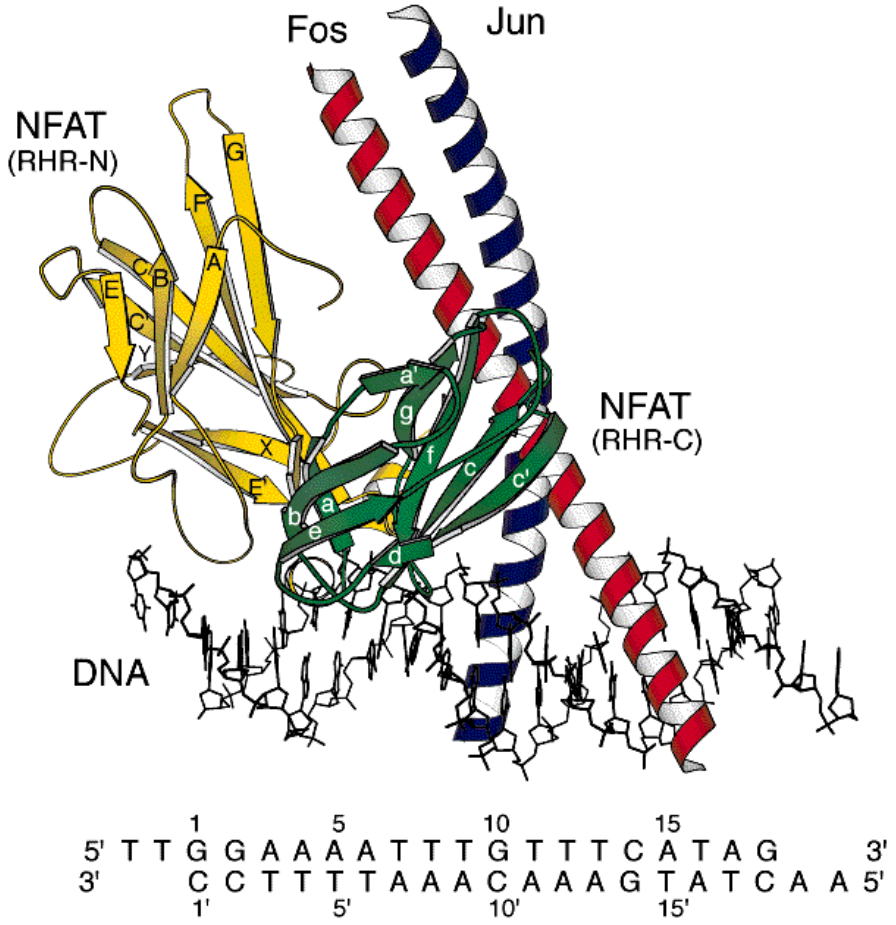
Current technologies

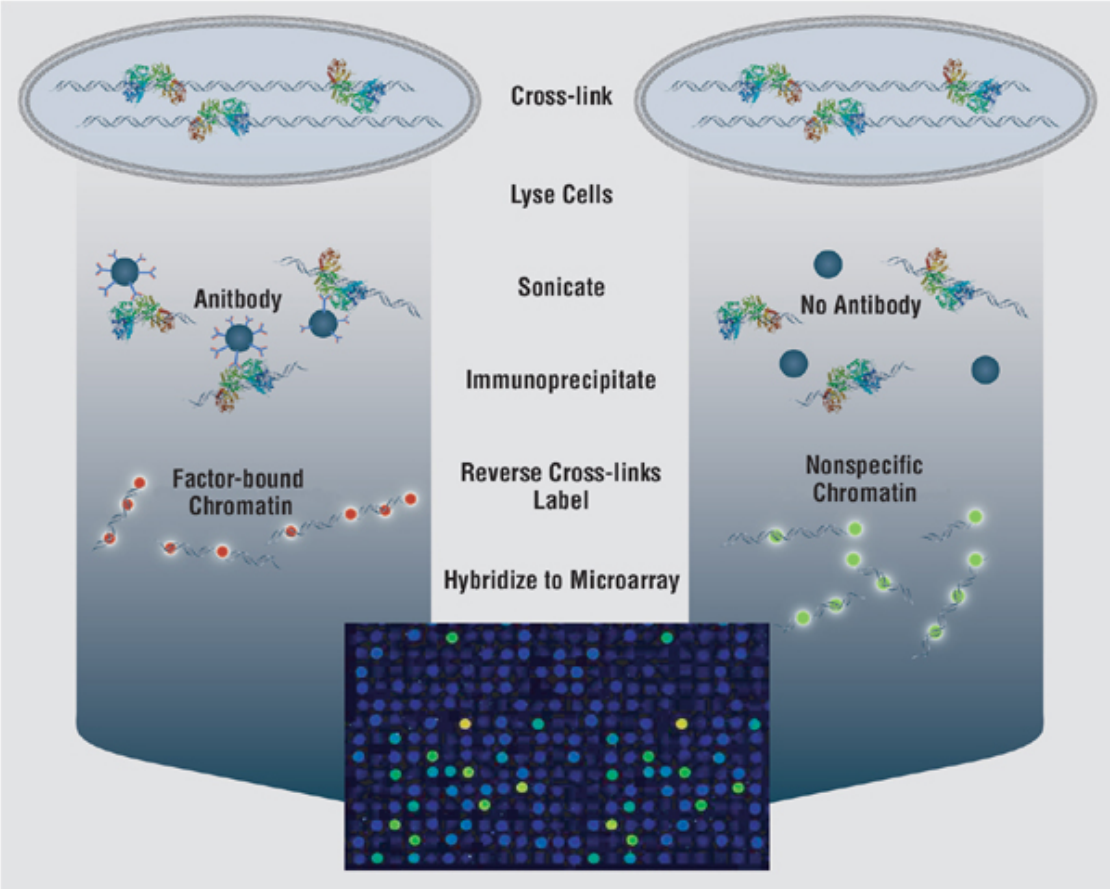
- Detecting TF binding: ChIP + sequencing
- ChIP-seq for TFs, Histone modifications
- Mustersuche
- DNA Methylierung
  - meDIP
  - Bisulfitsequenzierung
- 3D Struktur
- DNA accessibility
- Chromatin segmentation

