Computational translational epigenetics: concept-driven omics analyses

Part 1: The awakening of silent genes in malignancies : a new biomarker discovery strategy

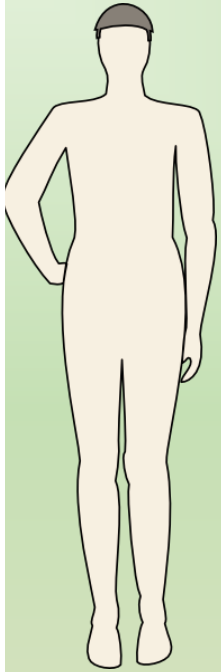
Part 2: Concept driven omics analyses : EpiMed information system and pipelines

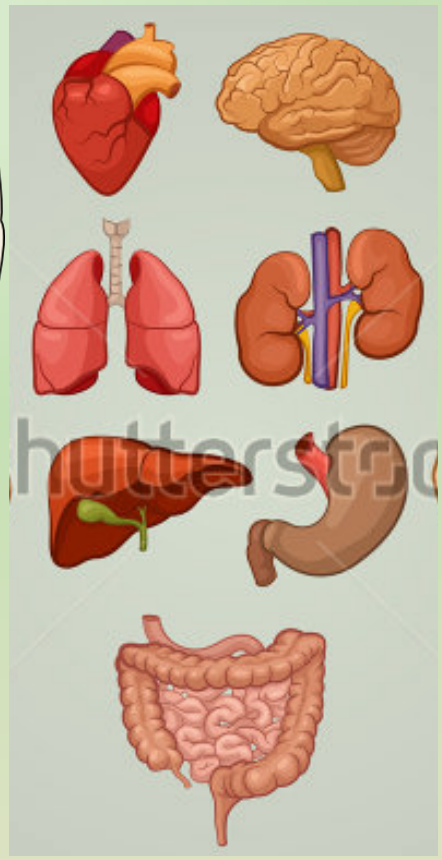
Part 1: The awakening of silent genes in malignancies : a new biomarker discovery strategy

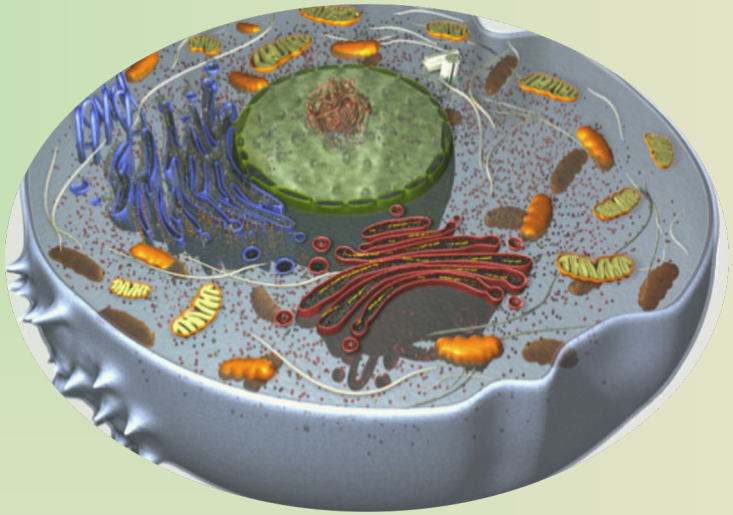
Sophie Rousseaux
Saadi Khochbin's team, EpiMed

Institute for **Advanced Biosciences**
CENTRE DE RECHERCHE UGA – INSERM U 1209 – CNRS UMR 5309

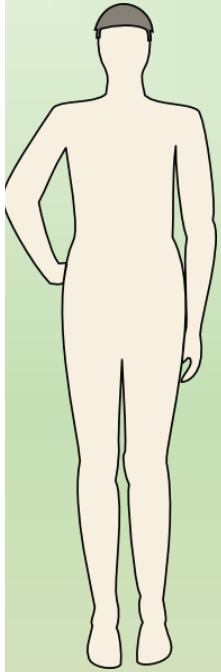
June 2020

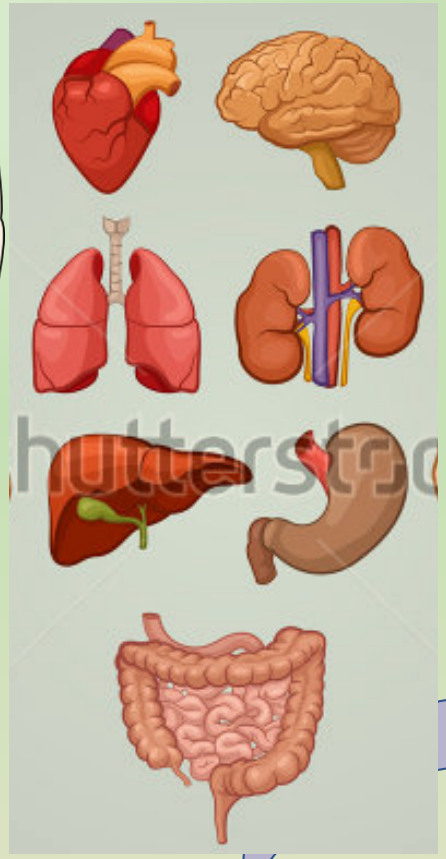


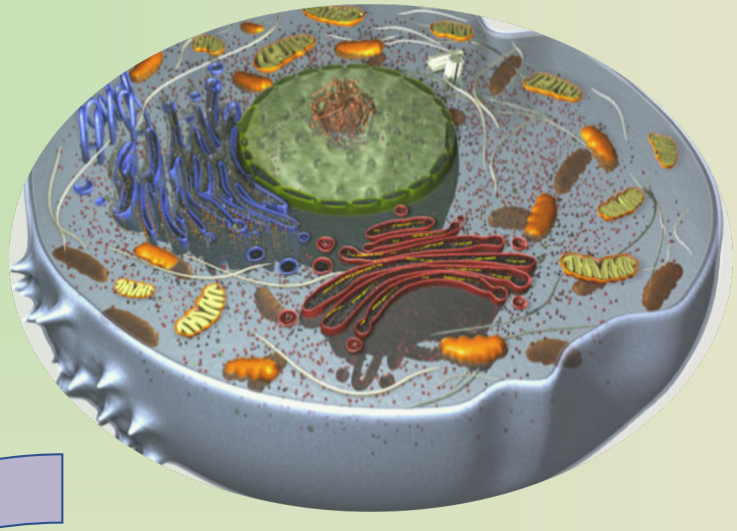




Cell identity

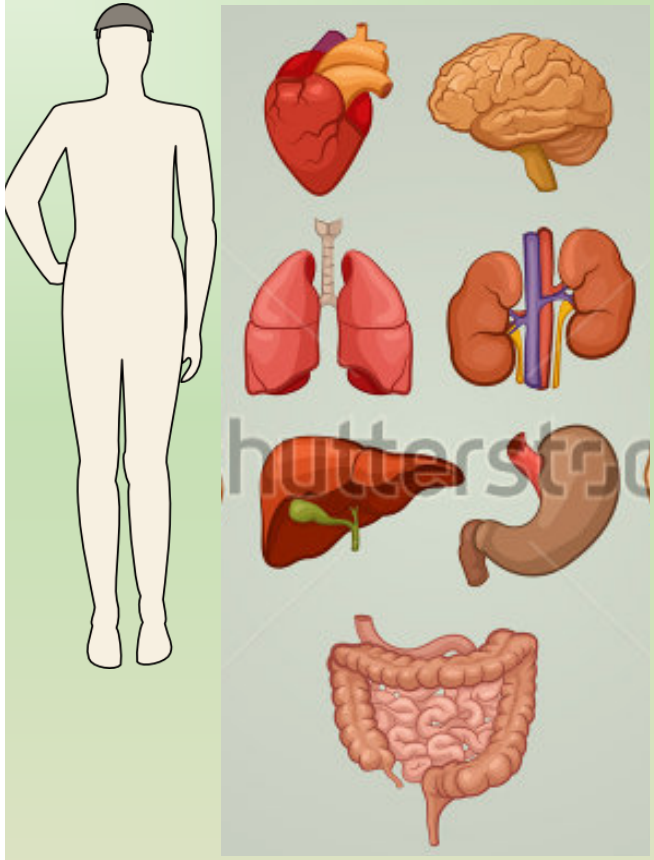




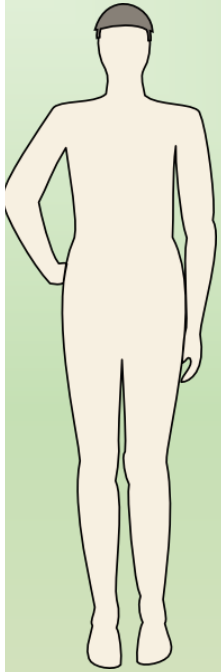


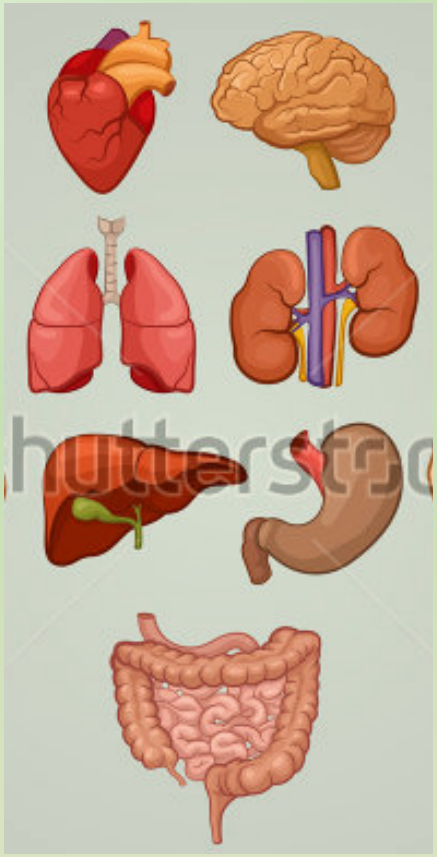
Cell identity/function is defined by

specific gene expression programs

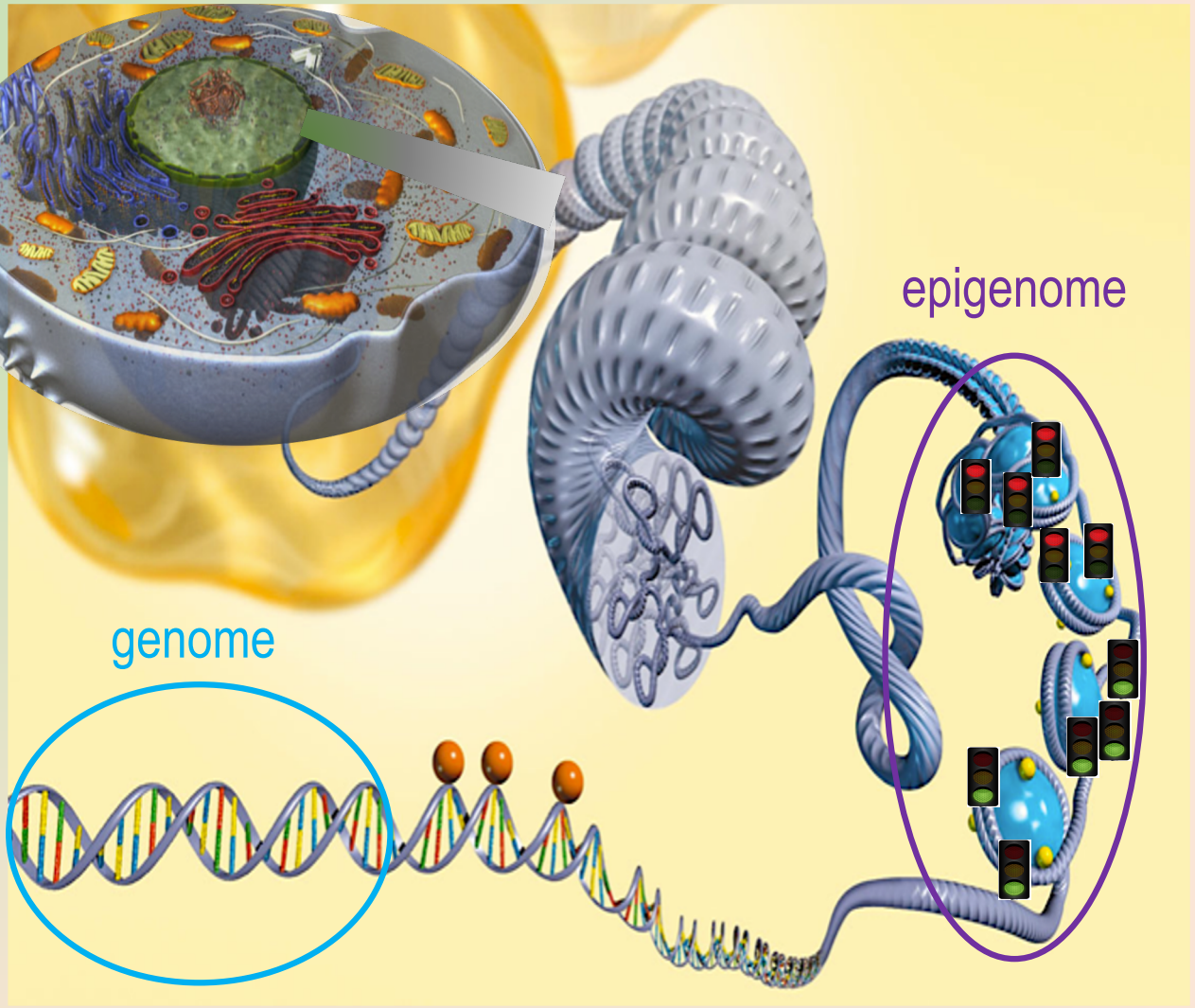


Gene expression programs are controlled by
the epigenome = genome signposting system





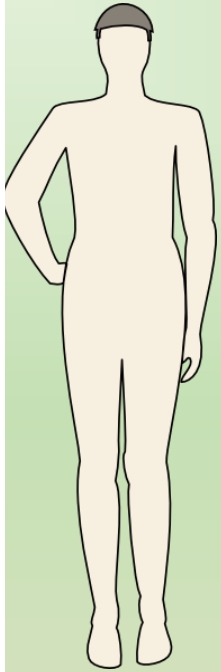
Gene expression programs are controlled by the epigenome = genome signposting system

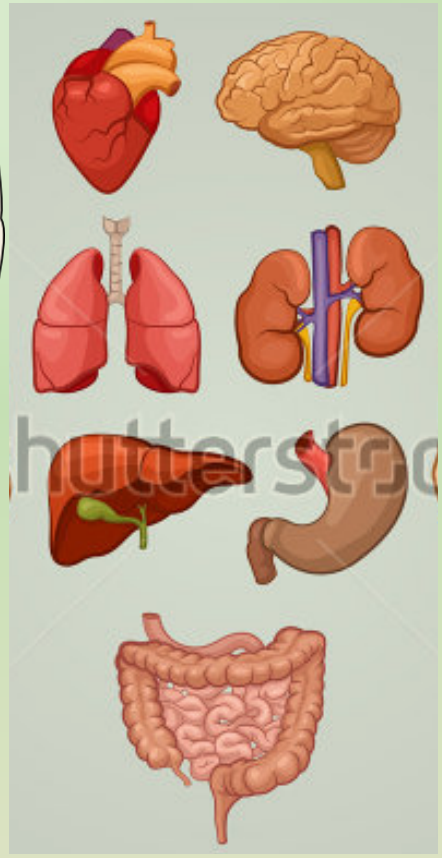


Epigenome = genome signposting system

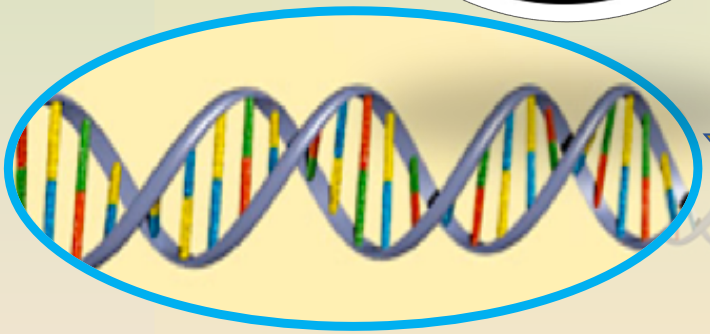
= chemical modifications of DNA and histones

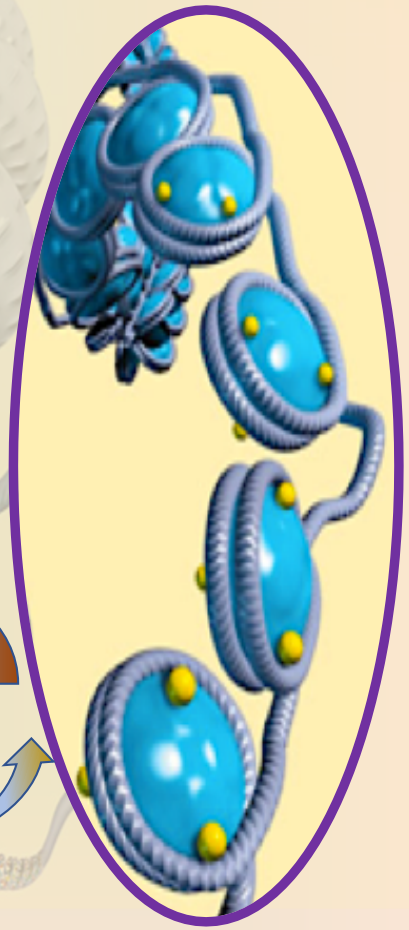
- Stabilizes a specific pattern of gene expression
- A barrier to reprogramming





