

Le rôle de la méthylation des protéines dans le cancer du sein

Ekaterina FLIN

Equipe « Régulations Epigénétiques »

Nicolas REYNOIRD

Equipe « Dynamique de la méthylation des protéines dans le cancer »

Plan

1. Contexte scientifique

- a. biologie / épigénétique
- b. médecine / cancer du sein

2. Données d'expression de gènes

- a. données « omiques » / RNA-seq

3. Objectifs du projet tutoré

- a. analyse d'expression
- b. analyse de survie
- c. analyse d'enrichissement

1.a Contexte biologique

Epigénétique

20-ème siècle – La génétique

La **génétique** (du grec « *donner naissance* ») est la science qui étudie **l'hérédité** et **les gènes**.

1909 – notion de gène, une base d'hérédité

1953 – découverte du double hélice d'ADN

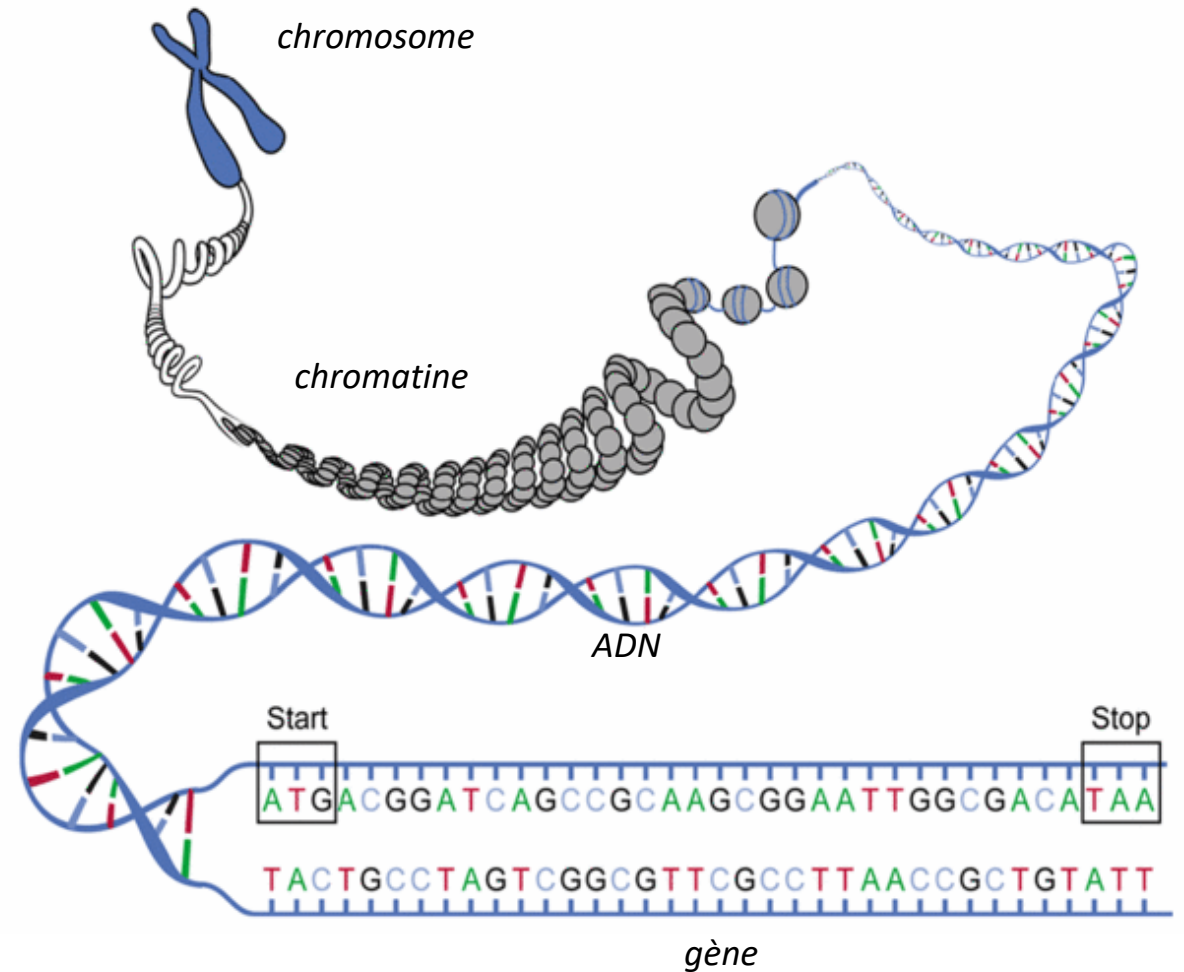
1977 – séquençage des fragments d'ADN

2003 – séquençage complet du génome humain

21-ème siècle – L'épigénétique

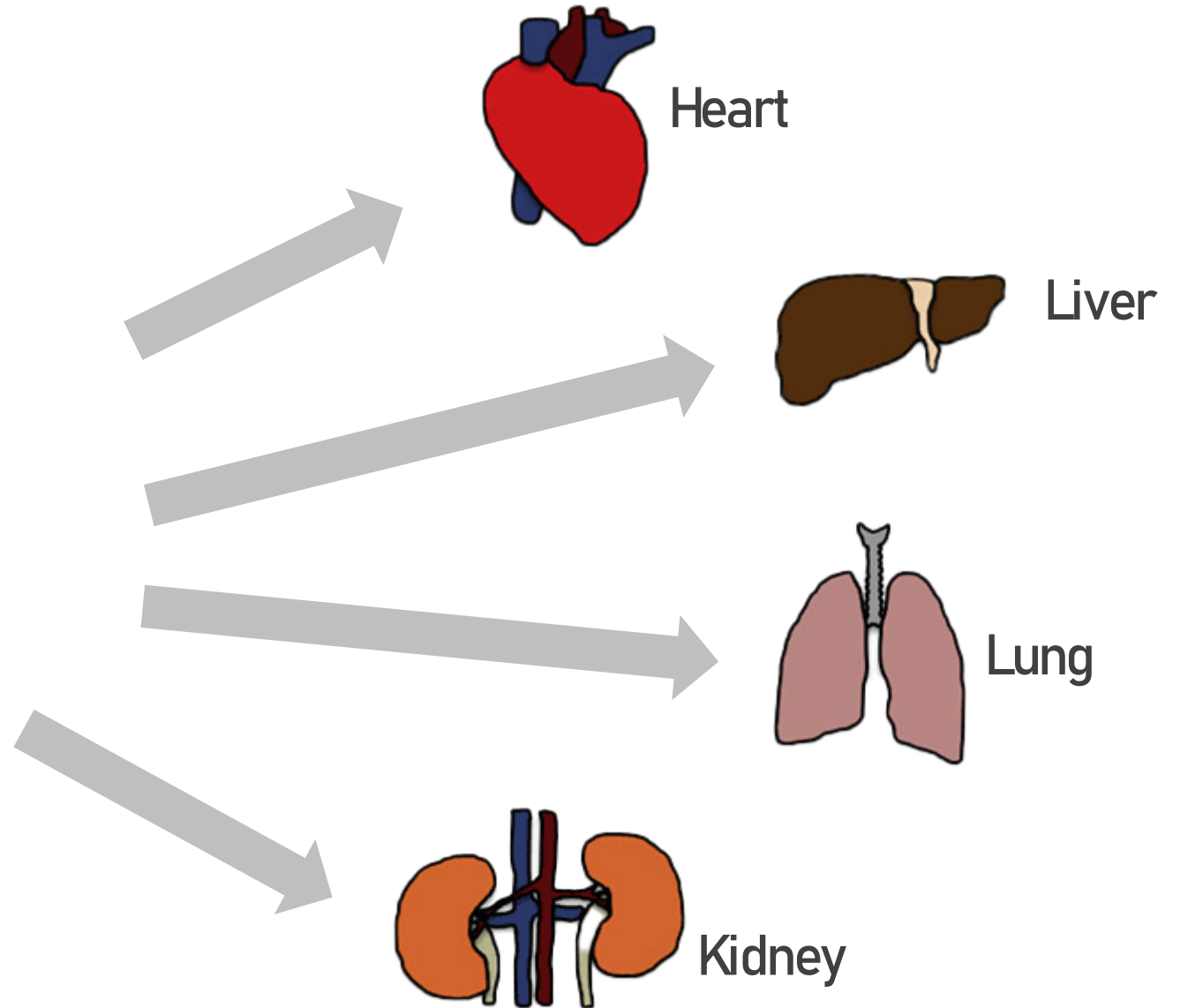
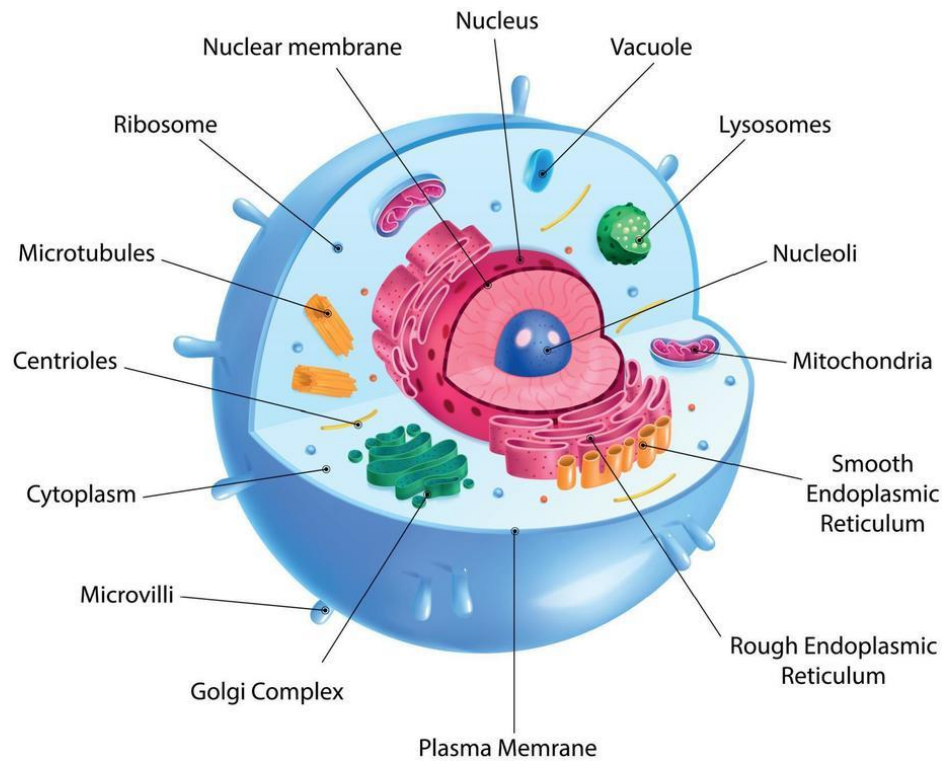
Épi : « *au dessus de* »

L'épigénétique étudie les mécanismes moléculaires qui modifient l'expression des gènes en fonction de l'environnement.

