

Analyse génomique intégrée des tumeurs

Eric Letouzé

Equipe "Génomique Fonctionnelle des Tumeurs Solides", INSERM U1162



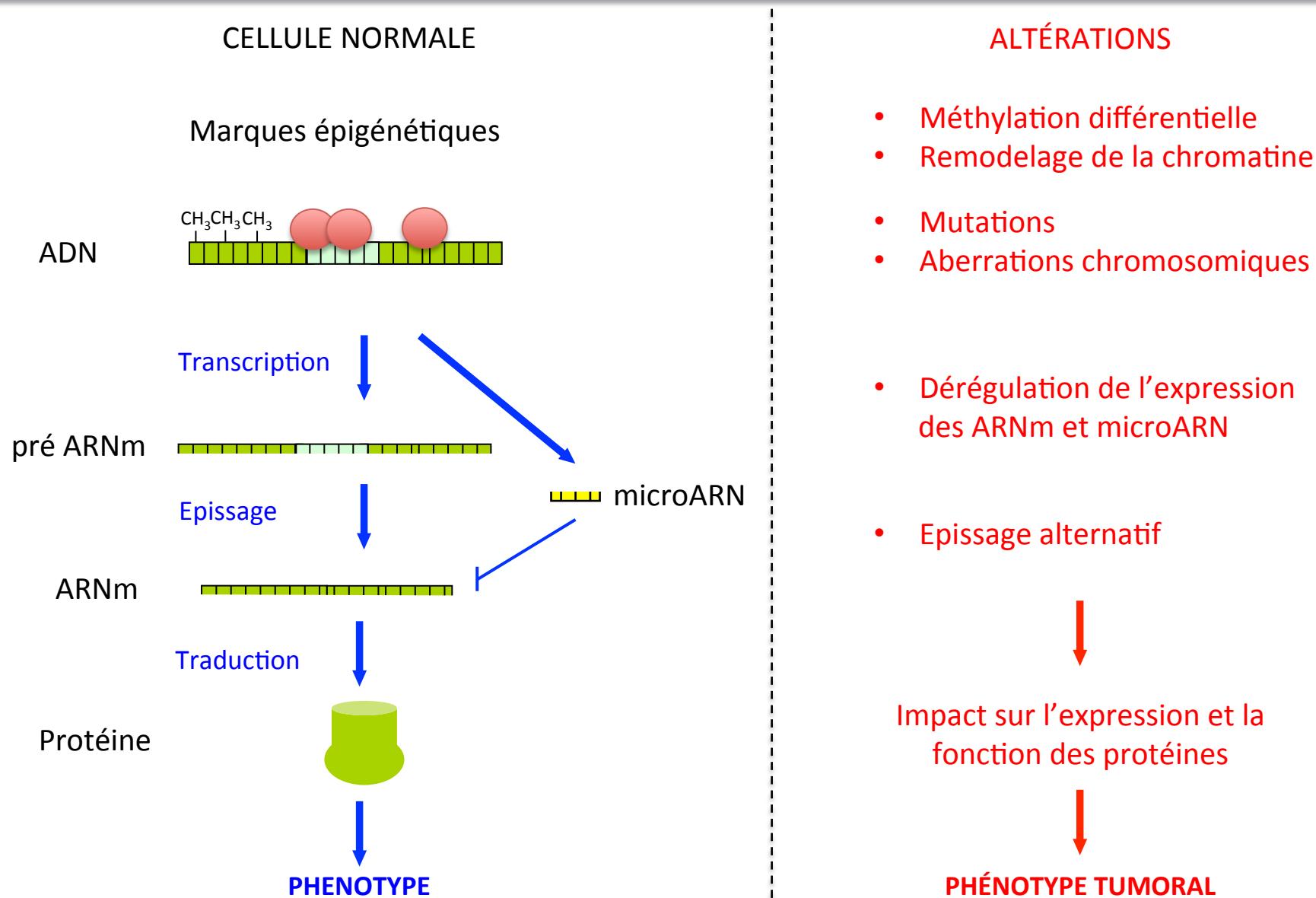
PLAN

1. Les technologies de la génomique à haut débit et leur application à l'étude des cancers
2. Analyse génomique intégrée des paragangliomes – lien entre méthylation de l'ADN et métabolisme cellulaire

PLAN

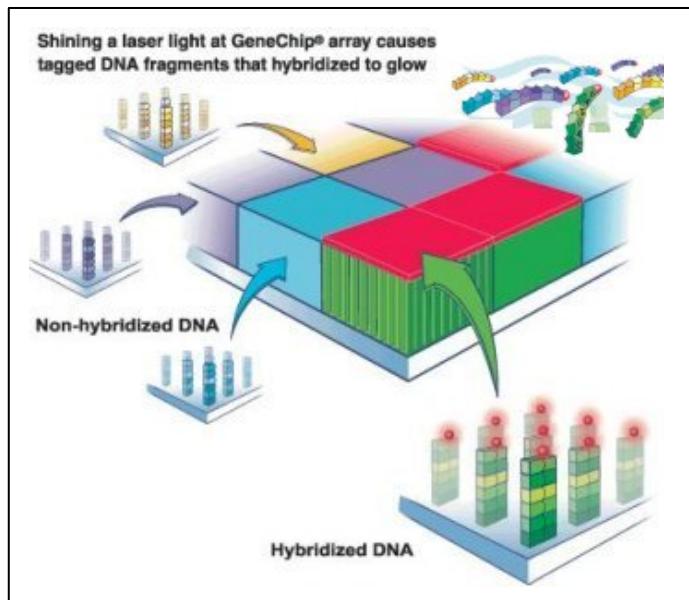
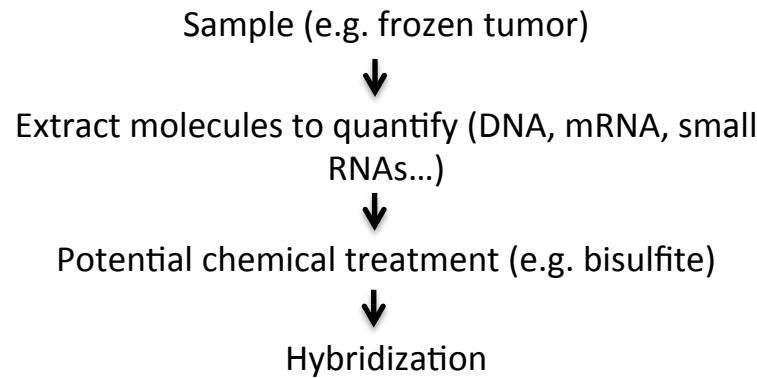
1. Les technologies de la génomique à haut débit et leur application à l'étude des cancers
2. Analyse génomique intégrée des paragangliomes – lien entre méthylation de l'ADN et métabolisme cellulaire

Analyses génomiques des tumeurs : que mesure-t-on, et comment ?



Puces vs. séquençage

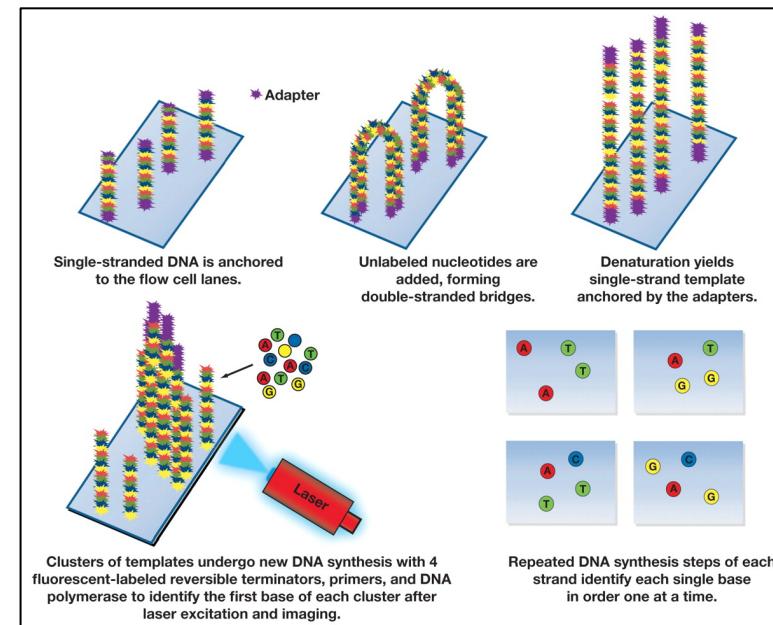
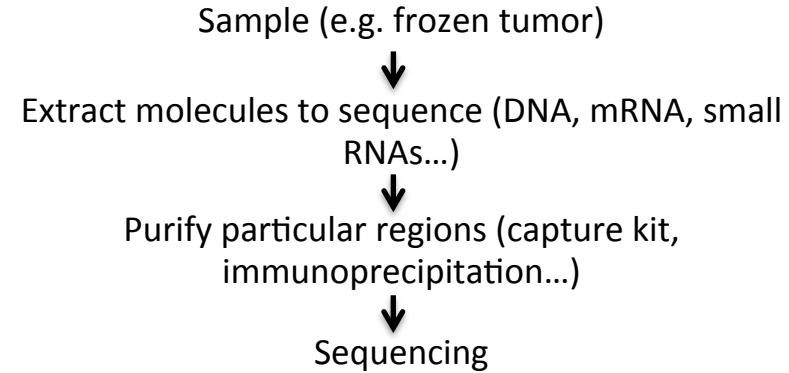
Puces à ADN



↓

Analysis (signal quantification, normalization...)

Séquençage à haut débit



↓

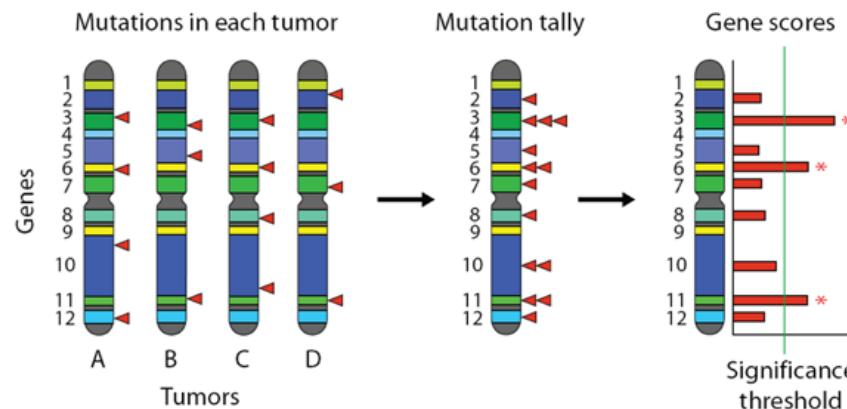
Analysis (alignement, quantification, variant discovery...)

Analyse des mutations

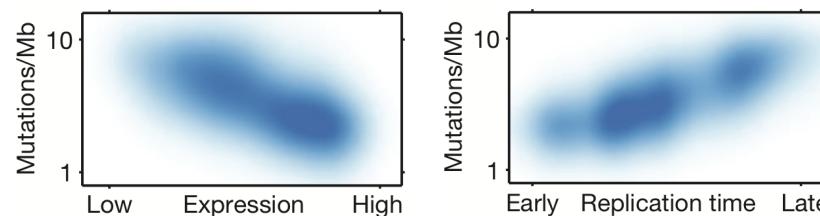
- **Goal:** Identify genes with recurrent mutations in cancer, which may correspond to driver genes.
- **Data:** Whole exome (~50 Mb) or whole genome (>3 Gb) sequences.
- **Analysis:**
 - 1) Identify somatic variants
 - Align reads to the human genome
 - Detect single nucleotide variant(s) and small indels
 - Remove low quality variants (read depth, Qphred...)
 - Compare normal and tumor sequences to detect somatic variants
 - Annotate the variants (referenced, predicted functional impact...)
 - 2) Interpret the variants
 - Examine the mutational spectrum
 - Find significantly mutated genes or pathways
 - Integrate with clinical annotations and other omics data

MÉTHODE : Identification des gènes significativement mutés

- We want to find genes that harbor **more mutations than expected by chance**, reflecting **selection** of driver genes.



- The number of mutations expected by chance depends on the **size of the gene** and on the **background mutation rate (BMR)**.
- The BMR depends on the **sample**, the **mutation category**, and the **genomic context** (expression level, replication time, chromatin context).

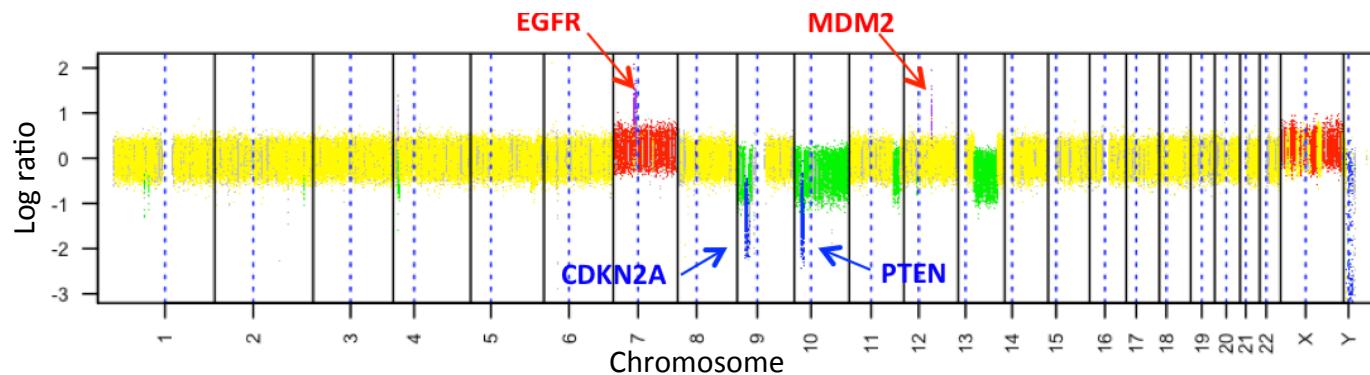


Lawrence *et al.*, Nature 2013

- The **MutSigCV algorithm** evaluates the mutation rate $\mu_{g,c,p}$ for each gene g , category c and patient p . A p-value is then computed for each gene, taking into account these mutation rates.

Analyse des aberrations chromosomiques

- **Goal:** Identify recurrent copy-number changes (gains, losses), which may point out tumor suppressors and oncogenes.

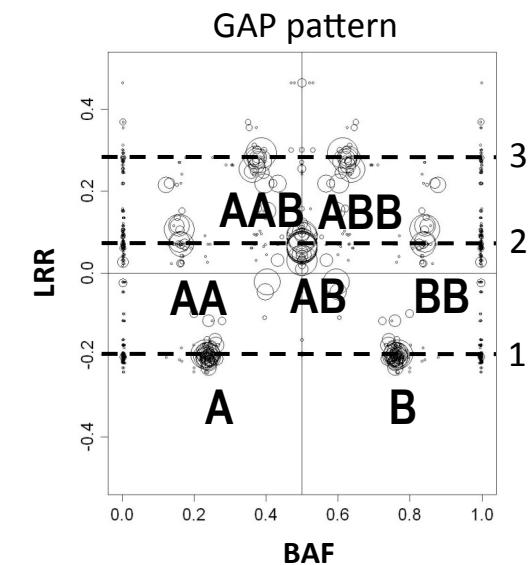
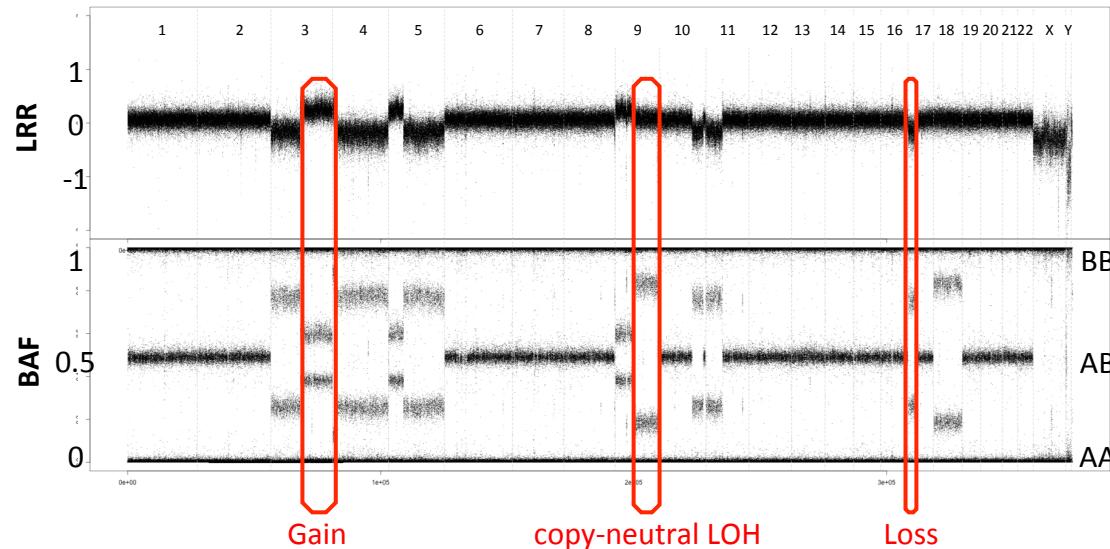


- **Data:** SNP array, exome or whole-genome sequences.
- **Analysis:**
 - 1) Identify copy-number alterations (CNAs) in each tumor
 - 2) Integrate CNAs across the cohort to find recurrent alterations and candidate driver genes

MÉTHODE : Utilisation des SNPs pour détecter les aberrations chromosomiques

- SNP profiles are analyzed using both the log R ratio (LRR) and B allele frequency (BAF) profiles:

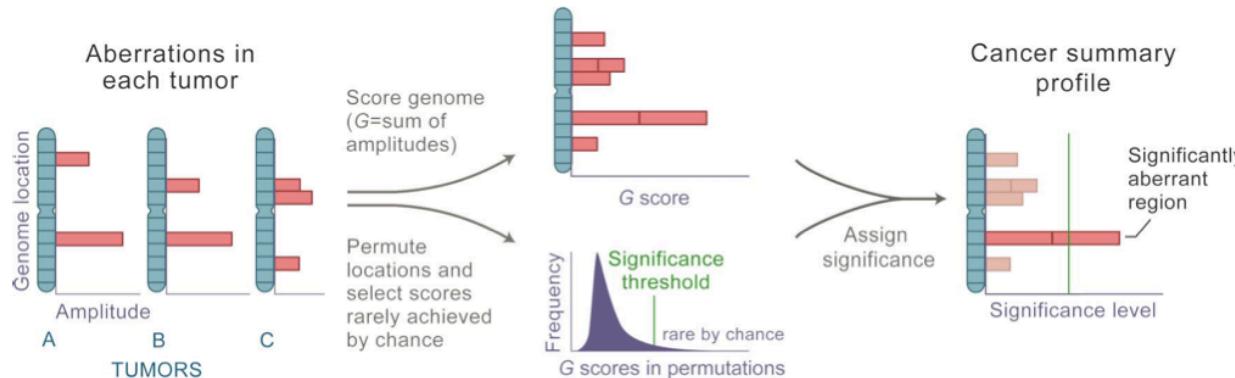
$$LRR = \log\left(\frac{(A + B)_{tum}}{(A + B)_{ref}}\right) \quad BAF = \frac{B}{A + B}$$



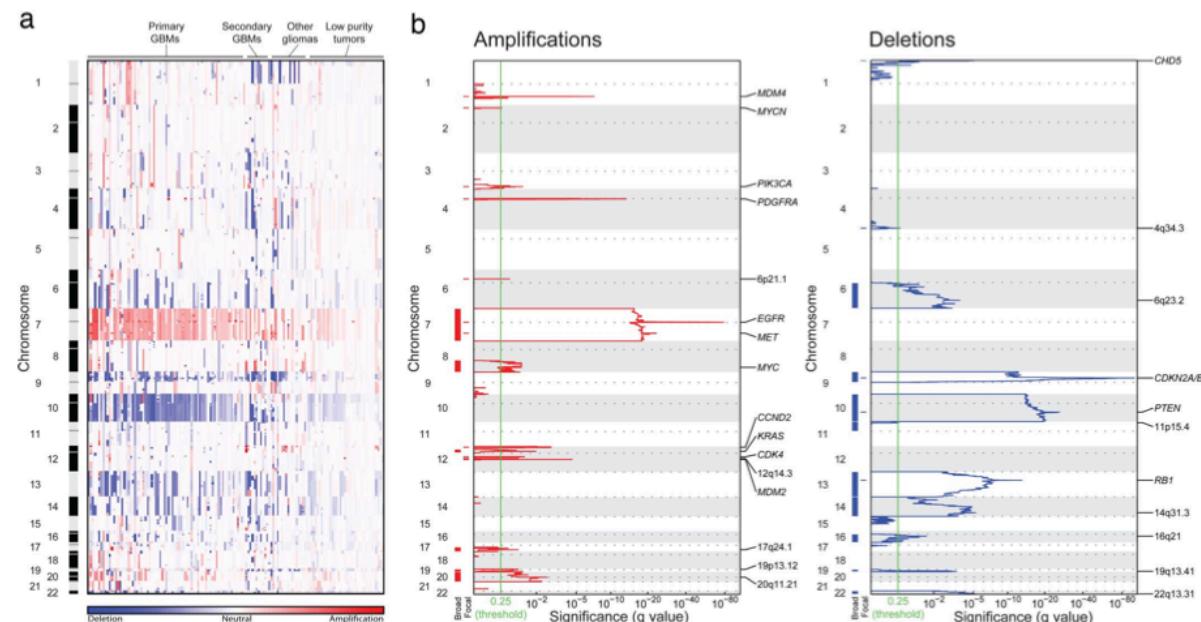
-> The **GAP method** estimates the **absolute copy-number** of each segment, the **ploidy** of the sample, and the level of **contamination** by normal cells (Popova *et al.*, Genome Biol 2009).

