

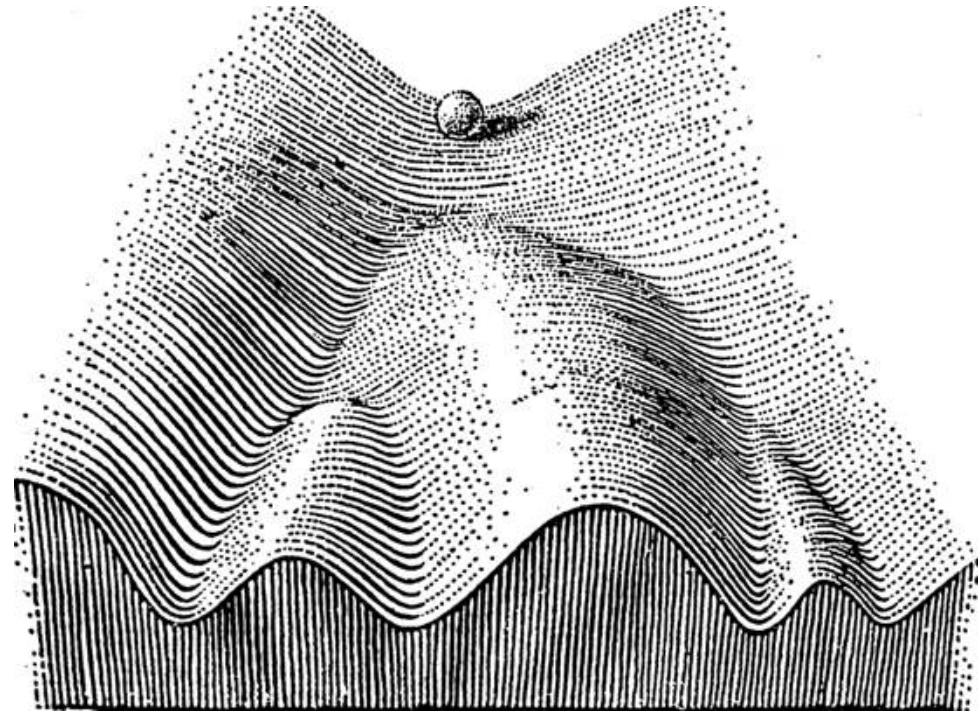
*Chapitre 4*  
***Des modifications chromatiniennes à  
l'épigénomique***

Université Populaire de Marseille, printemps 2019  
Cycle *Hérédité génétique et épigénétique*  
Jacques van Helden

## *Chapitre 4*

# *De l'épigénétique à l'épigénomique*

- Chaque facteur transcriptionnel active / réprime spécifiquement l'expression de certains gènes
- Les réseaux de régulation permettent de comprendre
  - L'adaptation cellulaire
  - Les voies de différenciation durant l'embryogénèse
- Questions ouvertes
  - Qu'est-ce qui explique la spécificité de la régulation ? (information des motifs de liaison)
  - Les boucles de rétroaction suffisent-elles pour expliquer
    - la stabilité des états épigénétiques ?
    - leur transmission au fil des divisions mitotiques ?
    - une éventuelle transmission trans-générationnelle des caractères phénotypiques ?

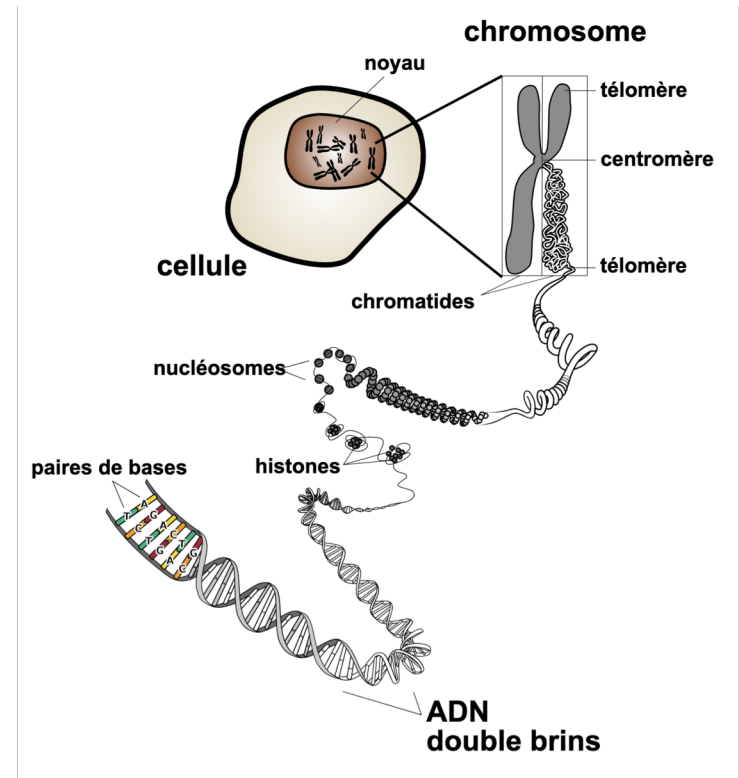


# *Organisation de la chromatine*



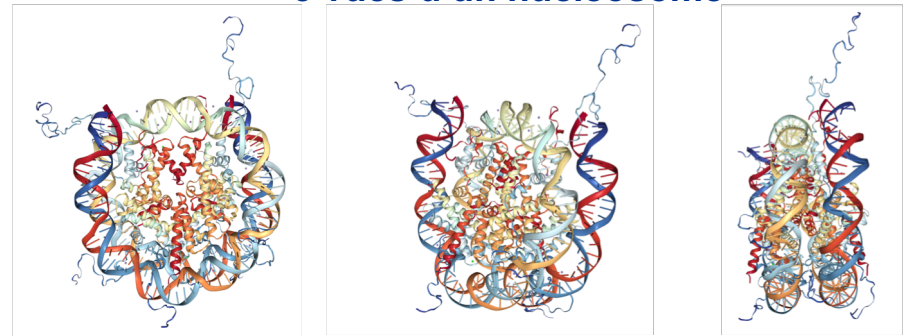
# La chromatine

- Chez les eucaryotes, l'ADN est empaqueté dans le noyau cellulaire sous une forme très compacte, qui résulte d'une série de mécanismes successifs.
- La double hélice d'ADN s'enroule autour d'un complexe formé de 8 protéines appelées **histones**.
- On appelle « **nucléosome** » l'unité formée par un segment d'ADN enroulé autour d'un complexe d'histones.
- Pendant la mitose (division cellulaire), la chaîne de nucléosomes forme à son tour en enroulement hélicoïdal, qui en forme un second, ce qui produit la forme la plus compacte des chromosomes en forme de « X ».
- On appelle « **chromatine** » la forme compacte de l'ADN, qui est composée d'ADN, de protéines et d'ARN.



[https://upload.wikimedia.org/wikipedia/commons/2/2a/Chromosome\\_fr.svg](https://upload.wikimedia.org/wikipedia/commons/2/2a/Chromosome_fr.svg)

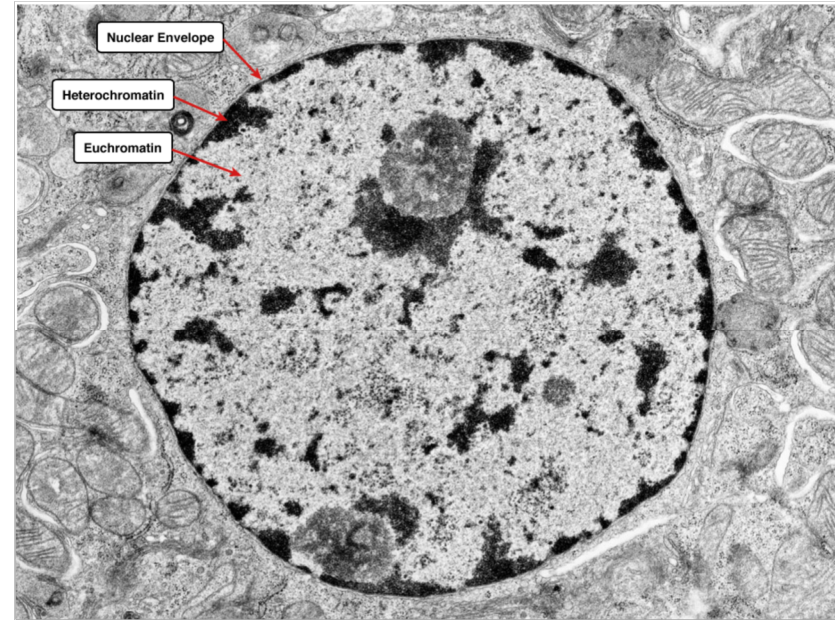
## 3 vues d'un nucléosome



<https://www.rcsb.org/3d-view/1KX5/1>

# Euchromatine et hétérochromatine

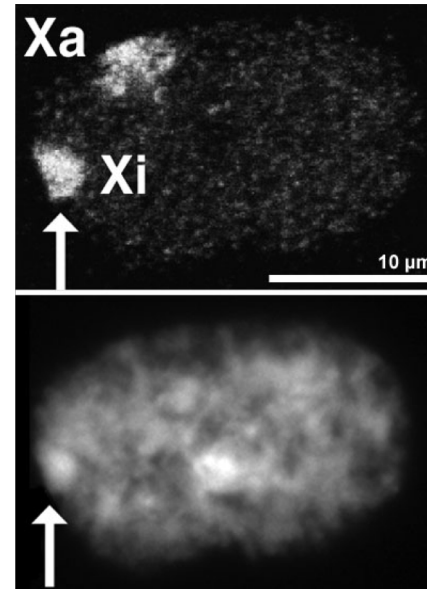
- On distingue l'**euchromatine**, qui peut passer d'une forme compacte à une forme ouverte, de l'**hétérochromatine**, qui est en permanence compactée.
  - Photo: microscopie électronique.
    - Zones sombres : hétérochromatine
    - Zones claires : euchromatine
- L'**activité transcriptionnelle** requiert une ouverture des chromosomes. On ne la trouve donc que dans l'**euchromatine**.
- Hétérochromatine **constitutive**
  - extrémités des chromosomes (**télomères**)
  - **centromères**
- Hétérochromatine **facultative**
  - associée à l'**inactivation transcriptionnelle** de certaines régions chromosomiques dans certains tissus.



[http://medcell.med.yale.edu/histology/cell\\_lab/euchromatin\\_and\\_heterochromatin.php](http://medcell.med.yale.edu/histology/cell_lab/euchromatin_and_heterochromatin.php)

# Inactivation du chromosome X

- Le **corpuscule de Barr** est une zone d'hétérochromatine qu'on trouve spécifiquement dans les cellules des femelles de mammifères.
- Il correspond au chromosome X inactif dans chaque cellule.
- L'inactivation du X est déclenchée à un stade précoce de l'embryogenèse, de façon indépendante dans chaque cellule.
- Chaque cellule transmet ensuite le caractère actif/inactif à ses descendantes.
- Le pelage « écaille de tortue » des chats provient de l'inactivation de l'un ou l'autre des chromosomes.
- Photo: microscopie à fluorescence
  - Haut: marquage des deux chromosomes
    - Xa: chromosome X actif
    - Xi et flèche verticale chromosome X inactif.
  - Bas: marquage du corpuscule de Barr



Marquage fluorescent des chromosomes X

- actif (Xa)
- inactif (Xi)