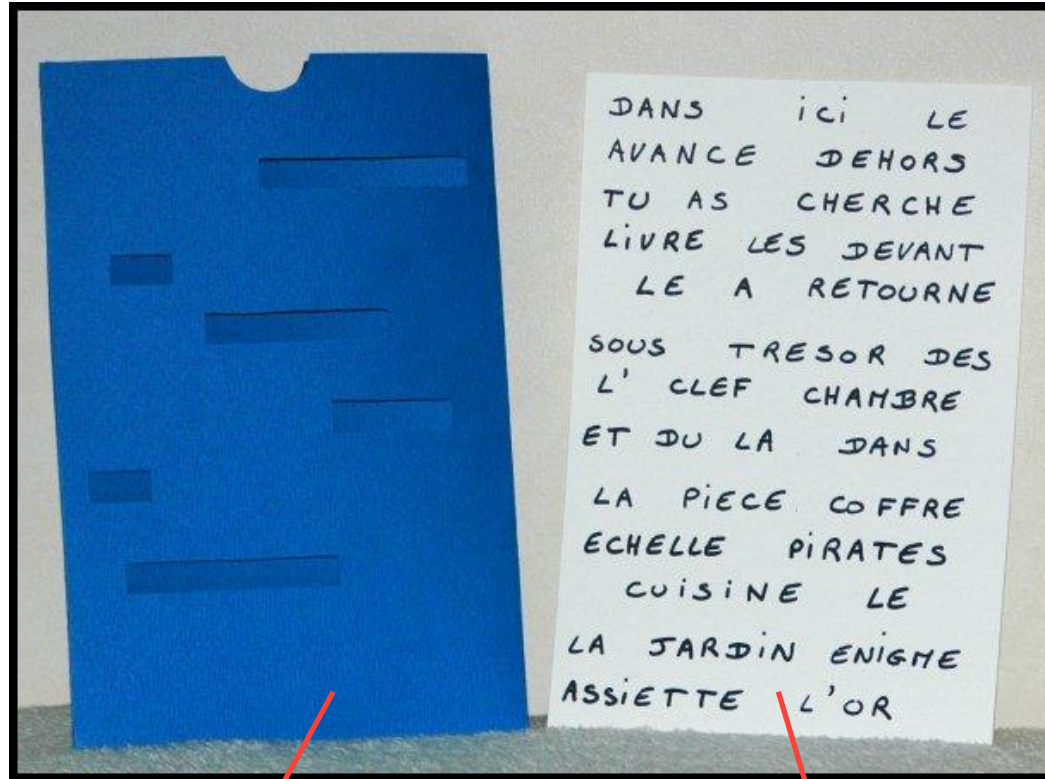


Régulation épigénétique

HUMAN CELL ANATOMY

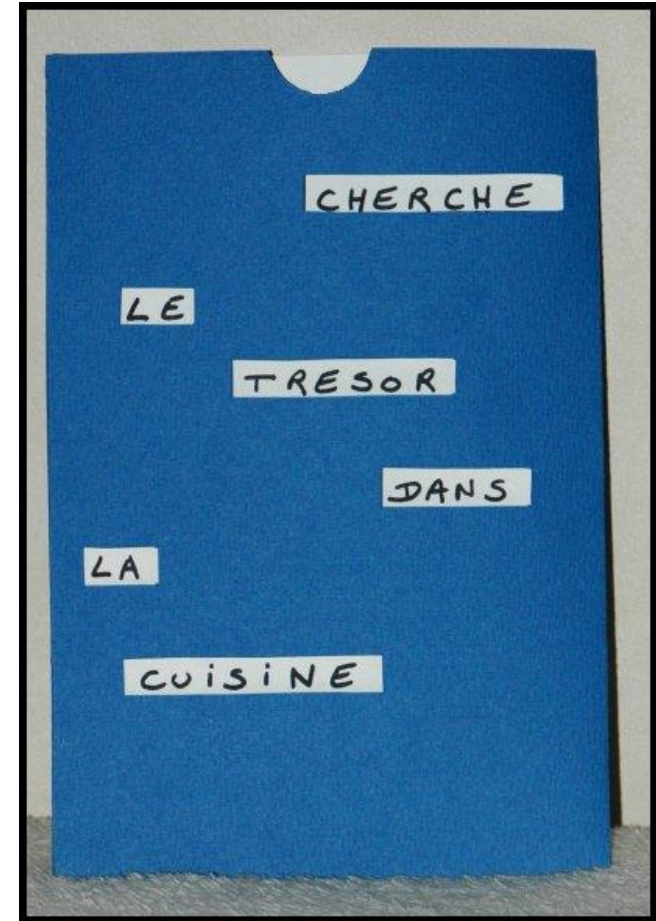


Régulation de l'expression des gènes par l'épigénome



Epigénome

ADN



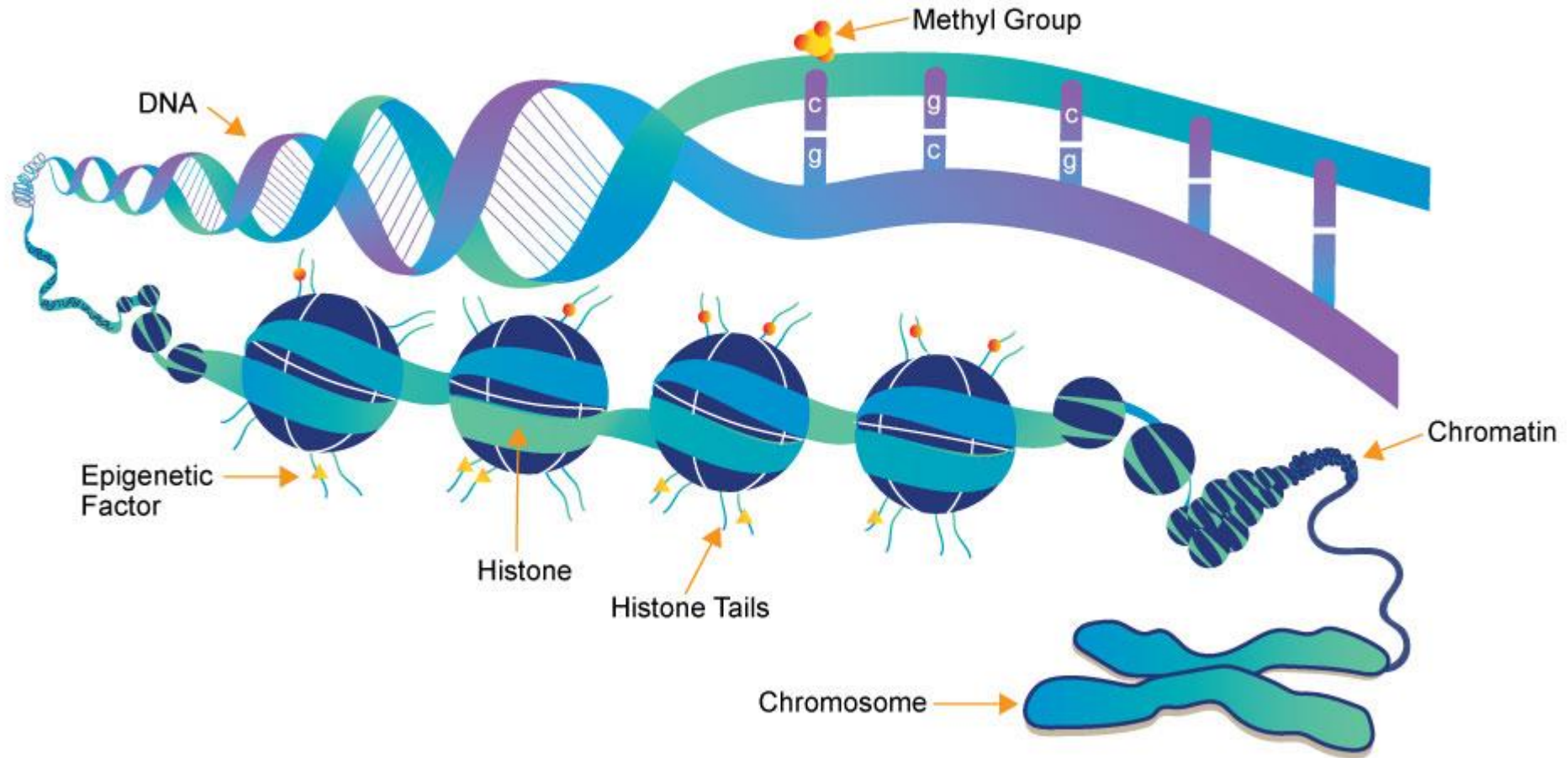
Gène « Avy » **ON**

Gène « Avy » **OFF**

