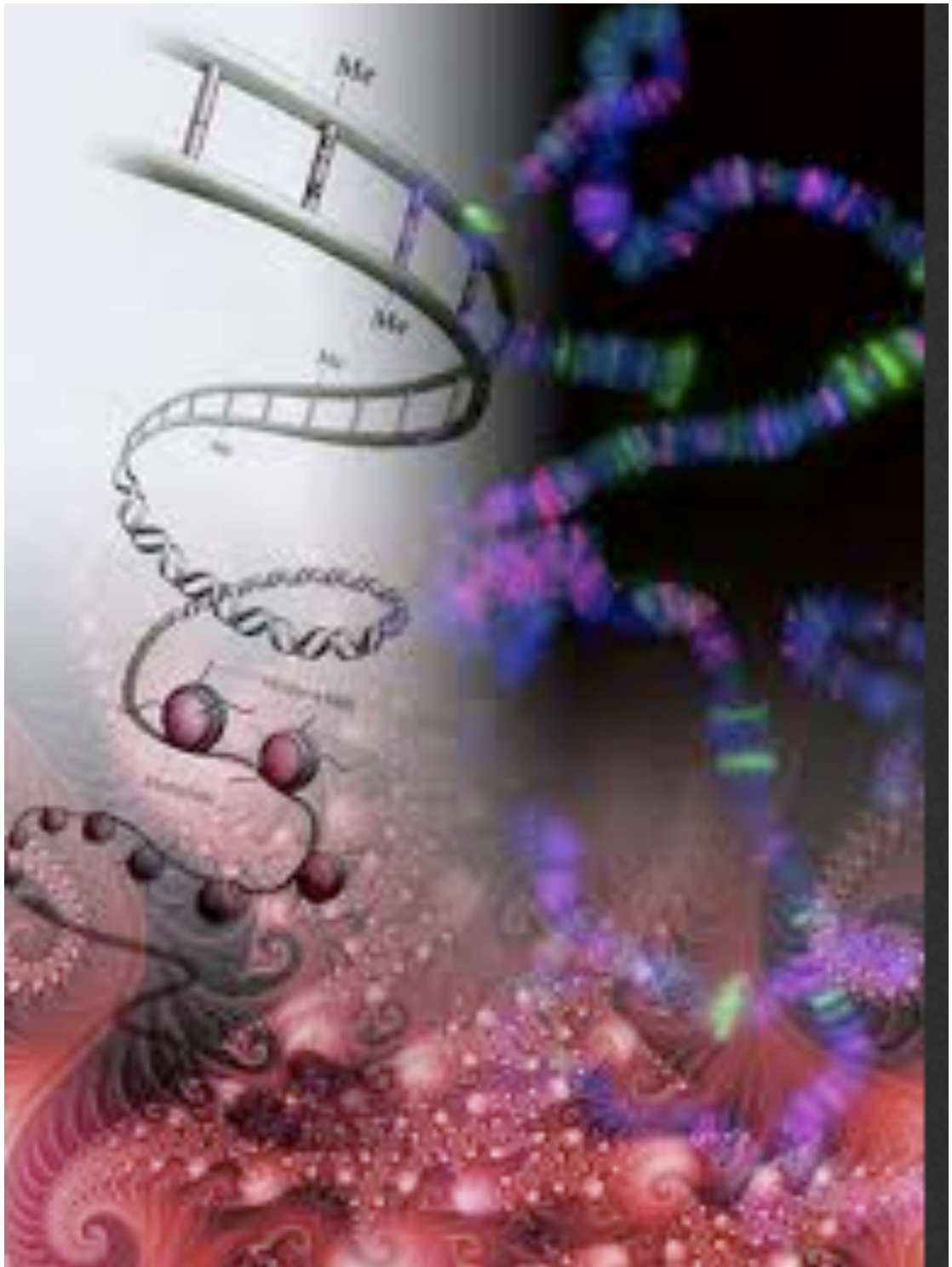
L'Epigénétique Part II

Chrystelle Carlo-Toumaniantz

l'Institut du thorax

2013



l'Institut du thorax

L'épigénétique

I – Introduction

II – Historique

III – Définition et rappels

IV – Modifications de l'ADN : la méthylation

V – Modifications de la structure de la chromatine/des protéines

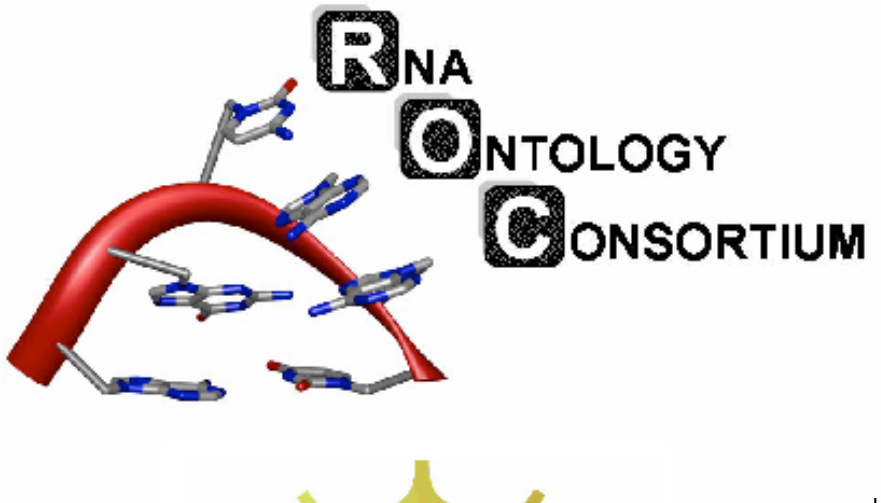
- **Modifications des histones/réorganisation du nucléosome**
- **L'acétylation des histones**
- **La méthylation des histones**
- **Autres modifications**

VI – Les ARN non-codants

Small, Non-coding RNAs

Cf cours de Gilles Toumaniantz

miRNA (microRNA)
snRNA (Small nuclear RNA)
sno (Small nucleolar RNA)
scRNA (Small cytoplasmic RNA)
siRNA (small interfering RNA)
stRNA (small temporal RNA)
tncRNA (tiny non-coding RNA)
smRNA (small modulatory RNA)
tasiRNA (trans-acting siRNA)
rasiRNA (repeat-associated siRNA)
scnRNA (small scan RNA)
Pi-RNA (Piwi-interacting RNA)
+Autres RNAs (télomerase-RNA, antisens, non-codant etc….)



Small, Non-coding RNAs

RNAs involved in protein synthesis

Type	Abbr.	Function	Distribution
Messenger RNA	mRNA	Codes for protein	All organisms
Ribosomal RNA	rRNA	Translation	All organisms
Signal recognition particle RNA	7SL RNA or SRP RNA	Membrane integration	All organisms
Transfer RNA	tRNA	Translation	All organisms
Transfer-messenger RNA	tmRNA	Rescuing stalled ribosomes	Bacteria

Small, Non-coding RNAs

RNAs involved in post-transcriptional modification or DNA replication

Type	Abbr.	Function	Distribution
Small nuclear RNA	snRNA	Splicing and other functions	Eukaryotes and archaea
Small nucleolar RNA	snoRNA	Nucleotide modification of RNAs	Eukaryotes and archaea
SmY RNA	SmY	mRNA trans-splicing	Nematodes
Small Cajal body-specific RNA	scaRNA	Type of snoRNA; Nucleotide modification of RNAs	
Guide RNA	gRNA	mRNA nucleotide modification	Kinetoplastid mitochondria
Ribonuclease P	RNase P	tRNA maturation	All organisms
Ribonuclease MRP	RNase MRP	rRNA maturation, DNA replication	Eukaryotes
Y RNA		RNA processing, DNA replication	Animals
Telomerase RNA		Telomere synthesis	Most eukaryotes

Small, Non-coding RNAs

Regulatory RNAs

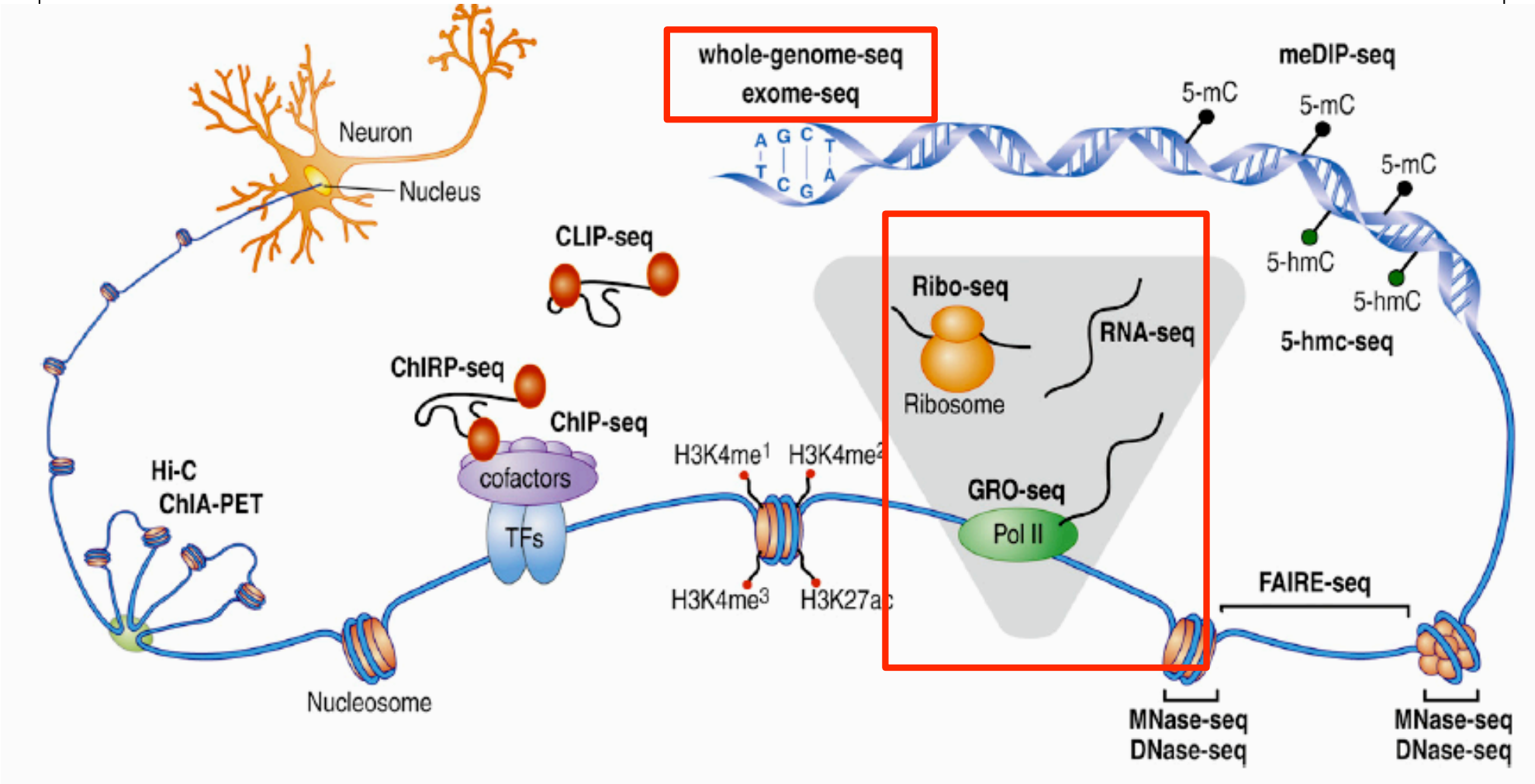
Type	Abbr.	Function	Distribution
Antisense RNA	aRNA	Transcriptional attenuation / mRNA degradation / mRNA stabilisation / Translation block	All organisms
Cis-natural antisense transcript		Gene regulation	
CRISPR RNA	crRNA	Resistance to parasites, probably by targeting their DNA	Bacteria and archaea
Long noncoding RNA	Long ncRNA	Various	Eukaryotes
MicroRNA	miRNA	Gene regulation	Most eukaryotes
Piwi-interacting RNA	piRNA	Transposon defense, maybe other functions	Most animals
Small interfering RNA	siRNA	Gene regulation	Most eukaryotes
Trans-acting siRNA	tasiRNA	Gene regulation	Land plants
Repeat associated siRNA	rasiRNA	Type of piRNA; transposon defense	Drosophila
7SK RNA	7SK	negatively regulating CDK9/cyclin T complex	

