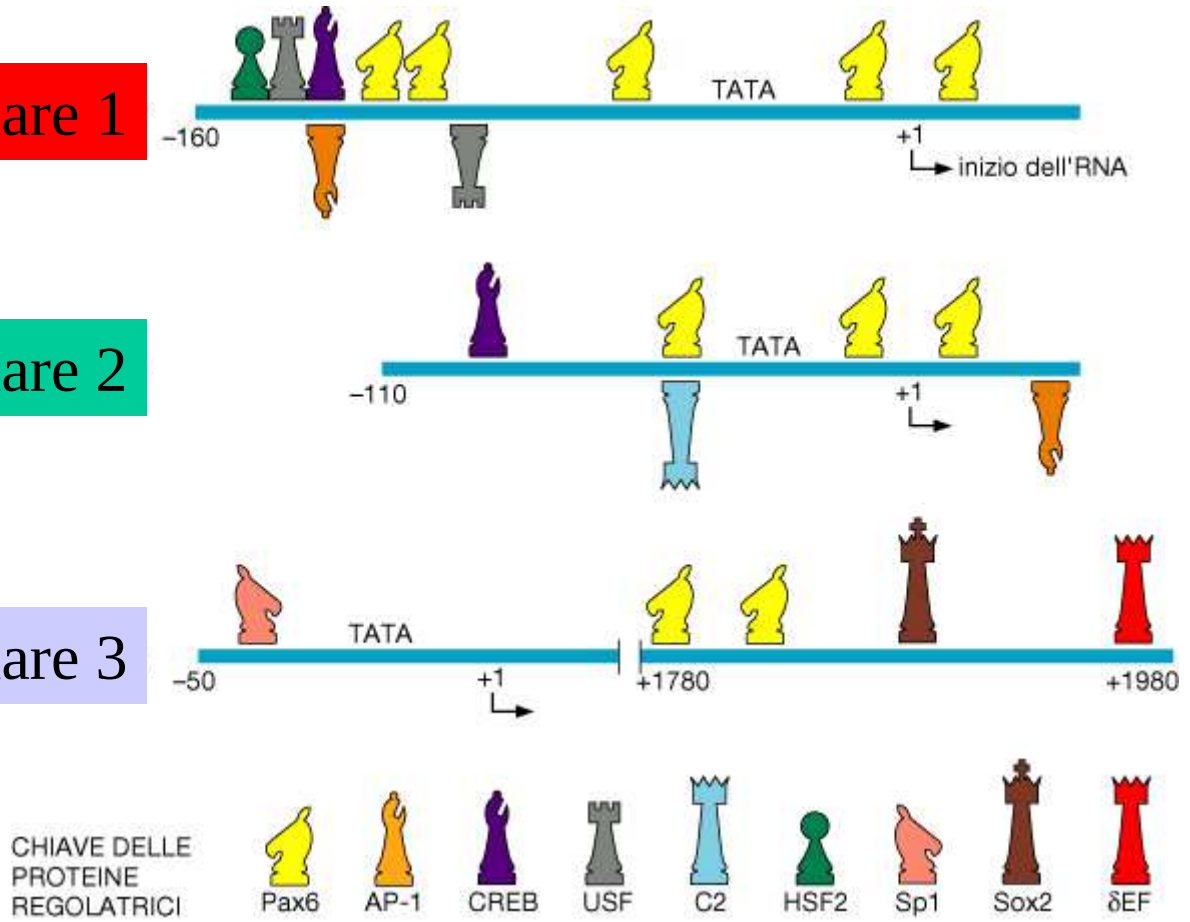
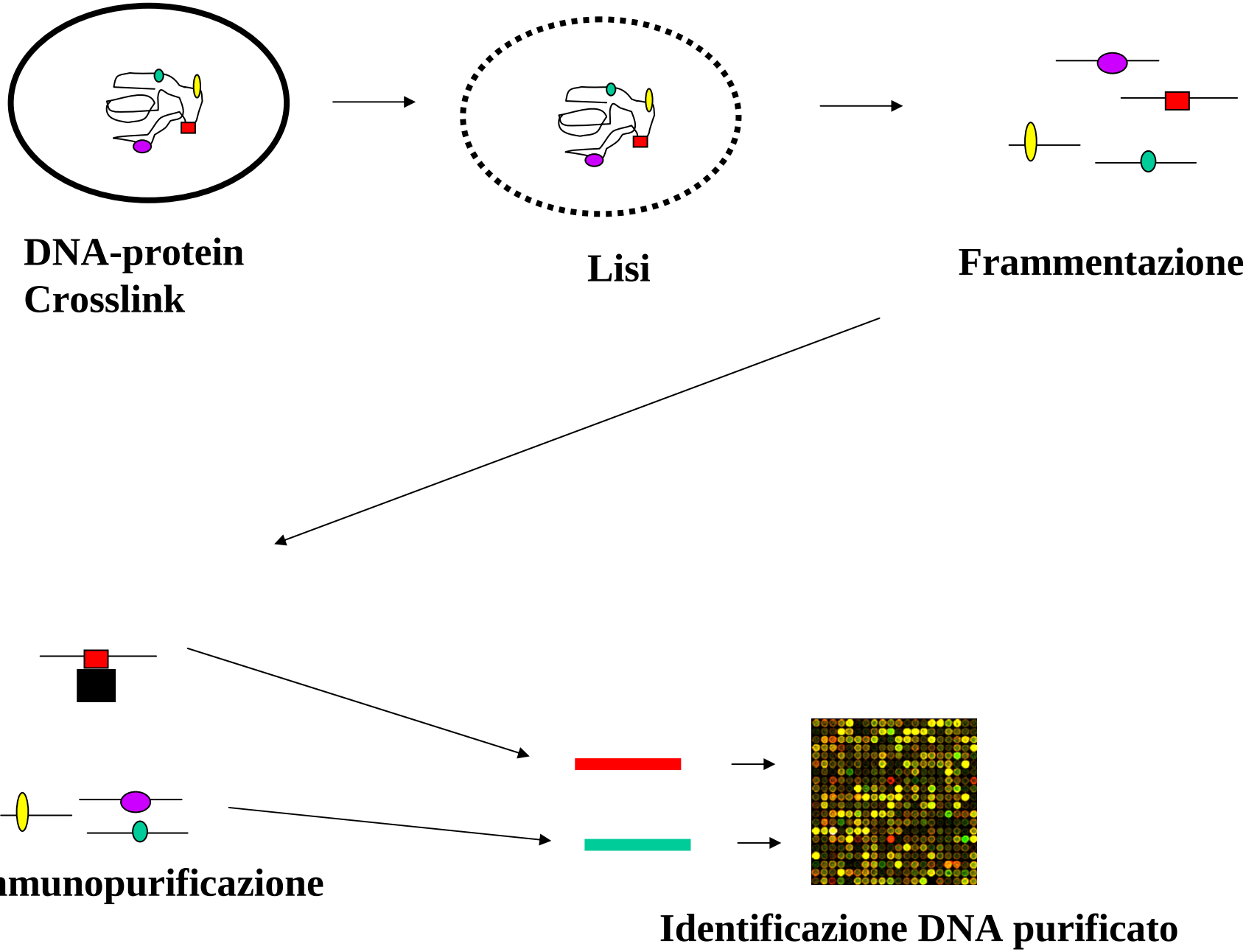
Le proteine regolative variano nei vari tipi cellulari e in funzione degli stimoli ambientali

Tipo cellulare 1

Tipo cellulare 2

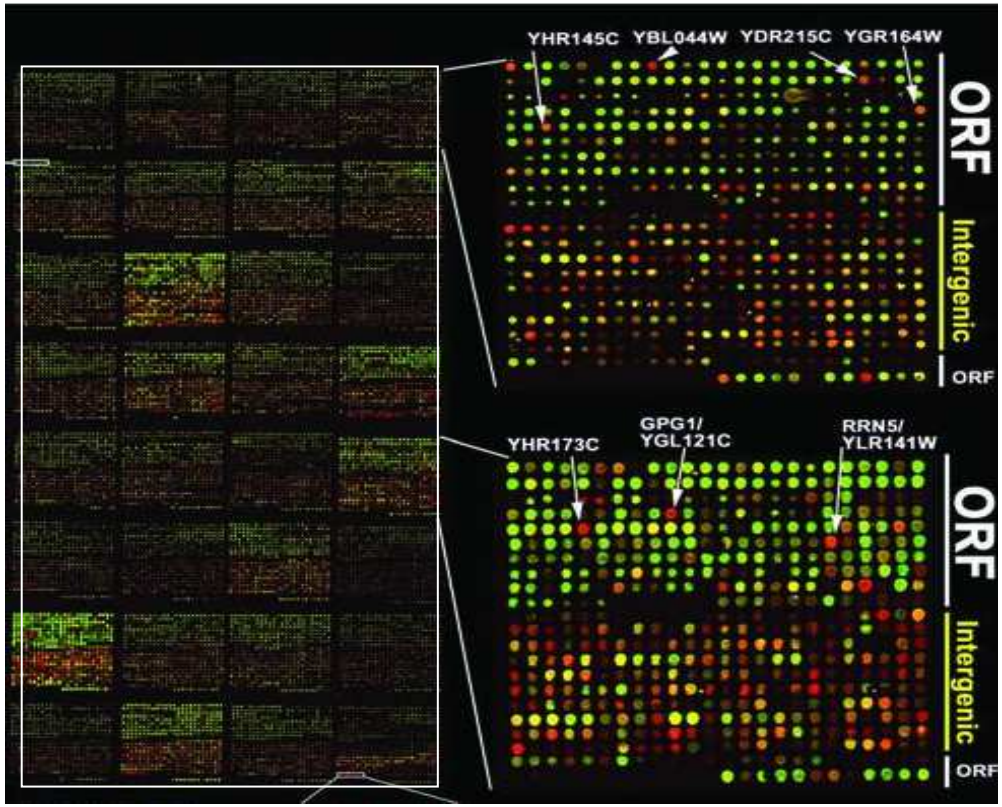
Tipo cellulare 3




DNA-protein
Crosslink
Lisi
Frammentazione
Immunopurificazione
Identificazione DNA purificato