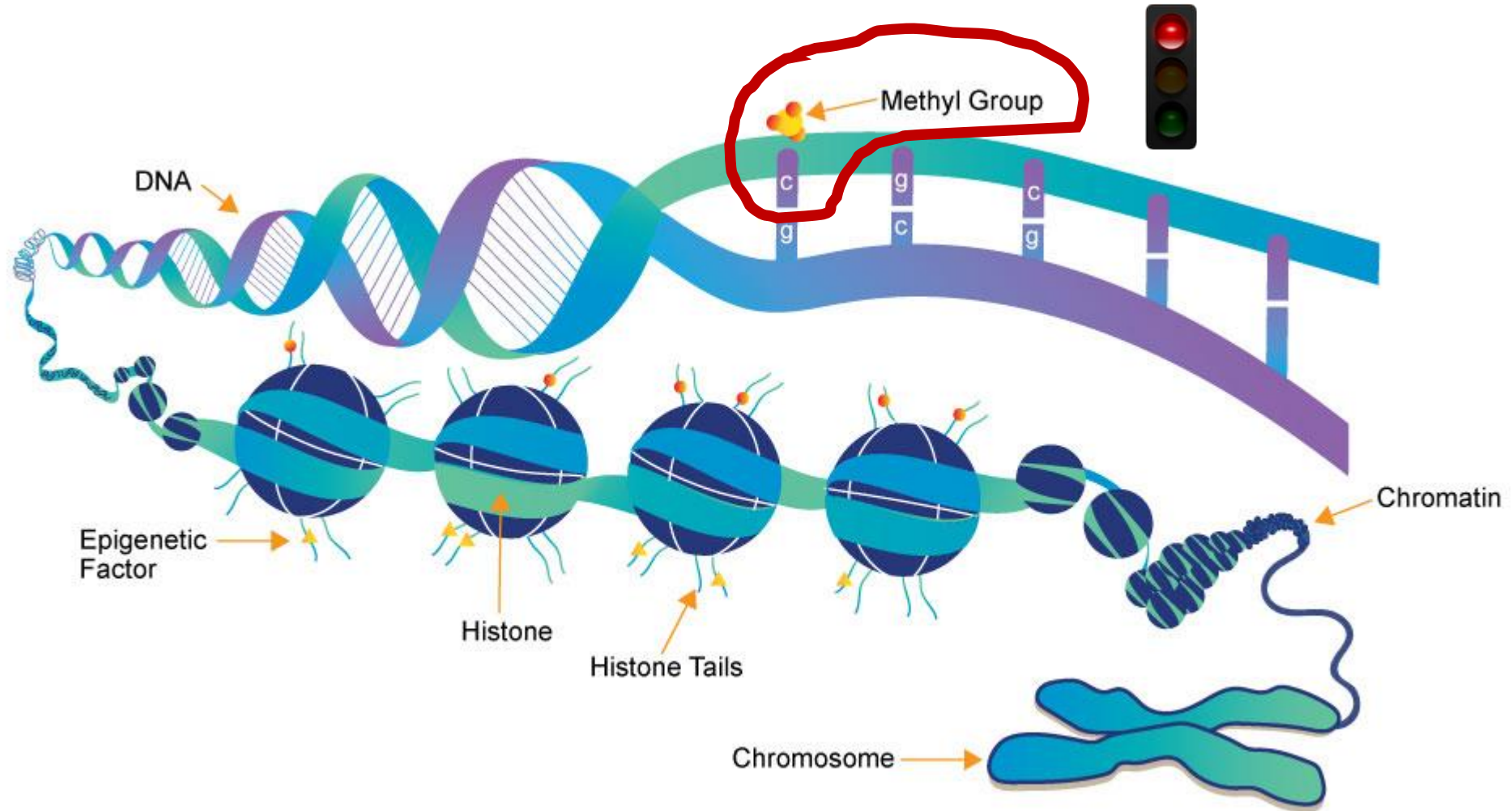


Deux souris génétiquement identiques

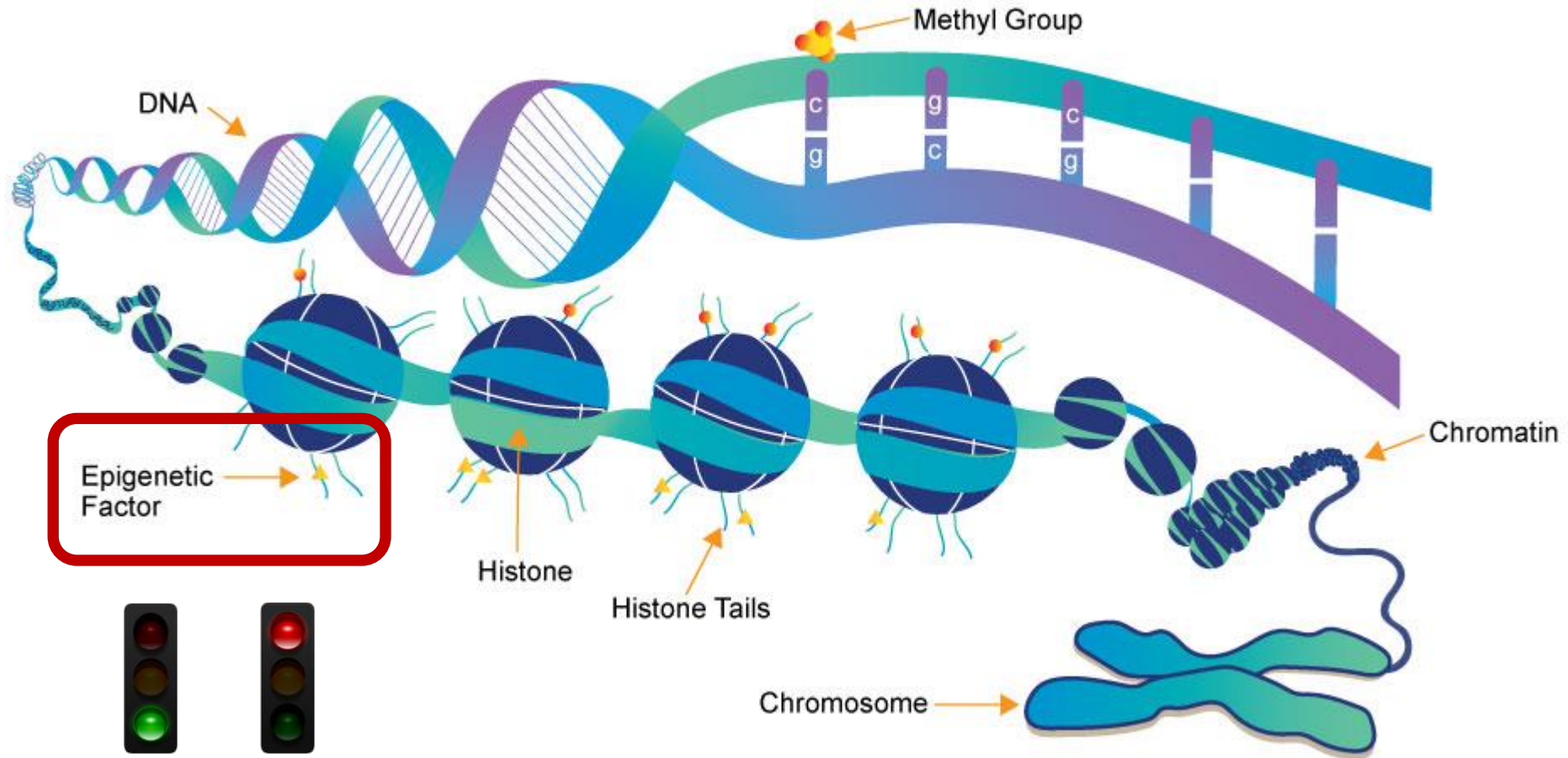
Régulation épigénétique



Régulation épigénétique



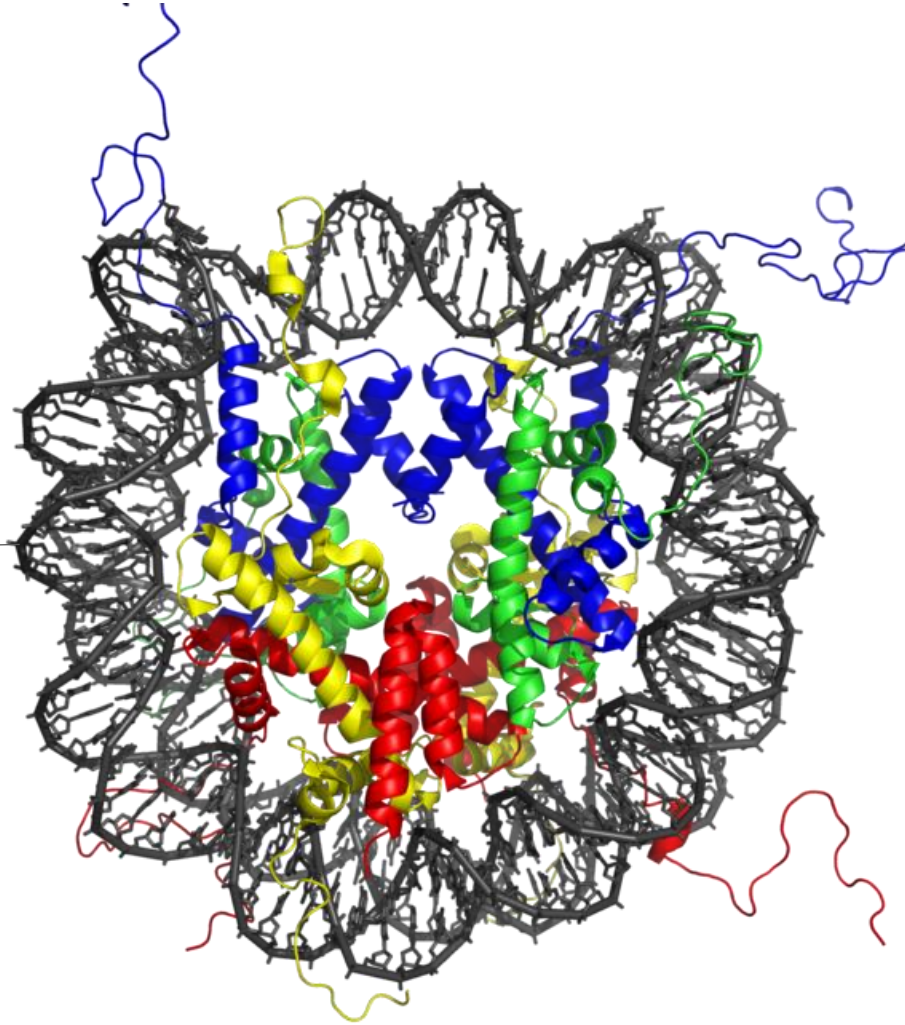
Régulation épigénétique



Régulation épigénétique

Nucléosome

ADN

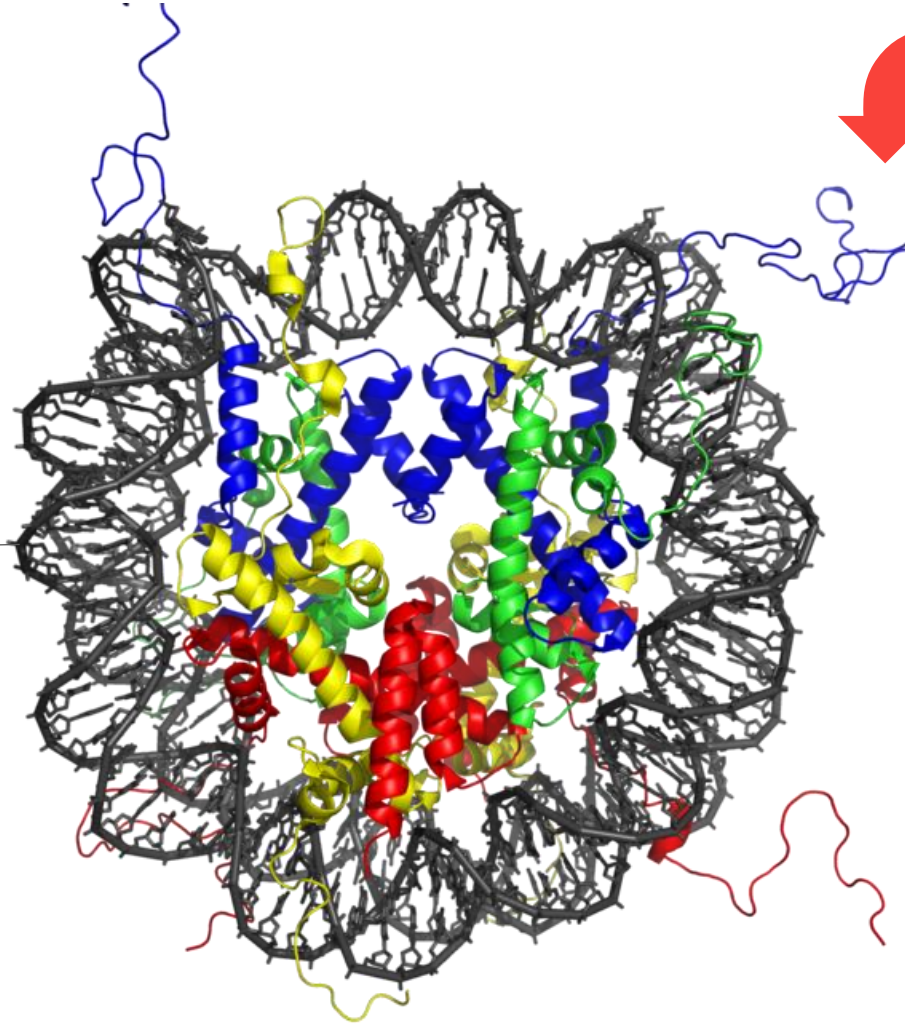


