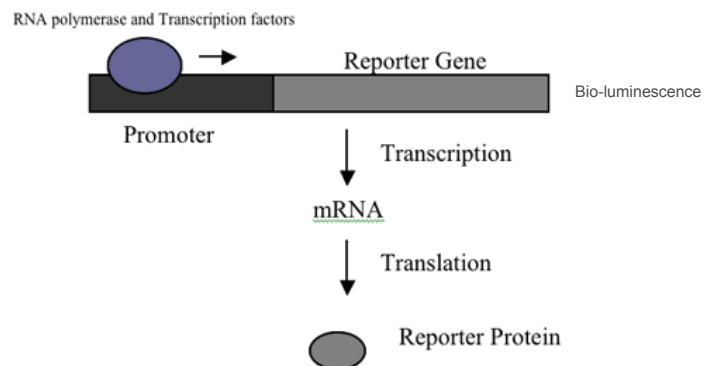




Plan

- Rappel : vision classique de la régulation
- Nouvelles possibilités à l'ère postgénomique
 - Omics : de nouvelles opportunités pour la biologie *in silico*
 - Epigénétique
 - Nouvelles approches expérimentales
- Nouvelle vision de la régulation
 - Cis-contexte
 - Trans-contexte
- Perspectives

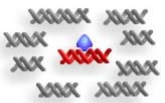
Méthodes d'étude ciblée : gène rapporteur



Méthodes d'étude ciblée : Immunoprécipitation de la chromatine (ChIP)



Fixation covalente *in vivo* des protéines sur leur site de fixation à l'ADN (formaldéhyde)



Fragmentation de la chromatine



Immunoprécipitation par anticorps spécifique de la protéine d'intérêt fixée à l'ADN



Réversion du pontage puis analyse de l'ADN
(ex: PCR avec amorces spécifiques de la région cible)

Nouvelles possibilités à l'ère postgénomique

- **Omics : de nouvelles opportunités pour la biologie *in silico***

Des milliers de génomes complets disponibles

=> possibilité de cartographier

=> possibilité de comparer les génomes

Approches in silico

- **Prédiction de sites de fixation dans le génome**

A partir d'un motif

A partir de séquences connues : méthodes d'apprentissage

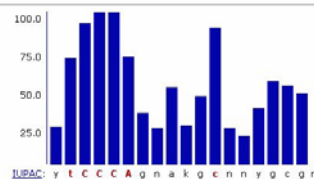
=> Construction d'un modèle décrivant le set d'apprentissage

=> Comparaison du modèle à de nouvelles séquences => scores

=> HMM (ex HMMsearch), profil = PSSM (ex: MatInspector, Match)

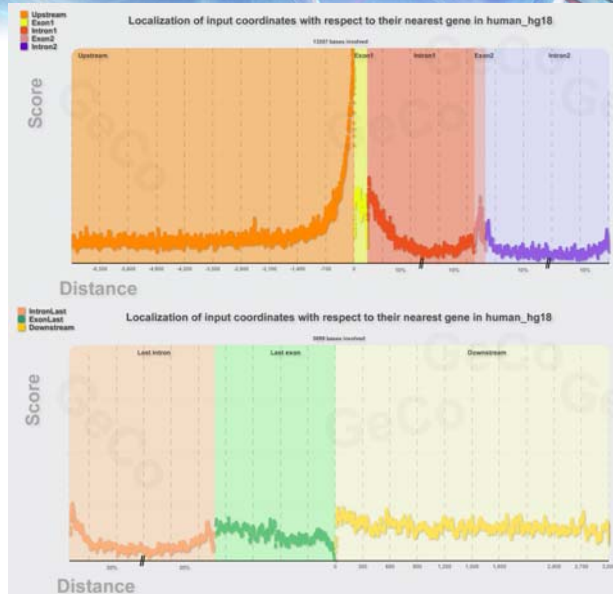
Ex : 347 sites de fixation de STAF

Pos.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	32	12	0	0	0	294	35	127	233	46	32	3	100	48	19	44	10	44
C	122	25	340	347	347	1	47	130	7	24	40	337	150	149	192	9	255	16
G	36	5	0	0	0	50	219	74	101	120	253	5	19	71	19	273	70	255
T	156	305	7	0	0	2	46	16	6	157	22	2	78	79	117	21	12	32
IUPAC	Y	T	C	C	C	A	G	N	A	K	G	C	N	N	Y	G	C	G
CI	25.5	70.2	93.9	100.0	100.0	71.0	34.1	25.0	51.0	26.0	45.7	90.0	24.5	19.3	37.1	55.6	52.3	47.2



Approches in silico

165 000 sites candidats !



