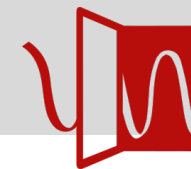


# Intégration de données multi-omiques

# Nous nous présentons

Comment en sommes-nous arrivés à intégrer des données ?

- Vincent Guillemot (Institut Pasteur) – « pour programmer en R de nouvelles supers méthodes »
- Cathy Philippe (Neurospin) – « on m'a forcée »
- Lucie Khamvongsa-Charbonnier (IFB) – « je voulais voir ce qui se passait en cis et après »
- Rachel Legendre (Institut Pasteur) – « Je connais bien Kohonen le barbare »



- **BAOBAB**
  - MRI physics, image analysis, population imaging, imaging-genetics.
- **UNIACT**
  - Clinical neuroscience
- **UNICOG**
  - Cognitive neuroscience

# Biostatisticien au Hub

- Analyser des données biologiques de toute sorte
- Transmettre ses connaissances (statistiques, R, etc.)
- Rechercher et développer de nouvelles méthodes



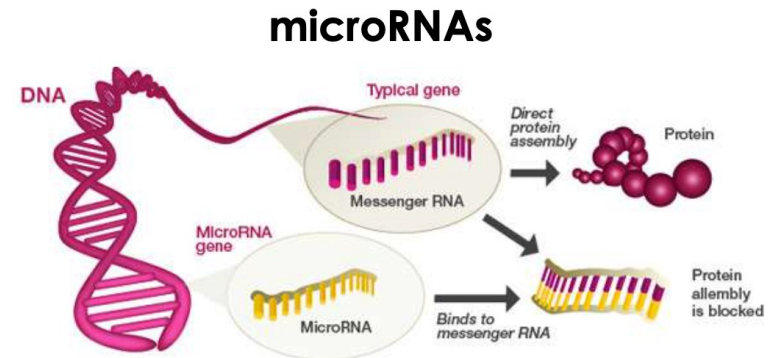
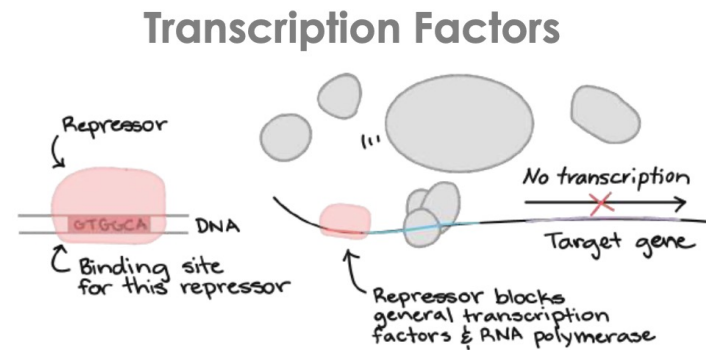
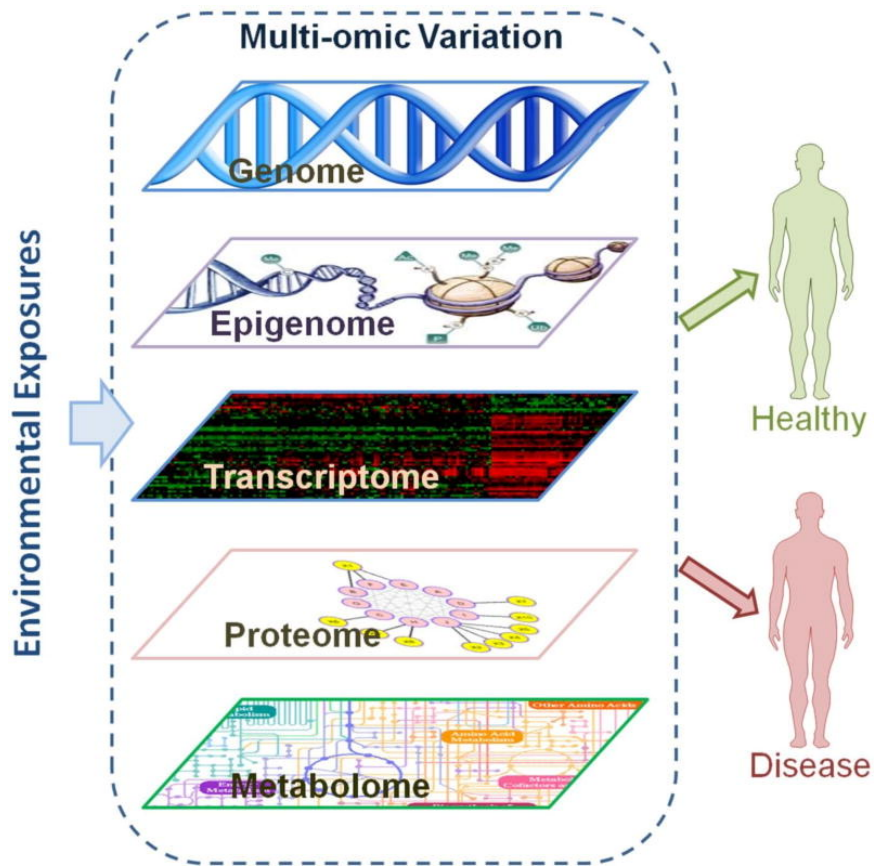
Bâtiment Yersin