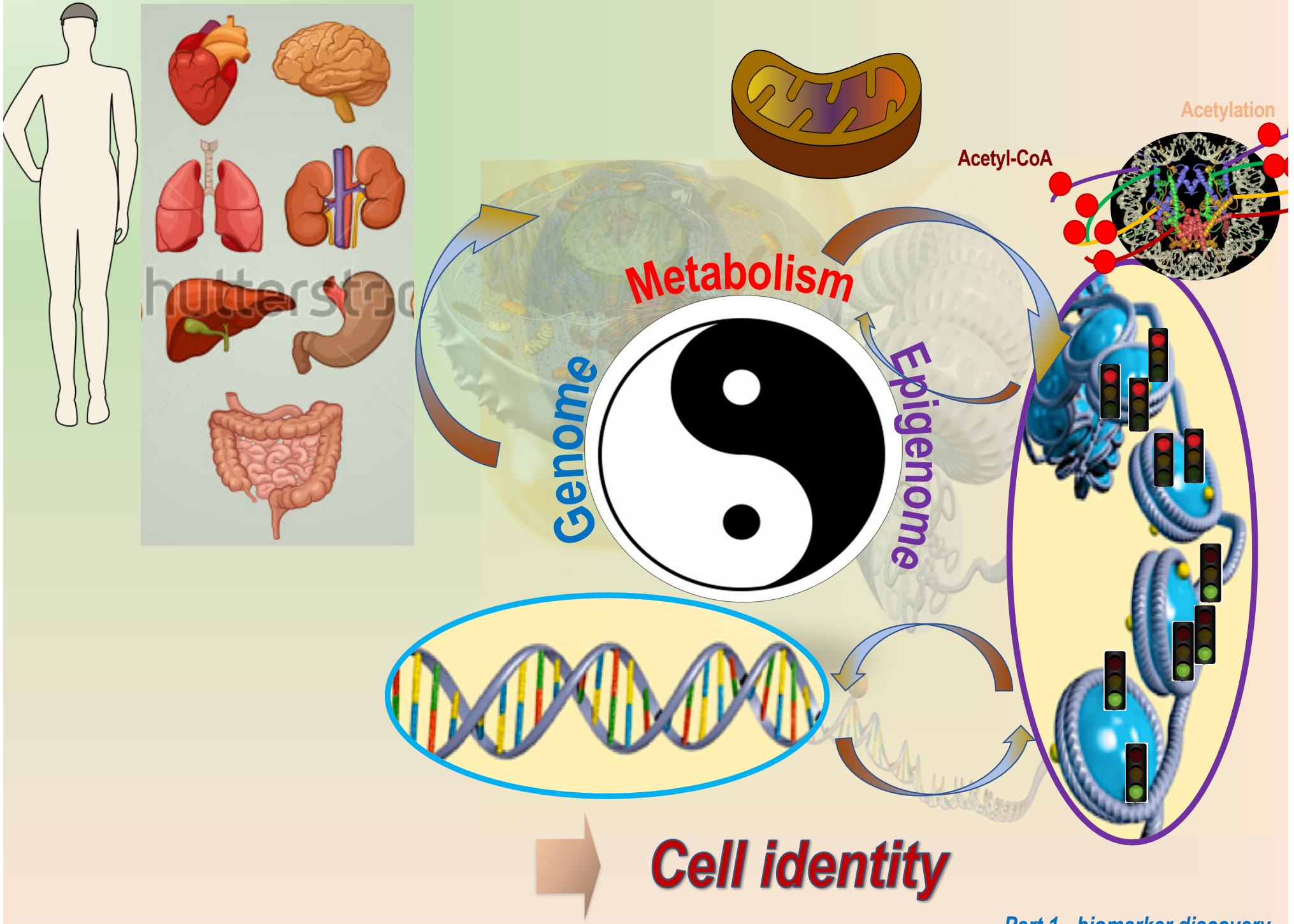
Gene expression programs are controlled by
the epigenome = genome signposting system



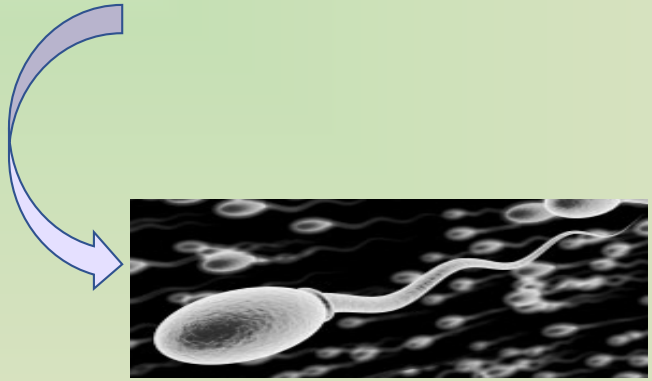


Cell identity

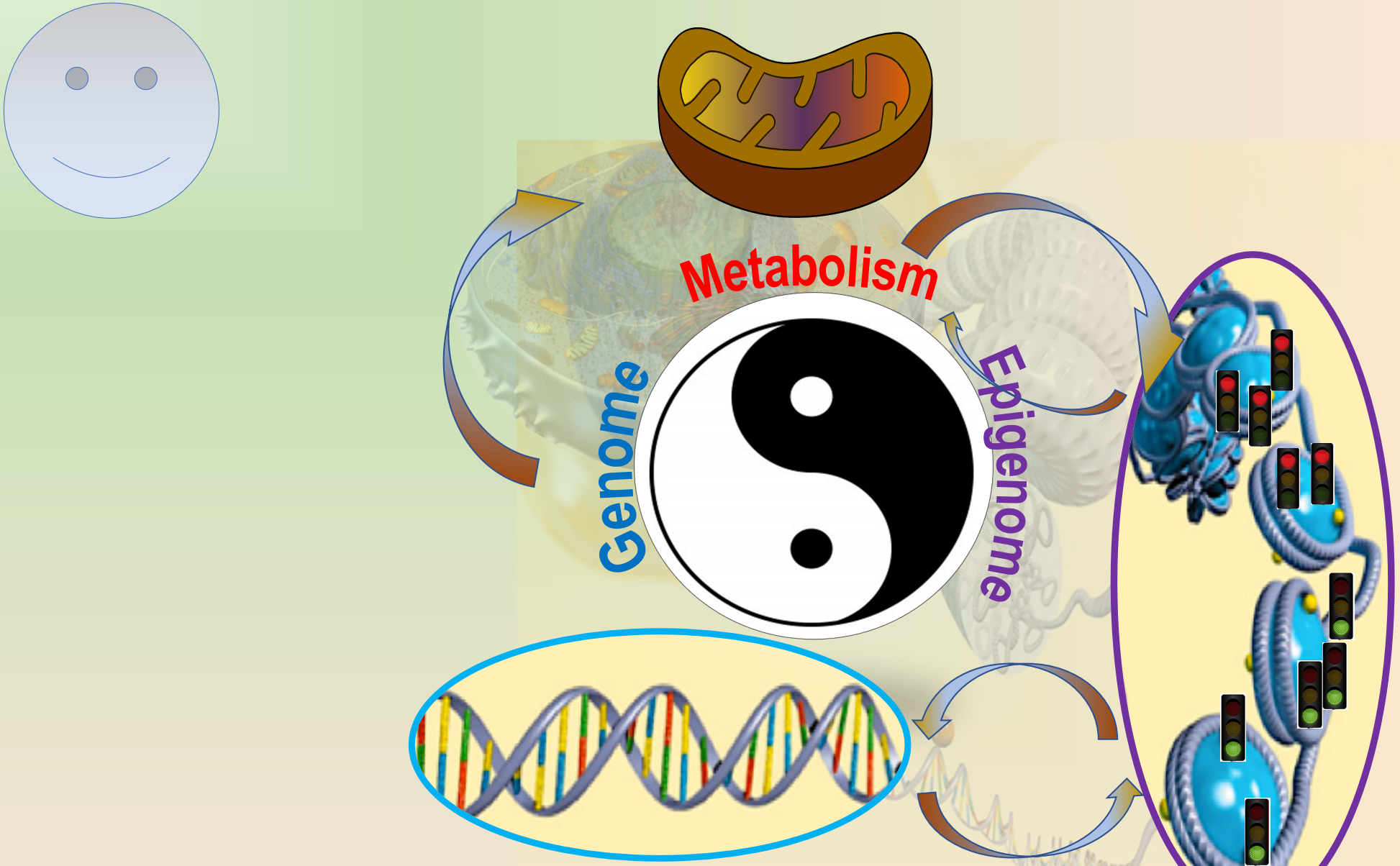
Part 1 - biomarker discovery


Acetylation
Acetyl-CoA
Metabolism
Genome
Epigenome
Cell identity
Part 1 - biomarker discovery

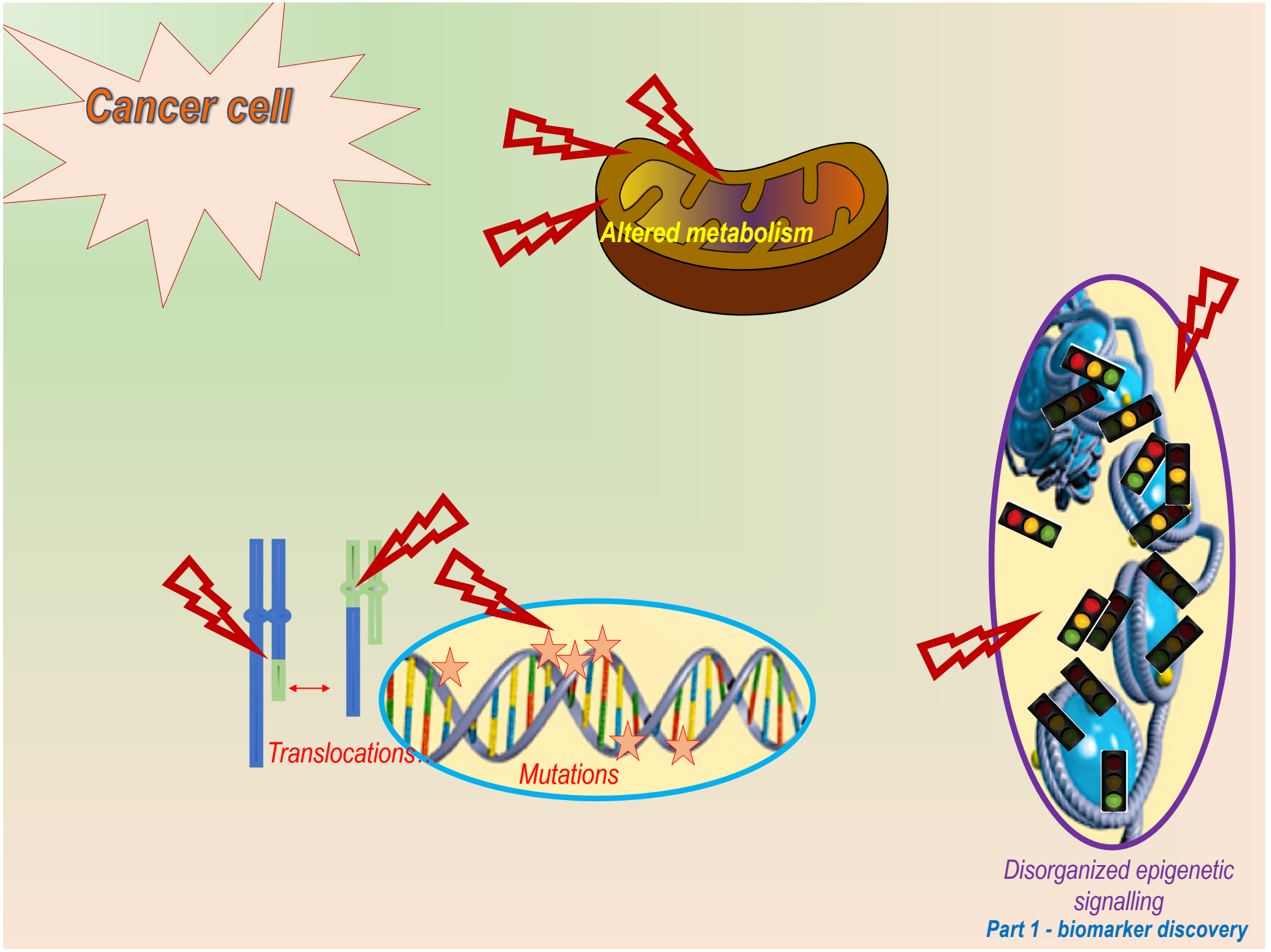
Some gene expression programs are highly TISSUE-SPECIFIC

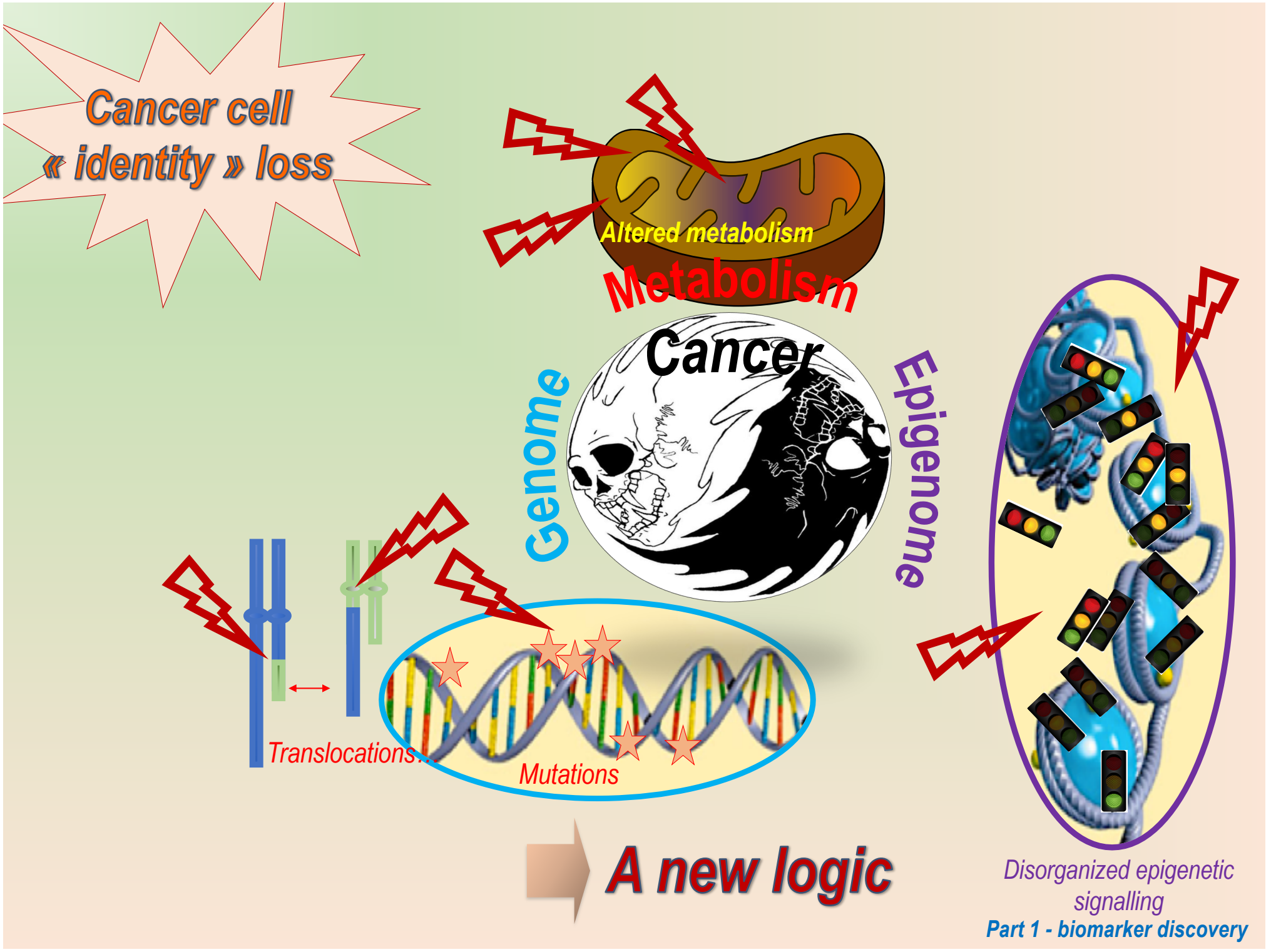


*Ex. male germ cells specific genes are **epigenetically** **"locked"** into a silent state in normal adult somatic cells*



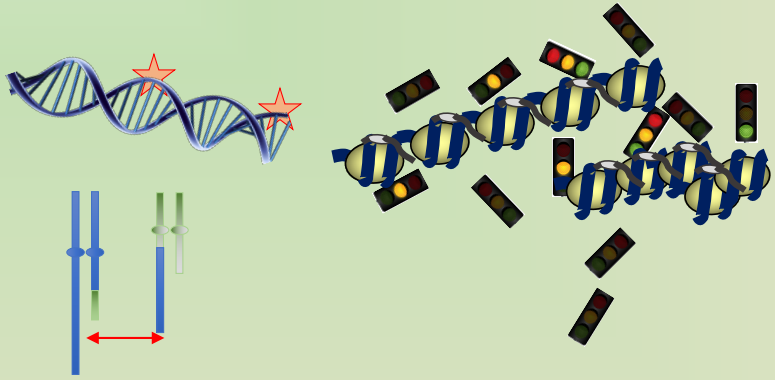
Cell type specific epigenome → **Cell identity**

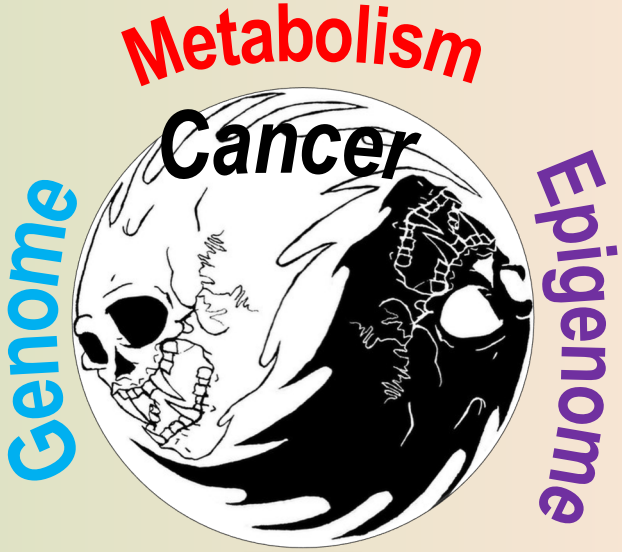

Cancer cell
Altered metabolism
Translocations...
Mutations
Disorganized epigenetic signalling
Part 1 - biomarker discovery


Cancer cell
« identity » loss
Altered metabolism
Metabolism
Cancer
Genome
Epigenome
Translocations…
Mutations
A new logic
Disorganized epigenetic signalling
Part 1 - biomarker discovery

Cancer cell « identity loss »

