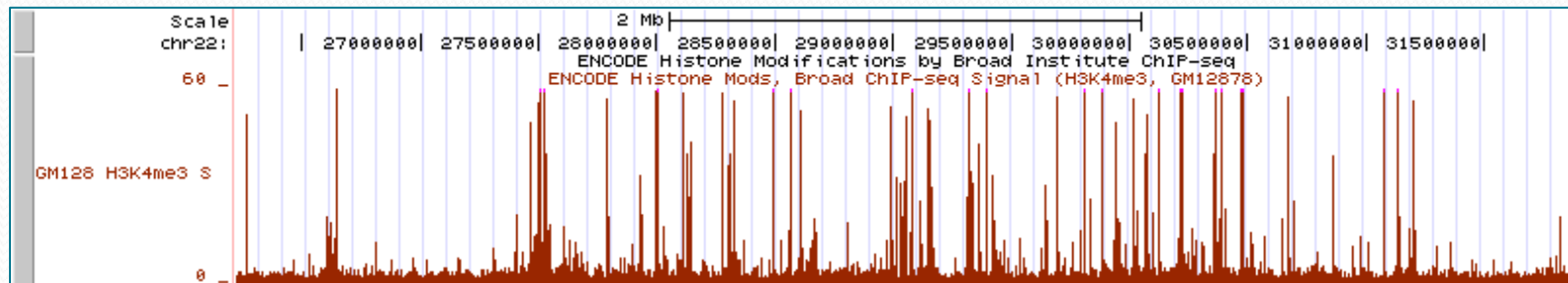
# Under-appreciated features

- Can save your browser state including position, what tracks and subtracks are displayed, wiggle heights, and indeed anything you can set in a “session” that can be emailed to people, or reused yourself later.
- All ChIP-SEQ tracks have reasonable scales set by wranglers so “autoScale,” which can be misleading, is no longer necessary.
- Can drag in ruler bar to select precise area to zoom into.