DNA-Protein interaction

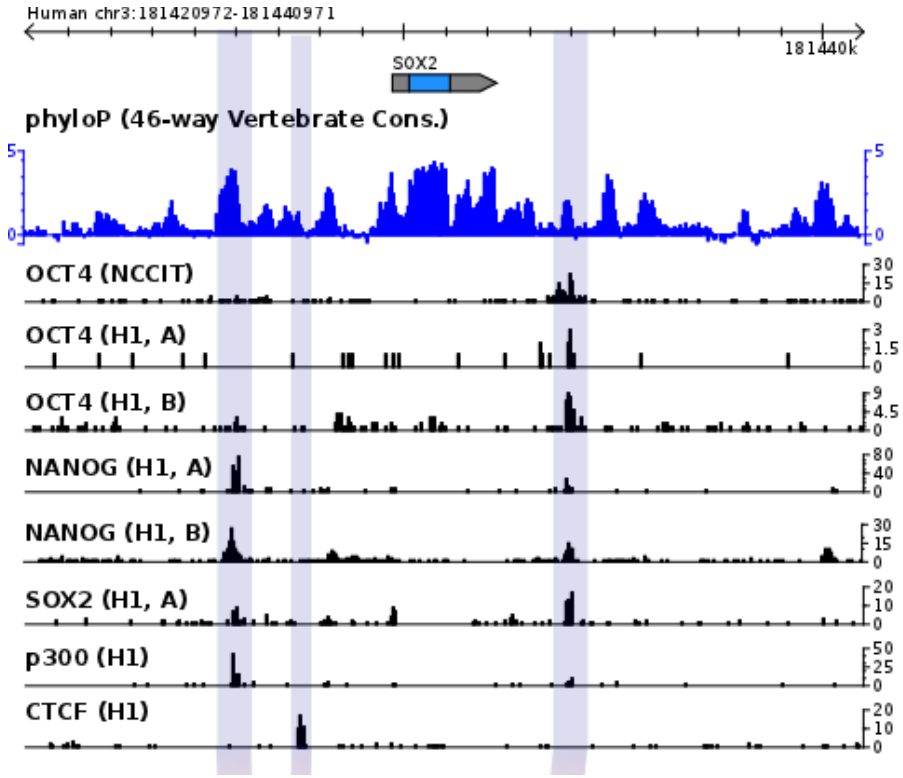


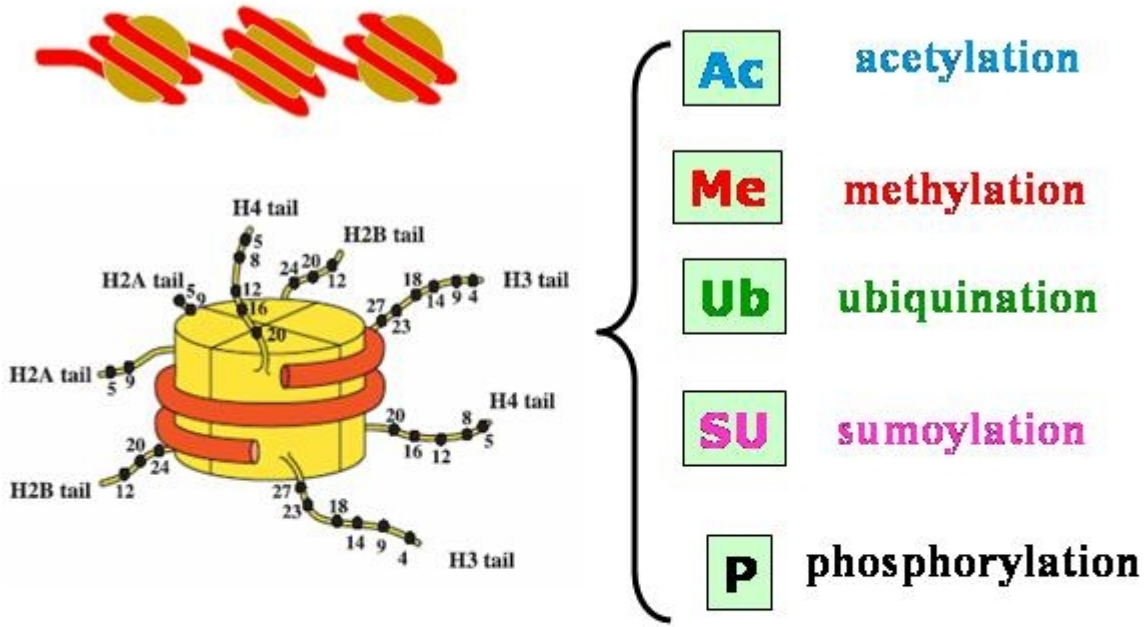
NFAT-AP1-DNA complex
[Harrison, Nature 1998]