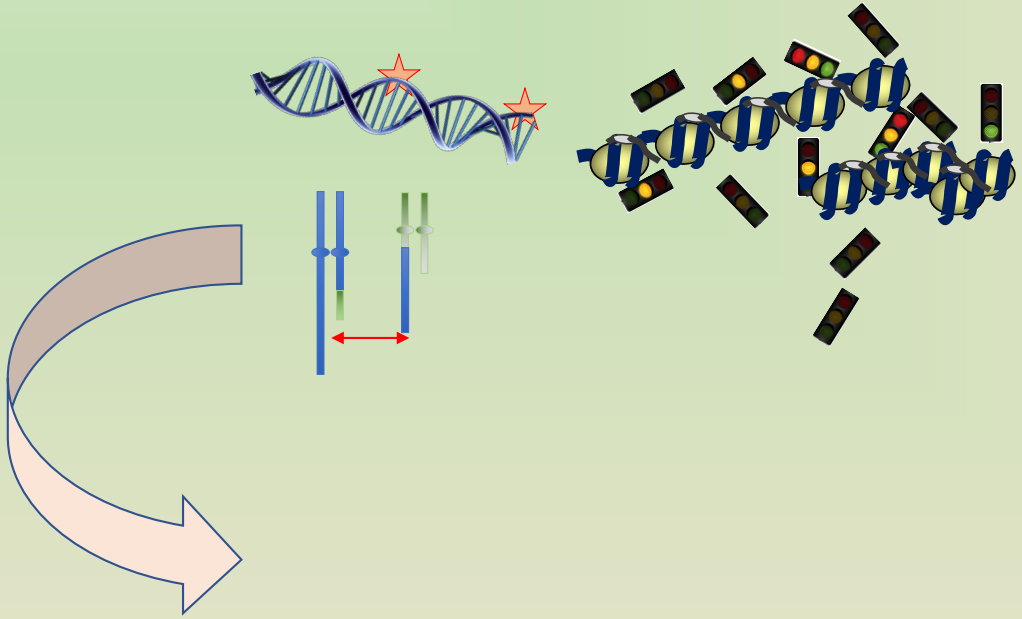
Genetic and epigenetic alterations

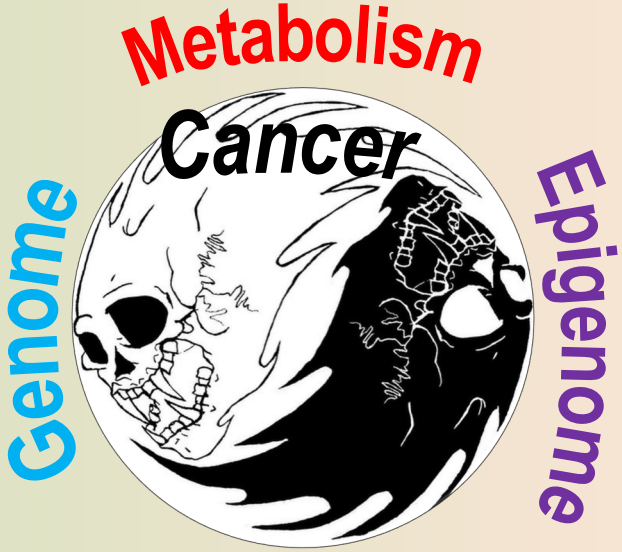




Cancer cell « identity loss »

Genetic and epigenetic alterations

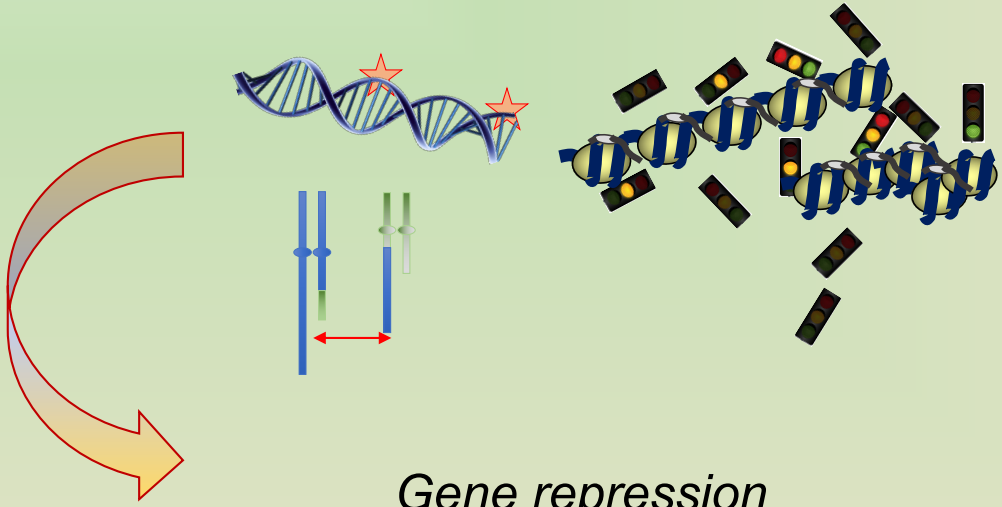




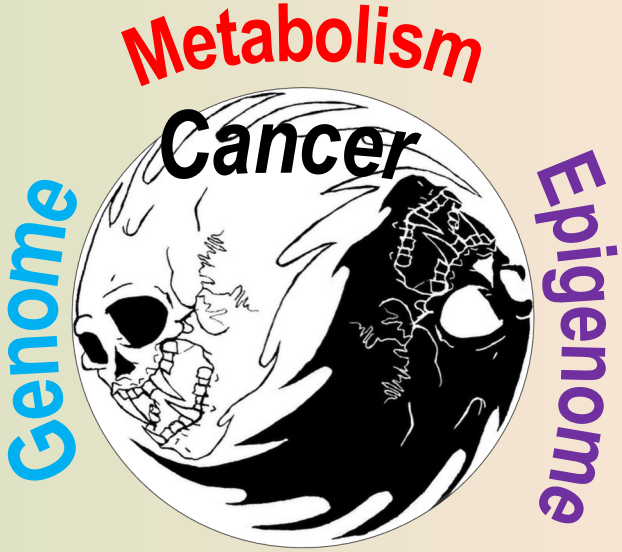
Major changes in gene expression programs

Cancer cell « identity loss »

Genetic and epigenetic alterations

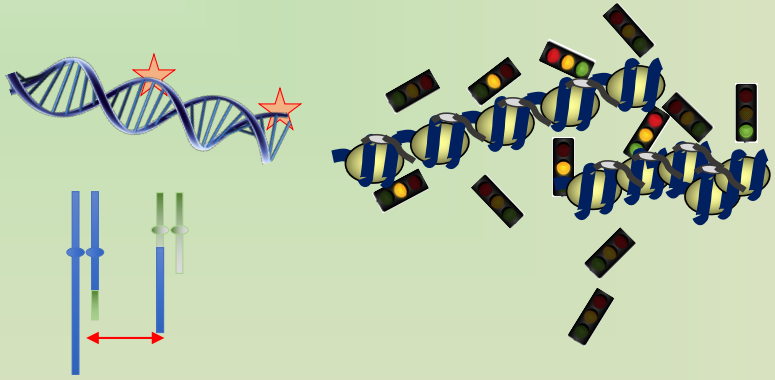


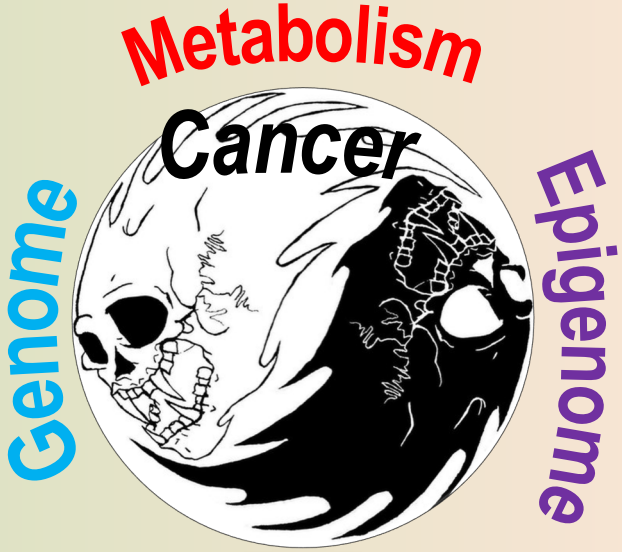
Gene repression
Tumour suppressors, etc..



Cancer cell « identity loss »

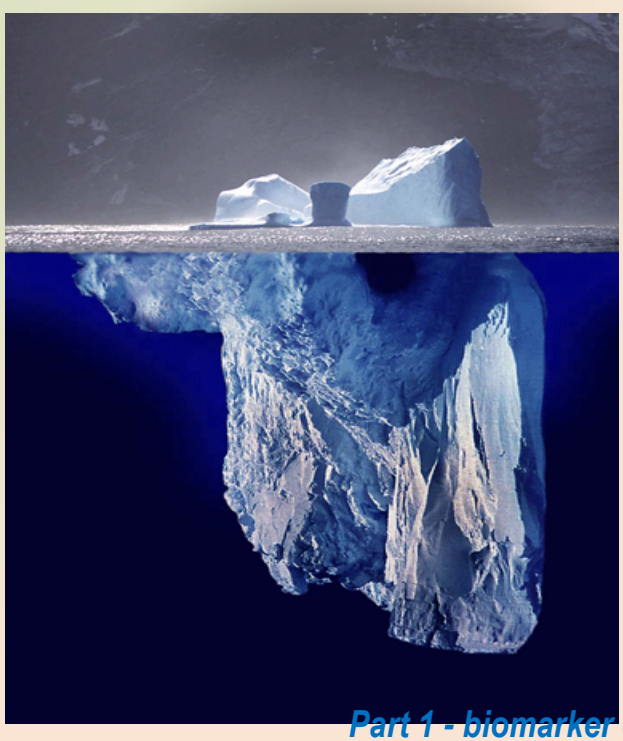
Genetic and epigenetic alterations





Gene repression
Tumour suppressors, etc..

Activation of tissue specific genes

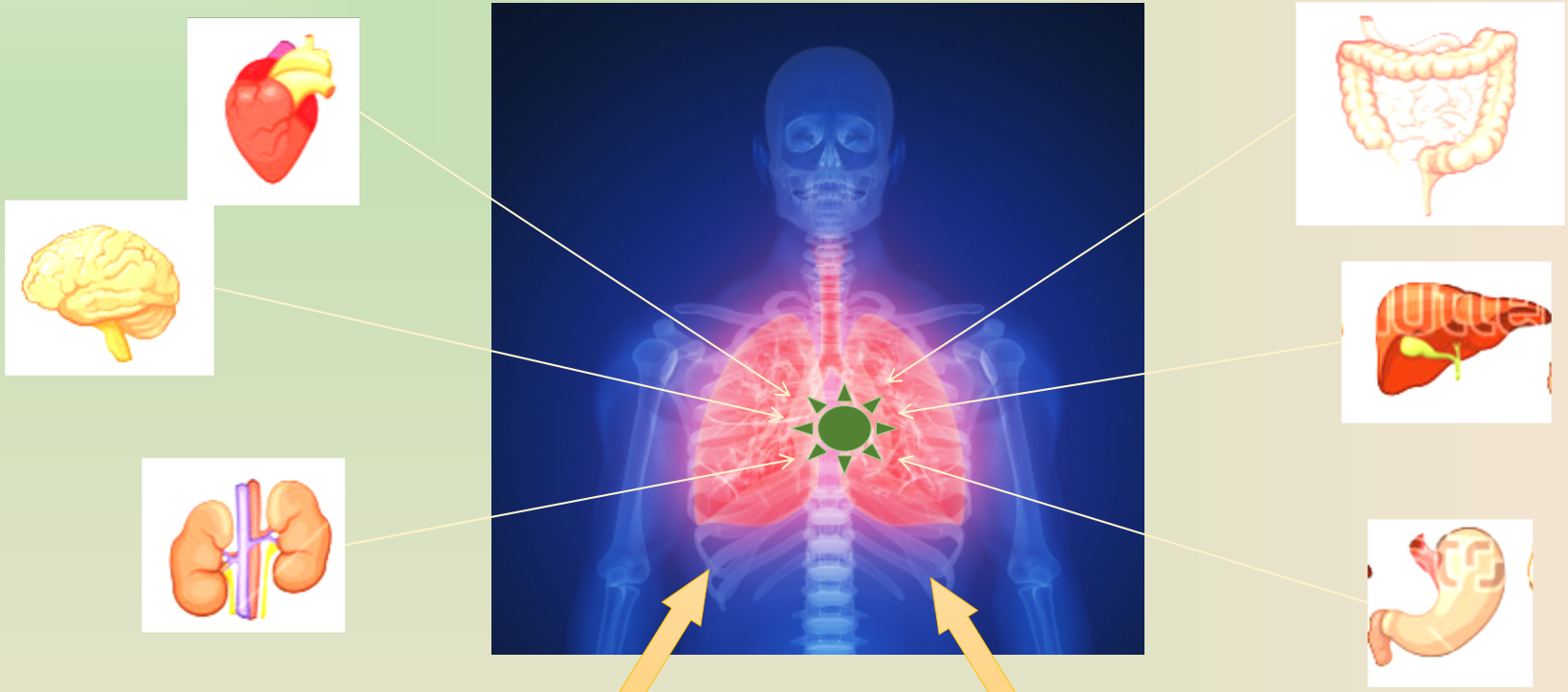


Cancer cell « identity loss »

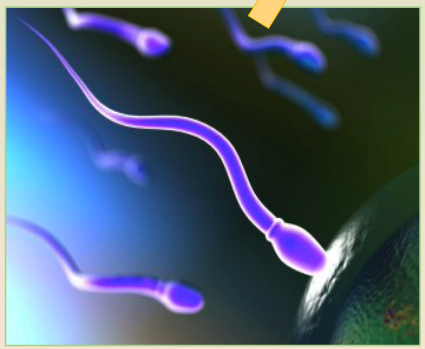
Activation of tissue specific genes

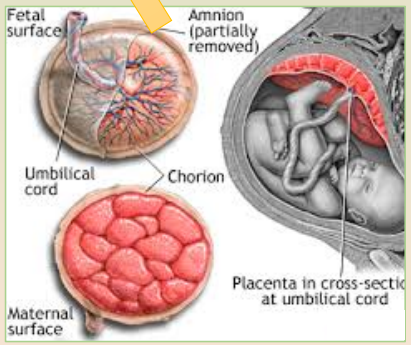
Cancer cell « identity loss »

Activation of tissue specific genes



Testis/
Spermatogenesis



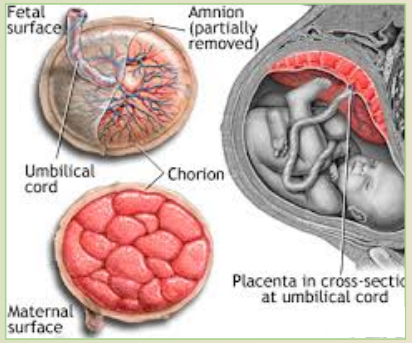


Placenta

Cancer cell « identity loss »

Activation of tissue specific genes

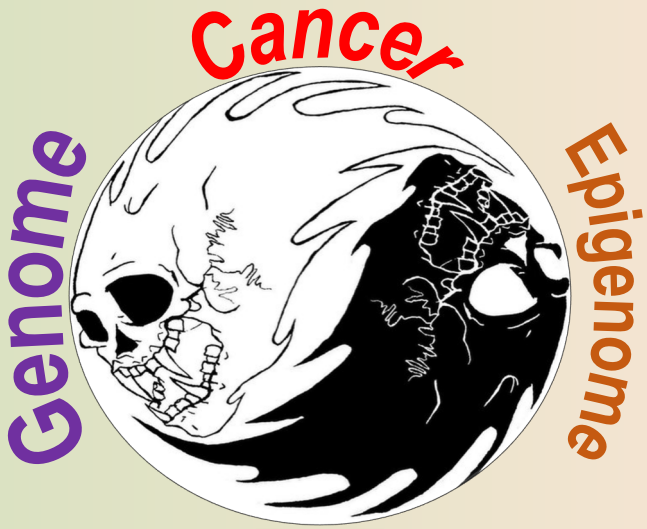
Testis/
Spermatogenesis



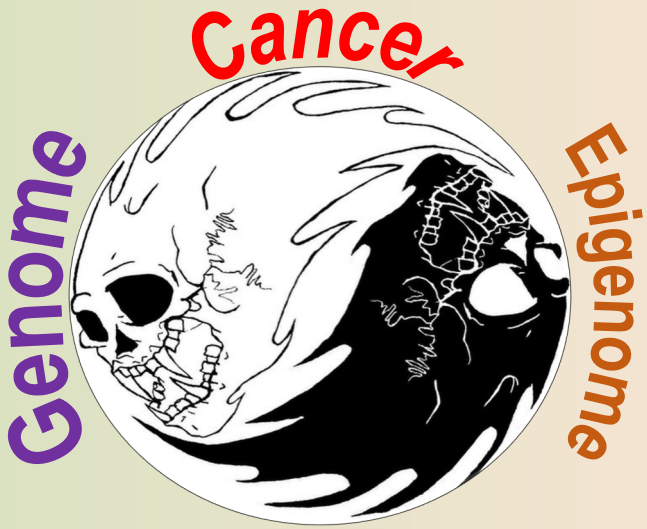
Placenta

1 – Very restricted pattern of expression

2 – Unknown to the immune system



Hypothesis: “out of context” activations of genes occur in all cancer types



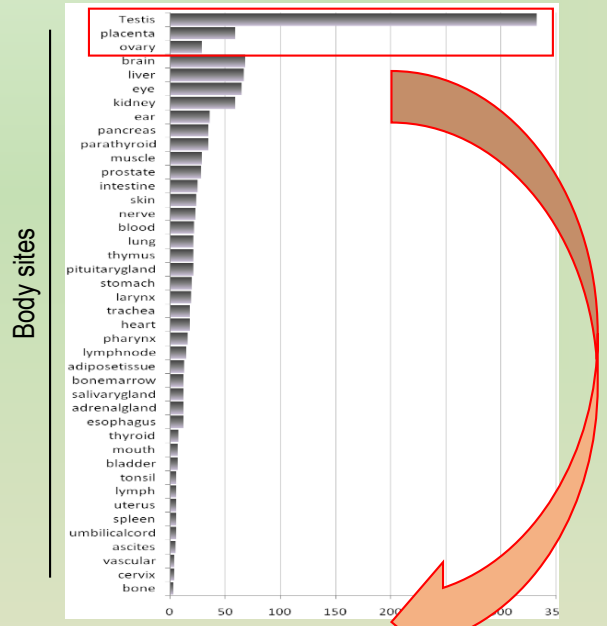
Hypothesis: "out of context" activations of genes occur in all cancer types