Approches in silico

- Prédiction de sites de fixation dans le génome
- Filtre par la conservation
 - Phylogenetic footprinting/shadowing

Phylogenetic footprinting/shadowing

Recherche d'éléments communs à un grand groupe

► Phylogenetic footprinting

Ex pour les mammifères :

comparaison homme/souris (séparation : 75 millions d'années)

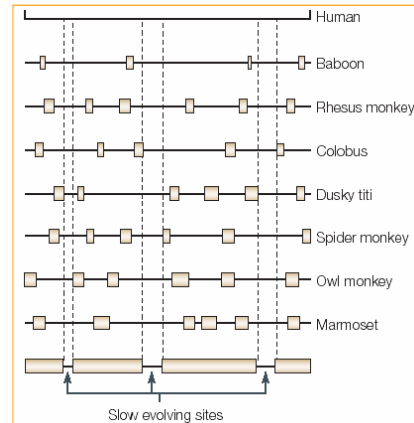
Recherche d'éléments plus récents

► Phylogenetic shadowing

Ex : trait apparu chez les primates

Pb: forte conservation

- ⇒ distinction impossible entre conservation fonctionnelle et conservation passive
- ⇒ augmenter la divergence en multipliant le nombre d'espèces



Approches in silico

- Prédiction de sites de fixation dans le génome
- Filtre par la conservation
 - Phylogenetic footprinting/shadowing
- Recherche d'un motif surreprésenté dans des séquences cibles

- Ex: logiciel MEME

```

ATGGTTAGATGGTTGTTCCCAAATTAGAA
ATGCCTCCCTAATTACCGATT
TTCCCAAATTAGAAATGGTTAGATGGTTG
GGGGCGCAACGCACCCATATCCCGGGTTTATCG
  
```

Approches in silico

- Prédiction de sites de fixation dans le génome
- Filtre par la conservation
- Recherche d'un motif surreprésenté dans des séquences cibles

Quelques ressources spécialisées :

- dbTSS
- Transfac
- Jaspar
- MatBase de Genomatix

Nouvelles possibilités à l'ère postgénomique

- Omics : de nouvelles opportunités pour la biologie in silico

- **Epigénétique**

étude de l'influence de l'environnement sur l'expression des gènes

Epigenetics may provide hope that we are more than just the sequence of our genes...

Epigenome: The chromatin states that are found along the genome, defined for a given time point and cell type. Thus, for a given genome there may be hundreds or thousands of epigenomes, depending on the stability of the chromatin states.

- **Avancées des techniques expérimentales**

Séquençage nouvelle génération => Approches globales, sans a priori
comprendre le fonctionnement du génome et les réseaux de régulation
qui le sous-tendent