Chromatin IP (ChIP chip)



Dynamic view: Specific measurements for a cell type





The figure illustrates nucleosome models and major posttranslational modifications which play essential roles in gene expression regulation and disease processes

Picture from: http://chemistry.gsu.edu/faculty/Zheng/research.html

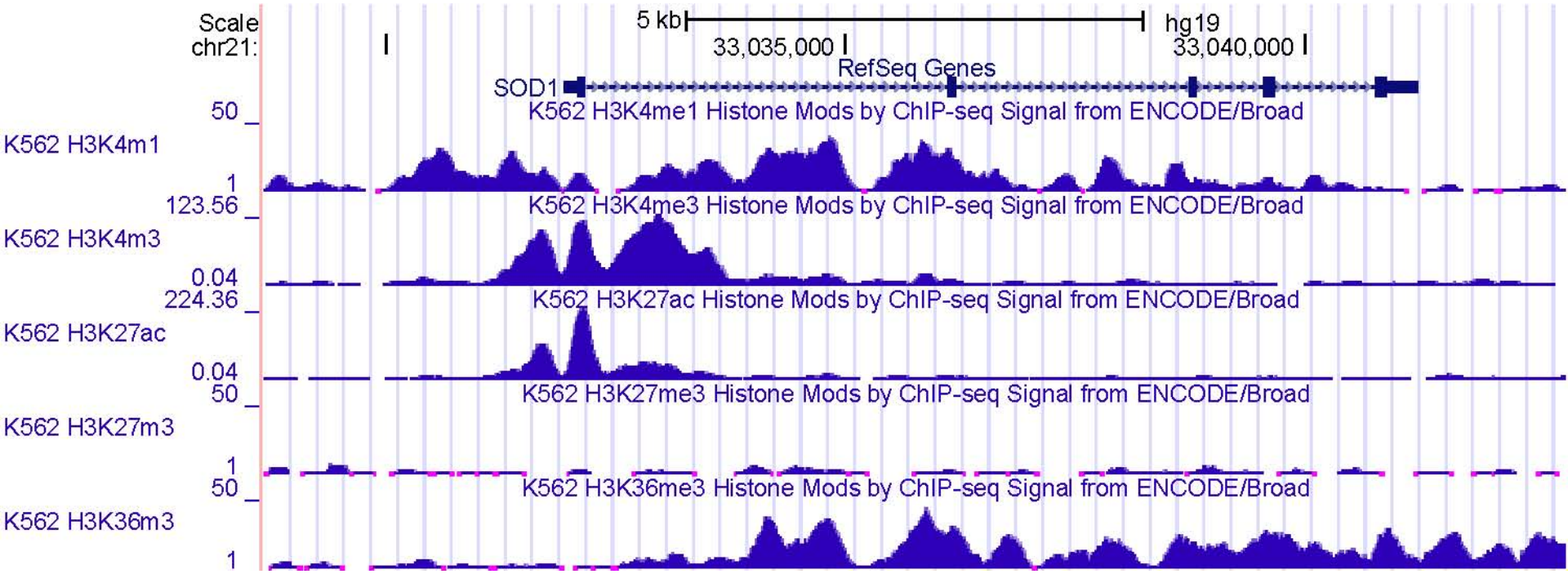

Scale
chr21:
5 kb
hg19
33,035,000
33,040,000
RefSeq Genes
SOD1
50
K562 H3K4me1 Histone Mods by ChIP-seq Signal from ENCODE/Broad
K562 H3K4m1
1
123.56
K562 H3K4me3 Histone Mods by ChIP-seq Signal from ENCODE/Broad
K562 H3K4m3
0.04
224.36
K562 H3K27ac Histone Mods by ChIP-seq Signal from ENCODE/Broad
K562 H3K27ac
0.04
50
K562 H3K27me3 Histone Mods by ChIP-seq Signal from ENCODE/Broad
K562 H3K27m3
1
50
K562 H3K36me3 Histone Mods by ChIP-seq Signal from ENCODE/Broad
K562 H3K36m3
1