# Objectif(s) de la journée

- Discuter autour de l'intégration de données
- Réviser l'ACP
- Faire « comme l'ACP » (aka RGCCA) mais avec plusieurs jeux de données
- Pouvoir appliquer RGCCA à des données multi-omiques
- Démystifier l'intégration

Intégration de données

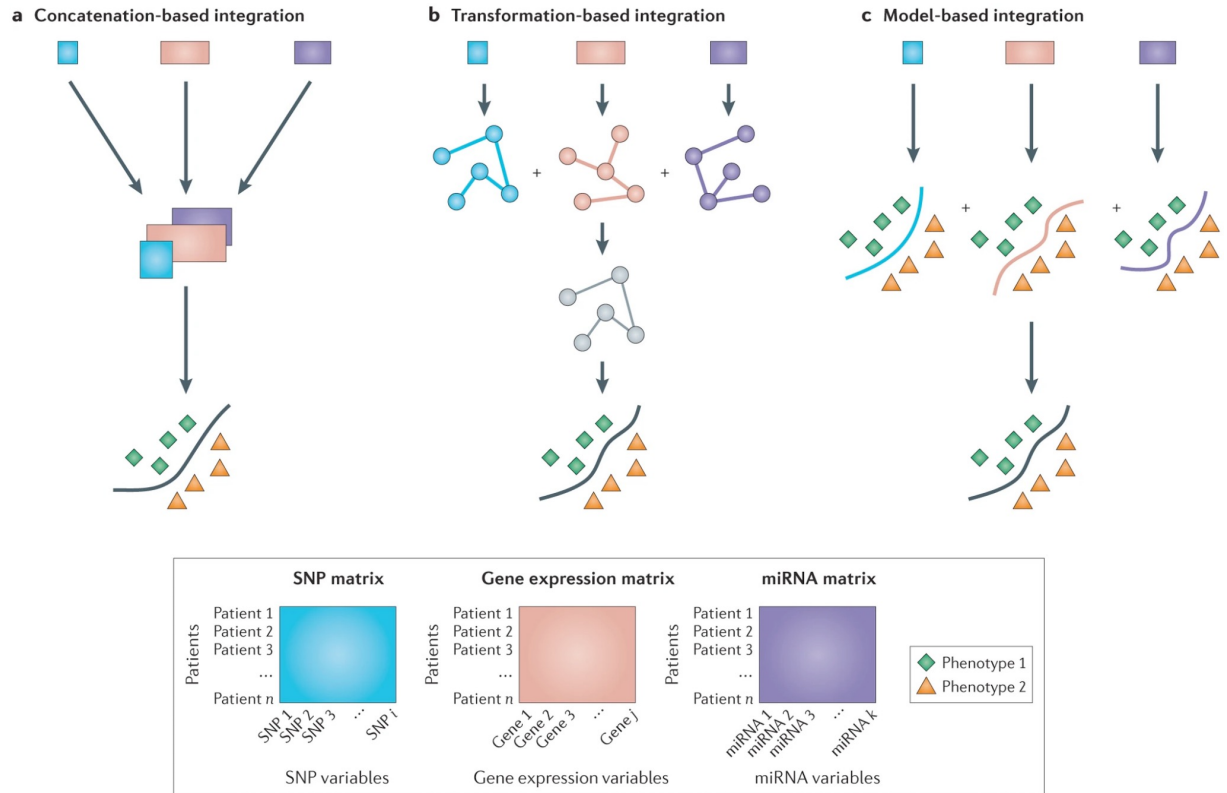
# Le contexte multi-omique



Credits: Laura Cantini.

<https://doi.org/10.1016/bs.adgen.2015.11.004>

# Typologie de l'analyse intégrative





Le  
benchmark  
de Laura  
Cantini et al.

jDR approach	Underlying approach	Constraints on the factors	Features or samples matching requirements	Implementation
<b>RGCCA</b>	Canonical Correlation Analysis (CCA)	omics-specific factors	matching of samples	R package RGCCA
<b>MCIA</b>	Co-Inertia Analysis (CIA)	omics-specific factors	matching of samples	R package omicade4
<b>MOFA</b>	Factor Analysis (FA) (Bayesian)	mixed factors	matching of samples (partial matching allowed)	R code on github bioFAM/MOFA
<b>MSFA</b>	Factor Analysis (FA) (Bayesian)	mixed factors	matching of samples	R code on github rdevito/MSFA
<b>intNMF</b>	Non-Negative Matrix Factorization (NMF)	shared factors	matching of samples	R package intNMF
<b>iCluster</b>	Gaussian latent variable model	shared factors	matching of samples	R package iCluster
<b>JIVE</b>	Principal Component Analysis (PCA)	mixed factors	matching of samples (partial matching allowed)	R package r.jive
<b>tICA</b>	Independent Component Analysis (ICA)	shared factors	matching of both samples and features (tensor)	R scripts in supplementary material of Teschendorff et al.
<b>Scikit-fusion</b>	Matrix tri-factorization	shared factors	matching of samples	Python scripts on github marinkaz/scikit-fusion