Systematic search for "out of context" (= ectopic) gene expression in cancer

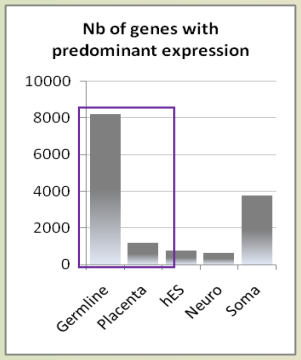
1st step: establishment of a list of « silent » genes

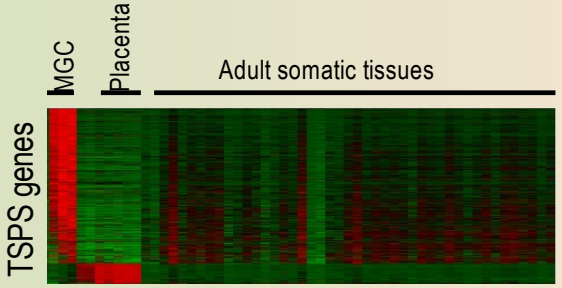
Expression databases

EST



Microarrays

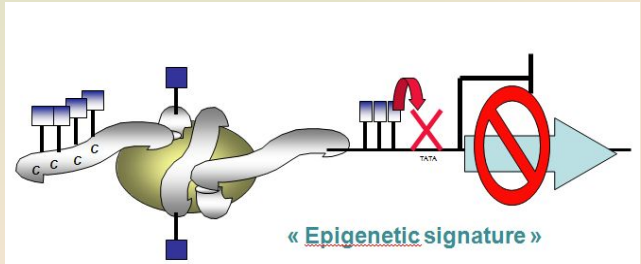




and

Epigenetic assessment

Methylomes, ChIP…



522 germline- and placenta- restricted genes (TSPS)

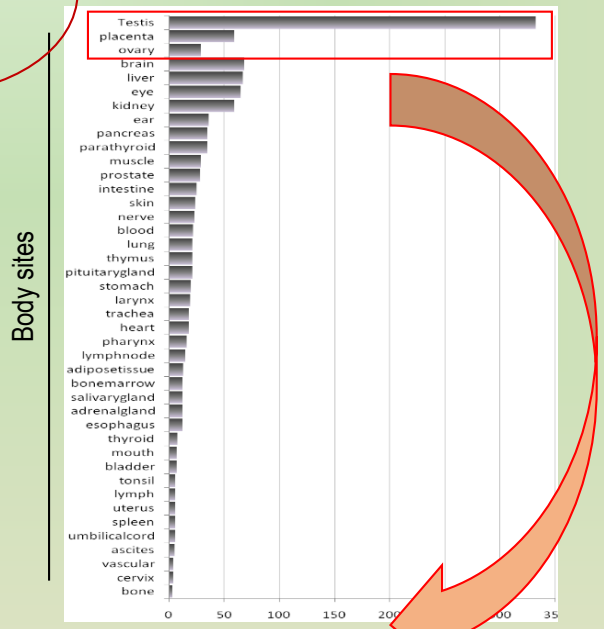
Part 1 - biomarker discovery

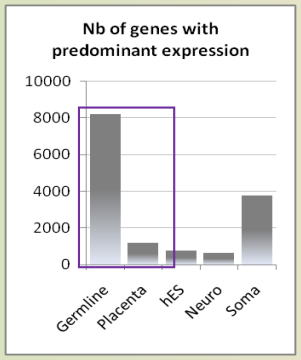
1st step: establishment of a list of « silent » genes

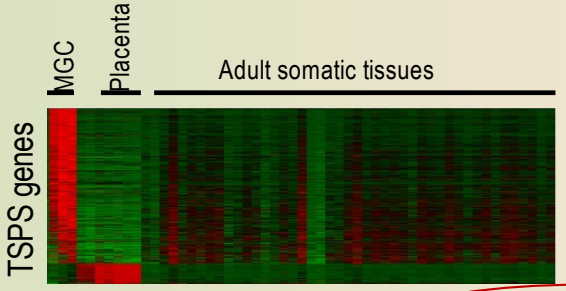
EST

Expression databases

Microarrays





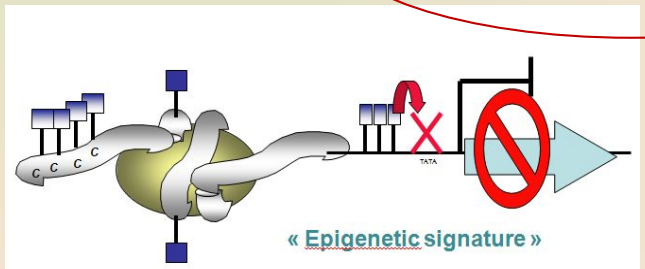


and

Epigenetic assessment

Methylomes, ChIP…

Matching gene IDs and corresponding tissues IDs

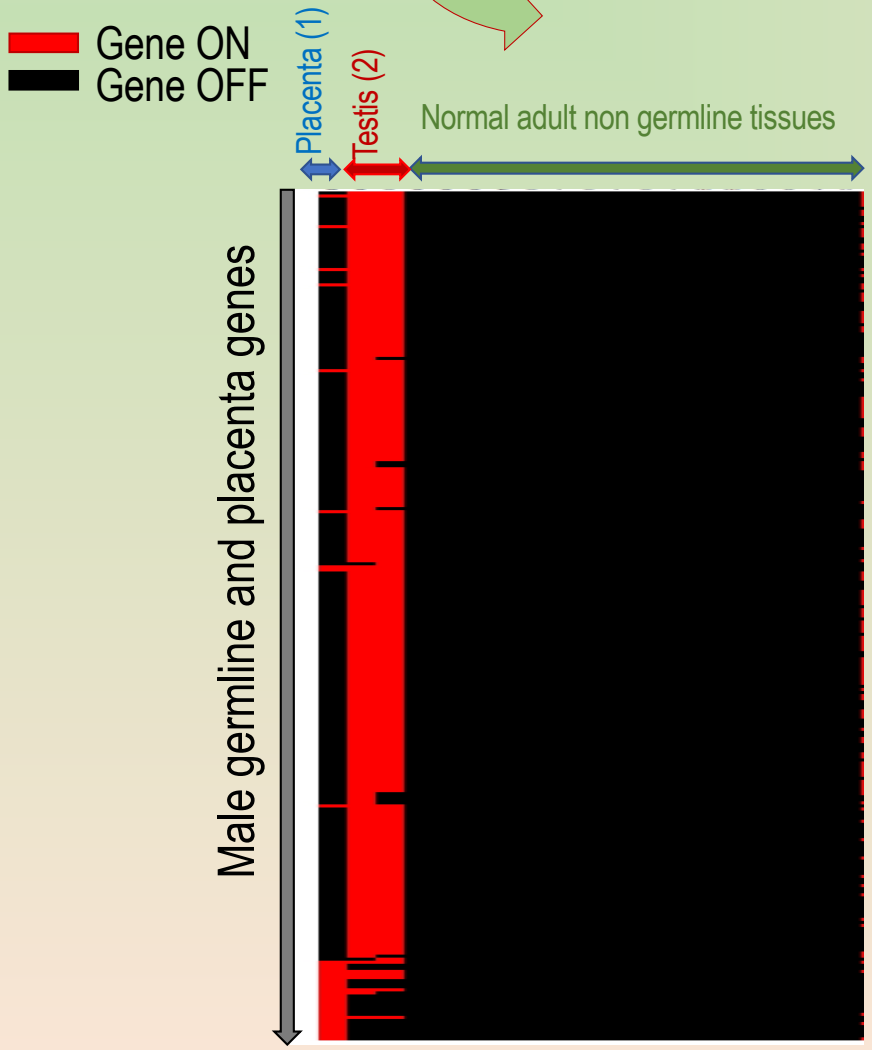


522 germline- and placenta- restricted genes (TSPS)

Part 1 - biomarker discovery

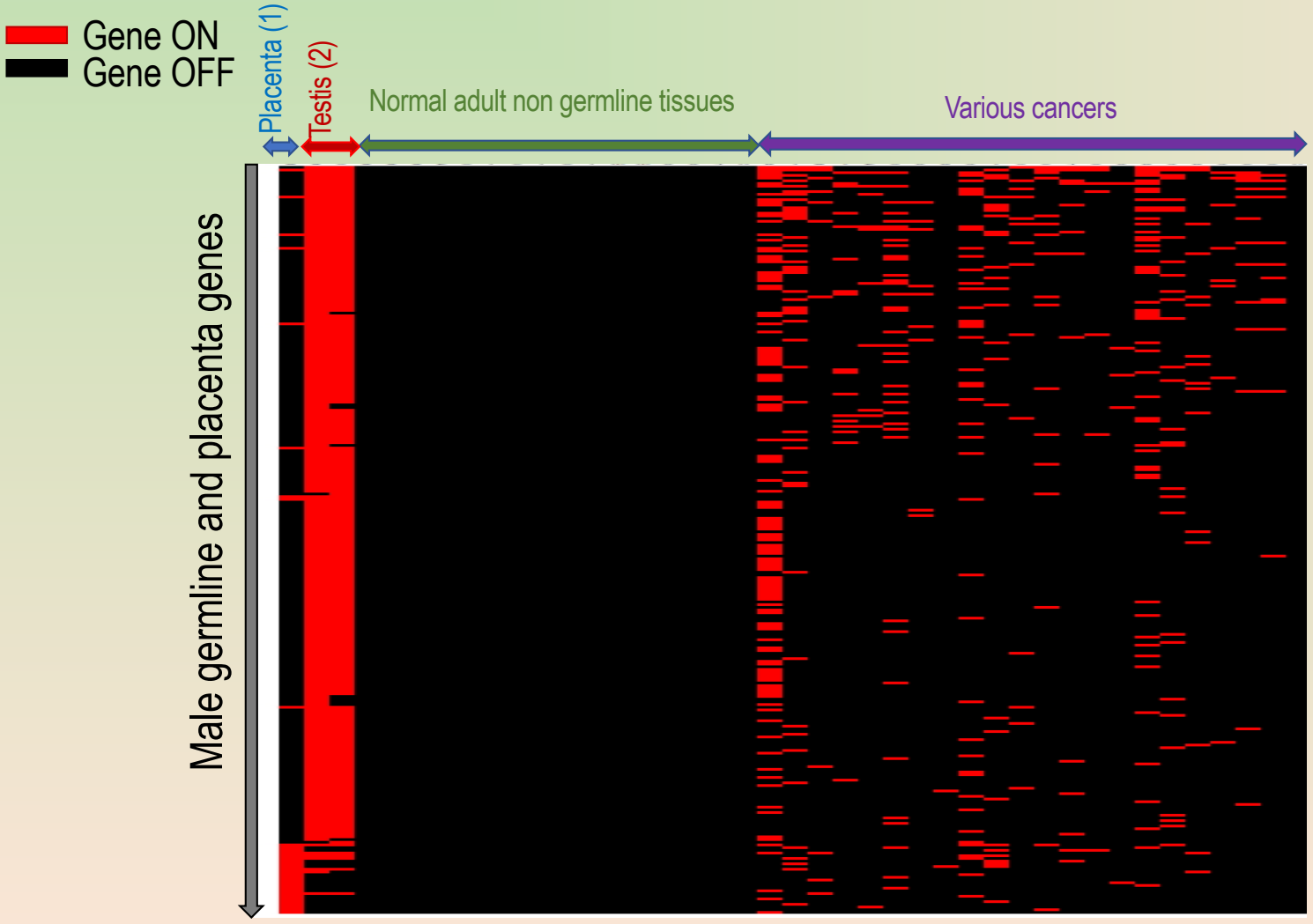
522 **germline and placenta « restricted » genes**

Dedicated array



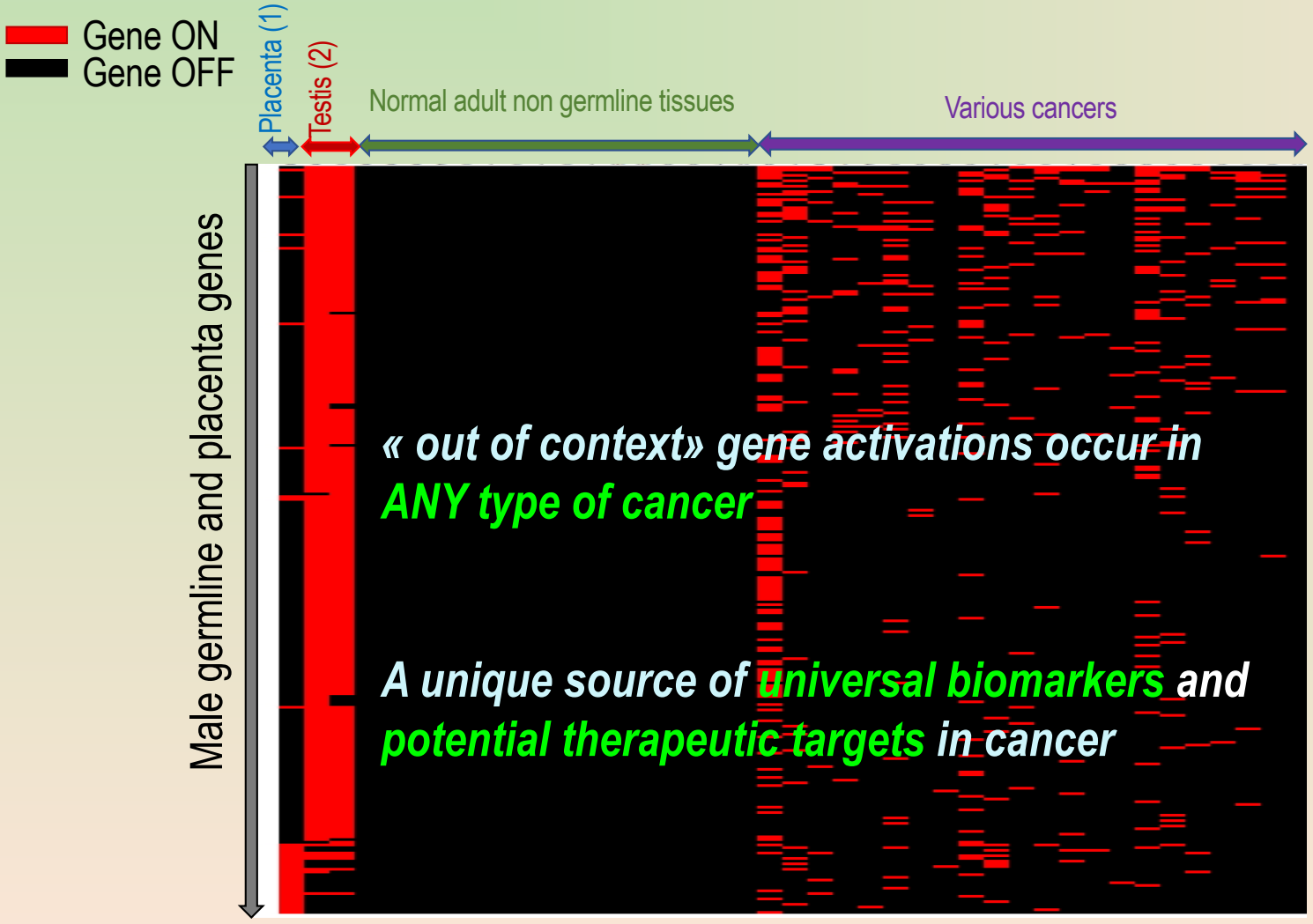
522 **germline and placenta « restricted » genes**

2nd step: detection of ectopic activation of germline/placenta specific genes in various cancer types



522 **germline and placenta « restricted » genes**

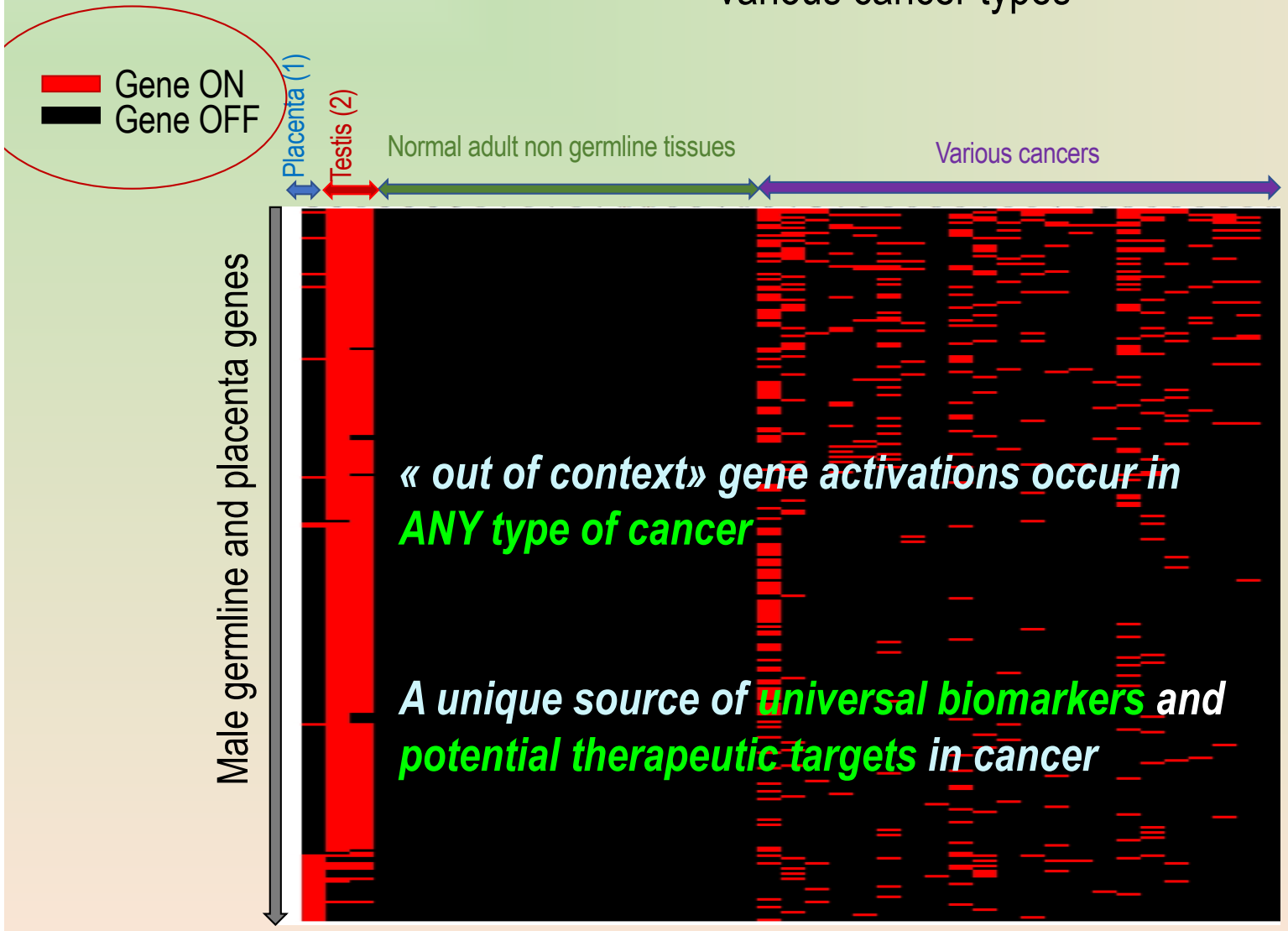
2nd step: detection of ectopic activation of germline/placenta specific genes in various cancer types



