

ChIP-seq/Functional Genomics/Epigenomics

CBSU/3CPG/CVG Next-Gen Sequencing Workshop

Josh Waterfall

March 31, 2010

# Outline

Introduction to ChIP-seq

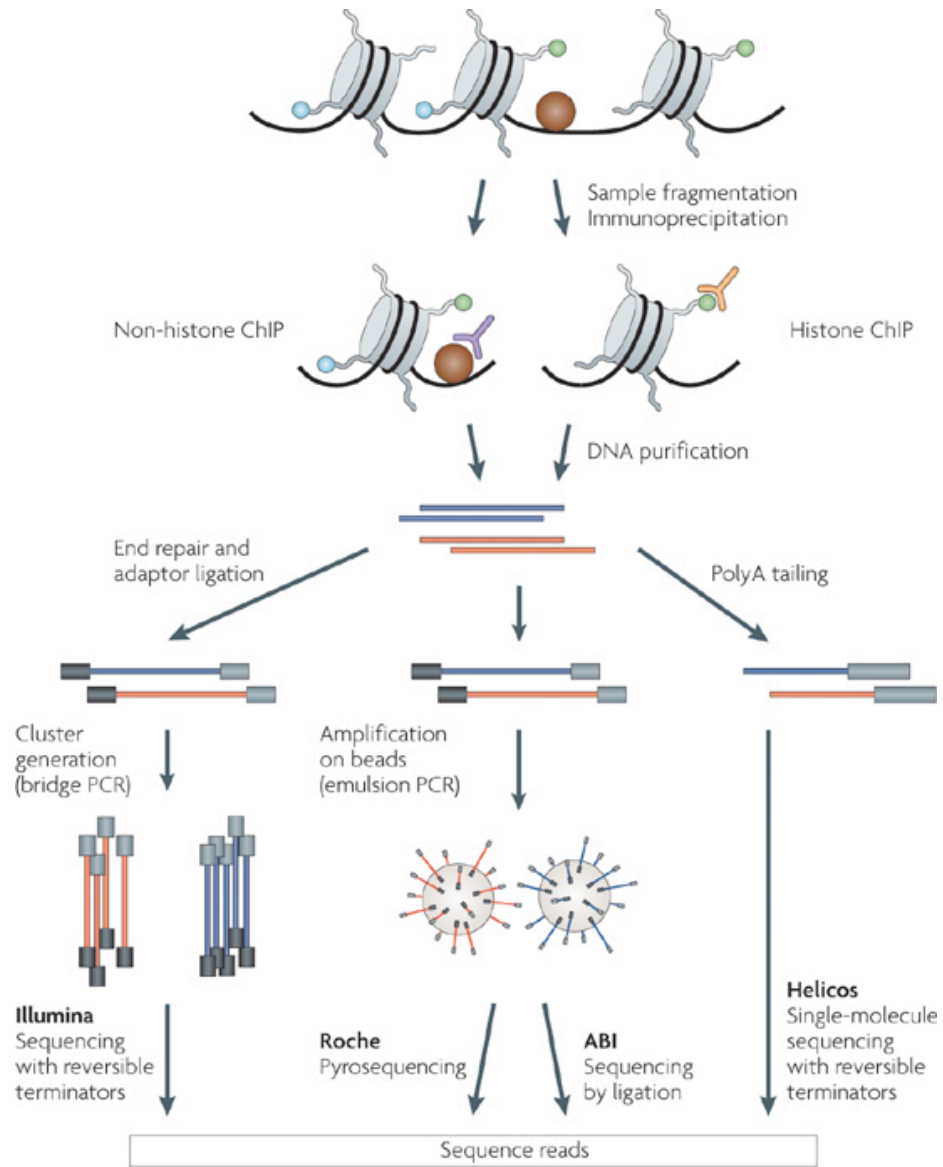
Control data sets

Peak/enriched region identification

Related functional genomics assays

Useful web resources

# Chromatin ImmunoPrecipitation (ChIP-seq)



## Advantages over array-based methods

No cross-hybridization background

Lower end of sensitivity largely dependent on just sequencing depth

More linear, quantitative assay

Unmappable portion of genome is distinct from, and much smaller than, repeat masked portion

No limit based on probe locations

Needs less starting material

Higher resolution

Important Issues:

Appropriate Controls

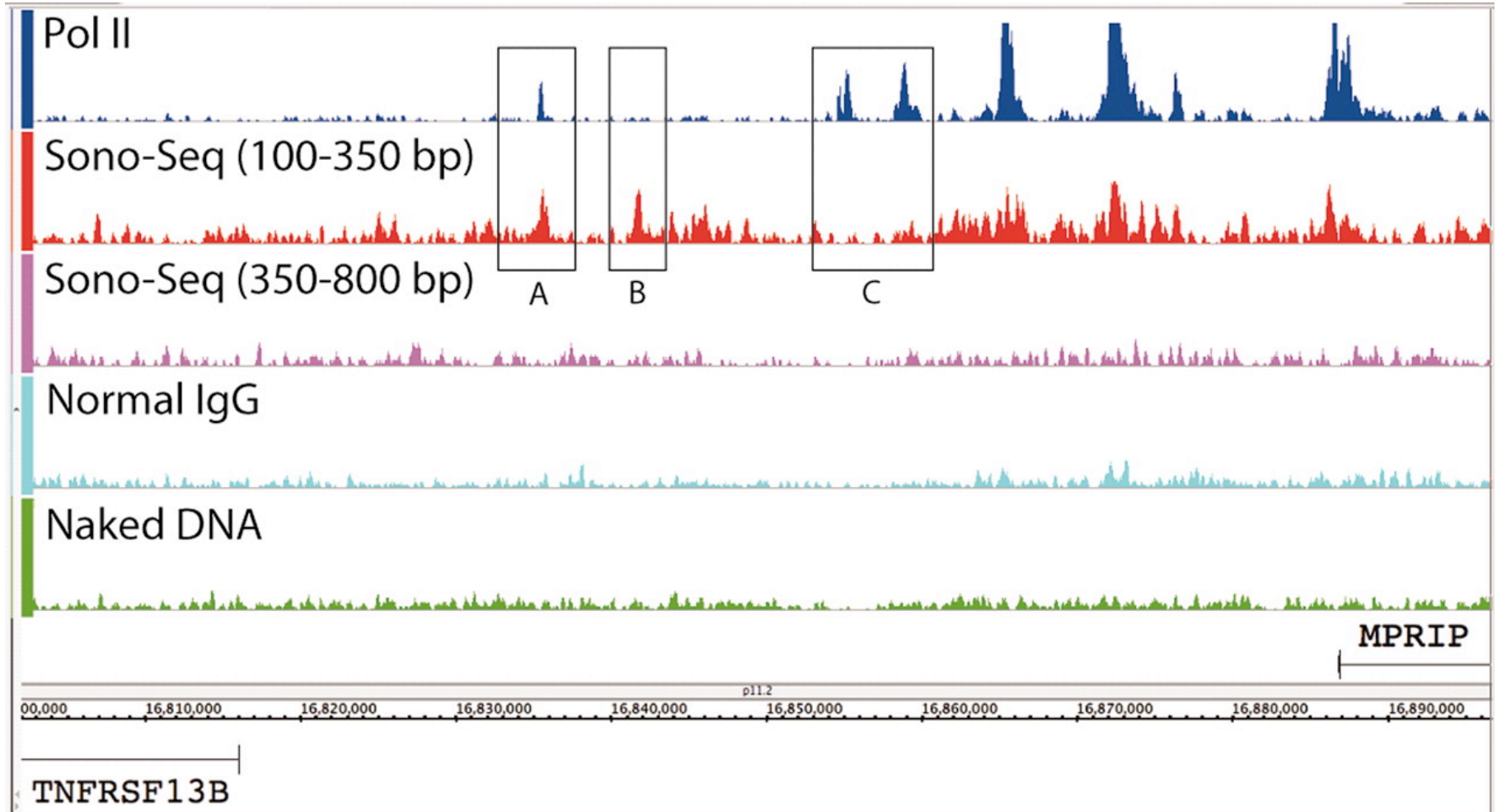
Identifying Enriched Regions

## Controls

Sequencing is such high-sensitivity that signals invisible in any other assay are now apparent. Need rigorous controls to be confident of enrichment.

- Input DNA has non-random pattern (open chromatin shears more easily) - Sono-seq is an actual assay.
- Mock-IP controls for more steps in ChIP protocol than input DNA but not antibody cross-reactivity.
- Different antibodies (to different epitopes) in separate experiments, or ChIP after target protein has been depleted (or in cell-line without tagged protein), help control for cross-reactivity.
- To characterize new antibody, IP and mass-spec everything that comes down to verify only expected binding partners are seen.