MÉTHODE : Détection des aberrations chromosomiques récurrentes

- We want to find chromosome gains and losses that are **significantly** recurrent across the dataset.
- The **GISTIC algorithm** uses both aberration frequencies and amplitudes to derive a G-score, which significance is determined as compared to a null distribution:

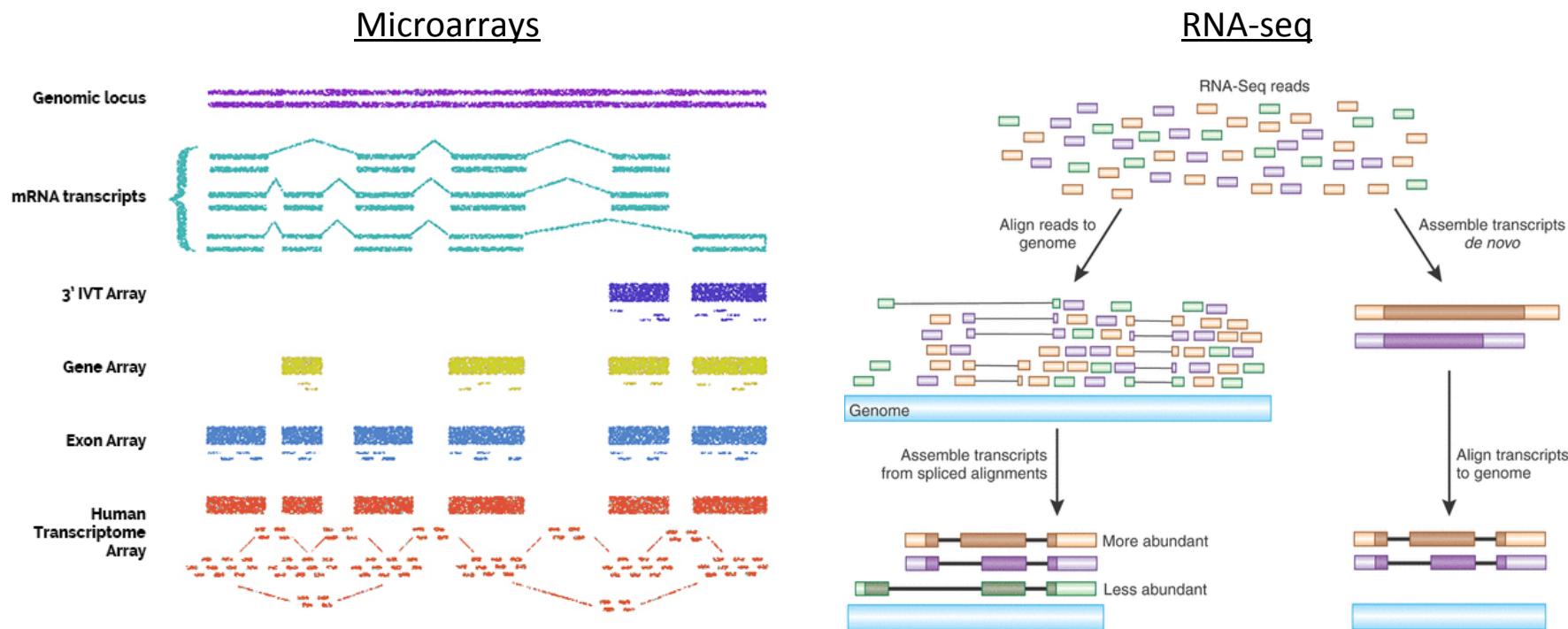


- GISTIC analysis of glioblastomas:



Analyse du transcriptome

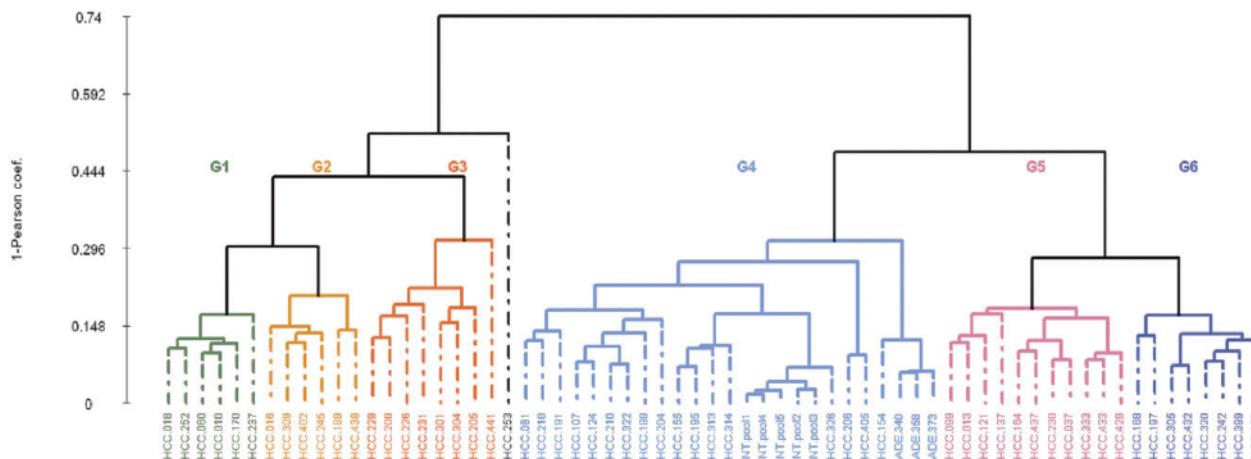
- **Goal:** Detect gene expression changes (mRNA or miRNA) and identify molecular subgroups.
- **Data:** mRNA or miRNA expression arrays, RNA-seq.



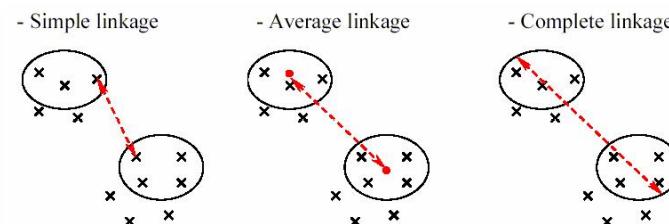
- **Analysis:**
 - Identify molecular subgroups (unsupervised)
 - Identify deregulated genes and pathways (supervised)
 - Define molecular predictors of diagnosis and prognosis

MÉTHODE : Clustering hiérarchique

- **Hierarchical clustering** is a classical approach to identify molecular subgroups of tumors.



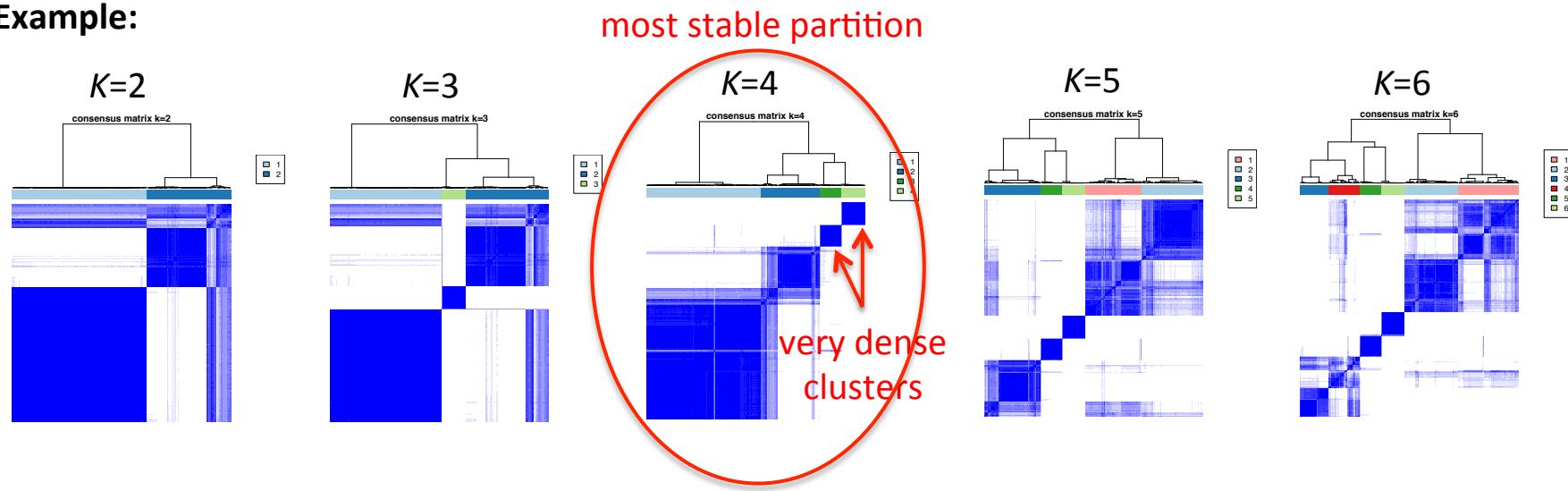
- This bottom-up agglomerative algorithm requires a **distance metric** (Euclidean, Pearson correlation...) and a **linkage criteria** (simple, average, complete...).



- BUT → How stable are the clusters?
→ What partition should be selected?

- **Consensus clustering** is an approach to investigate the stability of clusters and determine the best partition.
- **Method:**
 - 1) For each number K of clusters, repeat 1,000 times:
 - randomly select x % of probes and y % of samples
 - compute a partition in K clusters
 - 2) Calculate the co-classification value (0-1) for each pair of samples
 - 3) Clusterize the samples using the co-classification values.

- **Example:**

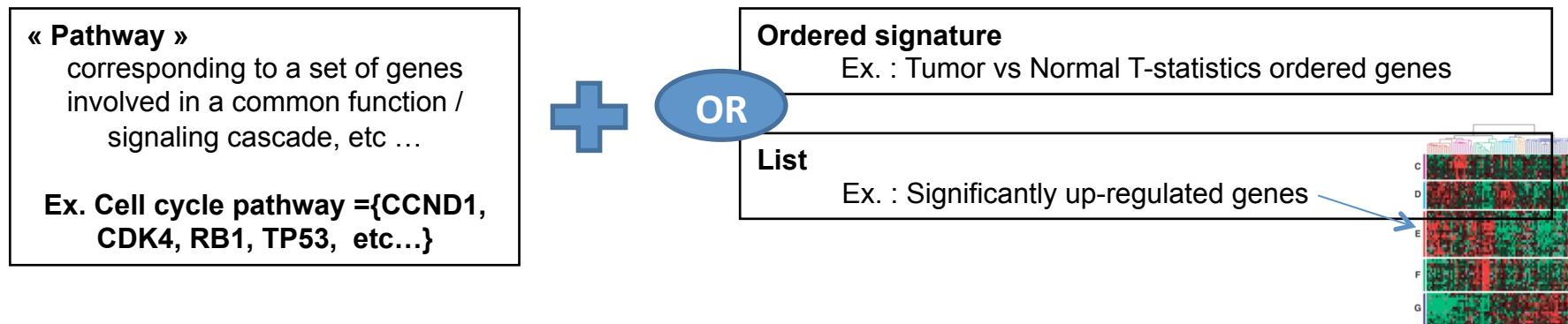


- Consensus clustering can also be applied to **other omics data** (e.g. DNA methylation profiles).

MÉTHODE : Analyse d'enrichissement – voies cellulaires

Goal: Extract the biological meaning of a set of genes deregulated in a cancer/subgroup.

Input



Two main approaches

Are the genes from this pathway uniformly distributed in the ordered signature ?



OR

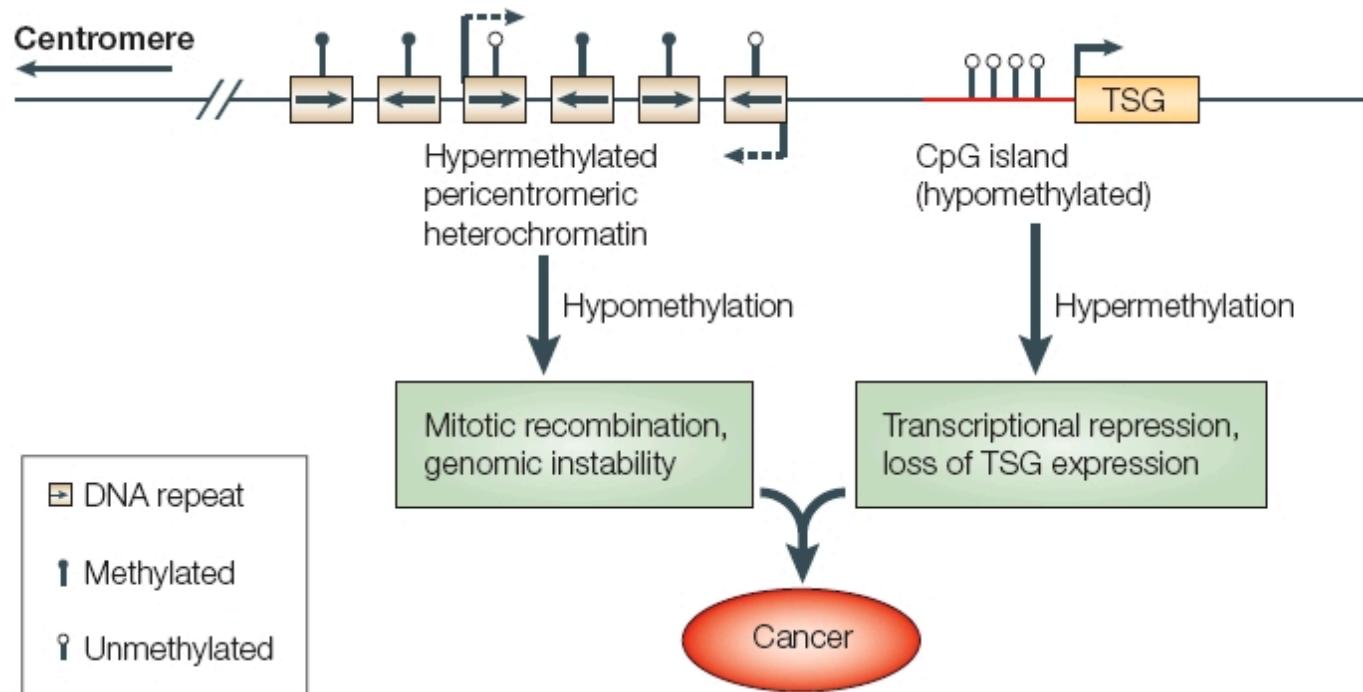
Are the genes from this pathway « enriched » in the list (vs outside) ?

	# in pathway	# out of pathway
# in list	a	b
# out of list	c	d

=> Hypergeometric test

ALTÉRATIONS DE LA MÉTHYLATION DE L'ADN DANS LES CANCERS

- General features:

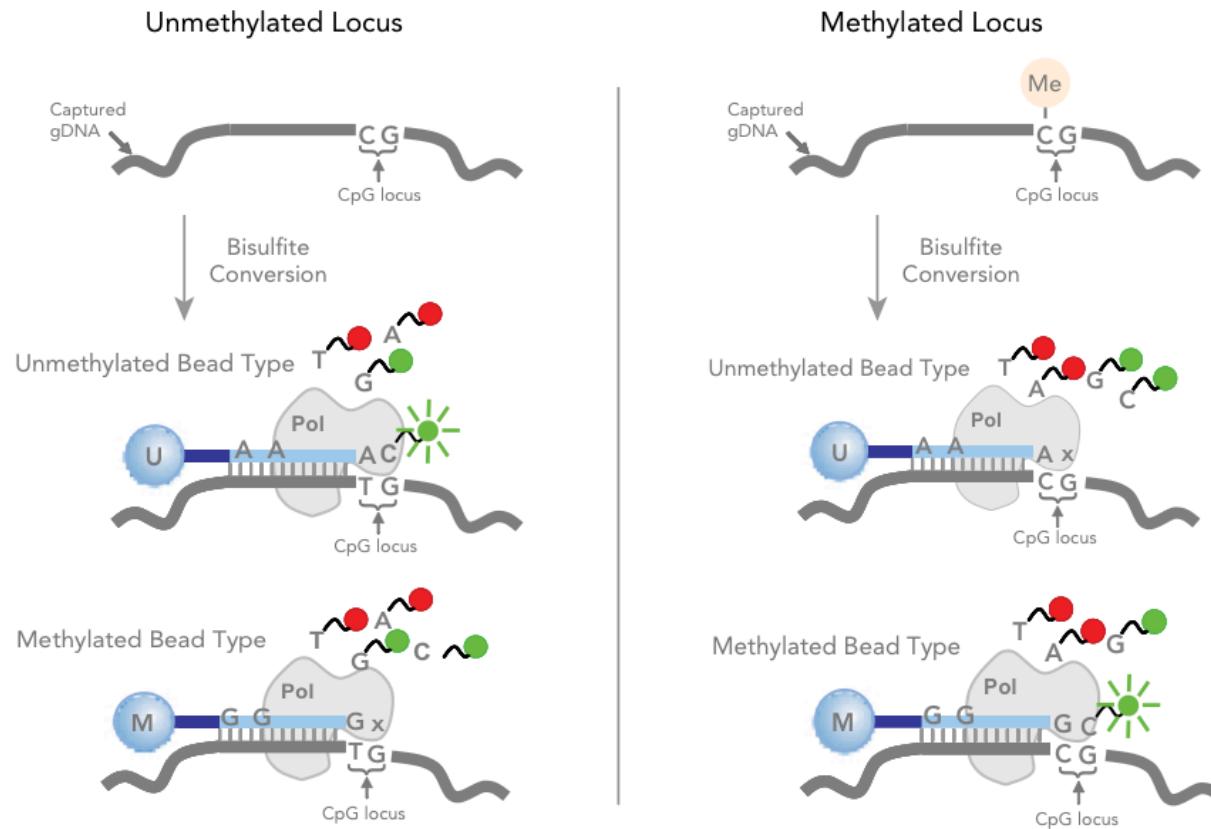


Phillips *et al.*, Nat Education 2008

- **CpG island methylator phenotypes (CIMP)** have been described in colorectal cancer (Toyota *et al.*, PNAS 1999), glioma (Nooshmehr *et al.*, Cancer Cell 2010) and other cancers, characterized by the concerted hypermethylation of a large number of genes.
- The glioma CIMP phenotype is associated with isocitrate dehydrogenase (IDH) mutations.