	Histones H2A
	Histones H2B
	Histones H3
	Histones H4

Régulation épigénétique

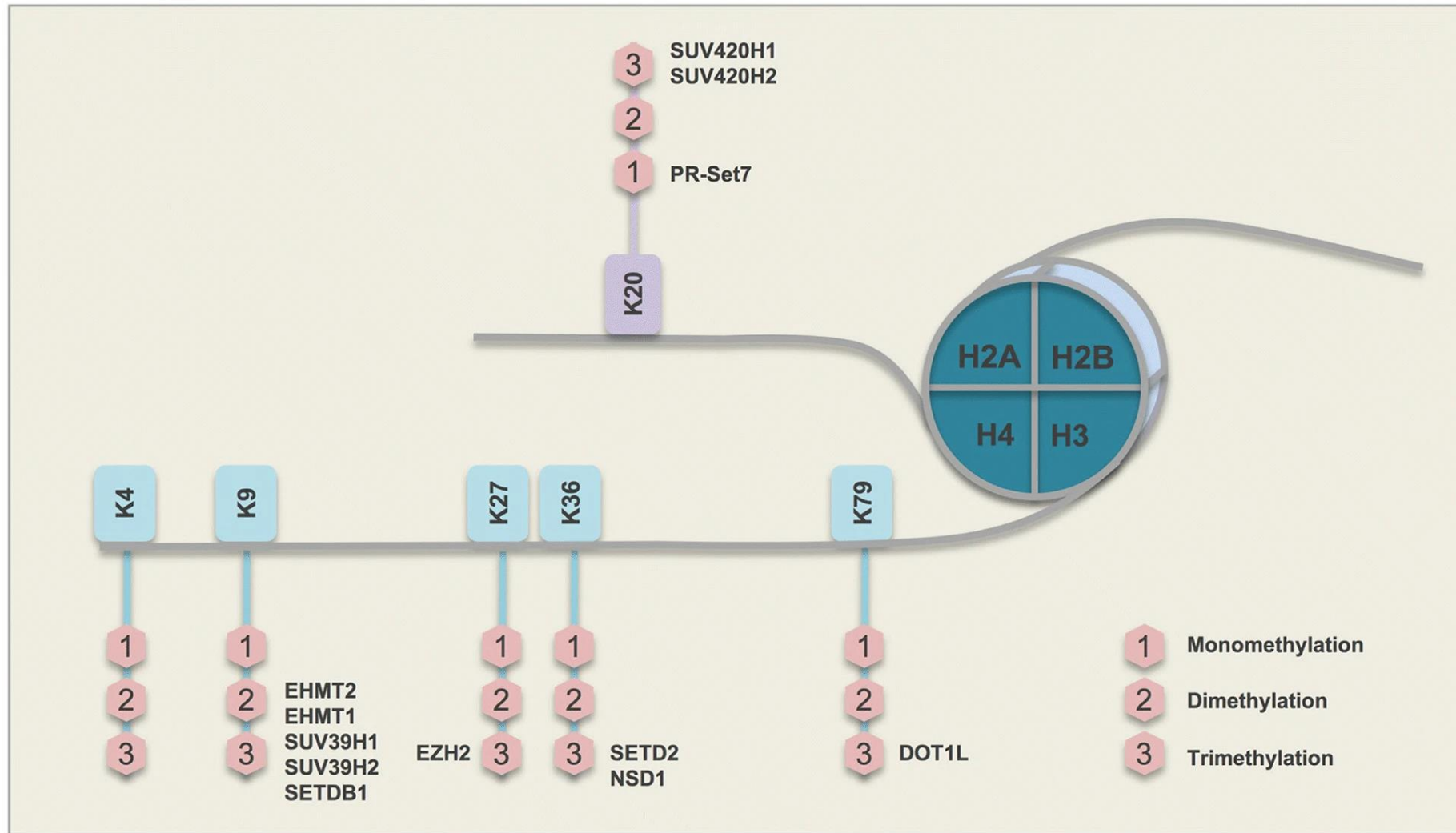
Nucléosome

ADN



	Histones H2A
	Histones H2B
	Histones H3
	Histones H4

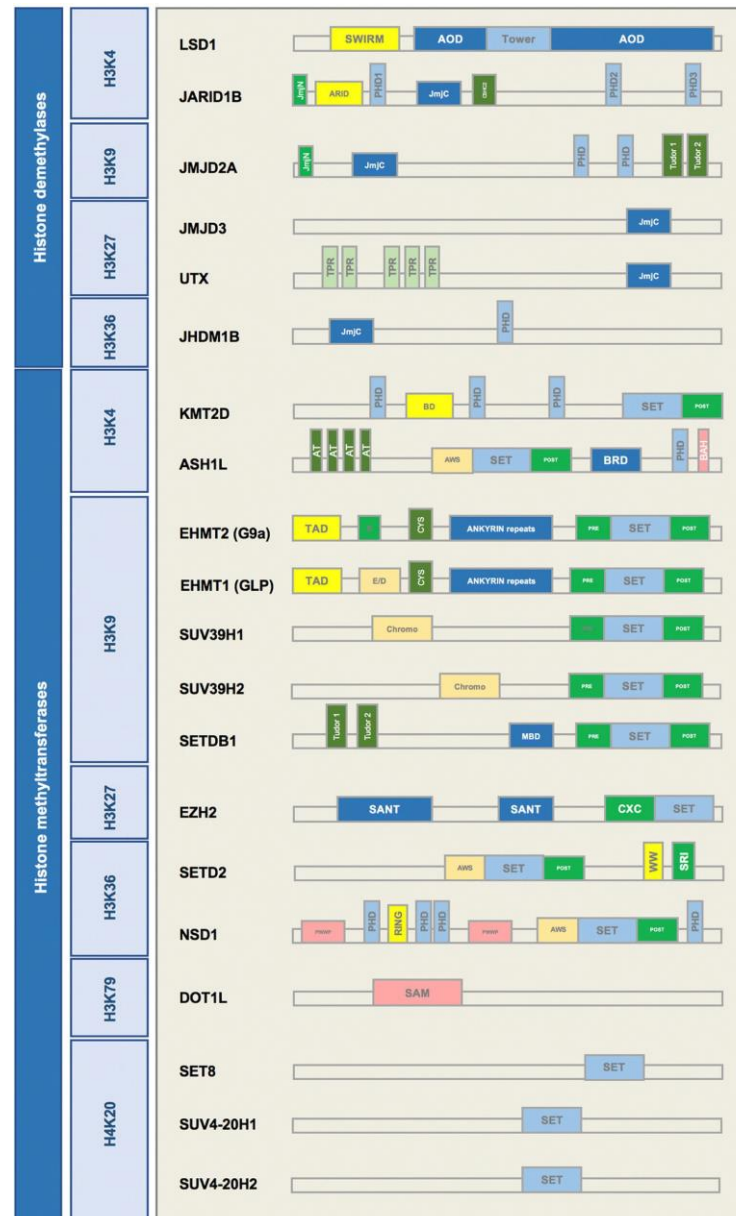
Sites majeurs de méthylation des histones H3 et H4