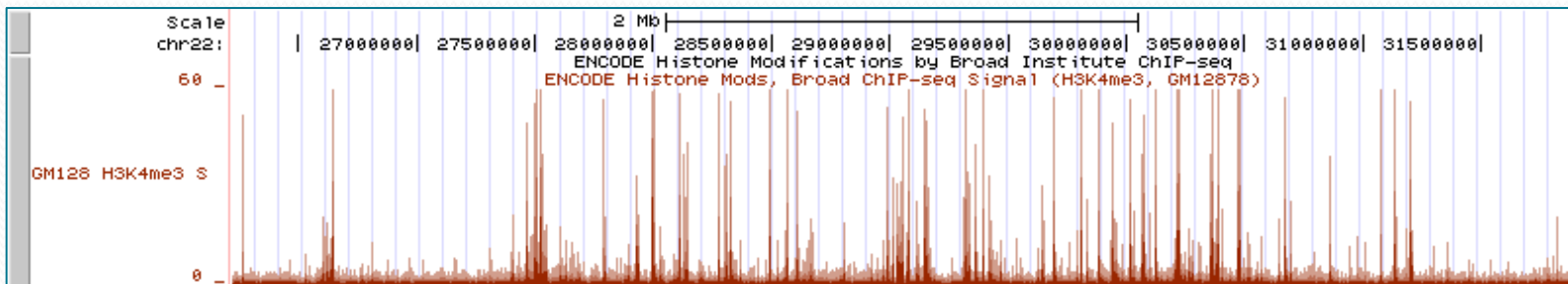
# Adding Whiskers to Wig



When zoomed out “mean” display fades to zero

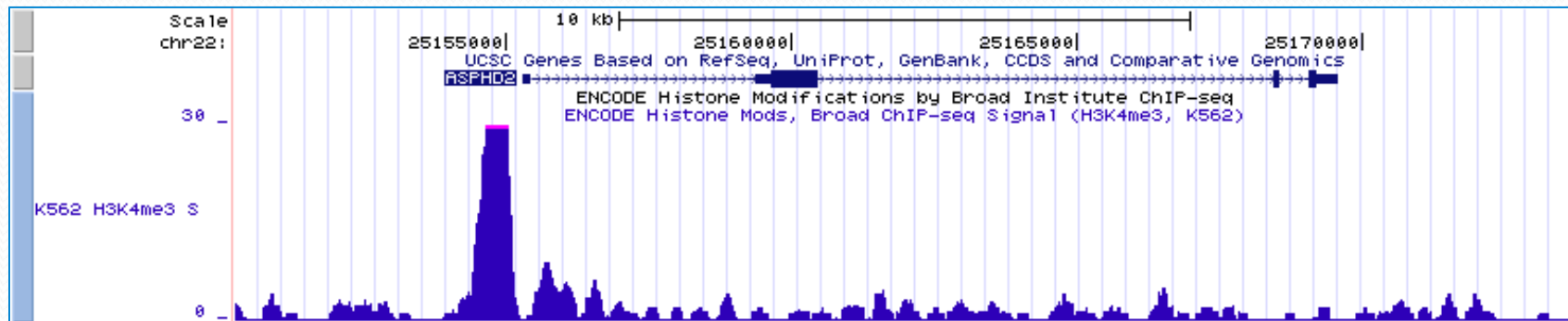


Alternatively can show max value under pixel, but this amplifies noise



Mean+whiskers shows mean in dark, mean+standard dev. in medium, max in light