Puces "méthylation"

- The most widely used DNA methylation arrays are Illumina bead arrays (450,000 CpG sites):

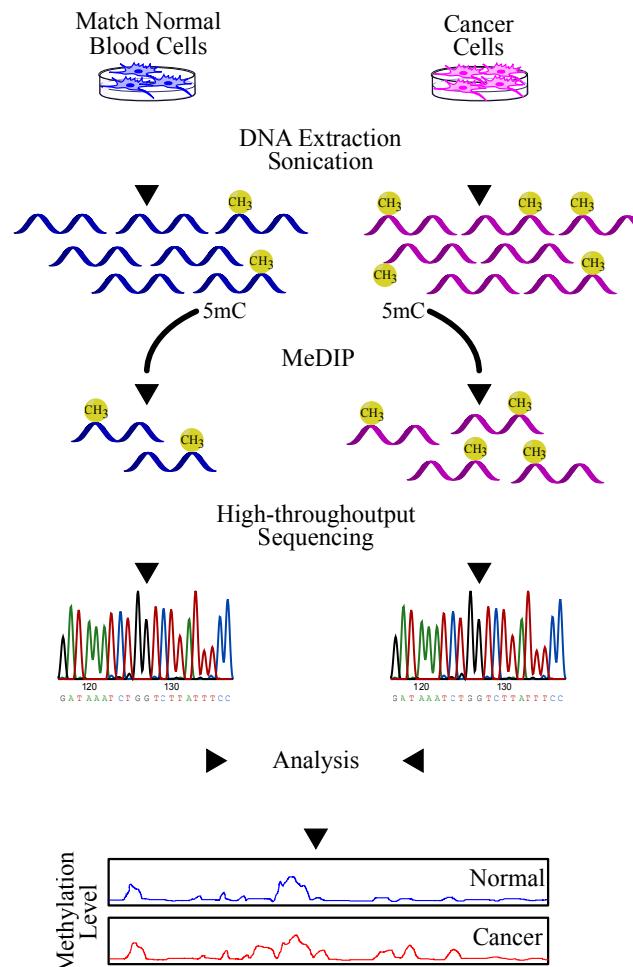


- Methylation at each CpG site is quantified as a beta value: $\beta = \frac{M}{M+U}$

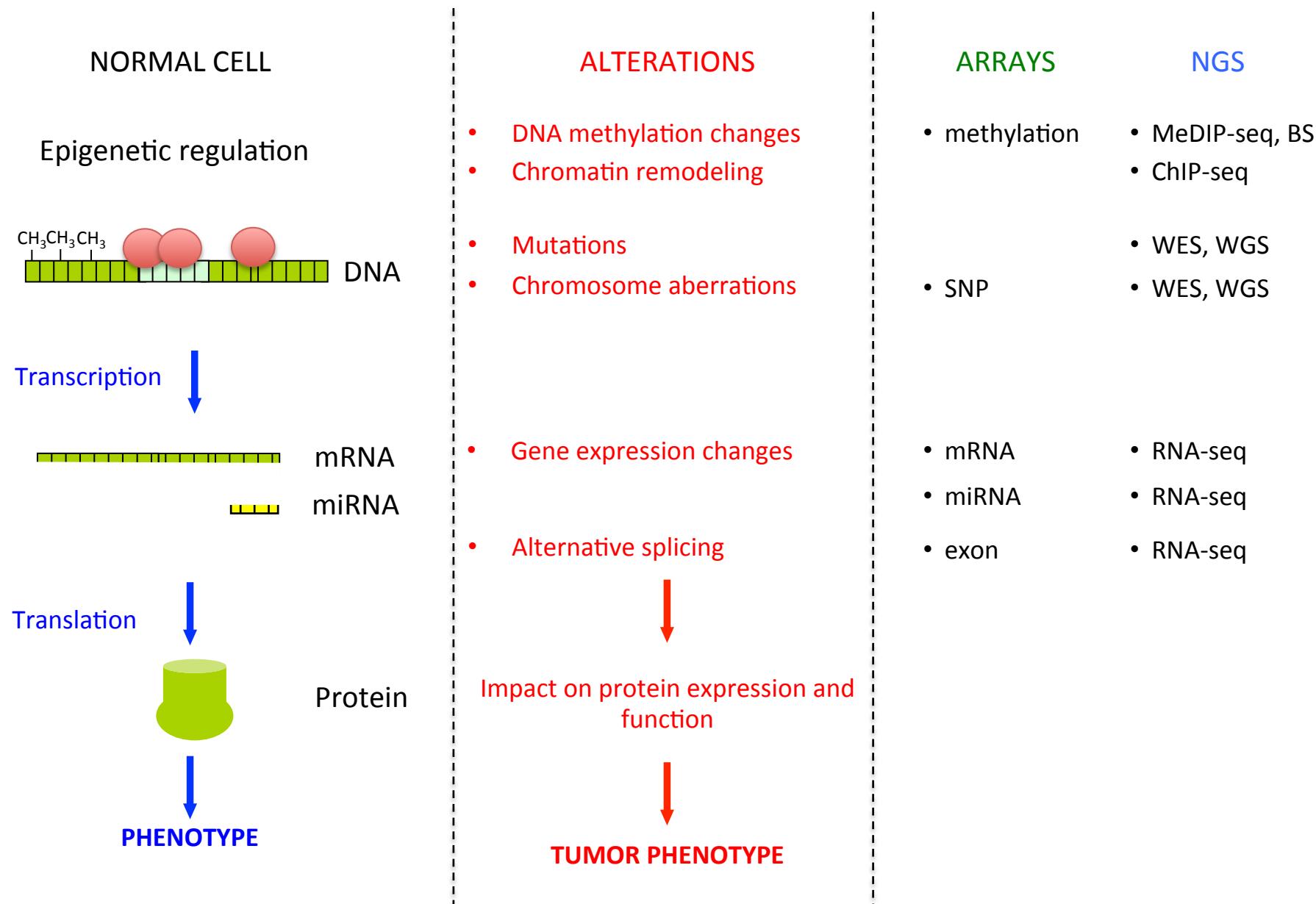
Analyse de méthylation par séquençage

Two main approaches:

- **Bisulfite sequencing** = bisulfite conversion followed by NGS (whole genome or RRBS)
- **MeDIP-seq** = immunoprecipitation-based approach.

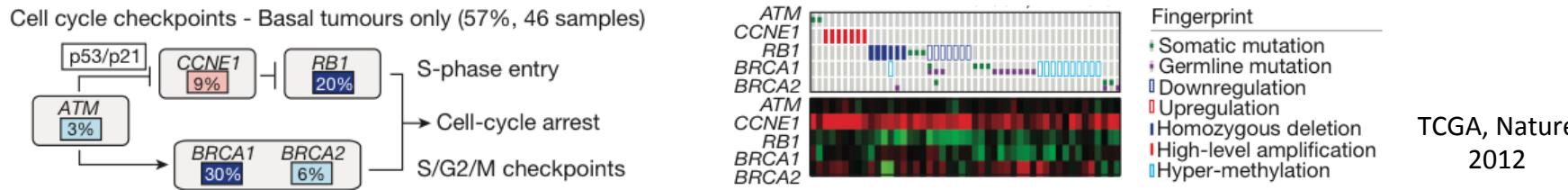


Bilan : Les technologies "omics" et leurs applications

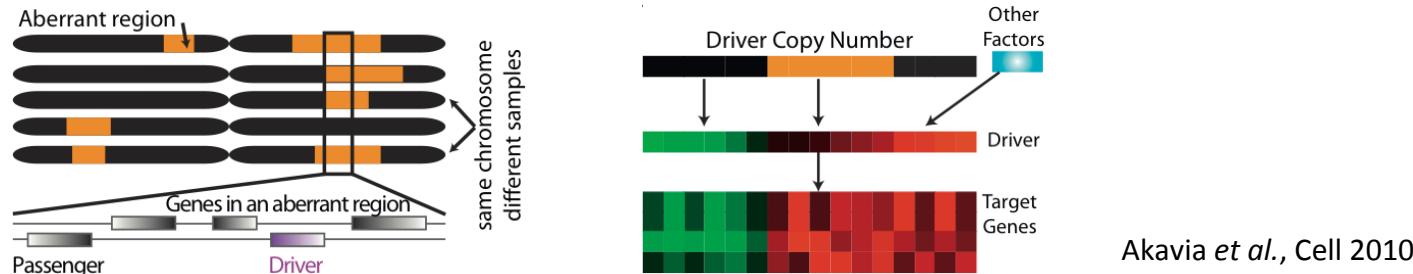


Intégration multi-omics

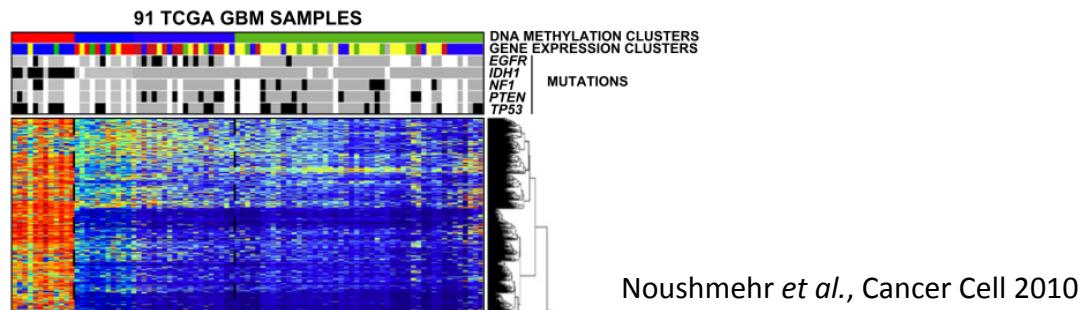
- Integrating several omics data is important to **refine oncogenic mechanisms** and **tumor subgroups**.
- **Example 1:** Several types of alteration can alter the same gene/pathway.



- **Example 2:** CONEXIC combines copy-number and expression data to identify driver genes.



- **Example 3:** Association between a mutation and a DNA methylation subgroup.

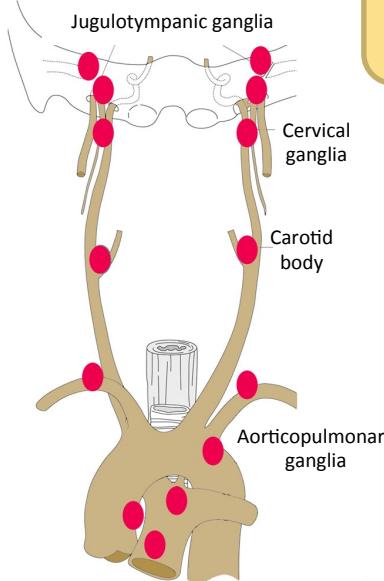


PLAN

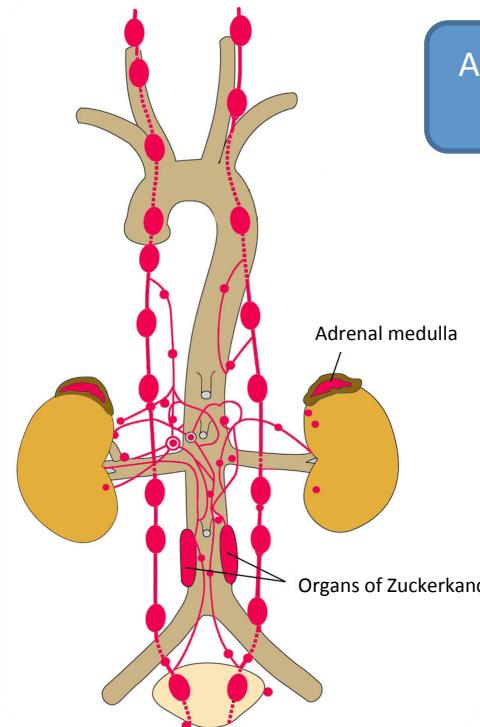
1. Les technologies de la génomique à haut débit et leur application à l'étude des cancers
2. Analyse génomique intégrée des paragangliomes – lien entre méthylation de l'ADN et métabolisme cellulaire

Phéochromocytomes et paragangliomes

- Pheochromocytomas (PCC) and paragangliomas (PGL) are rare tumors (2-8 cases per million) arising from neural crest-derived tissues.



PGL arise from
parasympathetic or
sympathetic ganglia



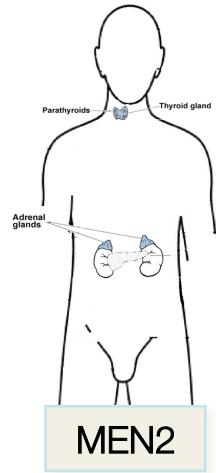
A PCC is a PGL that develops in
the adrenal medulla



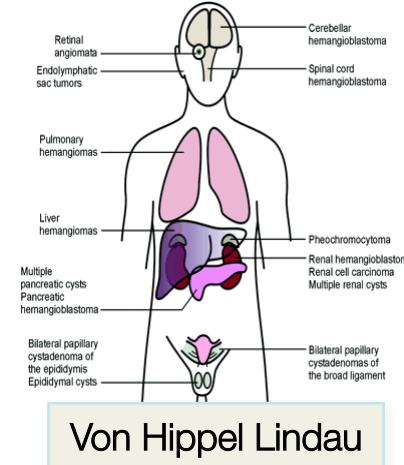
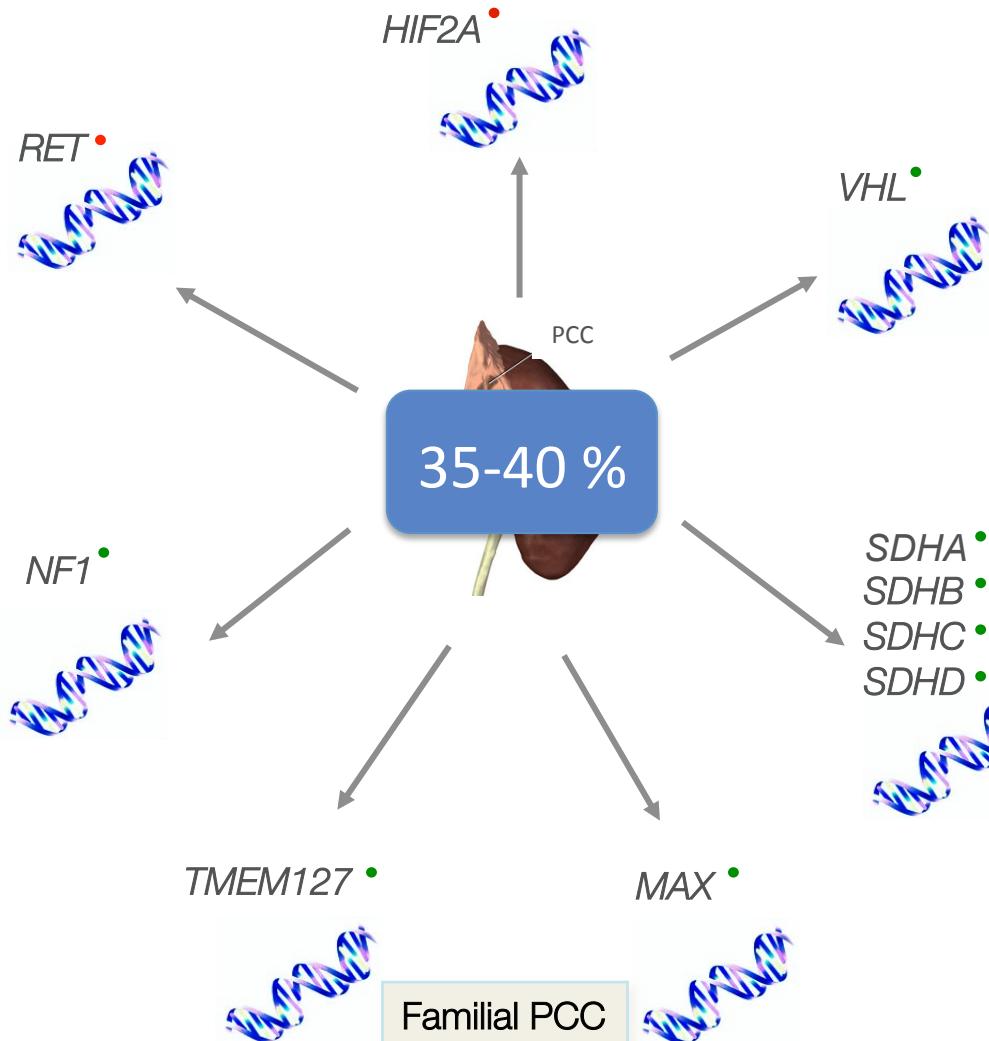
- PGL/PCC are neuroendocrine tumors often secreting catecholamines, resulting in hypertension, headache, sweating, palpitations...

Prédisposition génétique des paragangliomes

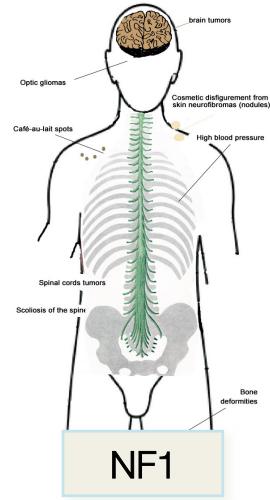
- PGL/PCC are often associated with hereditary syndromes:



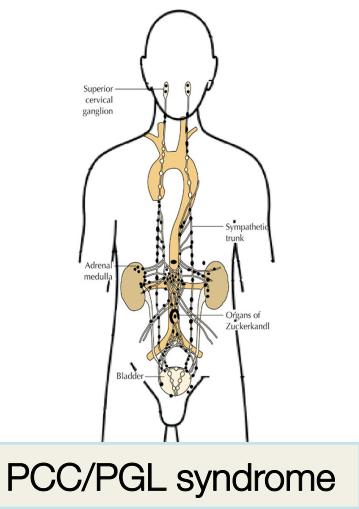
MEN2



Von Hippel Lindau



NF1

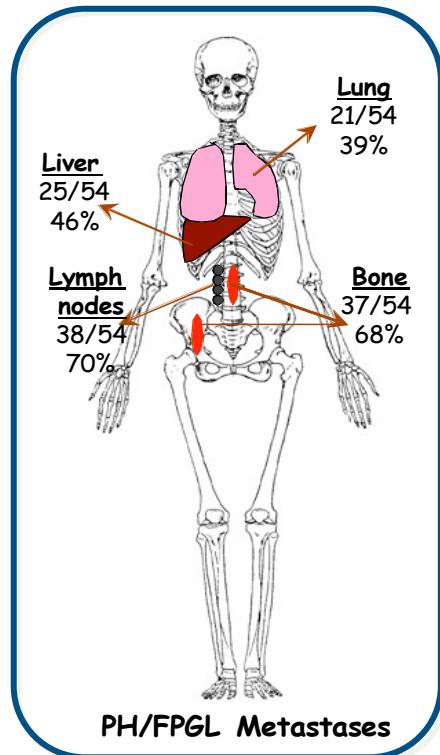


PCC/PGL syndrome

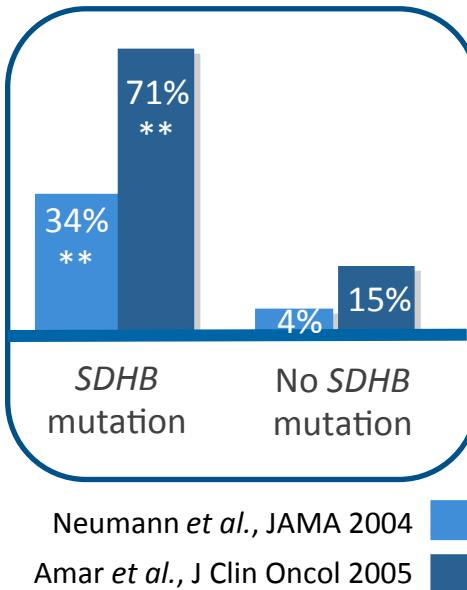
Les tumeurs mutées SDH sont particulièrement agressives ?

- PGL/PCC are generally benign but 10-15% of all cases develop metastases.
- SDHB mutations confer a much higher risk of metastasis and a shorter survival than patients with malignant tumors but no SDHB mutation.

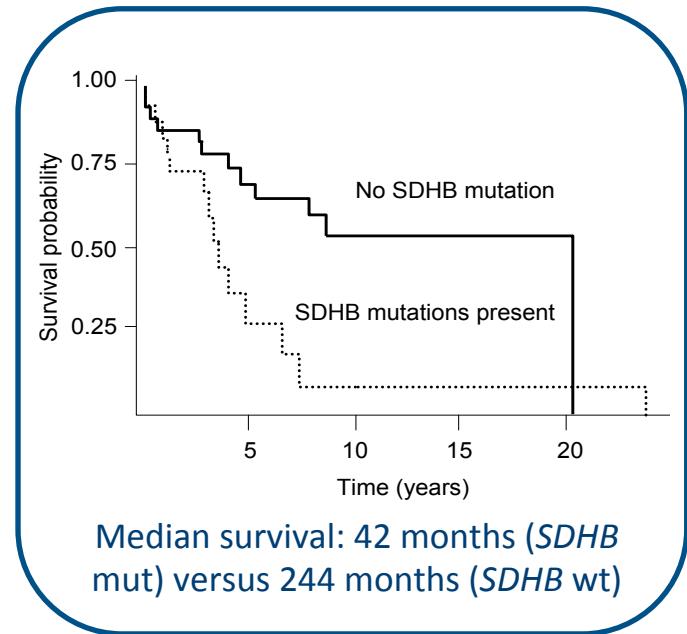
Malignancy: 10-15% of cases



Metastatic progression



Survival in malignant PGL/PCC



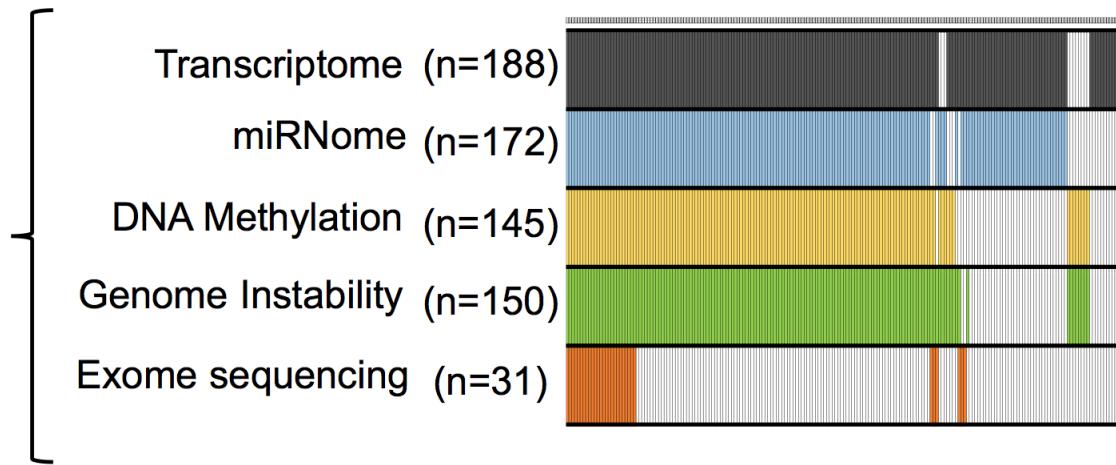
Gimenez-Roqueplo *et al.*, Cancer Res 2003

Amar *et al.*, J Clin Oncol 2005

Amar *et al.*, J Clin Endocrinol Metab 2007

Analyse génomique intégrée des paragangliomes (cohorte COMETE)

