



## PARIS7 BIOINFO CLUB

*Bioinformatics Club – IJM, Epigenetics, BFA*

# Next meetings

---

**13 September, 10am:** Twitter and other communication tools like Altmetrics  
(Pierre Kerner)

**11 October, 10am:** Three short talks on the status of bioinformatics in Epigenetics and cell fate unit (Paris Epigenetics, Olivier Kirsh), in BFA (Olivier Taboureau) and in IJM (Virginie Courtier-Orgogozo). Free discussion to define the themes to be addressed by the BioInfo Club.

**8 November, 10am:** Presentation of R and R Studio (Bénédicte Lefèvre) – bring computer

**15 November, 10am:** Presentation of ggplot to make great figures and drawings (Samuel Murail) – bring computer

**6 December, 10am:** (Kevin Cheeseman)

The BioInfo Club meets on Thursday from 10am to 11am, in François Jacob

# UMR7216 – Who are we?

The screenshot displays the website for UMR7216, featuring a header with the logo 'Epigénétique & Destin Cellulaire' and a search bar. The main banner reads 'EPIGÉNÉTIQUE ET DESTIN CELLULAIRE'. A central navigation menu lists research groups, and a sidebar on the right includes 'ENSEIGNEMENT' and 'PUBLICATIONS'. A 'Séminaires' dropdown menu is visible, and the footer contains social media icons and institutional logos.

Epigénétique & Destin Cellulaire

Recherche...

## EPIGÉNÉTIQUE ET DESTIN CELLULAIRE

GROUPES

- Ait-Si-Ali - Dynamique épigénétique et diff. cellulaire >
- Cosson - Régulations post transcr. & épigénétiques >
- Defossez - Dynamique de la Méthylation de l'ADN des Génomes Eucaryotes >
- Francastel - Méthylation de l'ADN et ARNnc en physiopath. >
- Mezger - Interface Dév. & Environnement >
- Polo - Intégrité de l'épigénome >
- Rougeulle - ARNs non-codants, Diff. et Développement >
- Weitzman - Plasticité des phénotypes cellulaires >

ENSEIGNEMENT

PUBLICATIONS

Séminaires

Annuaire

Liens utiles

RIS DIDEROT

SORBONNE PARIS CITE

<http://parisepigenetics.com/fr/>

# Epigenetic & cell fate - UMR7216 – Where are we?

UMR 7216  
**Epigenetics**  
& **Cell Fate**

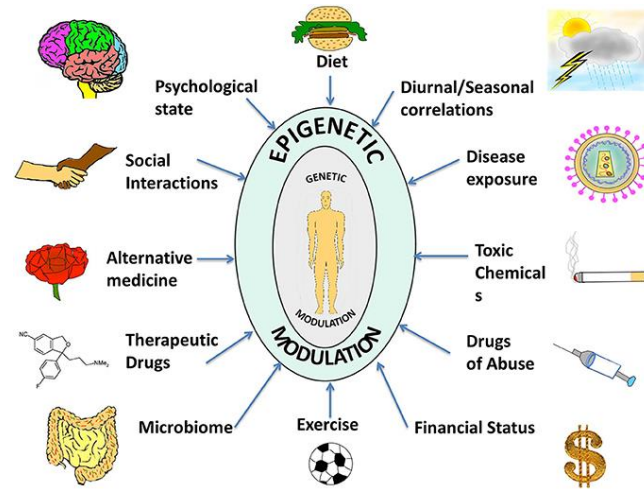
université  
**PARIS**  
**DIDEROT**  
PARIS 7



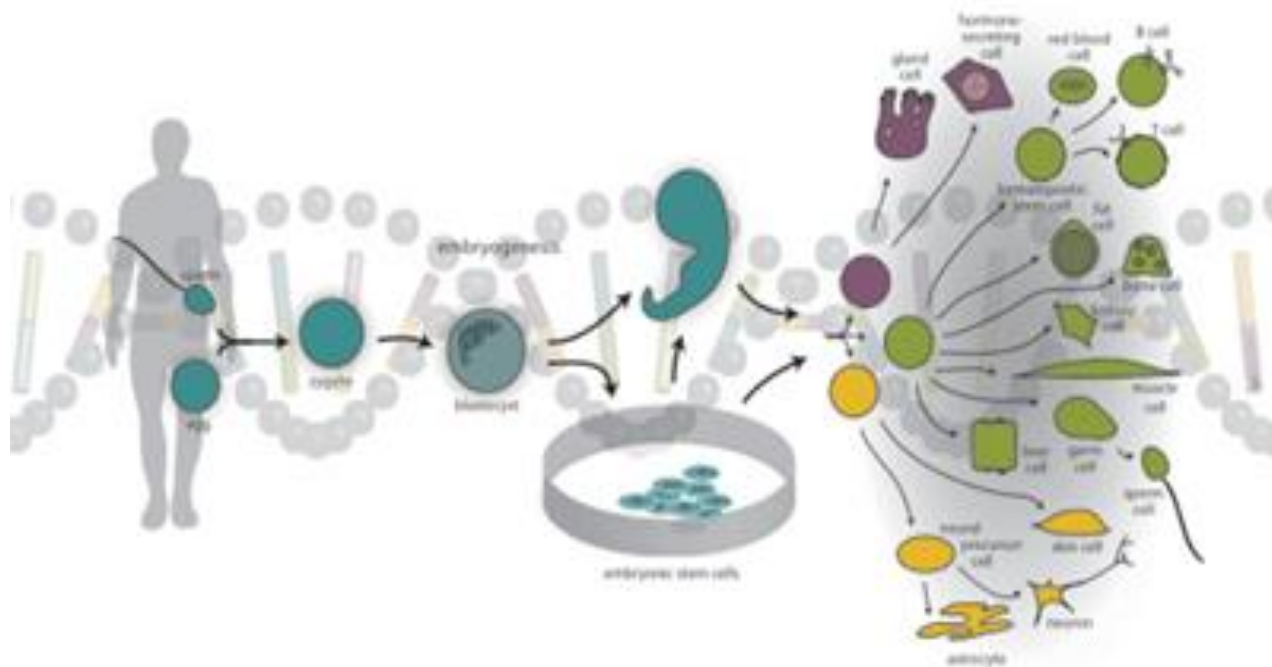
Lamark building  
4th floor

# Our biological questions

- Development
- Stem cell biology
- Response to stress
- Oncogenesis
- Human diseases
- Differentiation