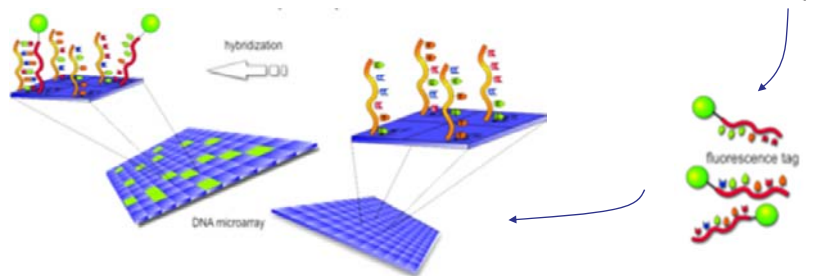
Méthodes globales : ChIP on chip

But : Identification des sites de fixation d'une protéine sur l'ADN

Même technique que pour ChIP

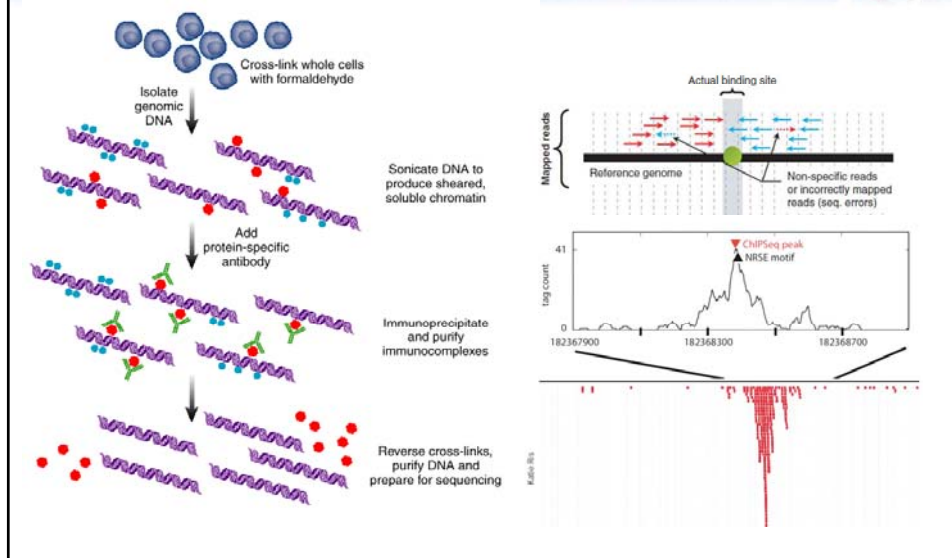
Fragments d'ADN isolés :

- ligation à des adaptateurs
- amplification et marquage de tous les fragments
- hybridation avec une puce ADN



Méthodes globales : Chip-seq

But : Identification des sites de fixation d'une protéine sur l'ADN

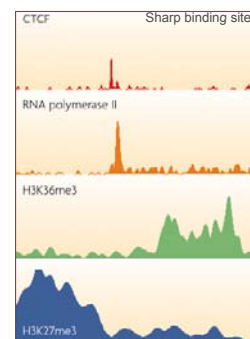
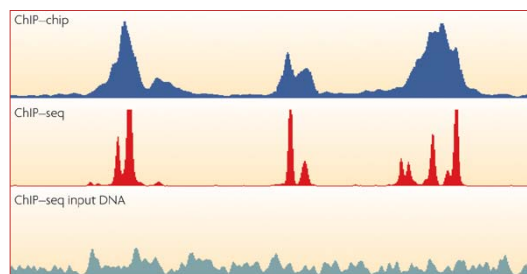


ChIP-chip et ChIP-seq

	ChIP-chip	ChIP-seq
Maximum resolution	Array-specific, generally 30–100bp	Single nucleotide
Coverage	Limited by sequences on the array; repetitive regions are usually masked out	Limited only by alignability of reads to the genome; increases with read length; many repetitive regions can be covered
Cost	US\$400–800 per array (1–6 million probes); multiple arrays may be needed for large genomes	Currently US\$1,000–2,000 per lane (using the Illumina Genome Analyzer); 6–15 million reads before alignment
Source of platform noise	Cross-hybridization between probes and nonspecific targets	Some GC bias can be present
Experimental design	Single- or double-channel, depending on the platform	Single channel
Cost-effective cases	Profiling of selected regions; when a large fraction of the genome is enriched for the modification or protein of interest (broad binding)	Large genomes; when a small fraction of the genome is enriched for the modification or protein of interest (sharp binding)
Required amount of ChIP DNA	High (a few micrograms)	Low (10–50 ng)
Dynamic range	Lower detection limit; saturation at high signal	Not limited
Amplification	More required	Less required; single-molecule sequencing without amplification is available