COMETE Network
HEGP/Cochin - 1993/2008
190 patients – 202 tumor samples

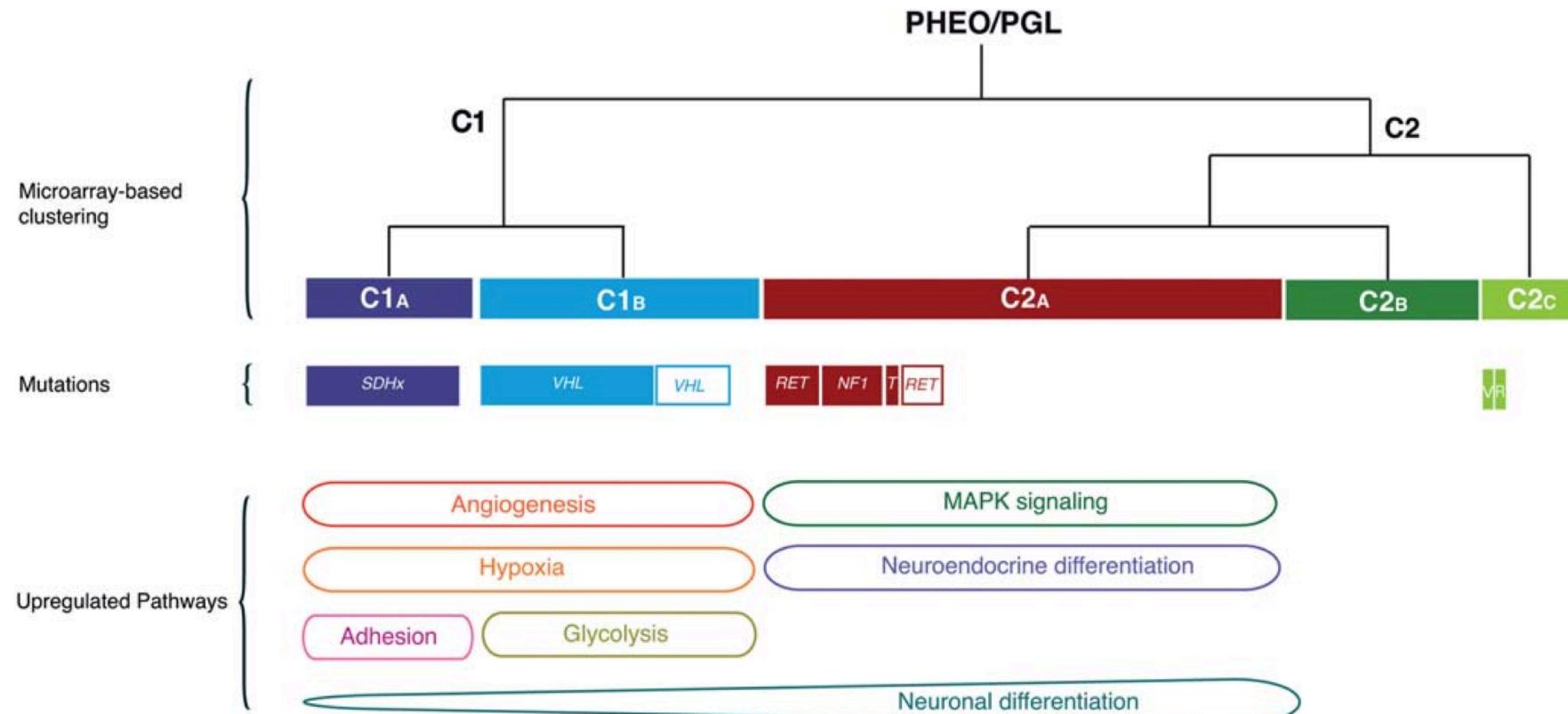


Main goals of the project:

- Identify new driver genes and oncogenic mechanisms
- Define the spectrum of alterations underlying tumorigenesis
- Explain the aggressiveness of *SDHB*-mutated case

Les groupes transcriptomiques de paragangliomes sont très associés aux gènes de prédisposition

- Transcriptome profiling of the CIT/COMETE cohort:

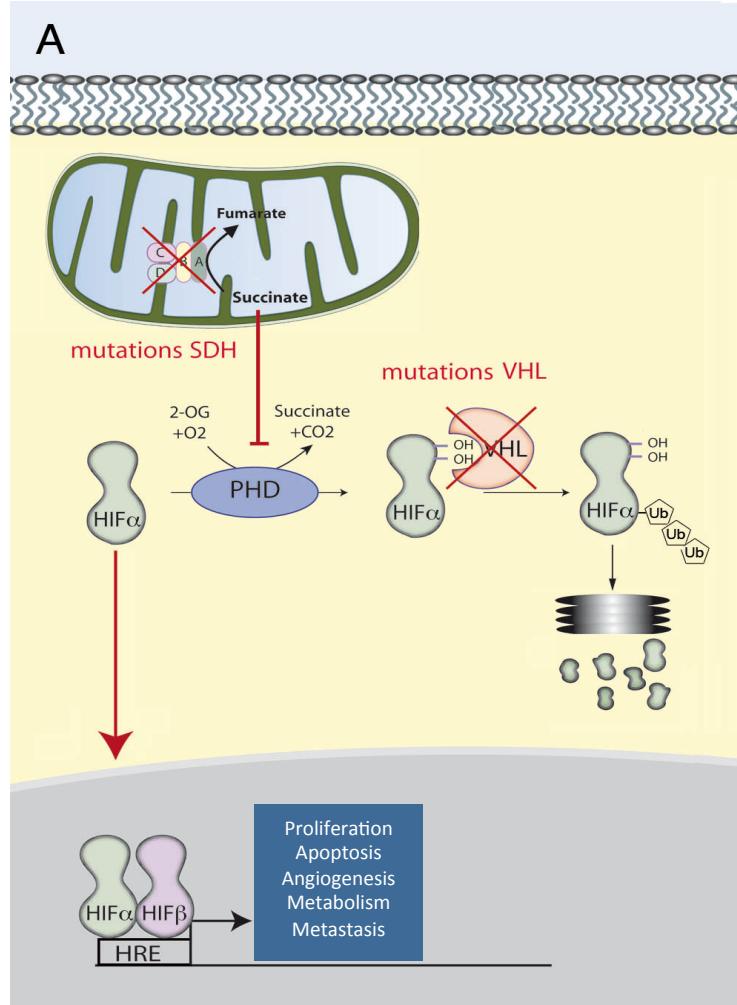


Burnichon *et al.*, HMG 2011

-> Tumors of the C1 cluster (*SDHx*, *VHL*) display an activation of hypoxia, whereas tumors of the C2A cluster (*RET*, *NF1*) display an activation of the MAP kinase pathway.

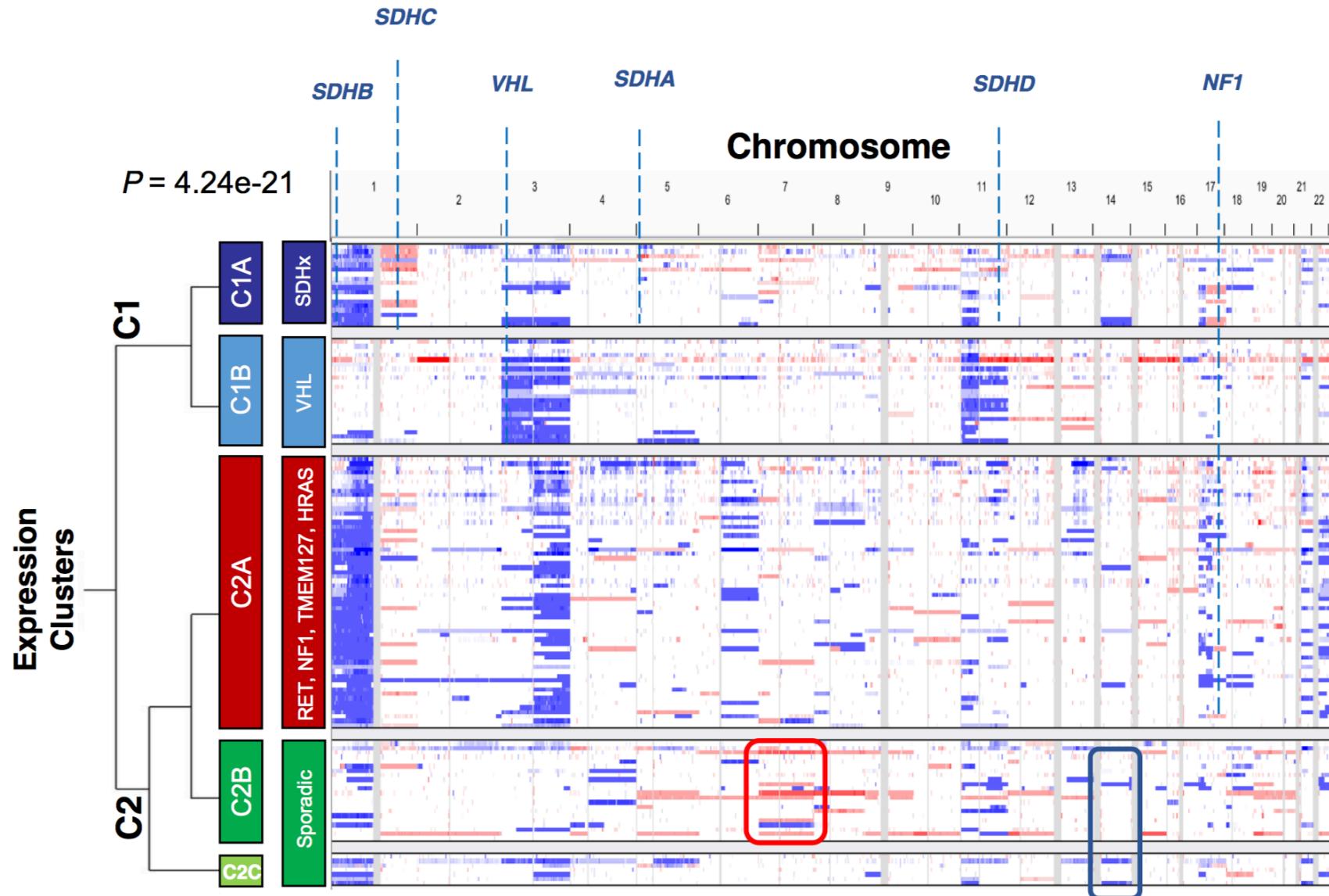
Deux voies cellulaires principales impliquées dans les paragangliomes

Cluster C1: hypoxia



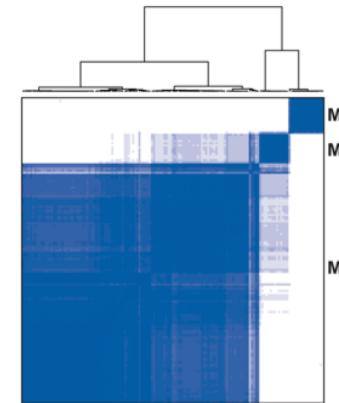
Adapted from Favier and Gimenez-Roqueplo, Médecine Sciences, 2012

Les aberrations chromosomiques sont très associées aux groupes moléculaires

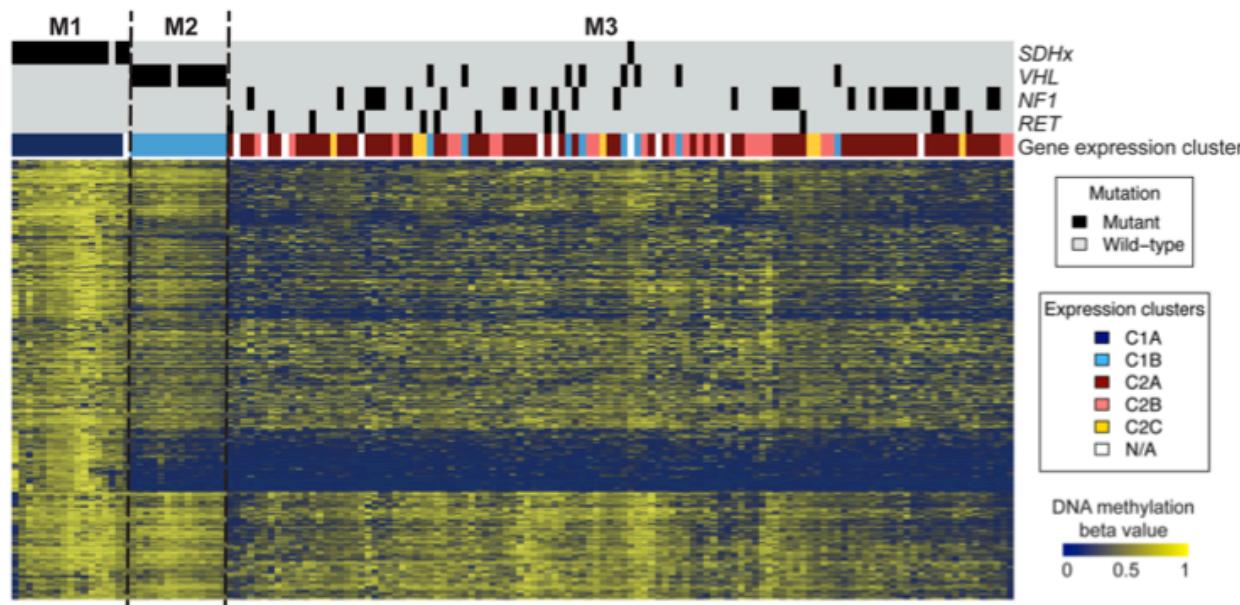


Classification du méthylome des paragangliomes

- Consensus clustering reveals 3 homogeneous tumor clusters:



- DNA methylation subgroups are highly associated with molecular subtypes:

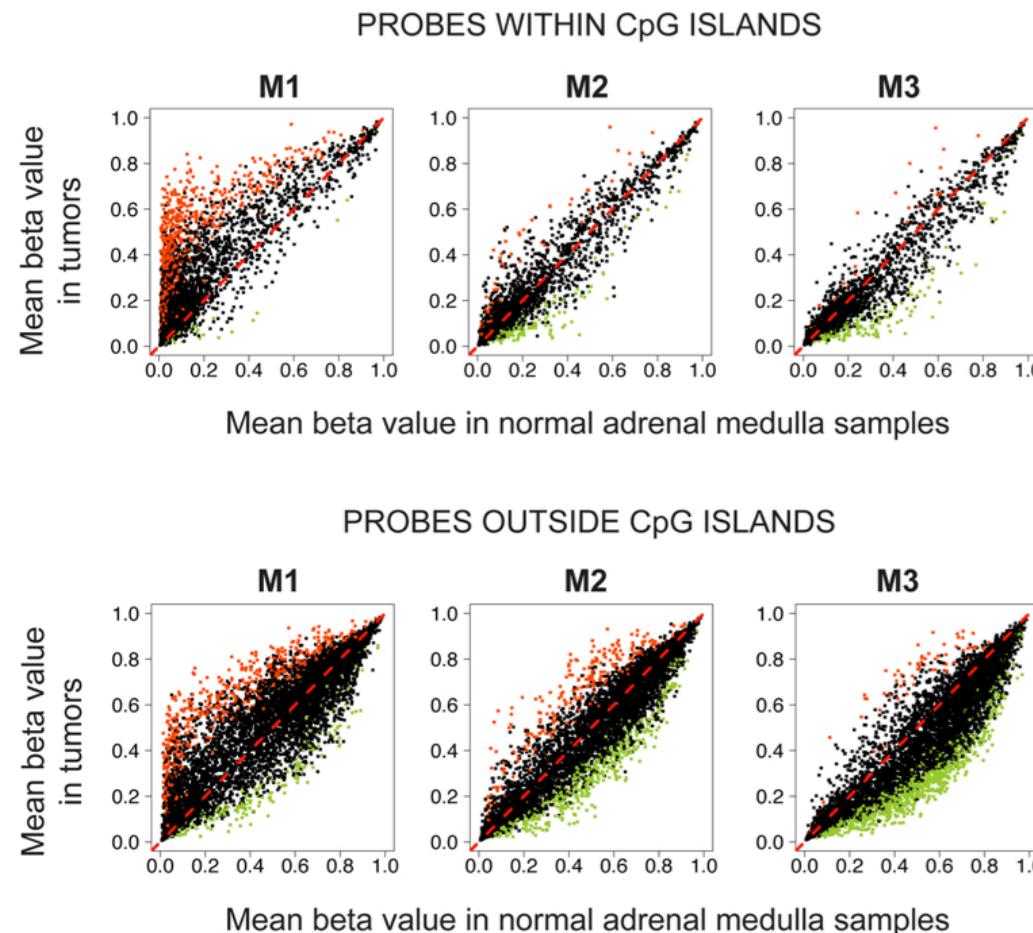


Letouzé*, Martinelli* et al.
Cancer Cell 2013

-> Clusters M1 and M2 respectively correspond to SDH and VHL-related tumors. Other tumors are comprised in cluster M3.

Les tumeurs mutées *SDHx* présentent un phénotype hyperméthylateur

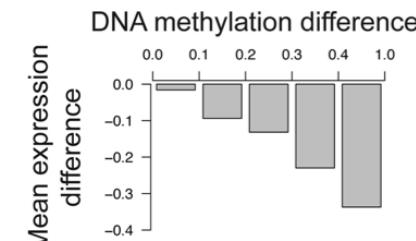
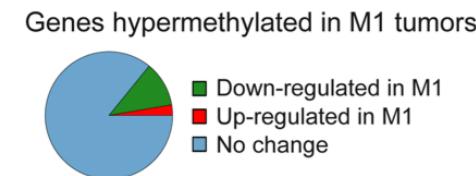
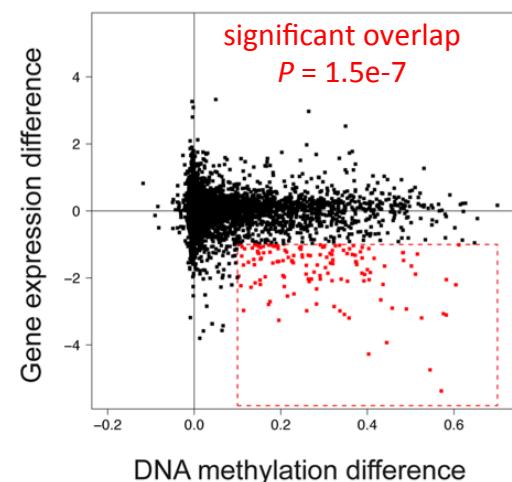
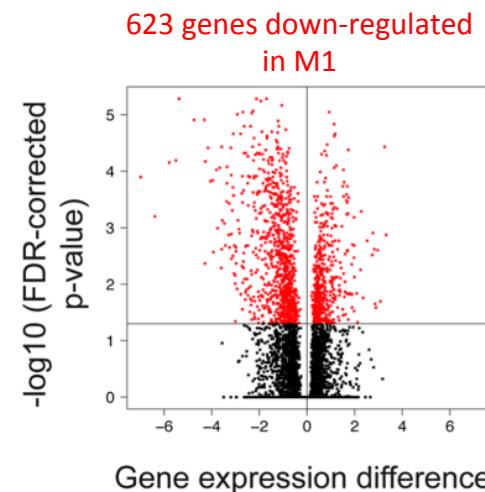
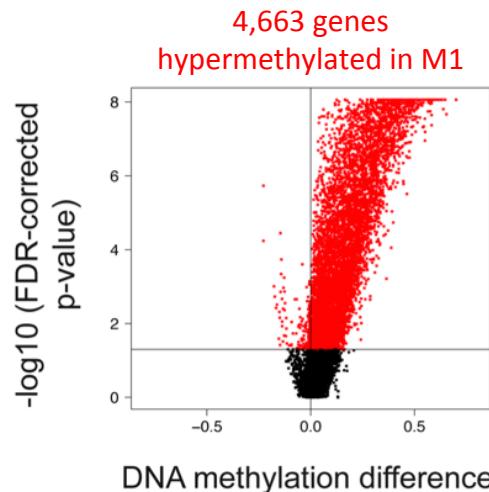
- DNA methylation changes in each tumor subgroup:



-> M1 tumors display widespread hypermethylation both within and outside CpG islands.

Les tumeurs *SDHx* présentent un phénotype hyperméthylateur associé à une inhibition transcriptionnelle

- DNA methylation and gene expression changes between M1 (*SDHx*) and non-M1 tumors:



-> 11.5% of hypermethylated genes (beta-value difference > 0.1) are down-regulated in M1 tumors.