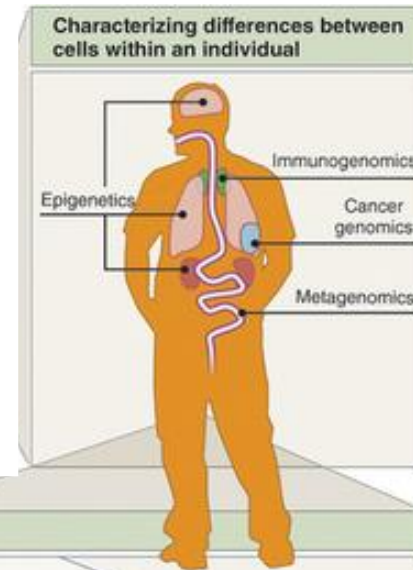


Kanherkar et al

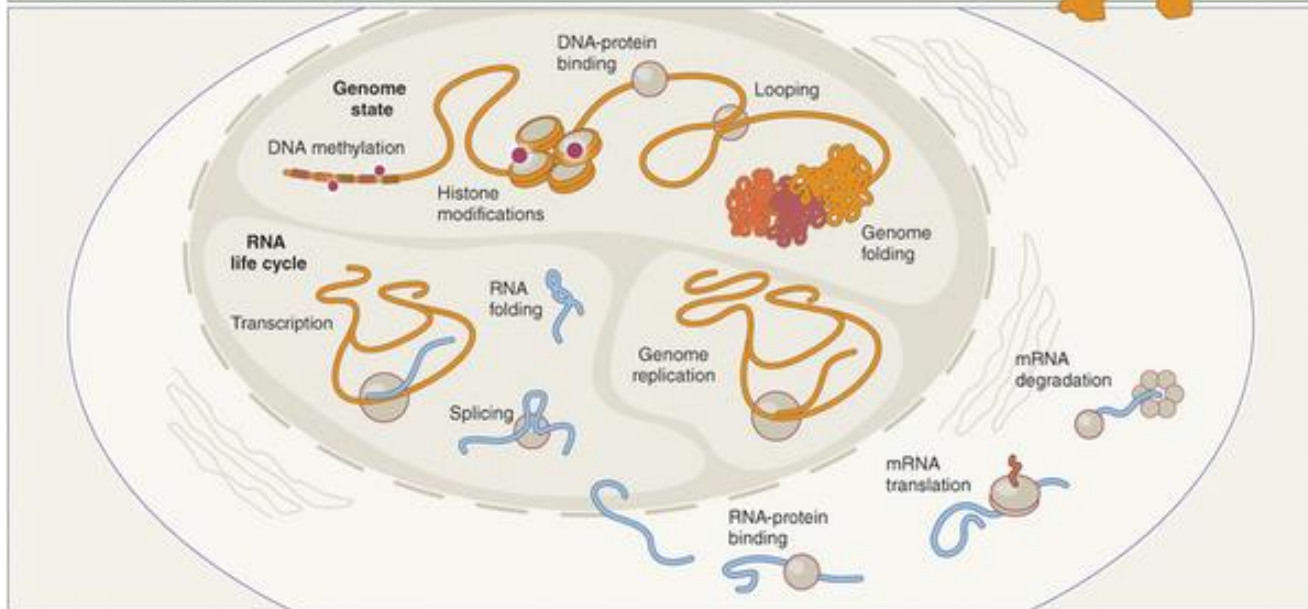


# Our biological questions

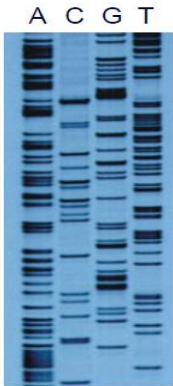
- Gene expression
- Chromatin modification
- Chromatin organisation
- DNA modification / structure
- RNA biology



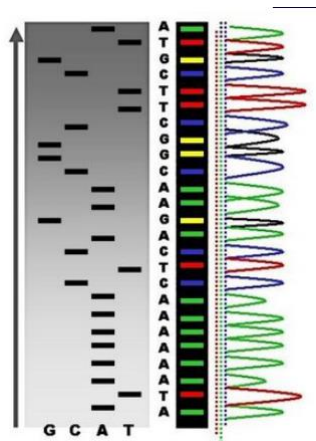
## Describing the underlying cellular mechanisms