Table S2A

Annotation of Regulatory Factors used in this analysis

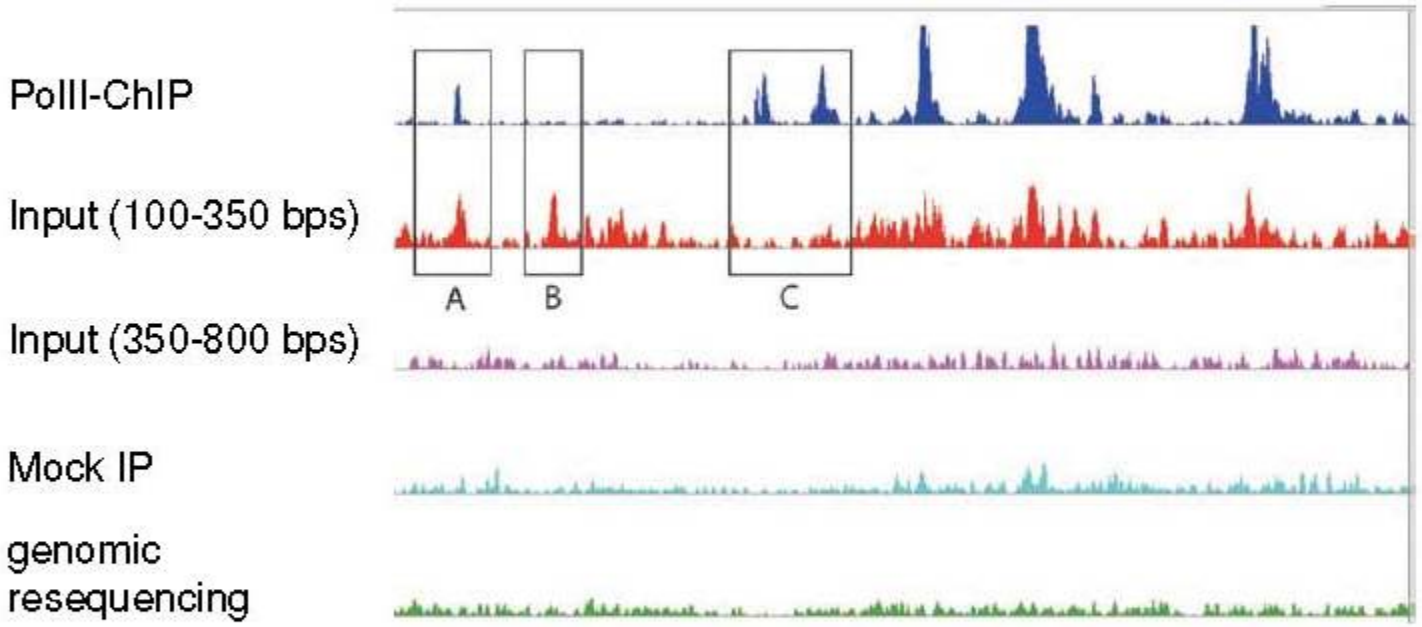
Factor	Alternate Factor Name	Major Class	TF family	TF Domains	Census category	Cell Line: GM 12878	K562	H1-hESC	HeLa-S3	Hep G2	Other	Total Lines	More detailed annotation for non-TFSS
ATF3	-	TFSS	bZIP		a	X	X	X		X		4	
BATF	-	TFSS	bZIP		a	X						1	
BCL11A	-	TFSS	ZNF	ZNF-C2H2	b	X		X				2	
BCL3	-	TFSS	Unknown		-	X	X	X				3	
BCLAF1	-	TFSS	Unknown		o	X	X					2	
BDP1	-	general	Homeodomain		x		X		X			2	Pol III associated
BHLHE40	-	TFSS	HLH		a					X		1	
BRCA1	-	TFSS	Unknown		-	X			X			2	TF and DNA repair
BRF1	-	general			-		X		X			2	Pol III associated
BRF2	-	general			-		X		X			2	Pol III associated
CCNT2	-	general			-		X					1	cyclin kinase
CEBPB	-	TFSS	bZIP		a		X		X	X		3	
CHD2	-	chromatin	Homeodomain		x	X	X			X		3	chromatin remodeling helicase
CTBP2	-	TFSS	Unknown		-			X				1	
CTCF	-	TFSS	ZNF	ZNF-C2H2	a	X	X	X	X	X	46	51	
CTCFL	-	TFSS	ZNF	ZNF-C2H2	a		X				0	1	
E2F1	RBAP-1	TFSS	wHTH	TDP | wHTH	a				X		0	1	
E2F4	-	TFSS	wHTH	TDP | wHTH	a		X					2	
E2F6	-	TFSS	wHTH	TDP | wHTH	a		X		X			2	
EBF1	-	TFSS	IPT/TIG		a	X						1	
EGR1	ZIF268	TFSS	ZNF	ZNF-C2H2	a	X	X	X				3	
ELF1	-	TFSS	ETS	ETS | wHTH	a	X	X			X		3	
ELK4	-	TFSS	ETS	ETS | wHTH	a				X		1	2	
EP300	p300	general			-	X	X	X	X	X	2	7	enhancer-related coactivator
ESR1	-	TFSS	NR		a					X		1	
ESRRA	ERRA	TFSS	NR		a						2	2	
ETS1	-	TFSS	ETS	ETS | wHTH	a	X	X					2	
FAM48A	SPT20	chromatin			-	X			X			2	chromatin remodeling complex
FOS	-	TFSS	bZIP		a	X	X		X		1	4	
FOSL1	FRA1	TFSS	bZIP		a		X					1	
FOSL2	FRA2	TFSS	bZIP		a					X		1	
FOXA1	-	TFSS	Forkhead	Forkhead | wHTH	a					X	2	3	
FOXA2	-	TFSS	Forkhead	Forkhead | wHTH	a					X		1	
GABPA	-	TFSS	ETS	ETS | wHTH	a	X	X	X	X	X		5	
GATA1	-	TFSS	ZNF	ZNF-GATA	a		X				1	2	
GATA2	-	TFSS	ZNF	ZNF-GATA	a		X				2	3	
GATA3	-	TFSS	ZNF	ZNF-GATA	a						1	1	
GTF2B	-	general			-		X					1	general Pol II-associated factor
GTF2F1	-	general	wHTH		x		X		X			2	general Pol II-associated factor
HDAC2	RPD3	chromatin			-		X	X		X		3	histone deacetylase
HMGN3	-	chromatin			-		X					1	chromatin structure factor
HNF4A	NR2A1	TFSS	NR		a					X		1	
HNF4G	-	TFSS	NR		a					X		1	
HSF1	-	TFSS	wHTH	Heat shock | wHTH	a					X		1	
IRF1	-	TFSS	wHTH	IRF | wHTH	a		X					1	
IRF3	-	TFSS	wHTH	IRF | wHTH	a	X			X	X		3	