# Background rate is non-uniform



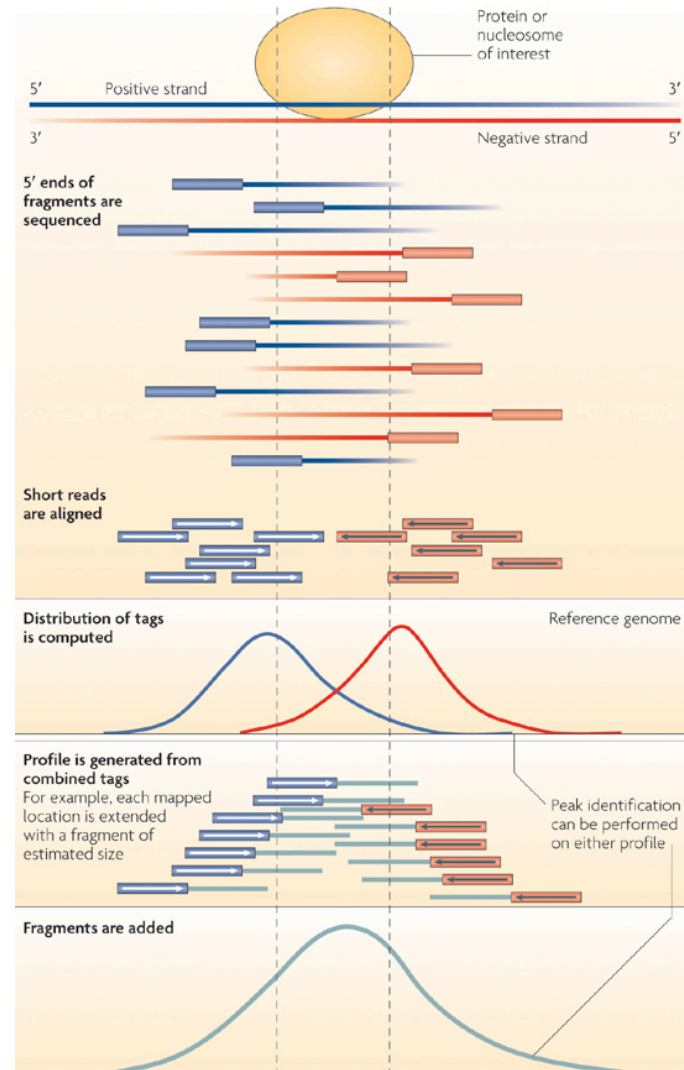
## Identifying enriched regions

Identifying narrow ChIP peaks is very different from:  
identifying broadly enriched regions in ChIP.  
identifying narrow peaks from other assays.

- Many well-worked out programs for identifying narrow ChIP peaks (e.g. from sequence specific binding factors).  
The best programs can exploit strand-specific patterns, local versus global background levels, mappability, etc.
- Other assays also result in localized peaks but without same strand pattern.
- Most work on broad regions (e.g. particular histone modifications or RNA polymerase) is based on sliding windows (either of fixed length or fixed read count).



# Strand specific pattern for localized peaks in ChIP-seq



# Peak Detection Software

Is strand information used?

If fragment size is used, is it defined by user or estimated from data?

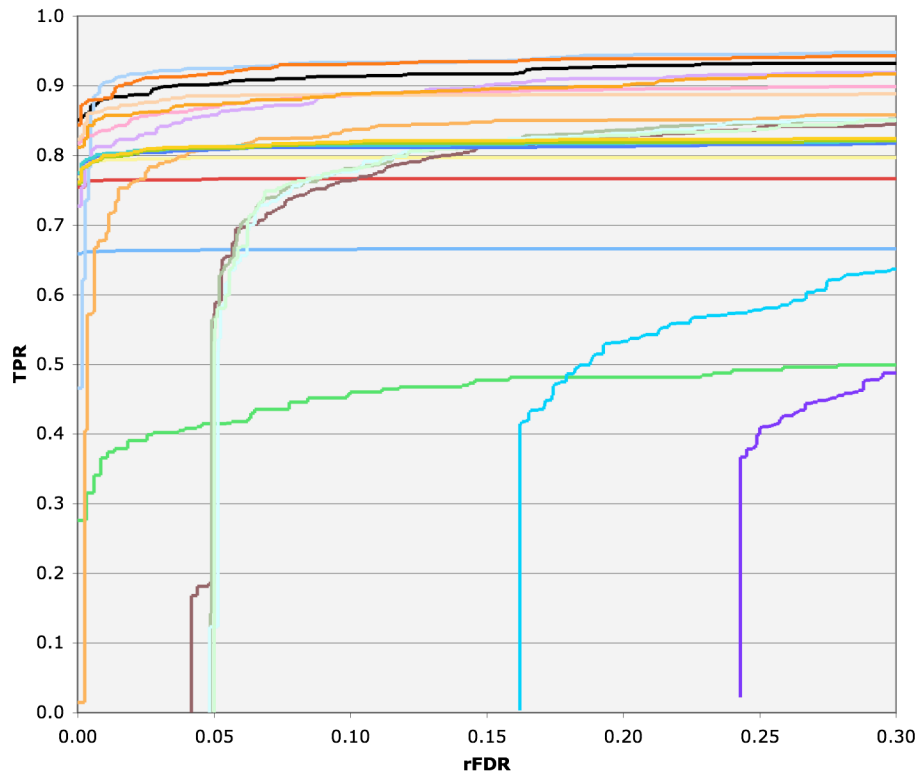
(How) is control data incorporated?

Is background defined locally or globally?

(How) are unmappable regions treated?

Table 1 | Publicly available ChIP-seq software packages discussed in this review

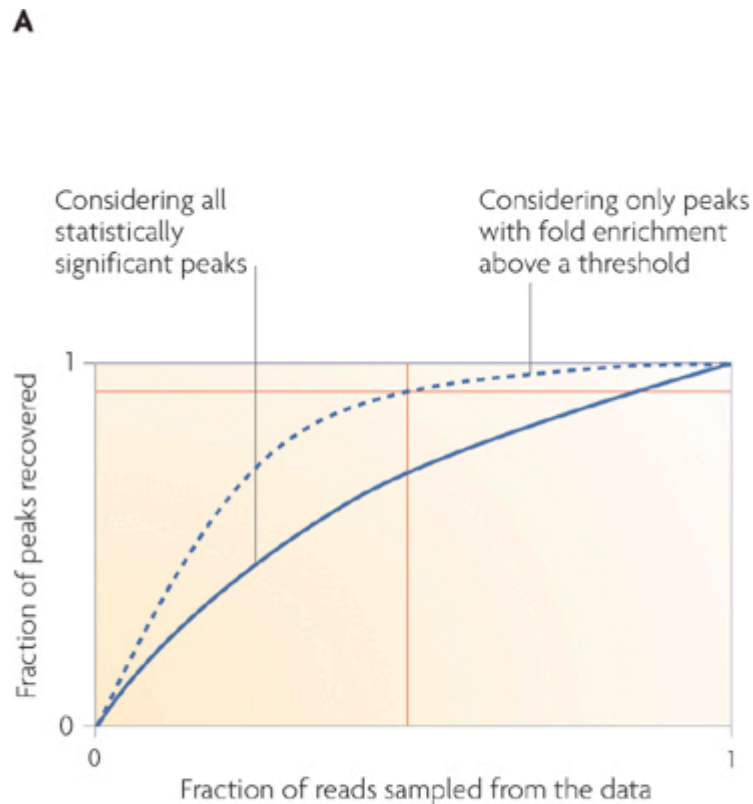
	Profile	Peak criteria <sup>a</sup>	Tag shift	Control data <sup>b</sup>	Rank by	FDR <sup>c</sup>	User input parameters <sup>d</sup>	Artifact filtering: strand-based/duplicate <sup>e</sup>	Refs.
CisGenome v1.1	Strand-specific window scan	1: Number of reads in window 2: Number of ChIP reads minus control reads in window	Average for highest ranking peak pairs	Conditional binomial used to estimate FDR	Number of reads under peak	1: Negative binomial 2: conditional binomial	Target FDR, optional window width, window interval	Yes / Yes	10
ERANGE v3.1	Tag aggregation	1: Height cutoff High quality peak estimate, per-region estimate, or input	High quality peak estimate, per-region estimate, or input	Used to calculate fold enrichment and optionally <i>P</i> values	<i>P</i> value	1: None 2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$	Optional peak height, ratio to background	Yes / No	4,18
FindPeaks v3.1.9.2	Aggregation of overlapped tags	Height threshold	Input or estimated	NA	Number of reads under peak	1: Monte Carlo simulation 2: NA	Minimum peak height, subpeak valley depth	Yes / Yes	19
F-Seq v1.82	Kernel density estimation (KDE)	<i>s</i> s.d. above KDE for 1: random background, 2: control	Input or estimated	KDE for local background	Peak height	1: None 2: None	Threshold s.d. value, KDE bandwidth	No / No	14
GLTR	Aggregation of overlapped tags	Classification by height and relative enrichment	User input tag extension	Multiply sampled to estimate background class values	Peak height and fold enrichment	2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$	Target FDR, number nearest neighbors for clustering	No / No	17
MACS v1.3.5	Tags shifted then window scan	Local region Poisson <i>P</i> value	Estimate from high quality peak pairs	Used for Poisson fit when available	<i>P</i> value	1: None 2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$	<i>P</i> -value threshold, tag length, <i>m</i> fold for shift estimate	No / Yes	13
PeakSeq	Extended tag aggregation	Local region binomial <i>P</i> value	Input tag extension length	Used for significance of sample enrichment with binomial distribution	<i>q</i> value	1: Poisson background assumption 2: From binomial for sample plus control	Target FDR	No / No	5
QuEST v2.3	Kernel density estimation	2: Height threshold, background ratio	Mode of local shifts that maximize strand cross-correlation	KDE for enrichment and empirical FDR estimation	<i>q</i> value	1: NA 2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$ as a function of profile threshold	KDE bandwidth, peak height, subpeak valley depth, ratio to background	Yes / Yes	9
SICER v1.02	Window scan with gaps allowed	<i>P</i> value from random background model, enrichment relative to control	Input	Linearly rescaled for candidate peak rejection and <i>P</i> values	<i>q</i> value	1: None 2: From Poisson <i>P</i> values	Window length, gap size, FDR (with control) or <i>E</i> -value (no control)	No / Yes	15
SiSSRs v1.4	Window scan	$N_+ - N_-$ sign change, $N_+ + N_-$ threshold in region <sup>f</sup>	Average nearest paired tag distance	Used to compute fold-enrichment distribution	<i>P</i> value	1: Poisson 2: control distribution	1: FDR 1,2: $N_+ + N_-$ threshold	Yes / Yes	11
spp v1.0	Strand specific window scan	Poisson <i>P</i> value (paired peaks only)	Maximal strand cross-correlation	Subtracted before peak calling	<i>P</i> value	1: Monte Carlo simulation 2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$	Ratio to background	Yes / No	12
USeq v4.2	Window scan	Binomial <i>P</i> value	Estimated or user specified	Subtracted before peak calling	<i>q</i> value	1, 2: binomial 2: $\frac{\# \text{ control}}{\# \text{ ChIP}}$	Target FDR	No / Yes	20