Gènes inhibés par hyperméthylation dans les tumeurs *SDHx*

- 191 genes are significantly hypermethylated and down-regulated in *SDHx* tumors:

Gene symbol	DNA METHYLATION			GENE EXPRESSION		
	Illumina CpG ID	Beta value difference (M1-non M1)	Methylation FDR-adjusted p-value	Affymetrix probeset ID	Log2 fold change (M1/non-M1)	Expression FDR-adjusted p-value
→ NPY	cg00355281	0.3	0.000298922	206001_at	-6.39	0.000631472
CRYBA2	cg02805994	0.35	0.000298922	220136_s_at	-5.37	5.22E-06
→ KRT19	cg02893823	0.45	0.000594842	201650_at	-4.74	1.23E-05
EFS	cg18454133	0.4	0.004512674	204400_at	-4.27	6.73E-05
SYT13	cg21074260	0.25	0.001551809	226086_at	-3.96	0.00014417
→ PNMT	cg06596543	0.25	0.000298922	206793_at	-3.93	0.002706443
→ DNAJA4	cg05392364	0.37	0.000408954	225061_at	-3.8	9.46E-05
GNG8	cg21072025	0.28	0.007184473	234284_at	-3.61	0.005169372
RET	cg06559368	0.39	0.00208515	205879_x_at	-3.58	3.69E-05
JAKMIP1	cg05382097	0.18	0.000298922	238600_at	-3.42	0.002065611
TMEM45B	cg27294678	0.04	0.157509306	230323_s_at	-3.37	0.000343782
GDF10	cg04110601	0.32	0.016436798	206159_at	-3.26	0.000373051
RPP25	cg16064478	0.42	0.000298922	219143_s_at	-3.19	3.85E-05
→ RBP1	cg13099330	0.51	0.000298922	203423_at	-3.11	0.003154997
→ SLC6A2	cg04874129	0.25	0.007184473	239394_at	-3.09	0.000268757
C2orf40	cg06499647	0.1	0.01098253	223623_at	-3.07	0.000268764
MAPK13	cg06100227	0.34	0.002718498	210059_s_at	-3.07	0.006474365
RAB27B	cg09052983	0.1	0.00208515	228708_at	-3.01	0.009740778
CALY	cg10780164	0.23	0.000408954	219896_at	-3.01	0.045421315
SHC3	cg27044506	0.08	0.016436798	229824_at	-2.98	0.002231716
→ SPOCK2	cg17958658	0.25	0.004512674	202524_s_at	-2.98	0.002604168

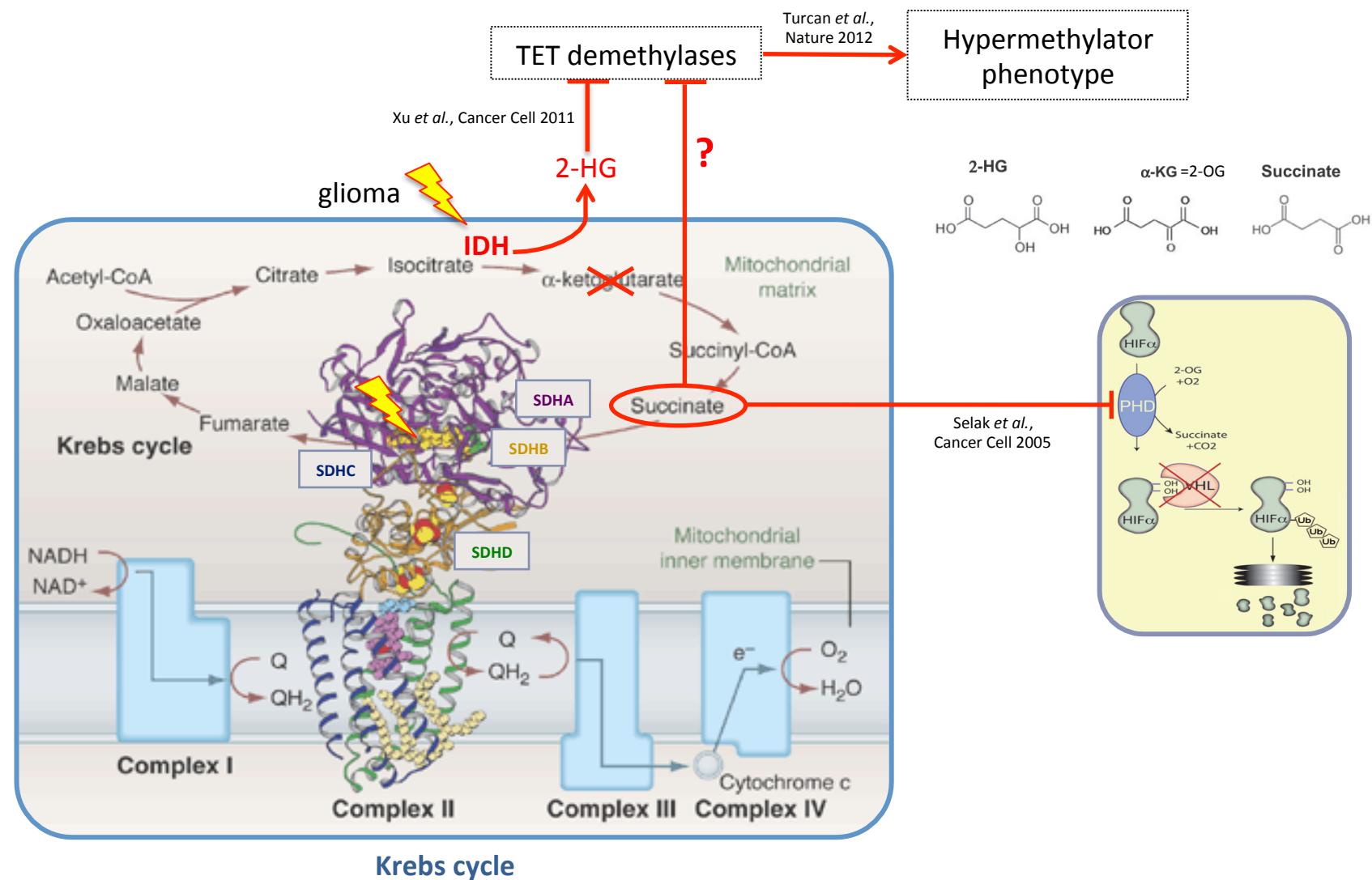
Catecholamine metabolic process

Epithelial-mesenchymal transition

Known tumor suppressors

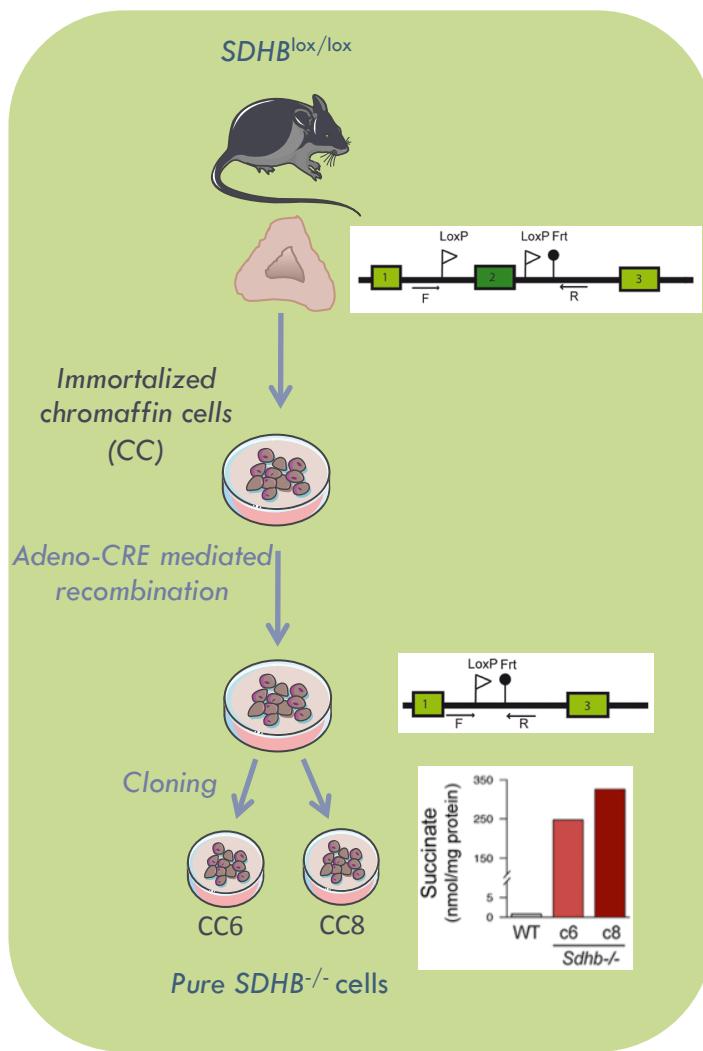
-> Epigenetic silencing of genes implicated in chromaffin cell differentiation and EMT explains the phenotypic characteristics of these tumors.

Quel mécanisme relie les mutations SDH et la méthylation de l'ADN ?

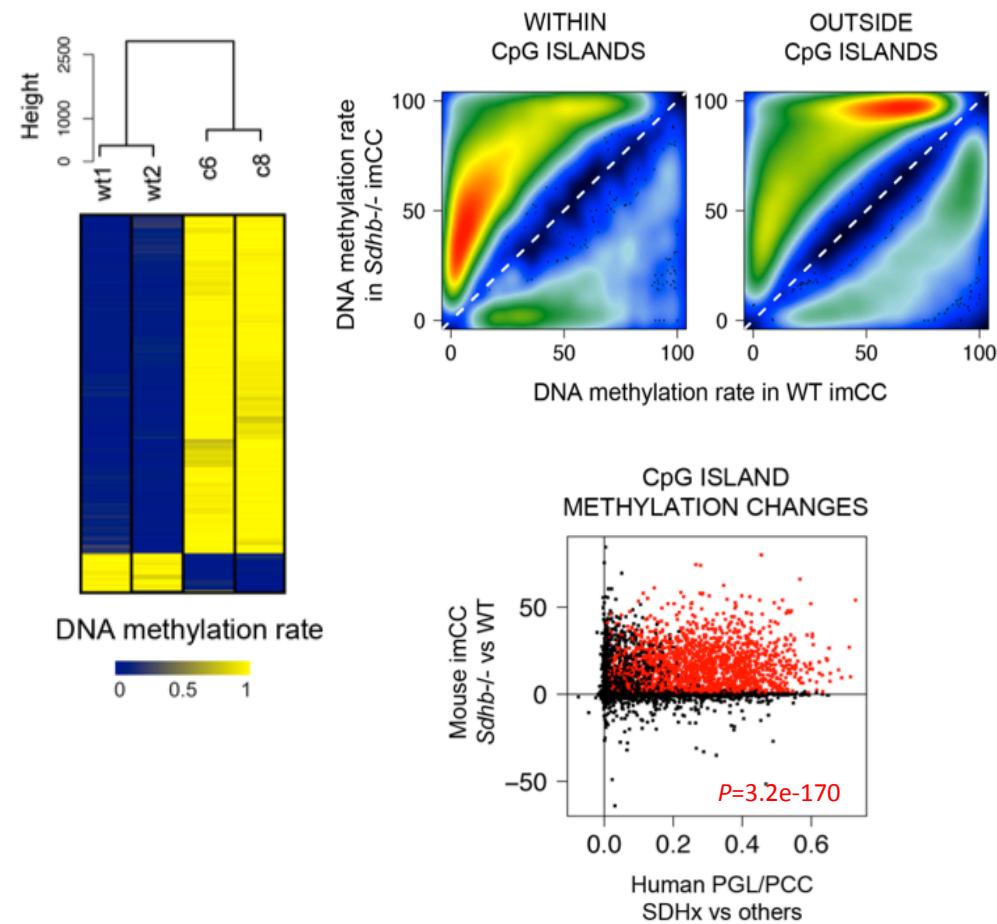


-> Does succinate also inhibit KDM and TET enzymes in SDH-related PGL/PCC?

Création et analyse de cellules murines *Sdhb*-/-



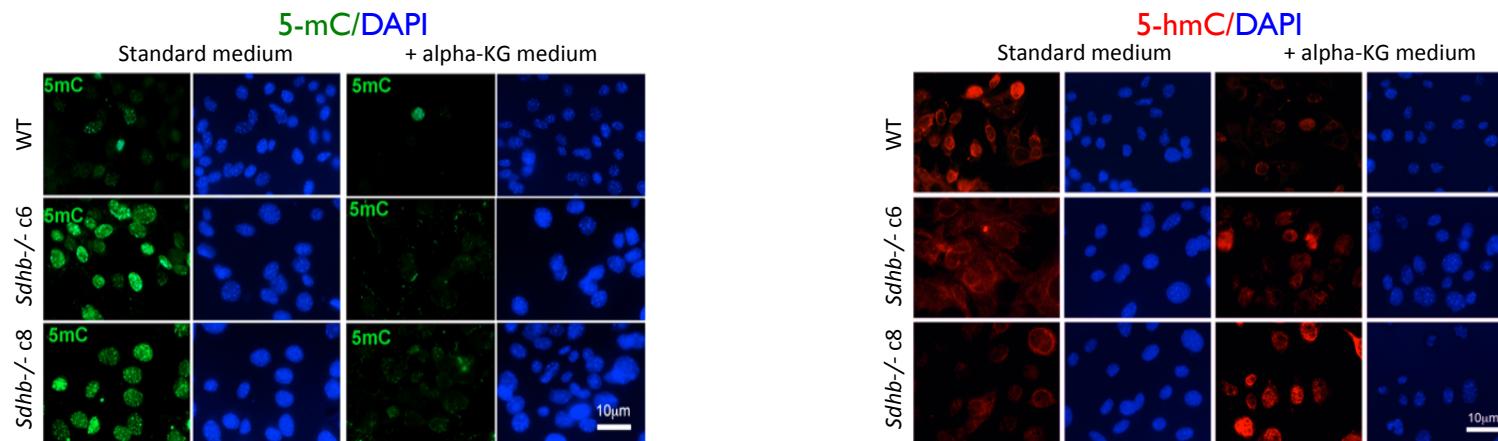
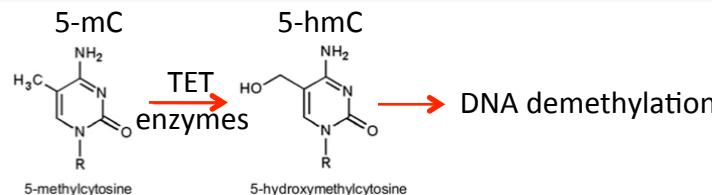
- Reduced representation bisulfite sequencing (RRBS) analysis of mouse chromaffin cells:



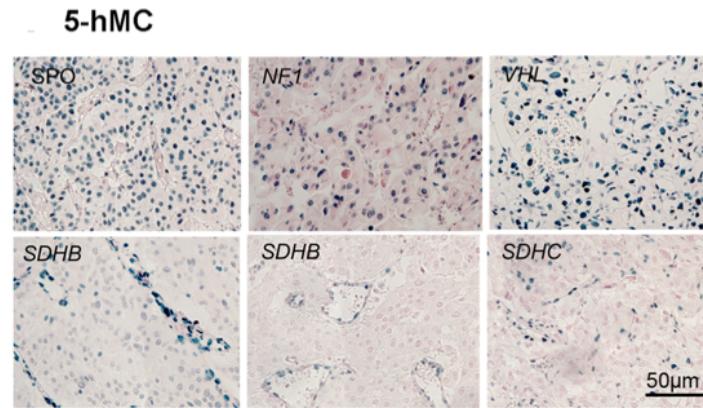
-> *Sdhb*-/- mouse chromaffin cells recapitulate the hypermethylator phenotype seen in human tumors, with 1,014 hypermethylated genes in common.

Ratio 5mC/5-hmC dans les cellules *Sdhb*-/-

- 5-mC/5-hmC ratio in *Sdhb*-/- cells:

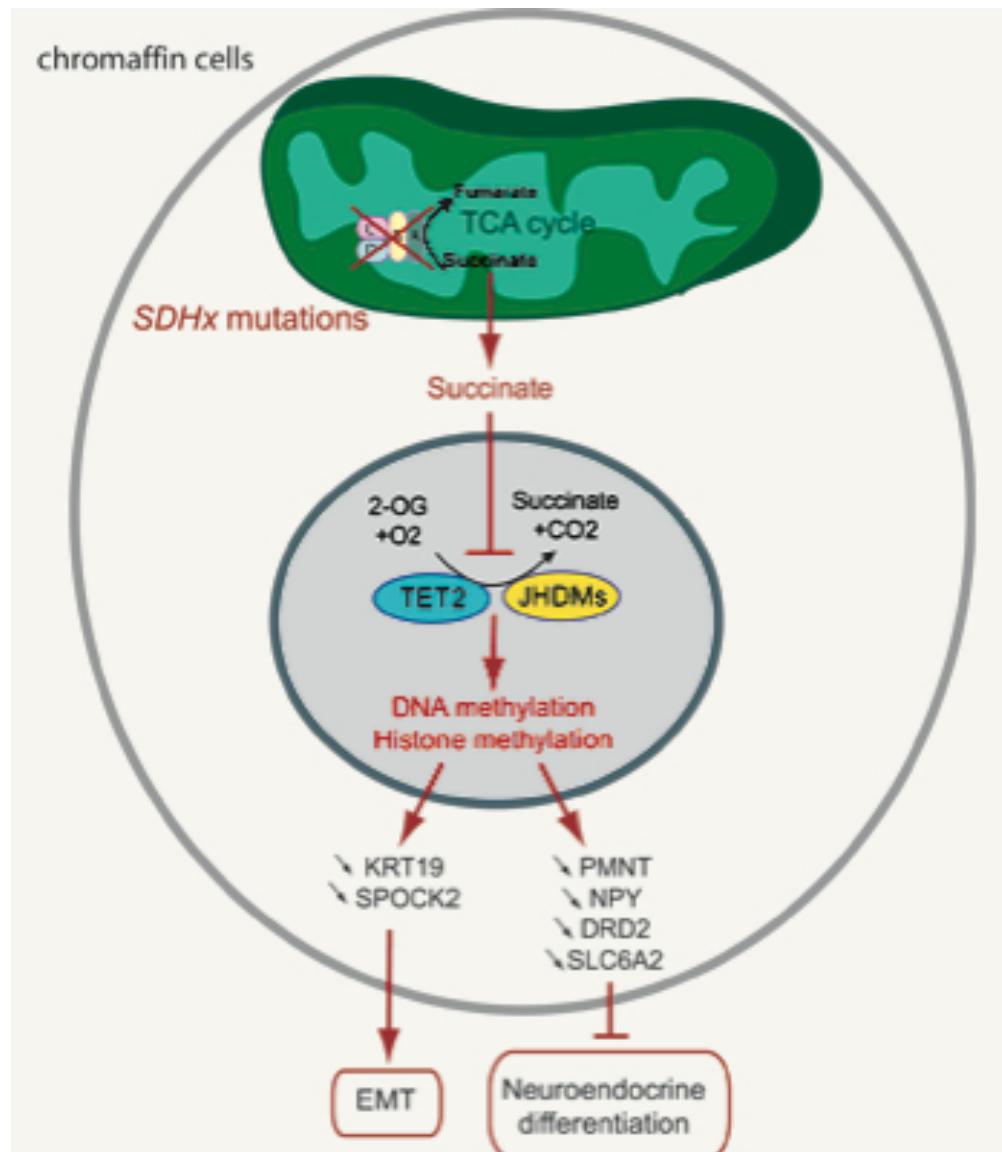


- 5-hmC staining of human PGL/PCC:



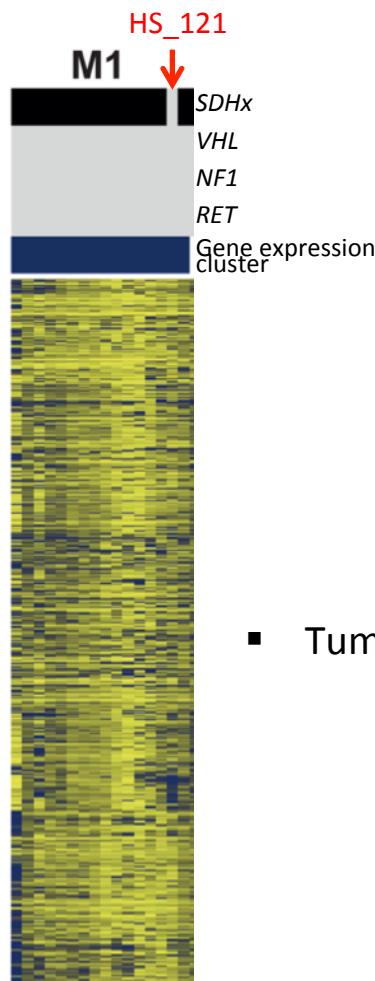
-> Inhibition of TET demethylases establishes the hypermethylator phenotype in SDH-deficient PGL/PCC.