l'Institut du thorax

L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

Analyses épigénétiques



l'Institut du thorax

L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

1 – identification des bases, séquençage NGS

Approches technologiques à l'étude du génome et transcriptome

Assays profiling the genome	Whole-genome-sequence Exome-sequence CGH arrays	Cf cours de génomique humaine
Assays profiling the transcriptome	RNA-seq and miR-seq	RNA-seq profiles transcripts genome-wide in their steady-state form. Different information can be obtained based on the specific pool of RNAs utilized, such as nuclear or cytoplasmic, or based on specific strategies of enrichment, such as polyadenylated RNAs or small RNAs (i.e., miR-seq).
	Ribo-seq (ribosome profiling)	Ribo-seq profiles transcripts undergoing translation, based on sequencing ribosome-associated RNAs. Tagging strategies of ribosomal proteins allow their affinity purification.
	GRO-seq (global run-on)	GRO-seq profiles nascent transcripts exclusively undergoing transcription. It is based on the rapid isolation of nuclei and subsequent addition of biotinylated nucleotides during a short period in which transcription is allowed to shortly proceed in vitro, effectively mapping the position, amount, and orientation of transcriptionally engaged RNA polymerases genome-wide.
	Pangenomic Microarrays	Cf cours de génomique humaine

l'Institut du thorax

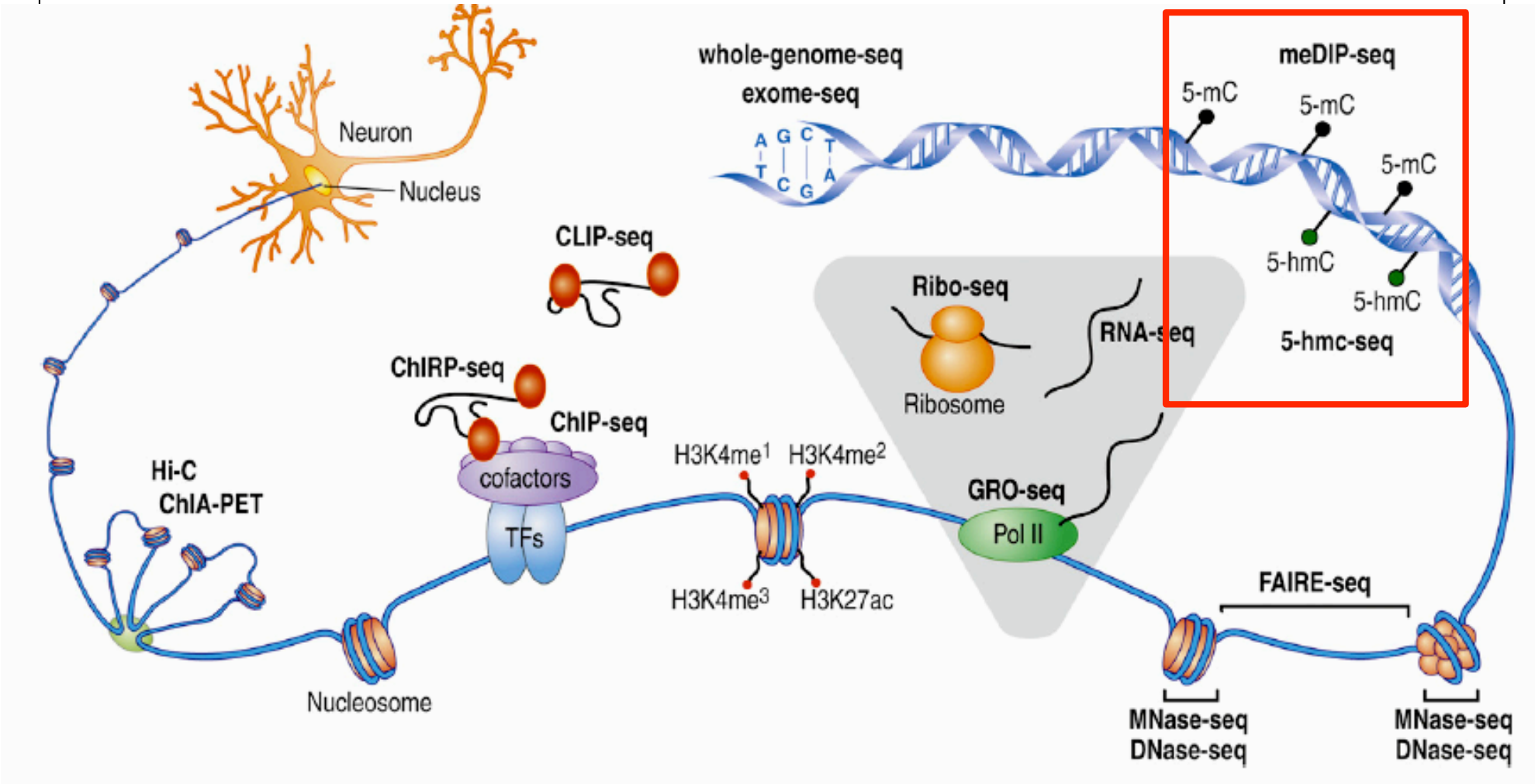
L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

- **1 – identification des bases, séquençage NGS**
- **2 – profil de méthylation de l'ADN**