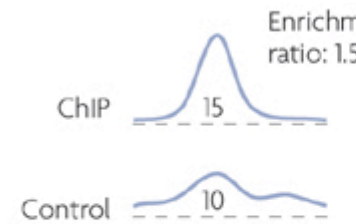
ChIP-seq peak calling software contest on <http://seqanswers.com>

Pepke, S., Wold, B., Mortazavi, A. (2009) Nat. Methods 6:S22-S32

# Assessing saturation and significance



**Ba** Not statistically significant

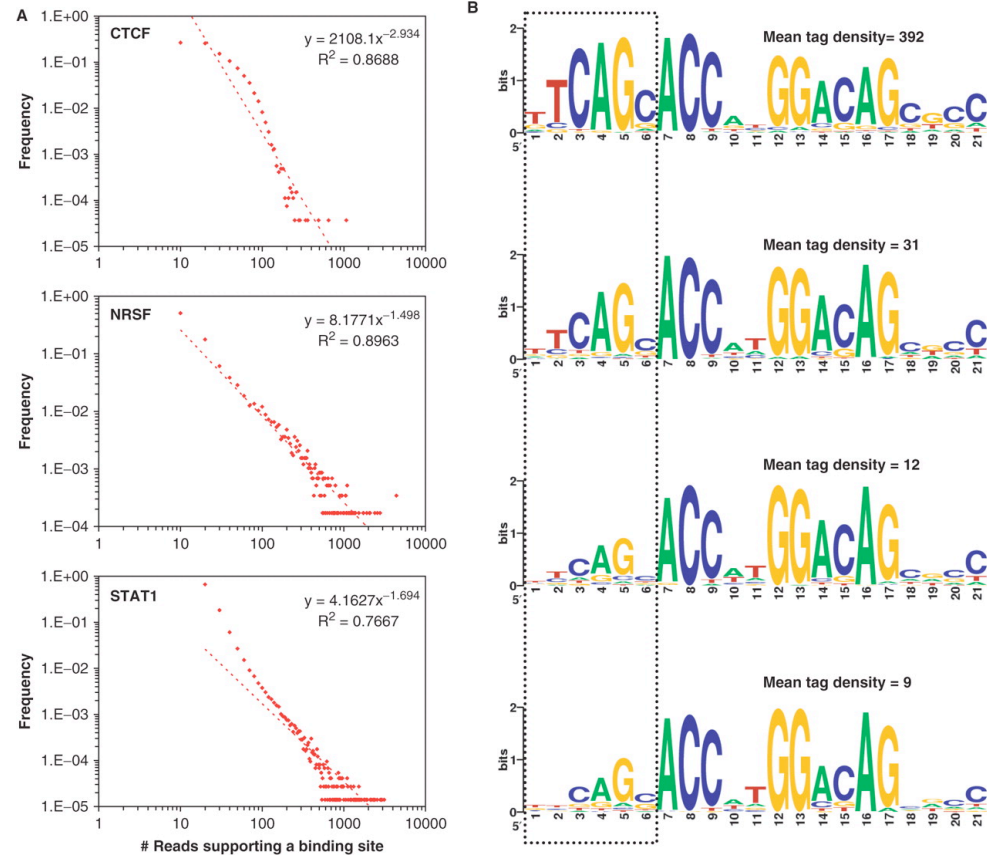


**Bb** Statistically significant

Saturation is dependent on both ChIP efficiency (globally) and factor:DNA affinity (locally)

Significance depends on both ChIP and control read counts

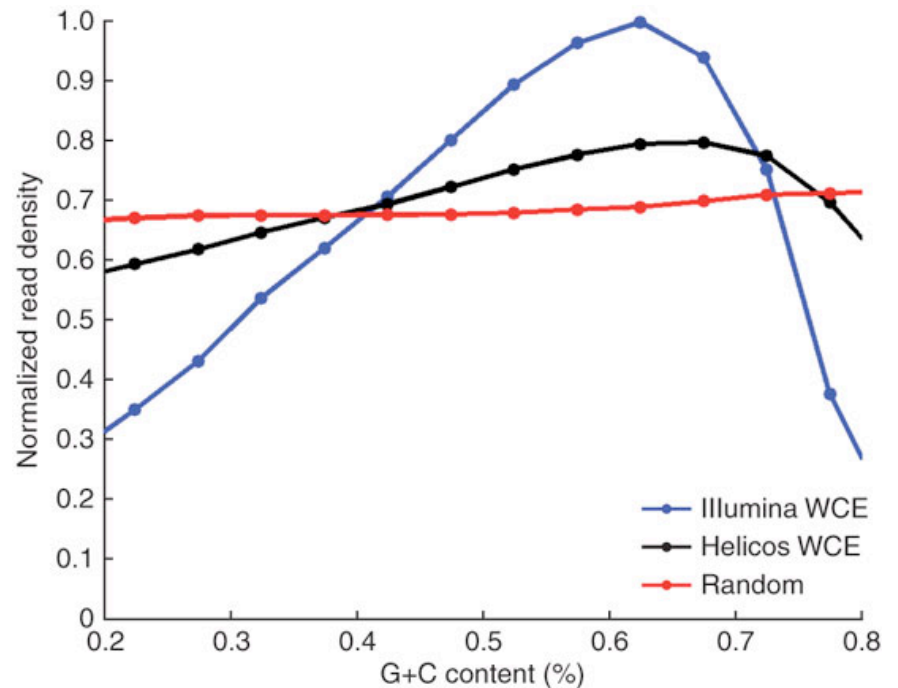
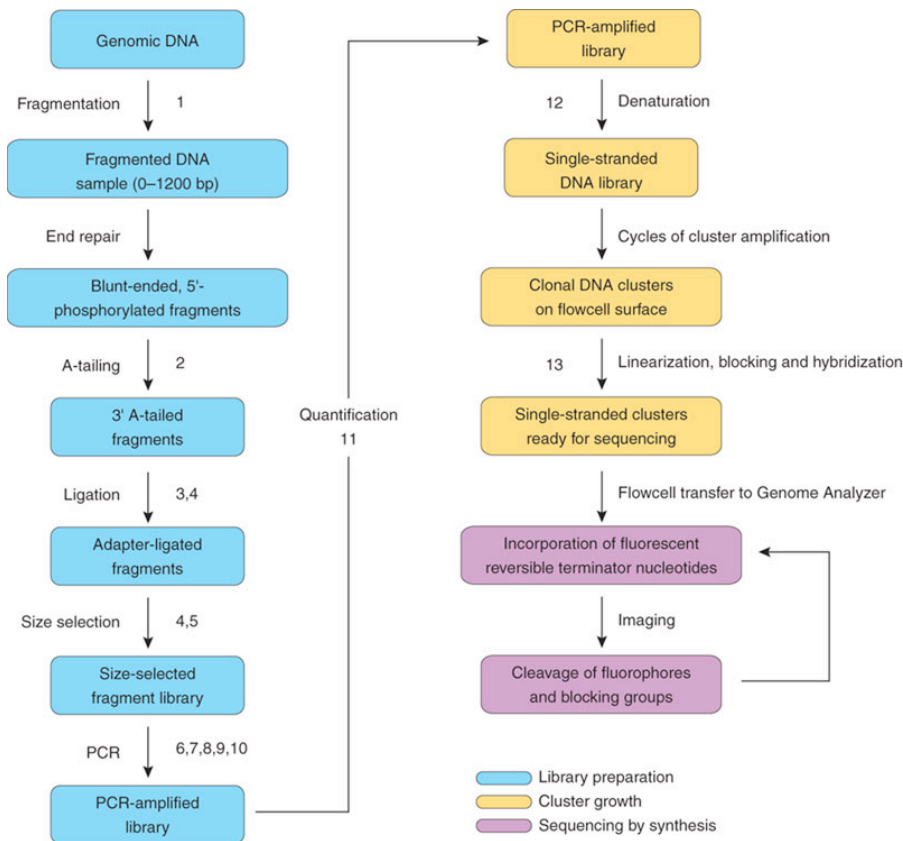
# ChIP-seq read density is quantitative measure of binding level



# Solutions to GC bias

Improve Illumina protocol  
(especially gel purification)