Le mécanisme oncogénique des mutations *SDHx*

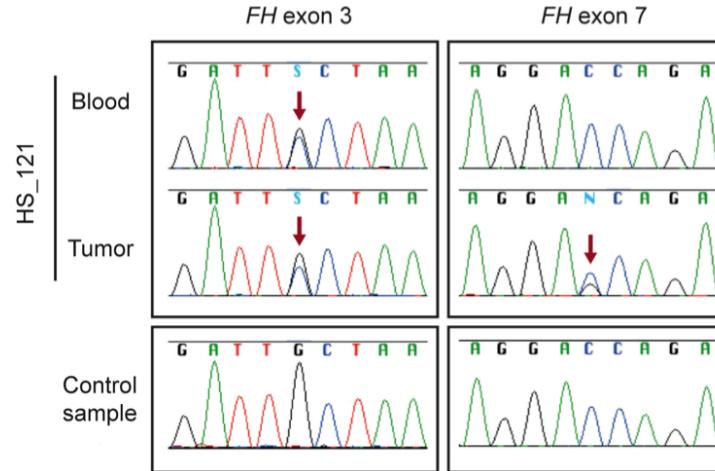


Letouzé*, Martinelli* et al. Cancer Cell 2013

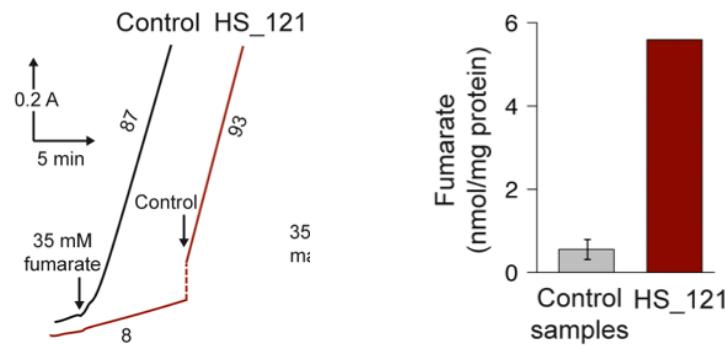
Le profil hyperméthylateur révèle un nouveau gène de prédisposition aux PGL/PCC



- Exome sequencing identifies two *FH* mutations in HS_121:



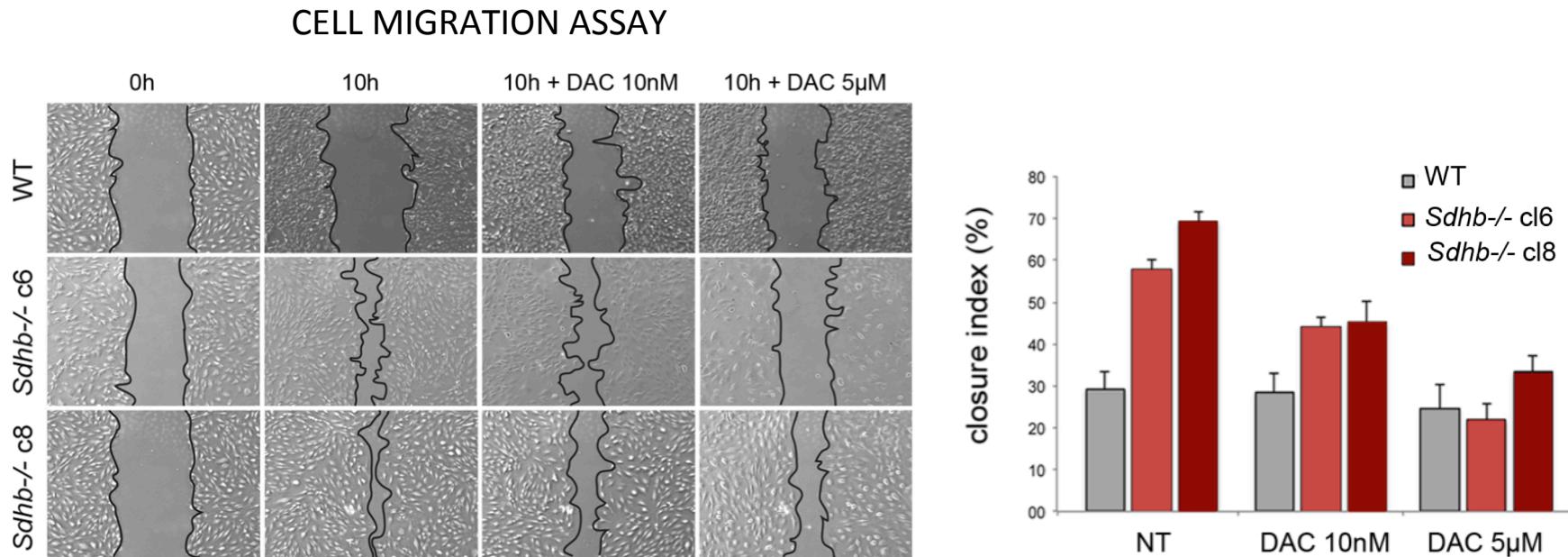
- Tumor HS_121 shows reduced fumarase activity and increased fumarate concentration:



-> *FH* mutations cause PGL/PCC by establishing a hypermethylator phenotype.

-> This new predisposing gene accounts for 0.8% of patients (Castro-Vega *et al.*, Hum Mol Genet 2014).

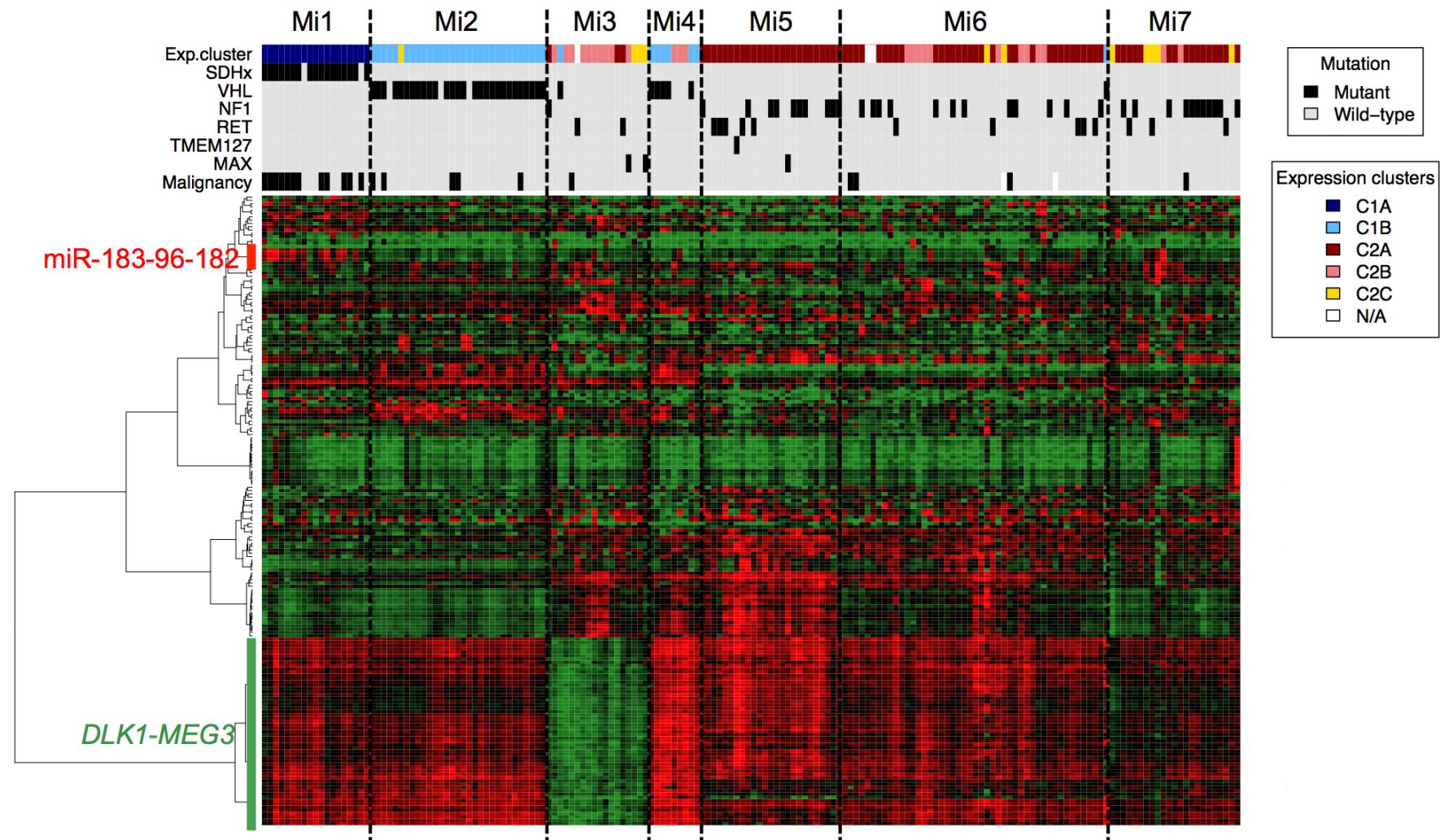
Une nouvelle option thérapeutique pour les tumeurs *SDHx*



- > Like SDH-related tumors, *Sdhb*^{-/-} cells do not proliferate fast but display a migratory phenotype.
- > Reversal of this phenotype by decitabin (DAC) suggests the use of demethylating agents to treat these aggressive tumors.

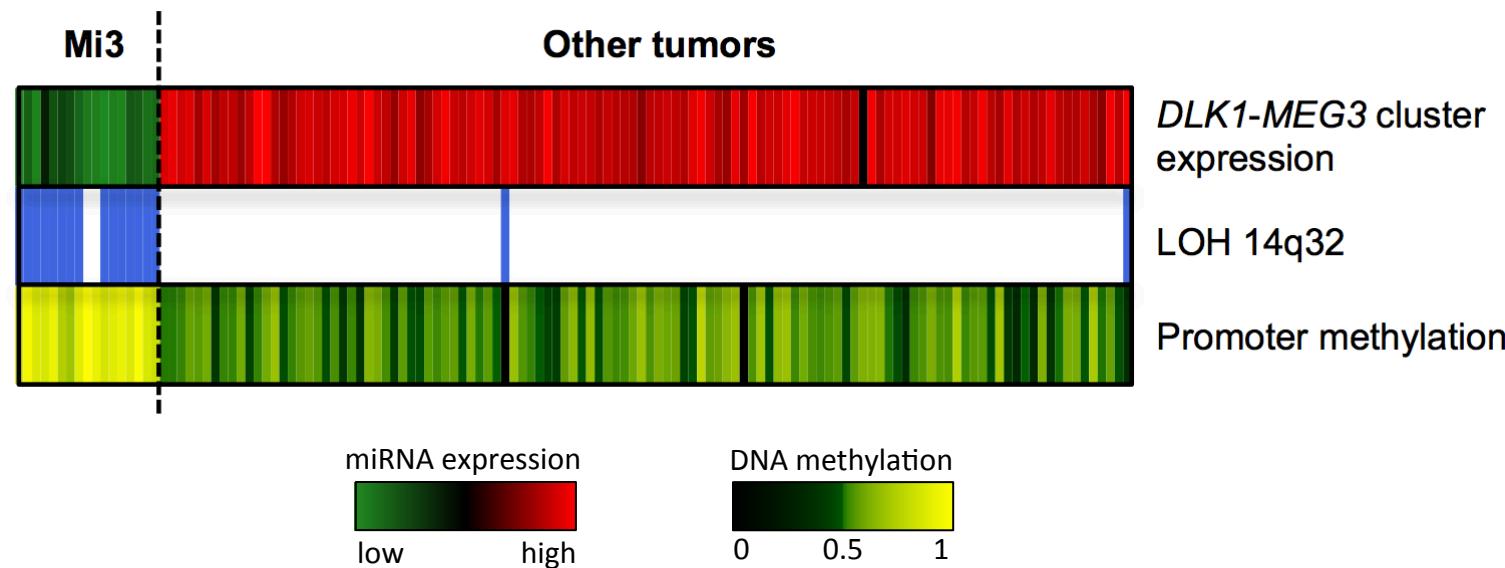
miRNA profiling of PGL/PCC

- miRNA profiling reveals 7 stable clusters with different miRNA expression patterns:



Mécanisme expliquant l'extinction du cluster *DLK1-MEG3* dans les tumeurs du groupe Mi3

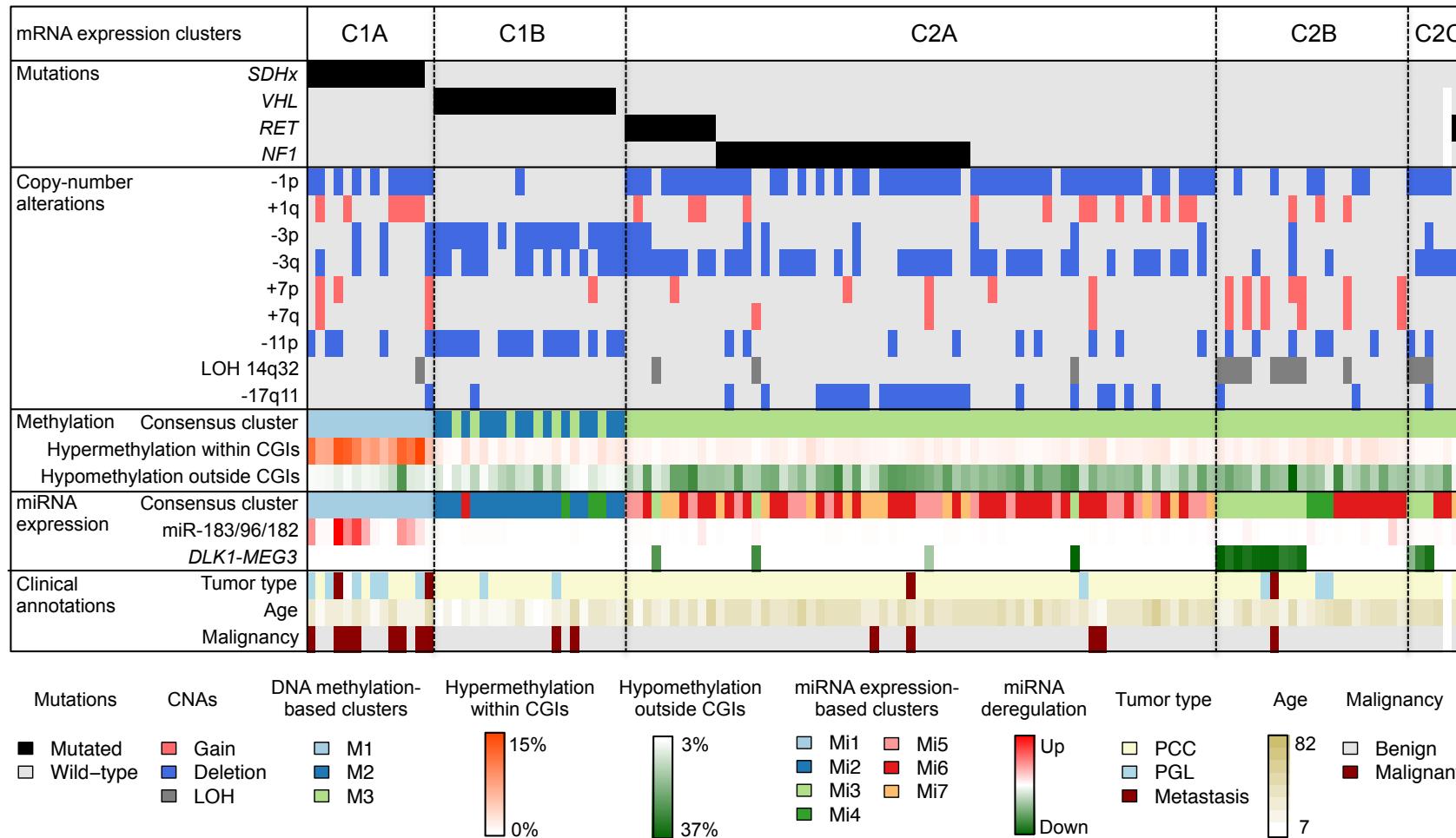
- The *DLK1-MEG3* cluster is **imprinted**: the maternal allele is expressed, whereas the paternal allele is methylated.
- Tumors of the Mi3 subgroup display an **LOH of chromosome arm 14q** associated with a **high methylation** of the *DLK1-MEG3* promoter:



-> In Mi1 tumors, loss of the maternal unmethylated allele leads to the inhibition of the *DLK1-MEG3* cluster.

Conclusion – Analyse génomique intégrée des paragangliomes

- Our analysis defines **5 molecular subgroups** of PGL/PCC, strongly associated with mutations in known driver genes, and displaying completely different molecular features:



Conclusion générale

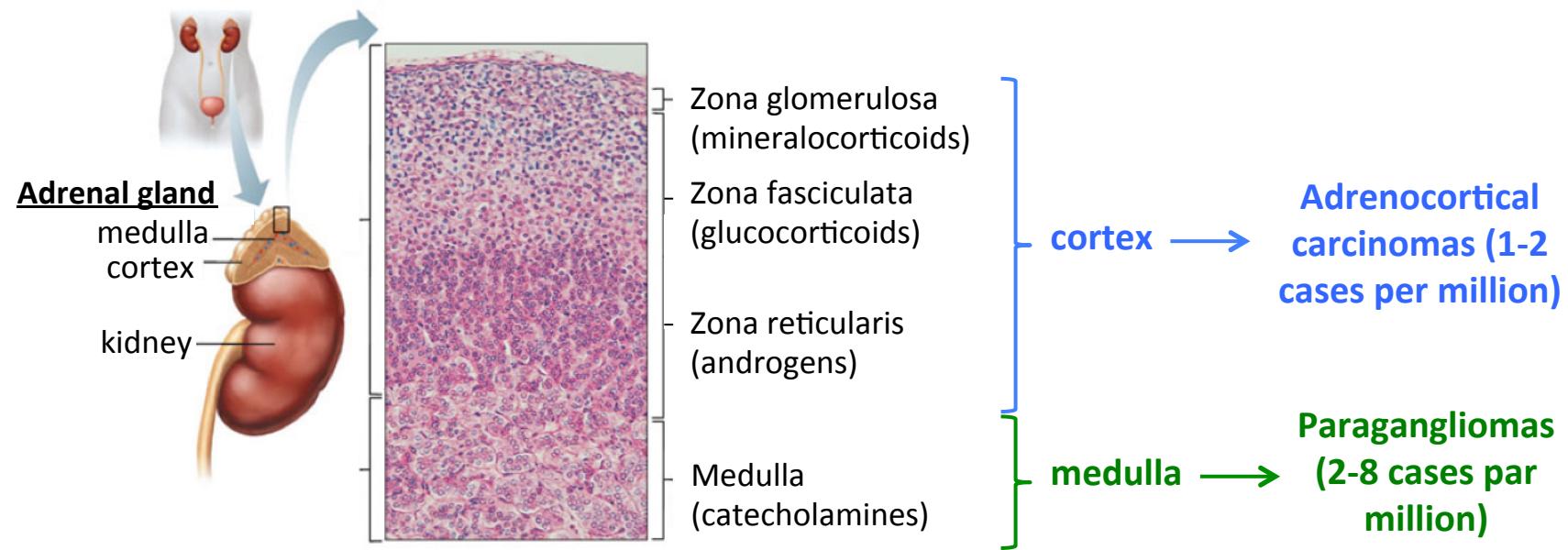
ANALYSE GÉNOMIQUE INTÉGRÉE

- Les données génomiques permettent d'étudier différents niveaux de dérégulations (épi)génétiques dans les cancers.
- Analyser de façon intégrée les différents omics est essentiel pour mieux comprendre la biologie des tumeurs.
- Les analyses multi-omics permettent de définir des sous-groupes homogènes de tumeurs et d'identifier les mécanismes oncogéniques impliqués.

APPLICATION AUX TUMEURS DE LA GLANDE SURRÉNALE

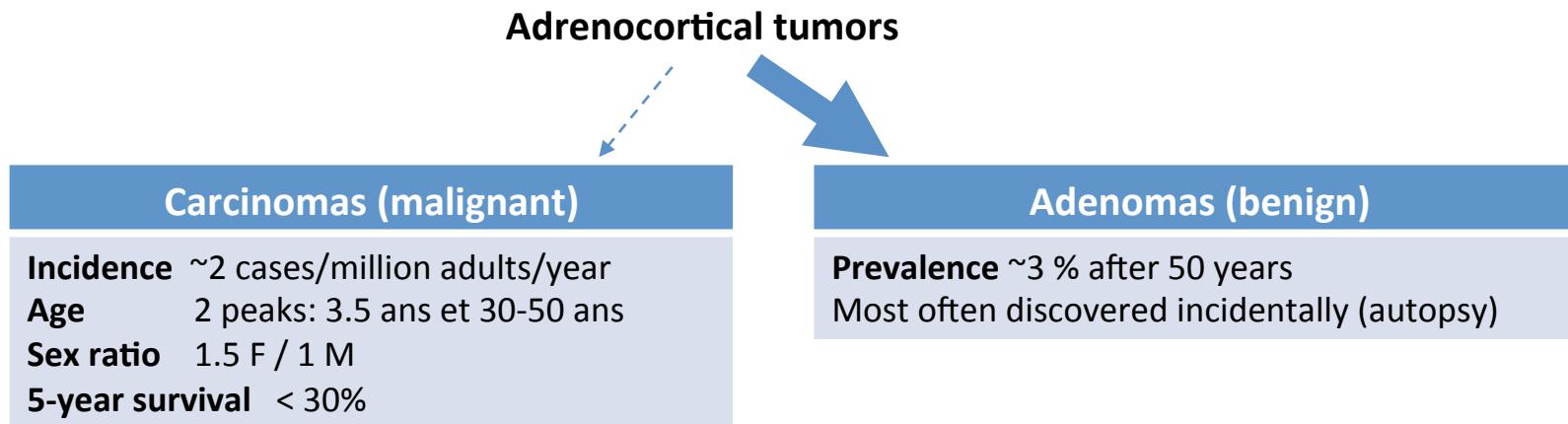
- Les données multi-omics révèlent des groupes de tumeurs très distincts d'un point de vue moléculaire et clinique.
- Dans les paragangliomes, l'inhibition des enzymes TET par le succinate relie méthylation de l'ADN et métabolisme.
- L'identification du phénotype hyperméthylateur ouvre de nouvelles perspectives thérapeutiques pour les paragnangiomes mutés *SDHx*.