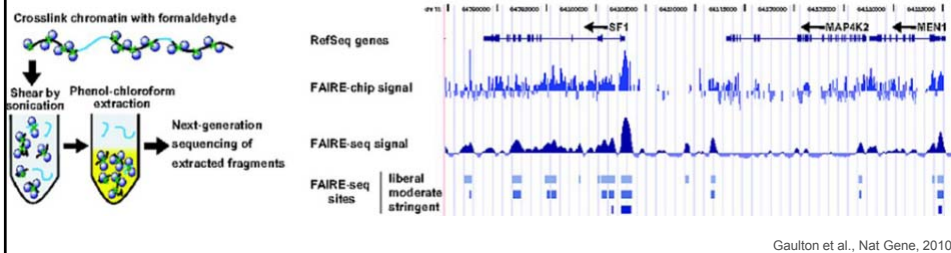
Park, Nature Rev Genetics, 2009

ChIP-chip et ChIP-seq



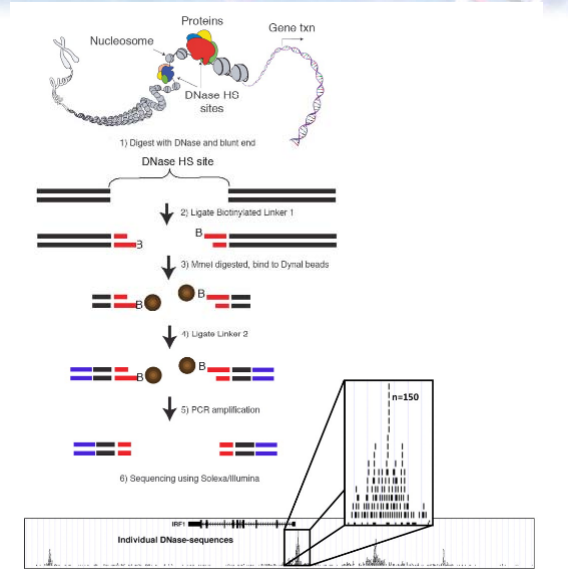
Faire-seq (Formaldehyde-Assisted Isolation of Regulatory Elements)

But : Identification des sites accessibles de l'ADN (euchromatine ouverte)



DNase-seq

But : Identification des sites accessibles de l'ADN (euchromatine ouverte)

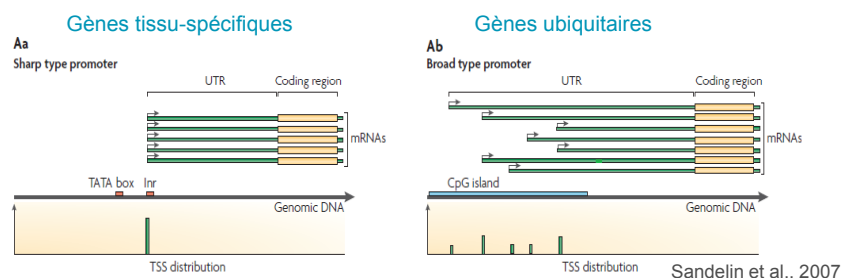


Plan

- Rappel : vision classique de la régulation
- Nouvelles possibilités à l'ère postgénomique
 - Omics : de nouvelles opportunités pour la biologie *in silico*
 - Epigénétique
 - Nouvelles approches expérimentales
- Nouvelle vision de la régulation
 - Cis-contexte
 - Trans-contexte
- Perspectives

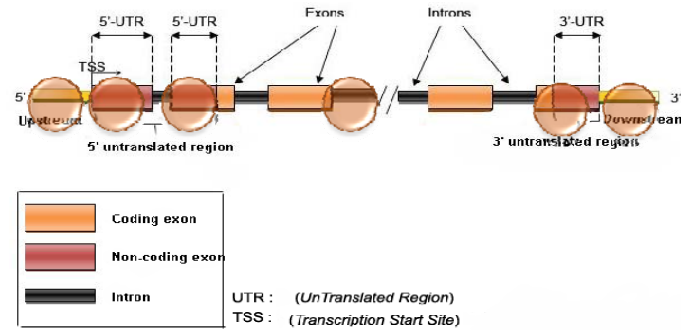
Cis-contexte : la séquence

- Promoteur proximal
 - ~76% of human core promoters lack TATA-like elements
 - ~54% lack the consensus INR
 - ~46% of human promoters lack both TATA-like and INR elements
 - Îlots CpG (régions génomiques présentant un enrichissement en dinucléotides CpG)
 - abondants au sein des promoteurs « larges » (gènes ubiquitaires)