Use single molecule  
sequencing technology



Quail, MA, et. al. (2008) Nat. Methods 5:1005-1010

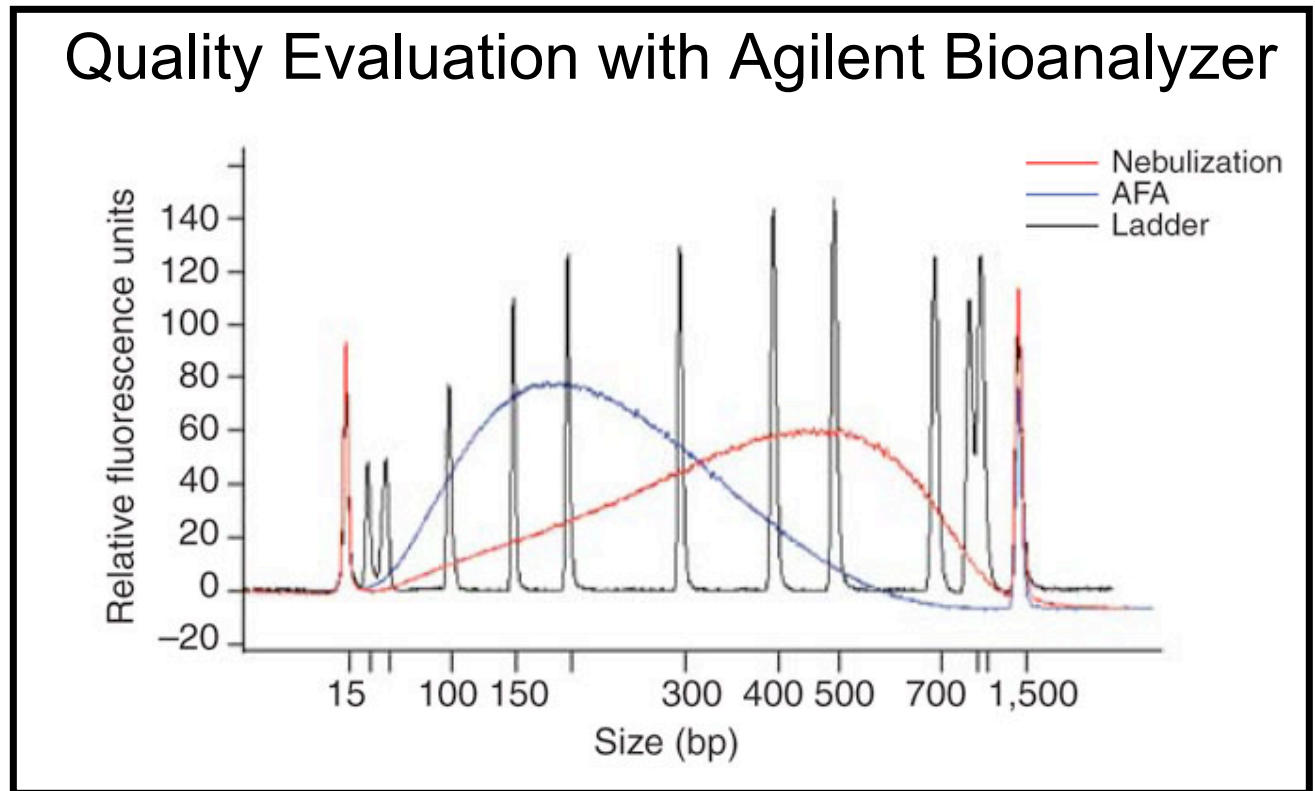
Goren, A, et. al. (2009) Nat. Methods 7:47-49

# Fragmenting DNA

AFA (Adaptive Focused Acoustics)/Covaris  
enzyme (MNase, NEB Fragmentase)

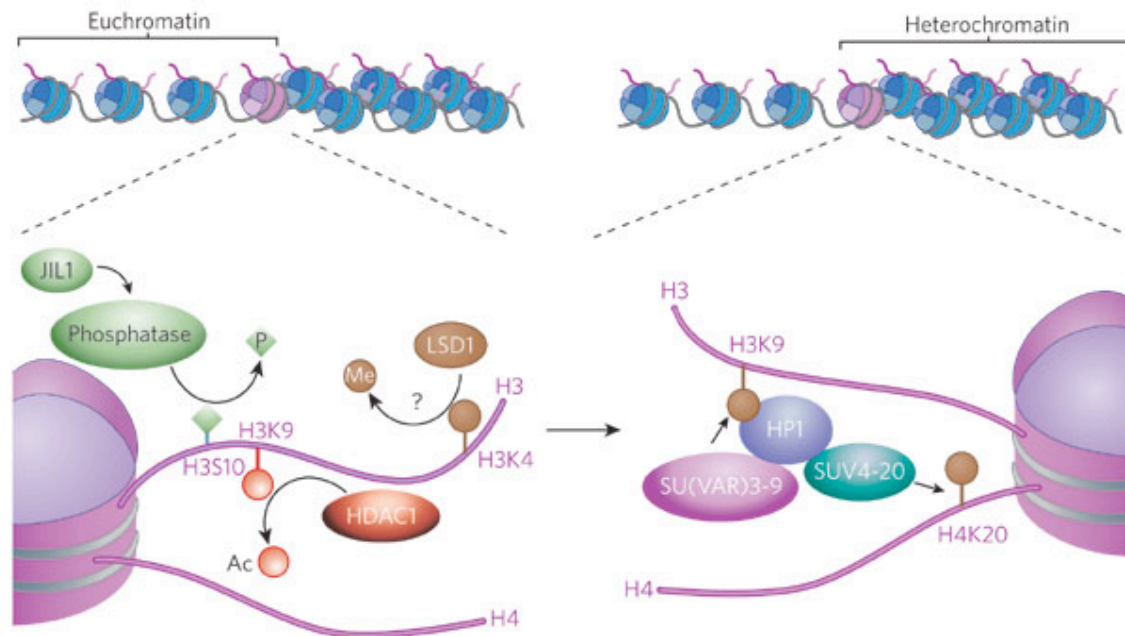
Nebulization

Sonication



# Fragmenting DNA

Open chromatin fragments more easily than closed chromatin -  
ChIP for factors associated with transcriptional repression or heterochromatin can be more difficult than for factors associated with transcriptional activation.



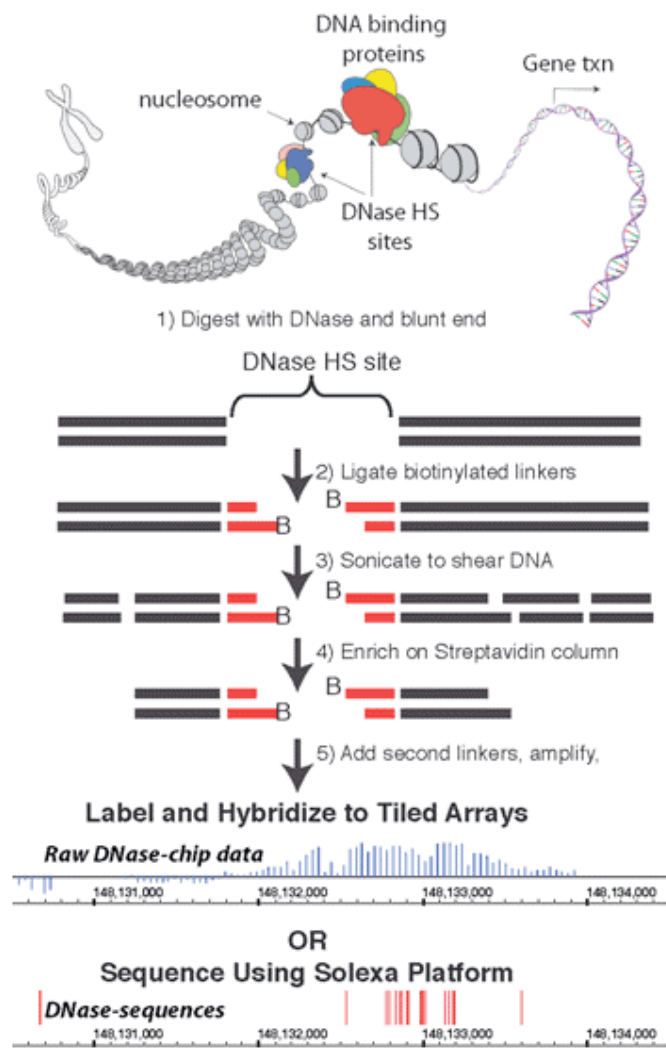
## Related Functional Genomic Assays

- Chromatin Immunoprecipitation (ChIP)
- DNase I Hypersensitivity
- Formaldehyde Assisted Isolation of Regulatory Elements
- DNA Methylation (bisulfite, affinity, or restriction enzyme based)
- Cap Analysis of Gene Expression (CAGE)
- Genomic Run On (GRO-seq)
- Many more...