l'Institut du thorax

Analyses épigénétiques



Inserm

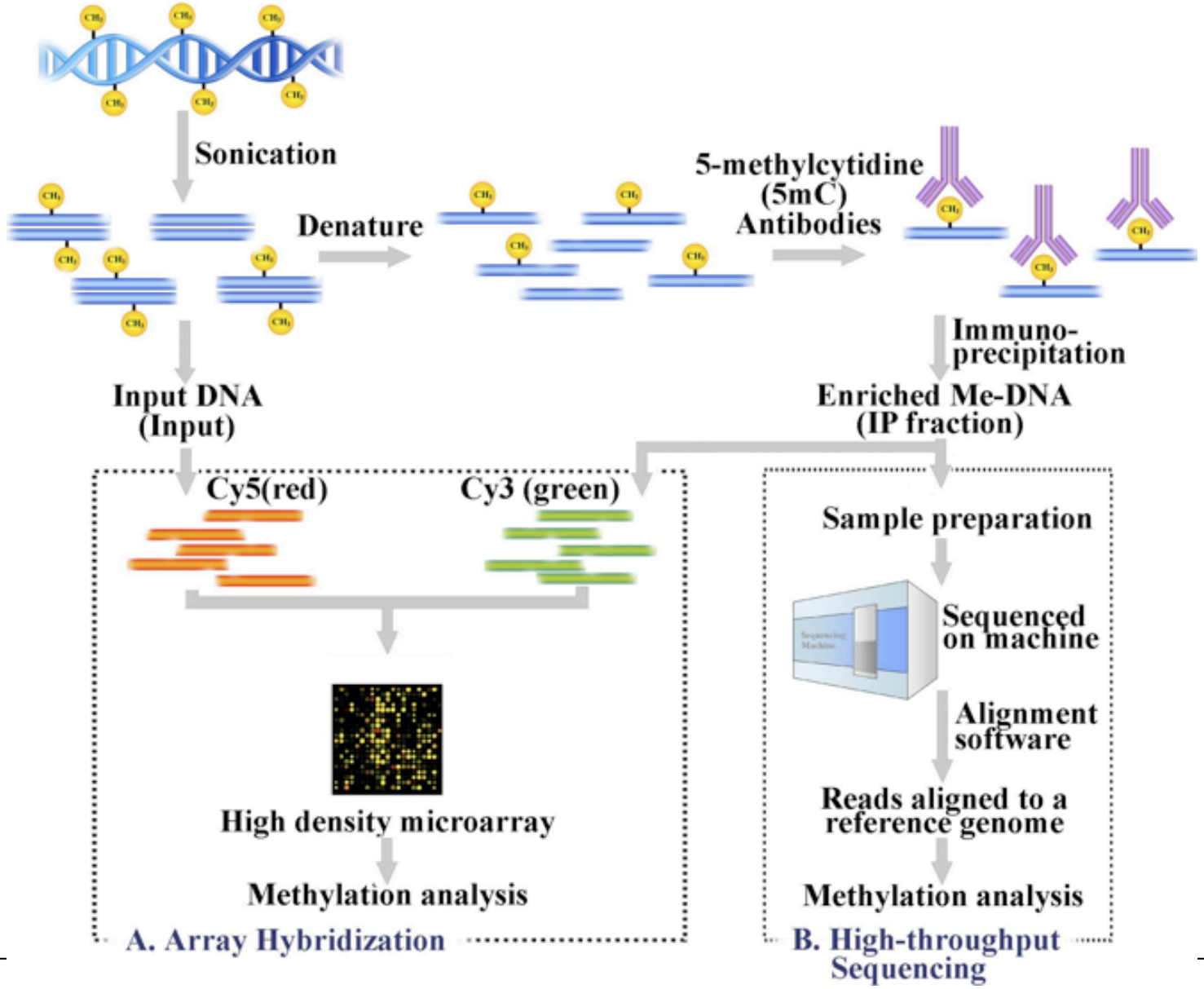
l'Institut du thorax

Profil de methylation de l'ADN

Affinity enrichment-based assays	MeDIP-seq (methylated DNA immunoprecipitation), MBD-seq	Based on immunoisolation of methylated DNA fragments using an antibody that recognizes 5-methylcytosine (MeDIP-seq) or using the methyl binding domains of MBD2 or MeCP2 (MBD-seq).
Enzyme digestion-based assays	MRE-seq (methylation-sensitive restriction snzyme)	Based on cleavage of DNA by methylation-sensitive restriction enzymes to fragment DNA at methylation sites before sequencing.
Bisulfite conversion-based assays	Bis-seq (bisulfite-sequencing), reduced representation Bis-Seq (RRBS)	Based on the chemical conversion of unmethylated cytosine into uracil. Identifies methylated cytosines at nucleotide resolution by comparison to untreated DNA samples or reference genome. Bisulfate-based approaches can be combined with pre-enrichment methods, such as array capture or bead technology, which can enrich specific DNA genomic regions for sequencing.
	OxBS-seq (oxidative bisulfite), TAB-seq (Tet-assisted bisulfite), and 5-hmCyt-seq (5hydroxy-methylcytosine)	Variants of Bis-seq that specifically discriminate between 5-mC and 5-hmC

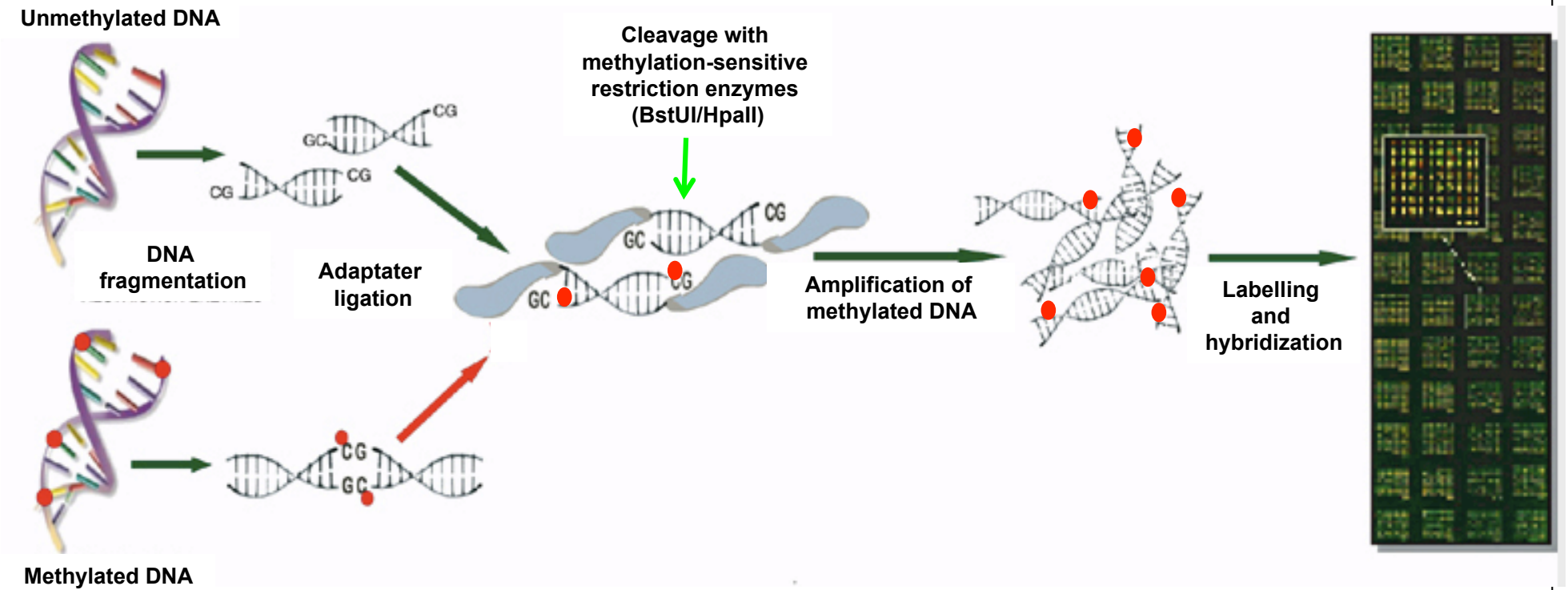
l'Institut du thorax

MeDIP-array et MeDIP-Seq = Methylated-DNA immunoprecipitation



l'Institut du thorax

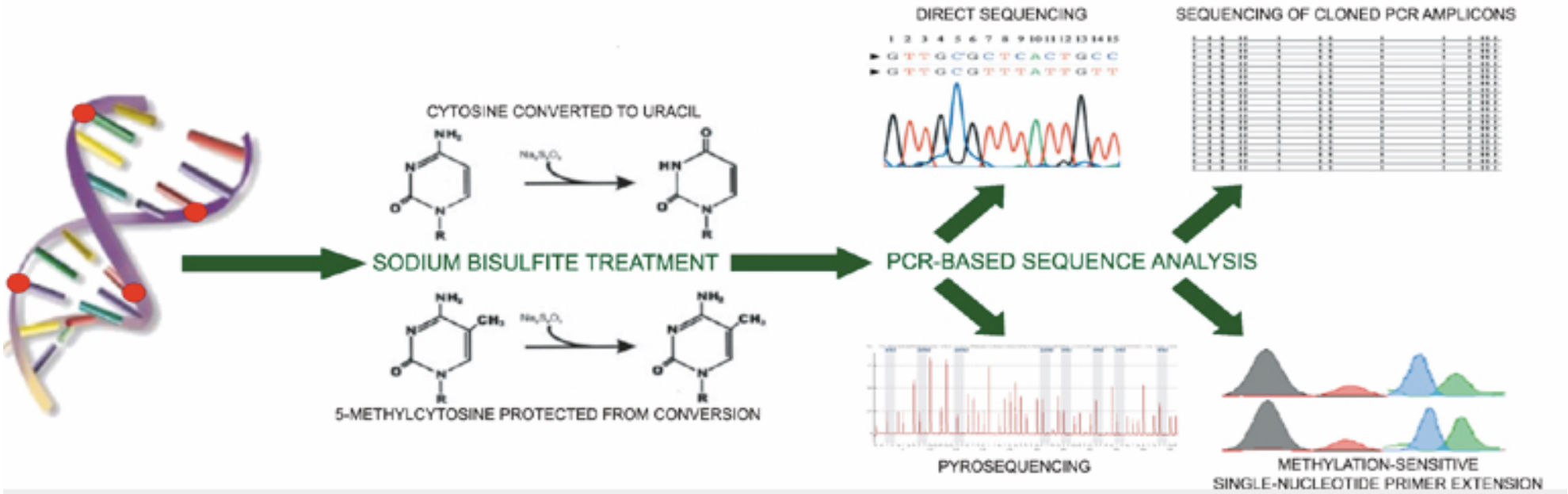
MeRE-array et MeRE-Seq = Methylated-Restricted Enzyme-array et Seq