Cis-contexte : la séquence

- Promoteur proximal
- Séquences régulatrices



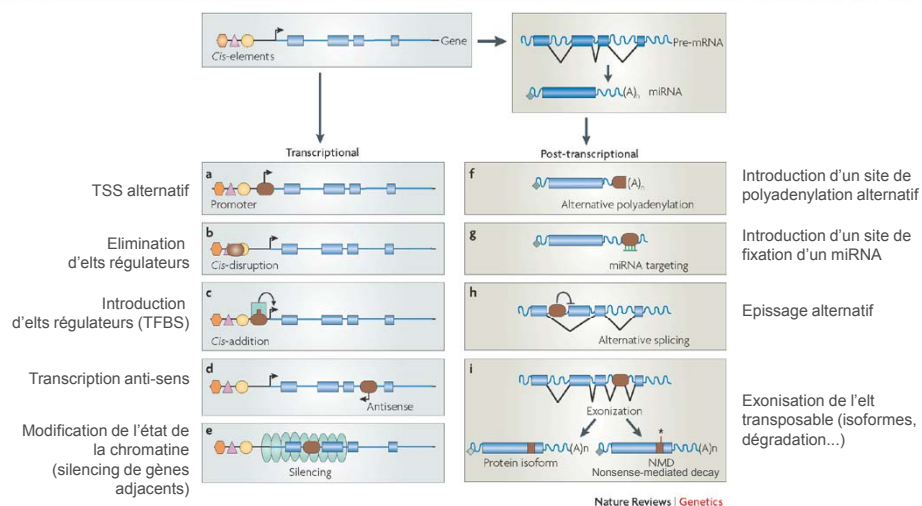
Cis-contexte : la séquence

- Promoteur proximal
- Séquences régulatrices
- Polymorphisme
- Elts répétés

Eléments répétés et régulation

- **Eléments répétés dispersés = 45% du génome humain : "junk DNA" ?**
- **Britten et Davidson (1971):**
Des systèmes de régulation coordonnés dans les génomes animaux seraient encodés par des réseaux provenant de séquences répétées
- **Etudes de génomique comparative :**
 - Proportion importante d'éléments conservés (autres que les exons) spécifiques aux mammifères (par ex) proviennent d'éléments mobiles
=> phénomène d'**exaptation**
Introduction par les elt transposables de signaux "prêts à servir" (utilisés pour réguler leur propre expression) : promoteurs pour ARN pol II ou III, enhancers, signaux de polyadénylation, sites d'épissage, TFBS,...
 - Transposons : force majeure dans évolution de la régulation des gènes de mammifères