# Sequencing technologies



Late 70's



90's  
1st generation



2007  
2<sup>nd</sup> generation

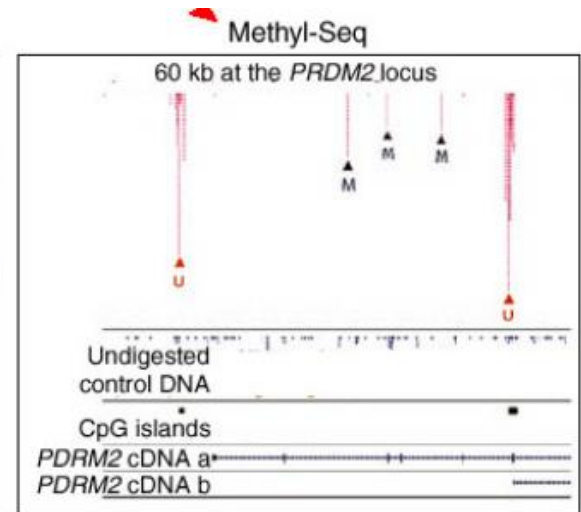
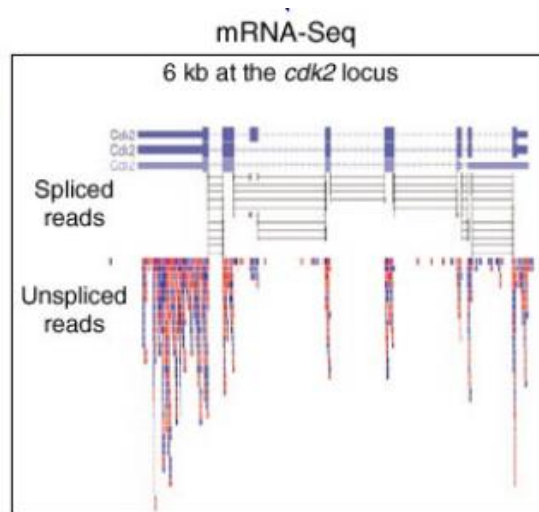
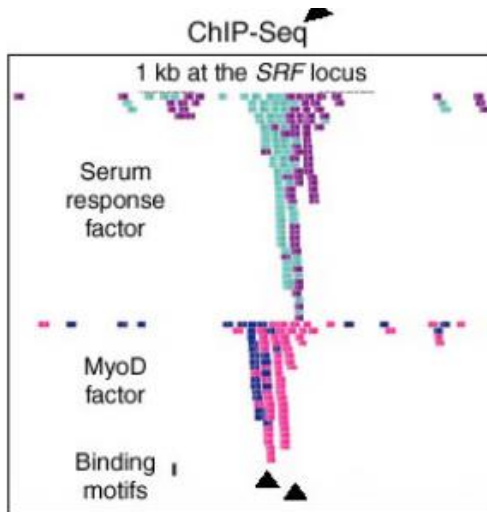
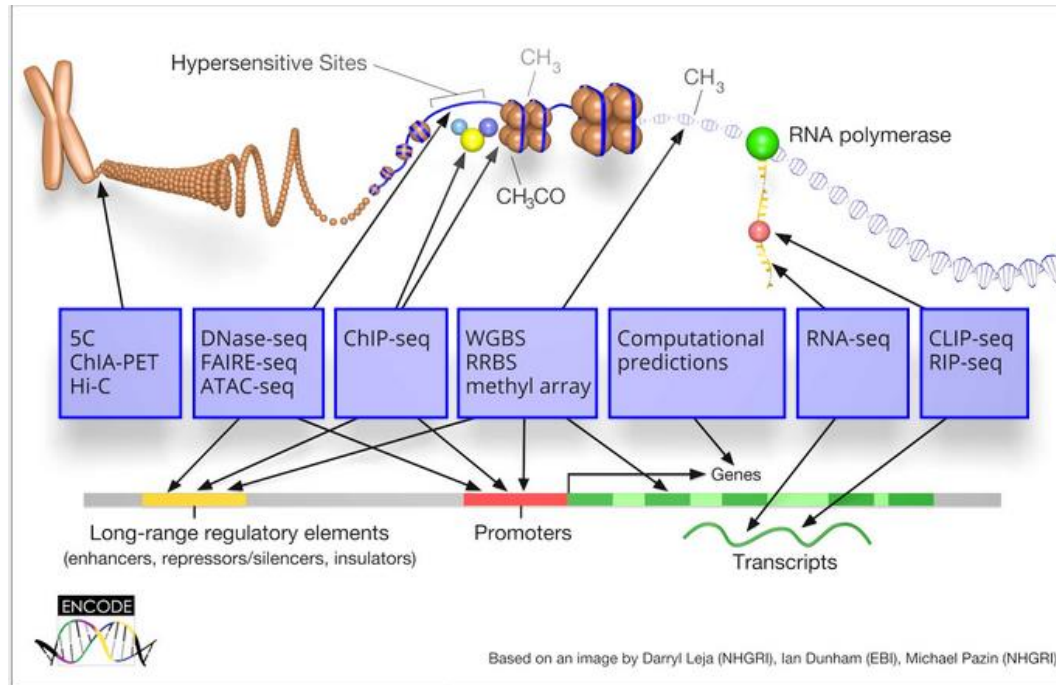