<https://www.nature.com/articles/s41467-020-20430-7>

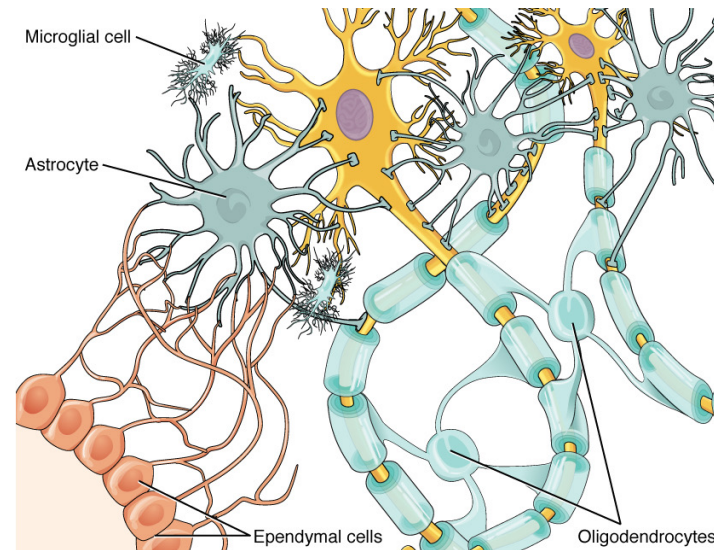
# Les gliomes malins pédiatriques

- Les tumeurs cérébrales malignes représentent 20 % des cancers pédiatriques
- Parmi elles, les tumeurs gliales sont les plus fréquentes (60%)
- Survie médiane de 12 à 14 mois
- Survie globale à 5 ans : 20 %
- Traitement : exérèse chirurgicale + radiothérapie



# Les gliomes malins pédiatriques

- Les tumeurs cérébrales malignes représentent 20 % des cancers pédiatriques
- Parmi elles, les tumeurs gliales sont les plus fréquentes (60%)
- Survie médiane de 12 à 14 mois
- Survie globale à 5 ans : 20 %
- Traitement : exérèse chirurgicale + radiothérapie

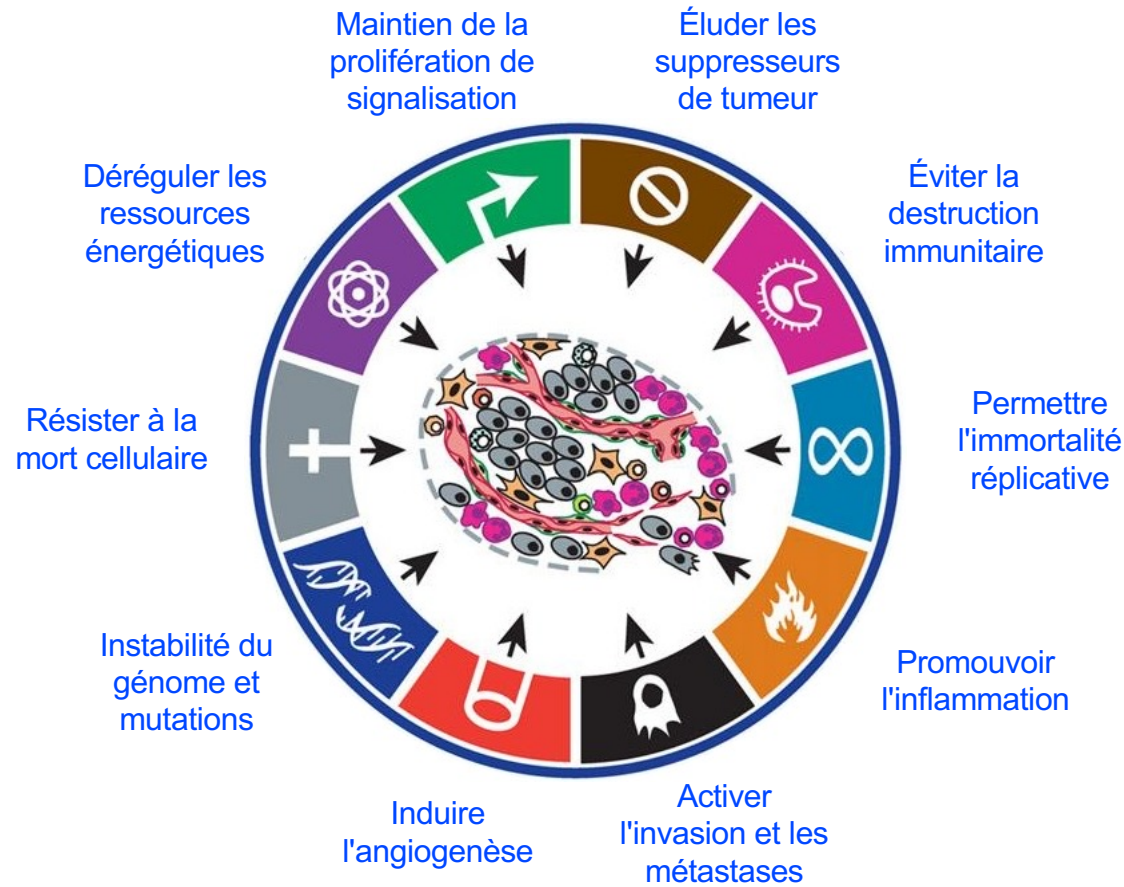


# Pourquoi une classification ?

- Classification de référence : OMS 2007 (adultes)
  - Rôle histo-pronostique
  - Choix d'un traitement adapté
  - Optimisation de la prise en charge du patient