Analysis of approx. 2000 samples from 16 different solid tumours (GSE2109: Expression Project for Oncology (expO) project) Affymetrix microarrays

522 **germline and placenta « restricted » genes**

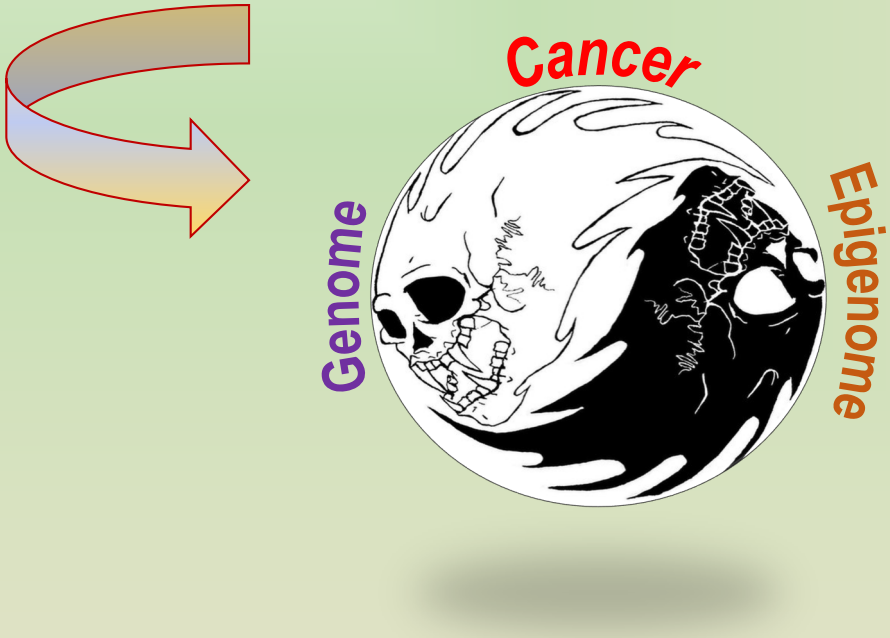
2nd step: detection of ectopic activation of germline/placenta specific genes in various cancer types



ON/OFF threshold for binary data

3rd step Clinical use of **ectopic gene expressions** in cancer

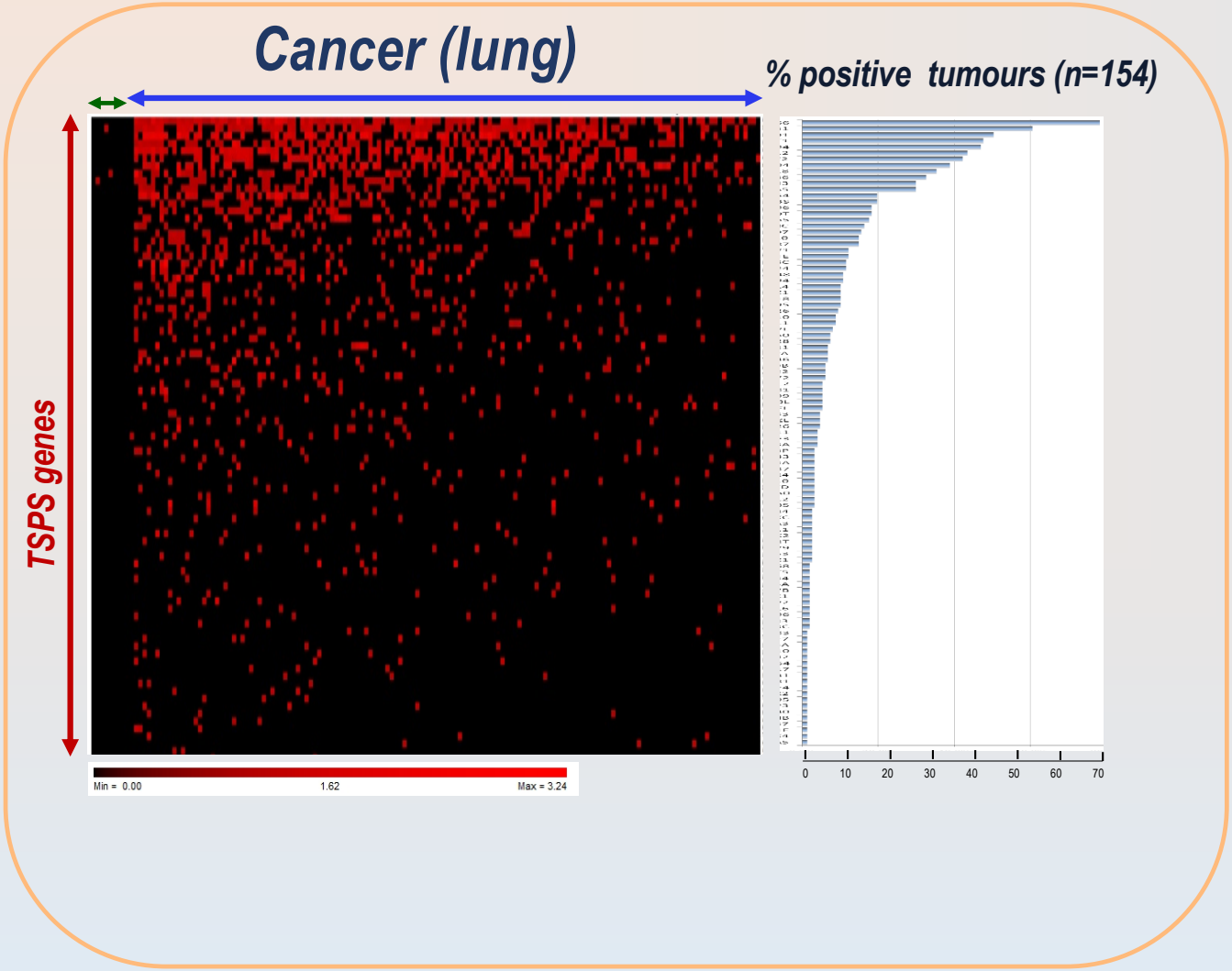
3rd step Clinical use of **ectopic gene expressions** in cancer



Biomarkers discovery

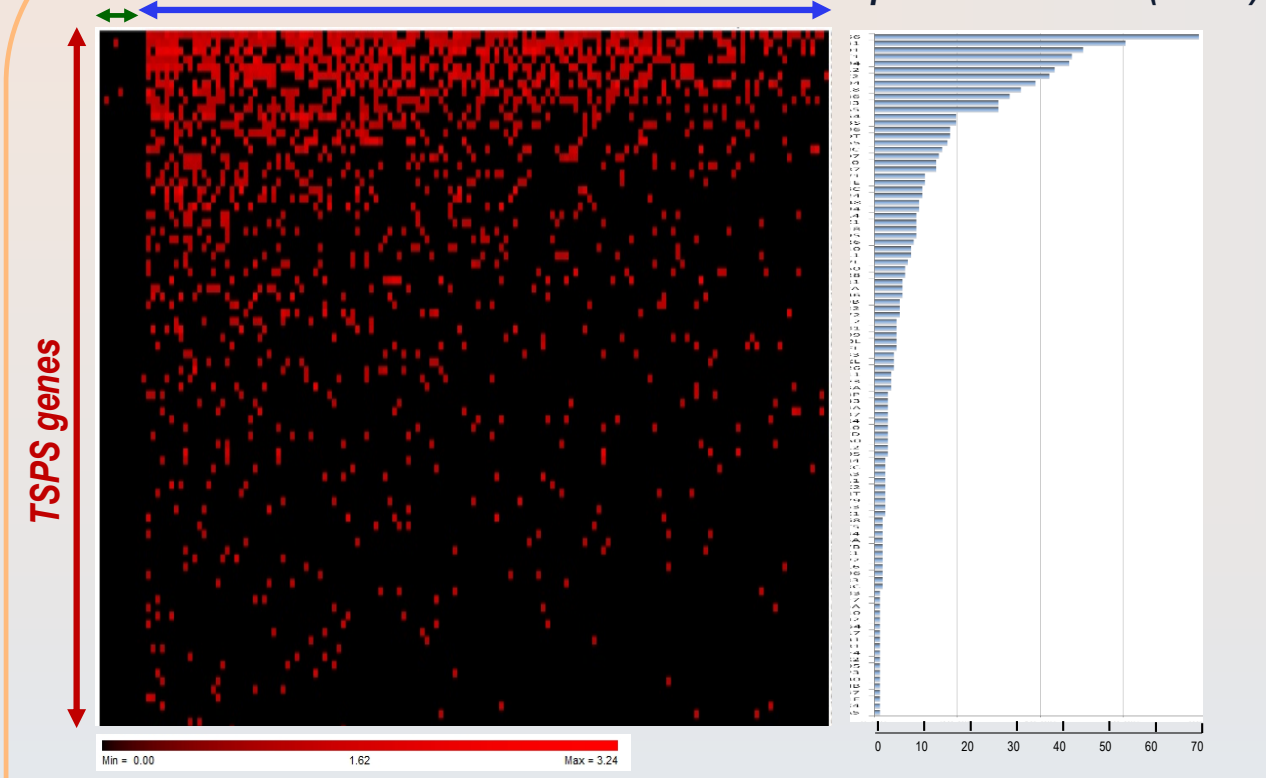
New approach to cancer prognosis and treatment

Biomarkers discovery pipeline

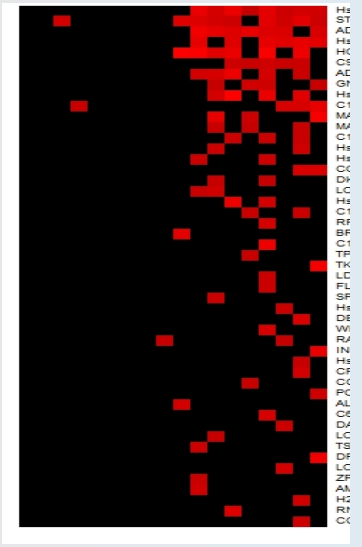


Biomarkers discovery pipeline

Cancer (lung)

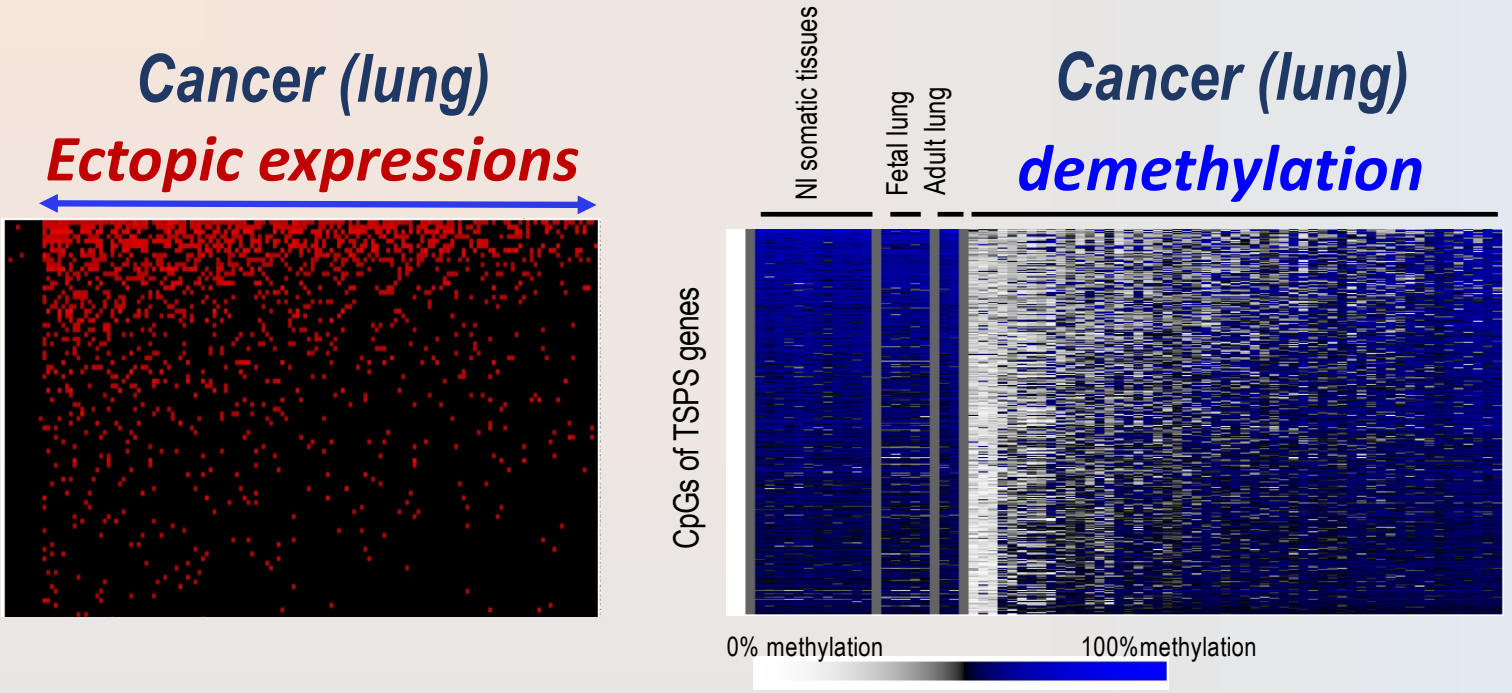


NTL paired T



=> *Non tumoral lung from patients with lung cancer do NOT have ectopic expressions*

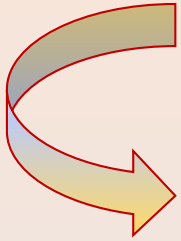
Epigenetic origin of ectopic expressions



Epigenetic abnormalities are responsible for the derepression and aberrant expression of germline/placenta-specific genes in cancer (lung)

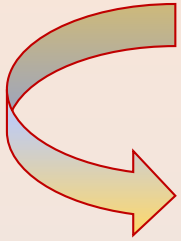
=> ***DNA demethylation of the promoter of germline/placenta genes*** *is associated with ectopic expression*