Inserm

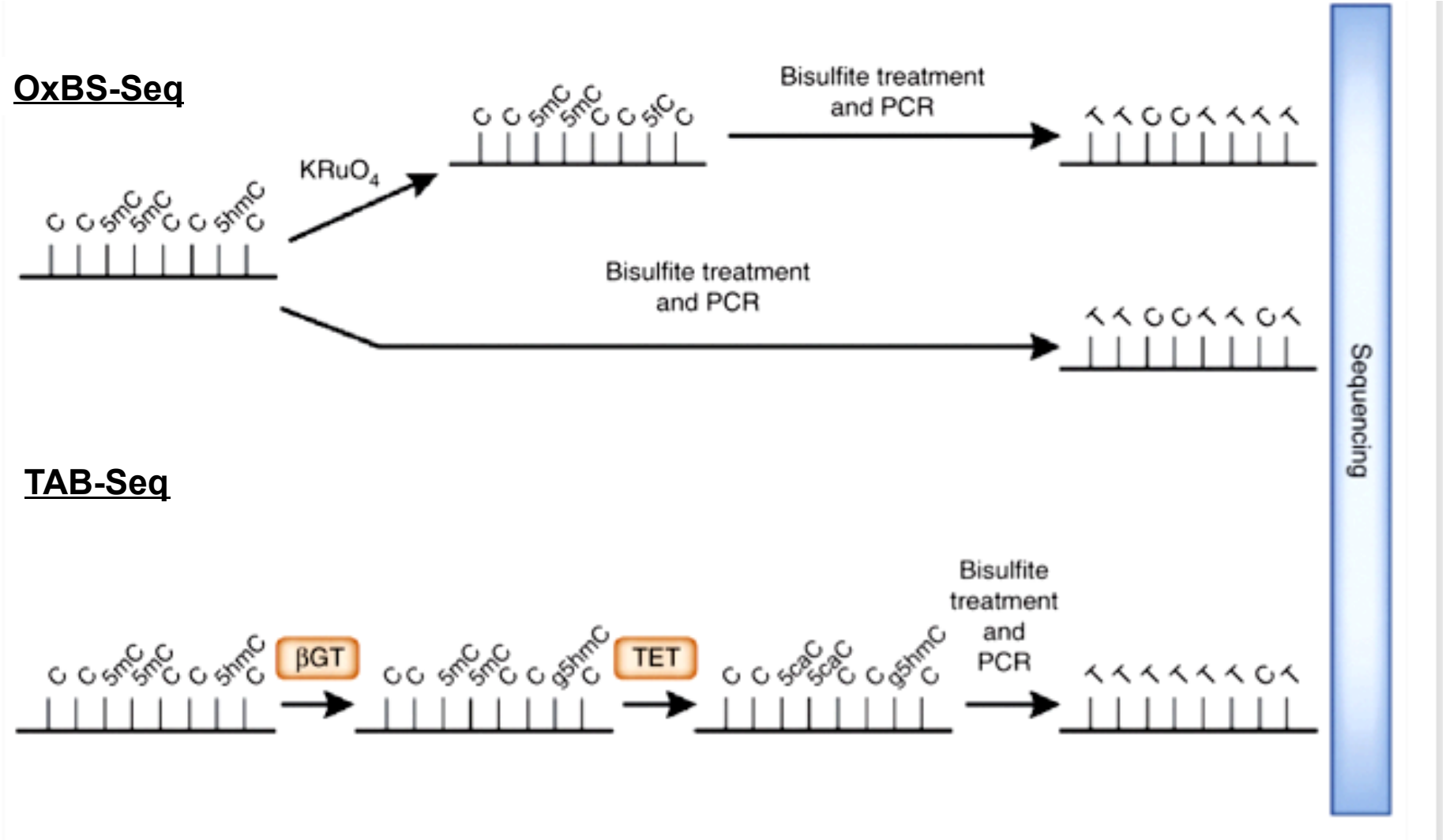
l'Institut du thorax

Bisulfite conversion-based analysis



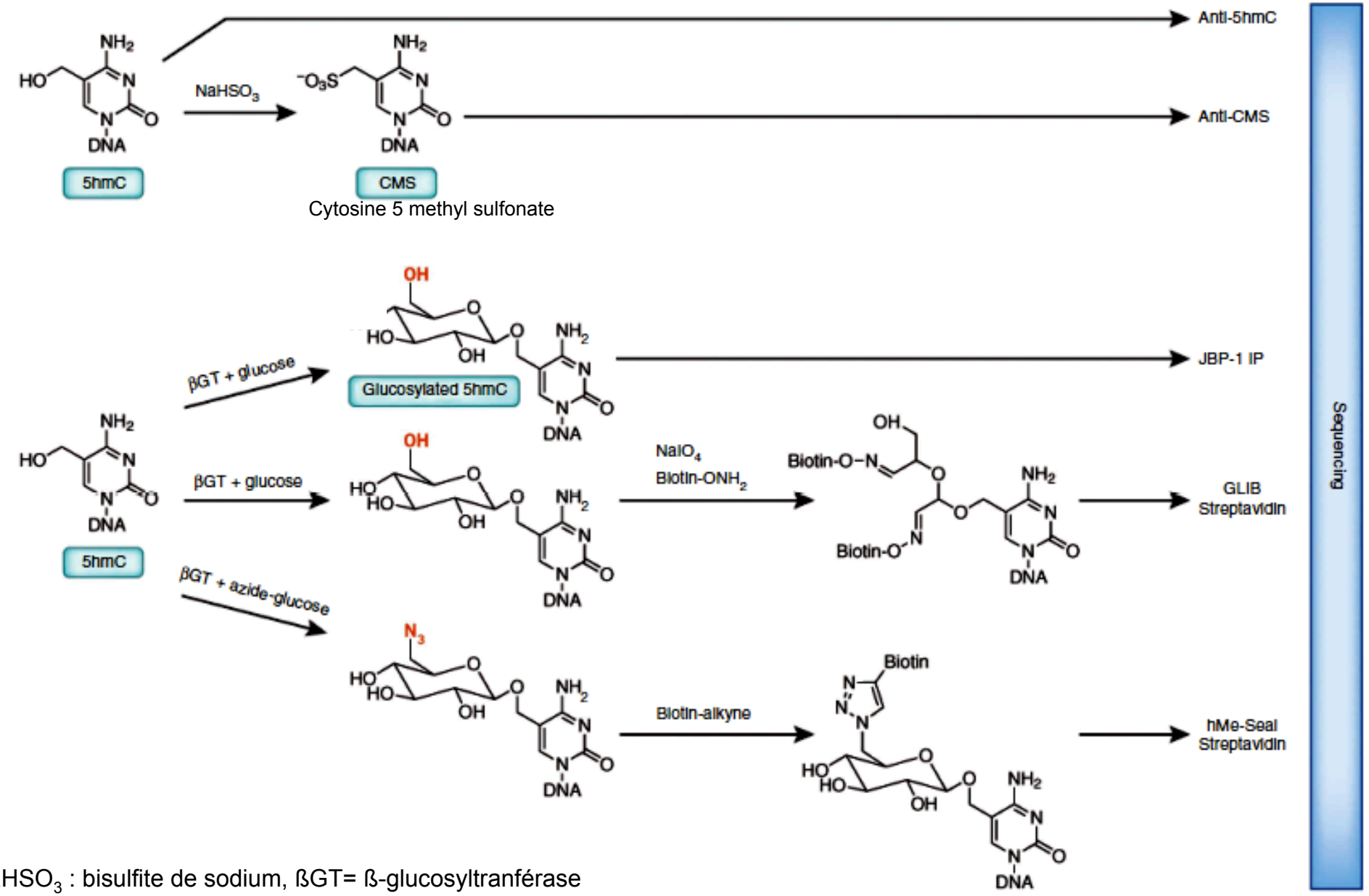
Inserm

OxBS-Seq (oxydative bisulfite) et TAB-Seq (Tet-assisted bisulfite)



l'Institut du thorax

5-hmC-Seq (5-hydroxyméthylCytosine –Seq)



$NaHSO_3$: bisulfite de sodium, ßGT= ß-glucosyltranférase
JBP-1 = J-binding protein 1, $NaIO_4$ = sodium periodate, GLIB = glucosylation, periodate oxidation and biotinylation

l'Institut du thorax

Analyses des modifications épigénétiques : softwares (1)

Software	Description	URL
Processing bisulphite-sequencing data		
B-SOLANA	Bisulphite aligner for processing bisulphite-sequencing data obtained in the two-base encoding of ABI SOLiD sequencers	http://code.google.com/p/bsolana
Bismark	Probably the most widely used three-letter bisulphite aligner; supports both Bowtie (fast, gap-free alignment) and Bowtie 2.0 (sensitive, gapped alignment)	http://www.bioinformatics.bbsrc.ac.uk/projects/bismark
Bis-SNP	Variant caller for inferring DNA methylation levels and genomic variants from bisulphite-sequencing reads that have been aligned by other tools	http://epigenome.usc.edu/publicationdata/bissnp2011
BRAT	Highly configurable and well-documented three-letter bisulphite aligner	http://compbio.cs.ucr.edu/brat
BS-Seeker	Basic three-letter bisulphite aligner based on Bowtie	http://pellegrini.mcdb.ucla.edu/BS_Seeker/BS_Seeker.html
BSMAP	Probably the most widely used wild-card bisulphite aligner	http://code.google.com/p/bsmap
GSNAP	Wild-card bisulphite aligner included in a widely used general-purpose alignment tool	http://share.gene.com/gmap
Last	Recent and well-validated wild-card bisulphite aligner included in a general-purpose alignment tool	http://last.cbrc.jp
MethylCoder	Three-letter bisulphite aligner that can be used with either Bowtie (high speed) or GSNAP (high sensitivity)	https://github.com/brentp/methylcoder
Pash	Wild-card bisulphite aligner included in a general-purpose alignment tool	http://brl.bcm.tmc.edu/pash
RMAP	Wild-card bisulphite aligner included in a general-purpose alignment tool	http://www.cmb.usc.edu/people/andrewds/rmap

l'Institut du thorax

Analyses des modifications épigénétiques : softwares (2)

Software	Description	URL
RRBSMAP	Variant of BSMAP that is specialized on reduced-representation bisulphite sequencing (RRBS) data	http://rrbsmap.computational-epigenetics.org
segemehl	Wild-card bisulphite aligner included in a general-purpose alignment tool	http://www.bioinf.uni-leipzig.de/Software/segemehl
Processing bisulphite microarray data		
ComBat	R script for correcting known or suspected batch effects using an empirical Bayes method	http://www.bu.edu/jlab/wp-assets/ComBat
Illumina BeadScan	Machine control and image processing software for Illumina Infinium microarray scanners	http://www.illumina.com/support/array/array_instruments/beadarray_reader.ilmn
Illumina GenomeStudio	Graphical tool for data normalization, analysis and visualization of Illumina Infinium microarrays (and other genomic data types)	http://www.illumina.com/software/genomestudio_software.ilmn
isva	R package for batch effect correction using an algorithm that is based on singular value decomposition	http://cran.r-project.org/web/packages/isva
methylumi	R/Bioconductor package for Infinium data normalization and general data handling	http://www.bioconductor.org/packages/release/bioc/html/methylumi.html
minfi	R/Bioconductor package for Infinium data normalization, analysis and visualization	http://www.bioconductor.org/packages/release/bioc/html/minfi.html
RnBeads	R package providing a software pipeline for Infinium data normalization, quality control, exploratory visualization and differentially methylated region (DMR) identification	http://rnbeads.computational-epigenetics.org
SVA	R/Bioconductor package for correcting batch effects that are directly inferred from the data using surrogate variable estimation	http://www.bioconductor.org/packages/release/bioc/html/sva.html

Inserm

l'Institut du thorax

Analyses des modifications épigénétiques : softwares (3)