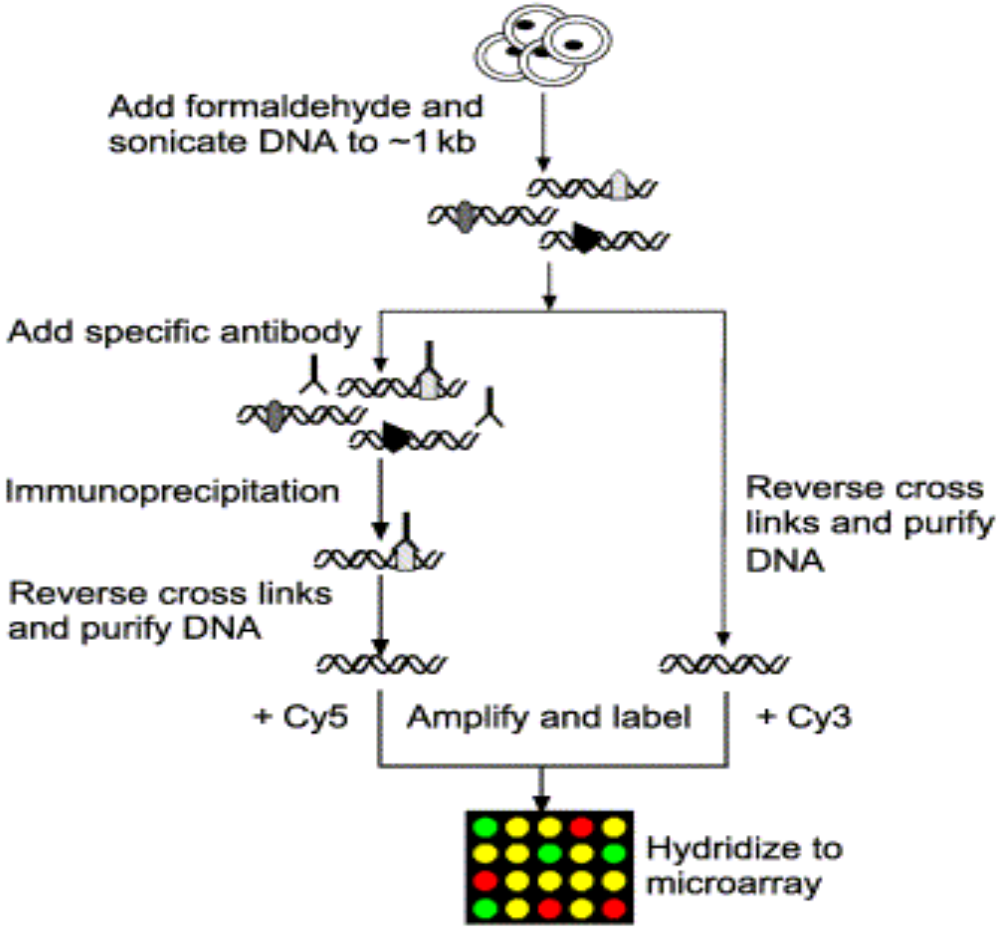
ChIP on chip = Chromatin IP + **microarray**

- Different from the arrays for transcriptome analysis : all the genome on the slide!



ChIP on chip procedure :
a tool to map protein /genome interaction

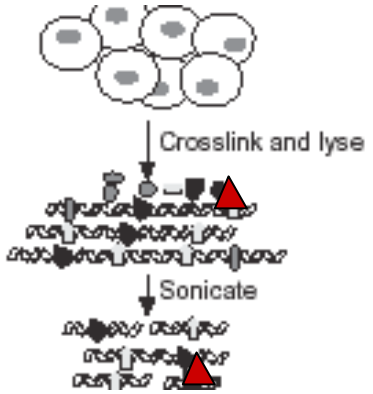
IP



Reference :
WCE

Buck and Lieb, 2004

1. Cross link and chromatin preparation

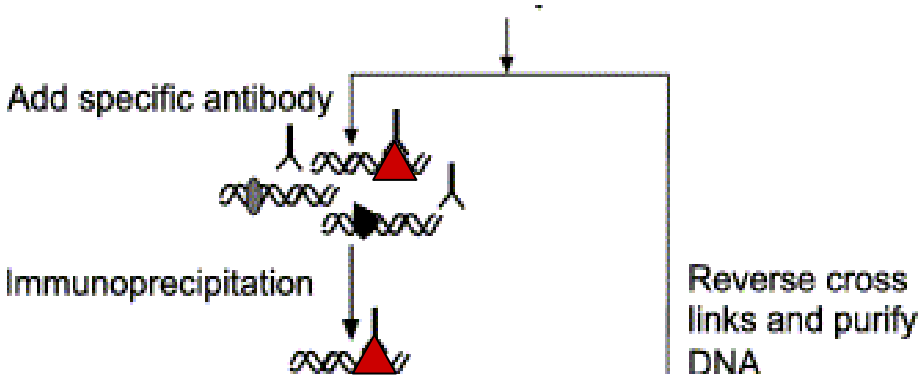


-Cross link with formaldehyde

-Cell lysis

-Sonication of the DNA into 1 kb fragments

2. Immunoprecipitation and cross link reversion



-Over-night IP with antibodies coupled with magnetic beads

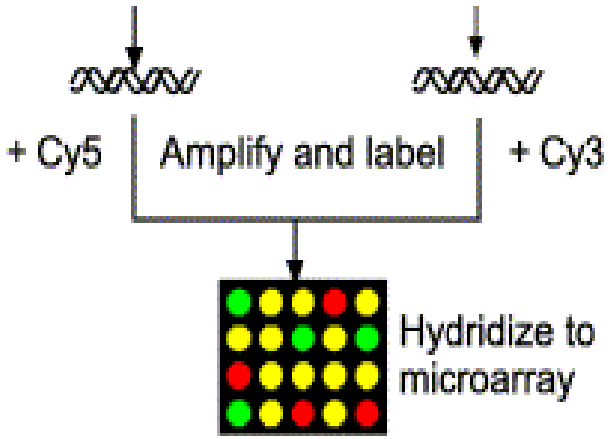
-Specific antibodies are required

-Reversion of the cross link by heating and agitating

IP **WCE**

3. Amplification and labeling of DNA

IP **WCE**



Microarray with the whole genome

-Proteinase and Rnase treatment : to « clean » the DNA

-Amplification (PCR) : random primers or ligation mediated

-Labeling : direct or indirect (aadUTP)

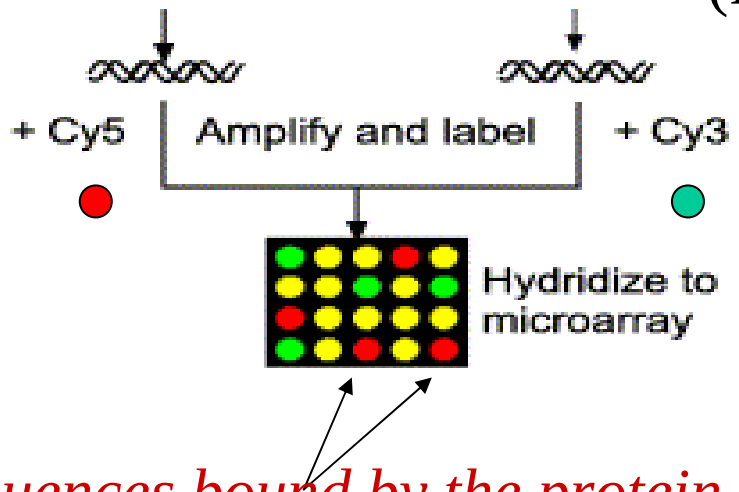
The results of ChIP on chip experiments

Enriched DNA

(from IP)

Input DNA

(from WCE)



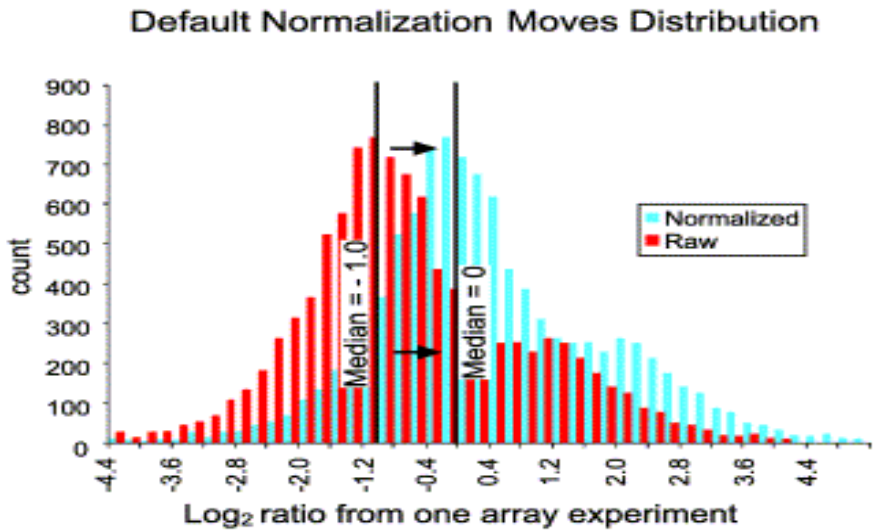
All the sequences bound by the protein of interest

- Global map of the protein fixation
- No need to know some of the targets before the experiment
- No need to determine the non enriched sequences before the experiment

Analysis of the results

For one experiment :

1. For each spot (each sequence) the enrichment ratio is determined : IP/WCE
2. Then, the data are normalized : so that the median log2 ratio =0