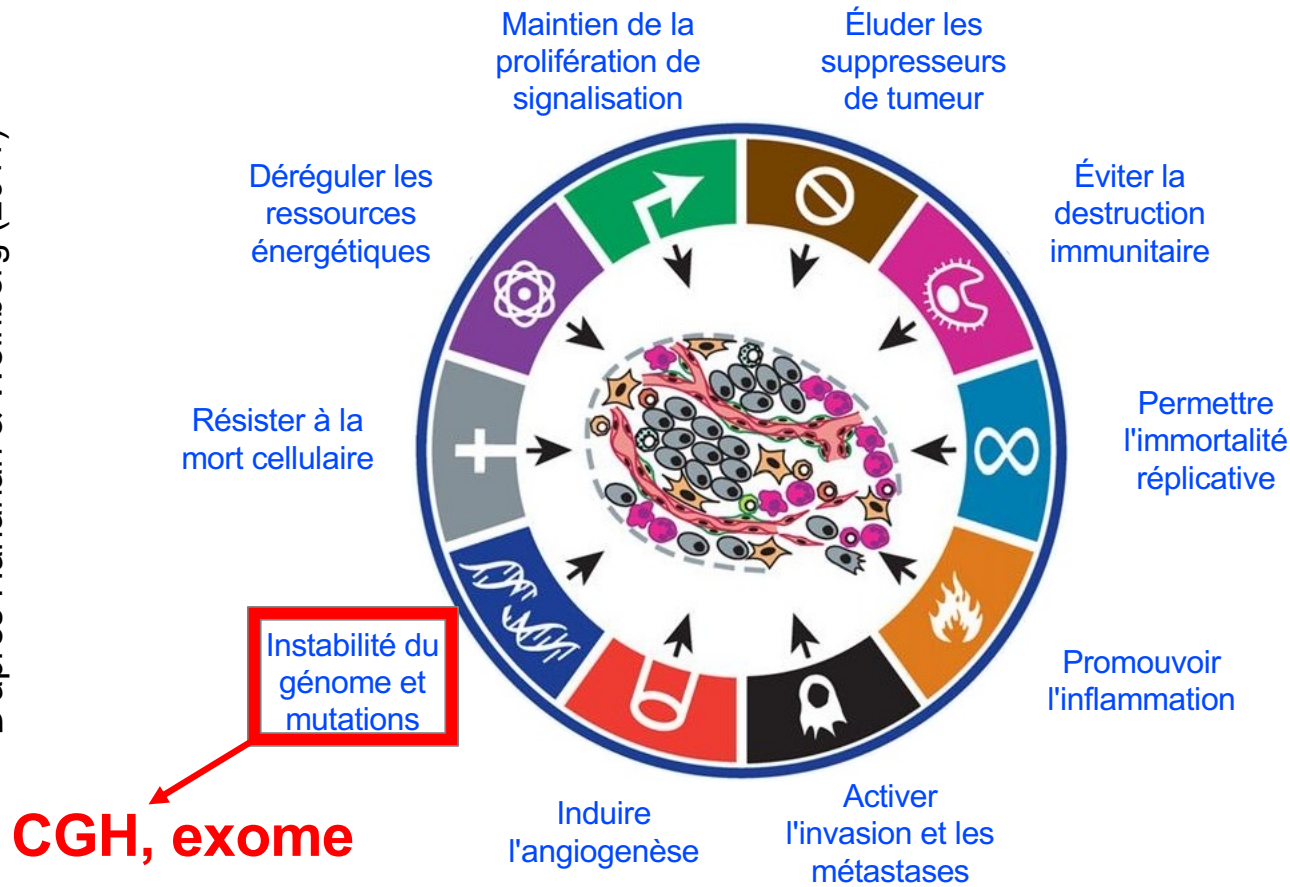
# Le cancer, maladie aux multiples facettes

D'après Hanahan & Weinberg (2011)