Eléments répétés et régulation



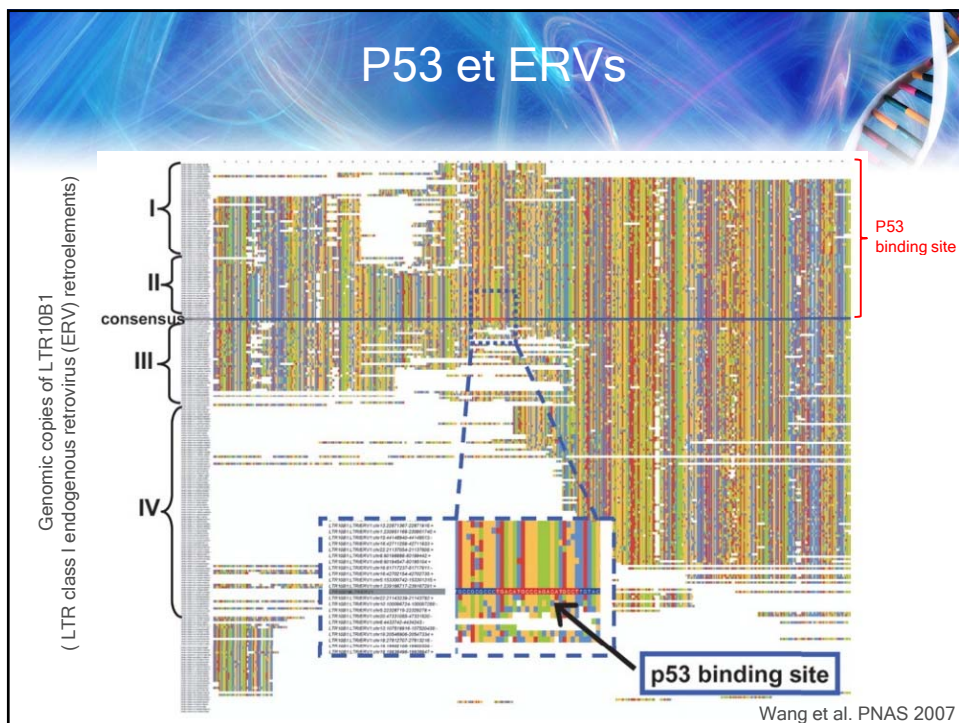
Feschotte, Nature. Rev Genet. 2008

Exemple : ERVs et p53

- Endogenous retroviruses (ERVs), retrotransposons à LTR : 8% du génome humain
- P53 (tumor suppressor protein p53) : facteur de transcription
=> réponse aux stress cellulaires en coordonnant l'expression de gènes impliqués dans l'arrêt du cycle cellulaire, l'apoptose...
- Enrichissement en sites de fixation p53 dans 2 sous-familles ERVs: LTR10 et MER61
- Estimation de l'âge de ces 2 familles
 - recherche dans autres espèces
 - Comparaison des séquences
 => Colonisation de la lignée anthropoïde par des ERVs possédant un site de fixation p53 (~40 millions d'années)

Wang et al. PNAS 2007

P53 et ERVs



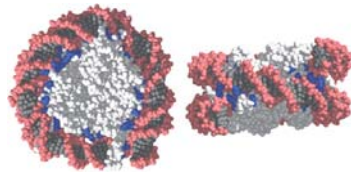
P53 et ERVs

- Annotation GO des nouveaux gènes cibles et recherche d'enrichissement fonctionnel :
 - Absence des fonctions classiquement régulés par p53 (ex: lésion de l'ADN)
 - Enrichissement en gènes impliqués dans l'adhésion cellulaire
 - Gènes impliqués dans l'adhésion cellulaire évoluent rapidement chez l'homme
- Introduction des ERVs
 - => nouveau sous-réseau de régulation primate-spécifique de p53

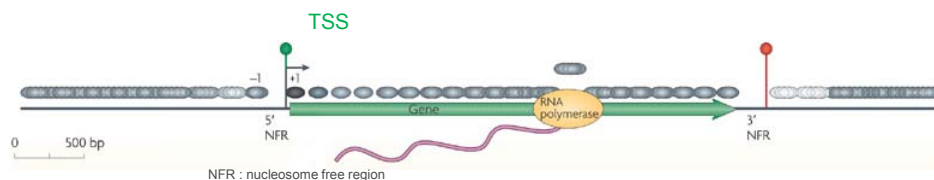
Wang et al. PNAS 2007

Trans-contexte : l'épigénétique

- Distribution des nucléosomes
 - Un nucléosome : ~147 bp d'ADN et 1 octamère de protéines histones)
 - Techniques : DNase-seq, FAIRE-seq...