Genome Biol. 2008; 9(9): R137.

Analysis

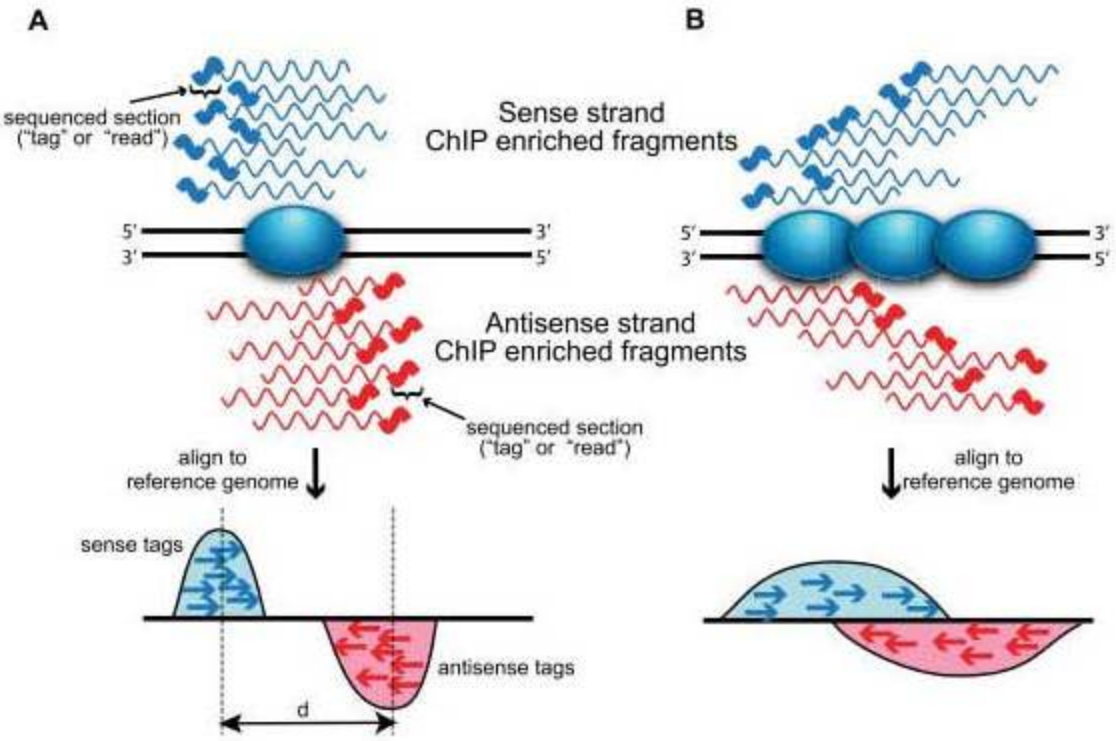
- Read mapping
- Peak detection
- Significance of peaks?
- Differential peaks?

... controls



Auerbach et al. (2009)

... strand dependent bimodality



Wilbanks et al. (2010)