# Le cancer, maladie aux multiples facettes

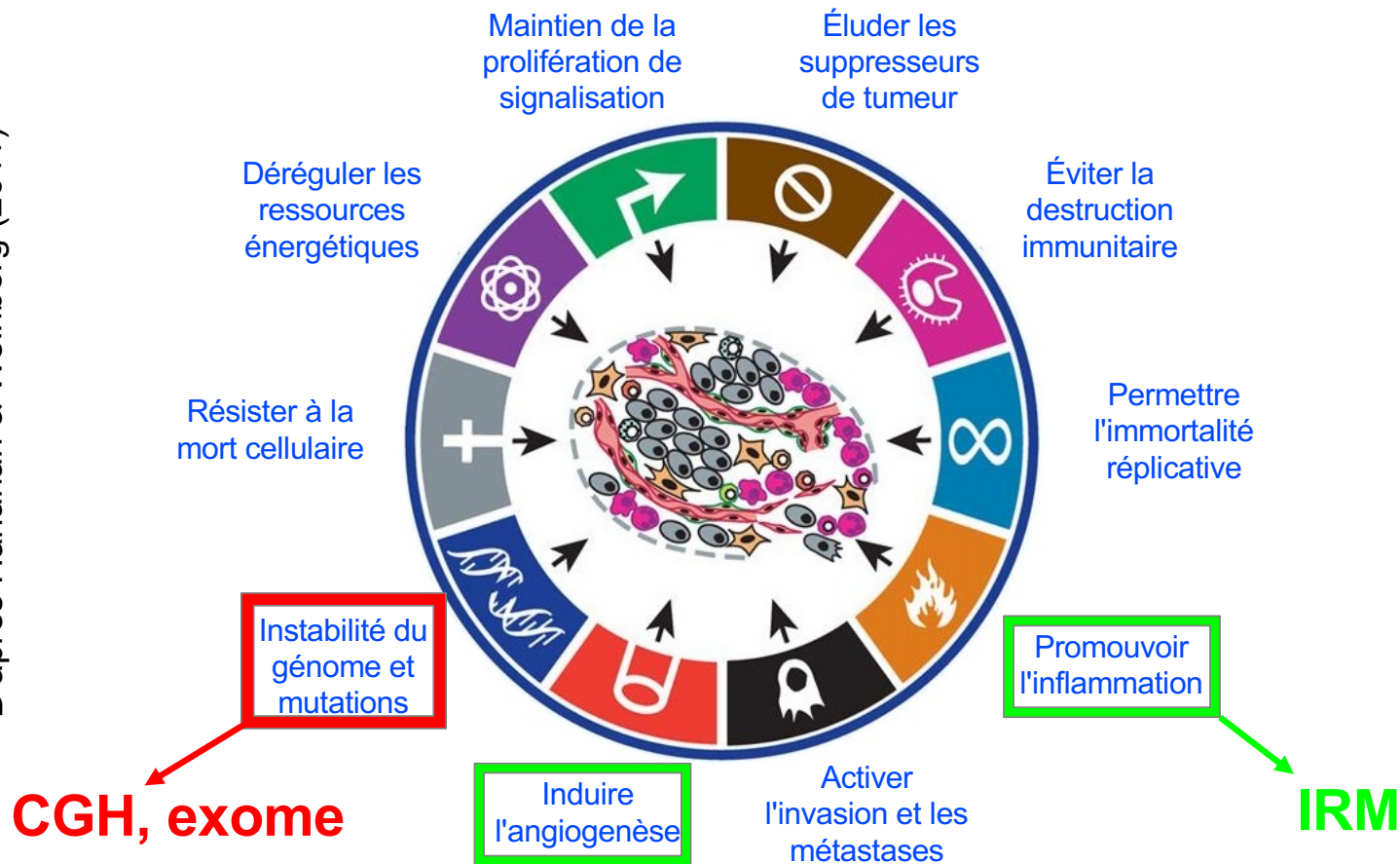
D'après Hanahan & Weinberg (2011)



**CGH, exome**

# Le cancer, maladie aux multiples facettes

D'après Hanahan & Weinberg (2011)