3rd generation

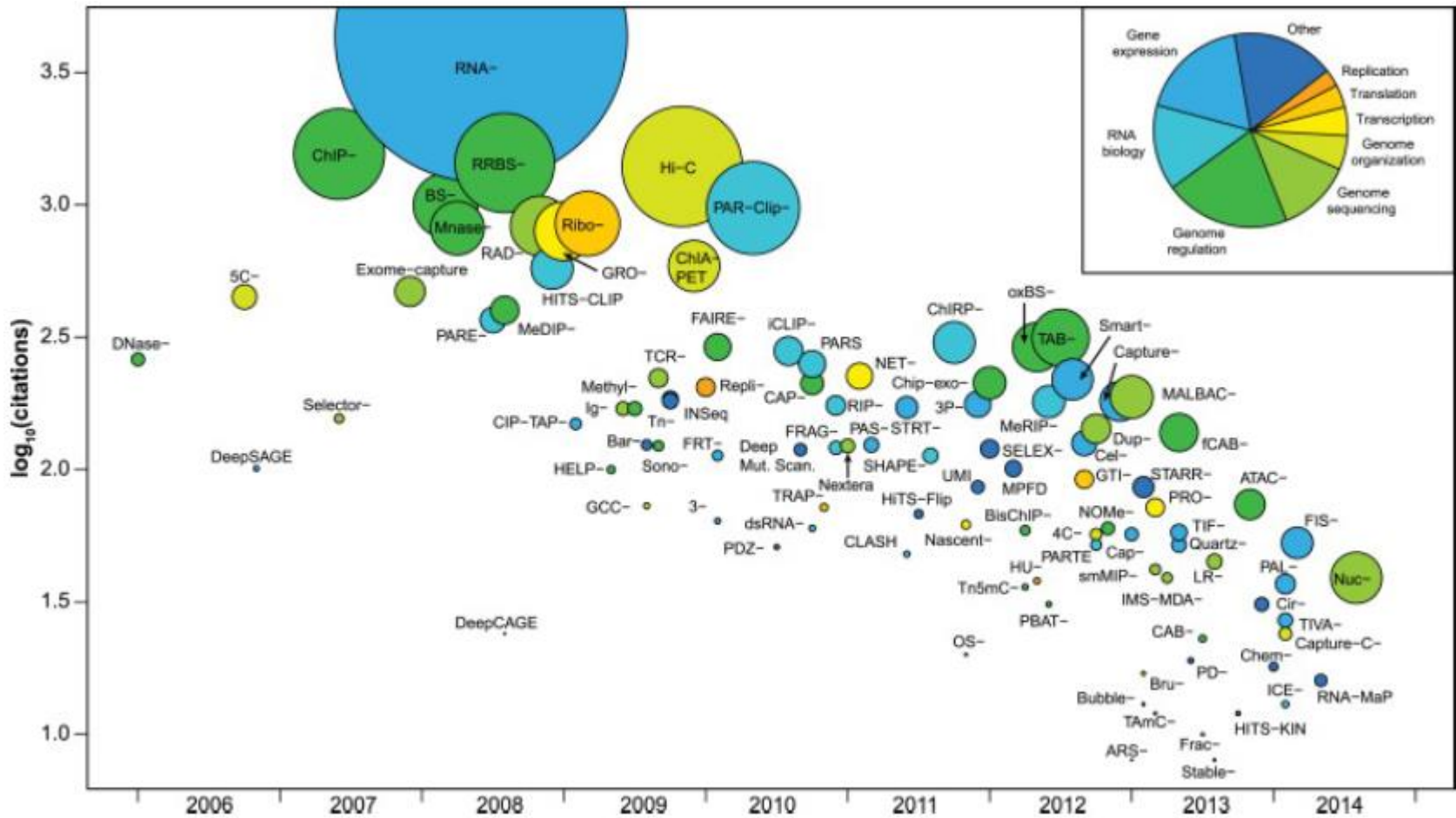
**HTS : High throughput sequencing**

**Novaseq =  $10 \cdot 10^9$  seq/44h**

# High throughput sequencing applications in biology

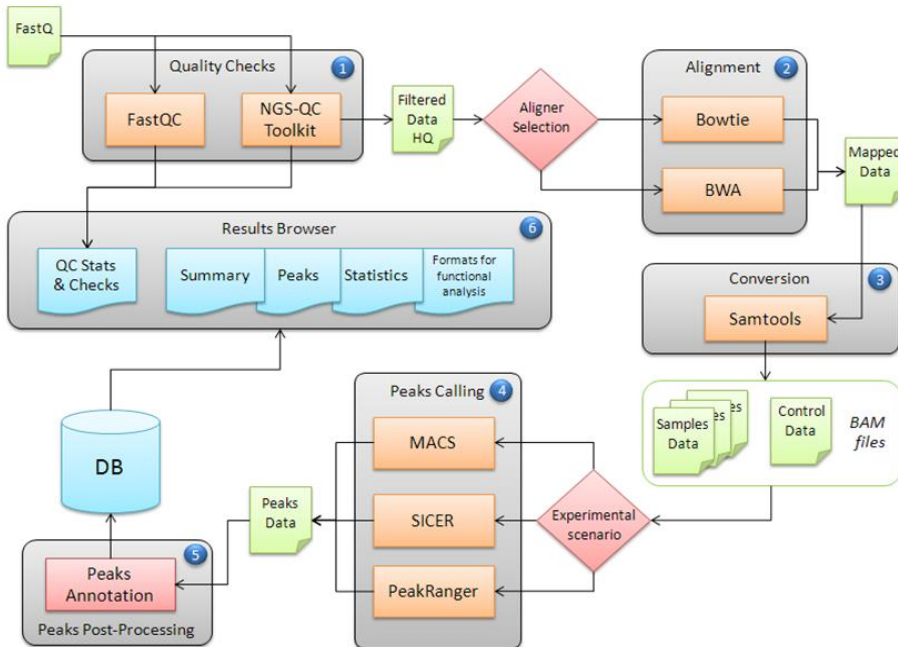


# High throughput sequencing applications in biology





## 1/ Design data analysis workflows



## 2/ Write/adapt scripts with

- open source tools / algo
- in house tools



## 3/ Perform data processing with

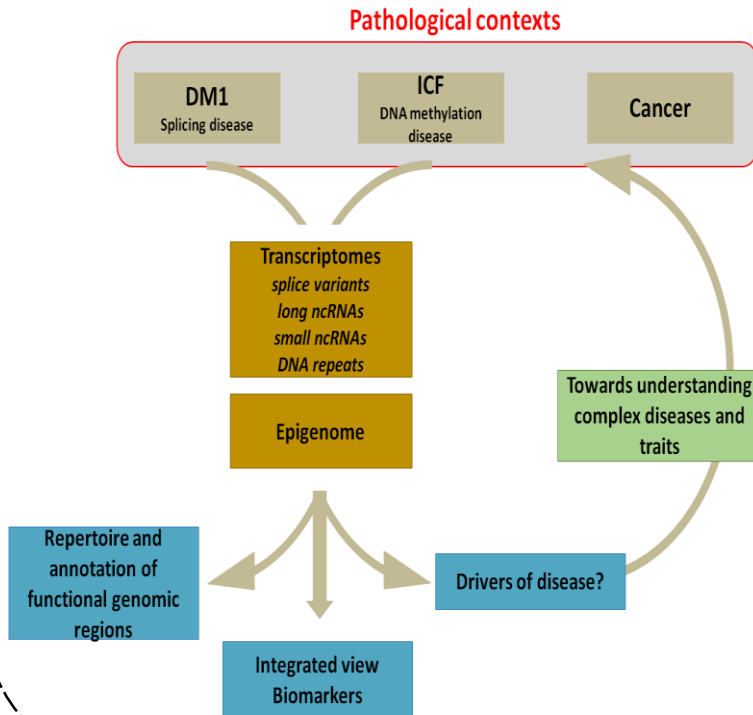
- (galaxy)
- local machine
- Cluster (RPBS)