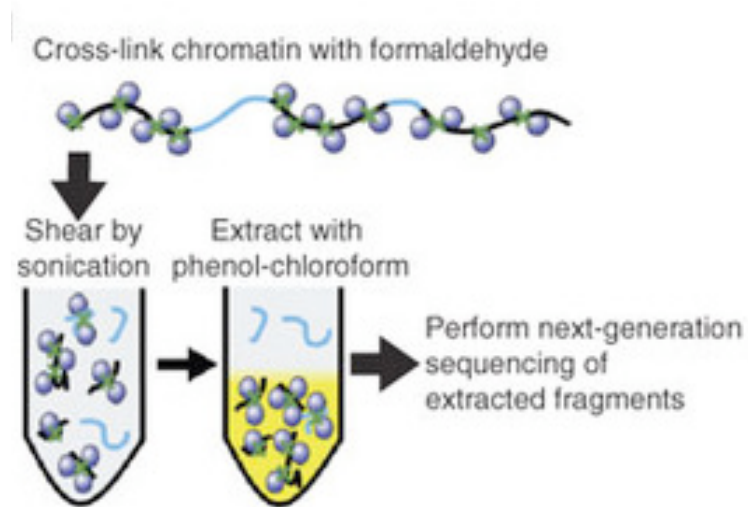
Each can couple to next-gen sequencing and entails analysis/identification of enriched regions



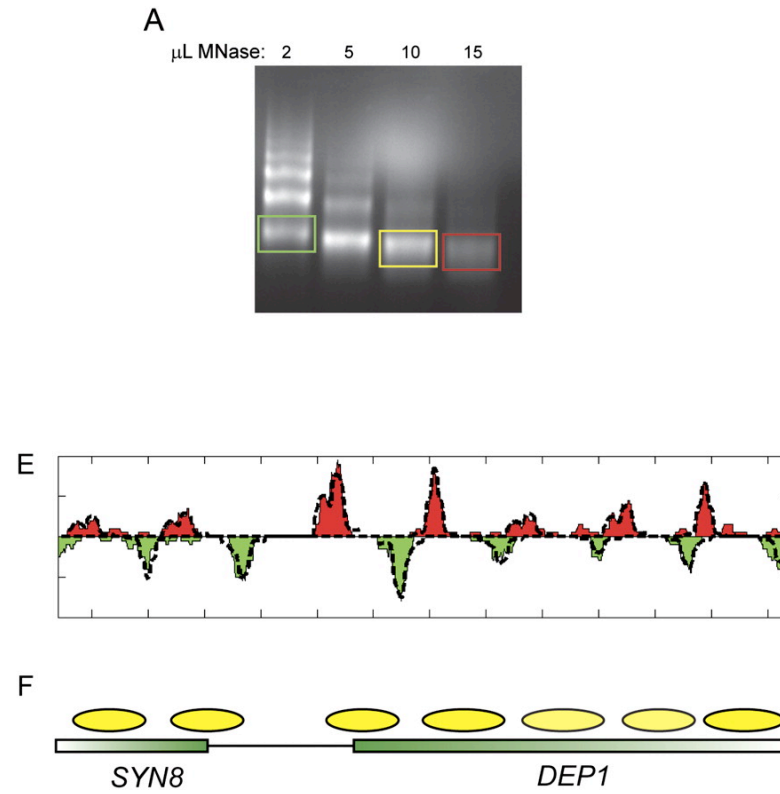
## High-throughput methods to identify DNase HS sites.



# Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE)



## Micrococcal Nuclease digestion (MNase-seq)



## DNA methylation

3 broad classes of assays:

Enzyme based

methylation sensitive restriction  
enzymes

Affinity based

antibodies or other meDNA binding proteins used in  
ChIP-like experiment

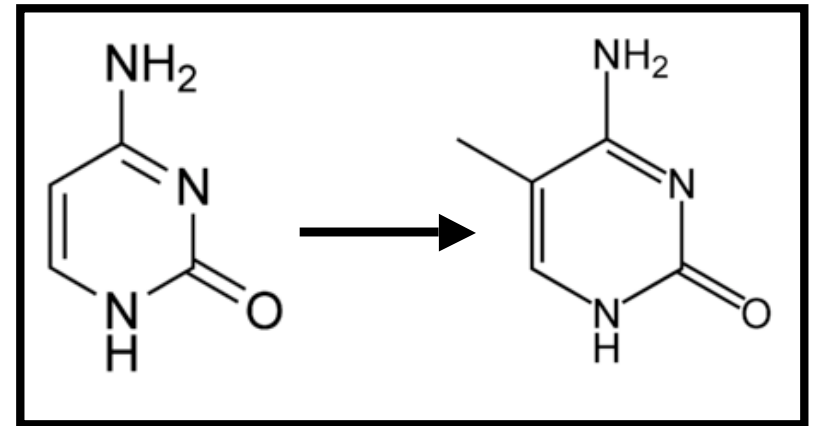
Bisulfite based

$\text{NaHSO}_3$  deaminates unmethylated cytosines to uracils  
but does not affect 5-methylcytosine.

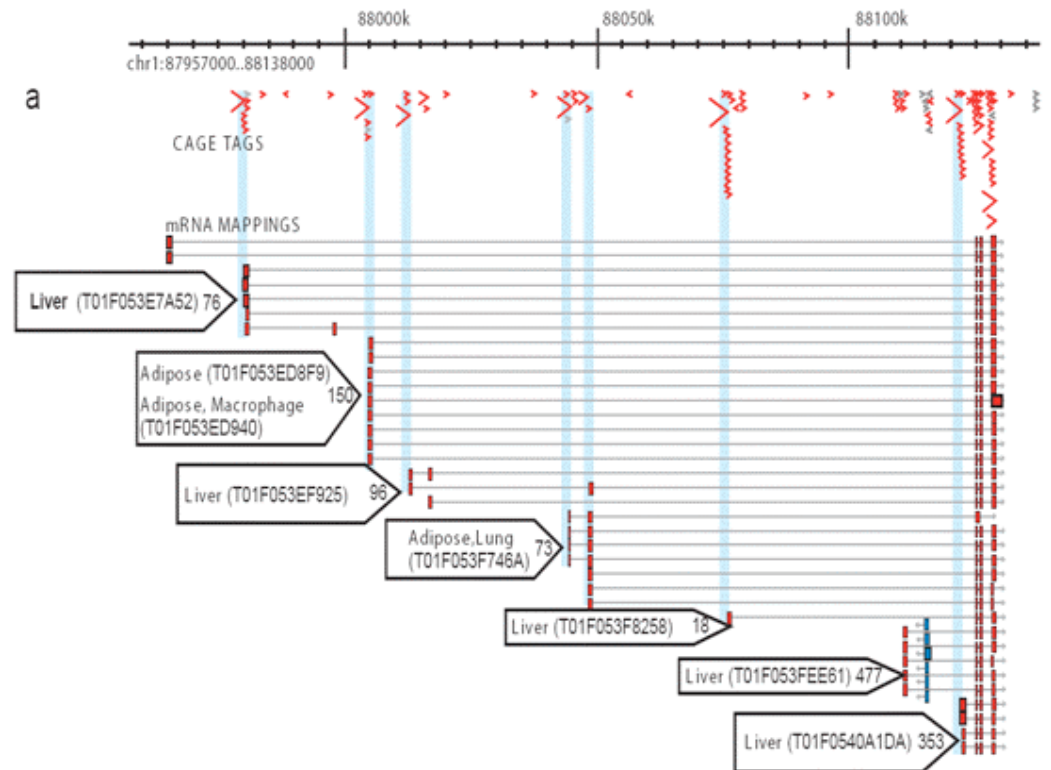
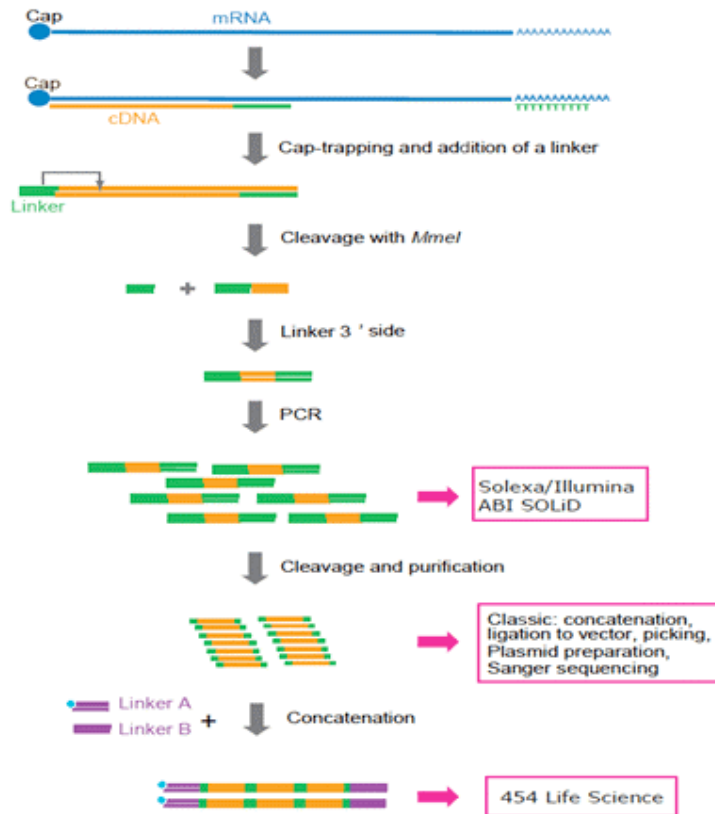
Reduced Representation based on enzymatic digest or  
hybridization enrichment common for cost efficiency in  
large, sparsely methylated (e.g. mammalian) genomes.

Read alignments done to *in silico* bisulfite-converted  
genome.