Corpuscule de Barr

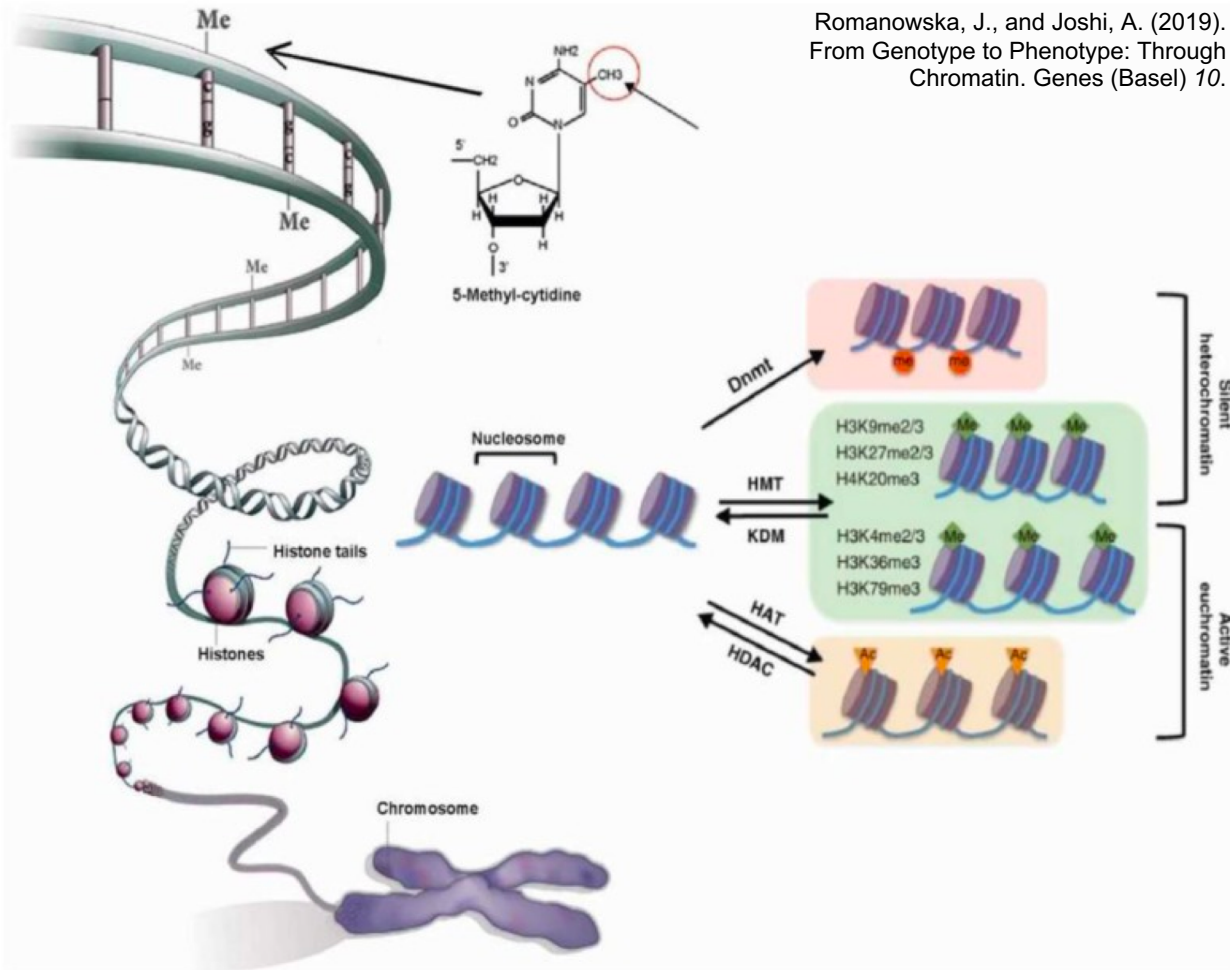
Steffen Dietzel, Dissertation an der Universität Heidelberg, 1996.  
[https://en.wikipedia.org/wiki/Barr\\_body#/media/File:Sd4hi-unten-crop.jpg](https://en.wikipedia.org/wiki/Barr_body#/media/File:Sd4hi-unten-crop.jpg)



[https://fr.wikipedia.org/wiki/Corpuscule\\_de\\_Barr#/media/File:Six\\_weeks\\_old\\_cat\\_\(aka\).jpg](https://fr.wikipedia.org/wiki/Corpuscule_de_Barr#/media/File:Six_weeks_old_cat_(aka).jpg)

# *Modifications chromatiniennes*

# Modifications chromatiniennes affectant la transcription



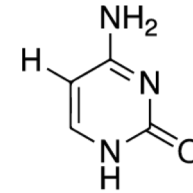
Romanowska, J., and Joshi, A. (2019). From Genotype to Phenotype: Through Chromatin. Genes (Basel) 10.

- Différents types de **modifications de la chromatine** affectent la transcription.
- Inactivation (« *silencing* »)
  - Méthylation ADN
  - Méthylation histones
    - H3K9me3/4
    - H3K27me2/3
    - H4K20me3
- Activité
  - Méthylation histones
    - H3K4me2/3
    - H3K36me3
    - H3K78me3
  - Acétylation histones

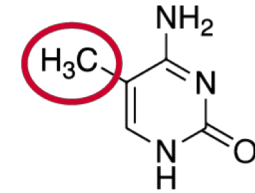


# Méthylation de l'ADN

- Des enzymes spécialisées, les **méthylases de la cytosine**, ajoutent un groupe méthyl aux cytosines.
- Méthylases de l'ADN chez l'humain
  - DNMT3 : méthylation « de novo »
    - Méthylation d'ADN non-méthylé sur les deux brins
  - DNMT1 : « maintenance »
    - lors de la réplication chaque copie de l'ADN est initialement méthylée sur un seul brin
    - DNMT1 reproduit les méthylations sur l'autre brin

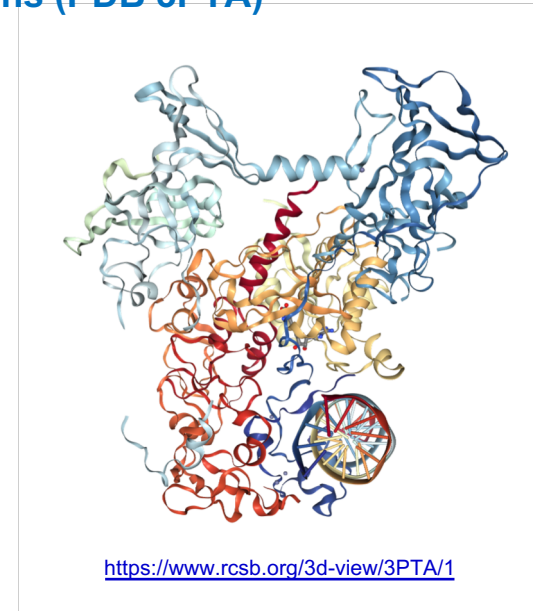
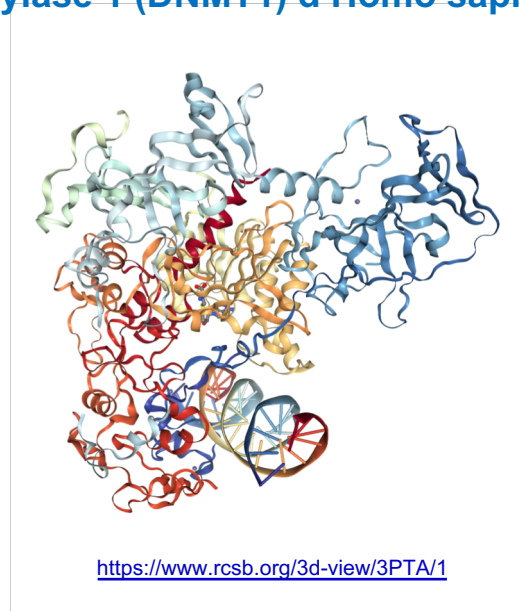
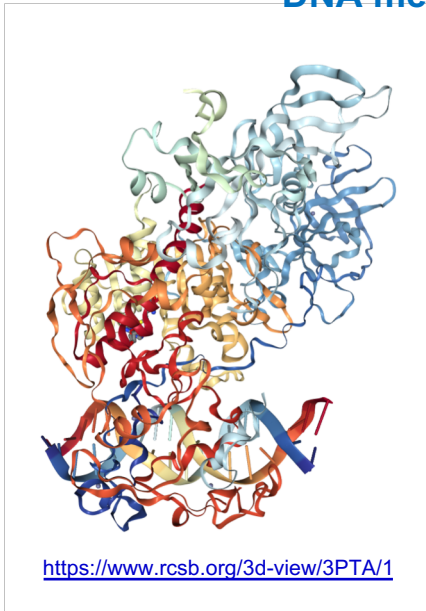


**Cytosine**



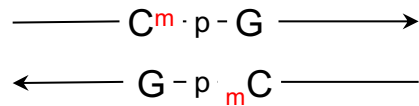
**methyated Cytosine**

## DNA methylase 1 (DNMT1) d'Homo sapiens (PDB 3PTA)

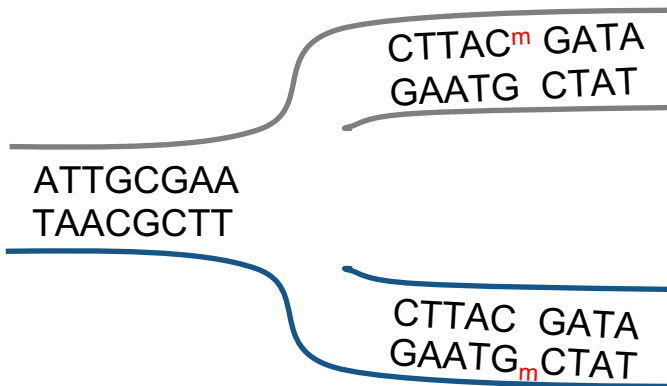


# Méthylation « de novo » et « de maintien » de la cytosine

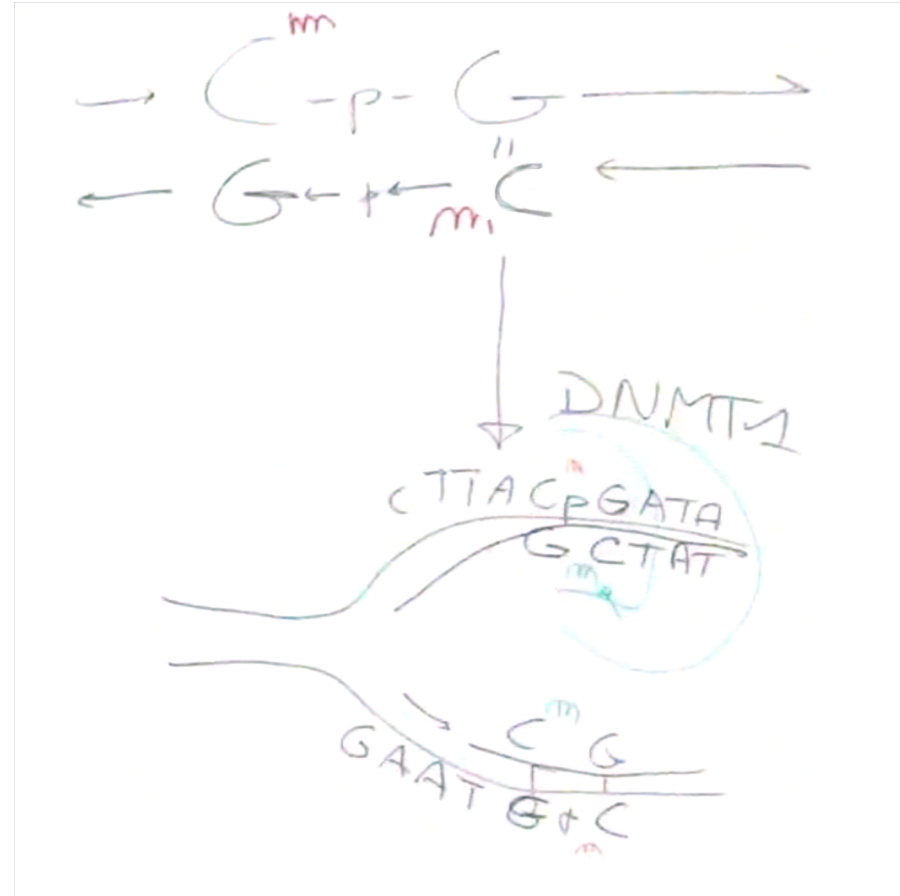
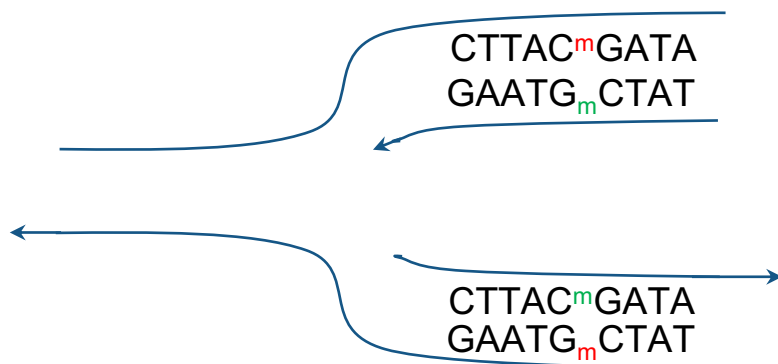
ADN avec cytosine méthylée sur les deux brins