Software	Description	URL
Processing enrichment-based data		
BATMAN	Command-line tool for methylated DNA immunoprecipitation (MeDIP) data normalization	http://td-blede.gurdon.cam.ac.uk/software/batman
Bowtie	General-purpose aligner based on the Burrows–Wheeler transform	http://bowtie-bio.sourceforge.net
BWA	General-purpose aligner based on the Burrows–Wheeler transform	http://bio-bwa.sourceforge.net
MEDIPS	User-friendly R package for MeDIP data normalization	http://medips.molgen.mpg.de
MEDME	R package for MeDIP data normalization	http://espresso.med.yale.edu/medme
MeDUSA	Command-line software pipeline for MeDIP read alignment, data normalization, quality control and DMR identification	http://www2.cancer.ucl.ac.uk/medicalgenomics/meduse
MetMap	Command-line tool for normalization of DNA methylation data obtained using restriction enzymes that specifically cut unmethylated DNA	http://www.cs.berkeley.edu/~meromit/MetMap.html
MeQA	Command-line software pipeline for MeDIP read alignment, data normalization and quality control	http://life.tongji.edu.cn/meqa
Repitools	R/Bioconductor package for quality control and visualization of enrichment-based DNA methylation data	http://www.bioconductor.org/packages/release/bioc/html/Repitools.html
Visualizing DNA methylation data		
bigWig/bigBed tools	Command-line tools for preparing genome browser tracks in a format that allows efficient visualization over the Internet	http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64
Ensembl	Widely used Web-based genome browser that includes a regulatory build with various epigenome data sets	http://www.ensembl.org
HilbertVis	Graphical tool and R package for visualizing genomic data as two-dimensional fingerprint diagrams	http://www.ebi.ac.uk/huber-srv/hilbert
IGB	Graphical genome browser that is run locally on the user's computer	http://bioviz.org/igb
IGV	Widely used graphical genome browser that is run locally on the user's computer	http://www.broadinstitute.org/igv
UCSC Genome Browser	Widely used Web-based genome browser hosting all ENCODE data	http://genome.ucsc.edu
WashU Epigenome Browser	Web-based genome browser focusing on the human epigenome	http://epigenomegateway.wustl.edu

l'Institut du thorax

Analyses des modifications épigénétiques : softwares (4)

Software	Description	URL
Identifying differentially methylated regions		
dmrFinder	Function for DMR detection that is a part of the charm package in R/Bioconductor	http://www.bioconductor.org/packages/release/bioc/html/charm.html
IMA	R package for exploratory analysis and DMR detection based on normalized Infinium data	http://www.rforge.net/IMA
NHMMfdr	R package for estimating false discovery rates (FDRs) using an explicit model of dependence between statistical tests performed for neighbouring CpGs	http://www.unc.edu/~pfkuan/softwares.htm
QDMR	User-friendly software tool for DMR identification based on Shannon entropy	http://bioinfo.hrbmu.edu.cn/qdmr
qvalue	R/Bioconductor package for calculating *q* value estimates of false discovery rates	http://www.bioconductor.org/packages/release/bioc/html/qvalue.html
Verifying and validating differences in DNA methylation		
BiQ Analyzer HT	Graphical tool for analysing locus-specific high-throughput bisulphite-sequencing data	http://biqanalyzer.computational-epigenetics.org
CpGassoc	R package for visualization and analysis of DNA methylation data	http://genetics.emory.edu/conneely
MassArray	R/Bioconductor package for processing Sequenom EpiTYPER data	http://www.bioconductor.org/packages/release/bioc/html/MassArray.html
MethLAB	Graphical interface for the visualization and analysis methods implemented in CpGassoc	http://genetics.emory.edu/conneely/MethLAB
MethMarker	Graphical tool for validating DMRs and designing DNA methylation biomarkers	http://methmarker.mpi-inf.mpg.de
PRIMEGENS	Web-based tool for large-scale primer design: for example, in the context of locus-specific high-throughput bisulphite sequencing	http://primegens.org

Analyses des modifications épigénétiques : softwares (5)

Software	Description	URL
Interpreting differences in DNA methylation		
AnnotationModules	Web-based tool for enrichment analysis based on genomic regions and using a diverse set of genome annotations	http://web.bioinformatica.cicbiogune.es/AM/AnnotationModules.php
EpiExplorer	Web-based tool for live exploration and interactive analysis of genomic region data in the context of public reference epigenome data sets	http://epiexplorer.mpi-inf.mpg.de
EpiGRAPH	Web-based tool for enrichment analysis based on genomic regions and using a diverse set of genome annotations	http://epigraph.mpi-inf.mpg.de
EVORA	R package for quantifying variation in DNA methylation as a cancer biomarker	http://cran.r-project.org/web/packages/evora
Galaxy	Widely used Web-based tool for genomic data processing and analysis	http://main.g2.bx.psu.edu
Genomic HyperBrowser	Web-based tool for statistical hypothesis testing based on genomic data sets	http://hyperbrowser.uio.no
GREAT	Web-based tool for Gene Ontology enrichment analysis based on genomic regions	http://great.stanford.edu

l'Institut du thorax

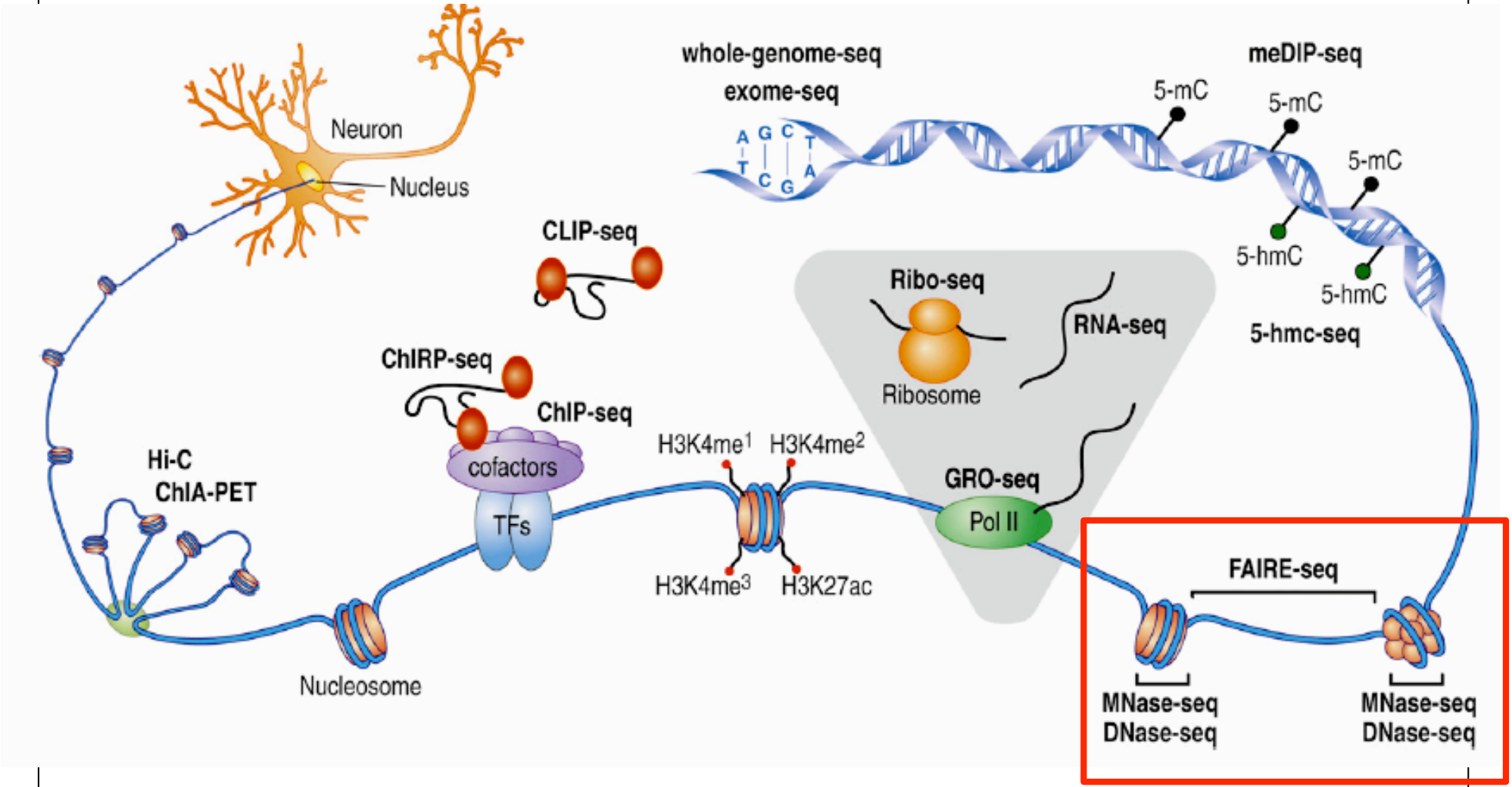
L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

- **1 – identification des bases, séquençage NGS**
- **2 – profil de méthylation de l'ADN**
- **3 – analyse du positionnement/organisation de la chromatine**

l'Institut du thorax

Analyses épigénétiques



Inserm

l'Institut du thorax

Analyses de l'organisation de la chromatine dans le noyau

Assays profiling nucleosome positioning

MNase-seq (micrococcal nuclease), DNase-seq (DNase I hypersensitive sites sequencing), and FAIRE-seq (formaldehyde-assisted isolation of regulatory elements)	Nucleosome positioning is identified by sequencing DNA that is protected by nucleosomes from digestion by micrococcal nucleases (MNase-seq). Alternatively, open chromatin regions are identified and sequenced based on their hypersensitivity to DNase I digestion (DNase-seq) or based on their solubility in the aqueous phase during phenol-chloroform extraction after formaldehyde crosslinking (FAIRE-seq).