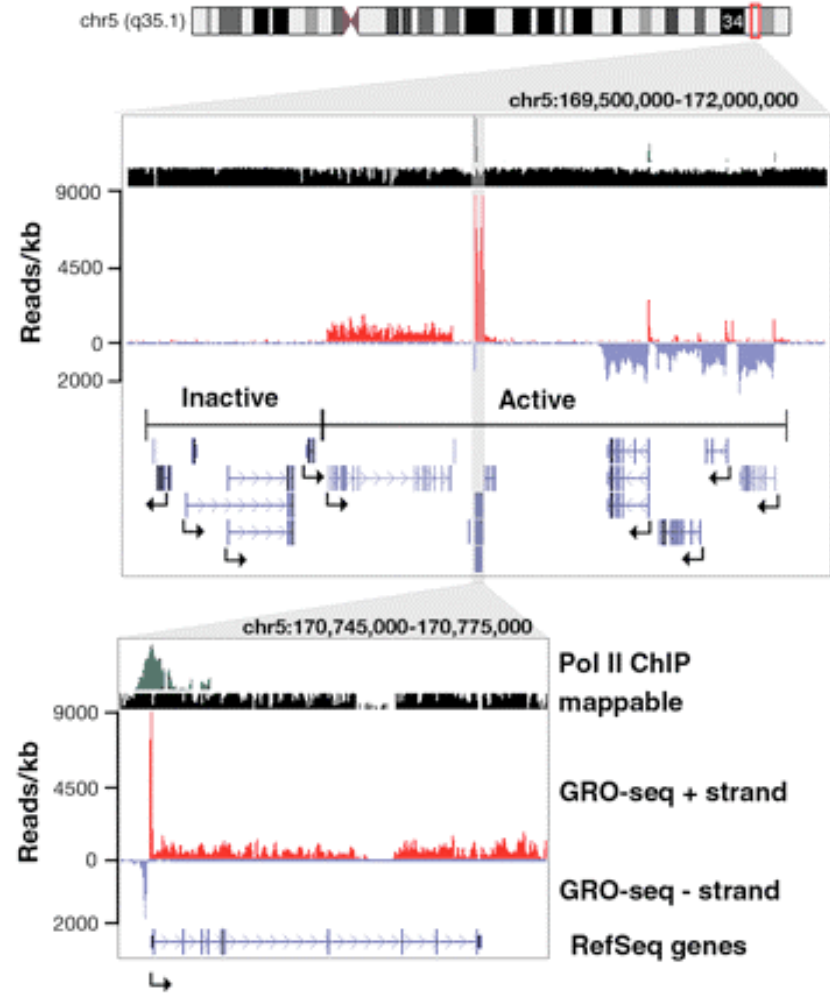
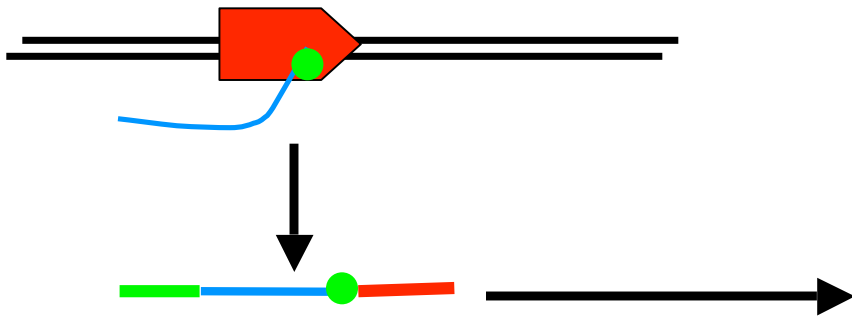
*Direct sequencing of 5th base (future technologies)*



# Cap Analysis of Gene Expression (CAGE)



# Genomic Run On (GRO-seq)



## Web resources for analyzing, viewing, sharing, and collecting genomics data

UCSC Genome Browser (<http://genome.ucsc.edu>)

Galaxy (<http://main.g2.bx.psu.edu>)

NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>)

# The UCSC Genome Browser

## http://genome.ucsc.edu

UCSC Genome Browser Home

http://genome.ucsc.edu/

Burroughs W...nformation NIH Guide: ...TION AWARD Cornell Homepage CALS-IT Help Desk Entrez PubMed UCSC Genom...wser Home Journals Seminars

## UCSC Genome Bioinformatics

**Genomes** - Blat - Tables - Gene Sorter - PCR - VisiGene - Proteome - Session - FAQ - Help

Genome Browser  
ENCODE  
Blat  
Table Browser  
Gene Sorter  
In Silico PCR  
Genome Graphs  
Galaxy  
VisiGene  
Proteome Browser  
Utilities  
Downloads  
Release Log  
Custom Tracks  
Microbial Genomes  
Mirrors

### About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

If you use [ENCODE](#) or [modENCODE](#) data, or are interested in exploring it in the future, we invite you to take the [2010 ENCODE/modENCODE Usability Survey](#). Your input will help us to make this data more accessible to the scientific community. Thank you!

### News News Archives ►

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

**24 Mar. 2010 - African Savannah Elephant Genome Browser Released**

We have released a Genome Browser for the African savannah elephant, *Loxodonta africana*. This assembly (UCSC version loxAfr3, Broad loxAfr3) was produced by the [Broad Institute](#), Cambridge, MA, USA. The elephant was the first member of Afrotheria to be sequenced. Afrotheria is the deepest node of Eutheria, and the elephant sequence should be useful in reconstructing the ancestral eutherian genome.

This draft of the elephant genome has a size of approximately 3 Gb with 7X coverage. The assembly comprises 2352 scaffolds and chrM (mitochondrial DNA). For more information on the assembly, see the Broad Institute [Elephant Genome Project](#) page.

Bulk downloads of the sequence and annotation data are available via the Genome Browser [FTP server](#) or the [Downloads](#) page. These data have [specific conditions for use](#).

Many thanks to the Broad Institute for the elephant assembly data. The loxAfr3 annotation tracks were generated by UCSC and collaborators.



# Uploading data to UCSC

Human (Homo sapiens) Genome Browser Gateway

http://genome.ucsc.edu/cgi-bin/hgGateway

Burroughs W...nformation NIH Guide: ...TION AWARD Cornell Homepage CALS-IT Help Desk Entrez PubMed UCSC Genom...wser Home Journals Seminars

Home Genomes Blat Tables Gene Sorter PCR Session FAQ Help

## Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade genome assembly position or search term gene image width

Mammal Human Mar. 2006 (NCBI36/hg18) chr1:110,202,796-110,263,081 1020 submit

[Click here to reset the browser user interface settings to their defaults.](#) [Take ENCODE Survey](#)

manage custom tracks configure tracks and display clear position


### About the Human Mar. 2006 (NCBI36/hg18) assembly [\(sequences\)](#)

The March 2006 human reference sequence (NCBI Build 36.1) was produced by the International Human Genome Sequencing Consortium.

#### Sample position queries

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, or a cytological band, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175 15q11;15q13	Displays region between STS markers RH18061 and RH80175 or chromosome bands 15q11 to 15q13. This syntax may also be used for other range queries, such as between uniquely-determined ESTs, mRNAs, refSeqs, etc.



*Homo sapiens*  
(Graphic courtesy of [CBSE](#))

# Uploading data to UCSC

http://genome.ucsc.edu/cgi-bin/hgCustom?hgsid=155173412

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

## Add Custom Tracks

clade  genome  assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [BED](#), [bigBed](#), [BEDGRAPH](#), [GFF](#), [GTF](#), [WIG](#), [bigWig](#), [MAF](#), [BAM](#) or [PSL](#) formats. To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#). URLs for data in the bigBed and bigWig formats must be embedded in a track line in the box below. Publicly available custom tracks are listed [here](#). Examples are [here](#).

Paste URLs or data: Or upload:  no file selected

Optional track documentation: Or upload:  no file selected

Click [here](#) for an HTML document template that may be used for Genome Browser track descriptions.

## Loading Custom Tracks

An annotation data file in one of the supported custom track [formats](#) may be uploaded by any of the following methods:

- (Preferred) Enter one or more [URLs](#) for custom tracks (one per line) in the data text box. The Genome Browser supports both the HTTP and FTP (passive-only) protocols.
- Click the "Browse" button directly above the URL/data text box, then choose a custom track file from your local computer, or type the pathname of the file into the "upload" text box adjacent to the "Browse" button. The custom track data may be compressed by any of the following programs: gzip (.gz), compress (.Z), or bzip2 (.bz2). Files containing compressed data must include the appropriate suffix in their names.

# Uploading data to UCSC

Manage Custom Tracks

http://genome.ucsc.edu/cgi-bin/hgCustom

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

### Manage Custom Tracks

genome  assembly  [hg18]

Name	Description	Type	Doc	Items	Pos	delete	
<a href="#">MACS peaks for HeLa_STAT1</a>	MACS peaks for HeLa_STAT1	bed		5902	chr1:	<input type="checkbox"/>	<input type="button" value="add custom tracks"/> <input type="button" value="go to genome browser"/> <input type="button" value="go to table browser"/>

### Managing Custom Tracks

This section provides a brief description of the columns in custom track management table. For more details about managing custom tracks, see the [Genome Browser User's Guide](#).

- **Name** - a hyperlink to the update page where you can edit your track data.
- **Description** - the value of the "description" attribute from the track line, if present. If no description is included in the input file, this field contains the track name.
- **Type** - the track type, determined by the Browser based on the format of the data.
- **Doc** - displays "Y" (Yes) if a description page has been uploaded for the track; otherwise the field is blank.
- **Items** - the number of data items in the custom track file. An item count is not displayed for tracks lacking individual items (e.g. wiggle format data).
- **Pos** - the default chromosomal position defined by the track file in either the browser line "position" attribute or the first data line. Clicking this link opens the Genome Browser or Table Browser at the specified position (note: only the chromosome name is shown in this column). The Pos column remains blank if the track lacks individual items (e.g. wiggle format data) and the browser line "position" attribute hasn't been set.

# Modifying display settings

The screenshot displays the UCSC Genome Browser interface for Human chromosome 1, region 110,202,796-110,263,081. The browser title is "Human chr1:110,202,796-110,263,081 - UCSC Genome Browser v227". The main heading is "UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly".