Enzymes de méthylation des histones H3 et H4

Dé-méthylases
enzymes qui enlèvent CH_3

Méthyltransférases
enzymes qui posent CH_3



Les gènes d'intérêt dans ce projet



Une liste d'enzymes de méthylation a été préparée par Nicolas Reynoird

109 KMT

Lysine méthyltransférases

35 KDM

Lysine dé-méthylases

461 KMB

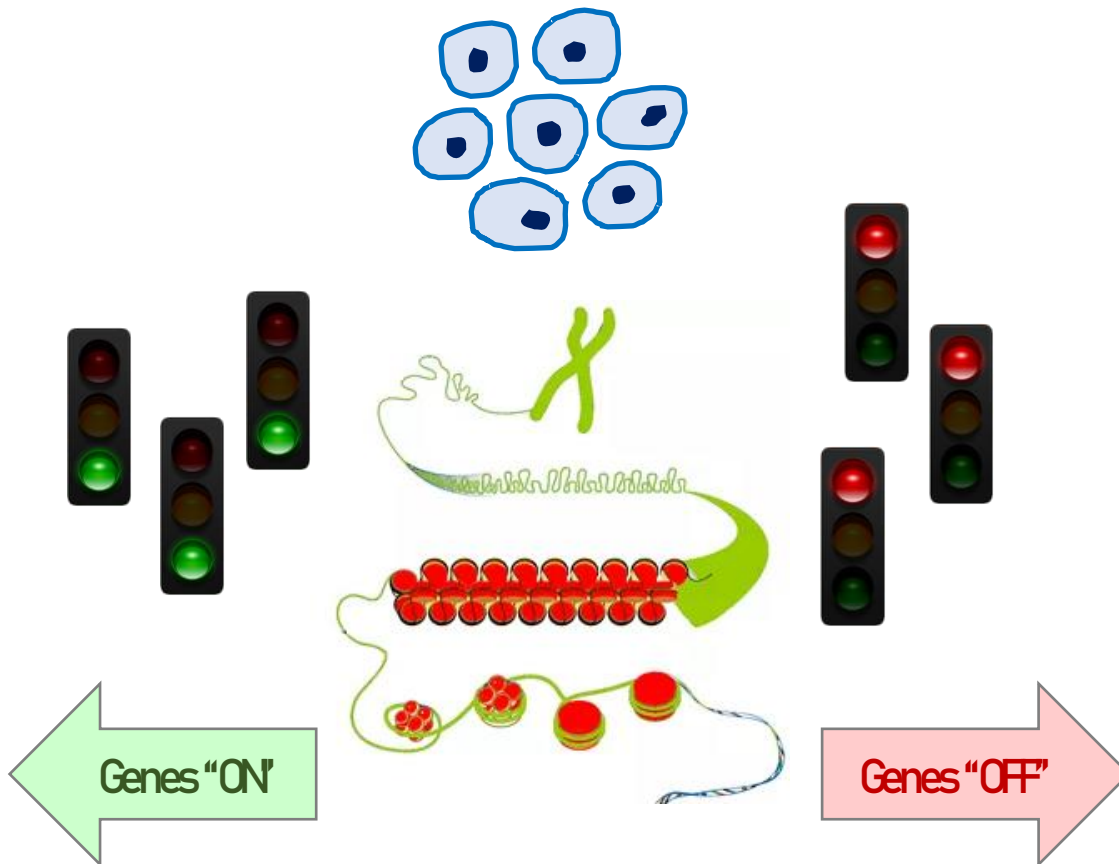
Methyl-lysine binding domain

605 genes au total

1.b Contexte médical

Dérégulations épigénétiques dans le cancer

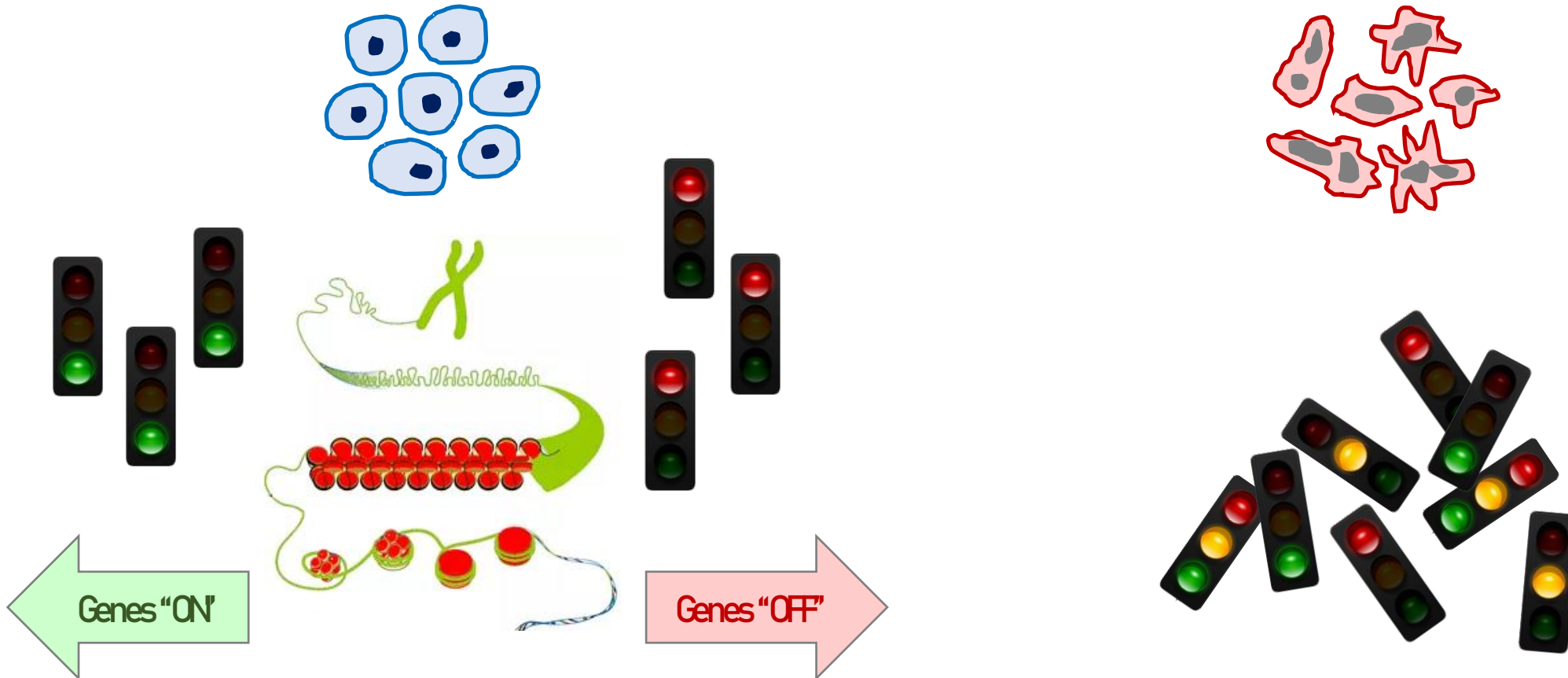
Cellules normales



Dérégulations épigénétiques dans le cancer

Cellules normales

Cellules tumorales



Cancer du sein

- 1-er cancer chez la femme en terme d'incidence (50000 nvx cas / an France)
- 2-ème cancer en terme de mortalité après le cancer du poumon (12000 décès / an France)
- Sous-types moléculaires
- Problème de santé publique

Stages of Breast Cancer



0

Abnormal cells in duct lining or sections of the breast. Increased risk of developing cancer in one or both breasts.

100%
SURVIVAL RATE



1

Cancer in breast tissue. Tumor is less than one inch across in size.

98%
SURVIVAL RATE



2

Cancer in breast tissue. Tumor is less than two inches across in size. Cancer may spread to the auxiliary lymph nodes.

88%
SURVIVAL RATE



3

Tumor is larger than two inches across in size and cancer has spread to auxiliary lymph nodes. Possible dimpling, inflammation or skin color change.

52%
SURVIVAL RATE



4

Cancer has spread beyond the breast to other nearby areas of the body.

16%
SURVIVAL RATE

T – taille
N – ganglions
M – métastases

Sous-types moléculaires du cancer du sein

4 sous-types principaux

Luminal-A

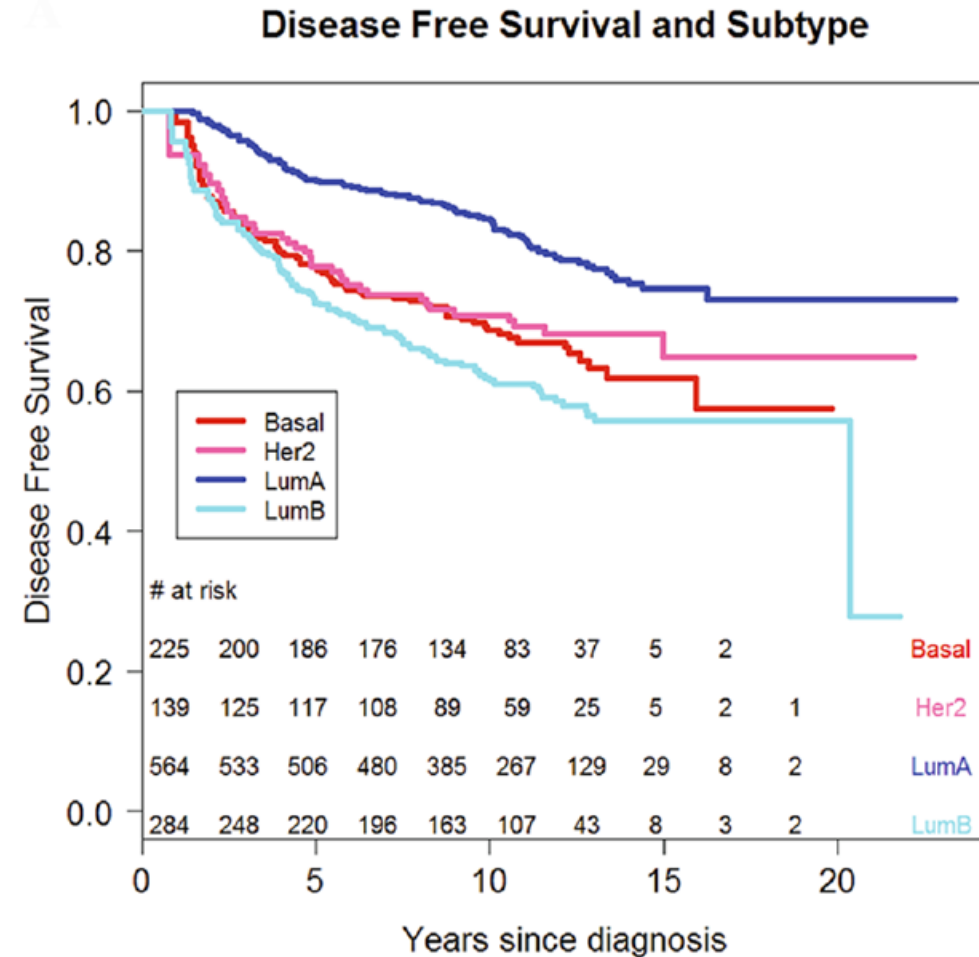
Luminal-B

HER2-enriched

Basal-like

PAM 50

Prediction Analysis for Microarrays



Pu et al., 2020. Breast Cancer Research and Treatment.