# Clipped peaks now “bleed”



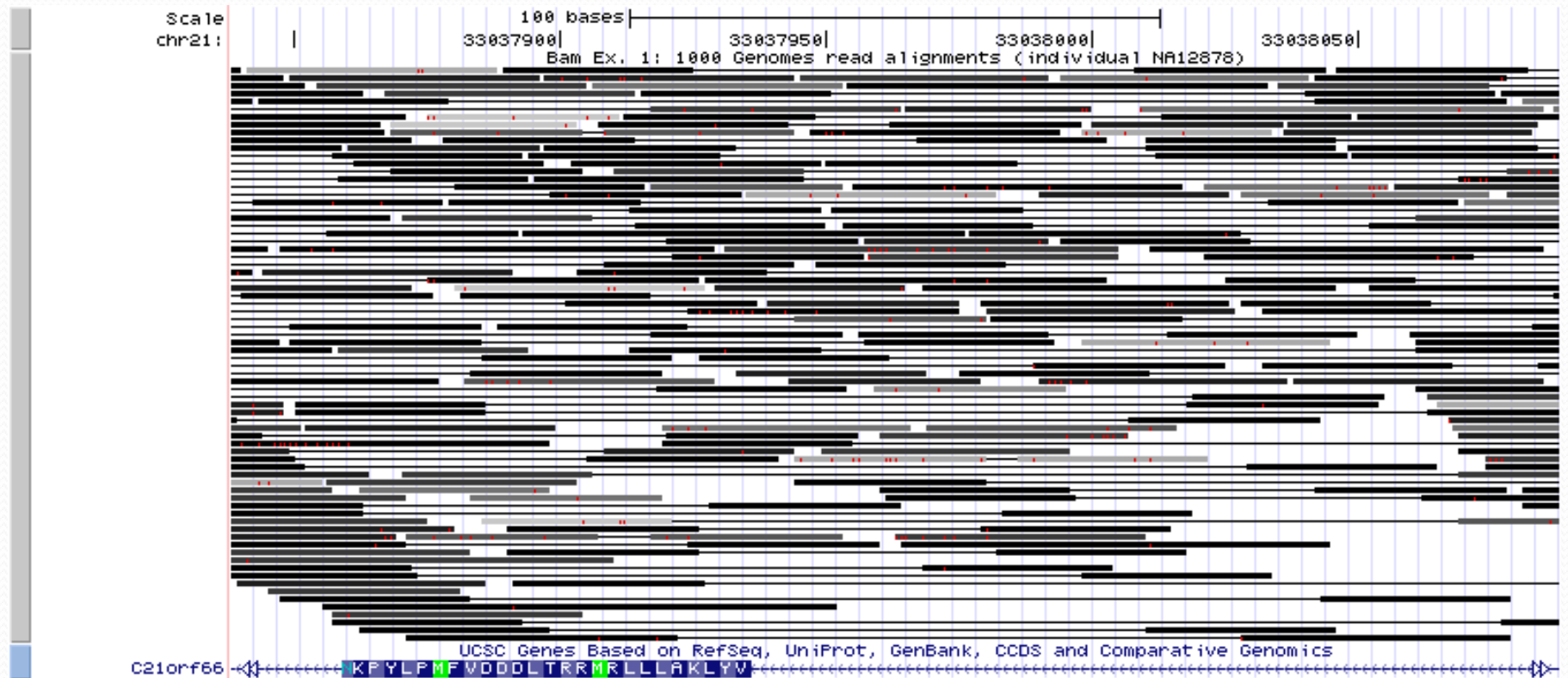


# Making use of BigWig

- 1) Create your file in wig or bedGraph format
- 2) Download wigToBigWig or bedGraphToBigWig
- 3) Run program to create indexed big\* file
- 4) Move big\* file to a http, https, or ftp site
- 5) Create custom track

Excellent idea to preview data this way before submitting to UCSC!