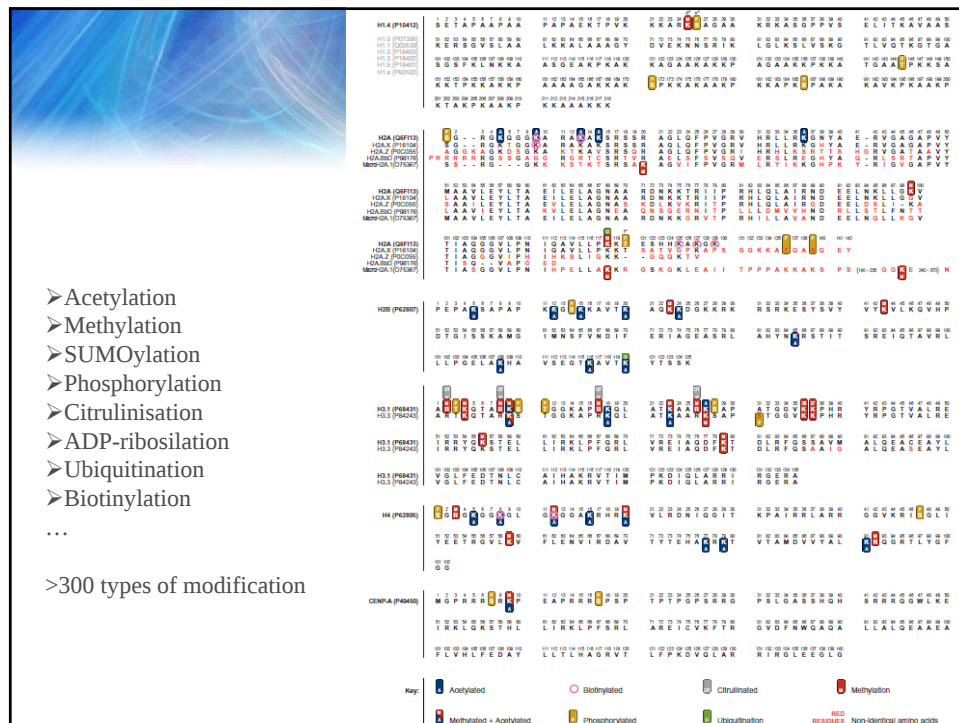


Distribution consensus des nucléosomes sur les gènes de *S. cerevisiae*.



Trans-contexte : l'épigénétique

- Distribution des nucléosomes
- Modification des histones (méthylation, acétylation...)
 - => compaction/décompaction de la chromatine, répression ou activation de la transcription des gènes voisins
 - => Cartographie de la modification des histones:
 - ChIP-seq avec anticorps dirigé contre la modification ciblée

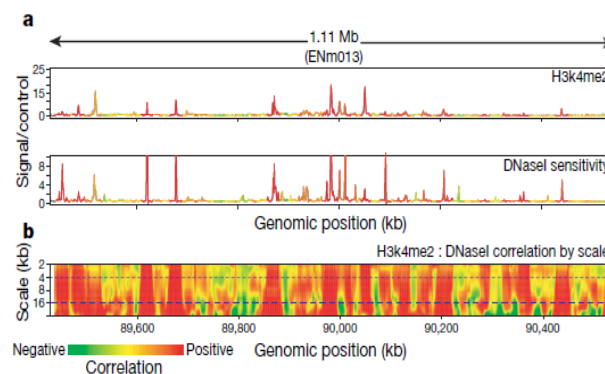


Modifications des histones

Type of Modifications	H3K4	H3K9	H3K27	H3K79	H4K20	H2BK5
monomethylation	activation	activation	activation	activation	activation	activation
dimethylation				activation		
trimethylations	activation	repression	repression	répression activation		
Type of Modifications	H3K9	H3K14				
acetylation	activation	activation				

Trans-contexte : l'épigénétique

Corrélation entre la bi-méthylation de la lysine 4 de H3 et la sensibilité à la DNase



Trans-contexte : l'épigénétique

- Distribution des nucléosomes
- Modifications des histones (méthylation, acétylation...)
- Cartographie de la polymérase II et de la polymérase III
ChIP-seq avec anticorps dirigé contre la polymérase
- Méthylation de l'ADN
 - Méthylation des cytosines qui entraîne généralement une répression de la transcription
 - Technique : methyl-seq

Encyclopedia of DNA Elements (ENCODE)

- Consortium international
- But : catalogue complet des éléments fonctionnels sur le génome humain
- 1ère phase : 1% du génome humain
The ENCODE Project Consortium, Nature 2007
- Sept 2007 : début de la phase "production"
 - Echelle du génome complet
 - Dvpt de protocoles expérimentaux standardisés
 - Unification des formats de données
 - Utilisation de séquençage de nouvelle génération

Encyclopedia of DNA Elements (ENCODE)