## Questions biologiques :

- Vision intégrée des altérations de l'épigénome et du transcriptome dans des contextes pathologiques :
- Compréhension du lien causal:
  - Entre ces différents mécanismes moléculaires
  - Avec les phénotypes cellulaires et signes cliniques
- Etablissement de biomarqueurs

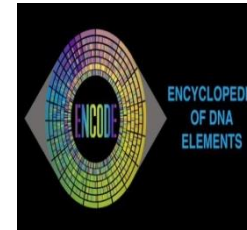


## Données générées : projet OMICS

- RNA-seq: total, small, « medium »
- ChIP-seq
- Methylome



Données publiques : Encode, GEO, EBI, UCSC



Chromatin segmentation (ChromHMM)  
RepeatMasker  
Repli-seq

**Approches** : Illumina methylchip 450-850K, RNA-Seq, ChIP-seq, et données cliniques

**Gestion des données** : Galaxy, R

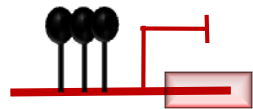
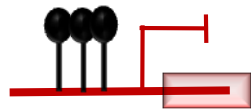


**Packages clés** : DESeq2, Minfi, ChAMP



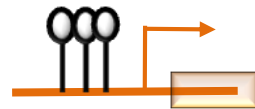
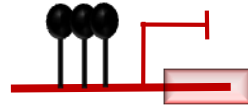
### Questions biologiques :

Somatic cell



Repressed C/T gene


Cancer cell



Re-activated C/T gene

DNA methylation dependant:

- Gene expression in cancer
- Gene expression in mES
- DNA methylation maintenance
- Functions of a MBD (ZBTB4)
- HIV latency mechanisms
- Interplay between signalling pathways & chromatin remodellers

Données publiques : TCGA   
cancers du colon, sein et poumons



GEO Dataset , SRA

**Données**

WGS, RNA-Seq, 450K et données cliniques

**Données générées au labo**

ChIP-Seq, RRBS, Nanostring, micro arrays

**Gestion des données :**

R, Shell unix, conda, docker en local ou sur le cluster, Rmd, Notebooks, jupyter

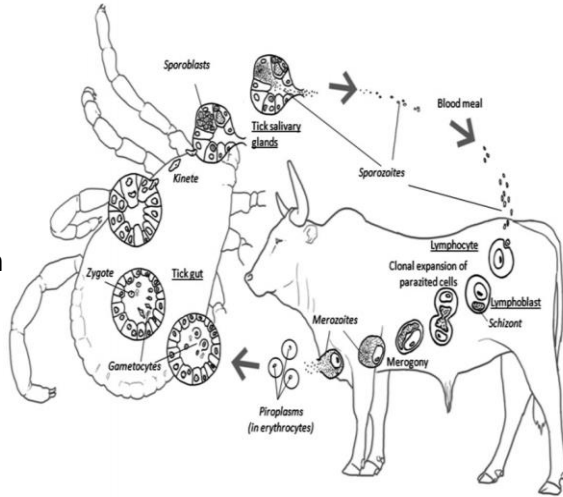
**Packages / outils :** (TCGAbiolinks), DESeq2, MCPcounter ++ scripts maisons  
HOMER, Samtools, bedtools, bedops, star, hisat2, htseq, methylkit, ggplot2, complexheatmaps etc....





**1. How does an intracellular parasite controls both host & parasite gene expression for its survival and the completion of its life cycle?**

A complex life cycle with different hosts and many cell types involving immortalisation, proliferation and differentiation processes



Comparative genomics

Functional (epi)genomics

data

Lab datasets & public db  
(EuPathDb, Pfam, Interpro, OrthoMCL-db)

work env.

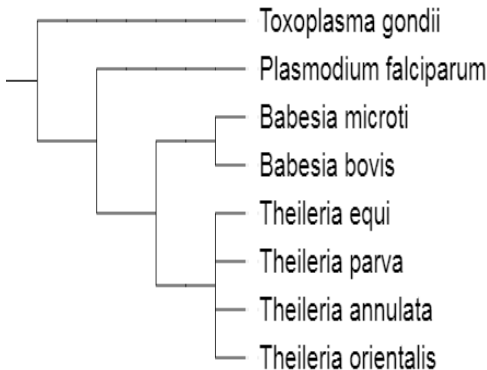
(local) Linux VM & cluster,  
Bash, R (Bioconductor) & Docker

approaches

Db, motifs & sequence mining,  
network analyses,  
MSA & phylogeny

RNA-seq,  
ChIP-seq,  
clustering,  
Ontologies/GSEA

**2. How did the parasite genome evolve these functions?**

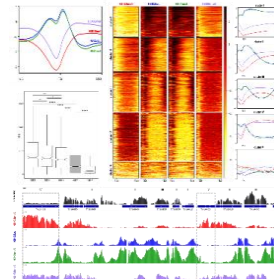
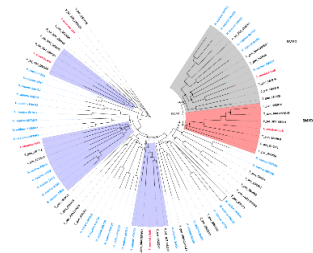