Analyse génomique intégrée des cortico-surrénalomes



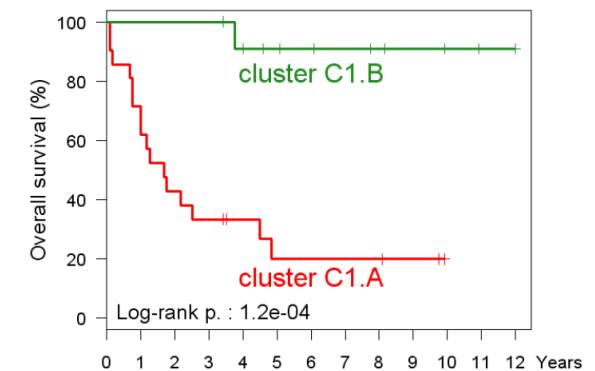
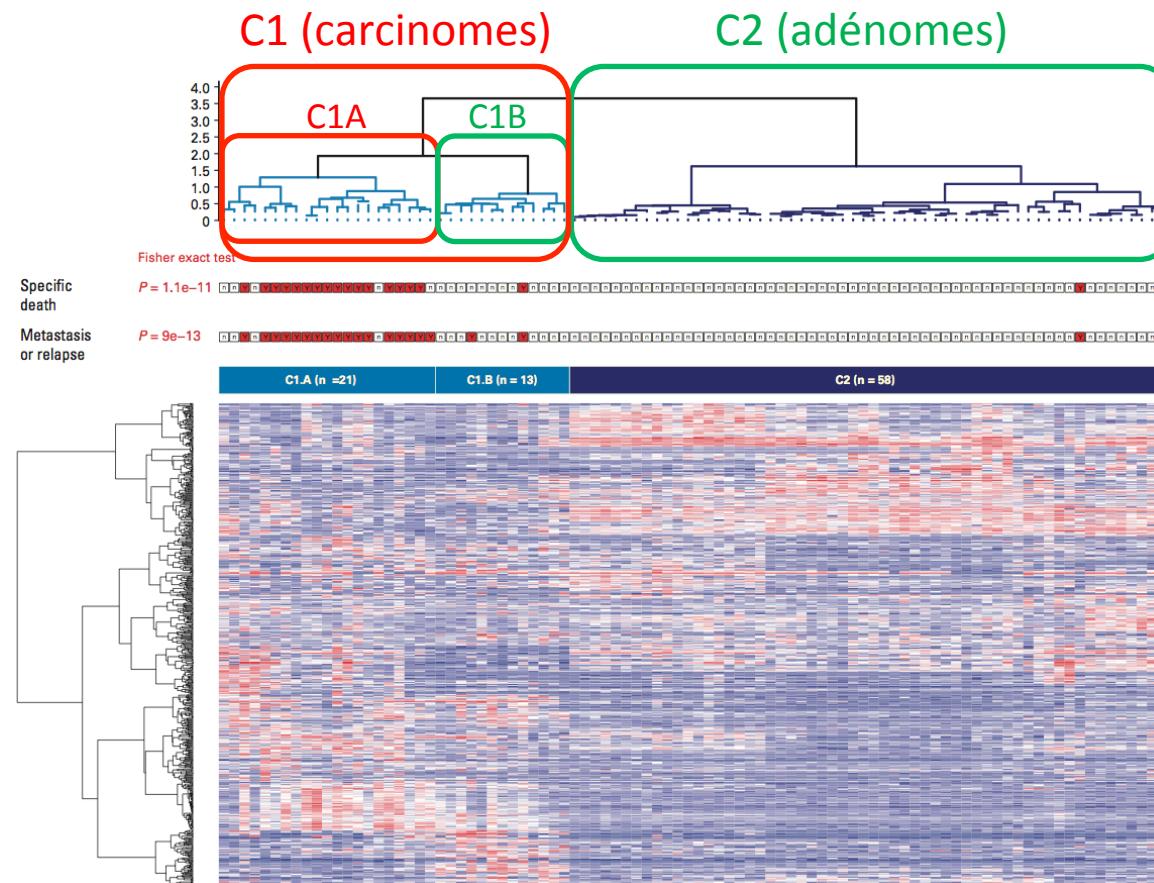
Introduction: Tumeurs du cortex surrénalien

Two types of adrenocortical tumors:



Classification des tumeurs basée sur le transcriptome

Transcriptome analysis of 92 adrenocortical tumors:



de Reyniès*, Assié* et al., JCO 2009

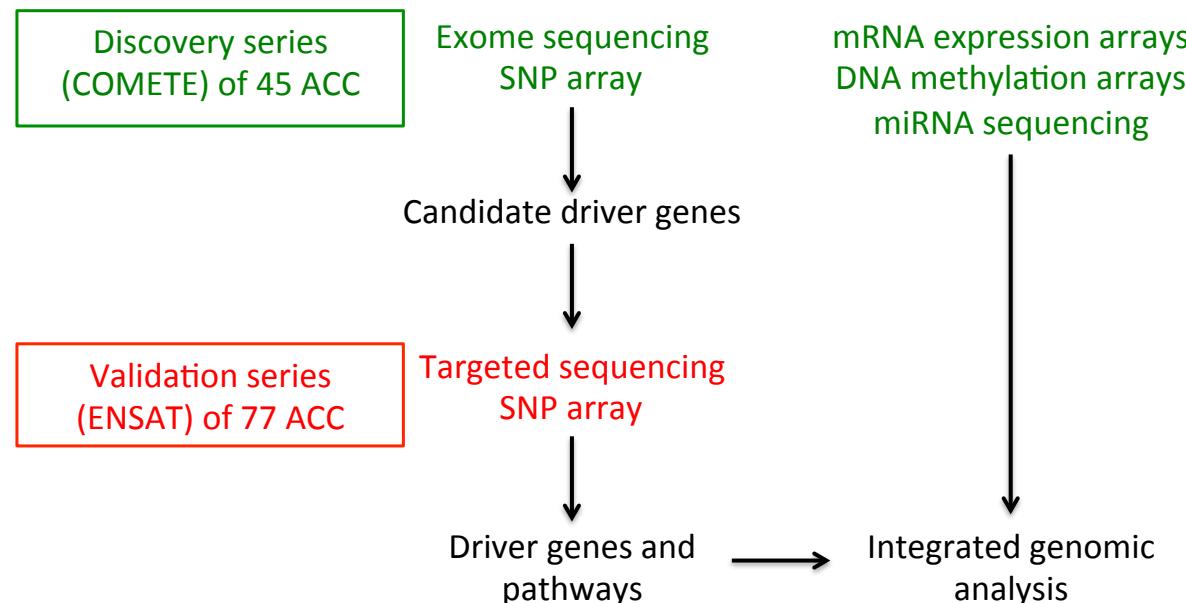
- > L'expression des gènes permet de séparer les adénomes des carcinomes, et révèle 2 groupes de carcinomes de pronostic totalement différent (**C1A** and **C1B**).
- > Un prédicteur basé sur seulement 3 gènes permet de distinguer les tumeurs C1A, C1B et C2.

Other known molecular features

- Recurrent activating mutations of ***CTNNB1*** and inactivating mutations of ***TP53***.
- A CpG island methylator phenotype (**CIMP**) has been described, associated with a poor prognosis..
- Several studies identified deregulated **miRNAs** associated with tumor type and prognosis.

Goals of the present study:

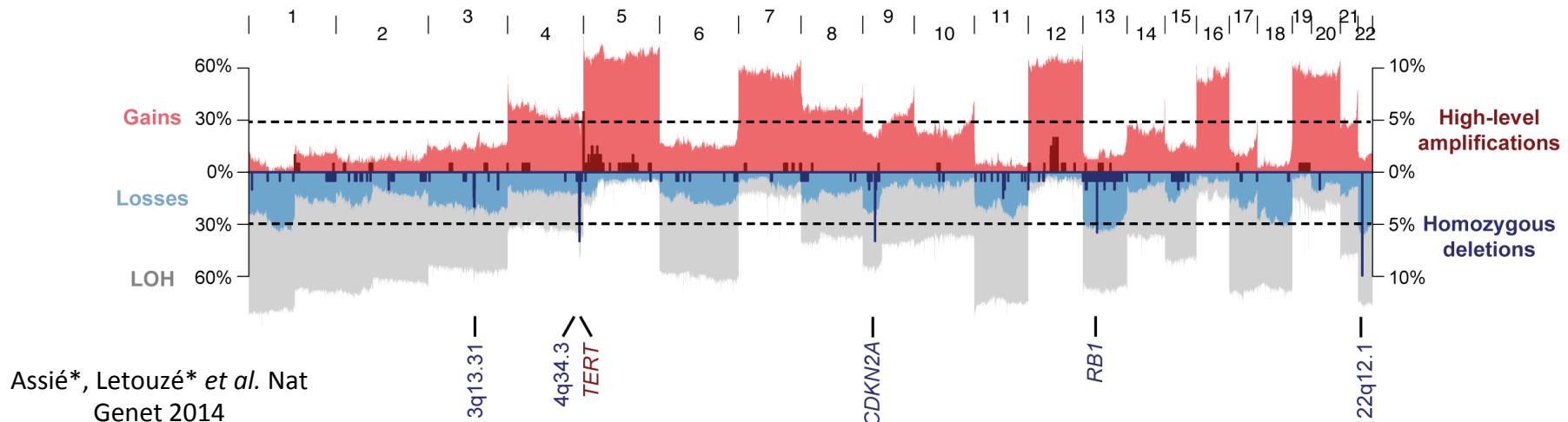
- > **PART 1:** Comprehensive identification of **driver genes and pathways** by a combined analysis of mutations (exome sequencing) and focal copy number changes (SNP array).
- > **PART 2: Integrated multi-omics analysis** of a single ACC cohort.



PARTIE 1 : Identification des gènes et pathways drivers

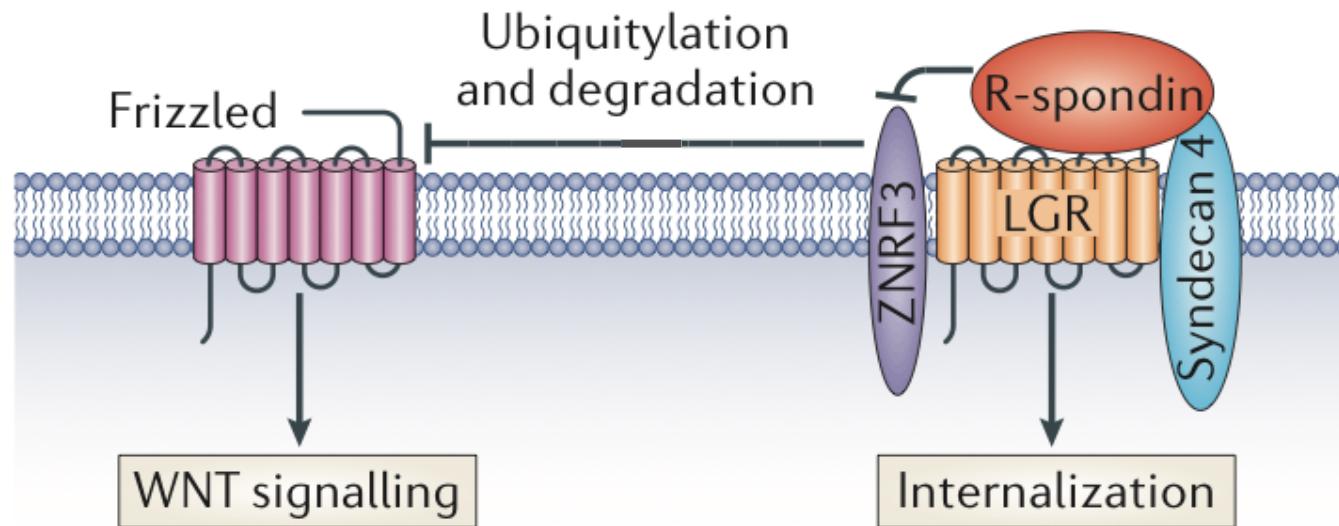
Analyse des aberrations chromosomiques

- ACC are extremely altered (mean=85% of the genome) and frequently polyploid (43%).



ZNRF3 est un régulateur négatif de la voie Wnt/β-caténine

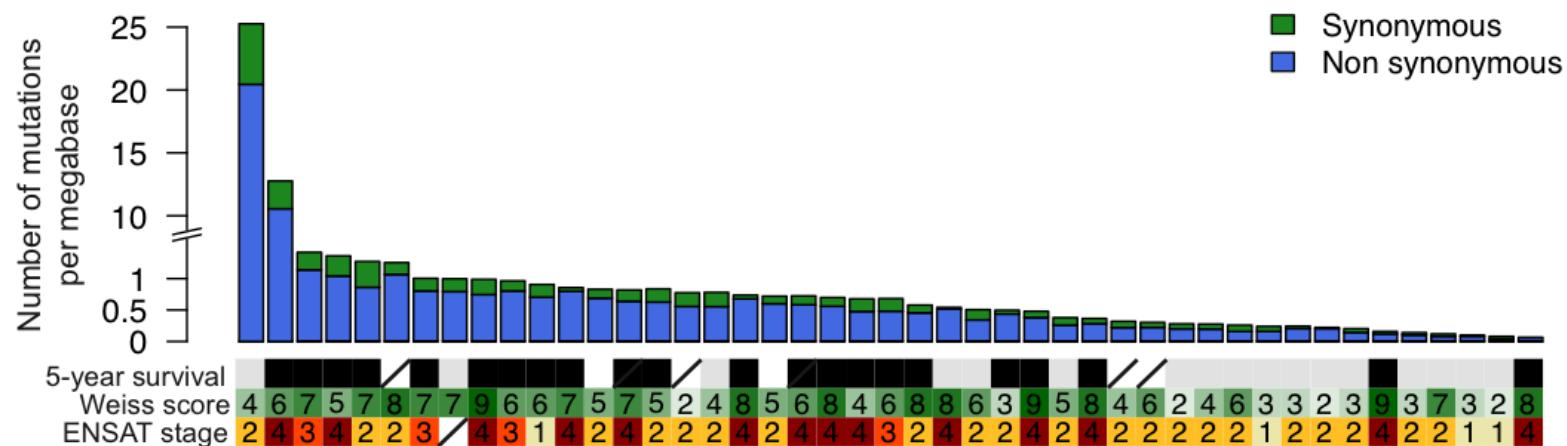
- ZNRF3 negatively regulates the Wnt pathway by promoting the turnover of Frizzled receptors (Hao *et al.*, Nature 2012):



-> ZNRF3 inactivation is another means to activate the Wnt/β-catenin pathway in ACC, in addition to *CTNNB1* mutations.

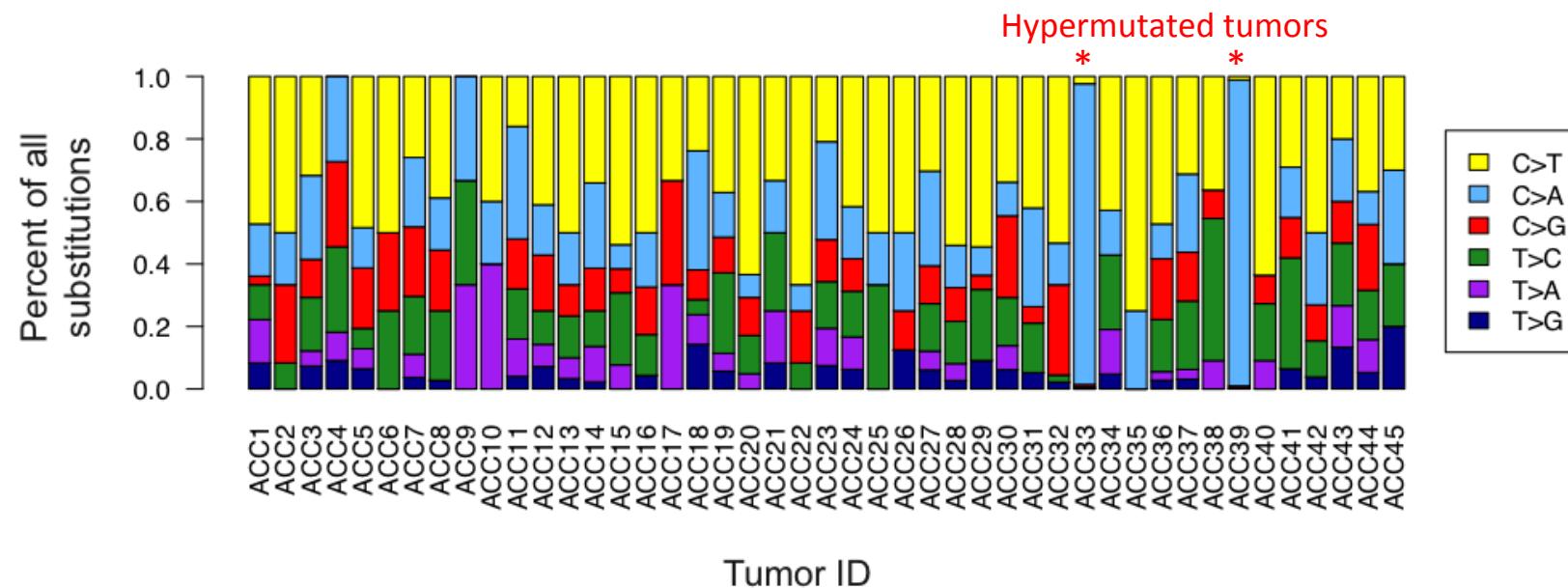
Analyse des mutations somatiques (séquençage exome)

- We sequenced 70 Mb of exonic sequences (20,975 genes) to a mean depth of 90x.
- We found a median of **30 mutations per tumor** in coding sequences. The mutation rate, relatively low (0.6 mutations/Mb), is significantly associated with tumor stage and prognosis:



Spectre mutationnel des ACC

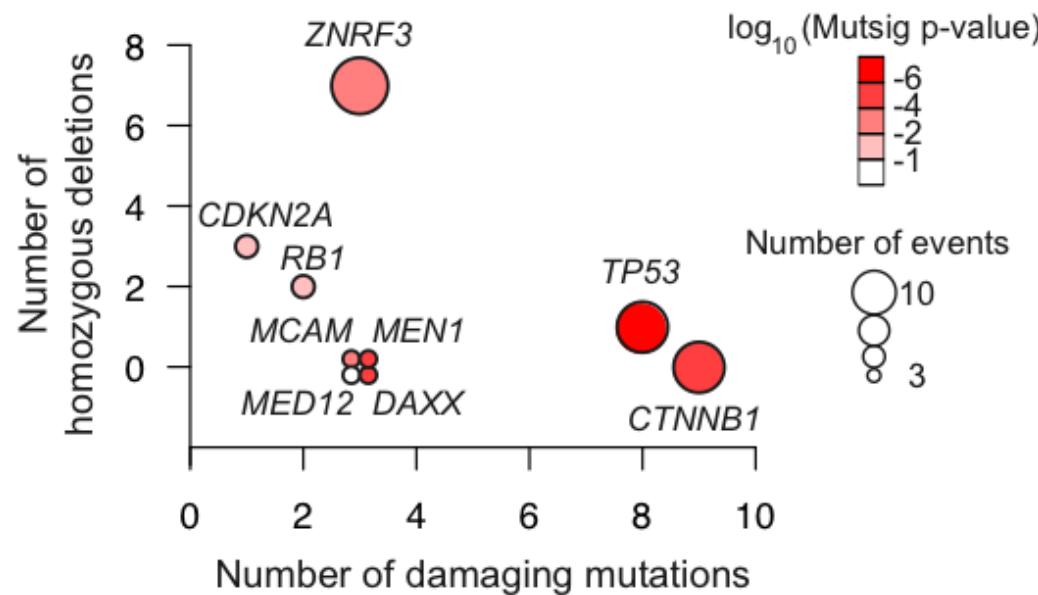
- Frequency of each type of substitution in 45 ACC:



-> As in most cancers, C>T transitions are the most frequent point mutations. Hypermutated tumors have a high rate of C>A transversions, presumably due to the exposure to a mutagen that remains to be identified.