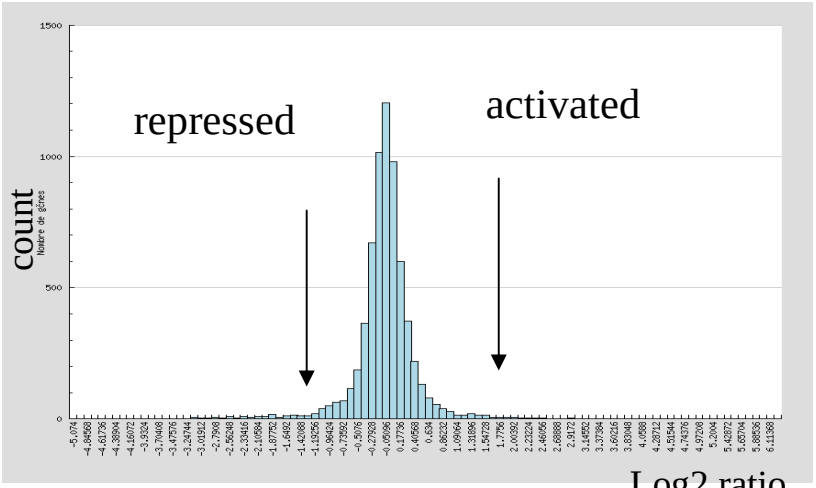


But one experiment is not enough…

Representation : Chip vs ChIP on chip

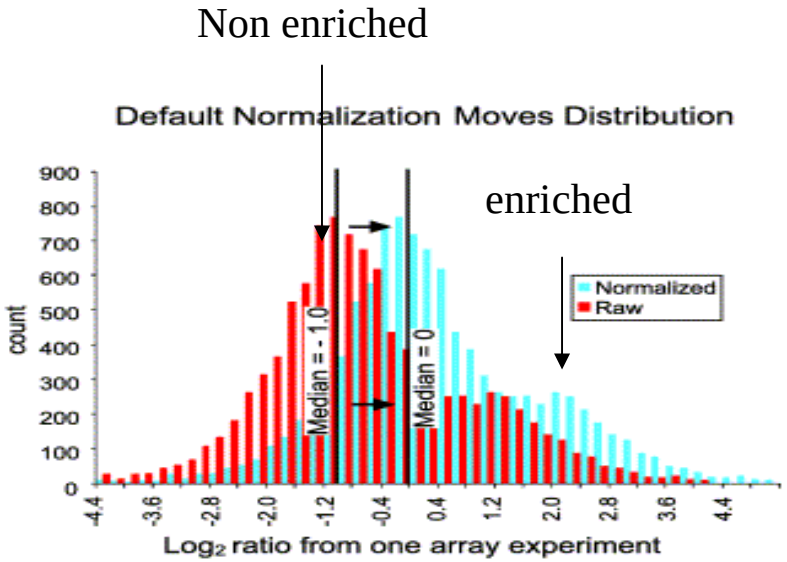
Transcriptome microarrays

- Symetric representation



ChIP on chip

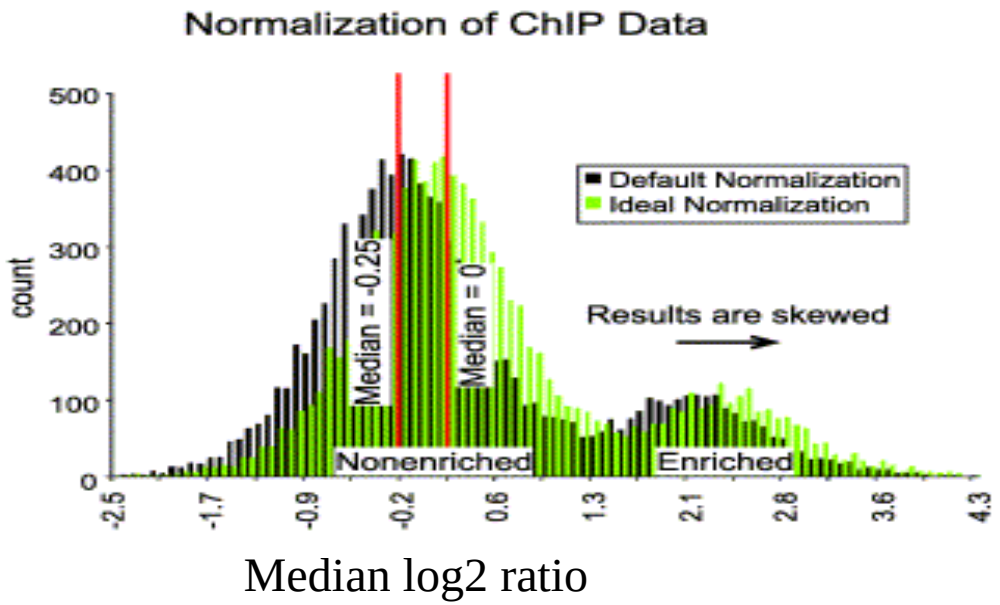
- There is only enrichment of DNA, there is no depletion!



Analysis of the results

For 4 experiments :

- At least 4 biological independent experiments are required to obtain significant results.
- For each sequence, the median of the normalized ratio is calculated



BUT, the amplitude of the ratios can change dramatically from one experiment to another for the same spot ➡ **RANK**

Analysis of the results

The median rank method

1. What is a rank?

The rank of a sequence is the position of this sequence in a list sorted by ratio in descending order.

Array Element	Replicate 1		Replicate 2		Replicate 3		Median % Rank
	Log Ratio	% Rank	Log Ratio	% Rank	Log Ratio	% Rank	
ITA(UGC)G	-1.39	0.00	-0.51	0.15	0.47	0.84	0.15
YGR151C	-0.67	0.09	1.20	0.99	-0.50	0.15	0.15
YGR152C	0.23	0.67	0.19	0.57	0.58	0.89	0.67
IYGR152C	-0.59	0.13	-0.61	0.11	0.58	0.90	0.13
YGR153W	-0.48	0.16	0.10	0.51	0.40	0.77	0.51
YGR154C	-0.10	0.40	0.06	0.49	-0.36	0.23	0.40
IYGR154C	0.74	0.91	-0.39	0.20	0.55	0.88	0.88
YGR155W	-0.78	0.08	0.04	0.46	-1.47	0.00	0.08
IYGR155W	-0.34	0.24	-0.63	0.10	0.17	0.63	0.24

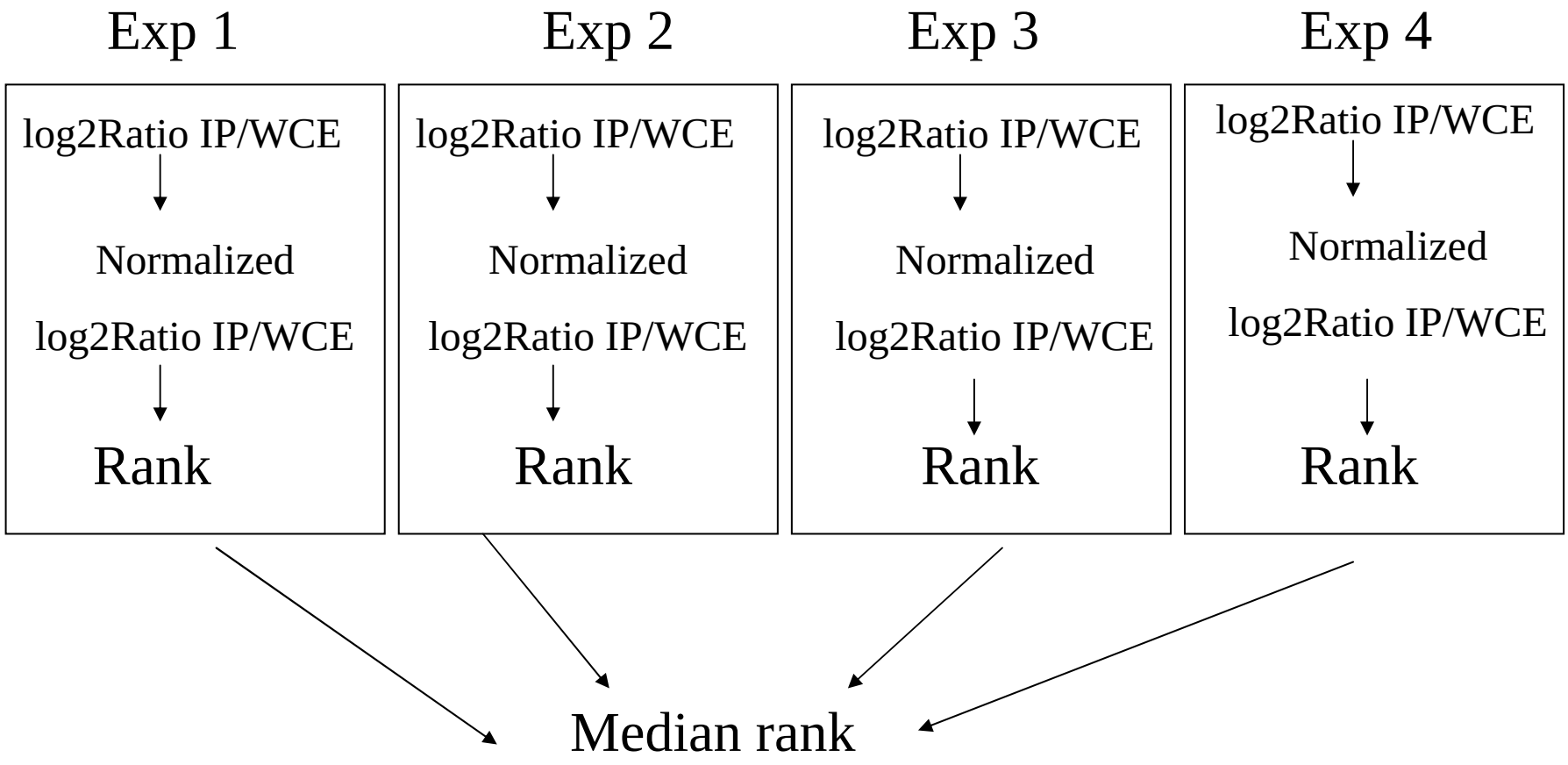
2. Why the ranks rather than ratios?