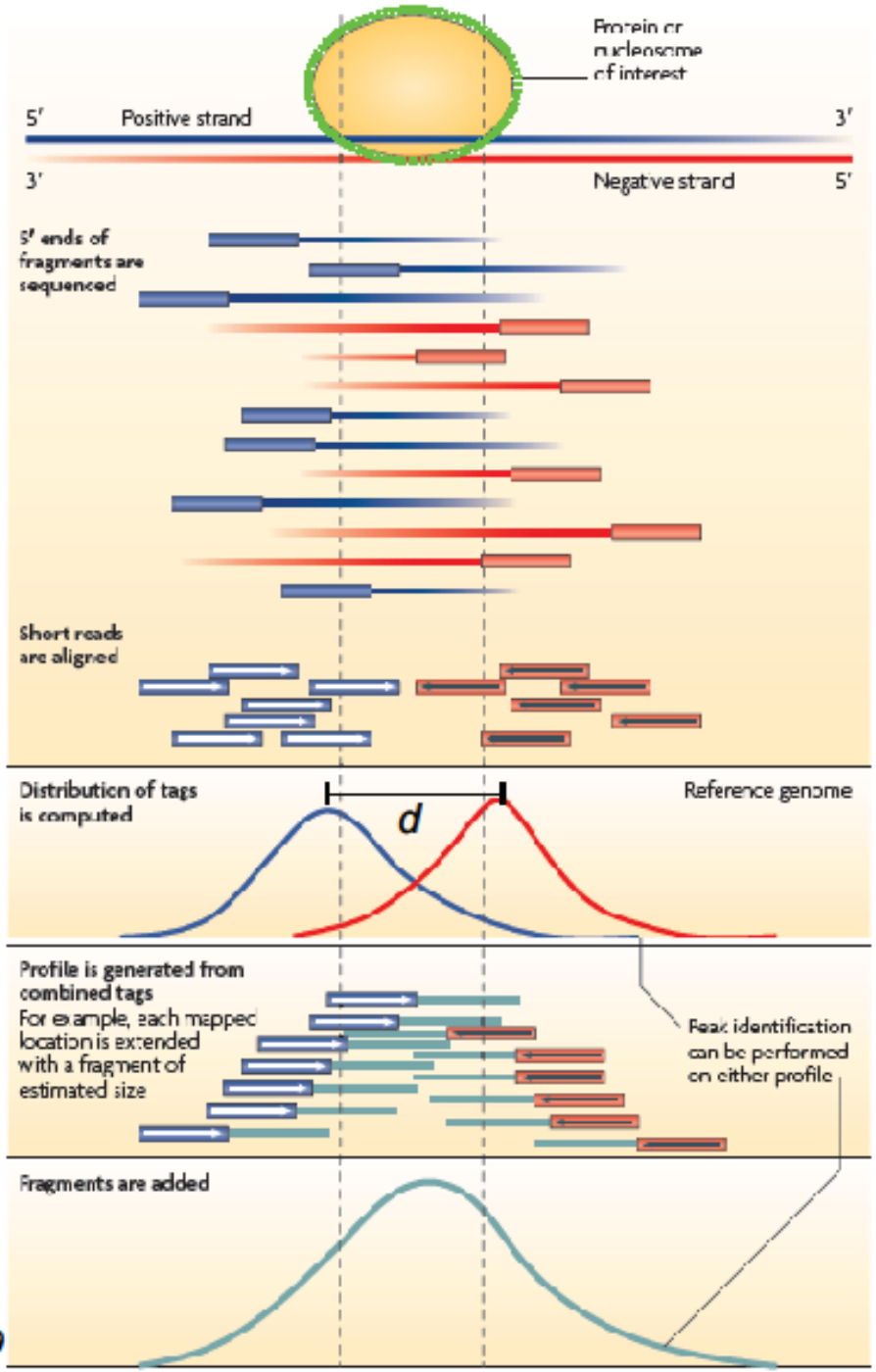
Detecting peaks

Scan the genomic DNA and look for enriched regions using a *window approach*.

Strand-specific patterns emerge and are used to locate the peaks

(1) **extend** the reads to the estimated fragment length, see two bottom panels at right.