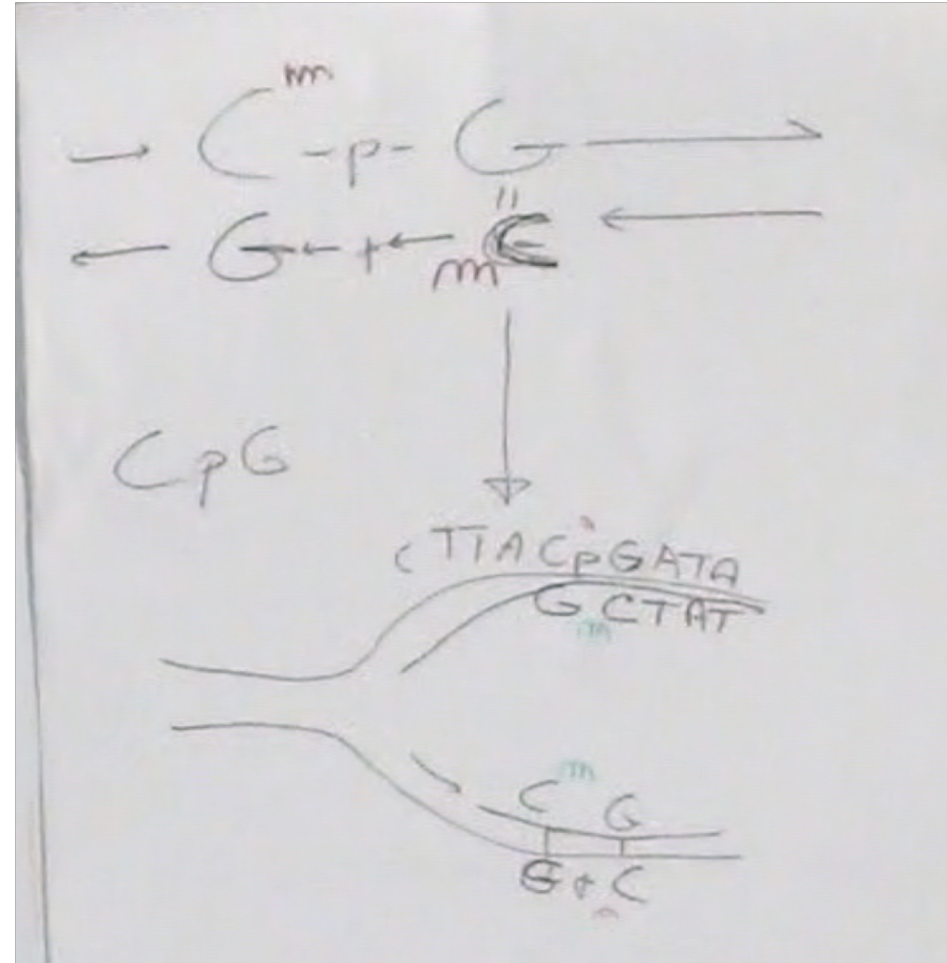
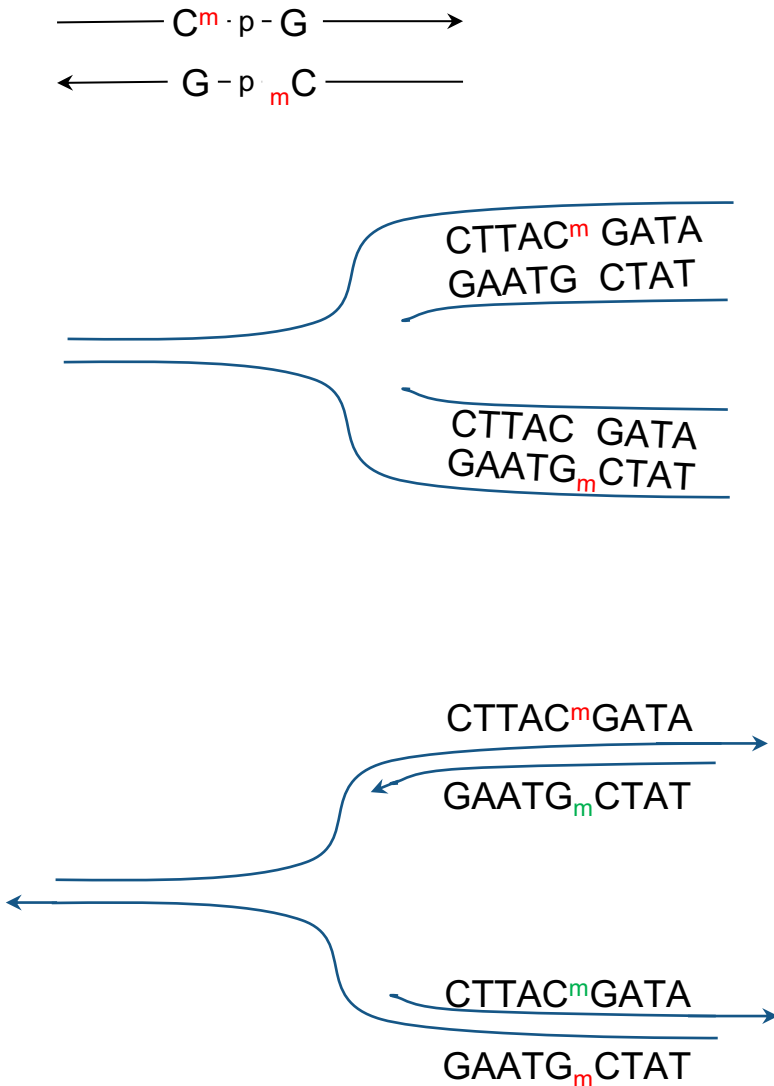
ADN hémi-méthylé suite à la réplication



Méthylation « de maintien » par DNMT1



# Méthylation « de novo » et « de maintien » de la cytosine

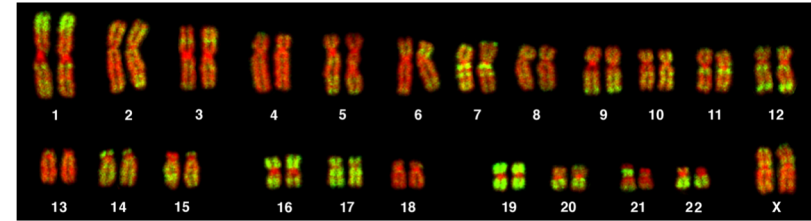






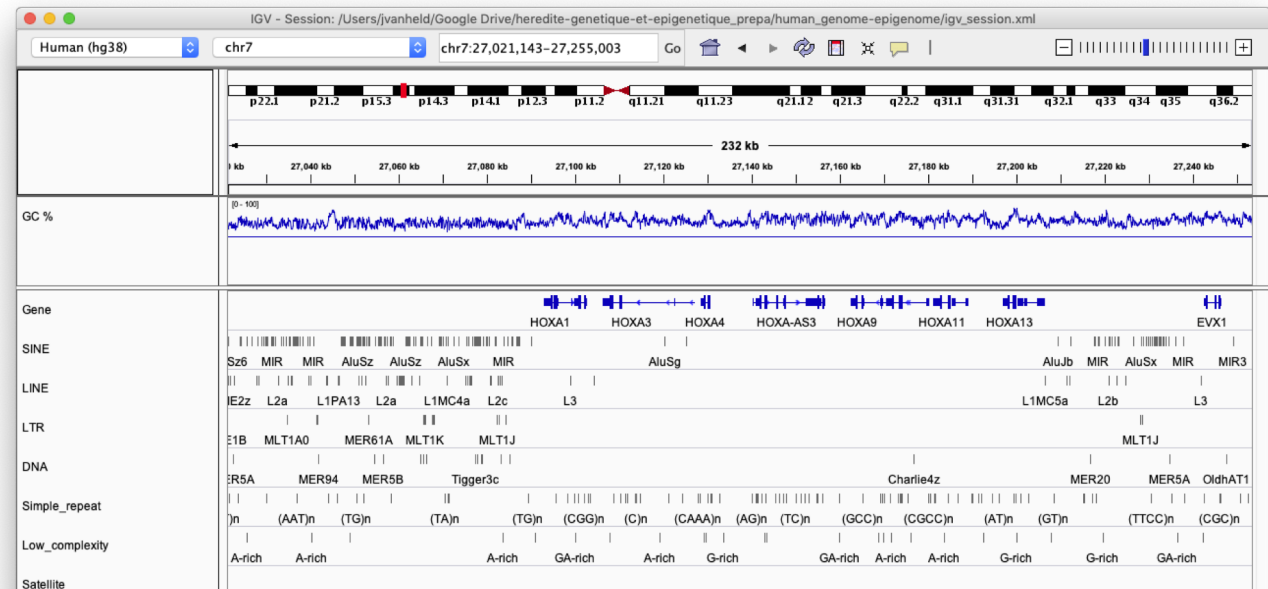
# Éléments répétés dans le génome humain

- Le génome humain est constitué à 45% d'éléments répétés
- L'un des éléments les plus fréquents est la séquence Alu (1 million de copies, couvrant ~10% du génome humain).
- Ces séquences proviennent d'une infection ancestrale par un virus à ARN, qui s'est ensuite répandu dans le génome par rétrotransposition.
- Certaines de ces séquences ont conservé la capacité de se déplacer dans les génomes, et de perturber le fonctionnement des gènes voisins.
- La méthylation de l'ADN réprime la rétrotransposition de ces éléments.



Vert : éléments Alu

Rouge : le reste des chromosomes



# Les nucléosomes

- Chez les eucaryotes, l'ADN est enroulé autour de complexes de 8 protéines appelées **histones**.
- Le complexe est formé de 4 types différents d'histones, chacune en 2 copies.
  - H2A, H2B, H3, H4
- On appelle **nucléosome** le corpuscule formé d'un complexe de 8 histones.
- D'autres histones interviennent dans la compaction de la chromatine en faisant le lien entre les nucléosomes.

Structure cristallographique (rayons X) d'un nucléosome (PDB 1KX5)

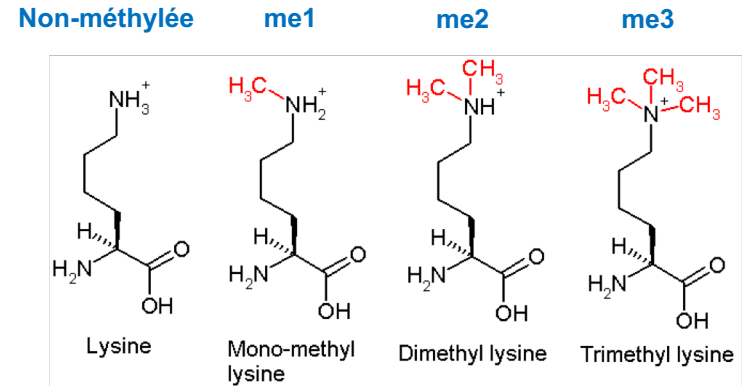


<https://www.rcsb.org/3d-view/1KX5/1>

# Méthylation des queues d'histones

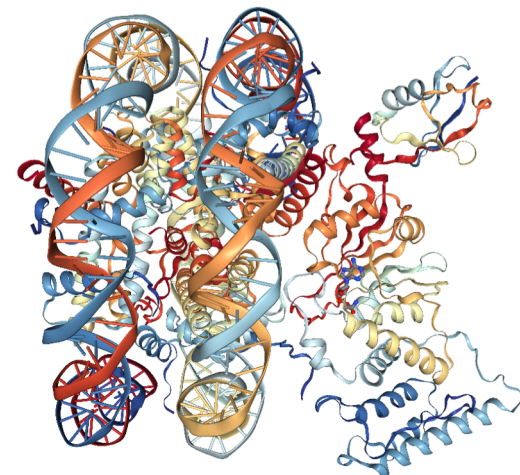
- Les méthyltransférases des histones se lient spécifiquement aux queues des histones et ajoutent un ou plusieurs groupes méthyle
  - à des résidus particuliers (généralement des **lysines**, dénotées par **K**)
  - à des positions spécifiques de la protéine (en particulier K4, K9, K27).

## Etats de méthylation de la lysine



<https://en.wikipedia.org/wiki/H3K27me3#/media/File:Methylation-lysine.PNG>

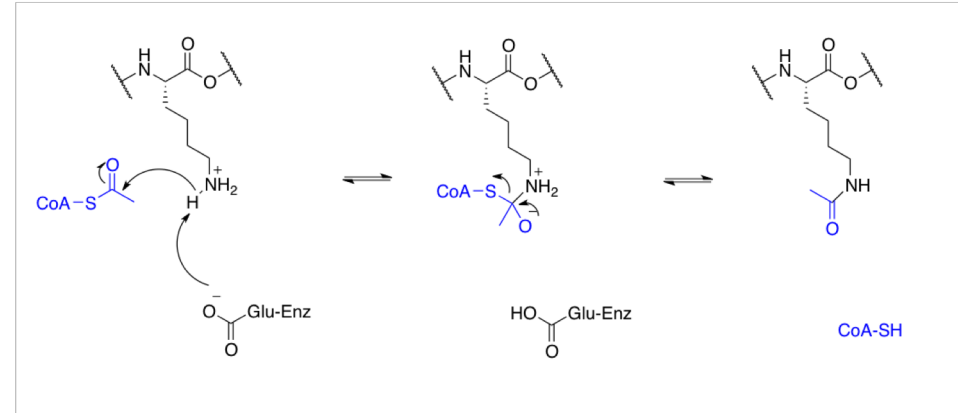
## Méthyltransférase DOT1L complexée avec un nucléosome



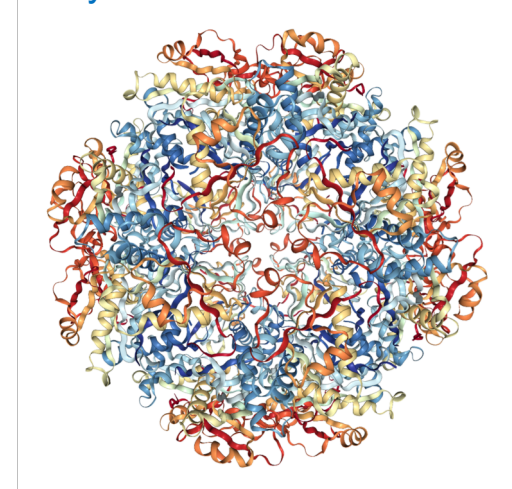
<http://www.rcsb.org/3d-view/6J99/1>

# Acétylation de la lysine des histones

- Des enzymes spécialisées peuvent ajouter (acétylation) ou supprimer (désacétylation) un groupe méthyle à une lysine de la queue des histones.
  - ▣ HAT histone acetyltransferase
  - ▣ HDAC histone deacetylase
- L'acétylation relâche la compaction de la chromatine.
- On l'observe fréquemment dans les **promoteurs des gènes activement transcrits**.