- To avoid normalization pb

- It’s easy

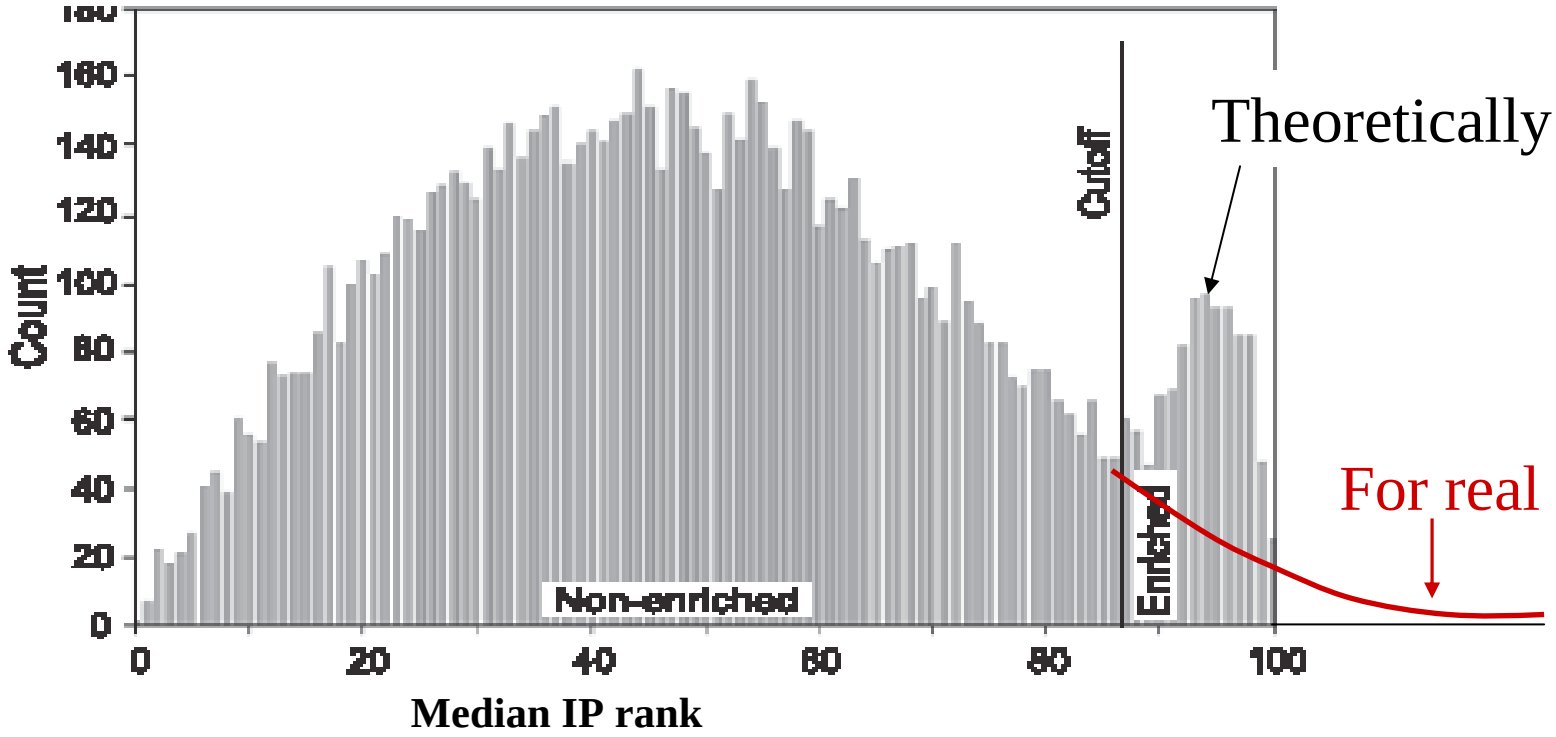
Analysis of the results

The Median rank method



Analysis of the results

Median percentile rank method

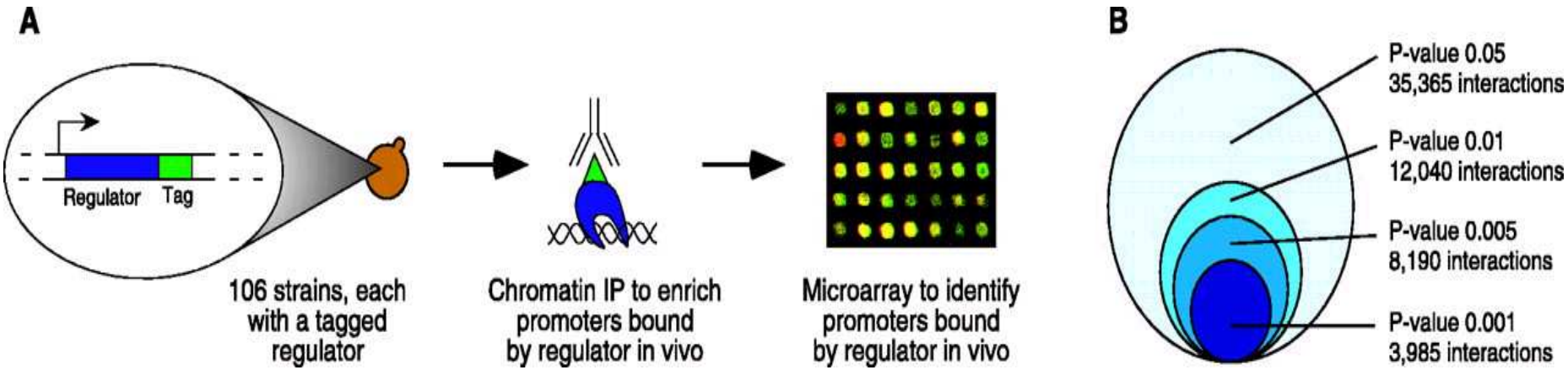


- That should give a bimodal distribution that makes the cut-off easier
- Efficient if more than 4% of the total elements are enriched
- Loss of amplitude information

Transcriptional regulatory networks in Saccharomyces cerevisiae.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA.

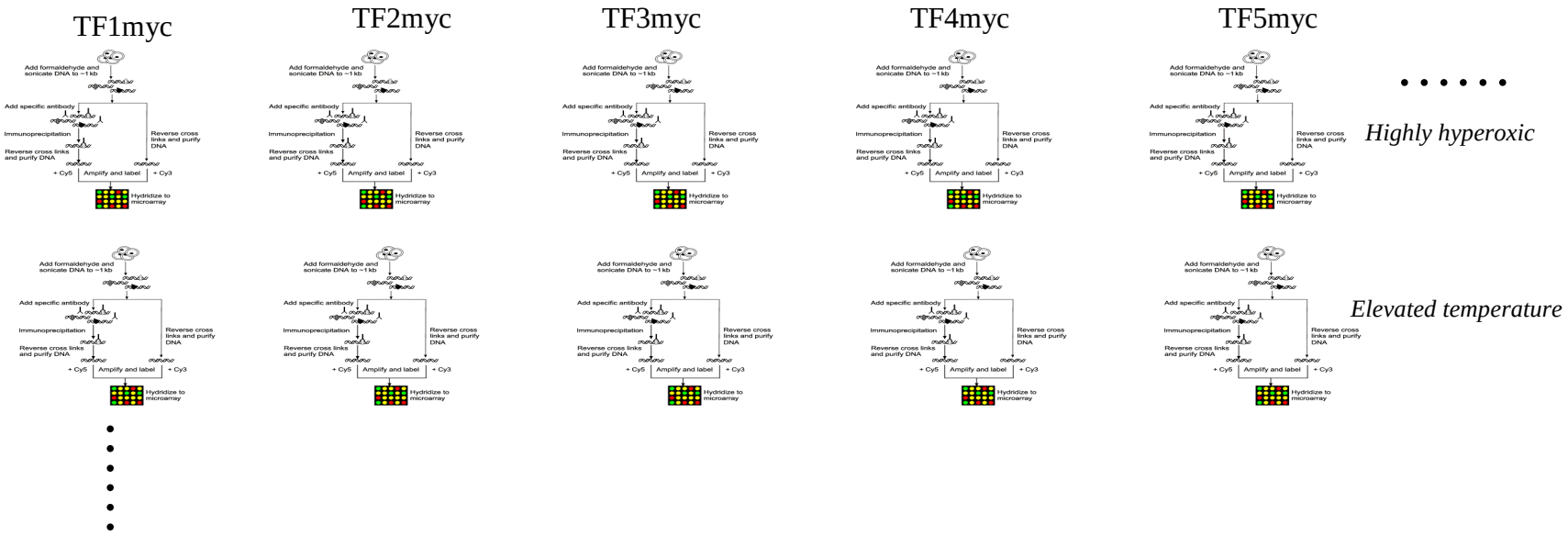
Science 2002 Oct 25;298(5594):763-4



1.Profiling the dynamics of TF binding

Genome-wide analysis of 84 TF in yeast…

…in many different growing conditions



. . . . . .

Highly hyperoxic

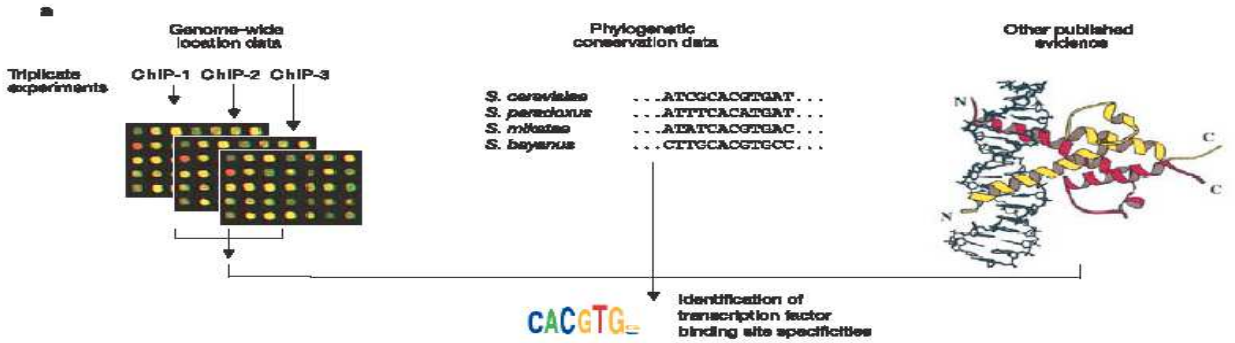
Elevated temperature

Harbison et al 2004

1.Profiling the dynamics of TF binding

Identification of cis-regulatory elements

By merging informations from ChIP on chip, phylogeny and prior knowledge

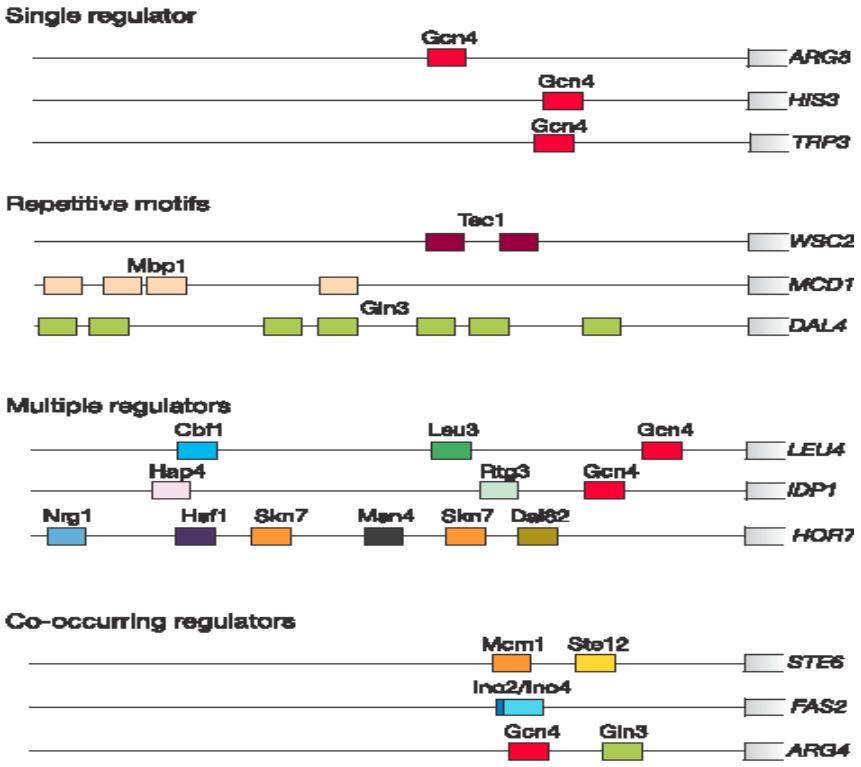


Harbisson et al 2004

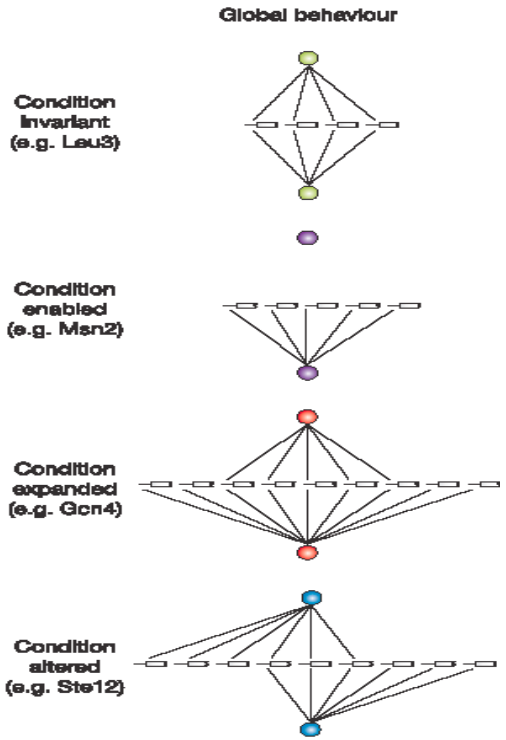
New regulatory sequences

1.Profiling the dynamics of TF binding

Transcriptional regulatory code of an eukaryotic genome



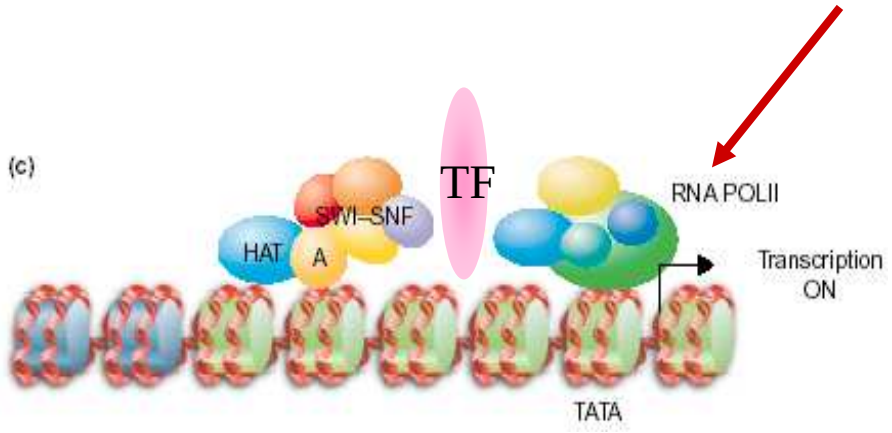
Static



Dynamics

2. Study of the general transcription machinery

Recruitment of general transcription machinery is often the end point in a cascade of regulatory events



ChIP on chip to have a genome-wide view of RNA polymerase (II and III) repartition along the chromosome in different conditions

2. Study of the general transcription machinery

Genome-wide analysis of RNA Pol II location during stationnary phase (SP) in yeast

- To survive to harsh conditions, yeast adopt a quiescent state (=GO for mammalian).
- Transition from and to this state were not well known, at the gene regulation level
- General shutdown of Pol II transcription has been reported.