(2) **shift** reads towards the middle of the two peaks; d/2



Park, Nat. Rev. Genet., 2009

Analysis

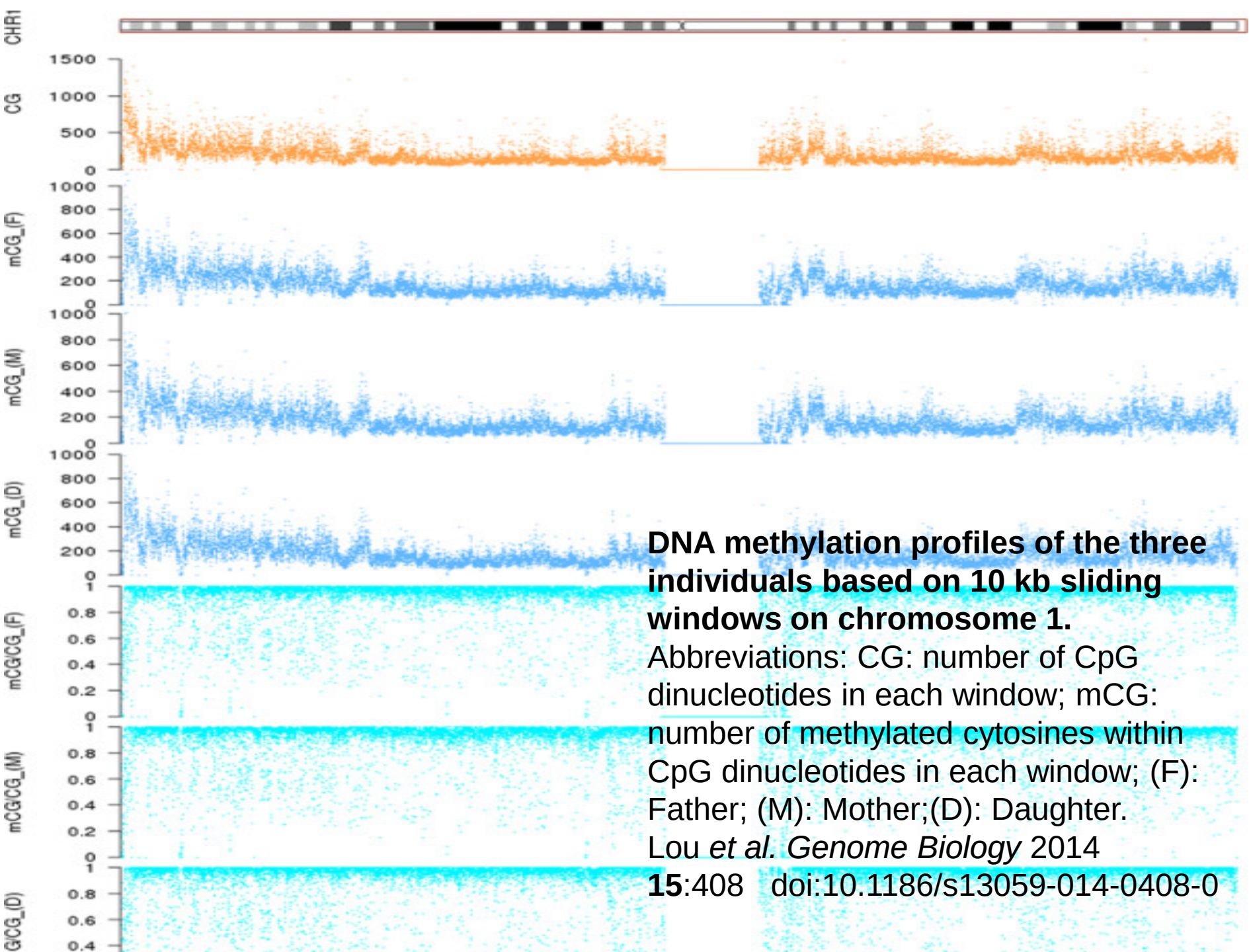
- Read mapping
- Peak detection
- Significance of peaks?
- Differential peaks?

With the current genome coverage of most ChIP-Seq experiments, tag distribution along the genome could be modeled by a Poisson distribution [7]. The advantage of this model is that one parameter, λ_{BG}, can capture both the mean and the variance of the distribution. After MACS shifts every tag by *d*/*2*, it slides *2d* windows across the genome to find candidate peaks with a significant tag enrichment (Poisson distribution *p*-value based on λ_{BG}, default 10-5). Overlapping enriched peaks are merged, and each tag position is extended *d* bases from its center. The location with the highest fragment pileup, hereafter referred to as the *summit*, is predicted as the precise binding location.

From: Zhang et al, **Model-based Analysis of ChIP-Seq (MACS)** , Genome Biol. 2008; 9(9): R137.

DNA methylation

- In CpG dinucleotides the Cytosine may be methylated
- Mapping methylated cytosines using an antibody: MeDIP
- Bisulfite sequencing: Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected -> sequence -> compare to true sequence -> determine methylated positions



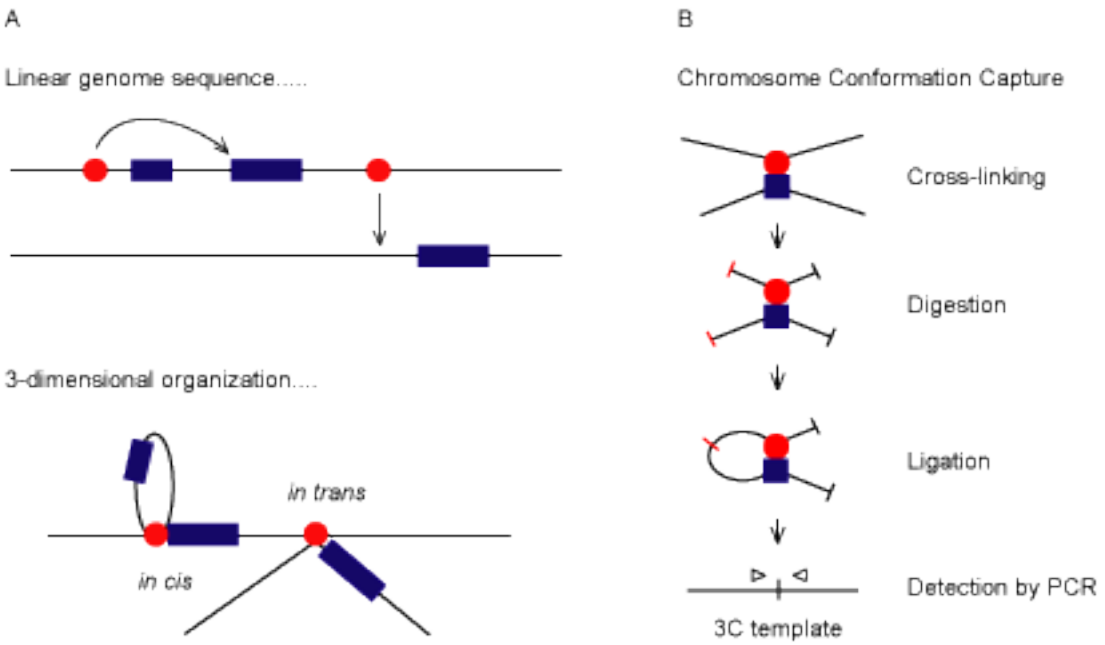
DNA methylation profiles of the three individuals based on 10 kb sliding windows on chromosome 1.
Abbreviations: CG: number of CpG dinucleotides in each window; mCG: number of methylated cytosines within CpG dinucleotides in each window; (F): Father; (M): Mother;(D): Daughter.
Lou *et al. Genome Biology* 2014 **15**:408 doi:10.1186/s13059-014-0408-0