compact & dynamic apicomplexan genomes

main tools

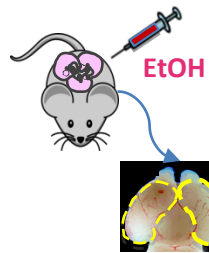
blast,  
mafft,  
ete3, Iqtree

Rsubread, HISAT2,  
EdgeR, Music,  
Deeptools





## Prenatal alcohol exposure during brain development : molecular mechanisms involved in DNA methylation defects

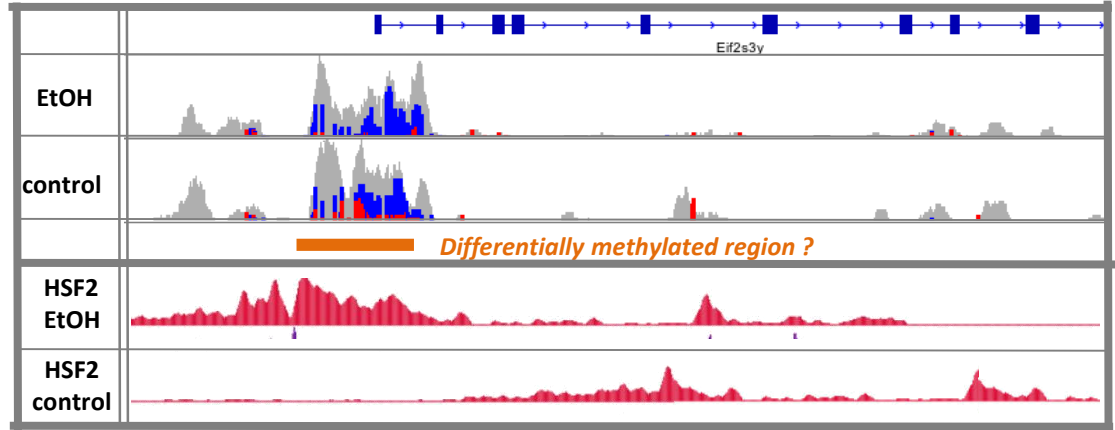


EtOH

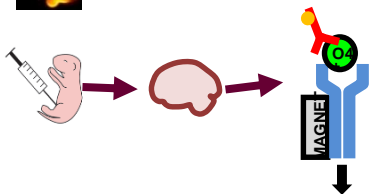
Personal data :  
mouse embryos cortices

BS-seq : targeted  
capture of DNA  
methylation (EpiCapture)

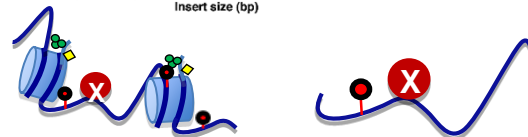
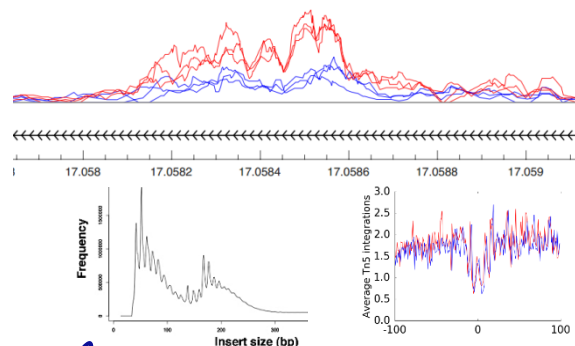
ChIP-seq of TF involved  
in stress response



## Impact of inflammatory stress on cell identity



Personal data :  
oligodendrocytes



ATAC-seq

Micro array

### •Environment - languages :

- bash
- R : packages + personal scripts



### •Project management :

- local computer or distant server
- Galaxy



### •People involved in BI project :

- Agathe Duchateau (PhD student)
- Délara Sabéran Djoneidi (MCF)
- Sascha Ott (collab. Warwick University)



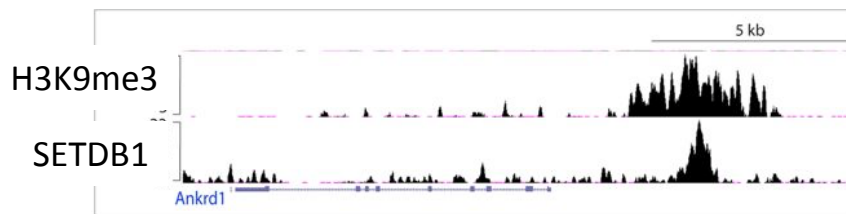


**Objectifs** : Conséquences de la dérégulation de SETDB1 sur la distribution pangénomique de H3K9me3, la localisation de CTCF/Cohésine et le transcriptome

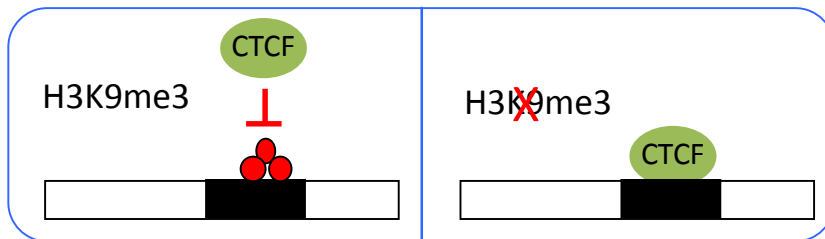
Conséquences épigénomiques ?

Immunoprécipitation de la chromatine (ChIP) -Seq

H3K9me3 et SETDB1



CTCF, Cohésine



Effet sur l'expression des gènes ?

Transcriptome  
RNA-Seq

Impact sur l'architecture  
3D du génome

Hi-C



L'amplification de SETDB1 pourrait jouer un rôle clé dans l'organisation 3D du génome et la régulation de l'expression des gènes



## Biological questions

- Roles of long non-coding RNAs, repeat elements and genomic organization in X chromosome inactivation
- X chromosome inactivation dynamic in the Human hematopoietic lineage : physiological and pathological conditions.
- Systematic identification and characterization of non-coding RNAs associated with bladder cancer progression

## Bioinformatics approach

- RNA-seq : hierarchical clustering, differential expression analysis, transcript reconstruction
- single-cell RNA-seq : allelic analysis
- ChIP-seq
- Hi-C
  
- "Homemade" datasets
- Databases : BLUEPRINT, TCGA, GEO, SRA, HPA



## Data analysis management

- MTI cluster + Migale (INRA)
- PCs Ubuntu / MAC
- NAS (32To)
- R packages : DE-Seq2, ConsensusClusterPlus, limma, ggplot2, RCircos...
- Other tools : fastqc, hisat2, samtools, StringTie, scallop, GATK4, DE-kupl, deeptools, HiC-Pro, Blast...
- Used languages : BASH, R, python



# Biological-bioinformatics questions

- Post transcriptional epigenetic regulation. The role and the fate of mature mRNA. Regulation on the translation level. MRNA RNA Binding Properties interactions.
- Large scale, genome wide, integration of multiple -omics datasets. Exploratory and predictive analysis.
- Network systems biology, software development of analytical tools for biological knowledge extraction.