Acétyltransférase de l'histone



<http://www.rcsb.org/3d-view/5TRM/1>

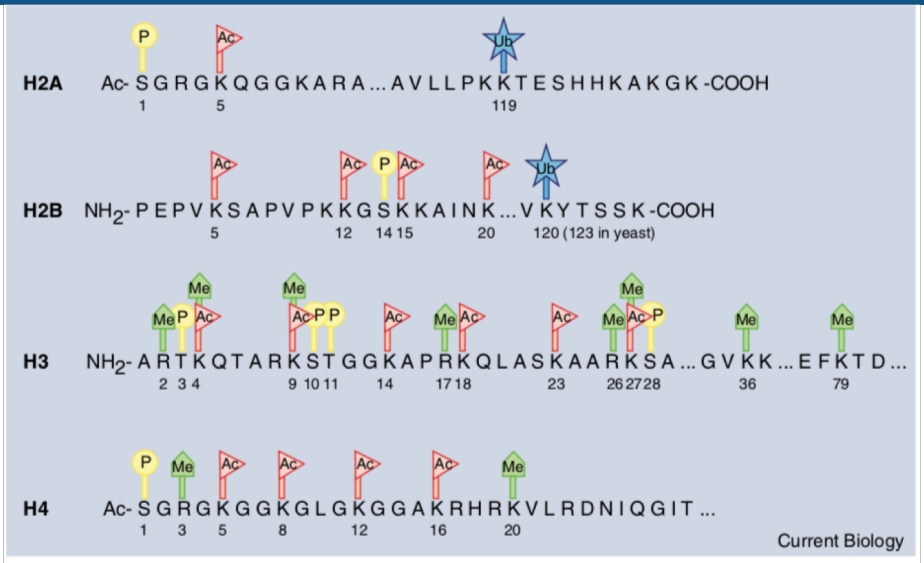
- Il est vraisemblable que les modifications d'histones soient transmises au cours des divisions cellulaires.
- On pense que cette transmission mitotique assure la stabilité de la différenciation cellulaire.
- Les mécanismes ne sont pas connus, et a priori pas évident
  - Lors de la réplication, les histones sont perdus ou bien se retrouvent sur un seul brin.
  - On ignore toujours la façon dont les profils de modifications d'histone sont rétablis lors de la réplication





# Quelques exemples de modifications d'histones indicatives

- Certaines modifications d'histones sont fréquemment associées
  - à des régions particulières de l'ADN
    - corps des gènes
    - promoteurs
    - modules cis-régulateurs
  - à des niveaux particuliers d'activité transcriptionnelle
    - répression
    - transcription active
    - polymérase en pause sur le promoteur, « prête à démarrer »



Peterson, C.L., and Laniel, M.-A. (2004). Histones and histone modifications. *Curr. Biol.* 14, R546-551.

Histone	Résidu	Modif	Effet	Localisations
H2B	K5	me1	activation	
H2B	K5	me3	répression	
H3	K122	ac	activation	promoteurs prêts à démarrer la transcription
H3	K14	ac	activation	promoteurs des gènes activement transcrits
H3	K27	ac	activation	promoteurs actifs et enhancers
H3	K27	me3	répression	
H3	K27	me3	répression	régions génomiques facultativement réprimées par formation d'hétérochromatine
H3	K36	me3	activation	corps des gènes activement transcrits
H3	K4	me1	activation	promoteurs actifs et enhancers
H3	K4	me2	activation	
H3	K4	me3	activation	
H3	K9	ac	activation	promoteurs des gènes activement transcrits
H3	K9	me3	répression	gènes constitutionnellement réprimés
H4	K20	me1	activation	



# *Epigénomique*

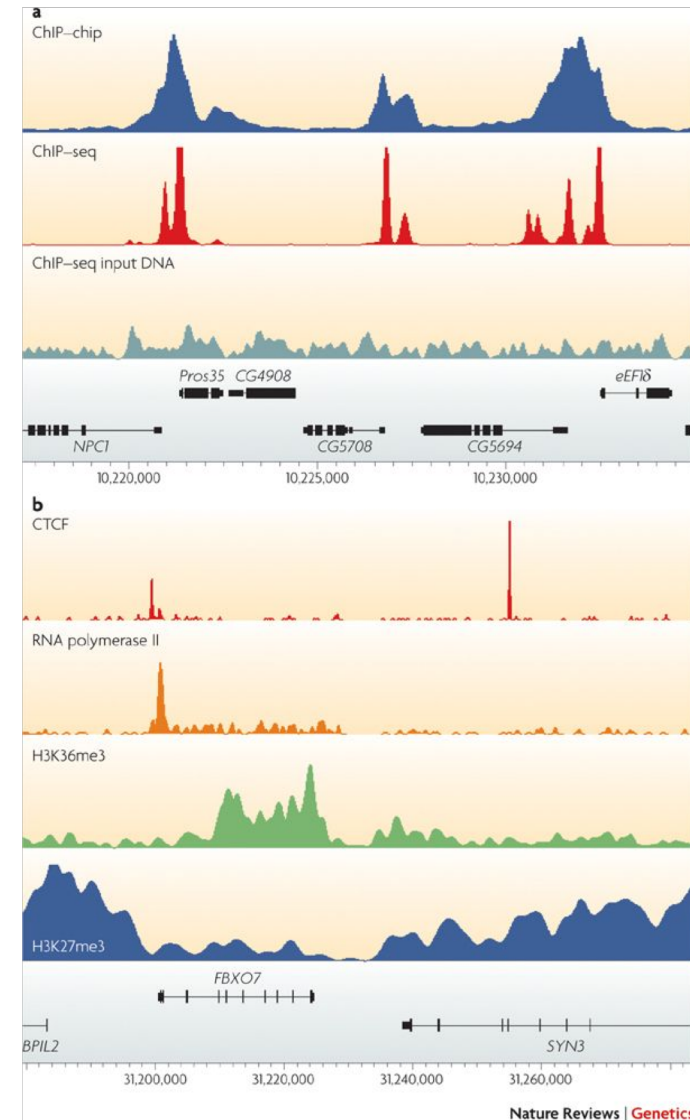
# Détection de modifications chromatiniennes dans les génomes complets

- Depuis 2001 on dispose de la séquence complète du génome humain.
- **Epigénomique**
  - Différentes méthodes à haut débit (ChIP-chip, ChIP-seq) permettent de détecter les régions génomiques à chaque position du génome.
  - On constate une corrélation entre les régions marquées et les différents types d'activité transcriptionnelle.

Polymérase de l'ADN

Régions géniques transcrits

Régions réprimées



# Roadmap – cartes épigénomiques des tissus humains

- Plusieurs projets internationaux ont visé à caractériser les profils épigénomiques et transcriptomiques à large échelle
  - **Épigénome** : détection de modifications chromatinienne tout au long du génome
  - **Transcriptome** : mesure du le niveau de transcription de chaque gène.
- Exemples de grands projets
  - **ENCODE**: épigénome et transcriptome dans différentes lignées de cultures cellulaires (cellules de foie, neurones, dérivées de cancer, ...)
  - **Roadmap epigenomics**: épigénome et transcriptome 111 tissus et types cellulaires
  - **GTEX**: transcriptome dans 53 tissus différents.

**ROADMAP epigenomics PROJECT**

HOME PARTICIPANTS BROWSE DATA PROTOCOLS COMPLETE EPIGENOMES TOOLS PUBLICATIONS

Data

DATA TABLE ADULT BRAIN FETAL STEM CELLS

**Nature, 2015: INTEGRATIVE ANALYSIS OF 111 REFERENCE HUMAN EPIGENOMES**

Uniformly re-processed data, integrative analysis products and interactive browser sessions can be found at the supplementary website for the 2015 Consortium paper at <http://compbio.mit.edu/roadmap>

VIEW DATA

**DATA BROWSER**

The NIH Roadmap Epigenomics Mapping Consortium aims to produce a public resource of epigenomic maps for stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease.

**Data Releases**

The current release 9 of the Human Epigenome Atlas is a product of the NIH Roadmap Epigenomics Consortium. Release 9 contains a total of 2,804 genome-wide datasets, including 1,821 histone modification datasets, 360 DNase datasets, 277 DNA methylation datasets, and 166 RNA-Seq datasets, encompassing a total of 150.21 billion mapped

**Download Data**

Data can be downloaded from the following sources:

- [Supplementary website for the 2015 Consortium paper](#)
- [NIH Roadmap Epigenomics - GEO - NCBI](#)
- [BCM Roadmap Epigenomics EDACC](#)

<http://www.roadmapepigenomics.org/data/>

# Détection de marques d'histones par la technologie ChIP-seq

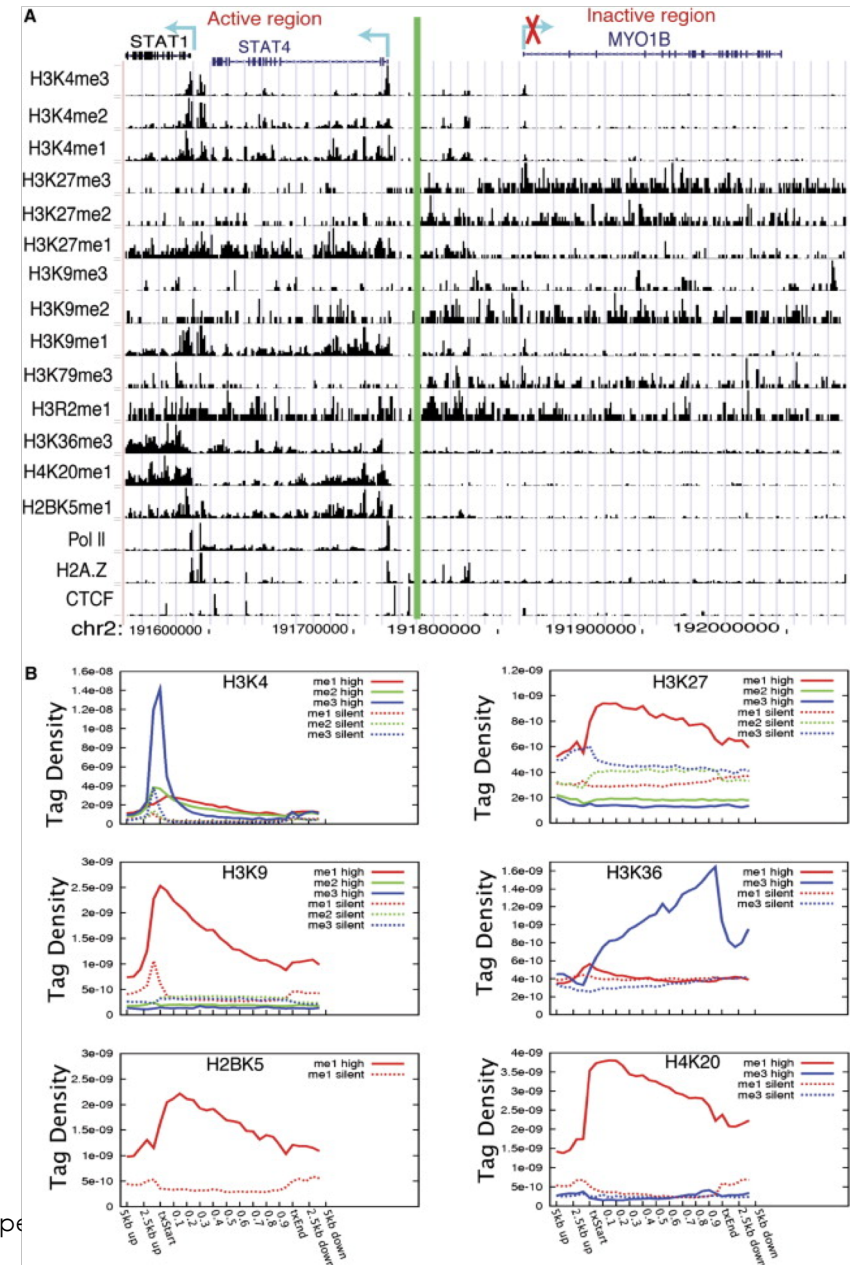
- La technologie ChIP-seq permet de détecter différentes marques d'histone et révèle l'état transcriptionnel des gènes

## Figure A

- Les gènes en cours de transcription montrent des profils radicalement différents des gènes réprimés.
- On retrouve à échelle génomique les associations caractéristiques des différentes modifications chromatinienne.

## Figure B

- La distribution moyenne d'une modification sur l'ensemble de tous les gènes confirme les tendances individuelles
- H3K4me3 est concentré au site d'initiation de la transcription
- H3K36me3 marque le corps des gènes actifs
- ...