The interface includes a navigation bar with links: Home, Genomes, Blat, Tables, Gene Sorter, PCR, DNA, Convert, Ensembl, NCBI, PDF/PS, Session, and Help. Below this is a search bar with "position/search" set to "chr1:110,142,510-110,323,367" and "gene" set to an empty field. The "size" is 180,858 bp. There are zoom controls (1.5x, 3x, 10x) and a "configure" link.

The main display area shows several tracks:

- Scale: chr1: 110200000 to 110250000 (50 kb)
- STAT1\_control\_chr1
- STAT1\_treat\_chr1
- MACS\_peak\_248, MACS\_peak\_249, MACS\_peak\_250, MACS\_peak\_251
- Gap Locations
- RefSeq Genes: CSF1, CSF1, CSF1
- Vertebrate Multiz Alignment & Conservation (44 Species)
- Multiz Alignments of 44 Vertebrates
- Simple Nucleotide Polymorphisms (dbSNP build 130)

Below the tracks are navigation controls: "move start" and "move end" with "2.0" zoom, and a "Click on a feature for details..." instruction. There are buttons for "default tracks", "hide all", "manage custom tracks", "configure", "reverse", and "refresh".

A section titled "Custom Tracks" contains a "refresh" button circled in red. Below it, a table lists tracks with their display modes:

Track Name	Display Mode
HeLa STAT1 control chr1	dense
HeLa STAT1 treat chr1	dense
MACS peaks for HeLa STAT1	full
ORRegAnno	hide

Below the "Custom Tracks" section is a "Mapping and Sequencing Tracks" section with a "refresh" button. It lists tracks like "Base Position", "Chromosome Band", "STS Markers", "FISH Clones", "Recomb Rate", and "Map Contigs", each with a "dense" or "hide" display mode.

# Modifying display settings

HeLa\_STAT1\_control\_chr1 Track Settings

## Shifted Merged MACS tag counts for every 10 bp

Display mode:

Type of graph:

Track height:  pixels (range: 11 to 128)

Vertical viewing range: min:  max:  (range: 1 to 122)

Data view scaling:  Always include zero:

Transform function: Transform data points by:

Windowing function:  Smoothing window:  pixels

Draw y indicator lines: at y = 0.0:  at y =

[Graph configuration help](#)

# Using publicly available tracks

Human chr1:110,142,510-110,323,367 - UCSC Genome Browser v227

http://genome.ucsc.edu/cgi-bin/hgTracks

Burroughs W...nformation NIH Guide: ...TION AWARD Cornell Homepage CALS-IT Help Desk Entrez PubMed UCSC Genom...wser Home Journals Seminars

**Phenotype and Disease Associations** refresh

**Genes and Gene Prediction Tracks** refresh

<a href="#">UCSC Genes</a> hide	<a href="#">Old UCSC Genes</a> hide	<a href="#">Alt Events</a> hide	<a href="#">Gencode Genes</a> hide	<a href="#">CCDS</a> hide	<a href="#">RefSeq Genes</a> pack
<a href="#">Other RefSeq</a> hide	<a href="#">MGC Genes</a> hide	<a href="#">ORFeome Clones</a> hide	<a href="#">TransMap...</a> hide	<a href="#">Vega Genes</a> hide	<a href="#">Ensembl Genes</a> hide
<a href="#">AceView Genes</a> hide	<a href="#">SIB Genes</a> hide	<a href="#">N-SCAN</a> hide	<a href="#">CONTRAST</a> hide	<a href="#">SGP Genes</a> hide	<a href="#">Genid Genes</a> hide
<a href="#">Genscan Genes</a> hide	<a href="#">Exoniphy</a> hide	<a href="#">Augustus</a> hide	<a href="#">RNA Genes</a> hide	<a href="#">ACEScan</a> hide	<a href="#">EvoFold</a> hide
<a href="#">sno/miRNA</a> hide	<a href="#">Pos Sel Genes</a> hide				

**mRNA and EST Tracks** refresh

<a href="#">Human mRNAs</a> hide	<a href="#">Spliced ESTs</a> hide	<a href="#">Human ESTs</a> hide	<a href="#">Other mRNAs</a> hide	<a href="#">Other ESTs</a> hide	<a href="#">H-Inv</a> hide
<a href="#">UniGene</a> hide	<a href="#">Gene Bounds</a> hide	<a href="#">SIB Alt-Splicing</a> hide	<a href="#">Poly(A)</a> hide	<a href="#">CGAP SAGE</a> hide	

**Expression** refresh

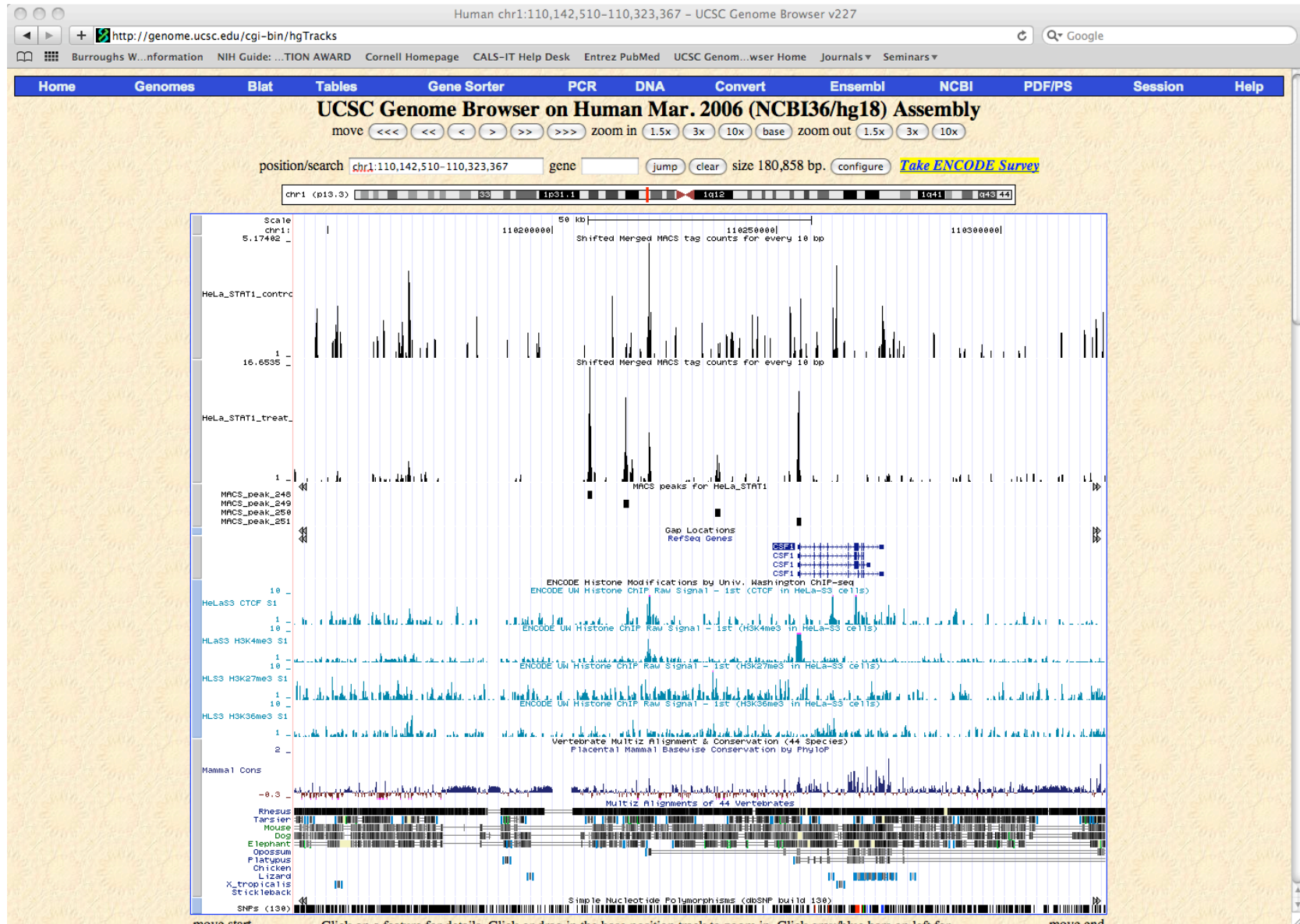
**Regulation** refresh

<a href="#">CpG Islands</a> hide	<a href="#">Broad Histone</a> hide	<a href="#">EIO/ICVINAS</a> hide	<a href="#">Eponine TSS</a> hide	<a href="#">FirstEF</a> hide	<a href="#">GIS ChIP-PET</a> hide
<a href="#">HAIB Methyl-seq</a> hide	<a href="#">HAIB Methyl27</a> hide	<a href="#">HAIB TFBS</a> hide	<a href="#">NHGRI Bi-Pro</a> hide	<a href="#">NHGRI NRE</a> [No data-mr1]	<a href="#">Open Chromatin</a> hide
<a href="#">SUNY RBP</a> hide	<a href="#">SwitchGear TSS</a> hide	<a href="#">TFBS Conserved</a> hide	<a href="#">TS miRNA sites</a> hide	<a href="#">UW DNaseI QS</a> hide	<a href="#">UW Histone</a> full
<a href="#">Vista Enhancers</a> hide	<a href="#">Yale TFBS</a> hide	<a href="#">7X Reg Potential</a> hide	<a href="#">FOX2 CLIP-seq</a> hide	<a href="#">LI/UCSD TAFI...</a> hide	<a href="#">NKT Nuc Lamina...</a> hide
<a href="#">Nucleosome Occupancy...</a> hide	<a href="#">Uppsala ChIP...</a> hide				