## Research approaches

- All different kinds of \*omics datasets. RNA-seq, ribo-seq, polysome profile, Hi-C, RIP-seq and any sorts...
- Sourced either locally, through collaborations or from big public databases. SRA, ENA, 4D genomes etc.
- Using the full repertoire of R – Bioconductor and scientific computing for data science in Python (Scipy, Pandas, biopython, sklearn etc.)
- Software development of user friendly, visualisation rich, bioinformatics applications and pipelines (nextflow, shell scripts, python and R packages, Docker - Singularity).
- Curation of the Paris Epigenetics GitHub repository.

<https://github.com/parisepigenetics/>

## Methods and competences.

- Exploratory data analysis. Dimensionality reduction, correlation matrices.
- Computational and statistical learning. Supervised/unsupervised machine learning
  - k-mean, hierarchical clustering.
  - Bi-clustering.
  - Building classifiers (ANN, random forests, Boosting etc.).
- Network inference, computational modelling.
- Visualisation as a tool.





# Training & teaching



B Cosson  
PR



C Bouyoukos  
MCF



O Kirsh  
MCF



JF Ouimette  
MCF



G Velasco  
MCF



A Duchateau  
PhD student



K Cheesman  
Post Doc



O Rospopoff  
Post Doc



M Laisné  
Phd Student

## UFR SDV

- teaching in the M1-2 BIB:
  - Network Systems Biology
  - Bioinformatics Genomics
  - Scientific Software Carpentry
  - Genomics week
  - R / Python
- Magistère de génétique –RNA-seq
- L3 BMA
- L3 OMIC
- L2 / L3 Biostatistics - R

## UMR7216

The screenshot displays the 'Formation BiInfo UMR7216 EDC' website. The main content area is titled 'Introduction Formation bioinfo niveau I' and lists three courses:

- \*cours 1 : par Bertrand et Jean François : conception du projet, séquençage jusqu'au raw data**  
thème abordé : but du séquençage/analyse haut débit, construction de la librairie, profondeur/couverture, les questions à se poser avant de séquençer, principe du séquençage...
- \*cours 2 : par Agathe, Olga et Olivier: workflow général**  
thème abordé : explication des grandes étapes du traitement des données haut débit, pour donner le vocabulaire de base, le format des fichiers et leur contenu général, la complexité des analyses (beaucoup d'options, choix d'outils différents en fonction de la question biologique posée...)  
On focalisera sur le RNA-seq et CHIP-seq pour donner des exemples concrets.
- \*cours 3 : par Jean-Charles Cadoret (IJM) : traitement et analyse des données avec les serveurs GALAXY et autres banques de données**

The right sidebar contains sections for 'RECHERCHE FORUMS', 'DERNIÈRES ANNONCES' (with a notice about a new subject), 'ÉVÉNEMENTS À VENIR', and 'ACTIVITÉ RÉCENTE' (with a report from October 2018).

# Formation continue - DU

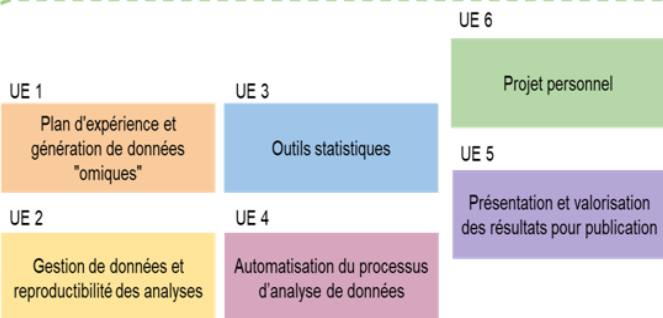
## Création, analyse et valorisation de données "omiques"

Gaëlle Lelandais, Pierre Poulain et Bertrand Cosson



Développer et renforcer des compétences en analyses de données  
expérimentales issues des techniques haut débit

- génomique, transcriptomique et protéomique



Diplôme Universitaire (DU), Université Paris Diderot

## Diplôme Universitaire Bioinformatique Intégrative



Organisateurs : Jacques Van Helden, Hélène Chiapello, Bertrand Cosson



Intégrer différents jeux de données, qui proviennent de différents niveaux d'analyse dans la cellule : génomes, transcriptomes, protéomes, métabolomes, réseaux d'interactions...

- > Environnement Unix
- > Initiation à la programmation et manipulation de données biologiques en Python
- > Analyse statistique avec R
- > Production des données à haut débit et sources de données pour la biologie intégrative
- > Méthodes et outils bioinformatiques pour l'analyse des données à haut débit
- > Bioinformatique intégrative
- > Projet tutoré