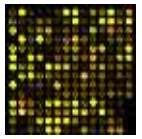
At what point does this shutdown occur?

Are there genes that escape repression?

Radonjic et al, 2005

2. Study of the general transcription machinery

FIRST : Global changing in mRNA levels during quiescence exit and entry studied with microarrays



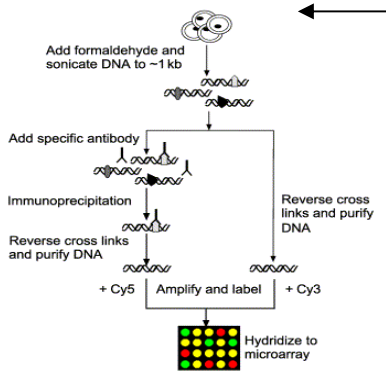
→ Extremely rapid response upon exit from SP (2500 genes upregulated within 6 min)

Hypothesis : this extremely rapid response can be explained by the fact that the transcription machinery is maintained on the promoter.

2. Study of the general transcription machinery

ChIP on chip experiments

SP



Tap Tag version of Rpb3, a PolII subunit

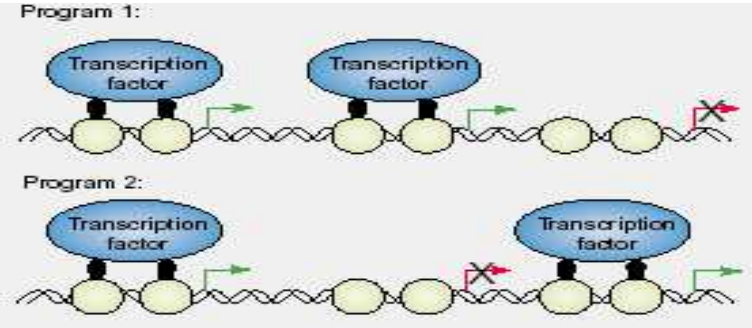
- Pol II is located upstream of rapid exit genes during SP
- Different from the current model : Pol II recruitment is the rate limiting step controled through chromatin remodeling

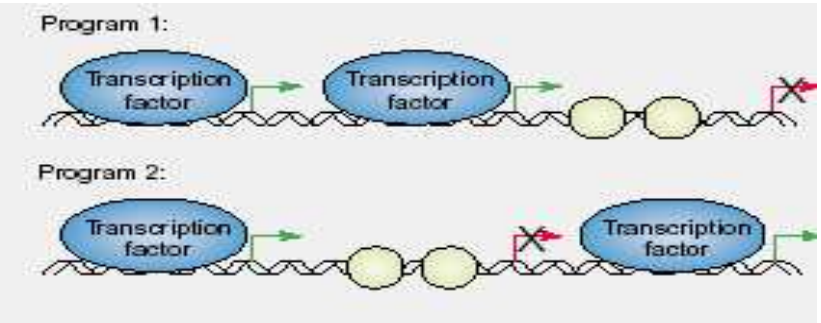
  →How this selective PolII binding occurs?

3. ChIP on chip analysis and chromatin dynamics

•Transcription occurs only if DNA is **accessible** !

The chromatin components are important regulators of transcription





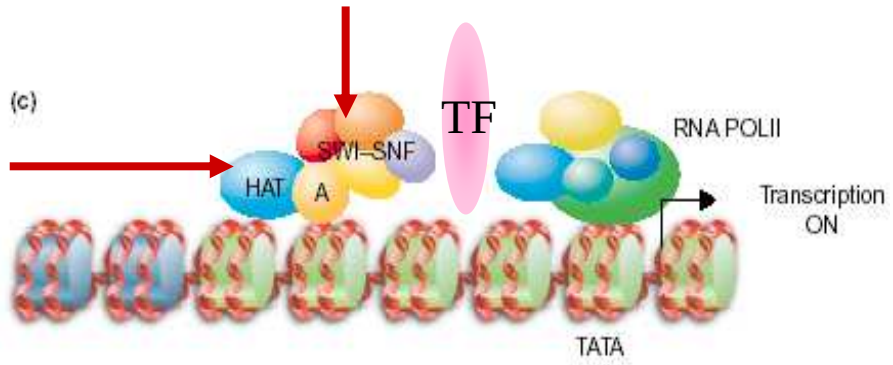
Changes in modification or occupancy → affect TF binding

3. ChIP on chip analysis and chromatin dynamics

- To increase accessibility :

  1. Nucleosome disruption by **remodeling complexes** (SWI-SNF)

  2. Chemical modifications of nucleosome by **histone modifying complexes** (HATs, HDACs)



Genome-wide correlation between chromatine modification and gene activation by using ChIP on chip : detection of **enzymes** and **histones modification**

3. ChIP on chip analysis and chromatin dynamics

Genome-wide studies of histone modifications

- Systematic ChIP on chip to map 11 different histone acetylation marks in yeast
- Compared with - each other
  - gene expression

Purpose : correlation between patterns of acetylation and gene activity?

•Conclusions :

Genes that share the same pattern of acetylation are biologicaly related

•For the future : basis for deciphering complex patterns of histone modification and their downstream effects on gene expression

Kurdistani et al, 2004

3. ChIP on chip analysis and chromatin dynamics

Analysis of histone-modifying proteins location

- Binding sites of HAT and HDAC

⟹ both global and gene specific control

- recruitment of HAT on active promoters

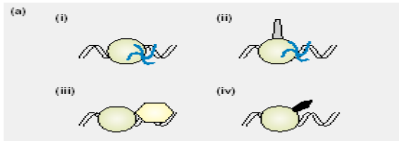
- recruitment of HDAC have a preference for distinct gene classes

Roberts et al, 2004

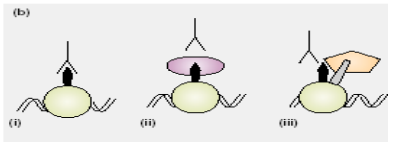
3. ChIP on chip analysis and chromatin dynamics

Challenges of applying ChIP on chip to chromatin

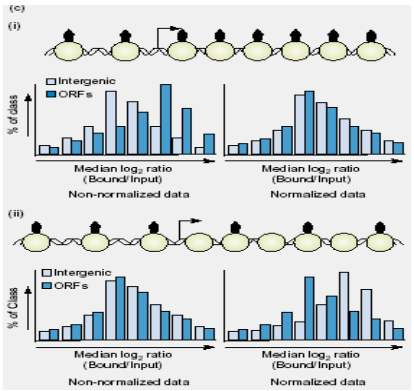
- Formaldehyde fixation



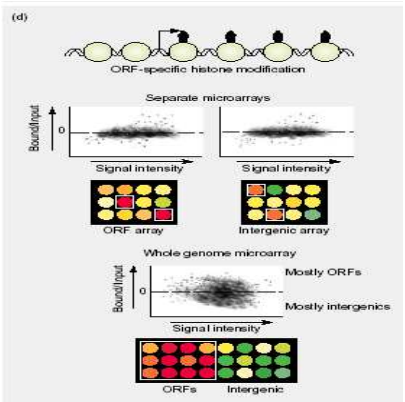
- Epitop accessibility



- Differential underlying nucleosome occupancy



- Choice of microarray design



3. Chip chip analysis and chromatin dynamics

ChIP on chip is not efficient to detect direct modifications of DNA

- Two microarray-based methods to detect methylations

➢ Methylation-sensitive restriction endonucleases
high MW fragments are not methylated
microarray

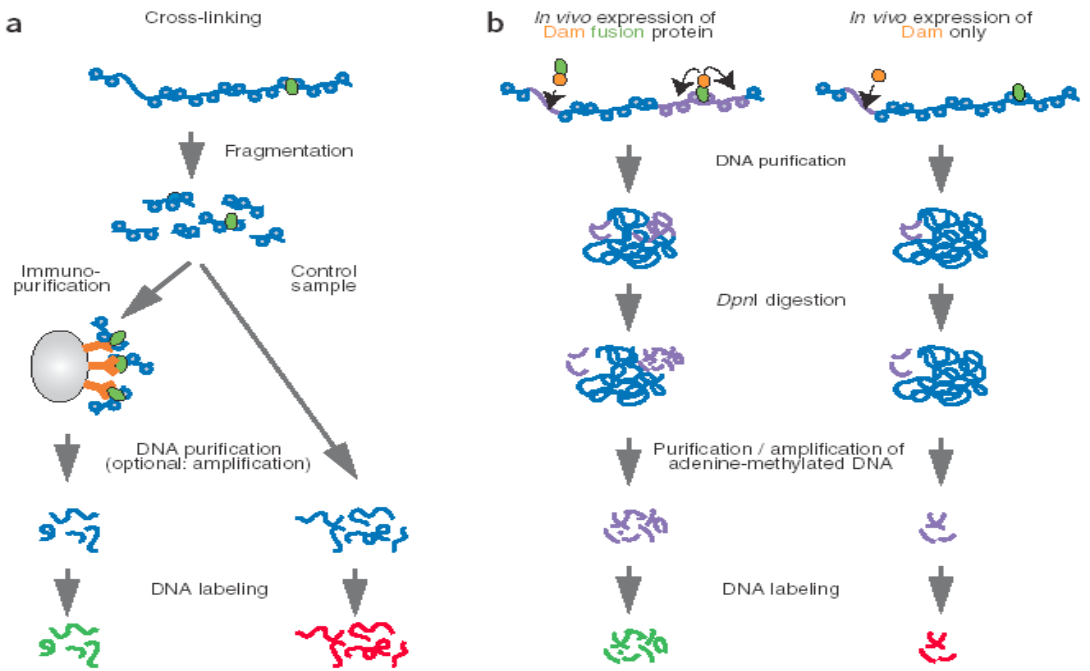
➢ Treatment C→U but Cmeth→Cmeth
after PCR : C→T
use of special design oligonucleotide array to quantify the C→T changes

4. Alternatives to ChIP on chip

DAMID

The protein of interest is fused to DNA methyl adenine transferase (DAM)