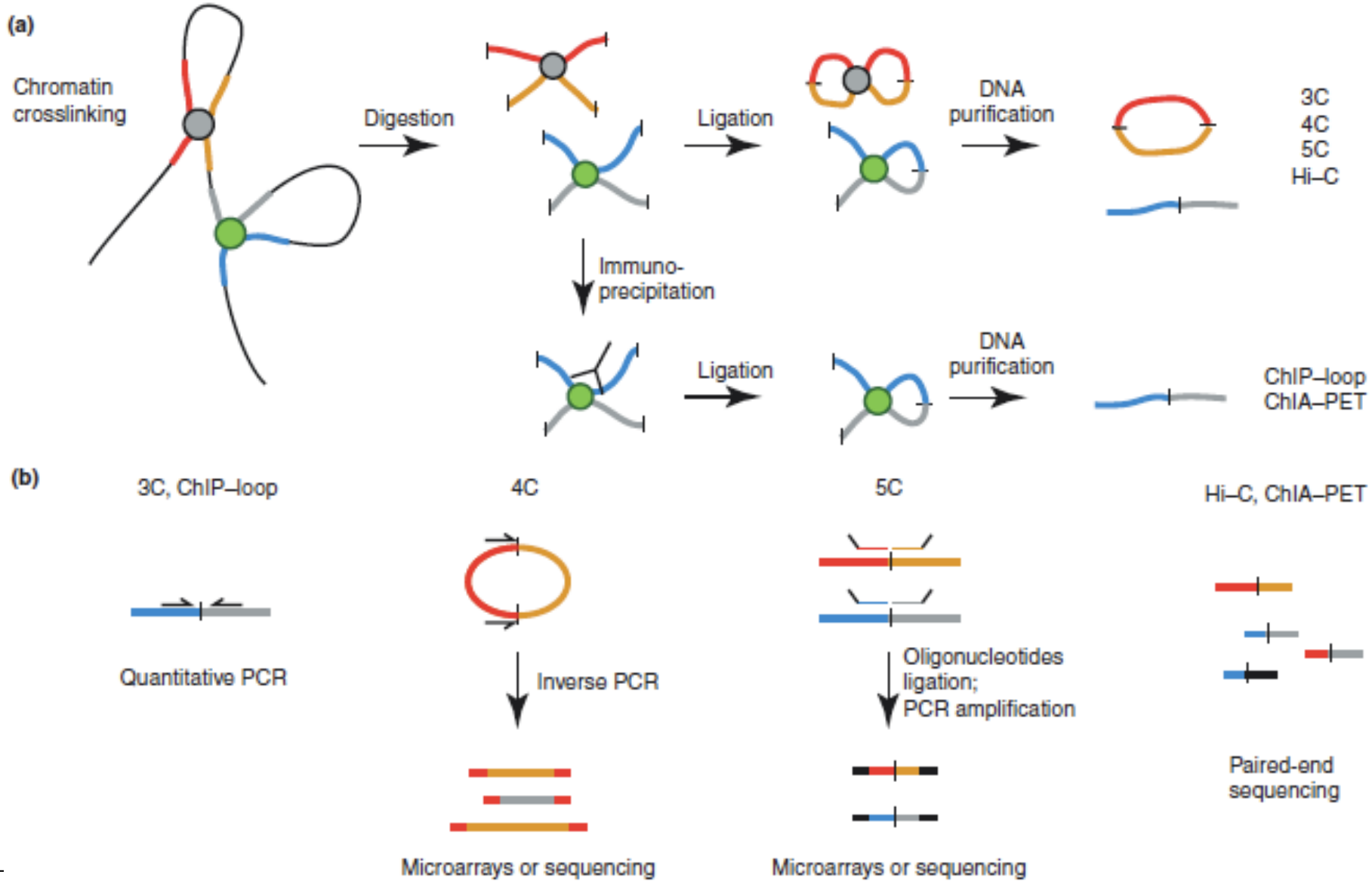
Assays profiling the organisation of the genome in the nuclear space

3C, 4C, Hi-C, and ChIA-PET	Chromosome conformation capture techniques determine the physical interaction between known genome regions (3C), unknown regions of the genome and a known bait (4C), or all genome-wide-occurring interaction in an unbiased fashion (Hi-C). ChiA-PET involves an immunoisolation step that allows the identification of the genomic interaction sites of a specific protein.

l'Institut du thorax

Analyses de l'organisation de la chromatine dans le noyau :

chromosome conformation capture = 3C et méthodes dérivées



l'Institut du thorax

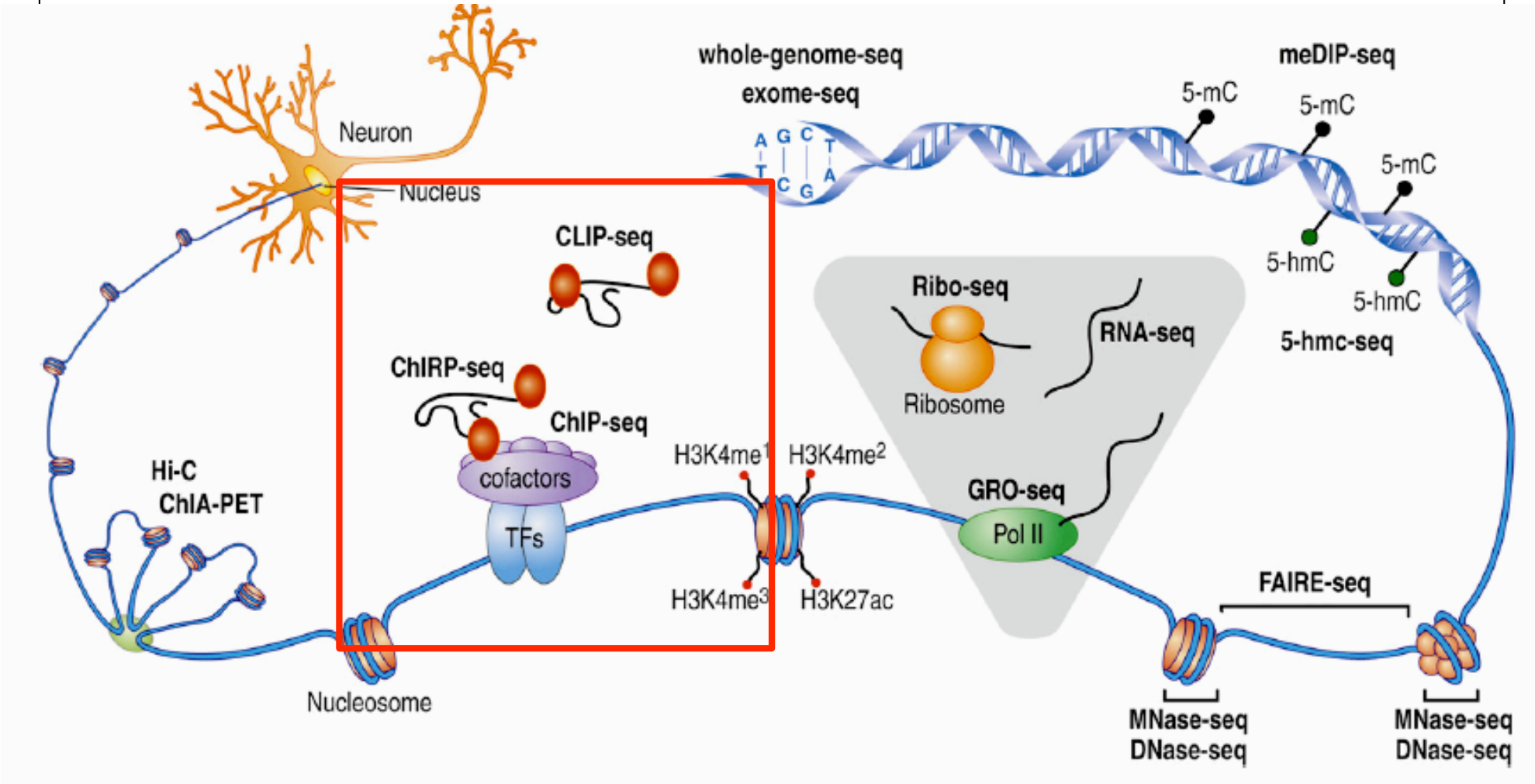
L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

- **1 – identification des bases, séquençage NGS**
- **2 – profil de méthylation de l'ADN**
- **3 – analyse du positionnement/organisation de la chromatine**
- **4 – analyse des histones**