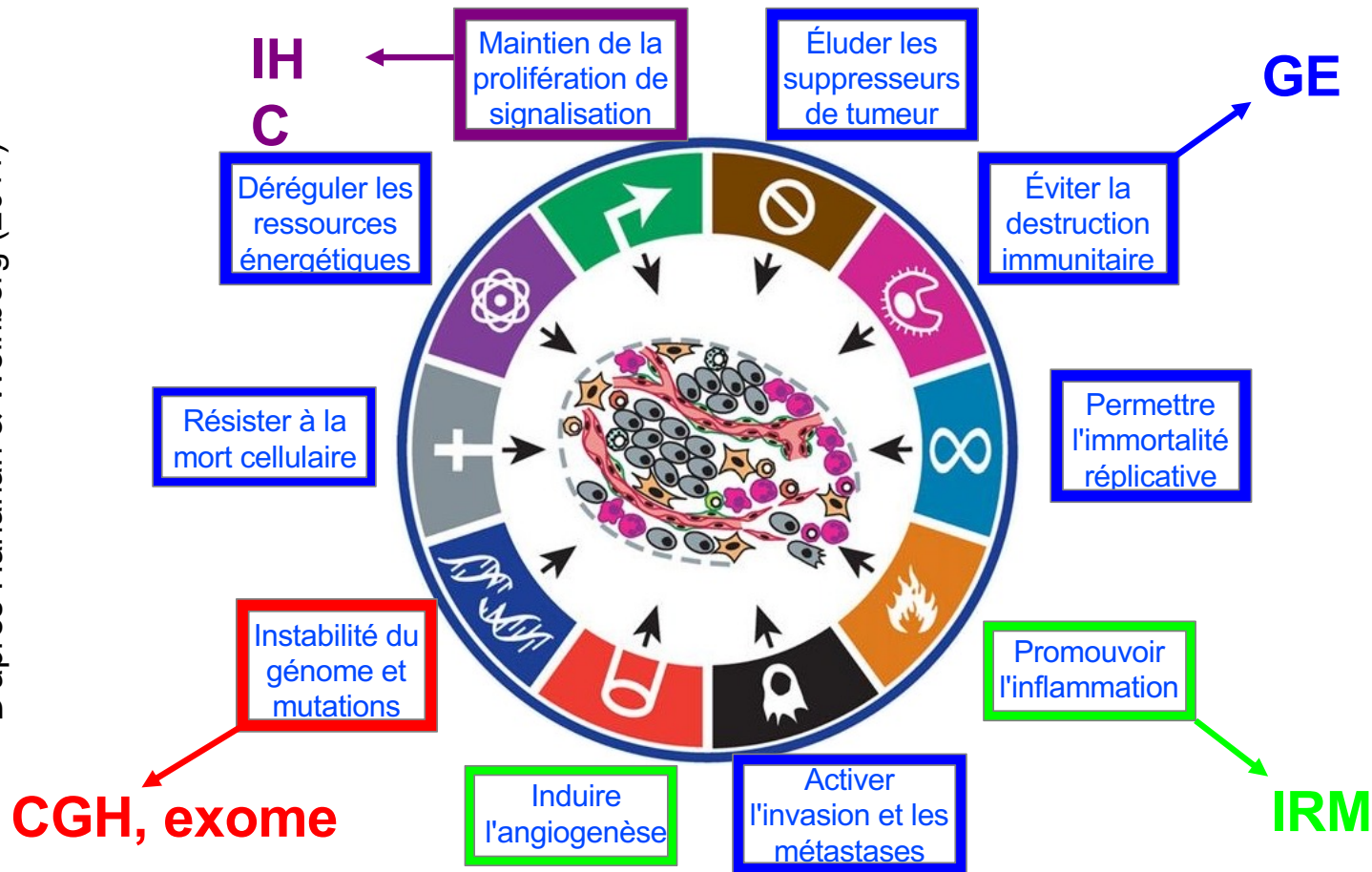






# Le cancer, maladie aux multiples facettes

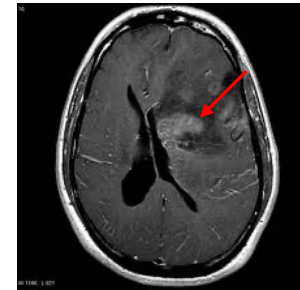
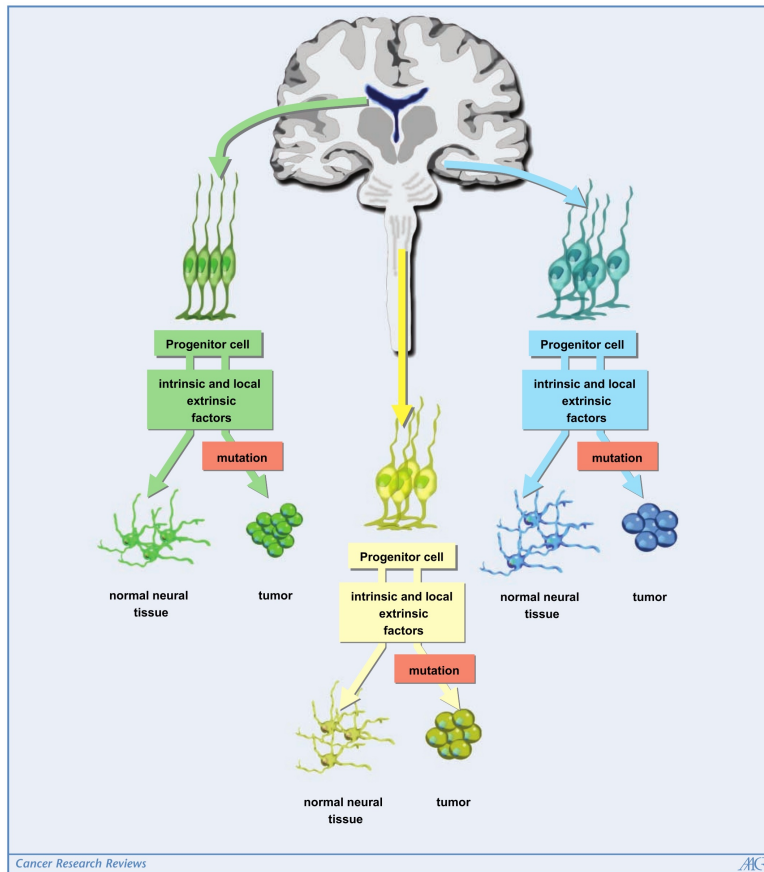
D'après Hanahan & Weinberg (2011)



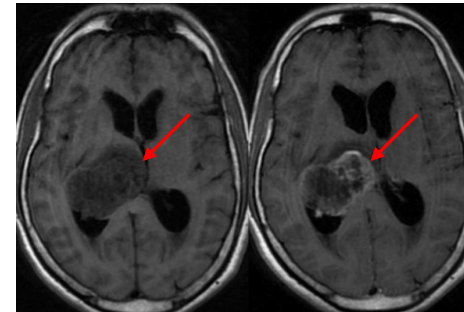


# Emplacement et niche tumorale

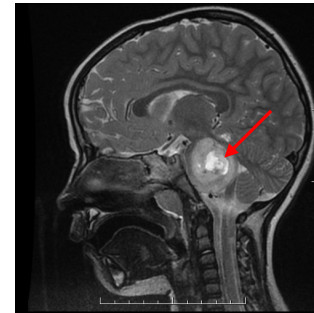
Gilbertson *et al.* (2007)