ChIP on chip DAM ID



Van steensel et al, 2004

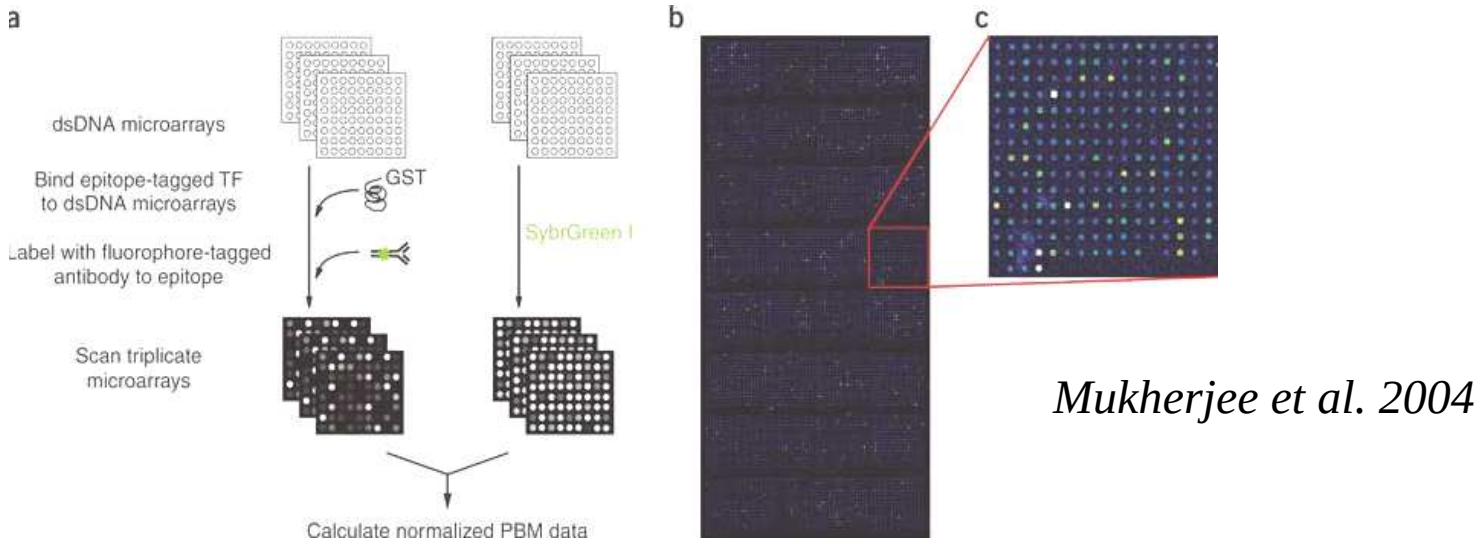
Advantage :No specific antibody required, no crosslink

Limits : no detection of histone modification

4. Alternatives to ChIP on chip

Protein Binding Microarray

- ChIP on chip : in vivo, but, it doesn't give all the potential targets…



Mukherjee et al. 2004

Advantages :

- No need for specific antibody

- No need to find the condition for

which the protein of interest is nuclear

Limits :

-in vitro

-Few overlap between ChIP on chip and PBM results

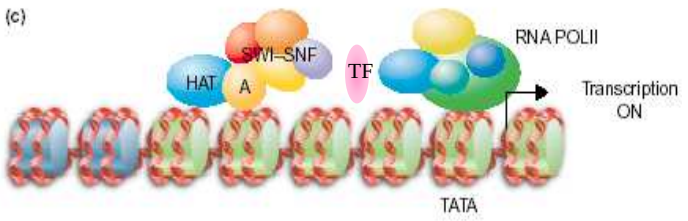
Conclusion

- ChIP on chip is a very powerful tool to explore gene regulation

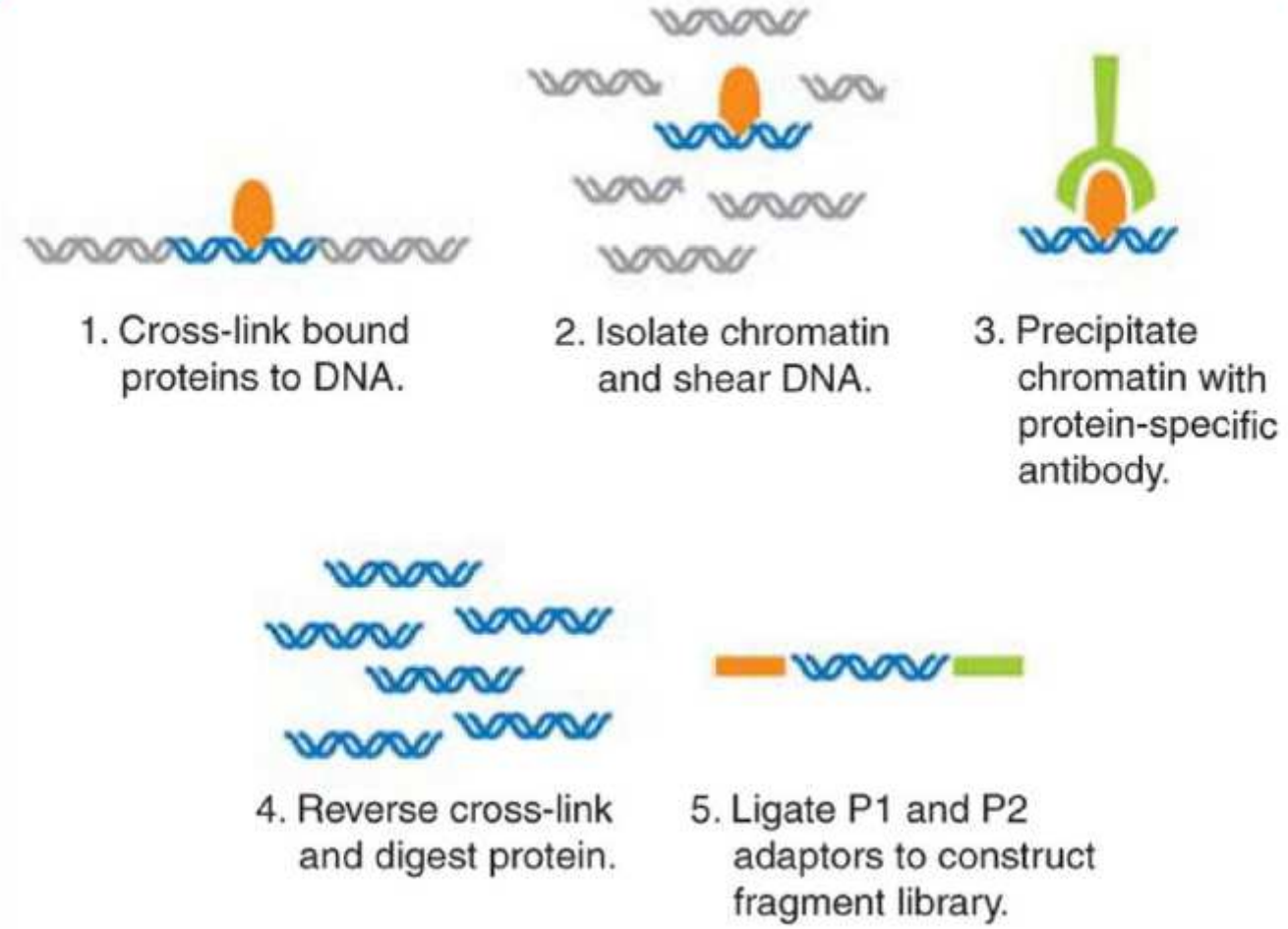
- ChIP on chip can also be useful to study the dynamics of replication and recombination

- ChIP on chip is not restricted to *S. cerevisiae*

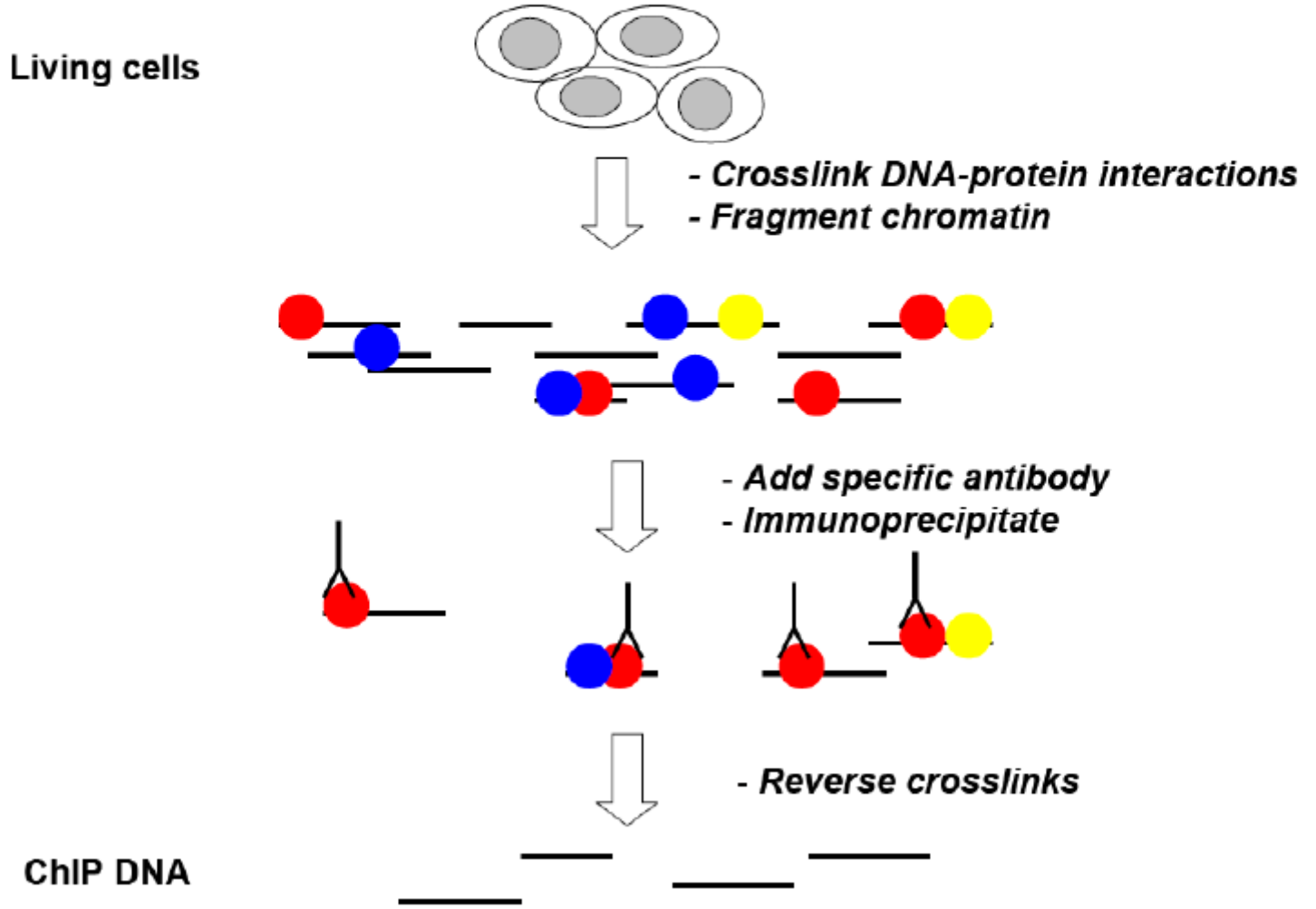
- Real challenge in the future :
  - Integrating all these data using bioinformatics
  - Studying the dynamics of the response to one condition at all the levels