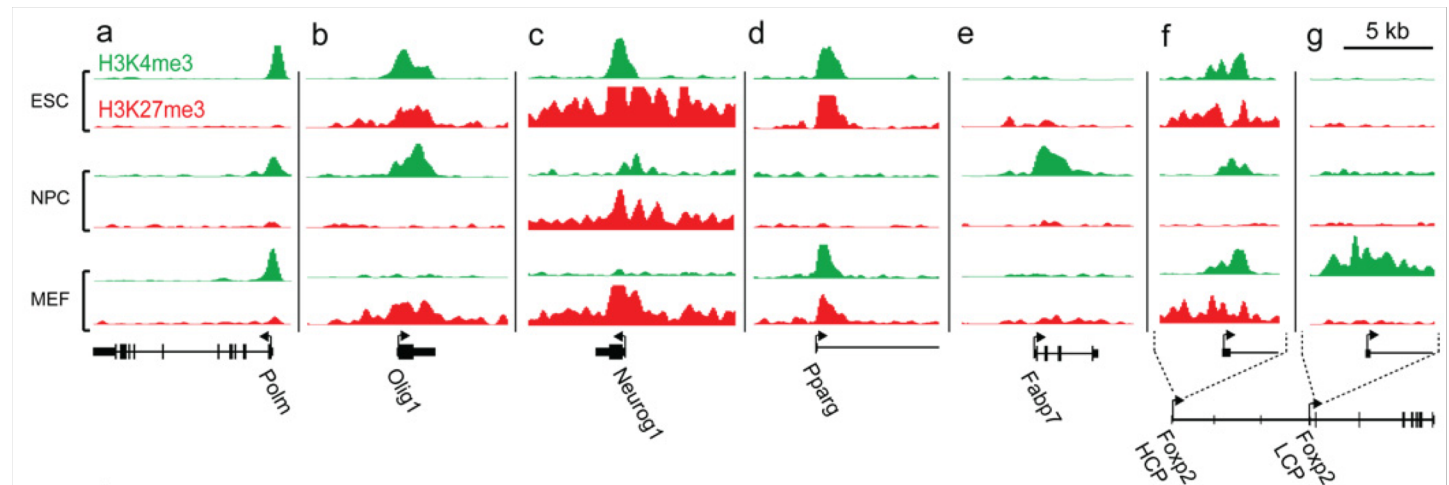


- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepe
- methylations in the human genome. *Cell* 129, 823-837.

# Modification spécifiques des tissus lors du développement

- Mikkelsen et collègues comparent les profils de H3K4me3 (activité transcriptionnelle) et H3K27me3 (répression) pour différents gènes, dans différents types de cellules.
  - ESC cellules souches embryonnaires
  - NPC précurseurs de neurones
  - MEF fibroblastes (cellules de soutien des tissus conjonctifs)

- (b) Olig1: facteur de transcription neural
- (c) Neurog1: FT de neurogénèse
- (d) Pparg: adipogenèse
- (e) Fabp7: marqueur de progéniteurs neuraux
- (f) Foxp2: cerveau et poumon
- (g) Foxp2 LCP



- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.

# Genotype-Tissue Expression (GTEx) project

- La spécificité tissulaire d'expression d'un gène reflète généralement sa fonction
- Le projet GTEx a caractérisé le profils transcriptomique dans 53 tissus humains

**Configuration** pack go

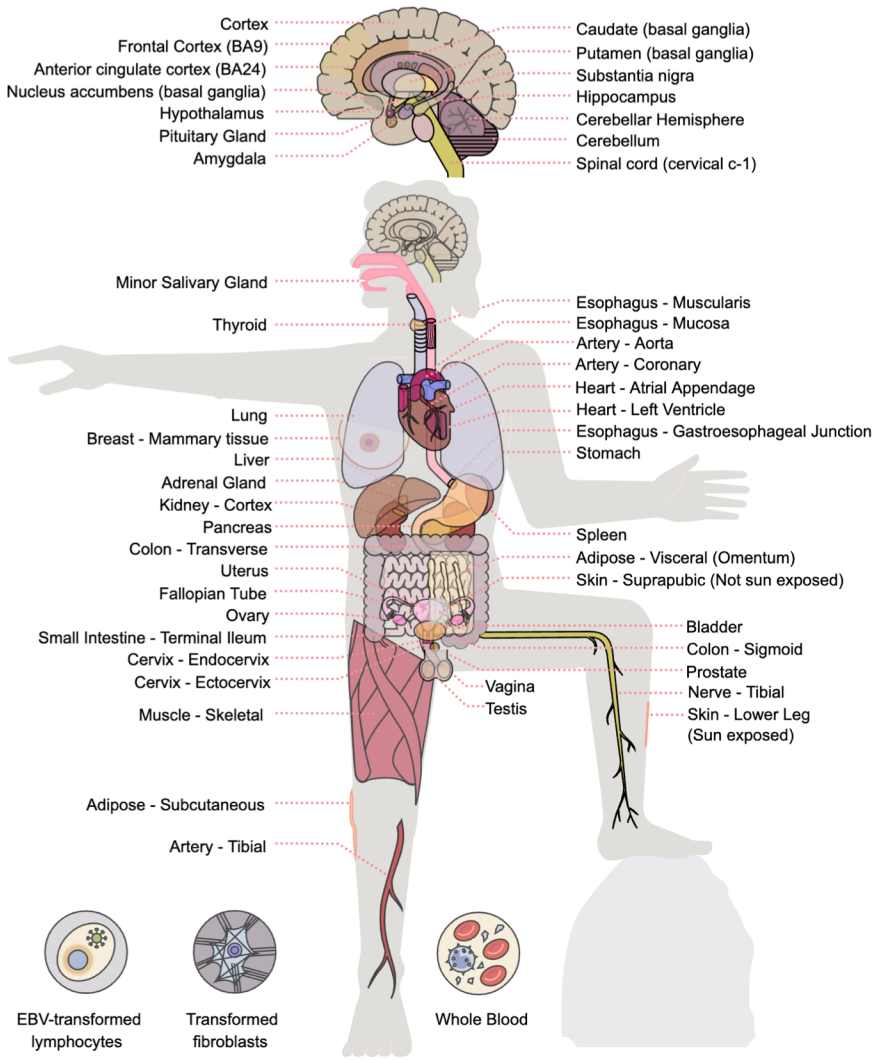
Label:  gene symbol  accession  both    Show GTEx gene model:

Log10 transform:     View limits maximum:  RPKM (range 0-711778)

Limit to protein-coding genes:     Limit to genes scored at or above:  (range 0-1000)

**Tissues** Click below or in Body Map to change tissues set all clear all

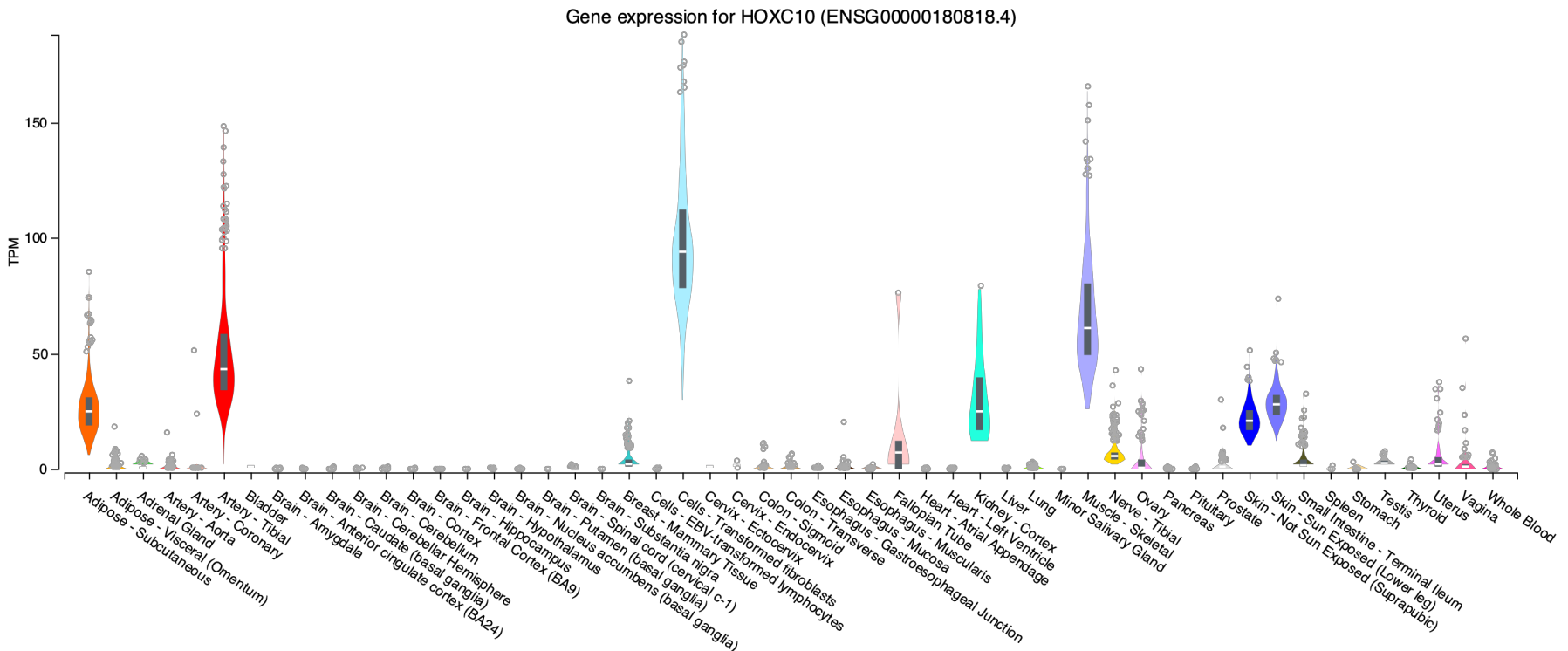
- |  |  |
|--|--|
| <input type="checkbox"/> Adipose - Subcutaneous                    | <input type="checkbox"/> Esophagus - Gastroesophageal Junction |
| <input type="checkbox"/> Adipose - Visceral (Omentum)              | <input type="checkbox"/> Esophagus - Mucosa                    |
| <input type="checkbox"/> Adrenal Gland                             | <input type="checkbox"/> Esophagus - Muscularis                |
| <input type="checkbox"/> Artery - Aorta                            | <input type="checkbox"/> Fallopian Tube                        |
| <input type="checkbox"/> Artery - Coronary                         | <input type="checkbox"/> Heart - Atrial Appendage              |
| <input type="checkbox"/> Artery - Tibial                           | <input type="checkbox"/> Heart - Left Ventricle                |
| <input type="checkbox"/> Bladder                                   | <input type="checkbox"/> Kidney - Cortex                       |
| <input type="checkbox"/> Brain - Amygdala                          | <input type="checkbox"/> Liver                                 |
| <input type="checkbox"/> Brain - Anterior cingulate cortex (BA24)  | <input type="checkbox"/> Lung                                  |
| <input type="checkbox"/> Brain - Caudate (basal ganglia)           | <input type="checkbox"/> Minor Salivary Gland                  |
| <input type="checkbox"/> Brain - Cerebellar Hemisphere             | <input type="checkbox"/> Muscle - Skeletal                     |
| <input type="checkbox"/> Brain - Cerebellum                        | <input type="checkbox"/> Nerve - Tibial                        |
| <input type="checkbox"/> Brain - Cortex                            | <input type="checkbox"/> Ovary                                 |
| <input type="checkbox"/> Brain - Frontal Cortex (BA9)              | <input type="checkbox"/> Pancreas                              |
| <input type="checkbox"/> Brain - Hippocampus                       | <input type="checkbox"/> Pituitary                             |
| <input type="checkbox"/> Brain - Hypothalamus                      | <input type="checkbox"/> Prostate                              |
| <input type="checkbox"/> Brain - Nucleus accumbens (basal ganglia) | <input type="checkbox"/> Skin - Not Sun Exposed (Suprapubic)   |
| <input type="checkbox"/> Brain - Putamen (basal ganglia)           | <input type="checkbox"/> Skin - Sun Exposed (Lower leg)        |
| <input type="checkbox"/> Brain - Spinal cord (cervical c-1)        | <input type="checkbox"/> Small Intestine - Terminal Ileum      |
| <input type="checkbox"/> Brain - Substantia nigra                  | <input type="checkbox"/> Spleen                                |
| <input type="checkbox"/> Breast - Mammary Tissue                   | <input type="checkbox"/> Stomach                               |
| <input type="checkbox"/> Cells - EBV-transformed lymphocytes       | <input type="checkbox"/> Testis                                |
| <input type="checkbox"/> Cells - Transformed fibroblasts           | <input type="checkbox"/> Thyroid                               |
| <input type="checkbox"/> Cervix - Ectocervix                       | <input type="checkbox"/> Uterus                                |
| <input type="checkbox"/> Cervix - Endocervix                       | <input type="checkbox"/> Vagina                                |
| <input type="checkbox"/> Colon - Sigmoid                           | <input type="checkbox"/> Whole Blood                           |
| <input type="checkbox"/> Colon - Transverse                        |  |