Biomarkers discovery pipeline



Evaluation of the clinical value of ectopic expressions as prognosis markers

Biomarkers discovery pipeline



Evaluation of the clinical value of ectopic expressions as prognosis markers

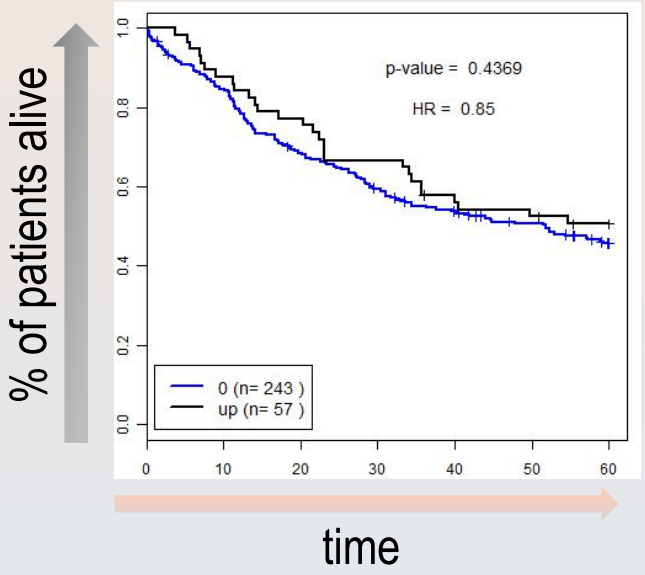
Tumours with ectopic expression

=> Compare survival

Tumours with no expression

Biomarkers discovery pipeline

Evaluation of the clinical value of ectopic expressions as prognosis markers



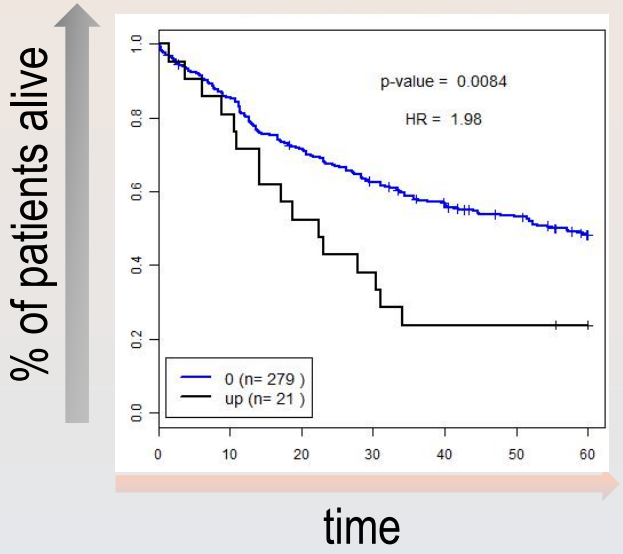
Tumours with ectopic expression

Tumours with no expression

=> Gene expression **not** associated with prognosis

Biomarkers discovery pipeline

Evaluation of the clinical value of ectopic expressions as prognosis markers



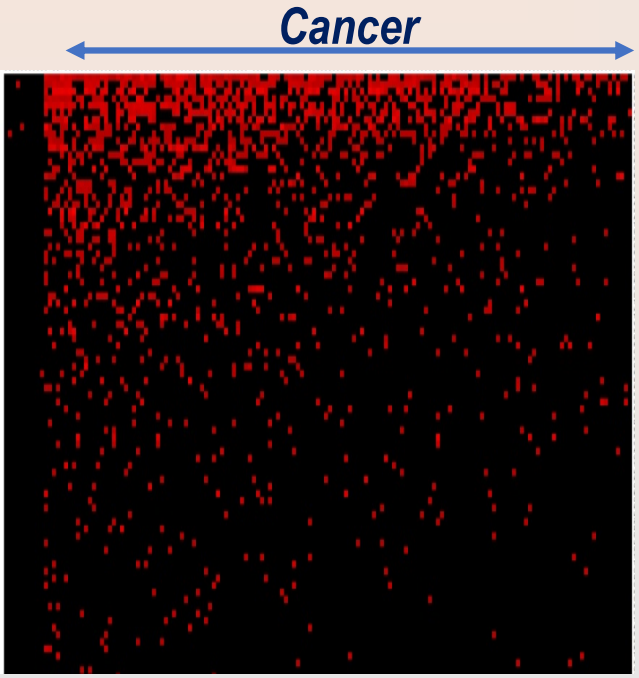
Tumours with no expression

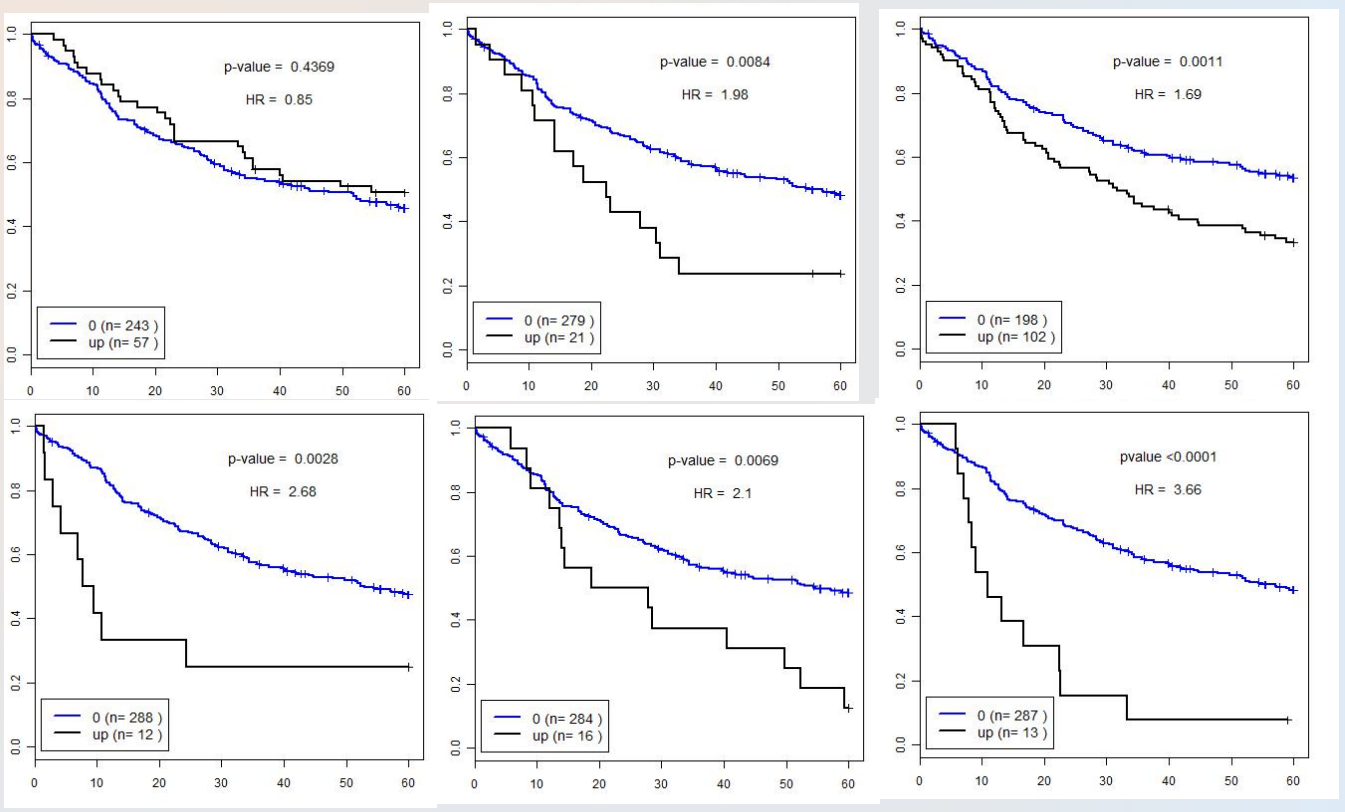
Tumours with ectopic expression

=> Gene expression associated with prognosis

Biomarkers discovery pipeline

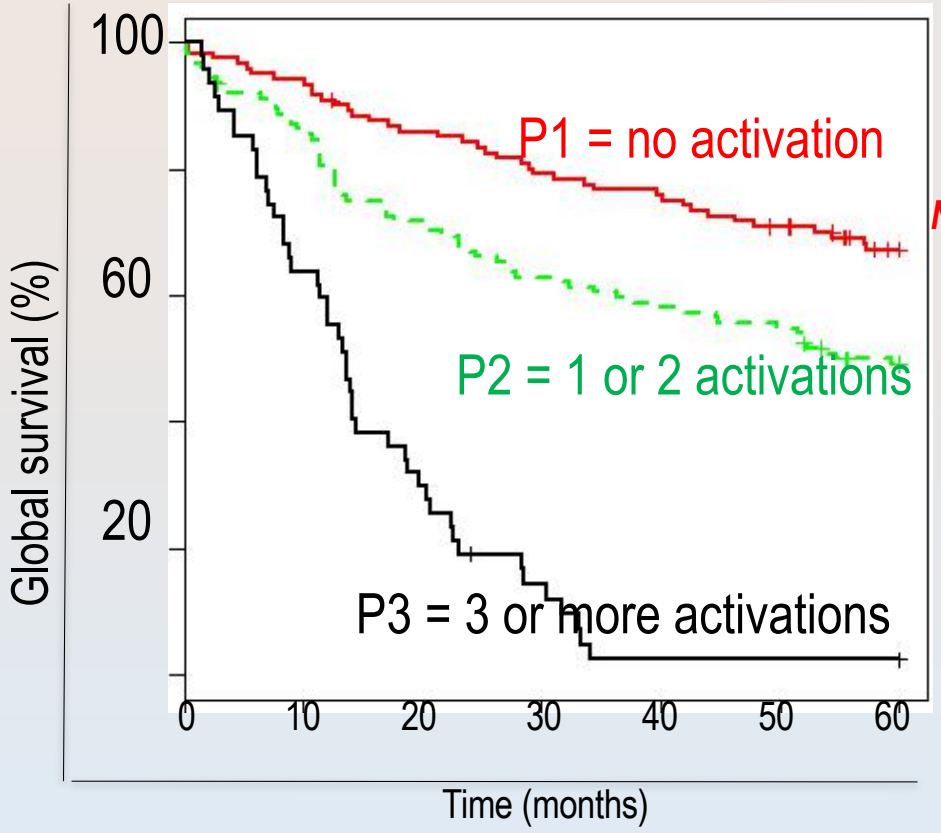
« off context » activation of germline restricted genes and prognosis:





Biomarkers discovery pipeline

Example of **gene combination** algorithm



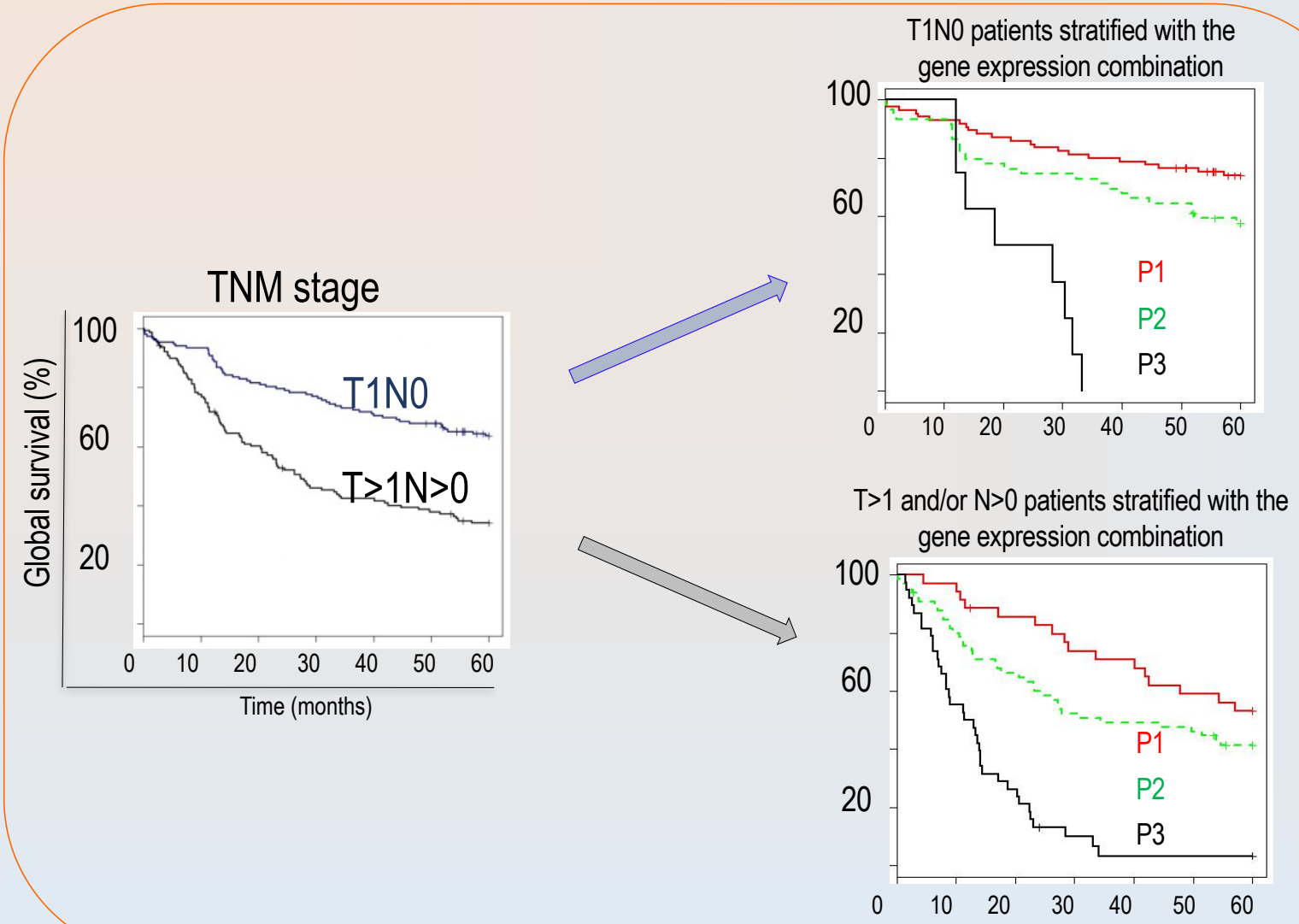
None of the genes are active

3 or more of the genes are active

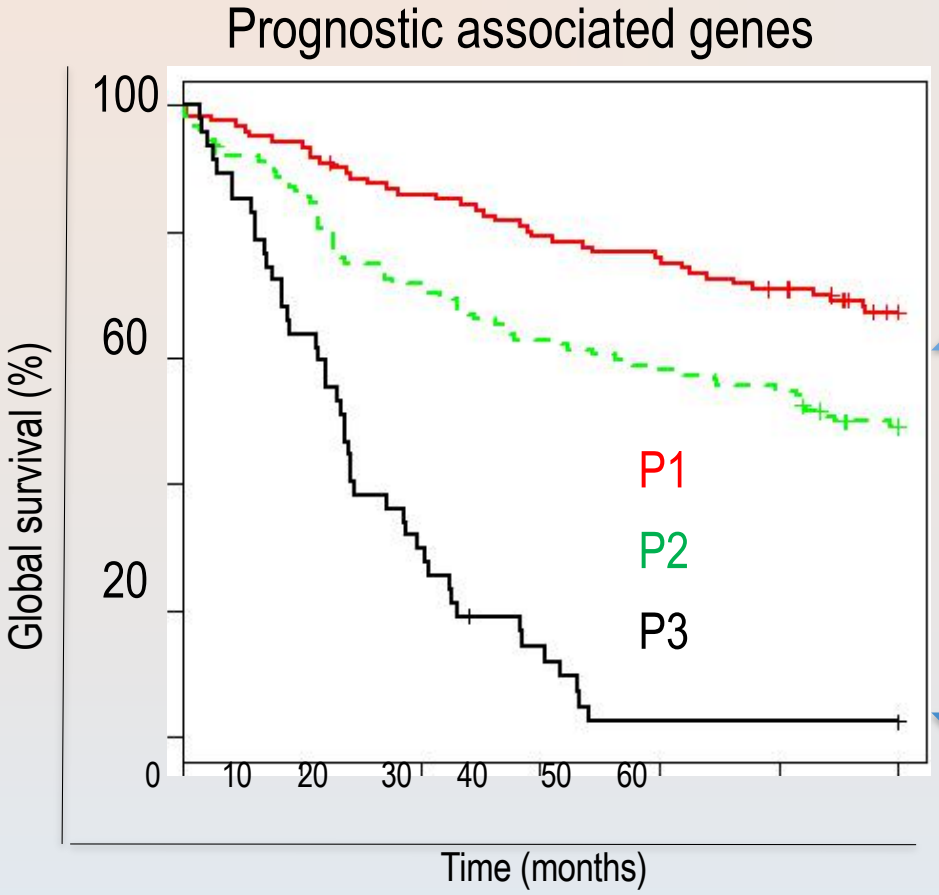
Prognosis stratifying method

Biomarkers discovery pipeline

Discriminating power of our ectopic genes classifying system: example (lung)