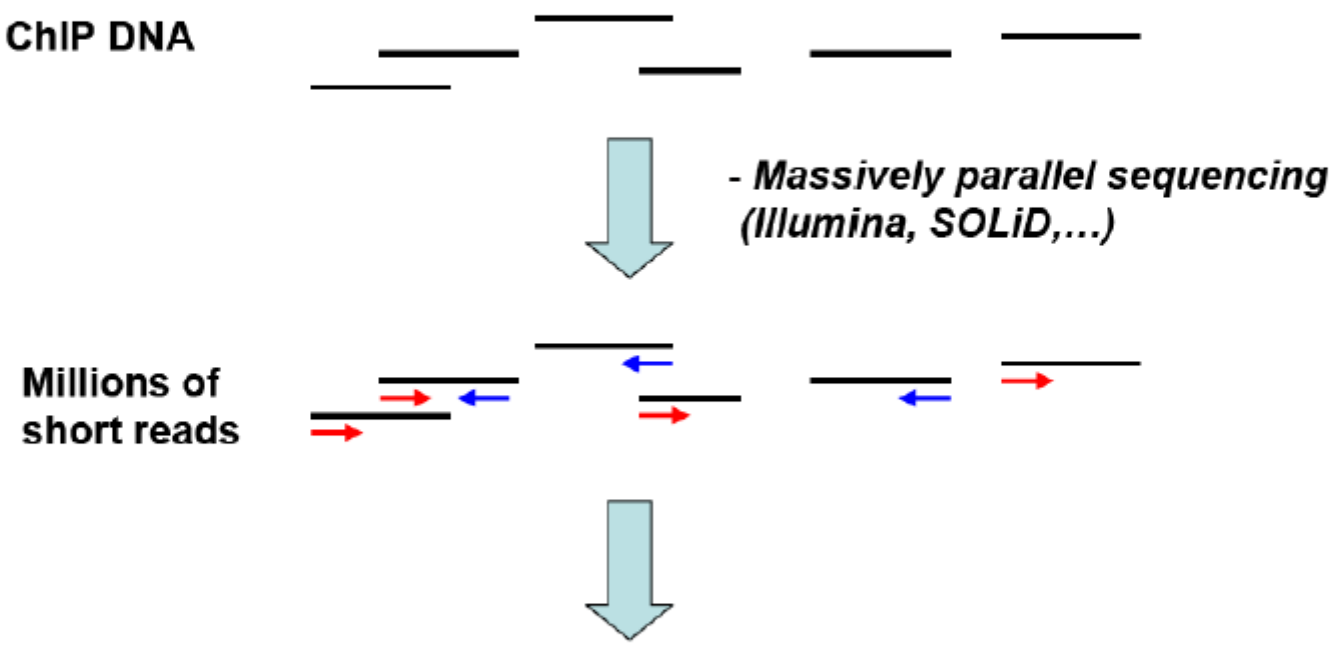
ChIP and SOLiD fragment library construction



ChIP – Chromatin Immunoprecipitation



Sequencing ChIP DNA

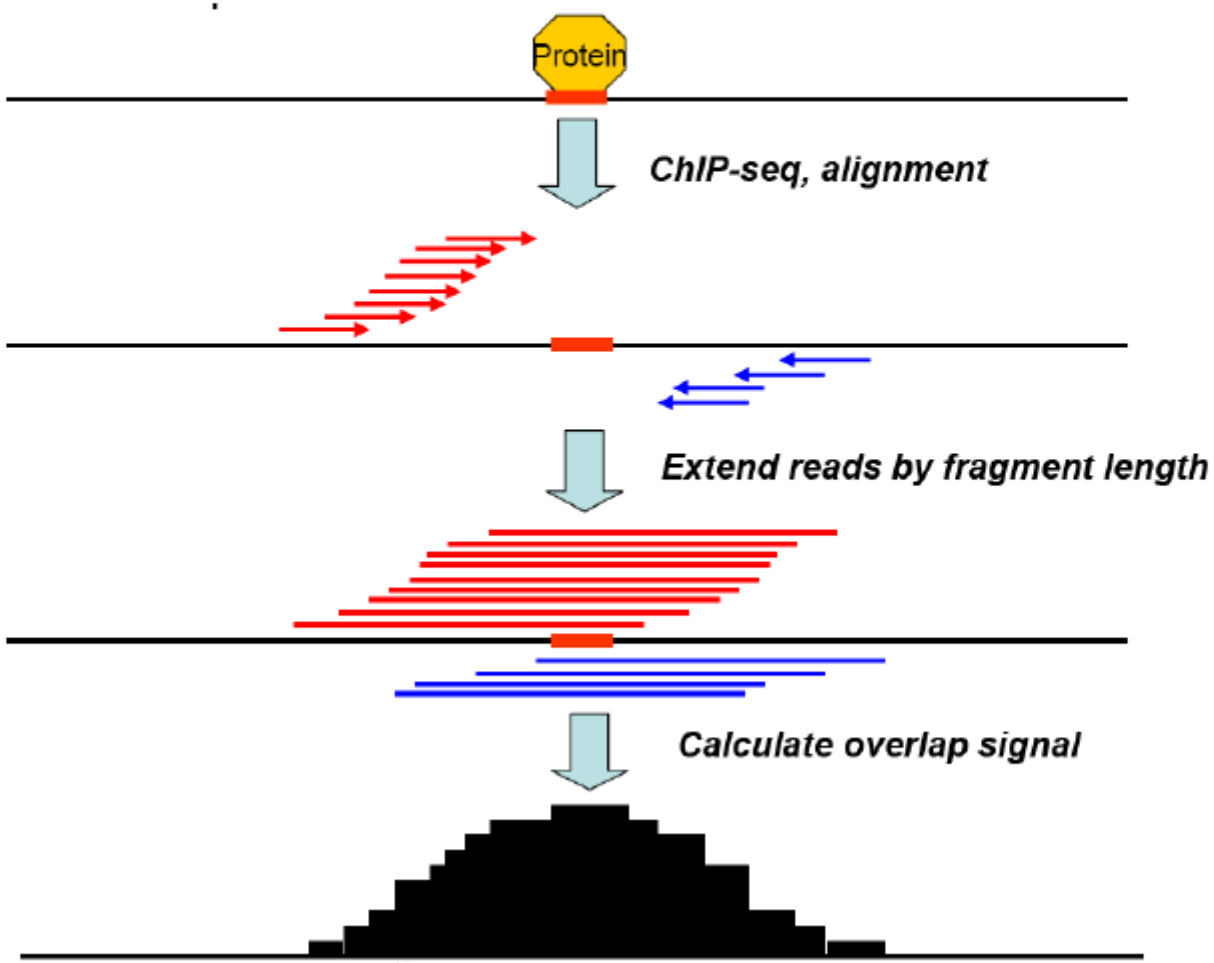


Alignment and further data analysis…

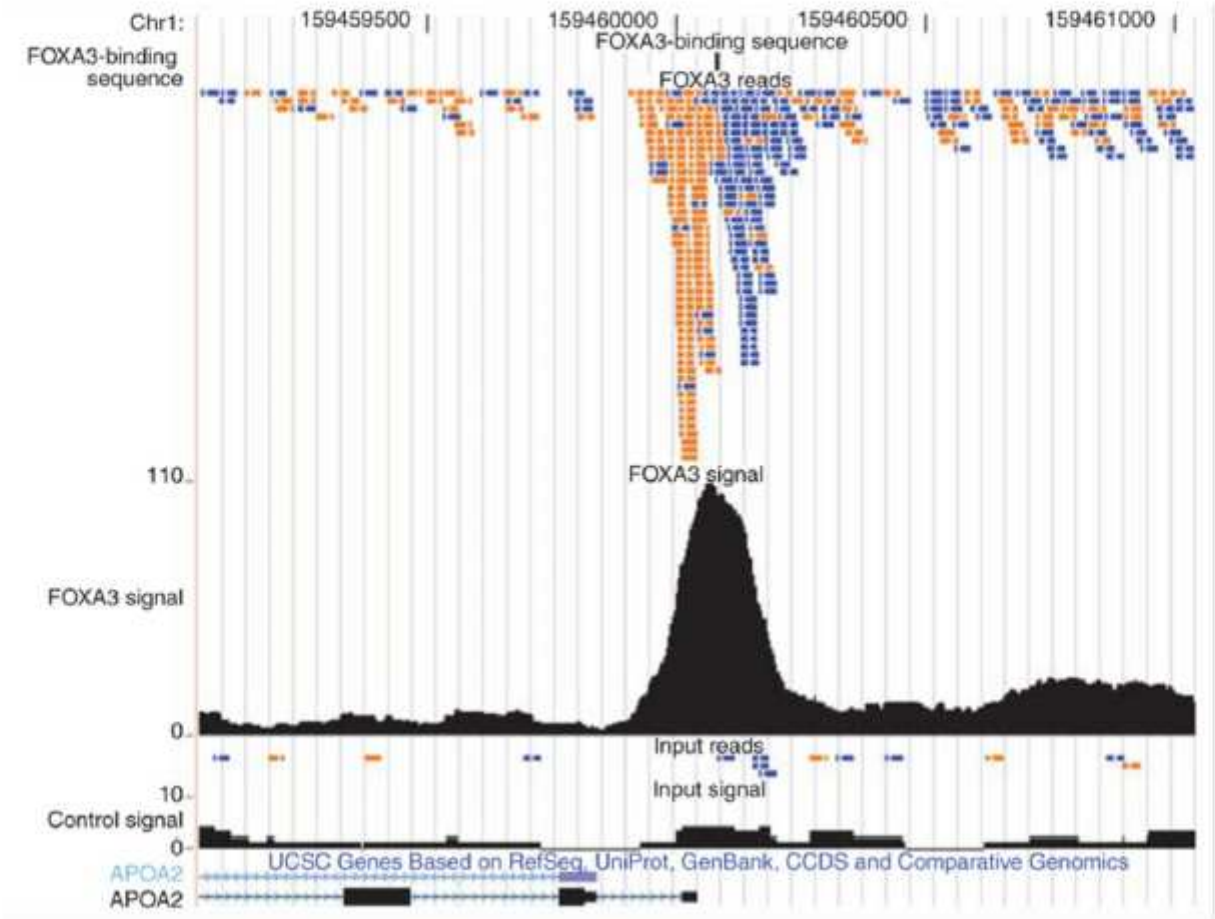
ChIP – downstream analysis

- **Where are the protein bound?**
  - Preprocessing
  - Peak detection
- **Can we indentify the exact binding sequences?**
  - Motif search
- Can we find allele-specific binding
- **Which genes are regulated by the proteins**
  - Genomic annotation
  - Signal footprints
  - ….