Sous-types moléculaires du cancer du sein

4 sous-types principaux

Luminal-A

Luminal-B

HER2-enriched

Basal-like

PAM 50

Prediction Analysis for Microarrays

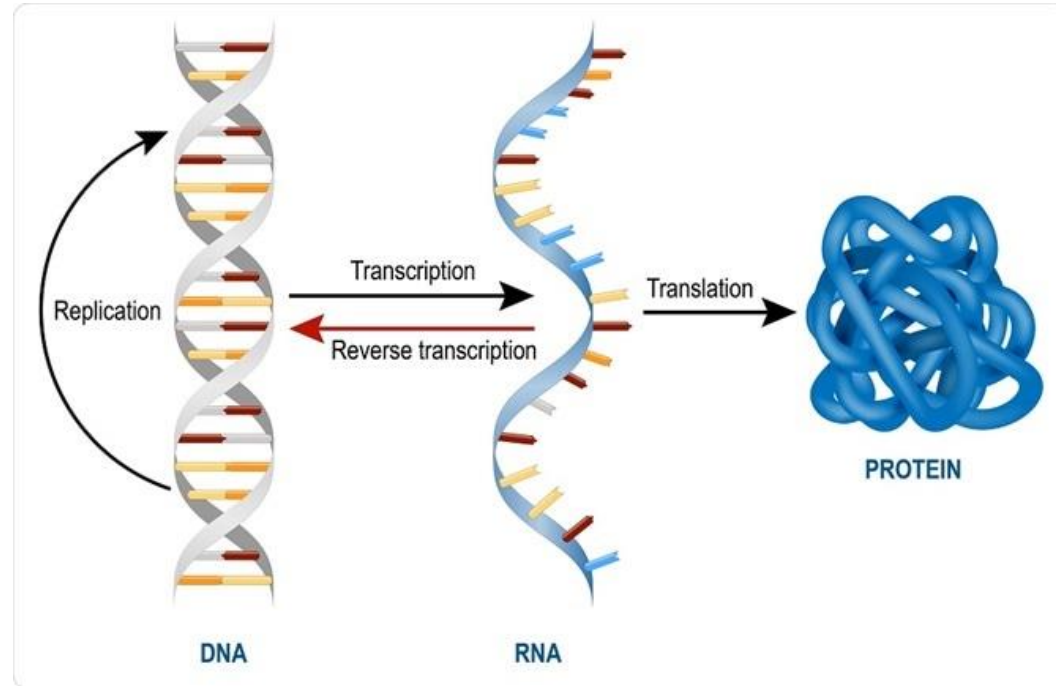


2. Données

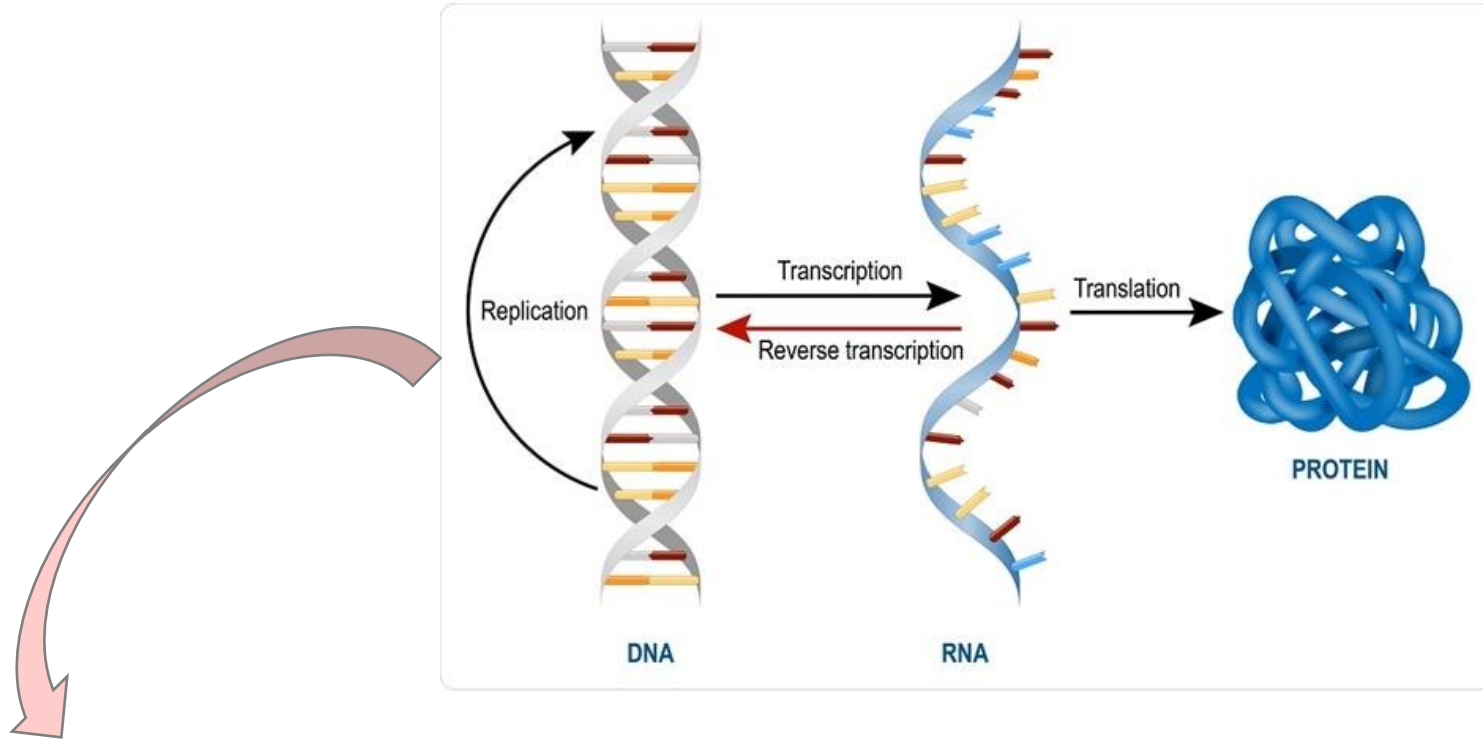
Données « omiques »

Dogme central de la
biologie moléculaire

ADN \Rightarrow ARN \Rightarrow protéine



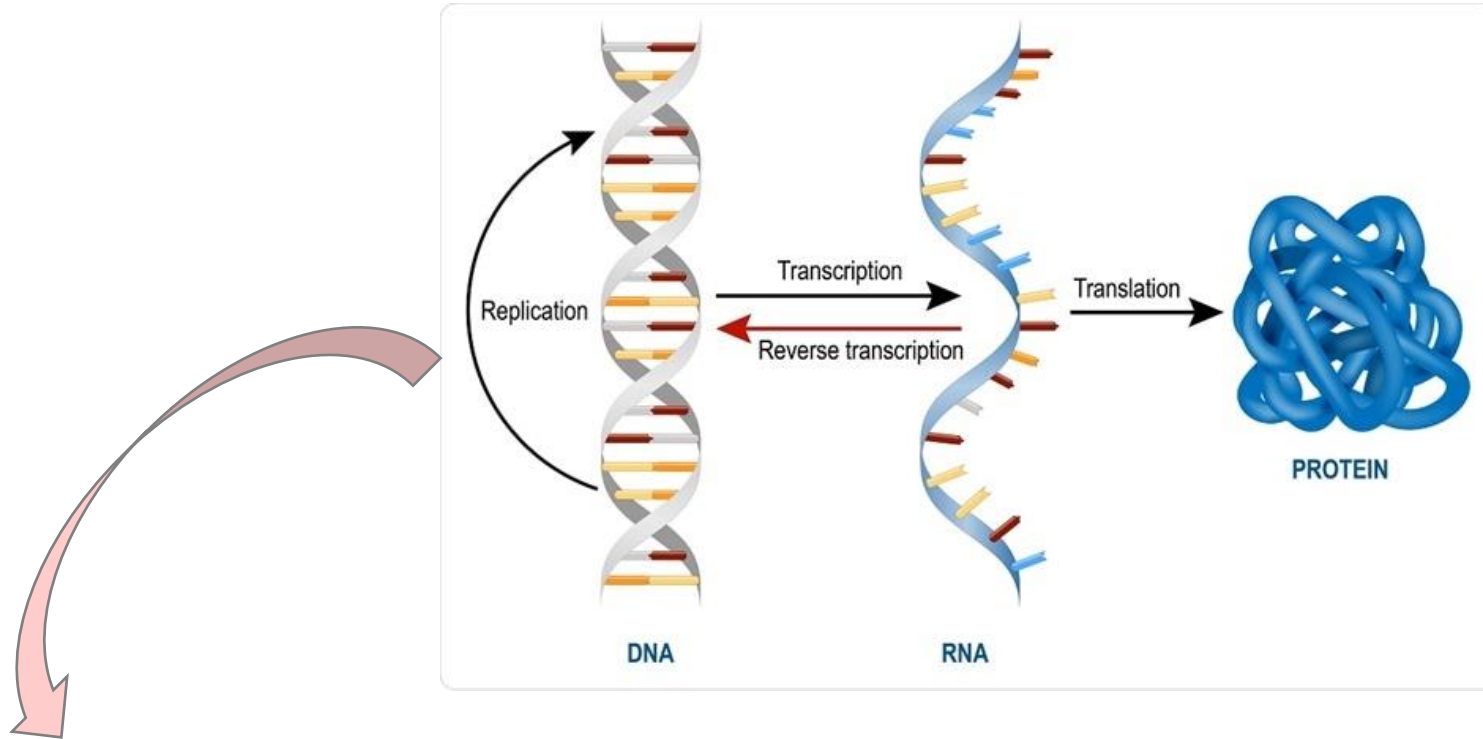
Données « omiques »



Genomics

Gene variants, mutations

Données « omiques »