**Comparative Genomics** refresh

<a href="#">Conservation</a> full	<a href="#">28-Way Cons</a> hide	<a href="#">28-Way Base Cons</a> hide	<a href="#">28-Way Most Cons</a> hide	<a href="#">17-Way Cons</a> hide	<a href="#">17-Way Most Cons</a> hide
<a href="#">Cons Indels MmCf</a> hide	<a href="#">Tetraodon Ecores</a> hide	<a href="#">Marmoset Chain/Net</a> hide	<a href="#">Rhesus Chain/Net</a> hide	<a href="#">Guinea pig Chain/Net</a> hide	<a href="#">Orangutan Chain/Net</a> hide

# Using publicly available tracks



# Galaxy

http://main.g2.mx.psu.edu

The screenshot shows the Galaxy web interface. The browser address bar displays <http://main.g2.mx.psu.edu/>. The navigation bar includes links for **Analyze Data**, **Workflow**, **Data Libraries**, **Help**, and **User**. A left sidebar lists various tools such as **Get Data**, **Send Data**, **ENCODE Tools**, **Lift-Over**, **Text Manipulation**, **Convert Formats**, **FASTA manipulation**, **Filter and Sort**, **Join, Subtract and Group**, **Extract Features**, **Fetch Sequences**, **Fetch Alignments**, **Get Genomic Scores**, **Operate on Genomic Intervals**, **Statistics**, **Graph/Display Data**, **Regional Variation**, **Multiple regression**, **Multivariate Analysis**, **Evolution**, **Metagenomic analyses**, **EMBOSS**, **NGS TOOLBOX BETA**, **NGS: QC and manipulation**, **NGS: Mapping**, **NGS: SAM Tools**, and **NGS: Peak Calling**.

The main content area features a central announcement: **Here is what's happening...** with a box containing **Galaxy Pages Spring 2010** and the text "A new standard for reproducible research". Below this is a link: "Access or analyze Southern African Genomes through Galaxy!".

A **Timeline for upcoming Quickies** section lists:

- Apr 5 - Mapping of single 454 reads with lastZ
- Apr 12 - Mapping of paired-end 454 reads with lastZ

The **Live Quickies** section shows four cards for different analysis tasks: **Mapping: ...**, **Illumina mapping: Paired Ends** (Galactic quickie # 12), **Basic fastQ manipulation:** (Galactic quickie # 13), and **Advanced fastQ manipulation:** (Galactic quickie # 14).

At the bottom, it states: "The Galaxy team is a part of BX at Penn State. This project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, and The Institute for CyberScience at Penn State. Galaxy build: \$Rev 1733:a4214de3752e\$".

The right sidebar shows the **History** panel with an "Options" dropdown. It contains a warning: "You are currently viewing a deleted history!". Below are job entries:

- 5: Base Coverage on data 4**  
1 lines, format: txt, database: dm3  
Info:  
39823406
- 4: Merge on data 3**  
6,722 regions, format: interval, database: dm3  
Info:  
| display at UCSC main | view in GeneTrack  

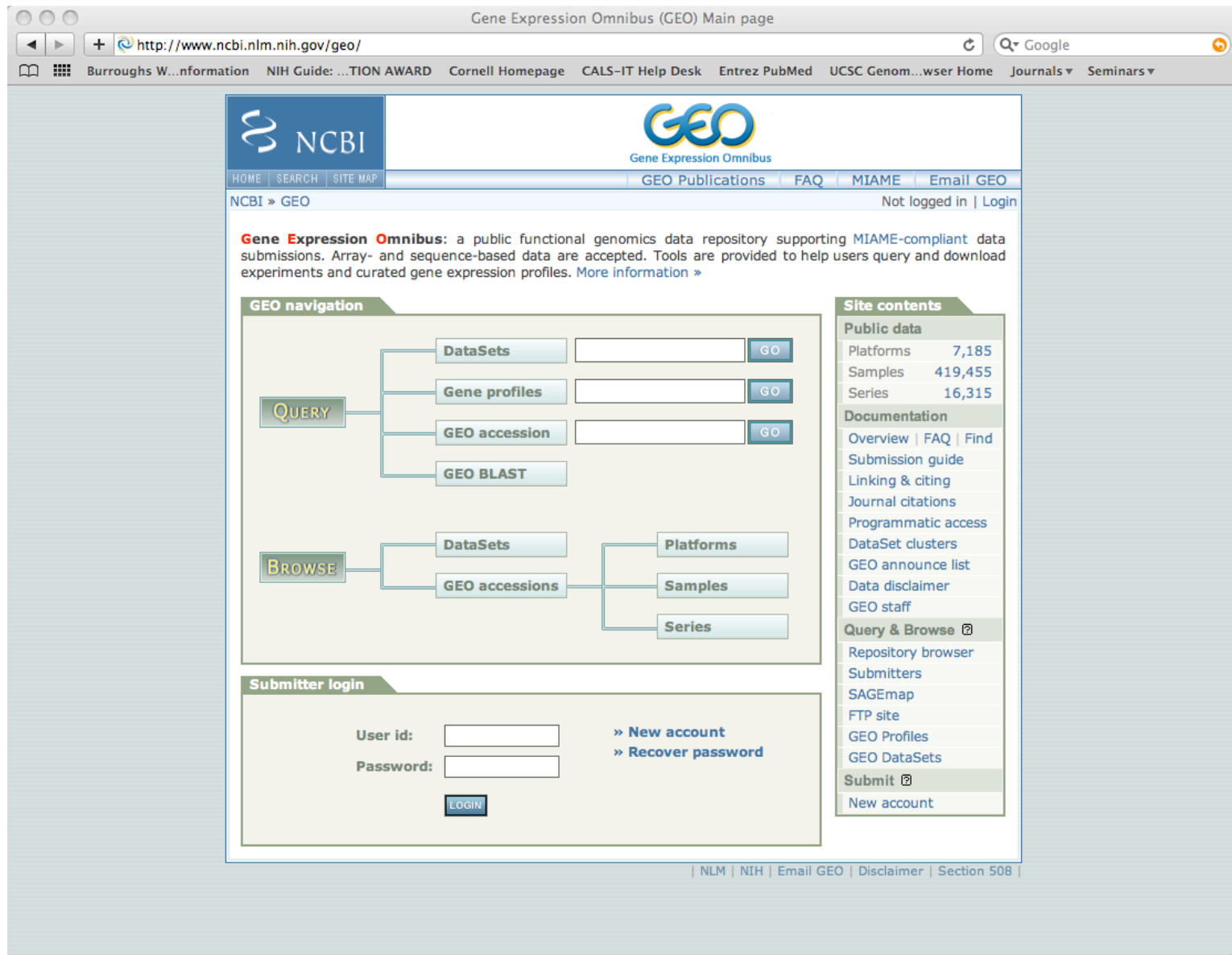
1.Name	2.Chrom	3.Start	5.End
chr2Rhet	.	41260	42381
chr2Rhet	.	117186	117594
chr2Rhet	.	198175	247809
chr2Rhet	.	373139	373379
chr2Rhet	.	390775	392988
- 3: UCSC Main on D. melanogaster: flyBaseGene (genome)**  
10,583 regions, format: interval, database: dm3  
Info: UCSC Main on D. melanogaster: flyBaseGene (genome)  
| display at UCSC main | view in GeneTrack  

1.Name	2.Chrom	3.Strand	4.Start
#filter: flyBaseGene.strand = '+'			



# NCBI Gene Expression Omnibus (GEO)

<http://www.ncbi.nlm.nih.gov/geo/>



Gene Expression Omnibus (GEO) Main page

<http://www.ncbi.nlm.nih.gov/geo/>

Burroughs W...nformation NIH Guide: ...TION AWARD Cornell Homepage CALS-IT Help Desk Entrez PubMed UCSC Genom...wser Home Journals Seminars

NCBI Gene Expression Omnibus

HOME SEARCH SITE MAP GEO Publications FAQ MIAME Email GEO

NCBI » GEO Not logged in | Login

**Gene Expression Omnibus:** a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

**GEO navigation**

**QUERY**

- DataSets  GO
- Gene profiles  GO
- GEO accession  GO
- GEO BLAST

**BROWSE**

- DataSets
- GEO accessions
  - Platforms
  - Samples
  - Series

**Submitter login**

User id:  [» New account](#)

Password:  [» Recover password](#)

**Site contents**

**Public data**

Platforms	7,185
Samples	419,455
Series	16,315

**Documentation**

- [Overview](#) | [FAQ](#) | [Find](#)
- [Submission guide](#)
- [Linking & citing](#)
- [Journal citations](#)
- [Programmatic access](#)
- [DataSet clusters](#)
- [GEO announce list](#)
- [Data disclaimer](#)
- [GEO staff](#)

**Query & Browse**

- [Repository browser](#)
- [Submitters](#)
- [SAGEmap](#)
- [FTP site](#)
- [GEO Profiles](#)
- [GEO DataSets](#)

**Submit**

- [New account](#)

NLM | NIH | Email GEO | Disclaimer | Section 508

## ChIP-seq exercise

Use MACS software to call peaks in STAT1 ChIP-seq experiment in human HeLA S3 cells after inteferon- $\gamma$  stimulation. Analyze diagnostics of run and upload data to UCSC genome browser to look at results.

Office hour:

Friday, April 2

1pm - 2pm

102 Weill conference room