Table 1. Summary of ENCODE datasets, as of 15 September 2009

Data type	Description	Investigators	Number of experiments
BiP	Bi-directional promoters	NHGRI	2
CAGE	5' cap analysis gene expression	Riken	11
ChIP-seq	TF and polymerase binding, histone marks by ChIP	Yale, UC Davis, HudsonAlpha, Broad, UW, UNC	185
DNA-seq	DNA fragment sequencing	Genome Inst Singapore	5
DNase-seq	DNaseI hypersensitivity	UW, Duke	20
Exon-array	Gene expression by all-exon microarray	Affymetrix/CSHL	10
FAIRE-seq	Formaldehyde Assisted Isolation of Regulatory Elements	U. Texas	5
Genes	High-quality gene annotations	Gencode/Sanger	3
Mapability	Uniqueness of short read nmers	Broad, Duke, UMass	5
Methyl27	DNA methylation by Illumina 27K	HudsonAlpha	3
Methyl-seq	DNA methylation by restriction enzymes	HudsonAlpha	15
NRE	Negative regulatory elements	NHGRI	6
PET	5'- and 3'-paired-end tags	Genome Inst. Singapore	13
RIP-chip	RNA-binding proteins	SUNY Albany	7
RNA-chip	RNA microarray	Affymetrix/CSHL	25
RNA-seq	RNA sequencing	Caltech, CSHL, GIS, Yale	23
TbaAlign	Multi-species alignment with TBA	NHGRI	1
CNV	Copy number variation	HudsonAlpha	3
DHS-SC	Chromatin interactions: DHS versus TSS	U Washington	2
5C	Chromatin interactions: pilot region	U Mass	2
Total			341

Rosenblum et al. NAR 2010

Encyclopedia of DNA Elements (ENCODE)

Visualisation à l'UCSC

The screenshot displays the ENCODE UCSC Genome Browser interface. It features two main sections: 'Expression' and 'Regulation'. Each section contains a grid of tracks with checkboxes and 'hide' buttons. The 'Expression' section includes tracks like Affy Exon..., Affy GNF1H, Affy RNA Loc, Affy U133, Affy U133Plus2, Affy U95, Agilent Array, Allen Brain, Bertone Yale TAR, CSHL Long RNA-seq, GIS PET RNA, GNF Atlas 2, GNF Ratio, Helicos RNA-seq, Illumina WG-6, RIKEN CAGE Loc, Sestan Brain, and UW Affy Exon. The 'Regulation' section includes tracks like Broad Histone, CpG Islands, EIO/ICV LNAS, Eponine TSS, FirstEF, GIS ChIP-PET, HAIB Methyl-seq, HAIB Methyl27, HAIB TFBS, NHGRI Bi-Pro, Open Chromatin, ORC2 Anno, SUNY RBP, SwitchGear TSS, TFBS Conserved, TS miRNA sites, UW DNaseI HS, Vista Enhancers, Yale TFBS, 7X Reg Potential, FOX2 CLIP-seq, LI/UCSD TAFI..., NKI Nuc Lamina..., and Nucleosome Occupancy... Each track has a 'hide' button and a checkbox to toggle its visibility.

Rosenblum et al. NAR 2010

Perspectives : un long chemin !

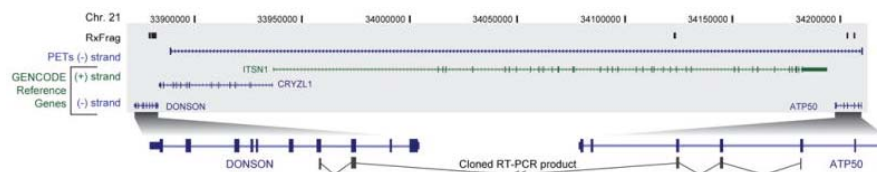
Deciphering gene regulatory networks in the postgenomic era will provide pivotal insights into genome function and human disease, but it will require a much improved knowledge of evolutionary forces that shape transcriptional networks.

Modifications d'histones dans diverses lignées cellulaires mappées par le projet EncODE

*- All	Cell Line	GM12878	H1-hESC	HepG2	HMEC	HSMM	HUVEC	K562	NHEK	NHLF
Antibody										
CTCF	+									
H3K4me1	+									
H3K4me2	+									
H3K4me3	+									
H3K9ac	+									
H3K9me1	+									
H3K27ac	+									
H3K27me3	+									
H3K36me3	+									
H4K20me1	+									

Beaucoup de questions !

Transsplicing : notion de gènes ?



Abondance des transcrits !