Sélection des gènes drivers candidats

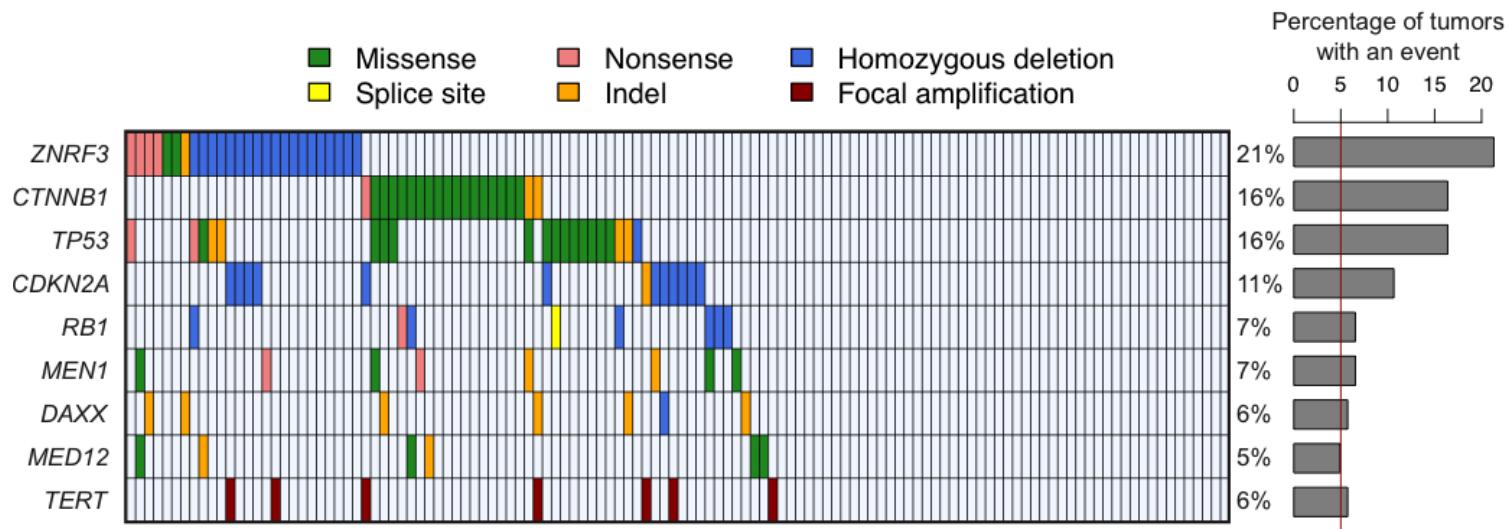
- To identify candidate driver genes, we integrated **damaging mutations** (nonsense, indel, or missense mutations with functional consequences predicted by Polyphen2 software) and **homozygous deletions**.
- We selected genes altered in ≥ 3 tumors of the discovery set (n=45):



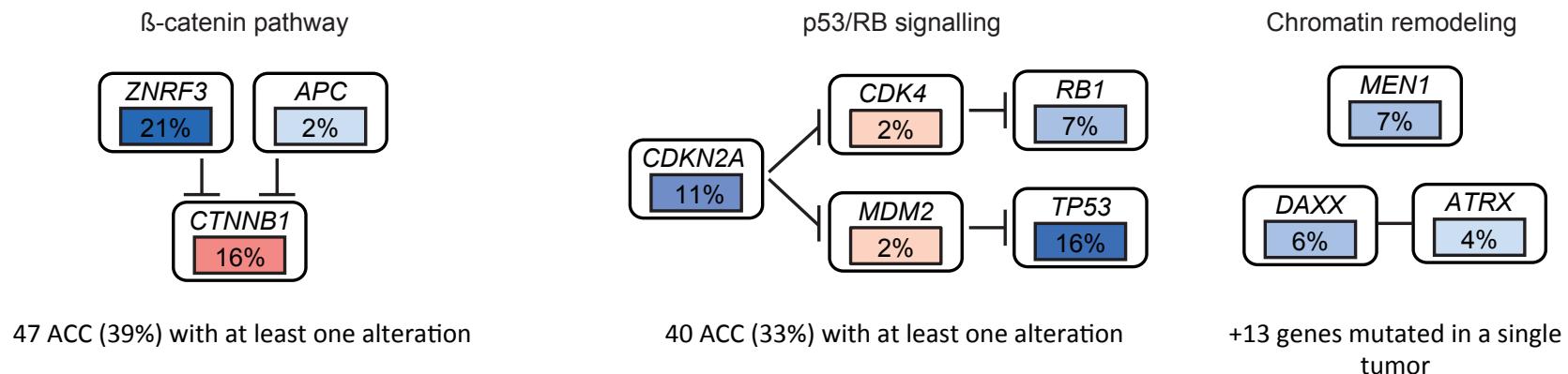
-> These genes were analyzed by targeted sequencing in the validation series of 77 ACC.

Gènes et voies cellulaires mutées dans la série complète de 122 ACCs

- In the entire cohort, 9 genes are altered by mutation, homozygous deletion or high-level amplification in $\geq 5\%$ of ACC :



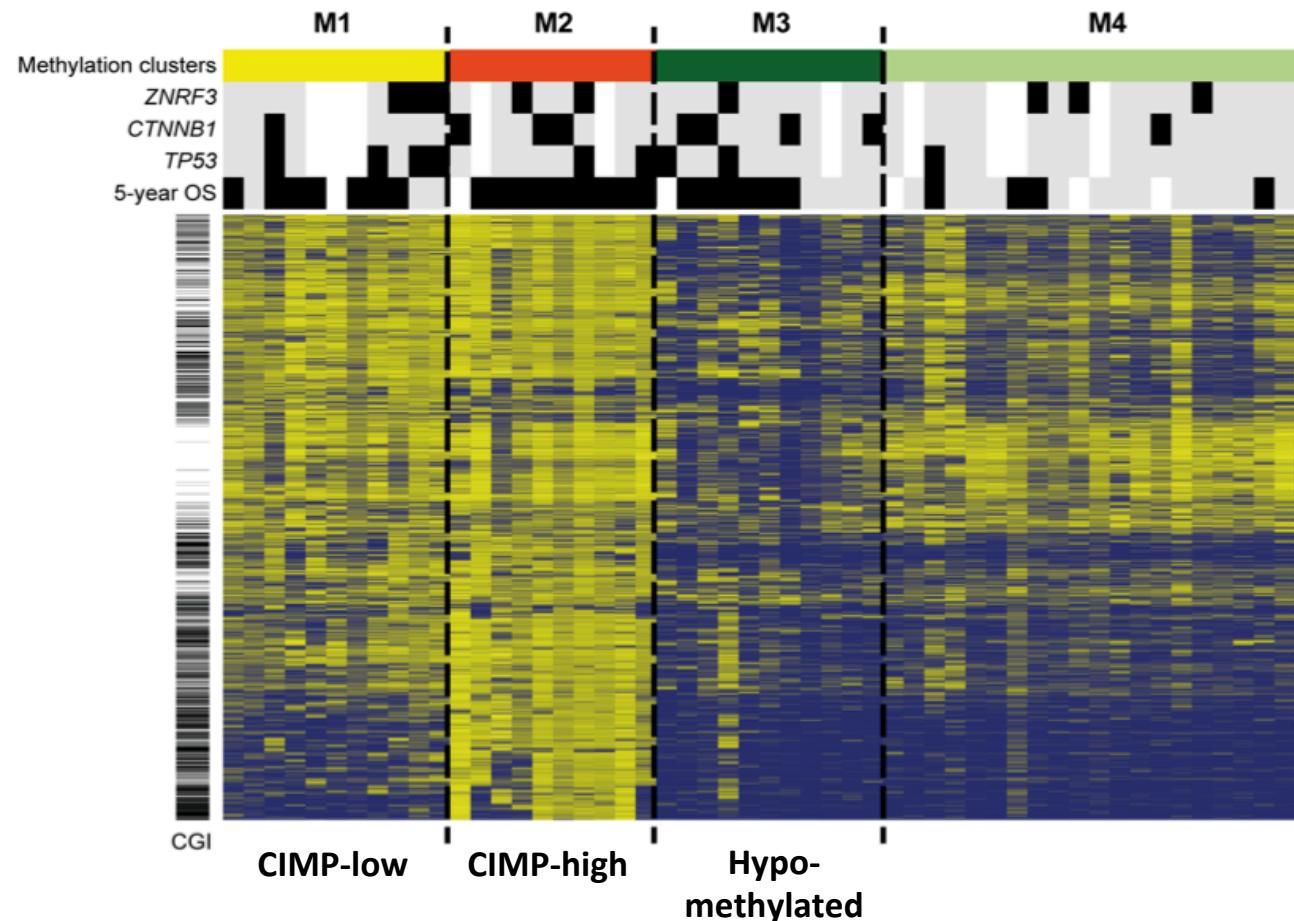
- These driver genes belong to three main pathways:



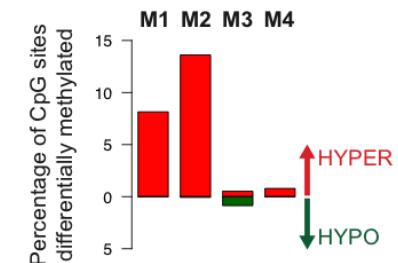
PARTIE 2 : Analyse multi-omics intégrée des ACC

Analyse de la méthylation de l'ADN

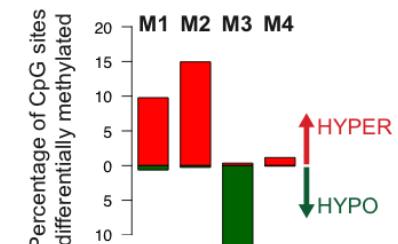
- Tumors of the discovery set were analyzed with Illumina Meth27 arrays (27,000 probes).
- We identified 4 homogeneous DNA methylation-based ACC subgroups using the RPMM method (Houseman *et al.*, BMC Bioinformatics 2008).



WITHIN CpG ISLANDS



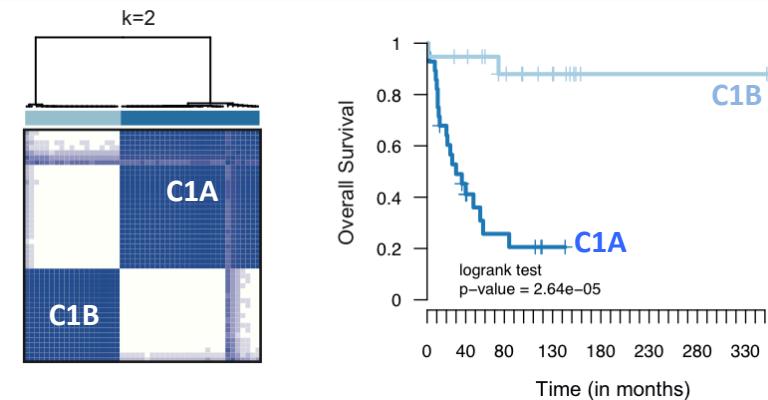
OUTSIDE CpG ISLANDS



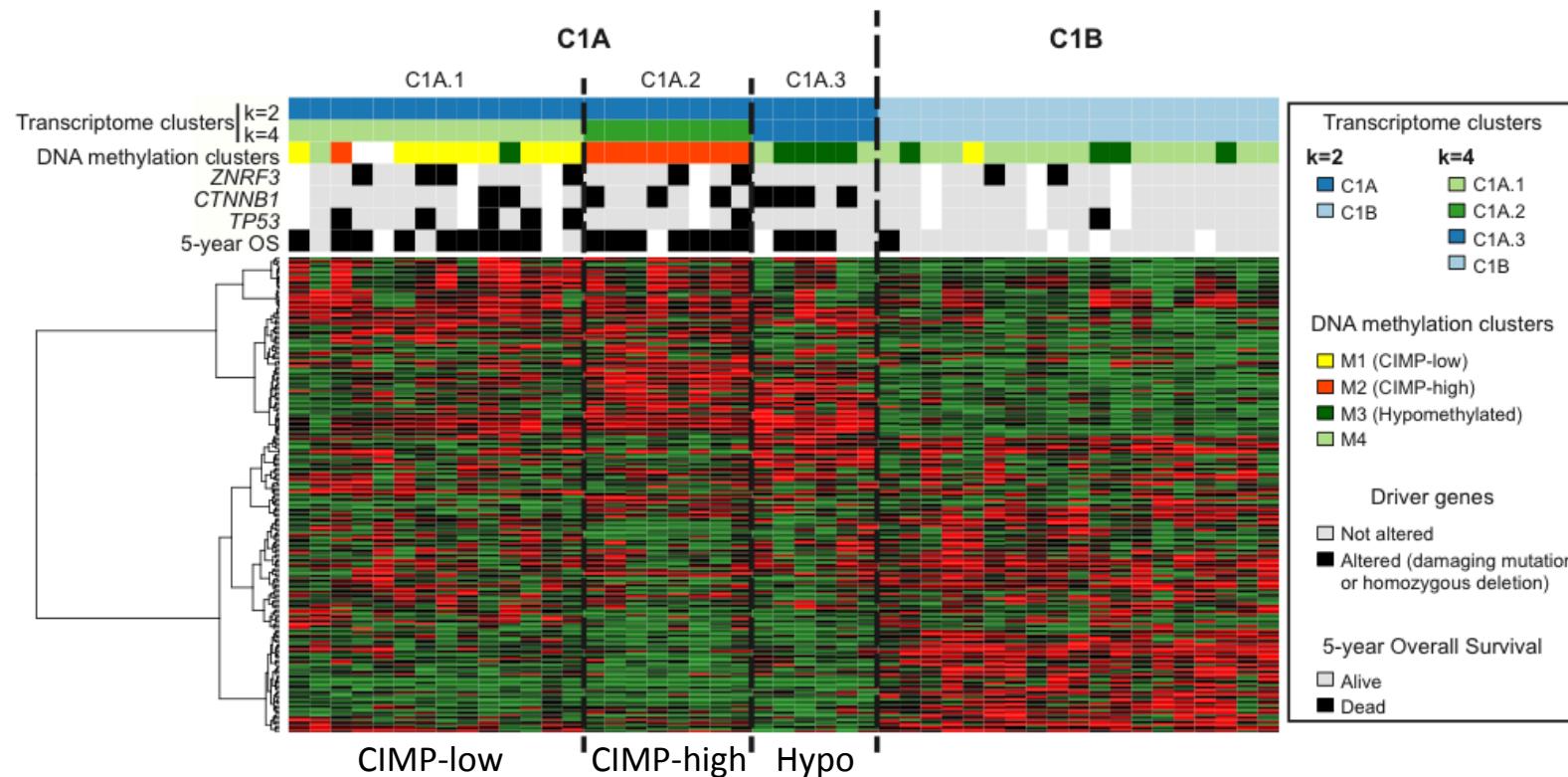
-> We identified the two previously described CIMP clusters, and a new ACC subgroup strongly hypomethylated outside CpG islands.

Analyse de l'expression des gènes (ARNm)

- A consensus clustering confirms the existence of 2 main clusters (C1A and C1B):

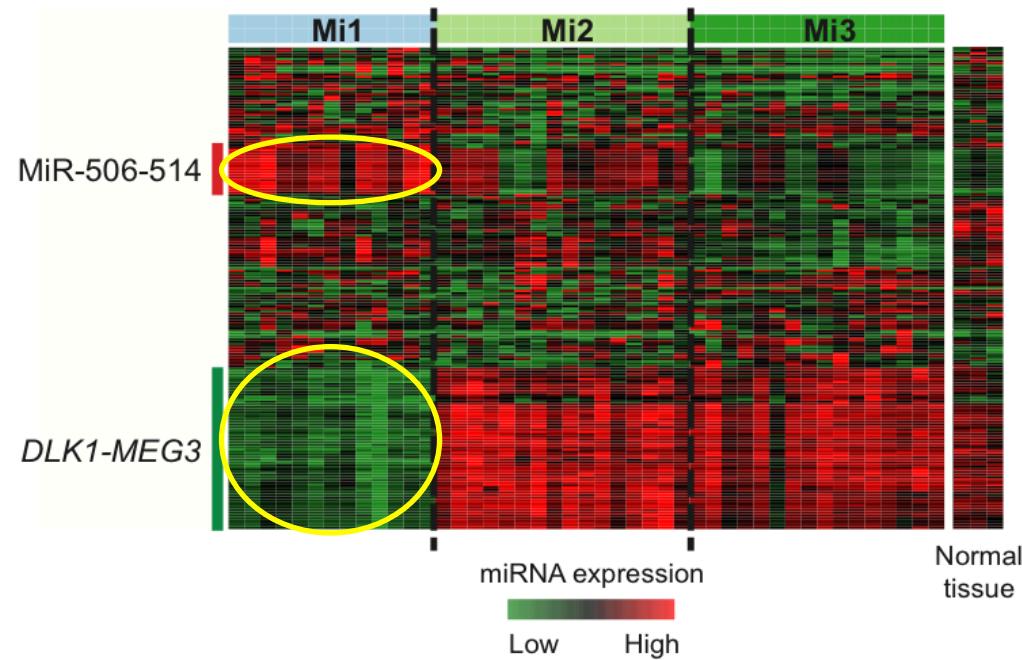
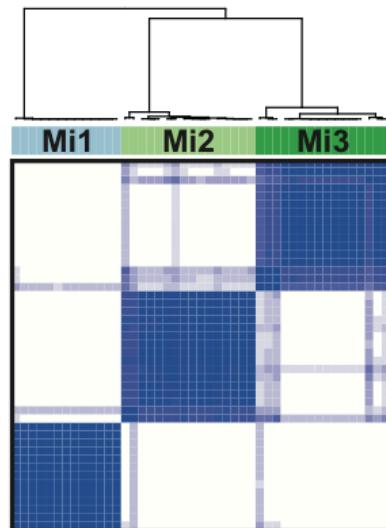


- The C1A group can be divided in 3 subgroups, strongly associated with DNA methylation clusters:



Analyse de l'expression des microARN

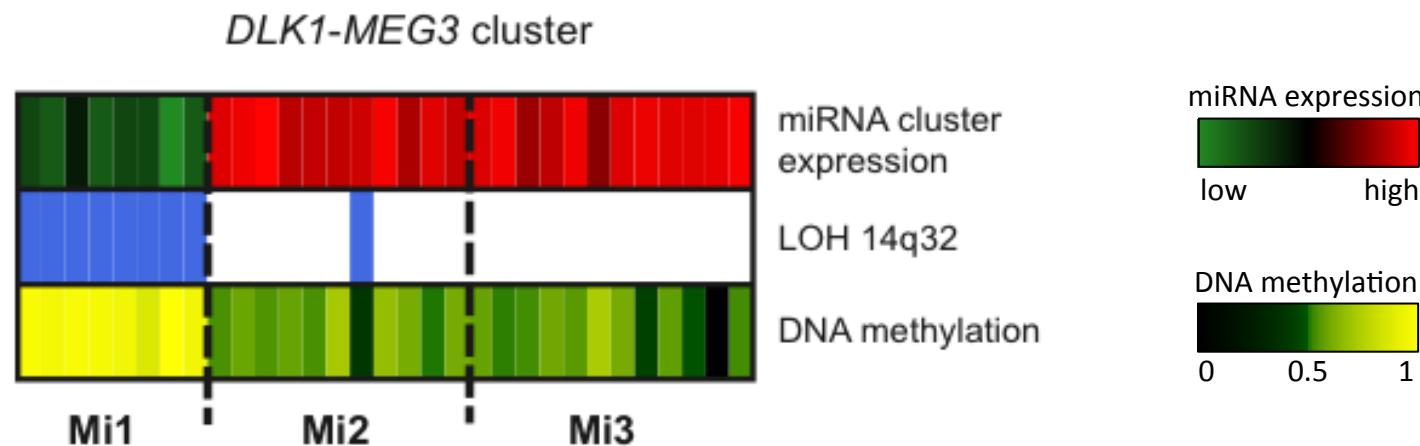
- A **consensus clustering** identifies 3 homogeneous ACC subgroups based on miRNA expression:



-> Mi1 tumors display the strongest miRNA expression changes with the deregulation of 2 miRNA clusters.

Mécanisme expliquant l'extinction du cluster *DLK1-MEG3*

- The *DLK1-MEG3* cluster is **imprinted**: the maternal allele is expressed, whereas the paternal allele is methylated.
- All tumors of the Mi1 tumors have an **LOH of chromosome arm 14q** vs only 1/23 non-Mi1 tumors ($P=1.1e-6$), and display a **hypermethylation** of the *DLK1-MEG3* promoter:



-> In Mi1 tumors, loss of the maternal unmethylated allele leads to the inhibition of the *DLK1-MEG3* cluster.

Conclusions

- Integrated exome/SNP analysis reveals **9 major driver genes** (*ZNRF3*, *CTNNB1*, *TP53*, *CDKN2A*, *RB1*, *MEN1*, *DAXX*, *MED12* and *TERT*), involved in **3 major pathways** (β -catenin, p53/Rb, and chromatin remodeling).
- Our **DNA methylation** analysis confirms the existence of CIMP-high and CIMP-low phenotypes, and reveals a new hypomethylated subgroup.
- **mRNA expression** analysis confirms the existence of 2 main clusters (C1A and C1B). The C1A cluster can be subdivided in 3 subgroups strongly associated with CIMP-high, CIMP-low and hypomethylated tumors.
- **miRNA expression** reveals a strong deregulation of miR-506-514 (oncogenic) and *DLK1-MEG3* (tumor suppressive) in a subgroup of ACC.
- ***DLK1-MEG3*** silencing results from the loss of the maternal allele by LOH of chromosome arm 14q.

Intégration multi-omics

- Integrative genomic analysis reveals the existence of **two distinct molecular and clinical entities** of ACC, driven by specific oncogenic alterations:
 - **C1A** tumors of bad prognosis, heavily mutated, often CIMP
 - **C1B** tumors of good prognosis, with a deregulation of miR-506-514 and *DLK1-MEG3* clusters.