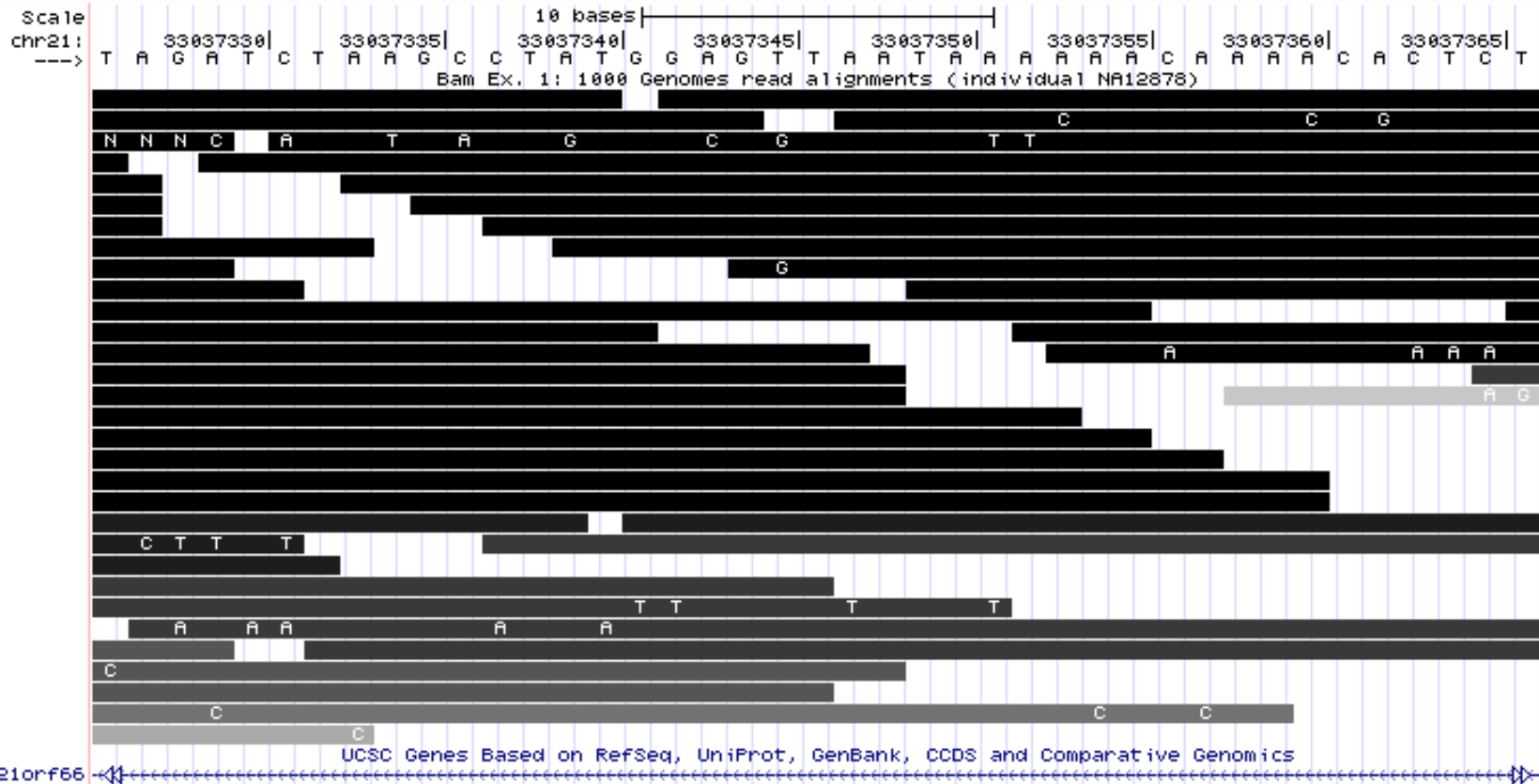
# BAM [custom] tracks



Gray-level indicates mapping confidence, red mismatches with reference. Lines connect mate pairs

BAM custom tracks cached much like BigWig and BigBed, though limits when you zoom out