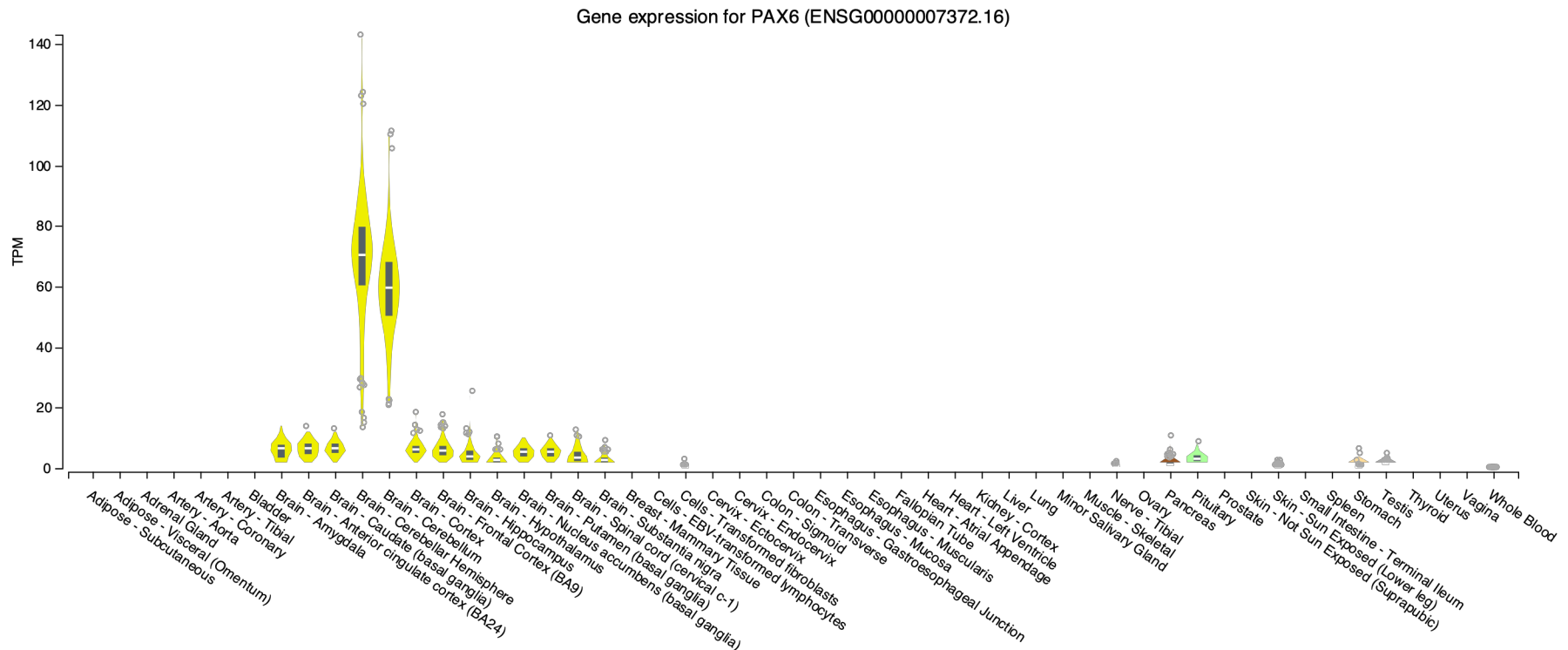
# Expression tissulaire de HOXC10

- Chacun des gènes homéotiques (complexe HOX) s'exprime dans un sous-ensemble spécifique de tissus, qui correspond à la fonction de différenciation tissulaire jouée par ces gènes



# Expression tissulaire de PAX6

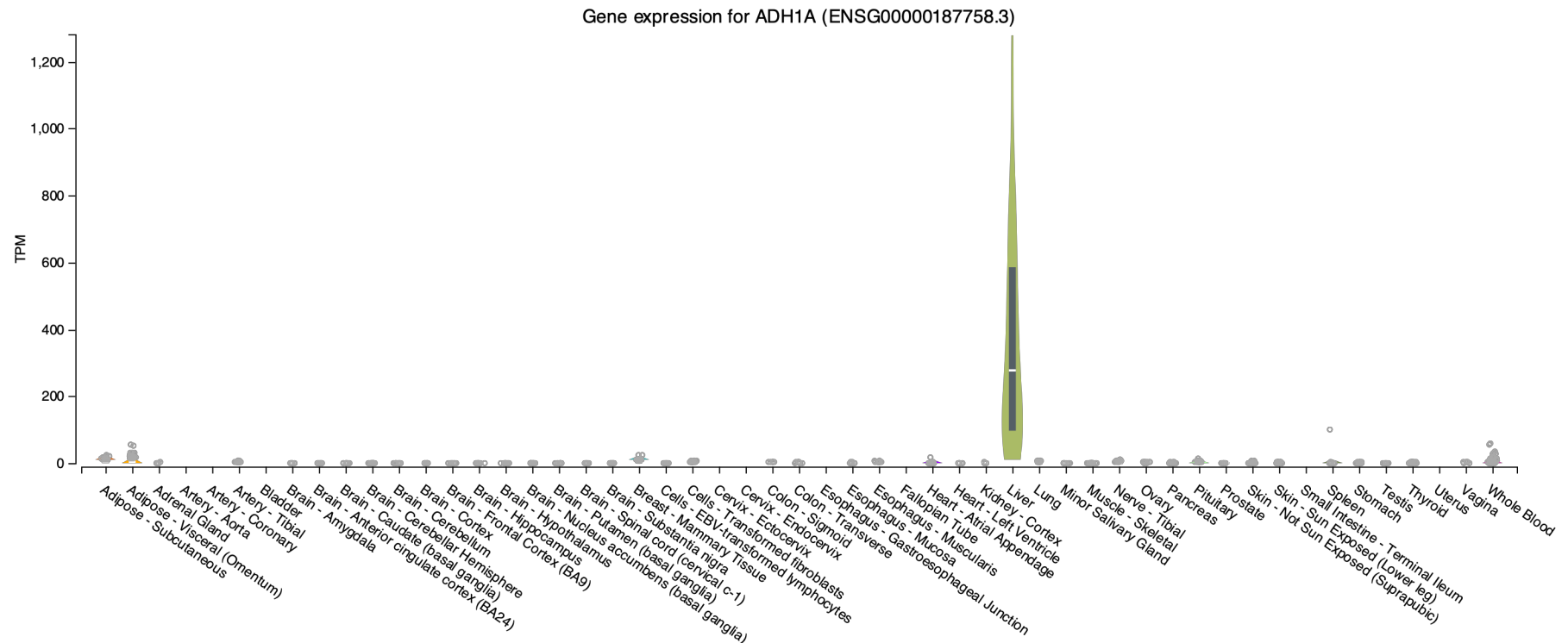
- Le gène PAX6, qui détermine la formation des yeux pendant le développement embryonnaire, s'exprime cerveau, et particulièrement dans le cervelet adulte.





# Expression tissulaire de l'alcool déshydrogénase

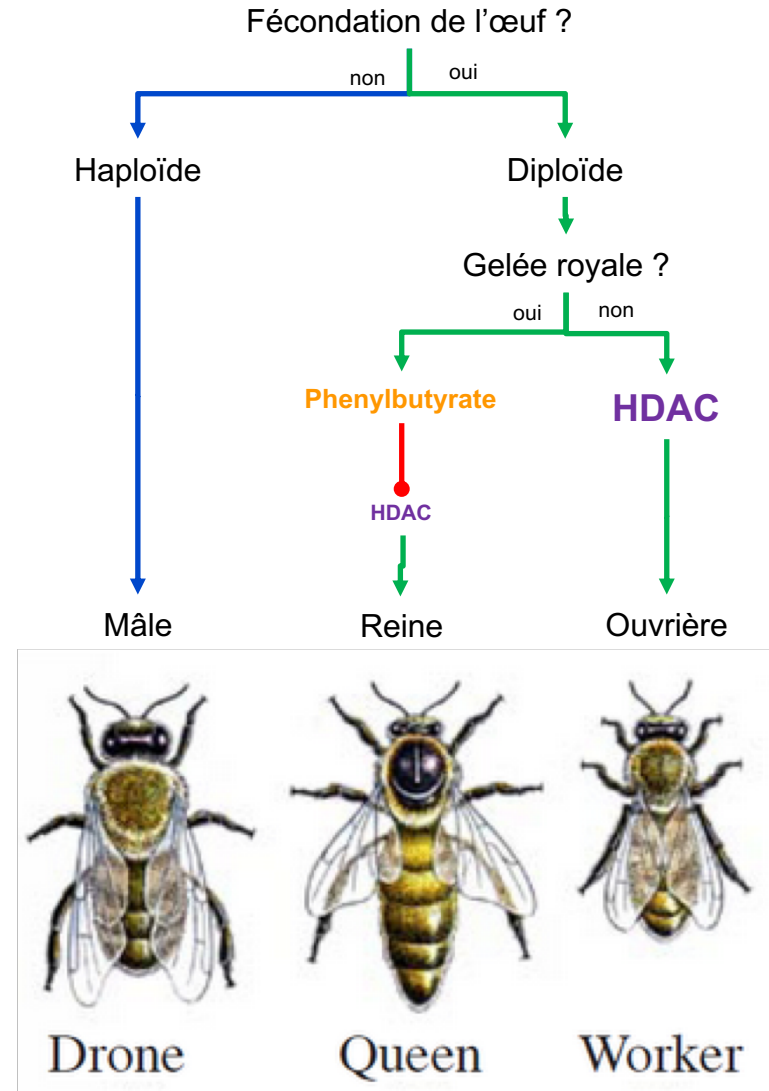
- L'alcool déshydrogénase est exprimée exclusivement dans le foie
- Ceci est parfaitement cohérent avec la fonction de détoxification de cet organe



# *Impact de l'environnement sur l'épigénome*

# Contrôle épigénétique de la différenciation des abeilles

- Chez les abeilles, les larves femelles peuvent se différencier soit en reines soit en ouvrières.
- Ce qui détermine la forme de différenciation ne résulte pas de différences génétiques, mais de l'alimentation (gelée royale).
- Observation : la gelée royale contient du **phénylbutyrate**, qui inhibe les **histone désacétylases (HDAC)**.
  - La différenciation pourrait donc provenir de différences d'acétylation des histones.
  - La méthylation *de novo* est catalysée par la DNA méthylase 3 (DNMT3).
- Expérience : **inhibition de la DNMT3**
  - les abeilles se différencient partiellement en reines (même en absence de gelée royale).



***Rappel: études – génétiques – d'association  
à échelle génomique***

# La séquence, et après ?

. . . .CGATGCTCAAACATTTCAATTTTTTAGGTCAAAAATG  
CCTTAGGTTTAGCACAGCAATGTAGGTGCCAAACTCATCGC  
AGTGAATTGCAGGCGGGAGCAACAAGGACGCCTGCCTCCTT  
TCTGCCTGCTTTTTTGAATAGTCCGATTTGAGAAGGGGACC  
CACGAGAGACAAAAATGCACGCCCCACGCCACATCCTTT  
TTACCCCGCAATGGGTAAAGACTGTCAACAGGCAGGCCACC  
TCGCAGCGTCCGCGGAGTTGCAGGCCCGCCCCGCCAGGGT  
GTGGCGCTGTCCCCCTGGCGCTGGGCGGGGGAGGAGGGGCG  
CGCGGCGGCCGAGGAGGGGCGCGCGGCGGGCGGGGCG  
AGCGGAGGCGAGTGGAGGACGCGTAGACGCGCCGCGGTCCC  
CGCCTGCCGCTGCTCCGCCGAGTCGCCGCTCCAGTCTATC  
CGGCACTAGGAACAGCCCCGAGCGGCGAGACGGTCCCCGCC  
ATGTCTGCGGCCATGAGGGAGAGGTTTCGACCGGTTCTGCA  
CGAGAAGAAGTGCATGACTGACCTTCTGGCCAAGCTCGAGG  
CCAAAACCGGCGTGAACAGGAGCTTCATCGCTCTTGGTGGG  
TGGCCGGGGGTCGCCGCCGCTGGTAGGGCCACGGGAGCCGC  
CGCTGCCCCAGCTGCTGGGGAAGGAAGCAGGGAGAGGACTC  
GGGAAAGGTGGAGTCGGAGACAGACGGGACAAGCAGCATAT  
TCAGGGATCAGGCTGGCCTCCCGGAAAGCGTGGGCATCGGA  
GGACCCCGCGGGGGCTGCCAGGCTGAGGGTTCGCGGGGCTG  
GAGGGCAGCTGCGGCGCCGGGCGCTGGCAGCTGGAAGGGCC  
AGCGCTGACGTATGTCTGCCCCGCGGCCCGGCGCCCTATTC  
CTGCTGTCTGCGGGTGGGCGCGGACGGCGGGGCCCTGC  
GGGCGGGCGCGTTGACGGAGGTACCCGGTCTTACCCGACC  
TCCGTGGAGCTCCGCCCGGAG . . . .

- Le séquençage ne constitue qu'une toute première étape pour l'analyse des génomes.
- Au terme d'un projet de séquençage, on obtient un "texte" formé de 4 lettres (une lettre par nucléotide), et il reste un énorme travail de décryptage pour pouvoir interpréter ce texte.
- L'exemple ci-contre montre un fragment de 1000 nucléotides du génome humain.
- Le génome complet comporte 3 milliards de caractères, soit 3 millions de fois plus de lettres que le texte ci-contre.