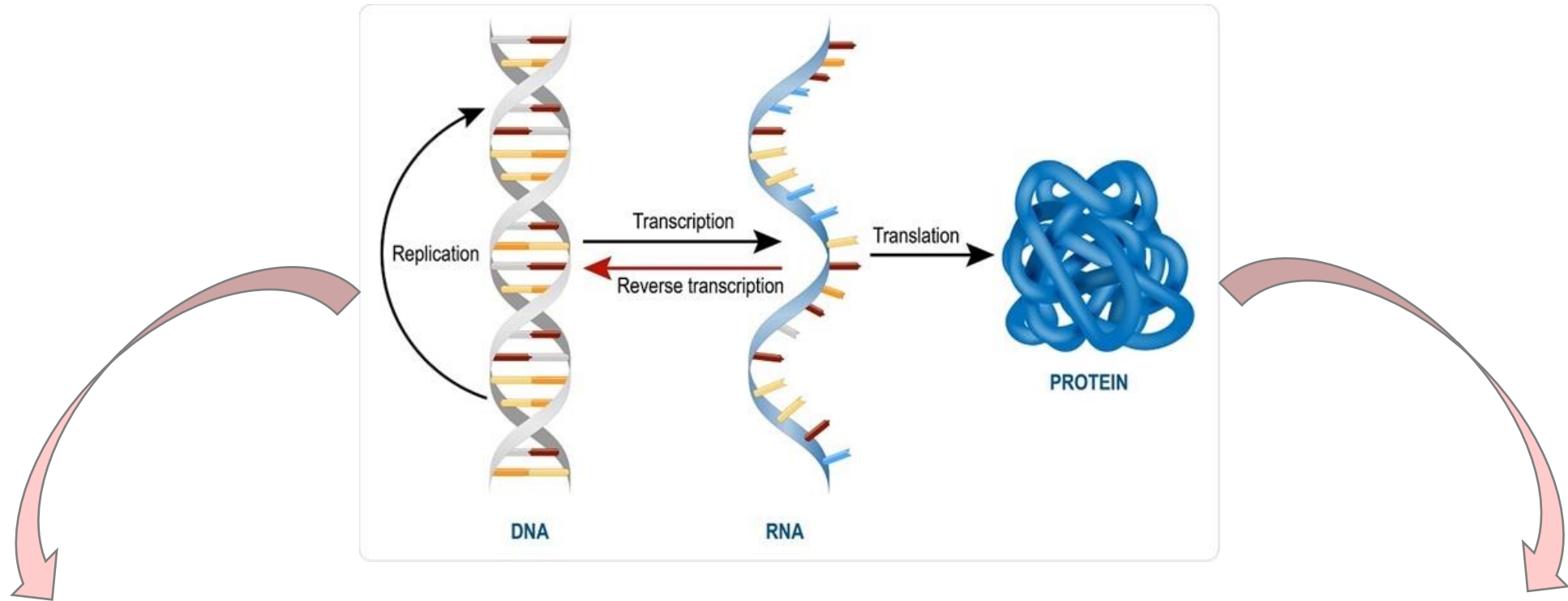
Genomics

Gene variants, mutations

Transcriptomics

Expression levels of genes

Données « omiques »



Genomics

Gene variants, mutations

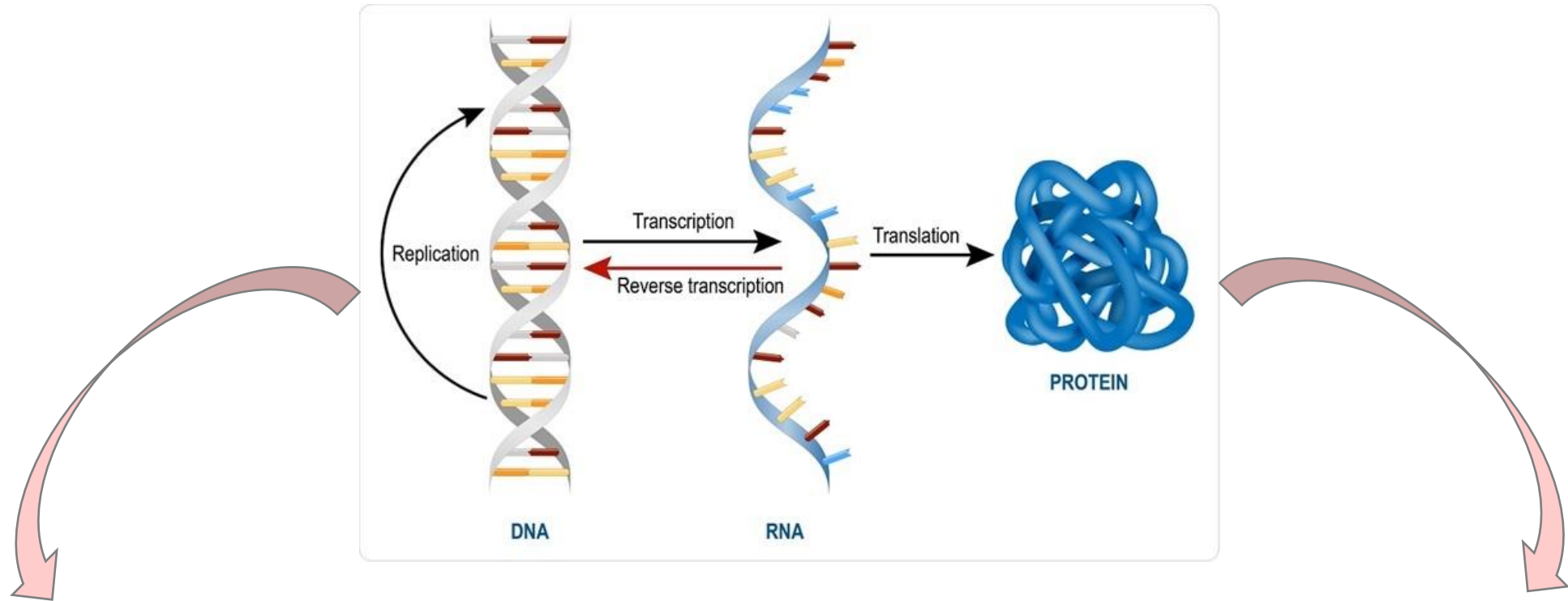
Transcriptomics

Expression levels of genes

Proteomics

Protein abundance

Données « omiques »



Genomics

Gene variants, mutations

Transcriptomics

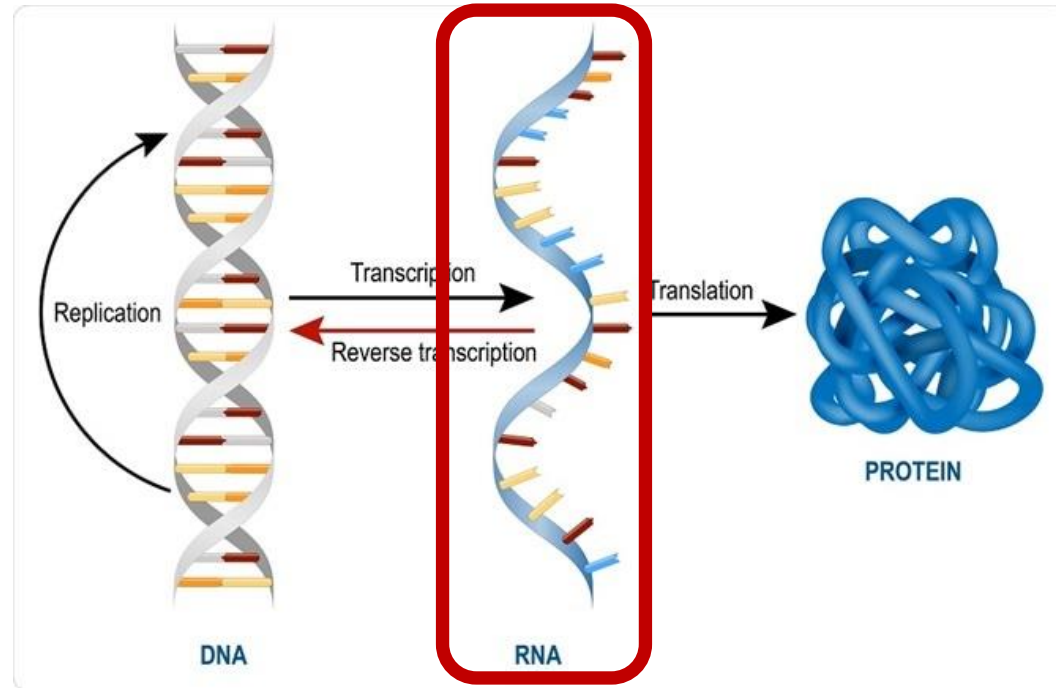
Expression levels of genes

Proteomics

Protein abundance

Epigenomics (methylome), metabolomics, lipidomics, glycomics, ...

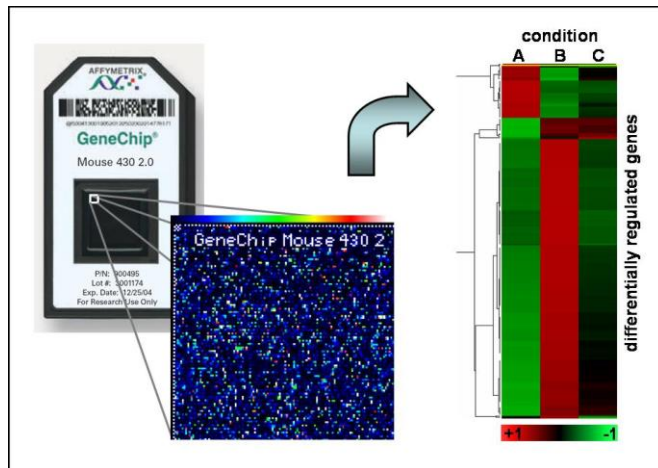
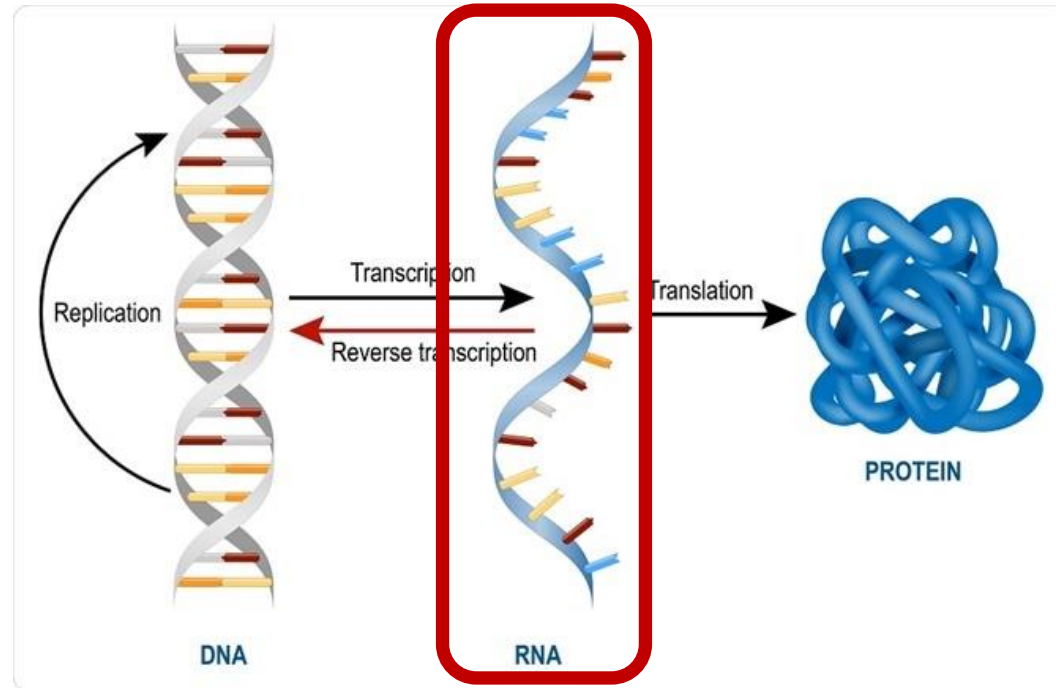
Données transcriptomiques : transcrits **ARN**