# BAM Close Up



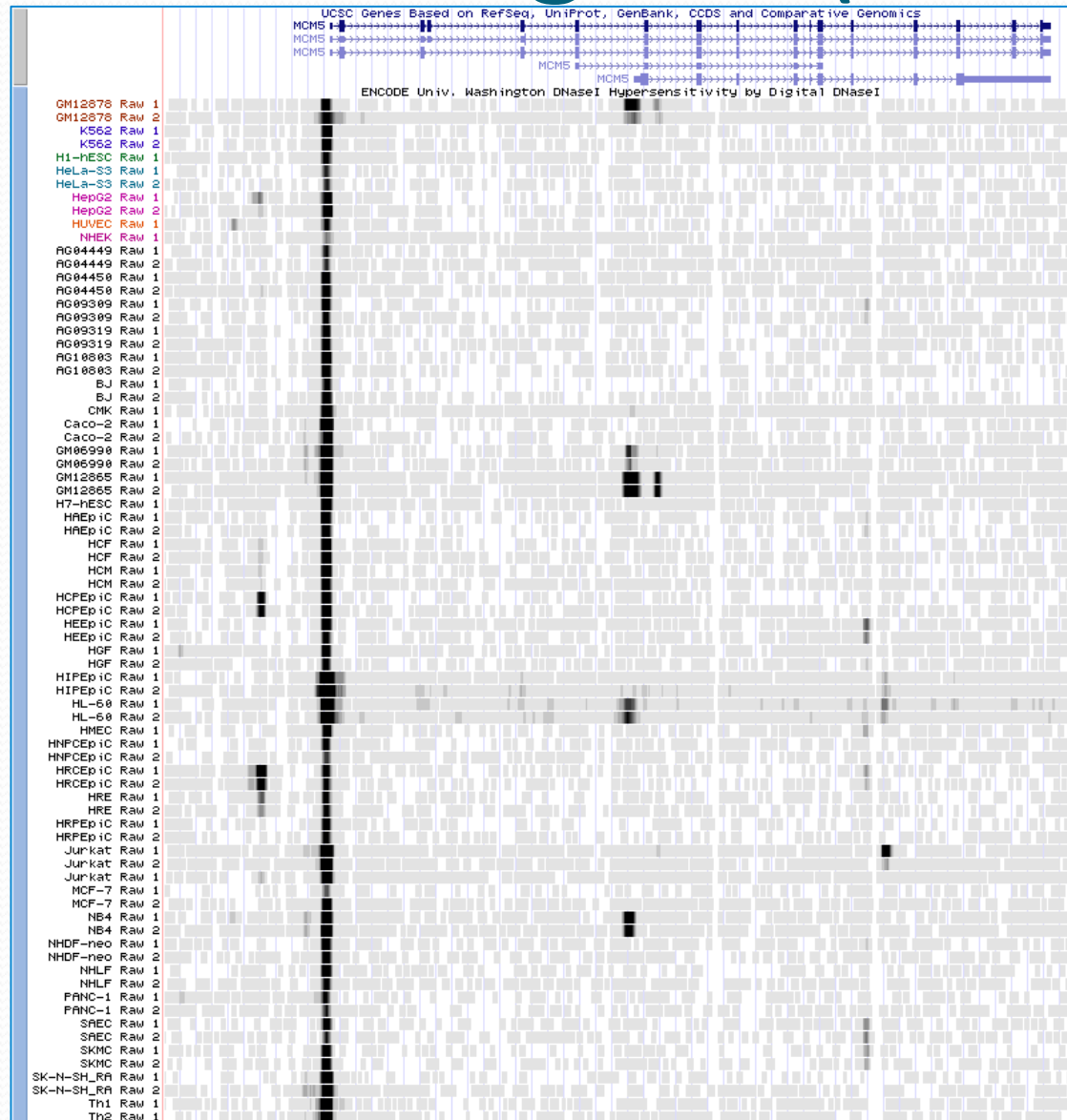
# Future Plans



# There is a need to do integrated tracks!

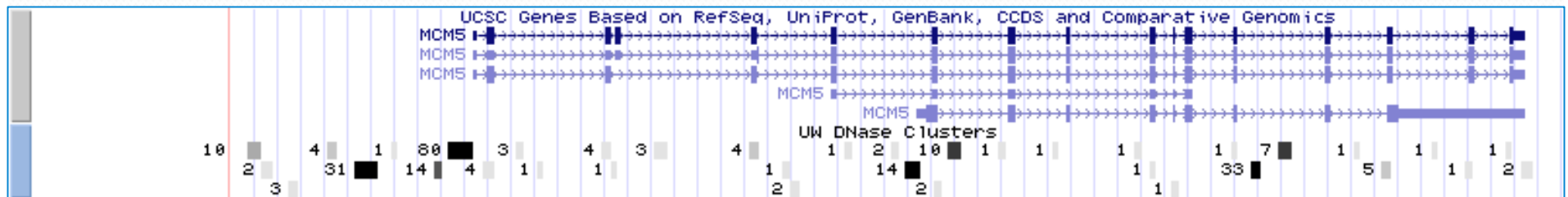
- Some work going on at UCSC
- Hope to bring in integrated tracks from analysis working group as well