l'Institut du thorax

Analyses épigénétiques



Inserm

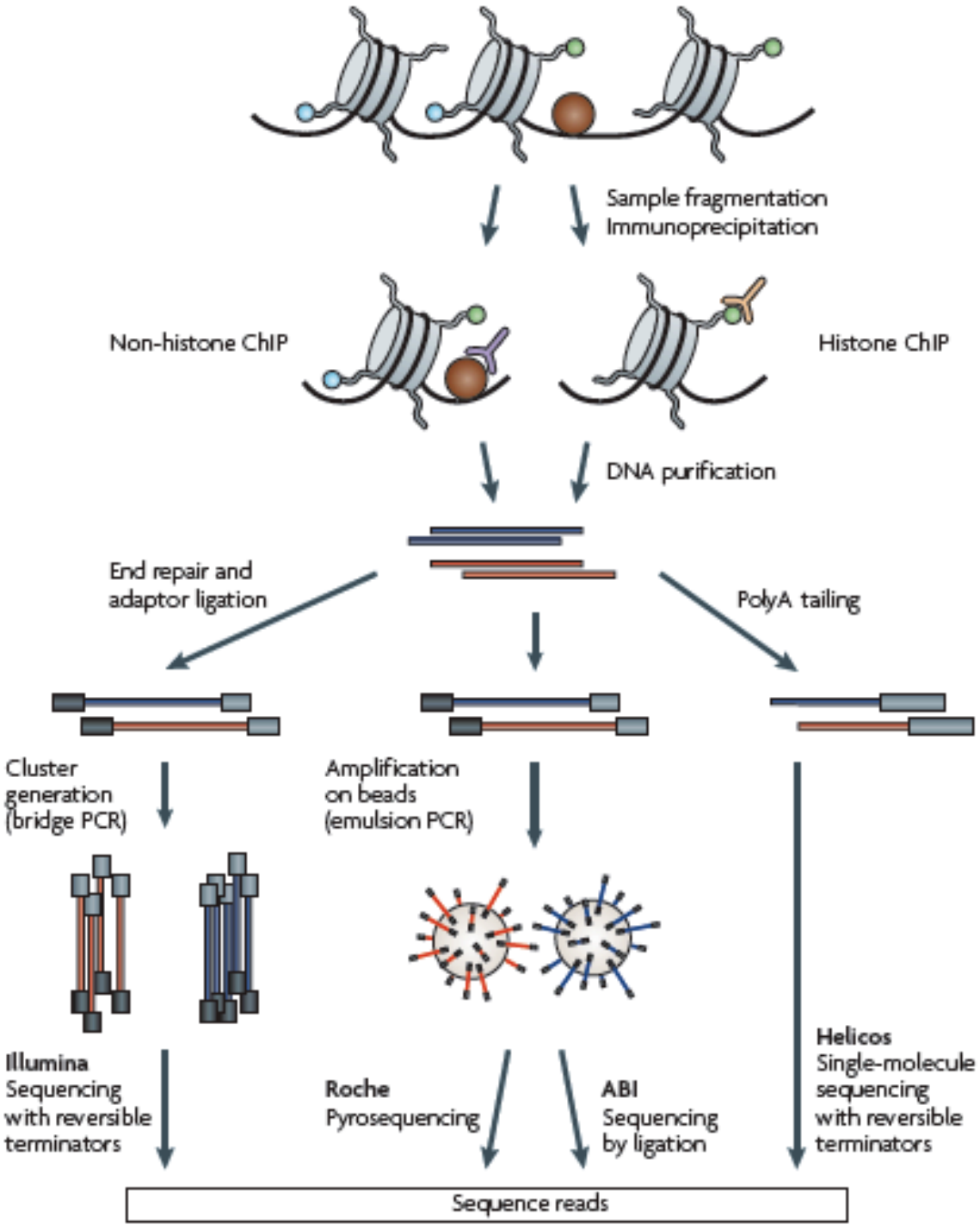
l'Institut du thorax

Analyses épigénétiques

Assays profiling the association of proteins/ histone marks/RNA to the genome	ChIP-seq (chromatin immunoprecipitation)	Based on enrichment of crosslinked DNA-protein complex isolated by an antibody raised against a protein or specific histone mark of interest. The bound chromatin fraction is identified by deep sequencing.
	ChIRP-Seq (chromatin RNA immunoprecipitation)	Variant of ChIP-seq that specifically detects association of RNA molecules to the genome. Based on hybridization of a pool of nonoverlapping, biotinylated oligonucleotides that complement to the sequence of the RNA of interest. The bound DNA are processed as in ChIP-seq.
Assays profiling the association of proteins to (RNA) the transcriptome	CLIP-seq (crosslinking and immunoprecipitation) and PAR-CLIP (photoactivatable ribonucleoside-enhanced-CLIP)	Based on UV-crosslinking to stabilize RNA and RNA binding protein interactions, followed by immunoisolation of the RNA-protein complex using a specific antibody for the protein. Enriched RNAs are then converted into cDNAs for sequencing (CLIP-seq). PAR-CLIP differs from CLIP-seq in the incorporation of photoactivatable ribonucleoside analogs into the nascent transcripts that facilitate UV-crosslinking between labeled RNAs and proteins.

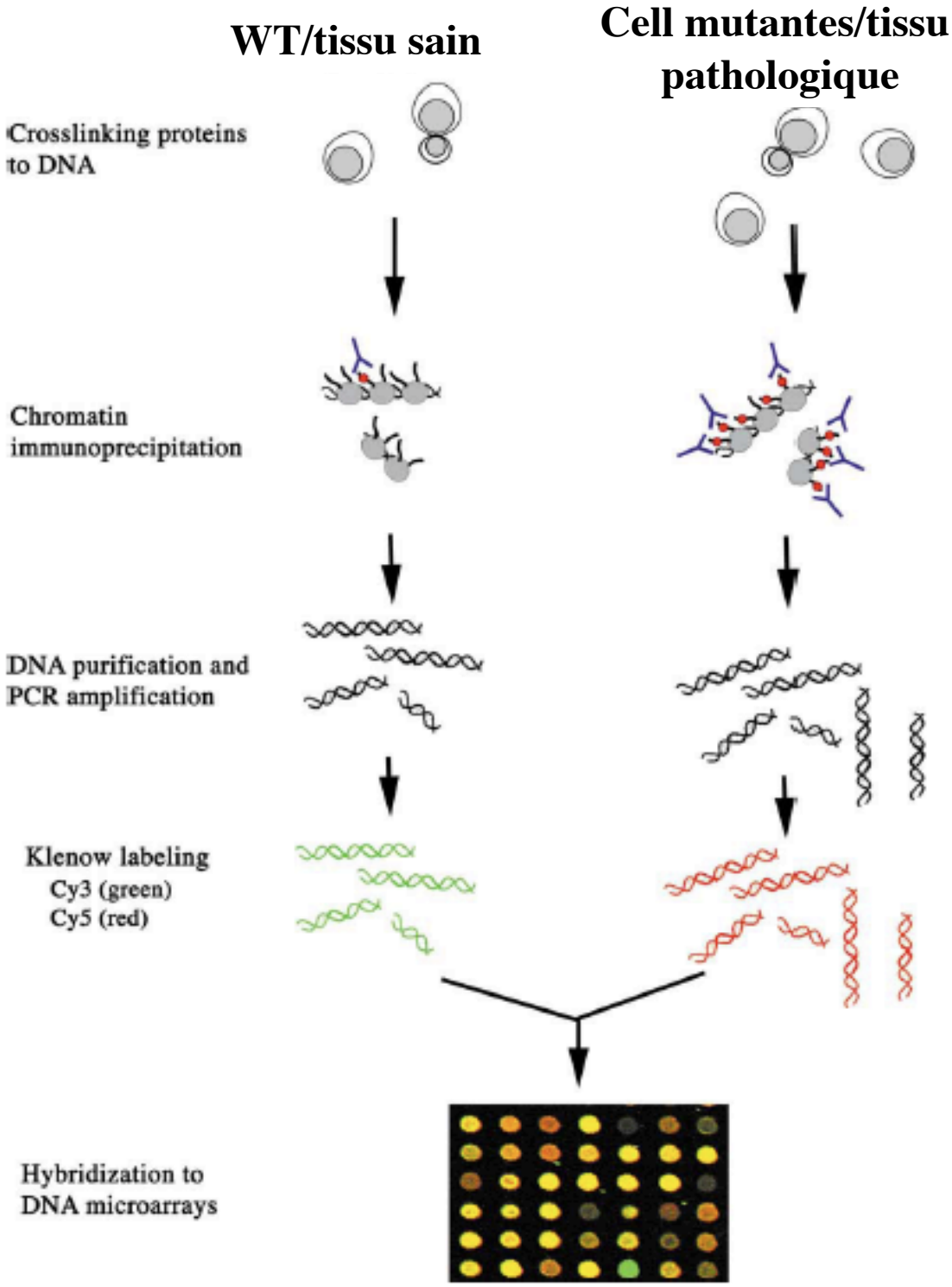
l'Institut du thorax

Histone Chip-Seq



l'Institut du thorax

Analyses des modifications des histones : Histone CHip



l'Institut du thorax

Analyses des modifications post-traductionelles des histones

Target Class	Detection Method	Assay
Protein Methyltransferases	Colorimetric	ELISA
	TRF	DELFIA
	Chemiluminescence	AlphaScreen
	Fluorescence	Microfluidic capillary electrophoresis
	Radioactive	Incorporation of radioactive methyl groups
	Enzyme-coupled fluorescence	Thioglo chromophore[31] or Ellman's reagent
Protein Demethylases	Enzyme-linked colorimetric	Formaldehyde dehydrogenase coupled reaction
	Colorimetric	Peroxide production
Histone-Binding Proteins	Chemiluminescence	AlphaScreen
Histone Acetyl Transferases and Histone Deacetylases	Fluorescence	Fluor-de-Lys® assay
	Luminescence	Prolumigenic peptide
	Fluorescence	Microfluidic capillary electrophoresis

l'Institut du thorax

L'épigénétique

I – Introduction
II – Historique
III – Définition et rappels
IV – Modifications de l'ADN : la méthylation
V – Modifications de la structure de la chromatine/des protéines
VI – Les ARN non-codants
VII – Approches technologiques à l'étude des modifications épigénétiques

- **1 – identification des bases, séquençage NGS**
- **2 – profil de méthylation de l'ADN**
- **3 – analyse du positionnement/organisation de la chromatine**
- **4 – analyse des histones**

VIII – Les pathologies liées aux modifications épigénétiques

Exemples de pathologies liées à des modifications épigénétiques

Syndrôme de RETT : Mutation du gène codant pour la MeCP2 (methyl-CpG-binding protein)

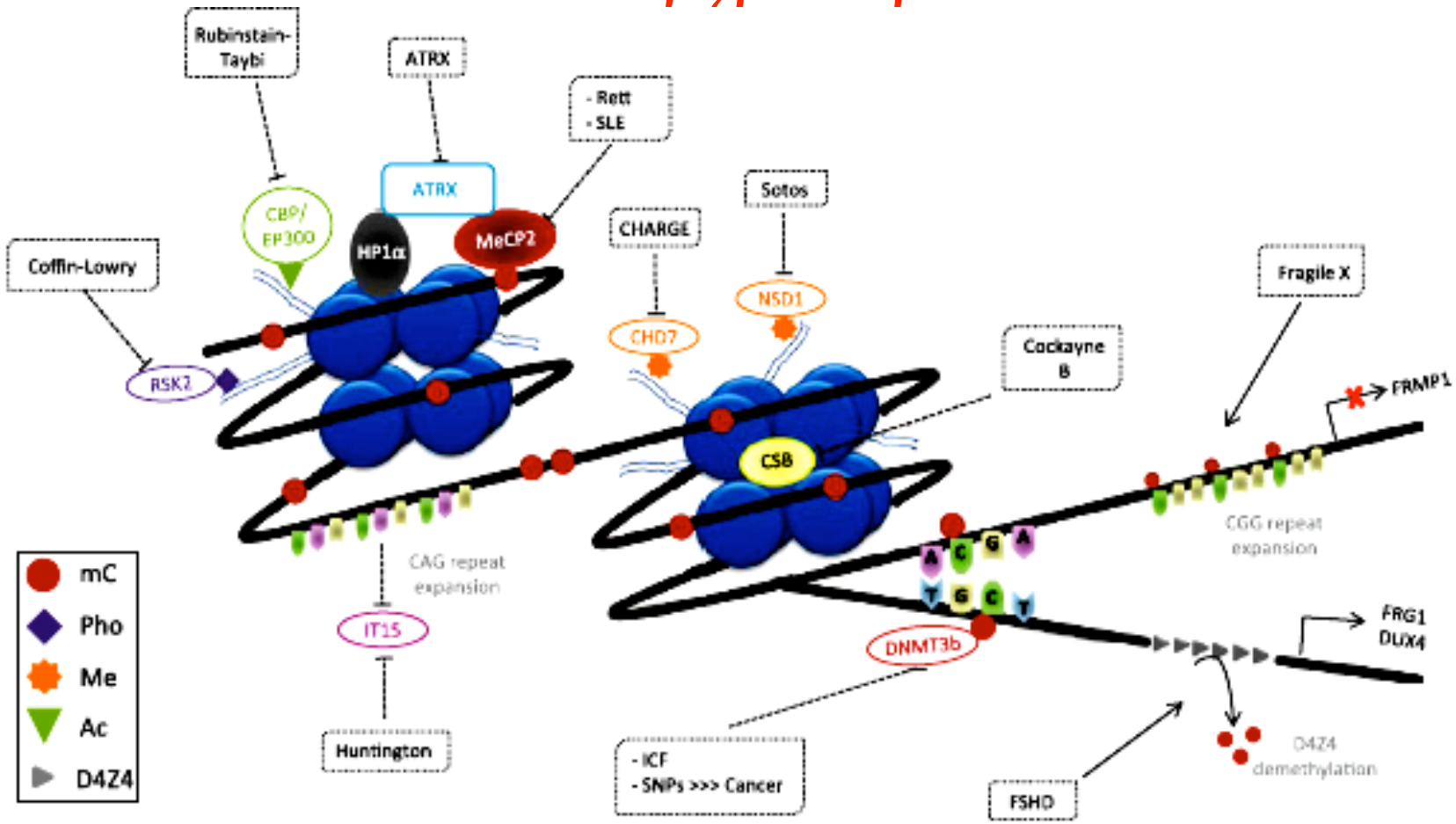
ICF syndrome (Immuno-deficiency, centromic region of instability et facial anomalies) : Mutation du gène codant pour la DNA méthyltransférase 3ß