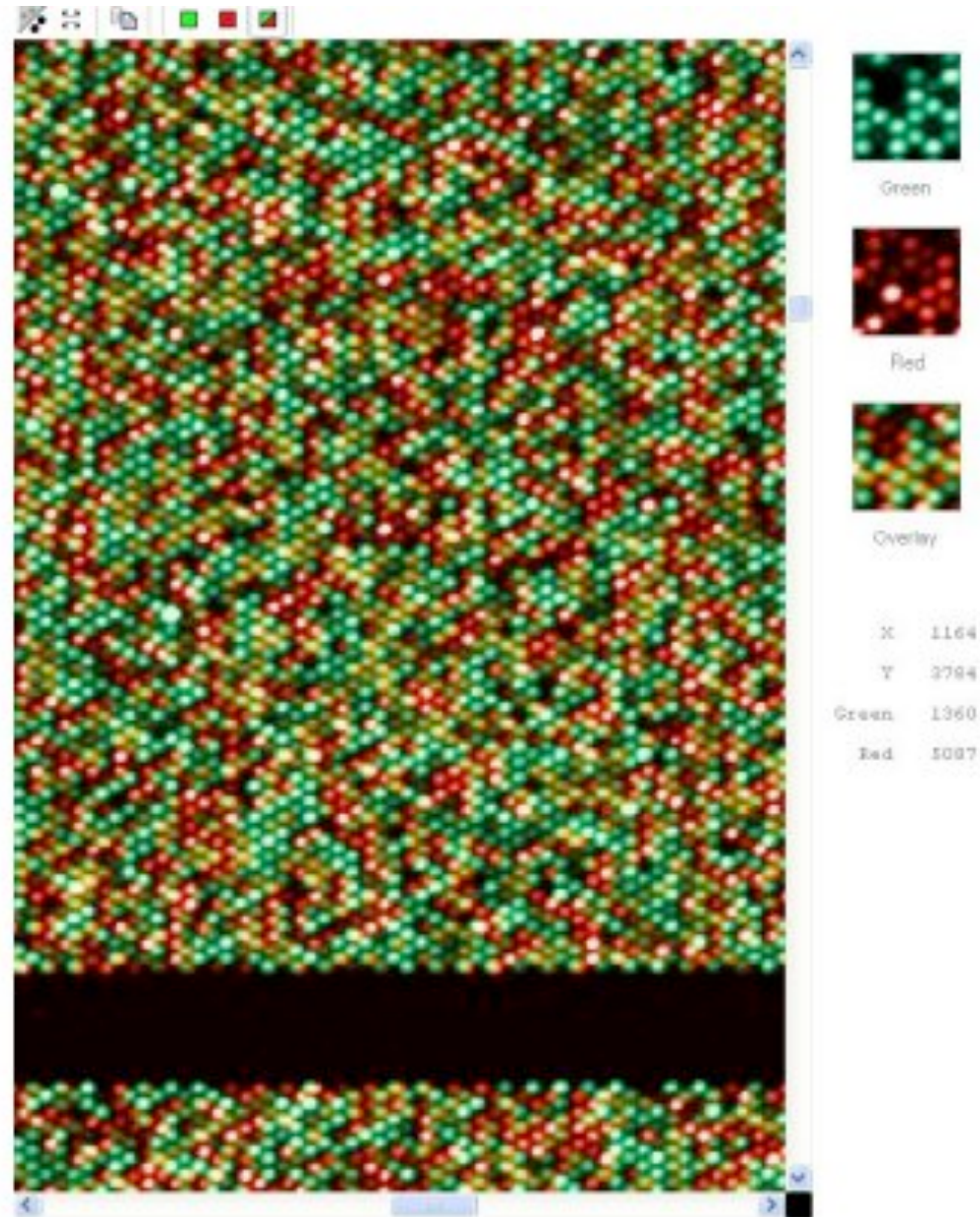
*Drew Sheneman, New Jersey -- The Newark Star Ledger*

- Les marqueurs génotypiques sont des petits fragments de génome qui présentent des variations interindividuelles.
  - **Micro-satellite**: fragment de séquence génomique comportant une courte séquence (2 à 5 résidus) répétée (typiquement 5 à 50 fois selon les cas). Le nombre de répétitions varie souvent d'un individu à l'autre.
    - Exemple:
      - Type 1      CAGCAGCAGCAGCAGCAG
      - Type 2      CAGCAGCAGCAGCAGCAGCAG
      - Type 3      CAGCAGCAGCAGCAGCAGCAGCAG
    - SNP: nucléotide variable d'un individu à l'autre. Ce type de variation est appelée « **SNP** » pour « **Single-Nucleotide Polymorphism** » (prononcez « *snip* »).
      - Exemple
        - Type 1      CAGTGCGAGGG
        - Type 2      CAGTGCGGGGG
- Certains marqueurs peuvent être associés à des maladies, ou à des caractères phénotypiques.
- D'autres marqueurs n'ont aucun effet détectable, mais sont utilisables pour identifier des personnes.



# Le génotypage à haut débit

- Depuis le début des années 2000, on a développé de nouvelles méthodes de génotypage à haut débit, basées sur la technologie des biopuces.
- Chaque point lumineux indique le type d'un SNP particulier.
- Sur une seule biopuce ("microarray"), on peut caractériser entre 500.000 (il y a ~10 ans) et 6.000.000 (2014) de SNP différents.
- Ceci permet de déterminer un nombre considérable de caractéristiques génétiques d'un individu.



<http://www.sanger.ac.uk/Teams/Team67/>

# Etudes d'association de type « cas - contrôle »

## ■ Principe des études d'association

- Etudes « cas-contrôles » : on caractérise le profil génétique d'un groupe patients souffrant d'une maladie donnée (les « cas »), et d'un groupe de personnes saines (les « contrôles »).
- On analyse ensuite chaque marqueur, en comparant la fréquence des différents génotypes dans les deux groupes.
- Ceci permet de délimiter des régions génomiques associées à la maladie.

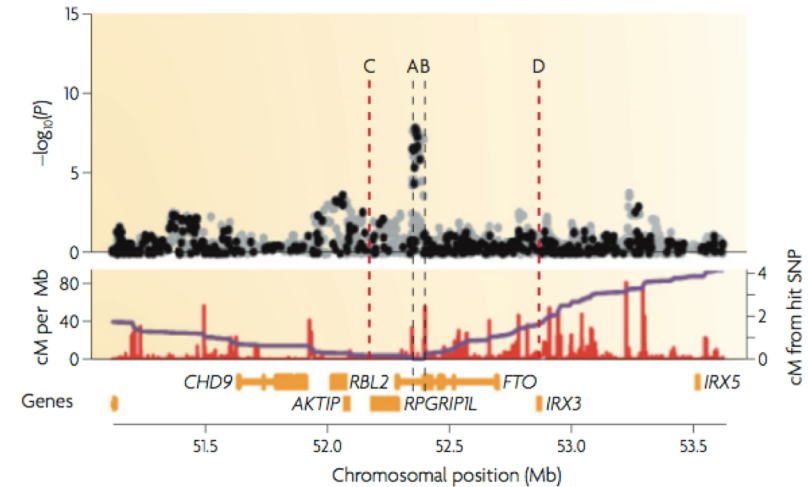
## ■ Attention, « corrélation n'est pas raison »

- Les études d'associations révèlent uniquement la corrélation entre un trait génotypique et une maladie, mais ne constituent pas un élément suffisant pour établir une relation de causalité

## ■ L'association peut provenir d'effets de « stratification »

- Si la maladie est plus fréquente dans une sous-population donnée, et que cette sous-population est caractérisée par un marqueur, on observera une corrélation fortuite.
- On risque donc de confondre les traits génétiques associés aux patients et ceux de la sous-population.
- Il existe des méthodes statistiques pour éviter ce piège de la stratification.

Génotype	aa	aA	AA	Total
Cas (patients)	X1	X2	X3	X
Contrôle	Y1	Y2	Y3	Y
Total	T1	T2	T3	T



McCarthy et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet (2008) vol. 9 (5) pp. 356-69

L'axe horizontal représente un fragment de chromosome. Chaque point noir ou gris représente un marqueur génétique (un site polymorphique), et la hauteur du point reflète la significativité de l'association entre ce site chromosomique et un trait phénotypique particulier.

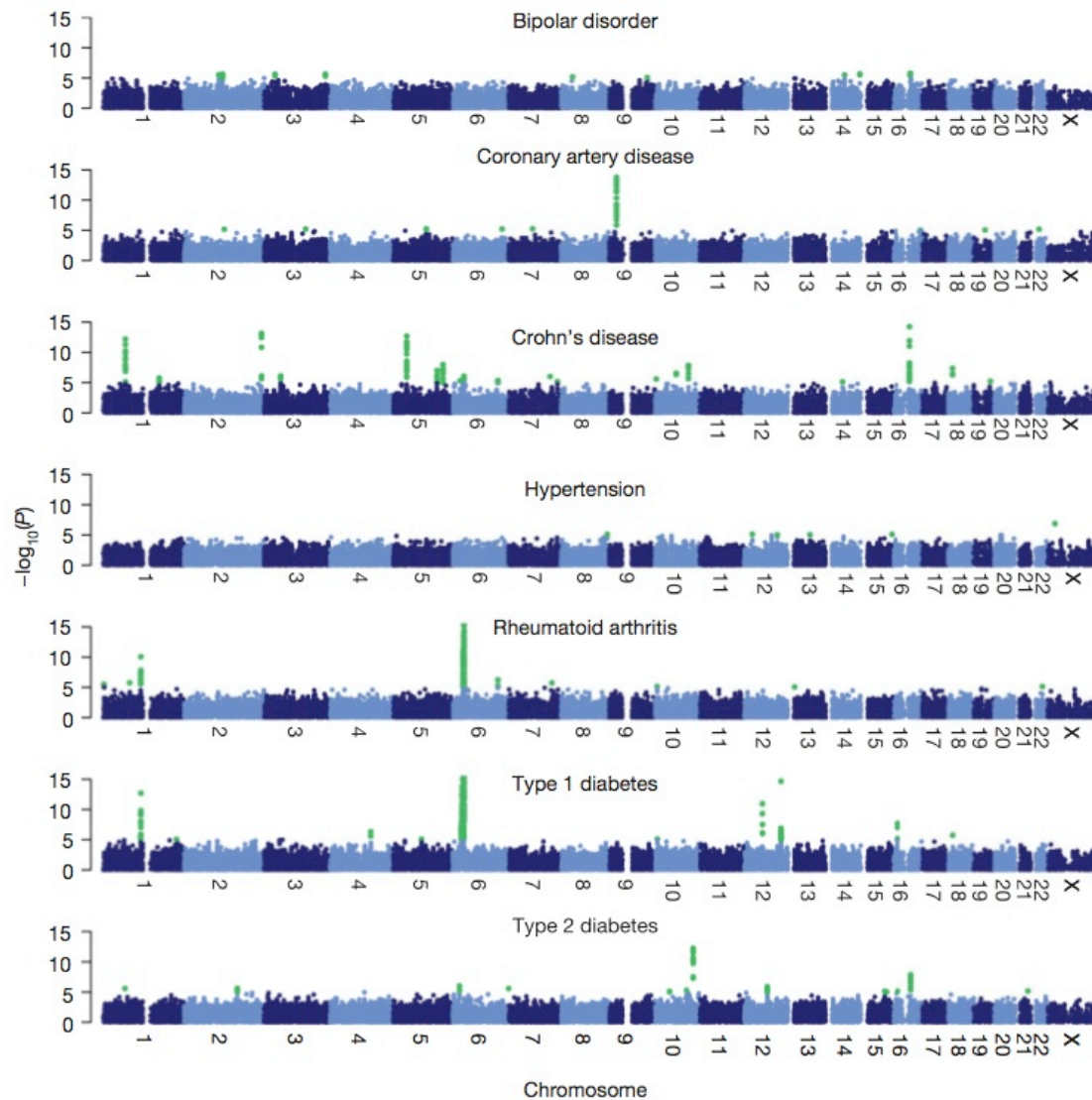
Sur cette figure, on observe une forte association entre le trait phénotypique et les marqueurs génétiques situés dans la région A-B.

- Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-78.
- McCarthy et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet (2008) vol. 9 (5) pp. 356-69.



# Etudes d'associations à l'échelle du génome complet

- Une étude a été menée sur 17.000 personnes afin d'identifier les régions génomiques associées à 7 maladies (2.000 patients par maladie) par rapport à un groupe de contrôle (3.000 personnes).
- La figure synthétise les résultats, en indiquant (en vert) les SNPs associés de façon significative à l'une des maladies.
  - Les zones bleues représentent les chromosomes.
  - Chaque point vert représente un SNP, et sa hauteur indique la significativité.



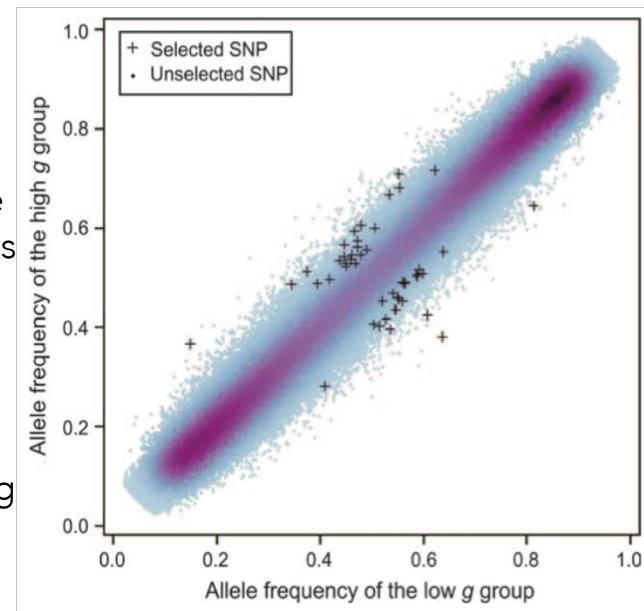
**Figure 4 | Genome-wide scan for seven diseases.** For each of seven diseases  $-\log_{10}$  of the trend test  $P$  value for quality-control-positive SNPs, excluding those in each disease that were excluded for having poor clustering after visual inspection, are plotted against position on each chromosome.

Chromosomes are shown in alternating colours for clarity, with  $P$  values  $< 1 \times 10^{-5}$  highlighted in green. All panels are truncated at  $-\log_{10}(P \text{ value}) = 15$ , although some markers (for example, in the MHC in T1D and RA) exceed this significance threshold.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-78.