Diplôme Universitaire (DU), Université Paris Diderot



# Diplôme Universitaire Bioinformatique Intégrative



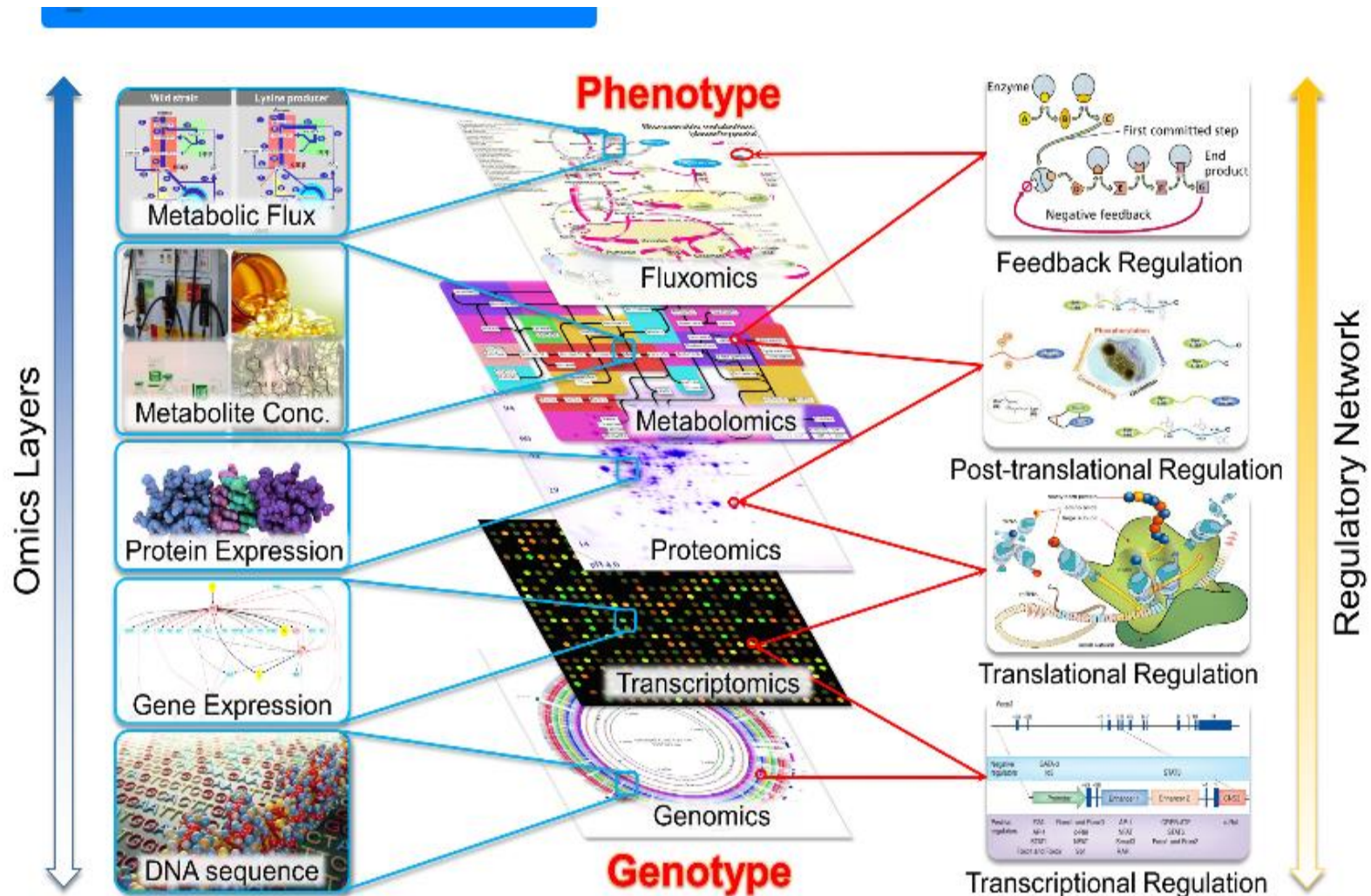
Organisateurs : Jacques Van Helden, Hélène Chiapello, Bertrand Cosson



Intégrer différents jeux de données, qui proviennent de différents niveaux d'analyse dans la cellule : génomes, transcriptomes, protéomes, métabolomes, réseaux d'interactions...

- > Environnement Unix
- > Initiation à la programmation et manipulation de données biologiques en Python
- > Analyse statistique avec R
- > Production des données à haut débit et sources de données pour la biologie intégrative
- > Méthodes et outils bioinformatiques pour l'analyse des données à haut débit
- > Bioinformatique intégrative
- > Projet tutoré

# Integrative bioinformatics - Multi-omics approaches - Data Analysis



# Save the date



≡ MENU

## PARIS7 BIOINFO CLUB

*Bioinformatics Club – IJM, Epigenetics, BFA*

## Next meetings

**13 September, 10am:** Twitter and other communication tools like Altmetrics  
(Pierre Kerner)

**11 October, 10am:** Three short talks on the status of bioinformatics in  
Epigenetics and cell fate unit (Paris Epigenetics, Olivier Kirsh), in BFA (Olivier  
Taboureau) and in IJM (Virginie Courtier-Orgogozo). Free discussion to define  
the themes to be addressed by the BioInfo Club.

**8 November, 10am:** Presentation of R and R Studio (Bénédicte Lefèvre) – bring  
computer

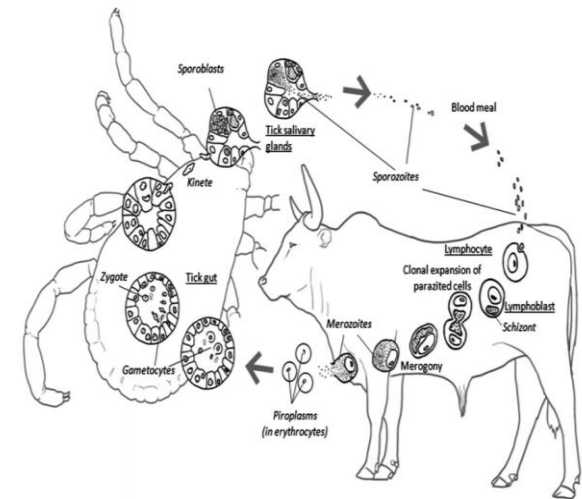
**15 November, 10am:** Presentation of ggplot to make great figures and drawings  
(Samuel Murail) – bring computer

**6 December, 10am:** (Kevin Cheeseman)



The BioInfo Club meets on Thursday from 10am to 11am, in François Jac

**How does an intracellular parasite control both host & parasite gene expression for its survival and the completion of its life cycle?**



## UMR7216 – people