Fragile X syndrome : Mutation $(CGG)_{n>55}$ en 5'UTR du gène FMR1 → méthylation du gène

Beckwith-Wiedemann syndrome : Perte de méthylation de la région DMR2, ou gain de méthylation de la région DMR1 du gène IGF2

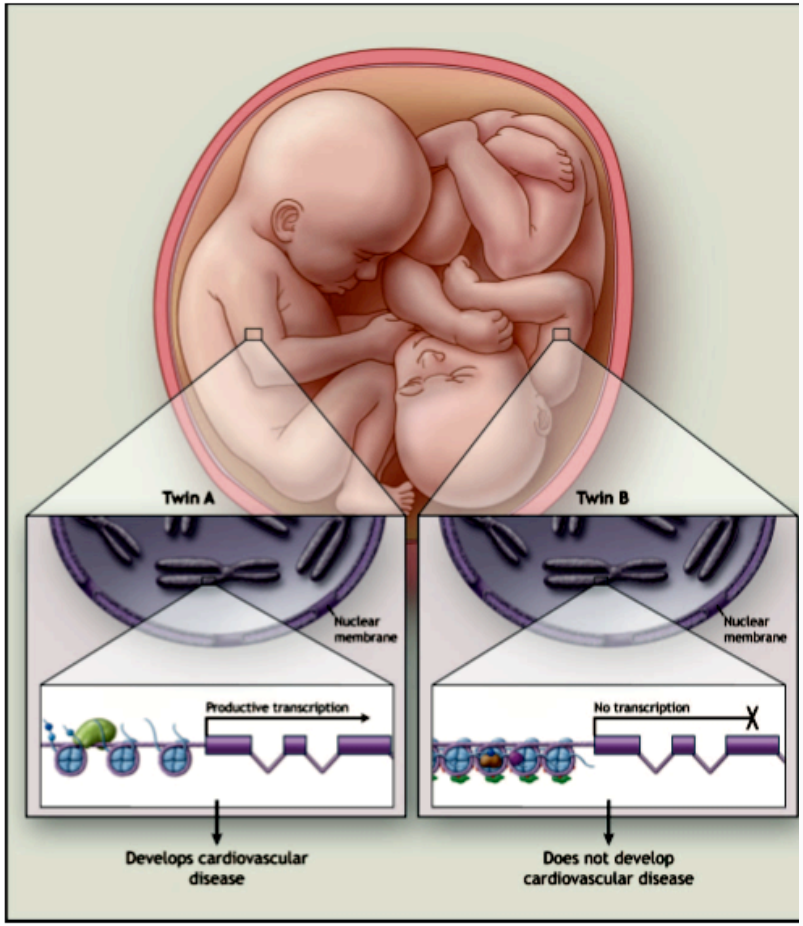
Prader-Willi syndrome, Angelman syndrome, Wilms tumor, Myotonic dystrophy (congenital) etc…

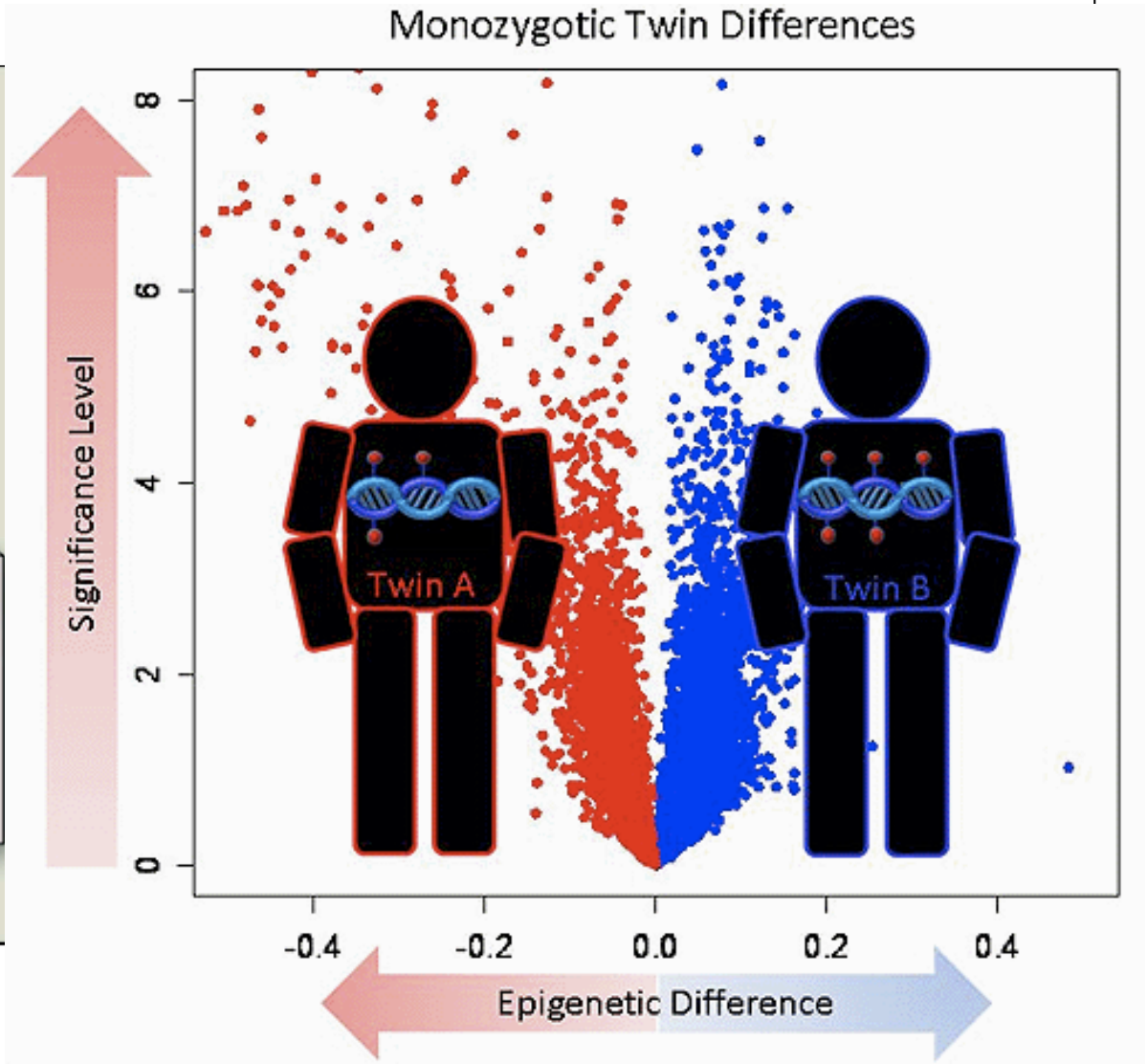
Exemples de pathologies liées à des modifications épigénétiques



l'Institut du thorax

Epigénétique et monozygotes





Inserm

Bases de données et épigénétiques

Epigenomic data resource	Description
Ensembl www.ensembl.org	Ensembl uptakes the epigenomic data tracks from ENCODE and the NIH Roadmap Epigenomics Project as well as smaller hypothesis-driven projects. Hundreds of tracks are available, primarily from ChIP-seq and DNase-seq assays for human and mouse cell types. Ensembl also contains a resource for the identification of human tissue-specific differentially methylated regions [109].
NCBI epigenomics portal www.ncbi.nlm.nih.gov/epigenomics	Curated, annotated and organized epigenetics-specific data selected from general-purpose archives, such as the Gene Expression Omnibus, and Sequence Read Archives, including the data from the NIH Roadmap Epigenomics Project. Includes 337 samples and over 1100 data tracks from five well-studied species.
NIH Roadmap Epigenomics Project www.roadmapepigenomics.org Human Epigenome Atlas www.epigenomeatlas.org	The largest current epigenomics project with an exclusive focus on producing a comprehensive catalog of human epigenomic maps, including methylomes, histone marks, chromatin accessibility and RNA-seq data for hundreds of human cell types. The epigenome maps produced by the NIH Roadmap and other projects are being integrated into the Human Epigenome Atlas.
UCSC genome browser www.genome.ucsc.edu	A collection of high-throughput, genome-wide data generated on a panel of human cell lines by the ENCODE project including histone marks, chromatin accessibility, DNA methylation, RNA expression, RNA binding and other cell state indicators.