Hémisphère



Ligne médiane



Tronc cérébral

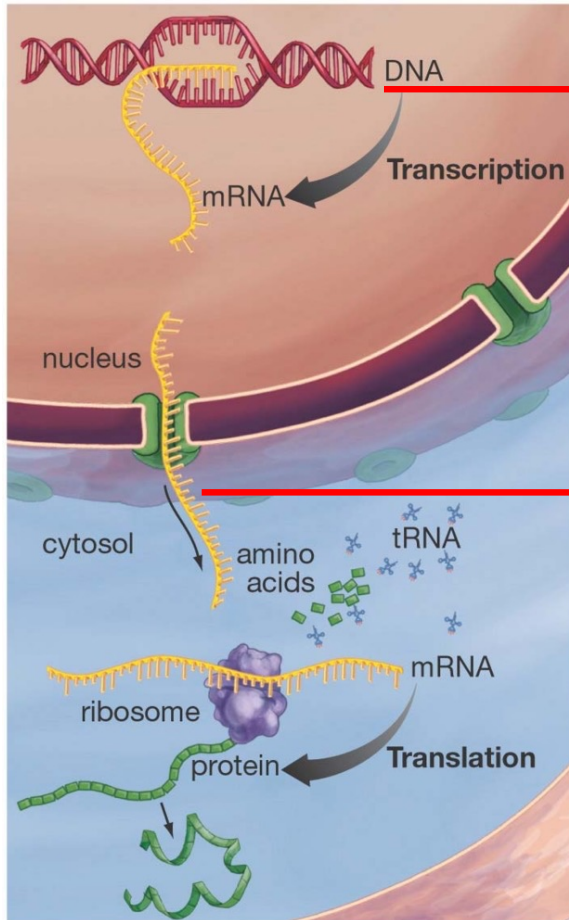
# Critères d'inclusion

- Tumeur gliale primaire
- Grade OMS III et IV + relecture
- $0 < \text{âge} < 18$  ans
- Toutes localisations
- ADN + ARN

N=53 patients

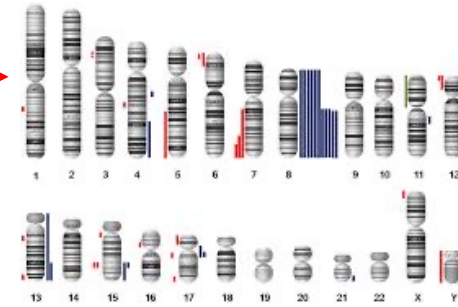
Survie médiane = 17 mois

# CGH et GE



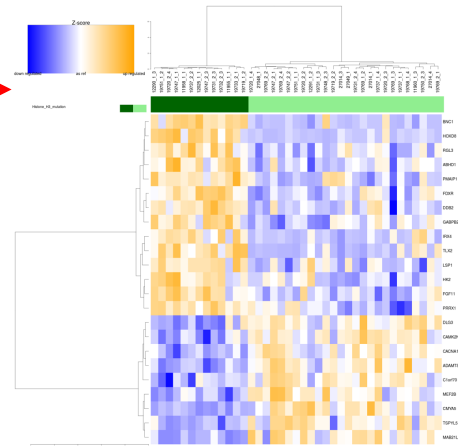
Aberrations  
du nombre  
de copies

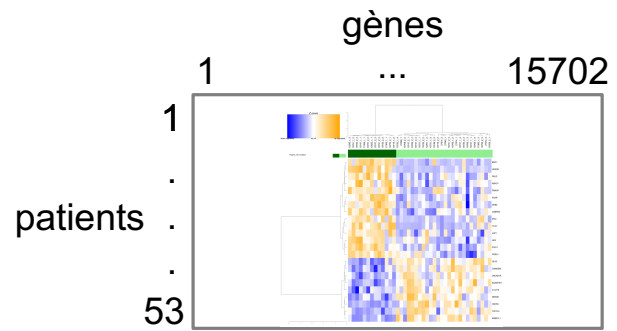
Comparative  
Genomique  
Hybridization



Expression  
des gènes

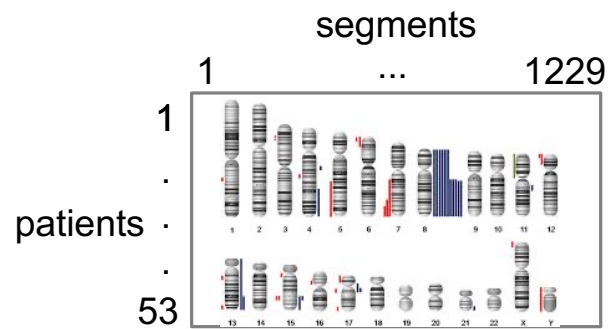
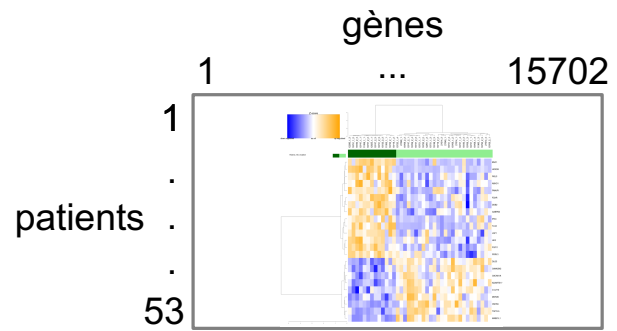
Gene Expression



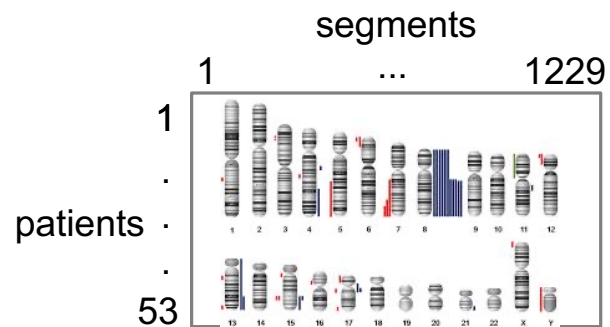
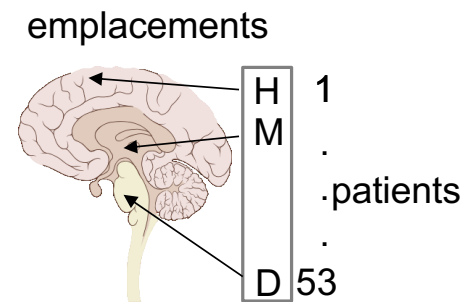
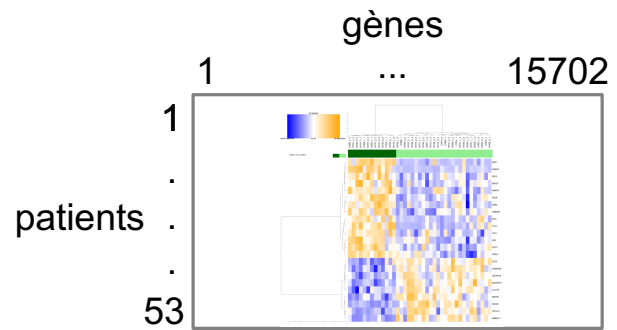