Accessible DNA segments

- DNAaseI hypersensitivity: mapping open regions in chromatin, contain many TFBSs, pinpoint enhancers
- Was performed in more than 100 cell-lines -> distinguish cell-line specific open regions

3D chromosome structure

- Methods to determine which genomic regions are spatially close to which other region: Chromosome Conformation Capture, (ChIA-PET)

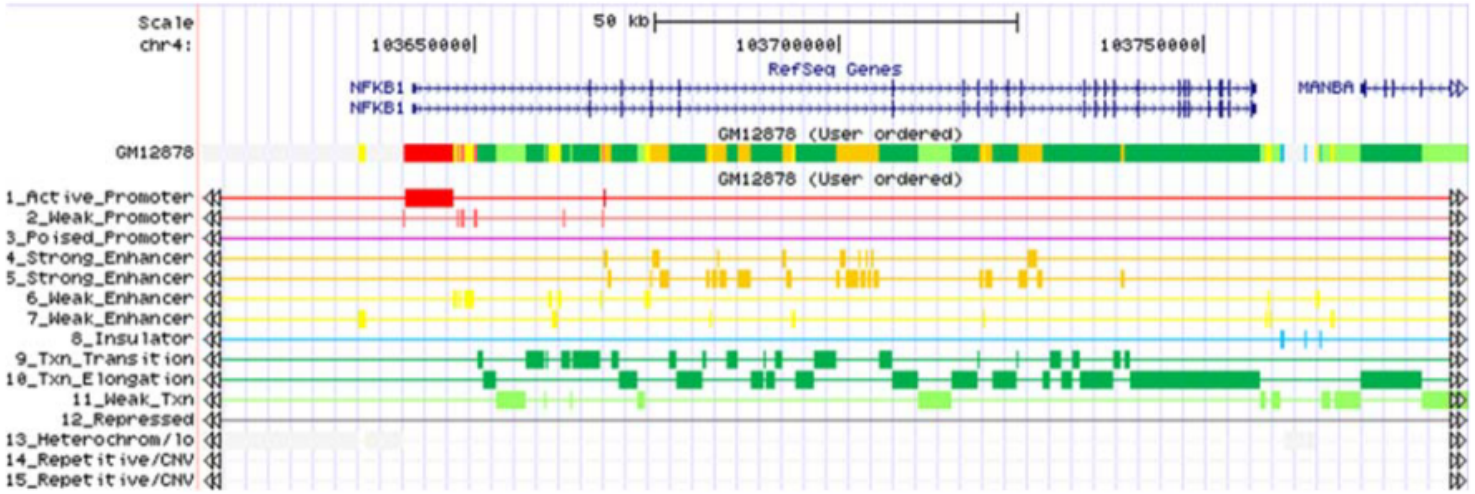


Von derWebseite von Job Dekker (U.Mass.)

Chromatin Segmentation

- **Unsupervised pattern discovery in human chromatin structure through genomic segmentation, by** Michael M Hoffman, …. William Stafford Noble
- Nature Methods 9, 473–476 (2012)
- We trained Segway, a dynamic Bayesian network method, simultaneously on chromatin data from multiple experiments, including positions of histone modifications, transcription-factor binding and open chromatin, all derived from a human chronic myeloid leukemia cell line. In an unsupervised fashion, we identified patterns associated with transcription start sites, gene ends, enhancers, transcriptional regulator CTCF-binding regions and repressed regions. Software and genome browser tracks are at http://noble.gs.washington.edu/proj/segway/.

ChromHMM: Chromatin state discovery and characterization



ChromHMM is software for learning and characterizing chromatin states. ChromHMM can integrate multiple chromatin datasets such as ChIP-seq data of various histone modifications to discover de novo the major re-occuring combinatorial and spatial patterns of marks. ChromHMM is based on a multivariate Hidden Markov Model that explicitly models the presence or absence of each chromatin mark. The resulting model can then be used to systematically annotate a genome in one or more cell types. By automatically computing state enrichments for large-scale functional and annotation datasets ChromHMM facilitates the biological characterization of each state. ChromHMM also produces files with genome-wide

www.nature.com/encode