Biomarkers based molecular characterization



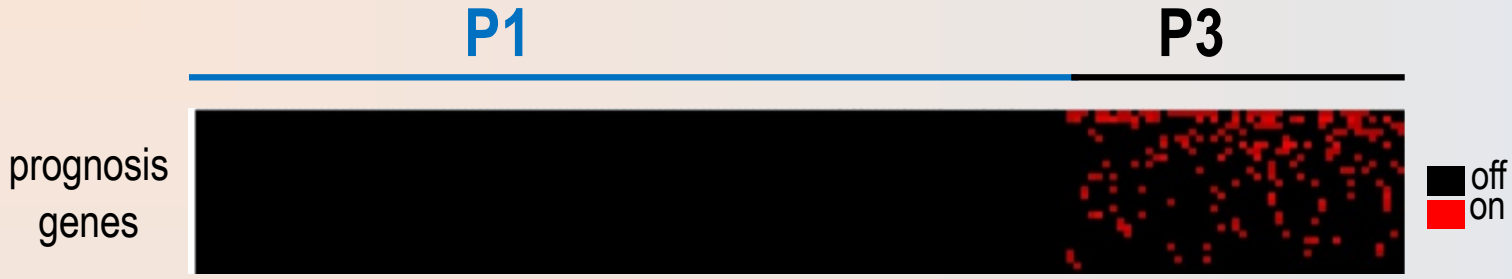
P1

P3

Supervised transcriptomic analysis

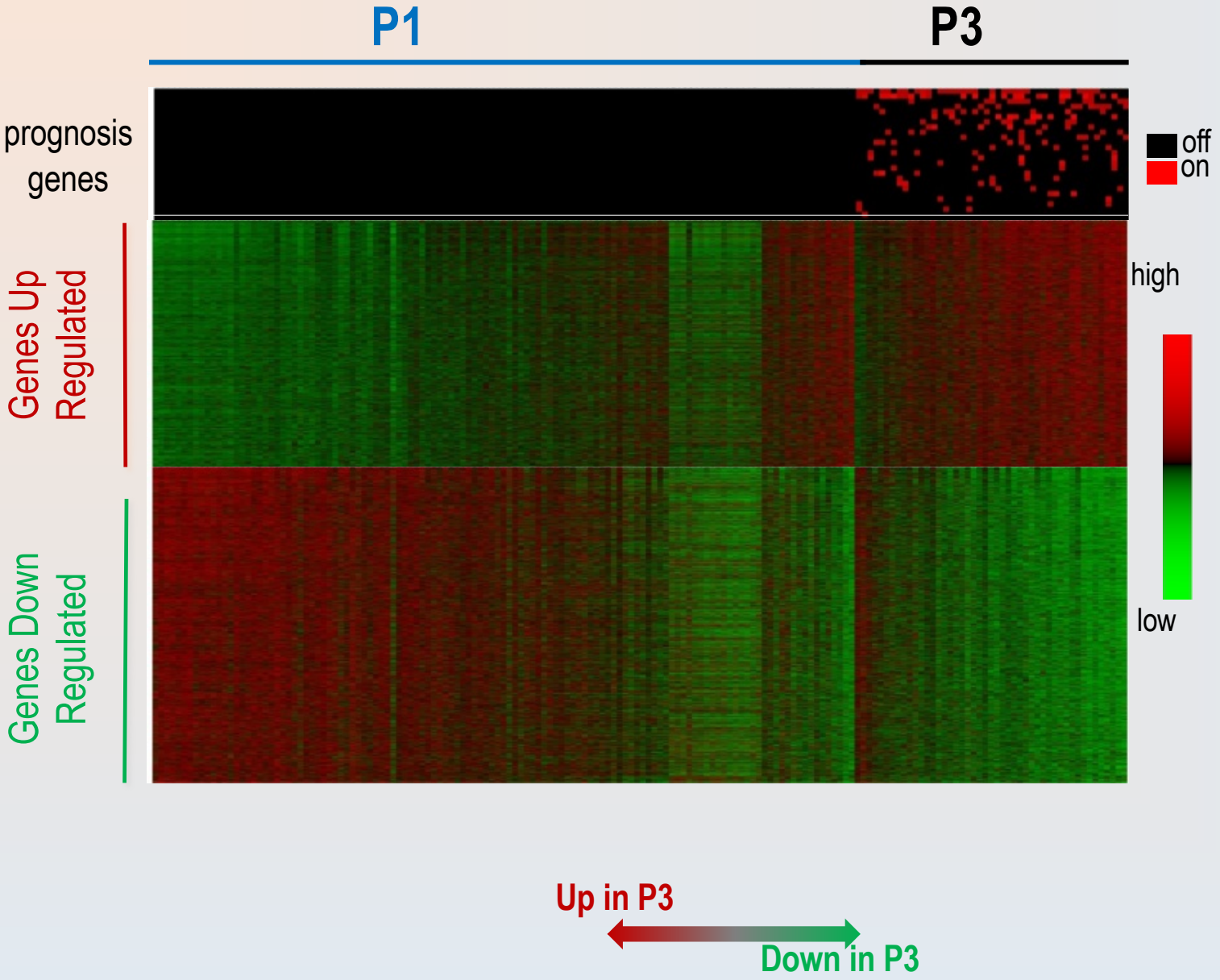
Gene expression profile

Biomarkers based molecular characterization

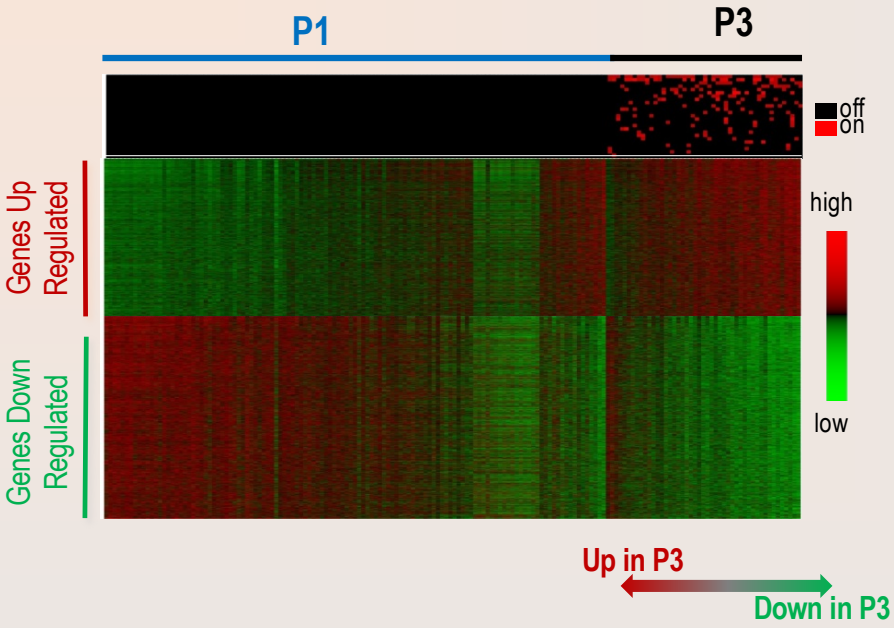


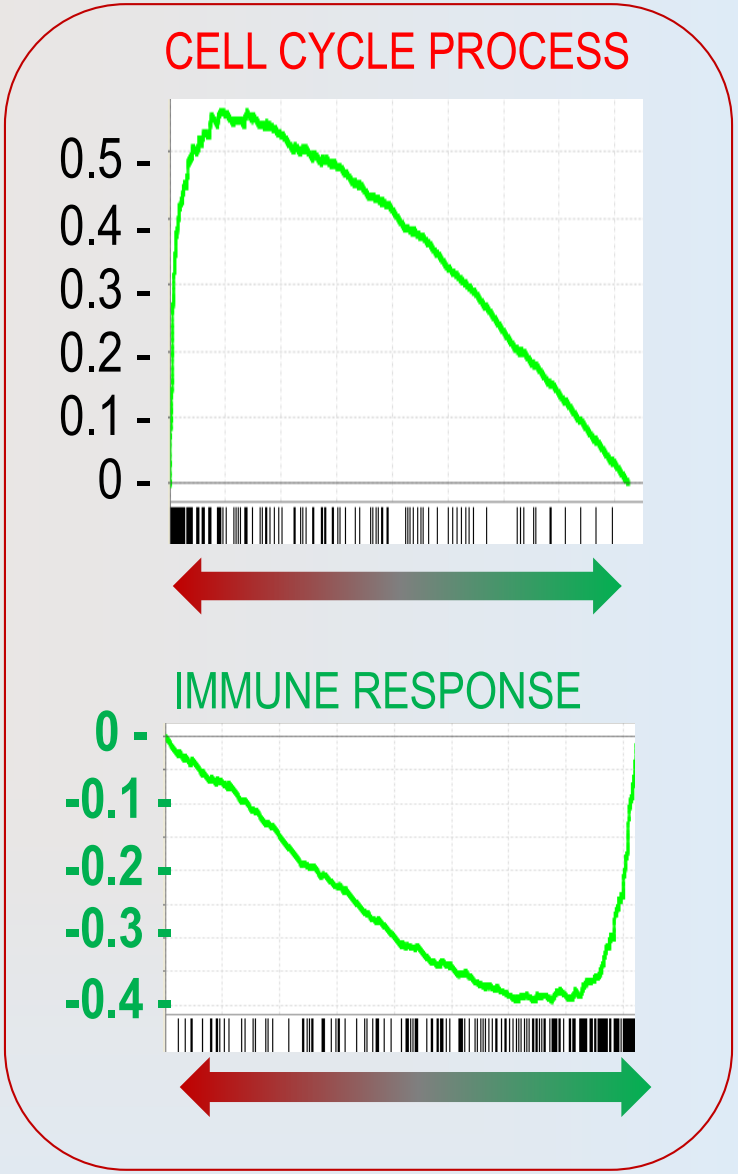
Supervised
transcriptomic analysis

Biomarkers based molecular characterization



Biomarkers based molecular characterization





GeneSet Enrichment Analysis

Molecular characterization of highly aggressive tumors

Computational translational epigenetics: concept-driven omics analyses

Part 1: The awakening of silent genes in malignancies : a new biomarker discovery strategy

Part 2: Concept driven omics analyses : EpiMed information system and pipelines

http://epimed.univ-grenoble-alpes.fr/database/

In silico Epigenetics

EpiMed

Computational
Translational
Epigenetics

Dept Signalling and chromatin

Saadi Khochbin

Saadi Khochbin

Epigenetics and cell signaling
Saadi Khochbin *(DRHC, CNRS)*

Sophie Rousseaux (*DR2, INSERM*)
Anne-Laure Vitte (*IE Bio, UGA*)
Ekaterina Flin (*IR bioinfo, UGA*)
Florent Chuffart (*IR Bioinfo, INSERM*)

Sophie Rousseaux

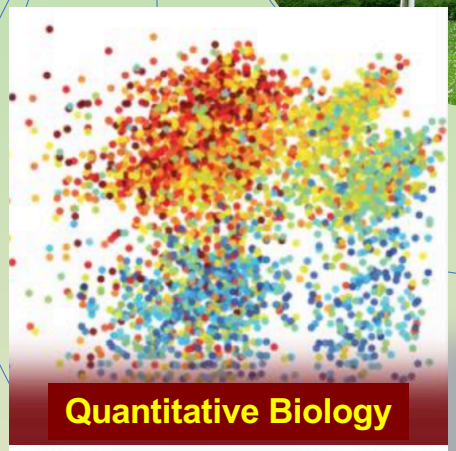
Ekaterina Flin

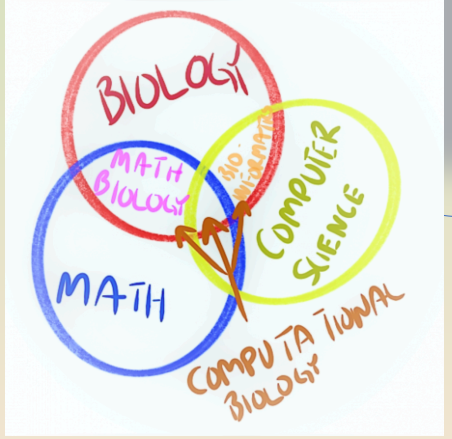
Florent Chuffard

Anne-Laure Vitte

In silico Epigenetics

EpiMed





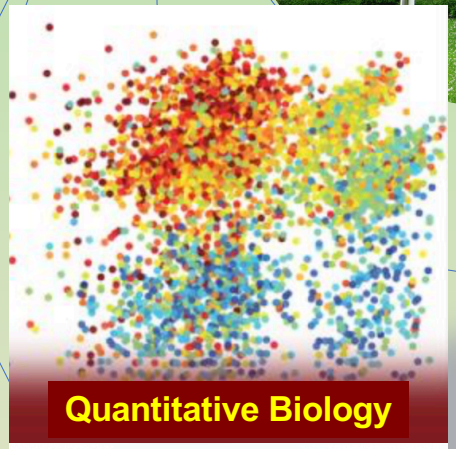
Computational biology to epigenetics

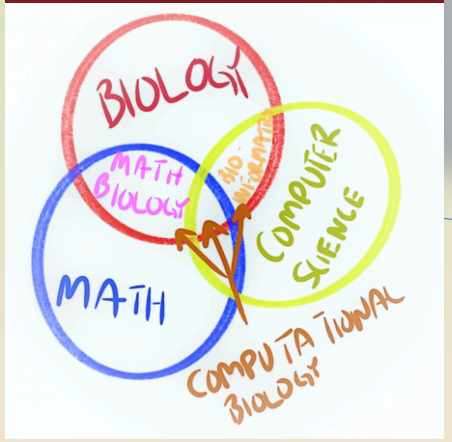
Ekaterina Flin

Florent Chuffart

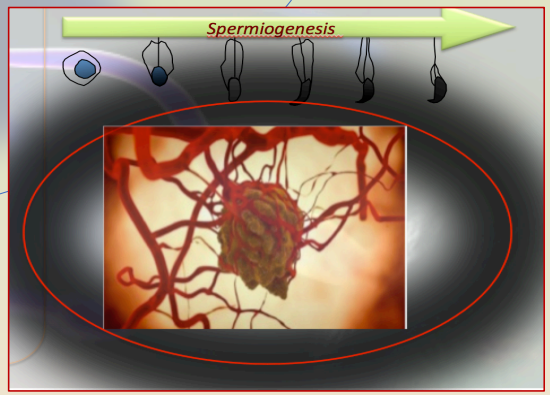
In silico Epigenetics

EpiMed





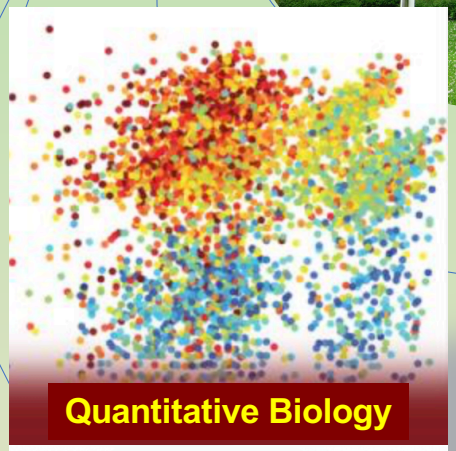
Computational biology to epigenetics

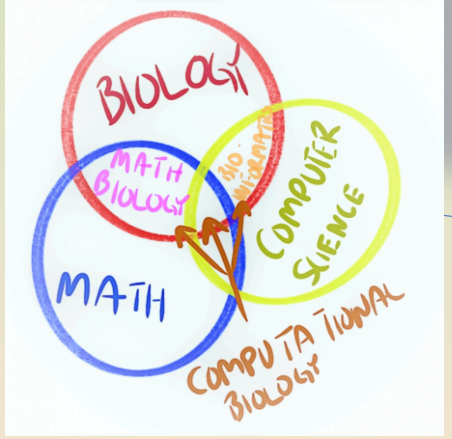


Cancer & chronic diseases biomarkers

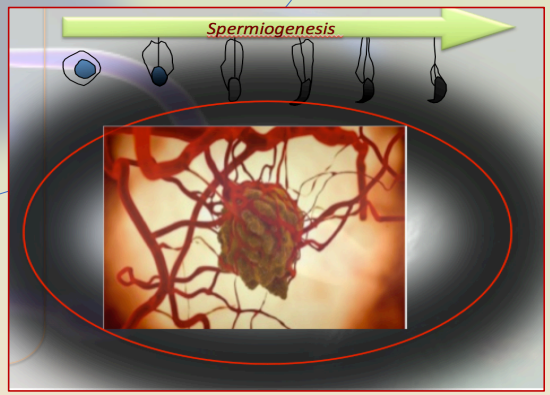
In silico Epigenetics

EpiMed





Computational biology to epigenetics



Cancer & chronic diseases biomarkers

Population Epigenetics

Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature

*1- identify **tissue-specific** genes*

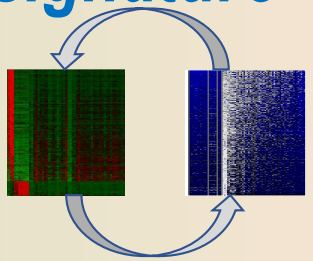
*2- detection of **ectopic expression** of tissue specific genes in cancer*

*3- correlate ectopic expressions with **clinical outcome***

*4- characterize **molecular profile of aggressive tumours***

*5- Understand **molecular basis of aggressive phenotype** by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)*

Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature



1- identify **tissue-specific** genes

cross different types of data

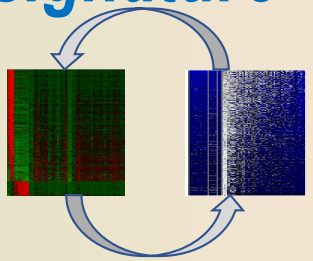
2- detection of **ectopic expression** of tissue specific genes in cancer

3- correlate ectopic expressions with **clinical outcome**

4- characterize **molecular profile of aggressive tumours**

5- Understand **molecular basis of aggressive phenotype** by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)

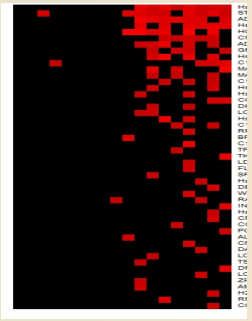
Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature



1- *identify* ***tissue-specific*** *genes*

cross different types of data

2- *detection of* ***ectopic expression*** *of tissue specific genes in cancer*



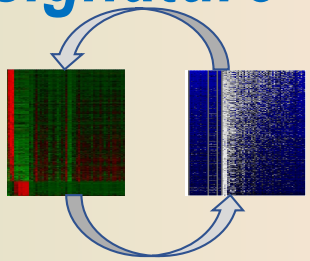
ON/OFF binary datasets

3- *correlate ectopic expressions with* ***clinical outcome***

4- *characterize* ***molecular profile of aggressive tumours***

5- *Understand* ***molecular basis of aggressive phenotype*** *by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)*

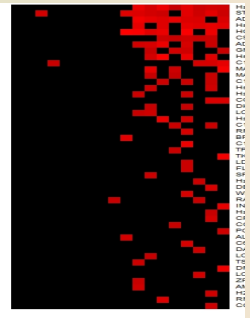
Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature



1- *identify* ***tissue-specific*** *genes*

cross different types of data

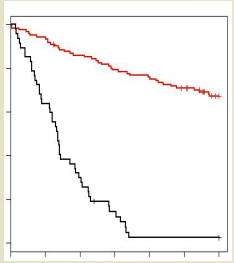
2- *detection of* ***ectopic expression*** *of tissue specific genes in cancer*



ON/OFF binary datasets

3- *correlate ectopic expressions with* ***clinical outcome***

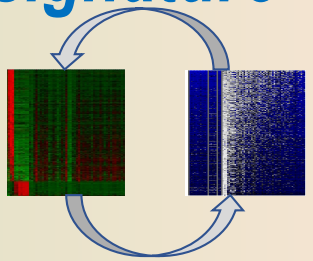
Correlate with clinical data and survival



4- *characterize* ***molecular profile of aggressive tumours***

5- *Understand* ***molecular basis of aggressive phenotype*** *by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)*

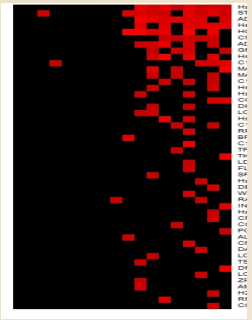
Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature



1- identify **tissue-specific** genes

cross different types of data

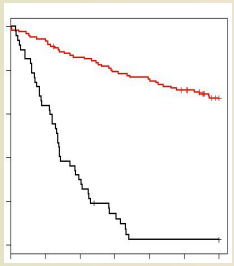
2- detection of **ectopic expression** of tissue specific genes in cancer

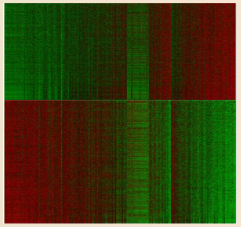


ON/OFF binary datasets

3- correlate ectopic expressions with **clinical outcome**

Correlate with clinical data and survival



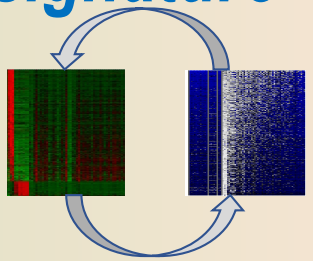


4- characterize **molecular profile of aggressive tumours**

Supervised transcriptomic analysis

5- Understand **molecular basis of aggressive phenotype** by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)

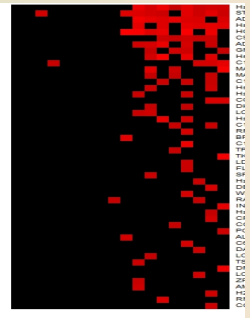
Ectopic gene activations for marker discovery and to explore aggressive tumor molecular signature



1- identify **tissue-specific** genes

cross different types of data

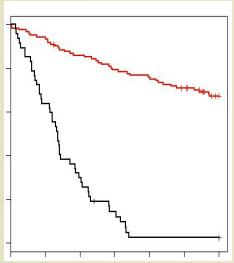
2- detection of **ectopic expression** of tissue specific genes in cancer

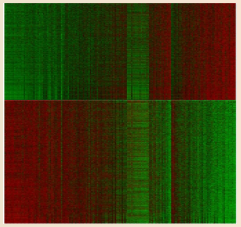


ON/OFF binary datasets

3- correlate ectopic expressions with **clinical outcome**

Correlate with clinical data and survival

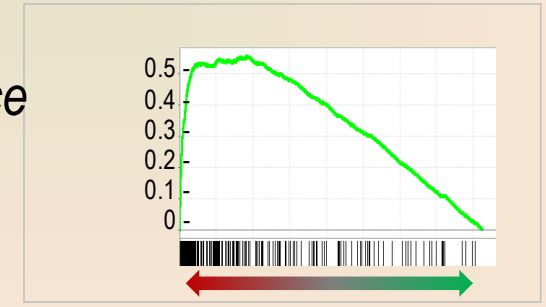




4- characterize **molecular profile of aggressive tumours**

Supervised transcriptomic analysis

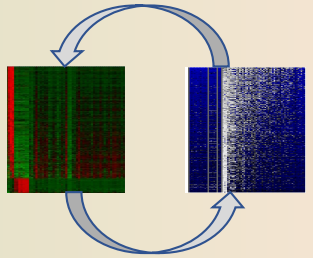
5- Understand **molecular basis of aggressive phenotype** by comparing with other expression profiles (Gene Set Enrichment Analysis), and using available tools to explore biological significance (Gene Ontology terms, pathways…)