# UW DNase all signals (so far!)





# UW DNase Merged



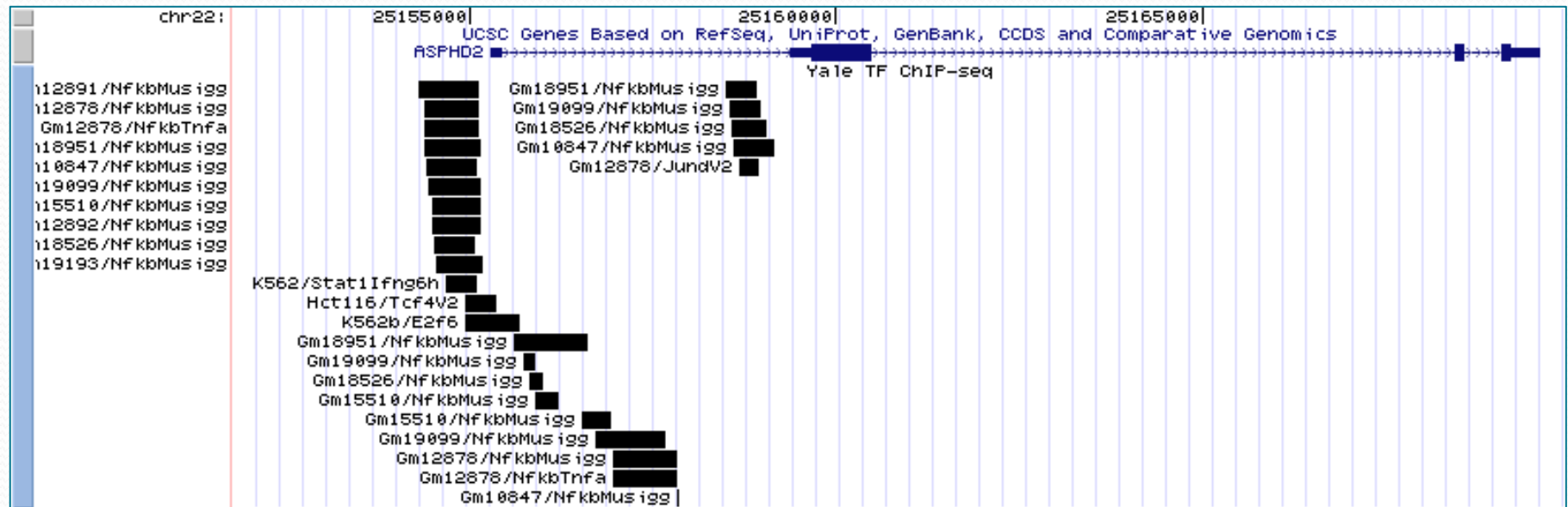
The number to the left indicates the number of samples (replicates of cell lines) in which the peak is seen.



# Yale TFBS Raw Signals (no Pol)

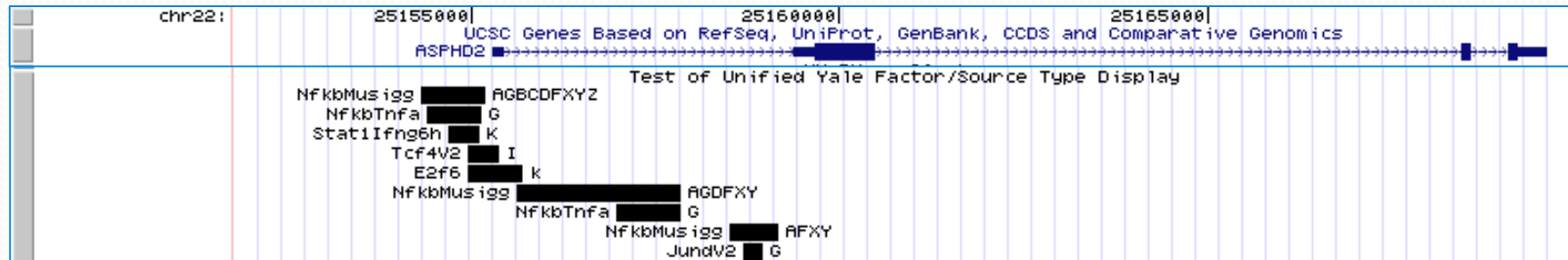


# Yale TFBS Combined Peaks



Various peak-called “tracks” combined into a single track with label being cell-line/factor. Good, but still get “stacks,” mostly of common factors seen in many cell lines.

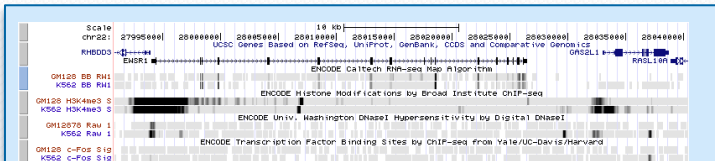
# Yale TFBS Merged Peaks



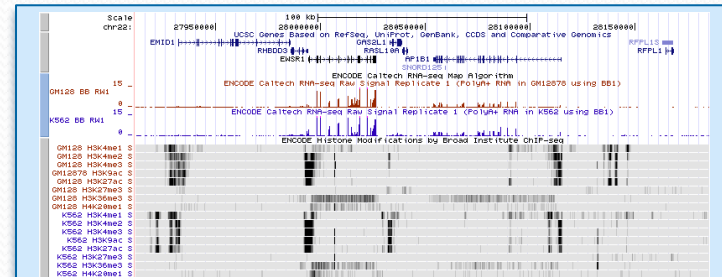
Letters to right indicate which cell lines peak is seen in:

- |           |           |
|-----------|-----------|
| A Gm10847 | I Hct116  |
| B Gm12891 | K K562    |
| C Gm12892 | L Hepg2   |
| D Gm15510 | M Mcf7    |
| E Gm18505 | N Nt2d1   |
| F Gm18526 | X Gm18951 |
| G Gm12878 | Y Gm19099 |
| H Helas3  | Z Gm19193 |

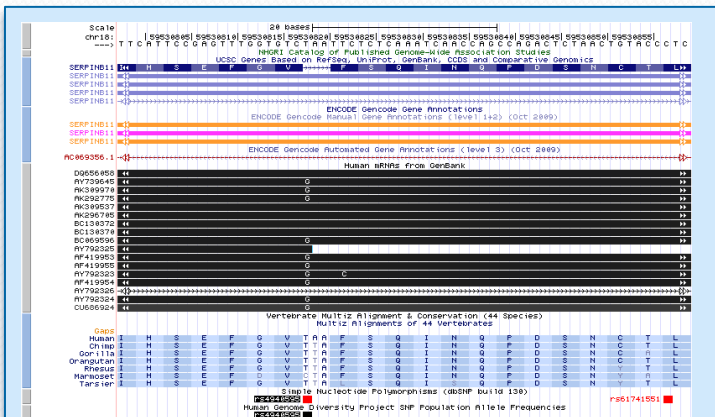
# Browser ENCODE Session Gallery



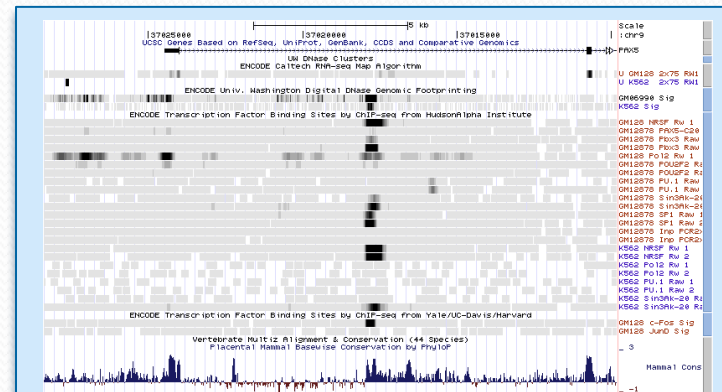
Basic ENCODE data on the tier 1 cell lines GM128, a lymphoid cell line and K562, a myeloid line. The tracks include RNA-seq which shows the level of expression of the gene *EWSR1*, the histone marks H3K4Me3 which is associated with promoters, DNase I hypersensitivity which is associated with regulatory regions in general, and ChIP-seq showing levels of occupancy by the transcription factor c-Fos.



All available histone ChIP-seq data on the two tier one cell lines. H3K4me1 is associated with enhancers and to an extent with promoters, H3K4me3 with promoters, and H3K36me3 with transcribed regions.



A premature stop codon that is found in the reference genome and about half of people of European descent. The UCSC Genes track is forced to skip the codon.