Transcriptomics

Expression levels of genes

Données transcriptomiques : transcrits **ARN**



Microarrays

Transcriptomics
Expression levels of genes



New Generation Sequencing (NGS)
RNA-seq

Format de données d'expression de gènes

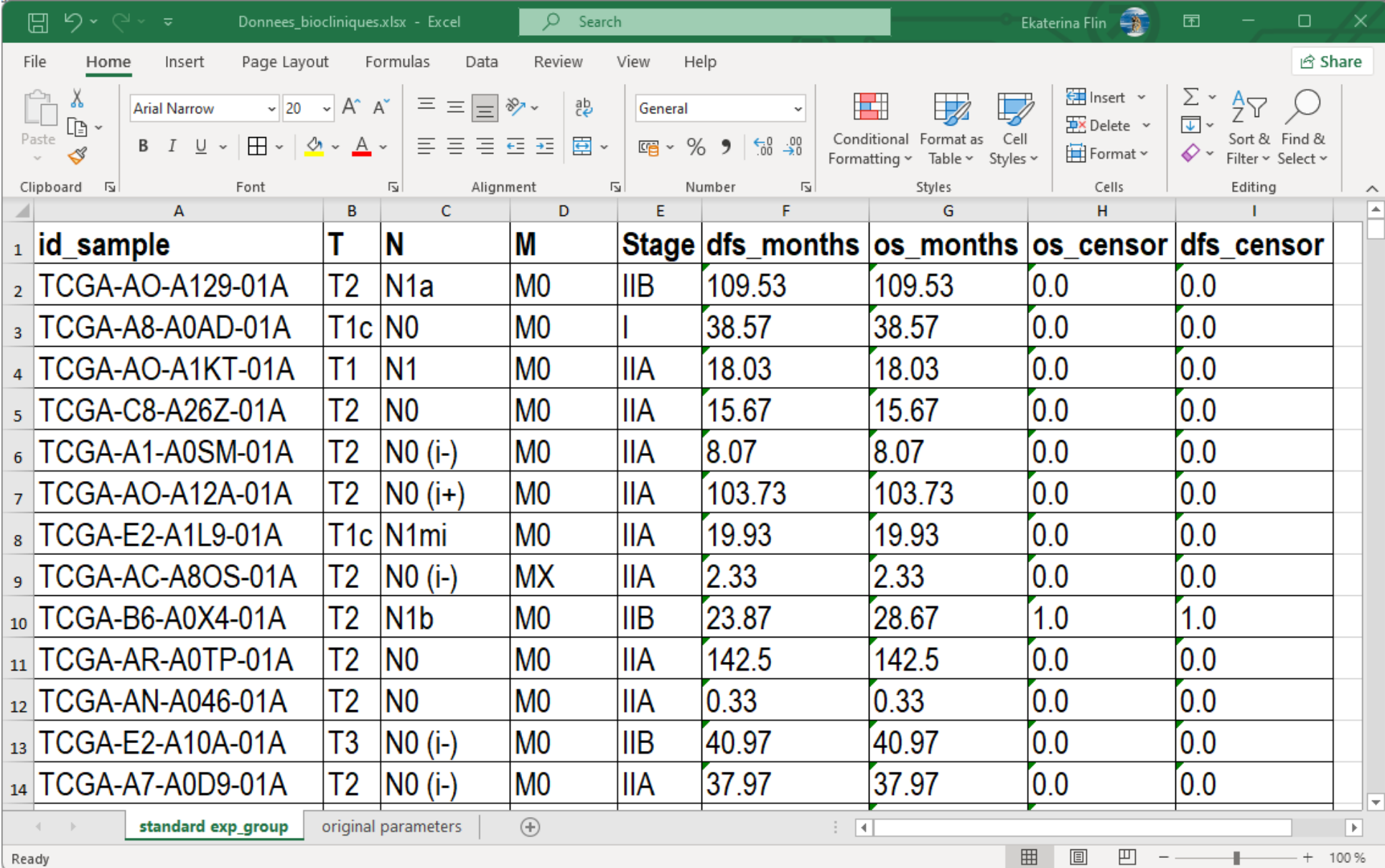


Florent Chuffart

	Identifiants de gènes		Noms de gènes		Identifiants d'échantillons	
	A	B	C	D	E	
1	id_gene	gene_symbol	TCGA-3C-AAAU-01A	TCGA-3C-AALI-01A	TCGA-3C-AALJ-01A	
2		1 A1BG	0.42708	0.443246	0.475529	
3		2 A2M	5.593322	5.9254	6.150867	
4		3 A2MP1	0.0	0.063505	0.165145	
5		9 NAT1	7.100449	3.45364	4.455574	
6		10 NAT2	0.568351	2.191191	0.04665	
7		11 NATP	1.174725	0.083619	0.0	
8		12 SERPINA3	1.244737	0.015039	1.415933	
9		13 AADAC	0.0	0.044619	0.037096	
10		14 AAMP	5.187316	5.669778	5.288264	
11		15 AANAT	0.01132	0.455035	0.187601	
12		16 AARS1	5.740998	5.751416	4.964529	
13		18 ABAT	2.156871	1.220817	1.623974	
14		19 ABCA1	1.282915	1.592122	2.130923	

Valeurs d'expression pour chaque couple (gène, échantillon).
Une valeur d'expression correspond à $\log_2(1+\text{FPKM})$.

Format de données biocliniques



Donnees_biocliniques.xlsx - Excel

File Home Insert Page Layout Formulas Data Review View Help

Clipboard Font Alignment Number Styles Cells Editing

Share

	A	B	C	D	E	F	G	H	I
	id_sample	T	N	M	Stage	dfs_months	os_months	os_censor	dfs_censor
1	TCGA-AO-A129-01A	T2	N1a	M0	IIB	109.53	109.53	0.0	0.0
2	TCGA-A8-A0AD-01A	T1c	N0	M0	I	38.57	38.57	0.0	0.0
3	TCGA-AO-A1KT-01A	T1	N1	M0	IIA	18.03	18.03	0.0	0.0
4	TCGA-C8-A26Z-01A	T2	N0	M0	IIA	15.67	15.67	0.0	0.0
5	TCGA-A1-A0SM-01A	T2	N0 (i-)	M0	IIA	8.07	8.07	0.0	0.0
6	TCGA-AO-A12A-01A	T2	N0 (i+)	M0	IIA	103.73	103.73	0.0	0.0
7	TCGA-E2-A1L9-01A	T1c	N1mi	M0	IIA	19.93	19.93	0.0	0.0
8	TCGA-AC-A8OS-01A	T2	N0 (i-)	MX	IIA	2.33	2.33	0.0	0.0
9	TCGA-B6-A0X4-01A	T2	N1b	M0	IIB	23.87	28.67	1.0	1.0
10	TCGA-AR-A0TP-01A	T2	N0	M0	IIA	142.5	142.5	0.0	0.0
11	TCGA-AN-A046-01A	T2	N0	M0	IIA	0.33	0.33	0.0	0.0
12	TCGA-E2-A10A-01A	T3	N0 (i-)	M0	IIB	40.97	40.97	0.0	0.0
13	TCGA-A7-A0D9-01A	T2	N0 (i-)	M0	IIA	37.97	37.97	0.0	0.0
14									

standard exp_group original parameters

Ready

100 %

8 jeux de données du cancer du sein

Dataset	Technologie
TCGA-BRCA	RNA-seq
GSE25066	Microarrays
GSE21653	Microarrays
GSE42568	Microarrays
Yau-2010 (PMID 16936776)	Microarrays
E-MTAB-365	Microarrays
Miller-2005 (alias GSE3494, GSE4922)	Microarrays
Naderi-Caldas-2007 (PMID 20946665)	Microarrays

3. Objectifs du projet

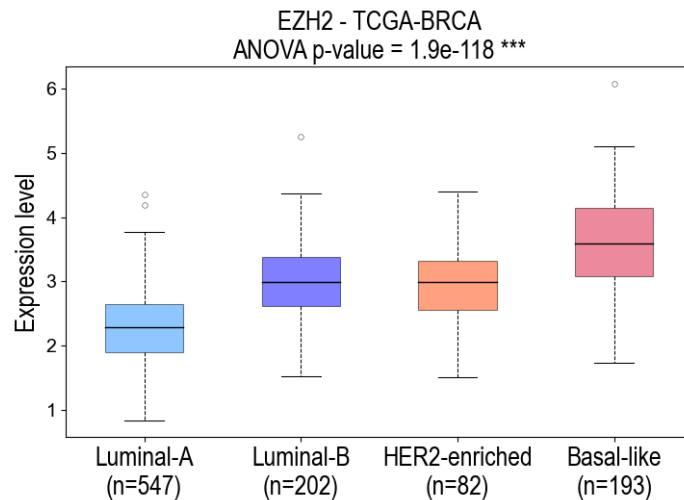
Objectif principal

Etudier le rôle des **acteurs de méthylation des lysines** (enzymes KMT, KDM, KMB) dans **le cancer du sein** à partir des données d'expression de gènes.

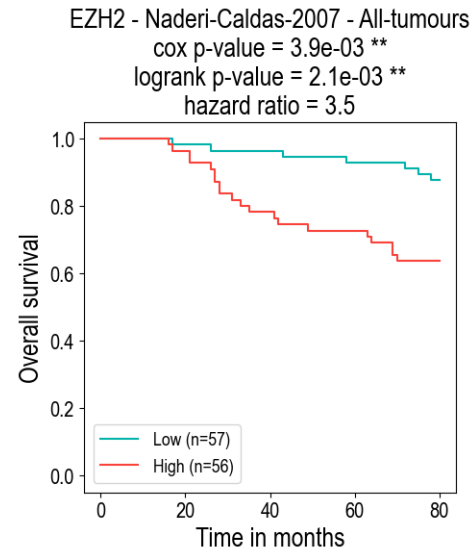
Définir leur **valeur pronostique** dans le cancer du sein et la possibilité de les utiliser en tant que **cibles thérapeutiques**.

Tâches concrètes

1. Analyse d'expression



2. Analyse de survie



3. Analyse d'enrichissement



Voir le document séparé « Cahier des charges »