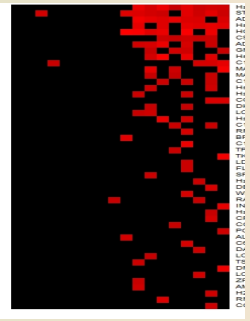
GSEA, …

Specificities of our bioinformatic pipelines

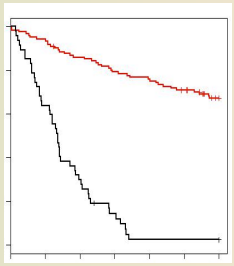
cross different types of data

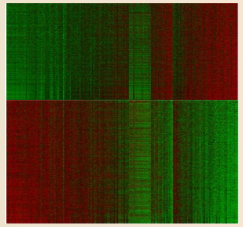


ON/OFF binary datasets



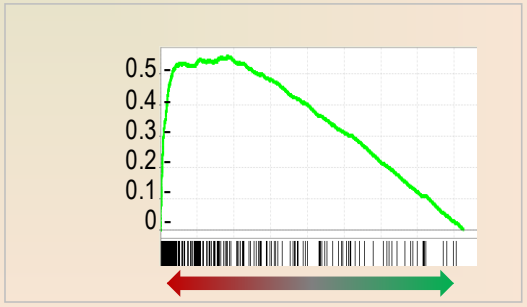
Correlate with clinical data and survival



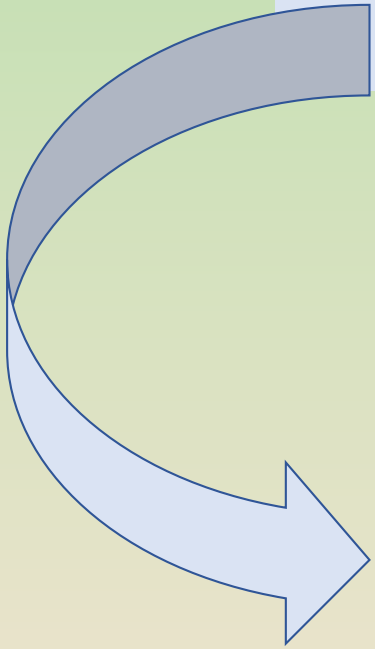


Supervised transcriptomic analysis

GSEA, …



More data = challenging opportunities



Increasing amounts of OMICs data

Update and scale-up strategies

+

Develop new pipelines

Increasing power of analysis

+

address new questions

Updating pipelines

Issues to be solved

1- Matching gene and tissues annotations and IDs

2- Increasing volumes of data (NGS)

3 - new types of data (RNAseq, ChIPseq..)

4- homogenize clinical annotations and data

5- Increase efficiency of pipelines for combined analysis of large and heterogeneous datasets

Updating pipelines

Issues to be solved

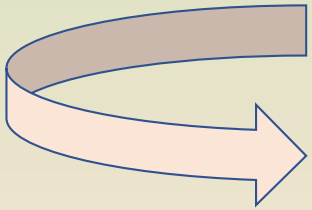
1- Matching gene and tissues annotations and IDs

2- Increasing volumes of data (NGS)

3 - new types of data (RNAseq, ChIPseq..)

4- homogenize clinical annotations and data

5- Increase efficiency of pipelines for combined analysis of large and heterogeneous datasets



Florent Chuffart

Ekaterina Flin

Updating pipelines

Ekaterina Flin

Navigate through omics and clinical data

1. Gene/genome annotations

Gene-related annotations are extremely **complex** and frequently modified

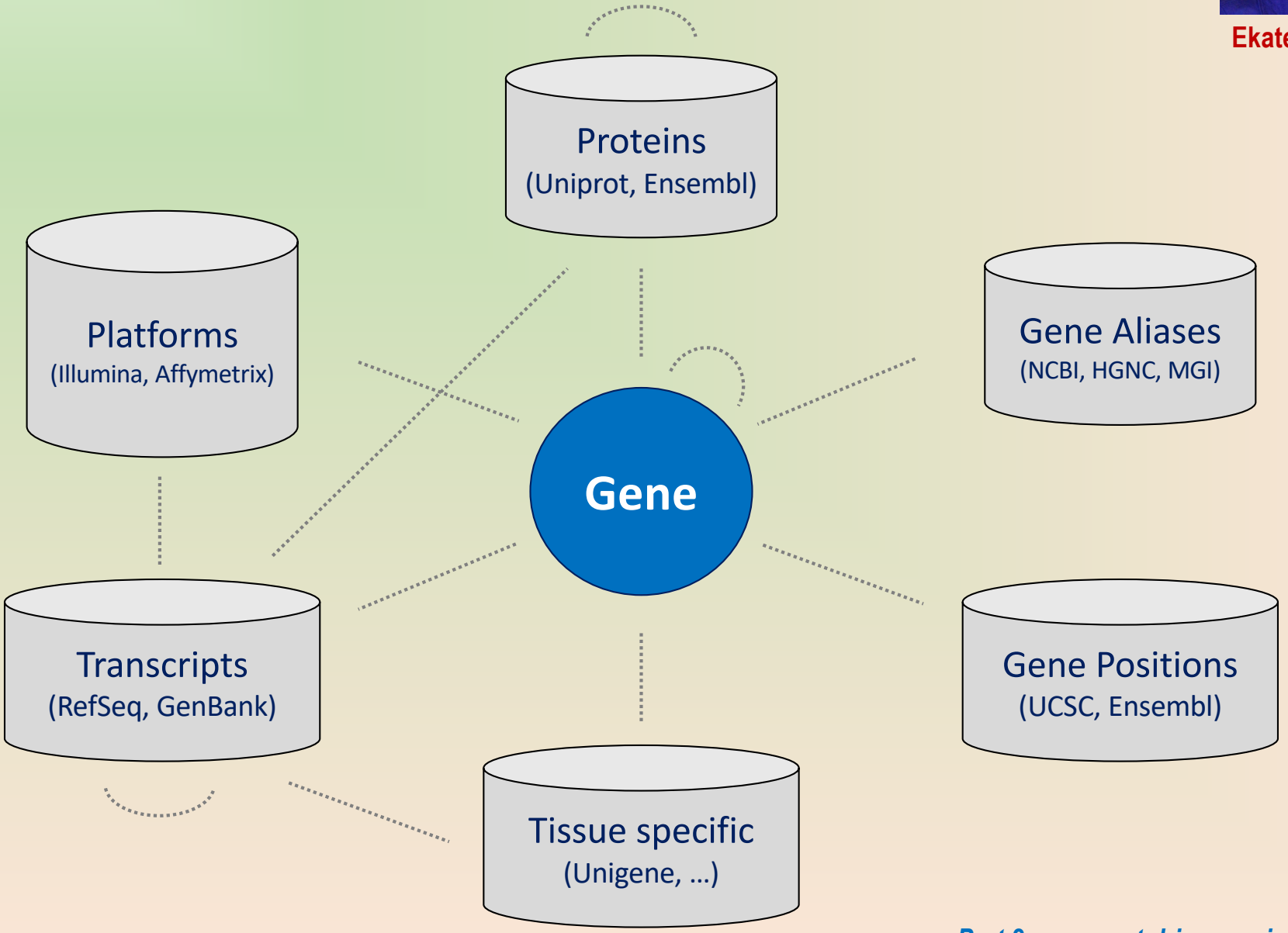
2. Clinical data and tissue annotations

Clinical data are **heterogeneous**, especially for tissue annotations which are usually different in different datasets

Updating pipelines

Navigate through gene-related data

Ekaterina Flin

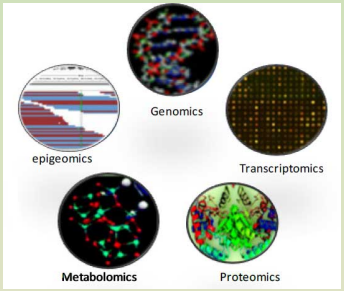


Updating pipelines

EpiMed Information System

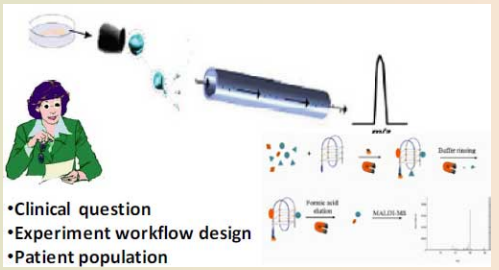
Ekaterina Flin

- Different data types (clinical, omics) from different platforms (Affymetrix, Illumina,…) with different data formats (txt, excel, pdf, raw data)
- Various access to public databases (sql, html, xml, json, web services)
- Scripts and pipelines for data analysis (Python, R)

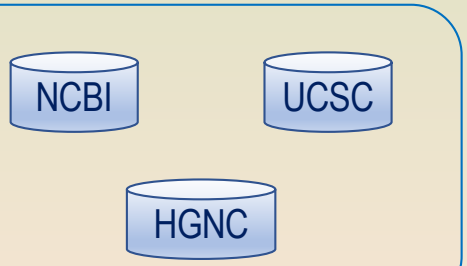


Omics / NGS Data

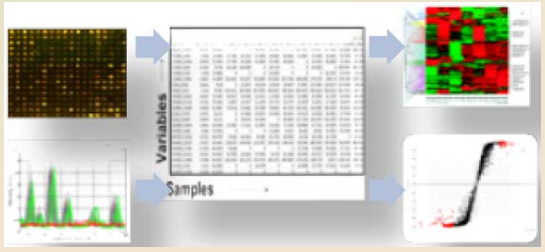
EpiMed



Clinical Data



External Databases



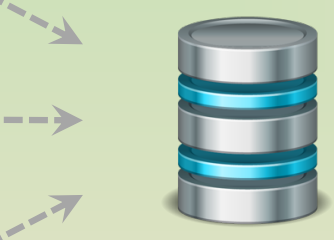
Data Analysis

Updating pipelines

… dedicated statistic and bioinformatic approaches

EpiMed Database: joint exploration of data

- Meta-data of omics platforms
- Clinical data (public/private)
- Dictionaries, classifications, nomenclatures
- Genes, proteins, assemblies



EpiMed Database

Ekaterina Flin

Updating pipelines

l dedicated statistic and bioinformatic approaches

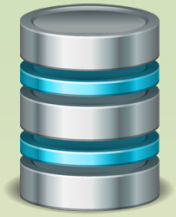
Meta-data of omics platforms

Clinical data (public/private)

Dictionaries, classifications, nomenclatures

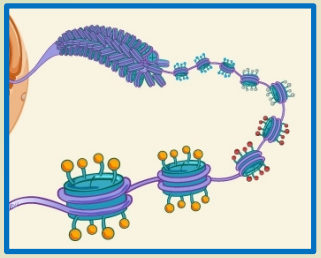
Genes, proteins, assemblies

EpiMed Database: joint exploration of data



EpiMed Database

Ekaterina Flin



Florent Chuffart

EpiMed tools: omic analysis toolbox

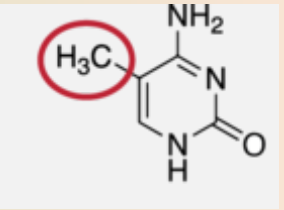
Dealing with more and more omic data…

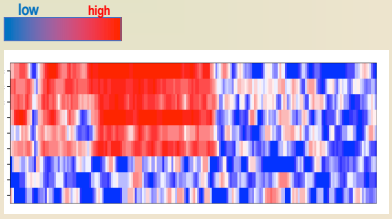
Update and implement **pipelines for analysis of omic data** of different origin
Improve analysis power and efficiency

Development of new **dedicated pipelines** to answer specific biological questions

=> Explore the epigenomic status/landscape of tissue-specific genes using NGS data

Florent Chuffart





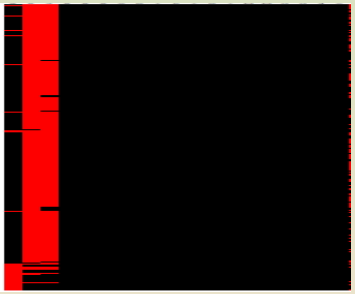
EpiMed tools : omic analysis toolbox

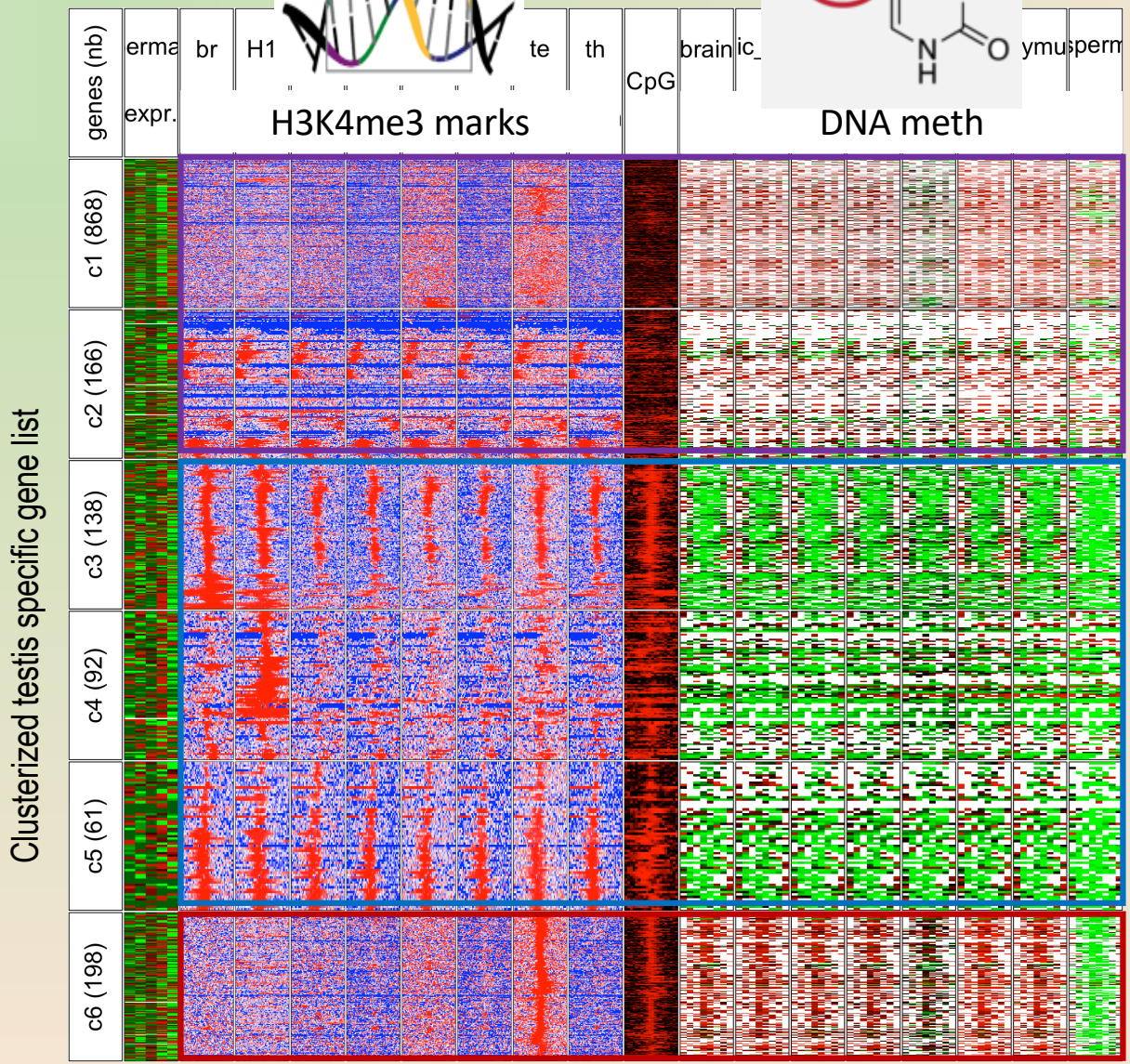
Tissue-specific genes have different epigenetic profiles

New pipelines : example

Florent Chuffart

Testis specific genes





=> Hypothesis : Different epigenetic mechanism are involved in the repression of silent genes

EpiMed information system and pipelines

Dedicated statistic and bioinformatic approaches

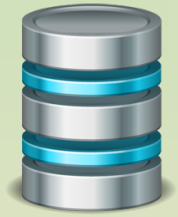
Meta-data of omics platforms

Clinical data (public/private)

Dictionaries, classifications, nomenclatures

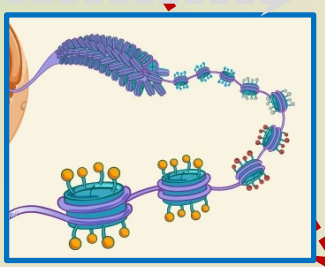
Genes, proteins, assemblies

EpiMed Database: joint exploration of data



EpiMed Database

Standard formats



EpiMed tools: omic analysis toolbox

Ekaterina Flin

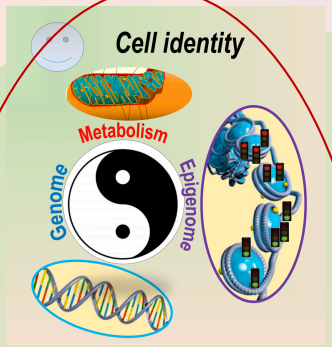
Florent Chuffart

http://epimed.univ-grenoble-alpes.fr/database/

Translational and multidisciplinary research

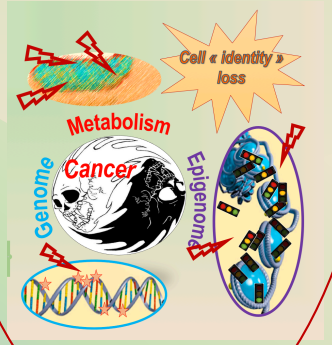
Environment

Metabolisms



Genetics

Epigenetics



http://epimed.univ-grenoble-alpes.fr/database/

http://epimed.univ-grenoble-alpes.fr/downloads/epimed_open_course/

https://epimed.github.io/

https://iab.univ-grenoble-alpes.fr/Plateformes/epigenetique-medicale-et-bioinformatique-epimed