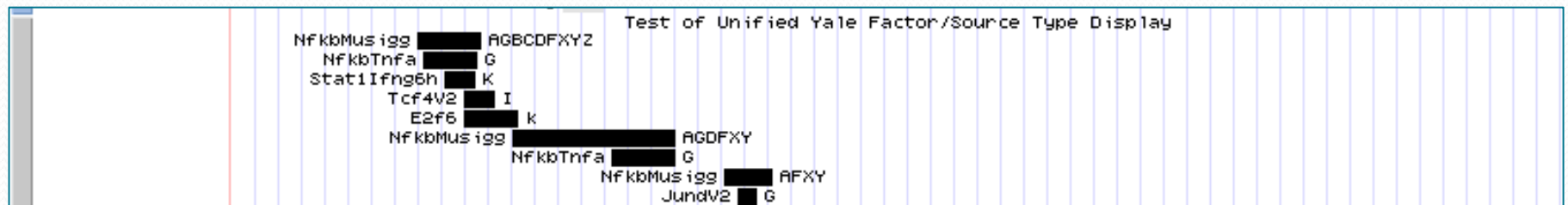
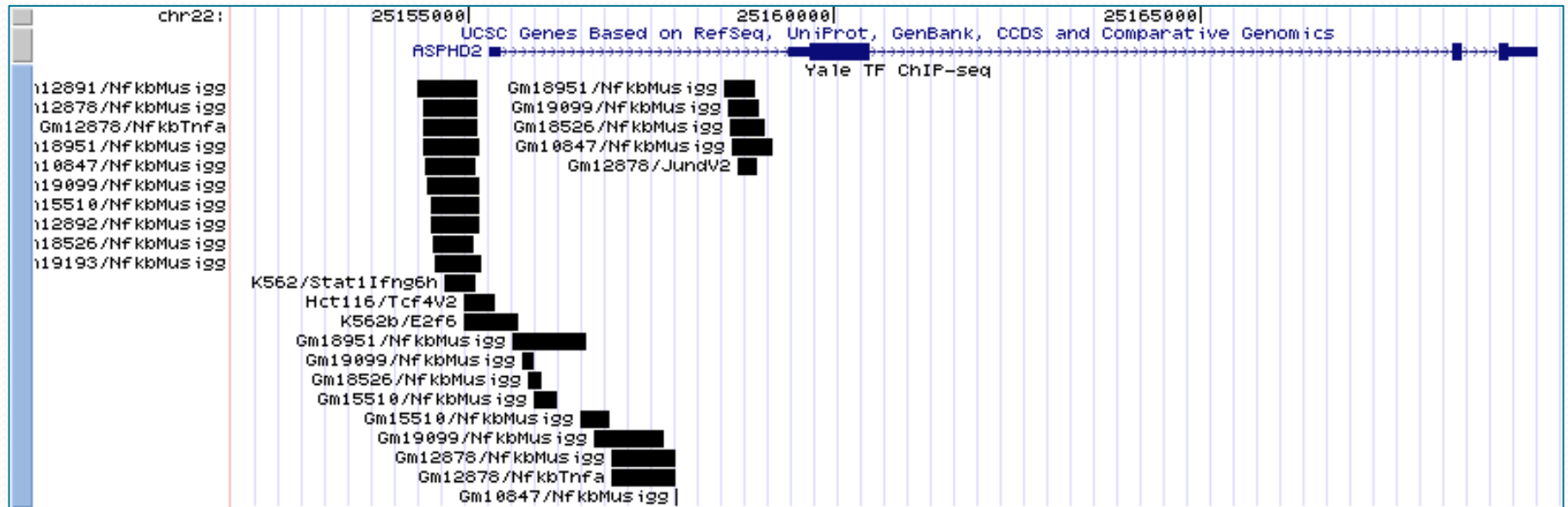
Regulatory elements in the promoter and first intron of the transcription factor *PAX5*, a gene expressed in GM128 but not K562 cells, suggestive of a complex regulatory circuit with some autoregulation.

Gallery linked to main ENCODE page. Clicking on image takes you to browser at that spot with those tracks configured as shown. Caption explains session to non-specialists.



The End

# Merging Process



Very simple minded – peaks from same factor in different cell lines that overlap are merged. Extents of merged peak encompass all peaks it is made of.  
 Would be happy to use something more sophisticated from analysis working group.