# Genetics and cognitive ability

- Depuis plusieurs décennies, des chercheurs ont tenté de détecter les gènes associés à des capacités cognitives spécifiques, ou générales (« intelligence »).
- En 2008, l'équipe de Richard Plomin a testé les capacités générales de 7000 enfants en leur faisant passer des tests cognitifs, dont les résultats sont synthétisés au moyen d'un score unique (le facteur « g »).
- Ils ont sélectionné les 1500 enfants ayant le meilleur score et les 1500 ayant le moins bon score, et caractérisé leur génotype au moyen de biopuces à 500.000 SNPs.
- Aucun des SNPs n'a montré d'association significative avec le facteur g
- Les 47 SNPs les plus significatifs (marqués par des croix vertes) ne passent pas le seuil qu'il faut appliquer quand on teste 500.000 hypothèses simultanées (correction de tests multiples).
- Remarques
  - Le facteur g est un outil rudimentaire pour mesurer l'intelligence. Lire *La malmesure de l'homme* (S.J. Gould, 1981) pour une critique du facteur "g".
  - En dépit des résultats négatifs de leur étude, les auteurs sont convaincus qu'une étude plus approfondie (avec plus de SNPs et des cohortes plus grandes) permettrait de prédire les risques génétiques et d'investiguer les effets fonctionnels des gènes sur le cerveau et le comportement.
  - Ils n'envisagent à aucun moment que les capacités cognitives pourraient être indépendantes des gènes.



Butcher et al. (2008), Figure 2: A scatter plot showing the 47 top-ranked SNPs (crosses) against the background of unselected SNPs comparing allele frequencies for the low g group (x-axis) and the high g group (y-axis).

- Butcher et al. Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav* (2008) vol. 7 (4) pp. 435-46.
- Gould, S. J. (1997). *La mal-mesure de l'homme*. Editions Odile Jacob

- Agrégation de données pour 5 troubles psychiatriques.
- 32.332 cas + 27.888 contrôles.
- Identification de plusieurs loci (positions génomiques) très significatives, qui étaient faiblement significatives dans chacune des maladies étudiées séparément.
- Certains gènes étaient préalablement ressortis séparément des études de schizophrénie et trouble bipolaire.

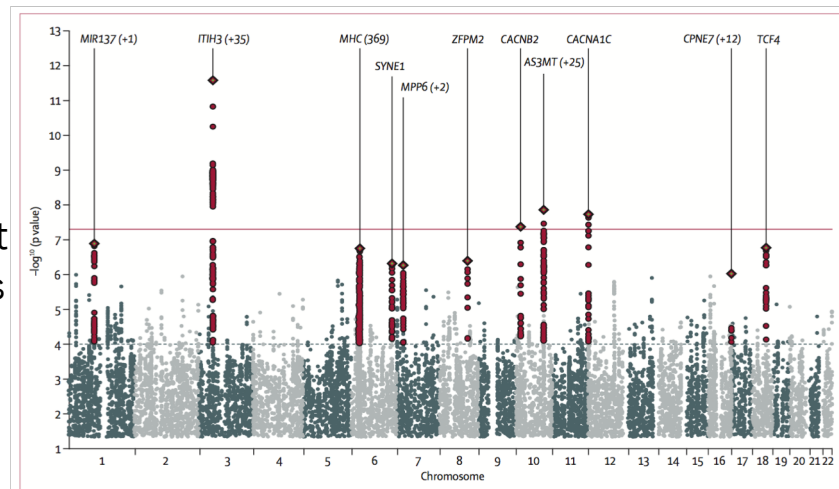


Figure 1: Manhattan plot of primary fixed-effects meta-analysis  
Horizontal line shows threshold for genome-wide significance ( $p < 5 \times 10^{-8}$ ).

	Chromosome	Base-pair position*	Nearest gene	Alleles	Frequency†	Imputation quality score (INFO)	p value	OR (95% CI)‡	Heterogeneity p value	Best-fit model (BIC)§
rs2535629	3	52808259	ITIH3 (+ many)	G/A	0.651	0.942	$2.54 \times 10^{-12}$	1.10 (1.07-1.12)	0.27	Five disorder¶
rs11191454	10	104649994	AS3MT (+ many)	A/G	0.910	1.01	$1.39 \times 10^{-8}$	1.13 (1.08-1.18)	0.32	Five disorder¶
rs1024582	12	2272507	CACNA1C	A/G	0.337	0.98	$1.87 \times 10^{-8}$	1.07 (1.05-1.10)	0.0057	BPD, schizophrenia
rs2799573	10	18641934	CACNB2	T/C	0.715	0.825	$4.29 \times 10^{-8}$	1.08 (1.05-1.12)	0.57	Five disorder¶

Most strongly associated single-nucleotide polymorphisms (SNP) in associated region after clumping—i.e., grouping SNPs within 250 kb of the index SNP that have  $r^2 > 0.2$  with the index SNP as implemented in PLINK. OR=odds ratio. BIC=Bayesian information criteria. BPD=bipolar disorder. \*Detected with University of California Santa Cruz Genome Browser (version hg18). †Risk allele frequency in controls. ‡Estimated OR from multinomial logistic regression used in the modelling analysis. §Best-fit multinomial logistic model by BIC criteria; appendix pp 38-45 provide a comparison of BIC and Akaike information criteria across models. ¶Best-fit model supports an effect on all five disorders.

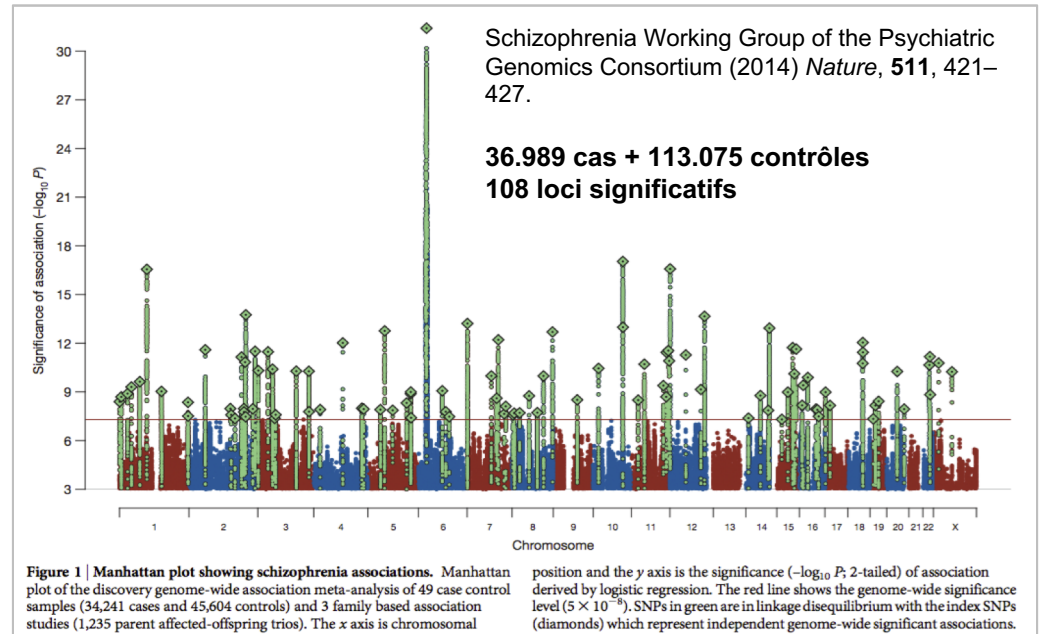
Table 1: Five disorder meta-analysis results for regions with  $p < 5 \times 10^{-8}$

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). [Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.](#) Lancet. 2013 Apr 20;381(9875):1371-9. Erratum in: Lancet. 2013 Apr 20;381(9875):1360.

- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). [Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis.](#) Lancet. 2013 Apr 20;381(9875):1371-9. Erratum in: Lancet. 2013 Apr 20;381(9875):1360.

# Une étude à très large échelle pour la schizophrénie

- 2014: une étude génomique d'association basée sur 37.000 cas de schizophrénie et 113.000 contrôles révèle 108 loci fortement significatifs.
  - 25 avaient déjà été détectés dans des études précédentes, mais à des niveaux de significativité limite, et variant selon les cohortes.
  - 83 nouveaux loci
- Quelques candidats notables
  - NOTCH4. Récepteur transmembranaire impliqué dans des voies de signalisation (communications entre cellules voisines par liaison d'une protéine-signal au récepteur).
  - Protéine à doigt de Zinc (ZNF804A), facteur transcriptionnel exprimé dans le cortex, hippocampe et cervelet, fonction méconnue.
  - CACNA1C, CACNB2, CACNA11: famille de protéines canaux à ions calcium



Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.

***Etudes d'association  
à l'échelle de l'épigénome***

# Impact de la tabagie sur l'épigénome

- Les études d'association à échelle génomique (GWAS) n'expliquent qu'une partie des caractères héritables.
- On peut maintenant mener des études au niveau des épigénomes (EWAS, epigenome-wide association studies)
- Des études d'association entre le tabagisme pendant la grossesse et la méthylation des CpG chez le nouveau-né révèlent plusieurs gènes dont la méthylation est affectée.

## Rotroff et al (2016)

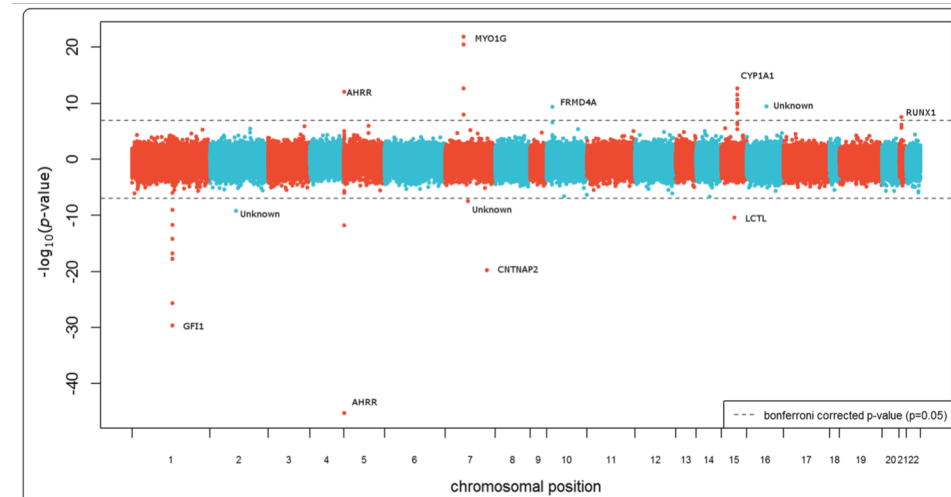
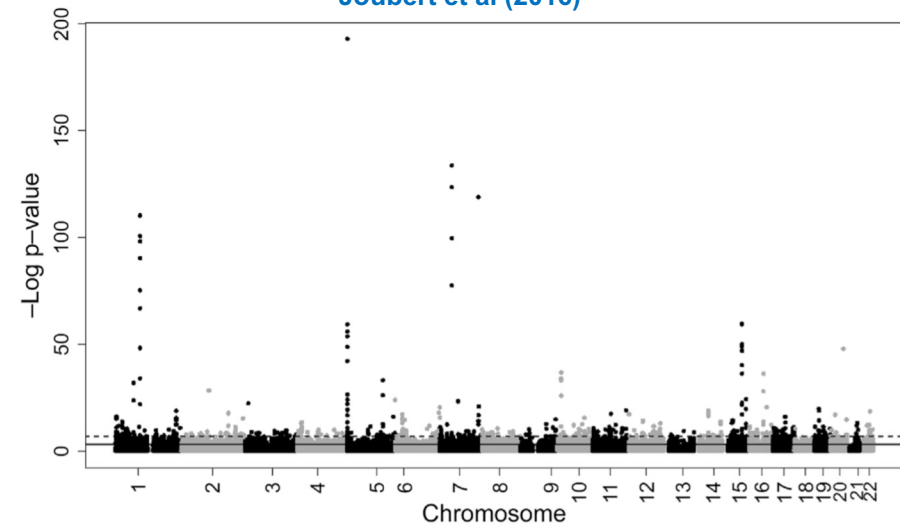


Fig. 2 Manhattan plot of univariate CpG results. The y-axis represents the  $-\log_{10}$  of the CpG  $p$ -values. CpGs with negative  $p$ -values corresponded to decreased methylation, whereas positive  $p$ -values corresponded to increased methylation. CpGs that reached genome-wide significance, with a bonferroni corrected  $p < 0.05$  are annotated with their corresponding genes

- L'interprétation requiert d'être prudents.
  - dans les études GWAS, on postule que les variations préexistent
  - les marques épigénomiques sont à la fois causes de modifications phénotypiques, et conséquences des facteurs de l'environnement

## Joubert et al (2016)



- Rotroff, D.M., Joubert, B.R., Marvel, S.W., Håberg, S.E., Wu, M.C., Nilsen, R.M., et al. (2016). Maternal smoking impacts key biological pathways in newborns through epigenetic modification in Utero. *BMC Genomics* 17, 976.
- Joubert, B.R., Felix, J.F., Yousefi, P., Bakulski, K.M., Just, A.C., Breton, C., Reese, S.E., Markunas, C.A., Richmond, R.C., Xu, C.-J., et al. (2016). DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* 98, 680–696.