From reads to peaks



Graphic representation of alignment of ChIP-seq reads (UCSC genomic browser)



GENOMNIA

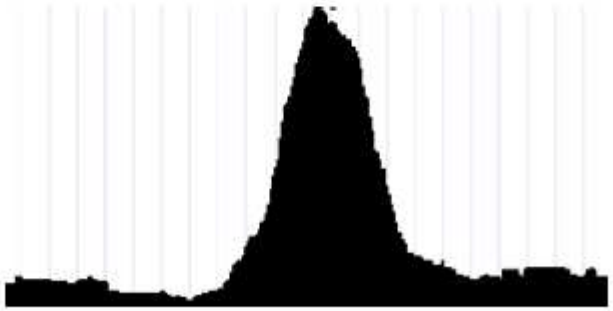
Peak detection

- Aim: identify bound regions from ChIP-seq data
- Some available methods:
  - QUEST (Volouev et al 2008)
  - SISSRS (Jothi et al 2008)
  - MACS (Zhang et al 2008)
- Common features of peak finding algorithms
  - Estimate average fragment length from data
  - Create overlap signals
  - Detect significant peaks
  - Take into account strand information of reads

TFBS detection

- Motif search algorithms can be used for TFBS detection
  - Several methods exists
  - Not all of them are suited for ChIP-seq data (1000s of regions)
- BCRANK (predicting Binding site Consensus from RANKed sequences)
  - Input: A list of DNA regions, ranked by ChIP enrichment signal
  - Idea: The true binding site should be…

frequently occurring in high peaks



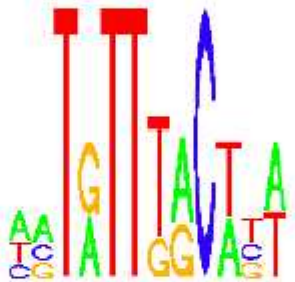
less frequent in low peaks

BCRANK is available from Bioconductor (www.bioconductor.org)

BCRANK results

Results for FOXA1 and FOXA3 (ABI/SOLiD sequencing)

FOXA1



Position	Start	End	Strand	Sequence
chr5 147194810 147195810	436	447	+	TGTGTTTACATT
chr5 147194810 147195810	904	915	+	CCTATTTGCTGT
chr5 147194810 147195810	214	225	-	TAAGTAAATATT
chr2 18844830 18845830	471	482	+	AATATTGACTTA
chr2 18844830 18845830	975	986	-	AGAGCAAATATT
chr1 8724366 8725366	529	540	-	AATGCAAACAGT
chr14 60004426 60005426	518	529	+	ACTGTTTACTGA
chr14 60004426 60005426	844	855	+	CCTATTGGCTTA
chr2 181549316 181550316	4	15	+	CATATTTGCACT
chr2 181549316 181550316	757	768	+	CCTGTTTACACT
chr2 181549316 181550316	563	574	-	AATGTAAATATA
chr4 155750862 155750862	416	427	+	TCTATTGACATT
chr12 26813071 26814071	444	455	+	AATGTTTGCTTT
chr12 26813071 26814071	492	503	-	ACTGTAAACAGG
chr2 65453673 65454673	485	496	-	AATGTCAATATT
chr2 65453673 65454673	513	524	-	TATGCAAACACT
chr14 38562819 38563819	466	477	-	TAAGTAAATATA
chr14 38562819 38563819	491	502	-	AATGTAAATATG
chr14 37085504 37086504	799	810	-	AAAGTAAATATT
chr10 33484610 33486610	628	639	-	AAAGTAAATATG
...	...	...	.	...
...	...	...	.	...
...	...	...	.	...

FOXA3

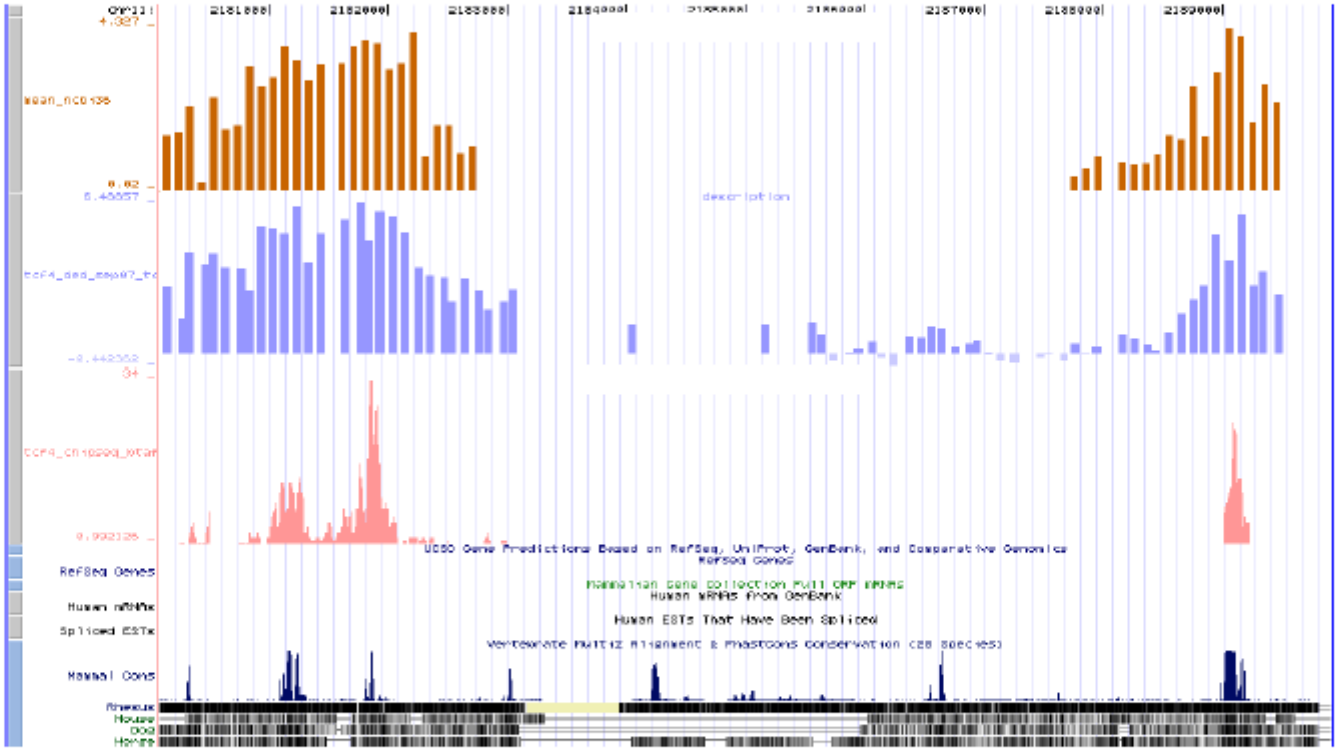
Position	Start	End	Strand	Sequence
chr5 147194810 147195810	436	447	+	TGTGTTTACATT
chr5 147194810 147195810	904	915	+	CCTATTTGCTGT
chr5 147194810 147195810	214	225	-	TAAGTAAATATT
chr2 18844830 18845830	471	482	+	AATATTGACTTA
chr2 18844830 18845830	975	986	-	AGAGCAAATATT
chr1 8724366 8725366	529	540	-	AATGCAAACAGT
chr14 60004426 60005426	518	529	+	ACTGTTTACTGA
chr14 60004426 60005426	844	855	+	CCTATTGGCTTA
chr2 181549316 181550316	4	15	+	CATATTTGCACT
chr2 181549316 181550316	757	768	+	CCTGTTTACACT
chr2 181549316 181550316	563	574	-	AATGTAAATATA
chr4 155750862 155750862	416	427	+	TCTATTGACATT
chr12 26813071 26814071	444	455	+	AATGTTTGCTTT
chr12 26813071 26814071	492	503	-	ACTGTAAACAGG
chr2 65453673 65454673	485	496	-	AATGTCAATATT
chr2 65453673 65454673	513	524	-	TATGCAAACACT
chr14 38562819 38563819	466	477	-	TAAGTAAATATA
chr14 38562819 38563819	491	502	-	AATGTAAATATG
chr14 37085504 37086504	799	810	-	AAAGTAAATATT
chr10 33484610 33486610	628	639	-	AAAGTAAATATG
...	...	...	.	...
...	...	...	.	...
...	...	...	.	...

ChIP-Seq vs ChIP-chip

Nimblegen
Whole genome

Agilent
Dedicated regions

SOLiD
Whole genome
4 M raw reads

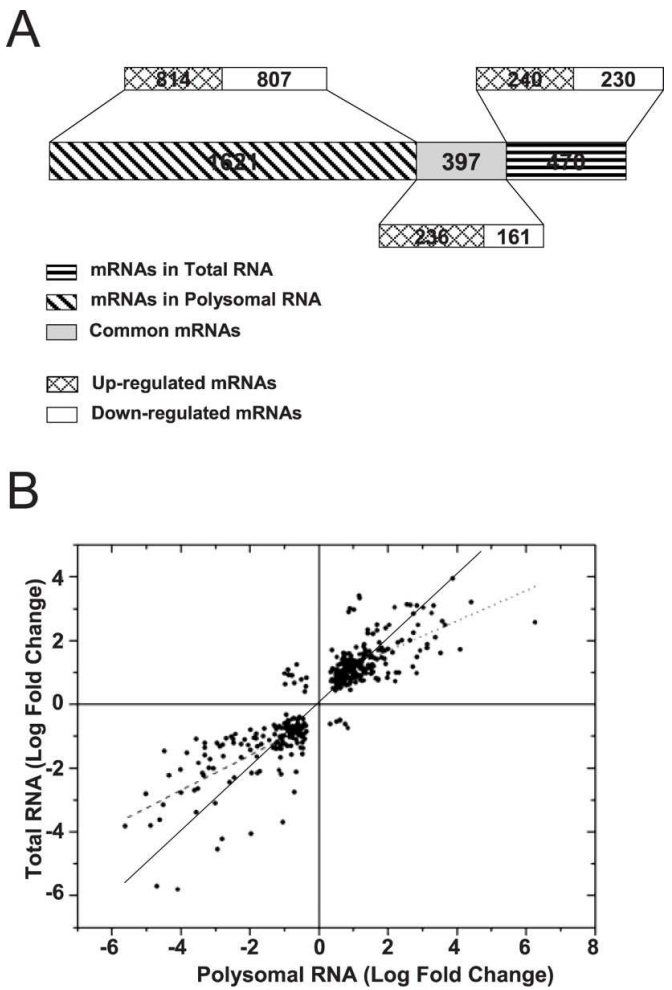


GENOMNIA

ChIP-seq on SOLiD vs ChIP-chip microarrays

Feature	SOLiD System (ChIP-Seq)	Microarray (ChIP-chip)
Resolution	>400 million sequence tags per run	2.1 million oligonucleotide features per array
Genome coverage	Unlimited: entire genome can be sequenced hypothesis free	Limited by probe design
Specificity	No cross-hybridization risks; identifies unique sequence tags	Cross-hybridization risks between closely related elements
Sample multiplexing	Yes	No

Comparative bidimensional transcriptome analysis of SW480 and SW620 cell lines.



Provenzani A et al. Carcinogenesis 2006;27:1323-1333

Carcinogenesis