# Le jeu de données pHGG

# Le jeu de données pHGG



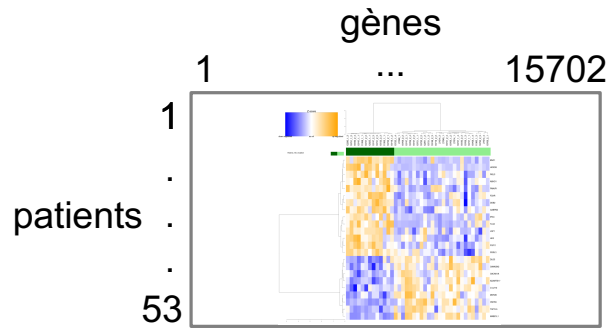
# Le jeu de données pHGG



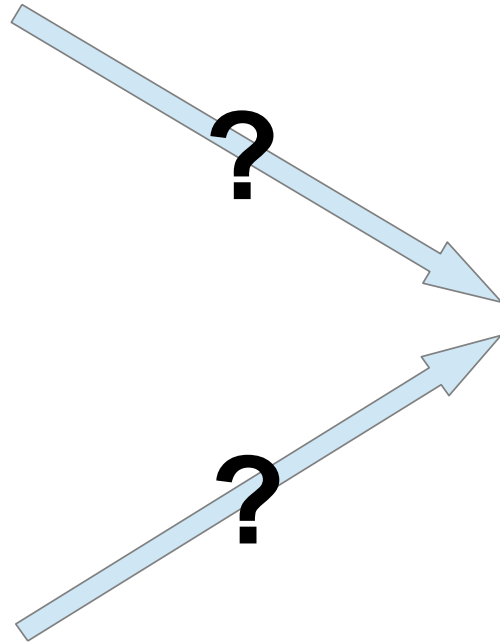
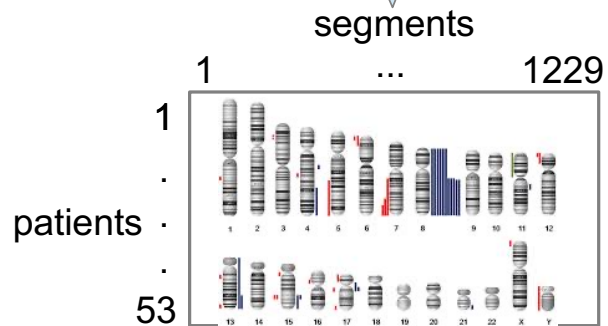




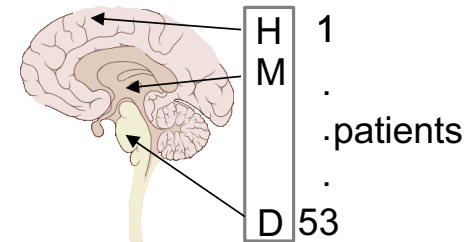
# Le jeu de données pHGG



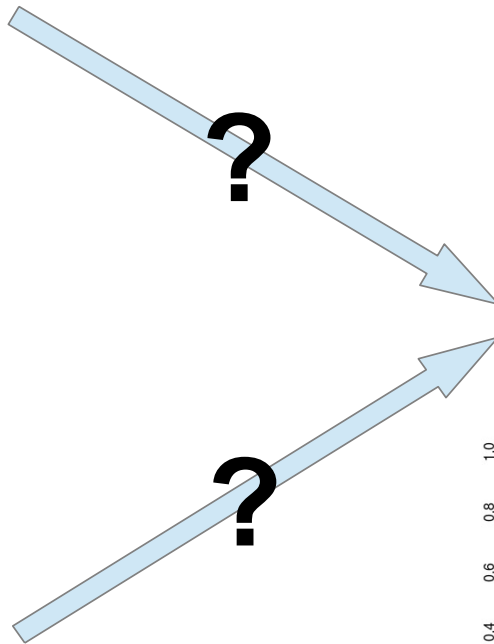
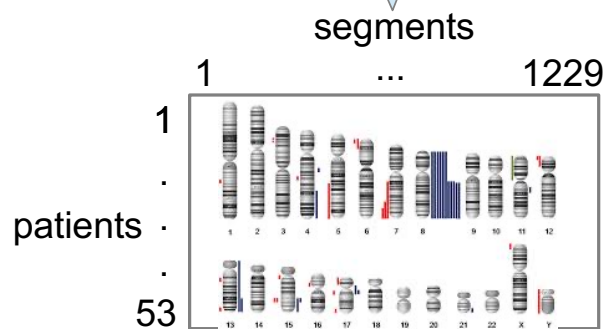
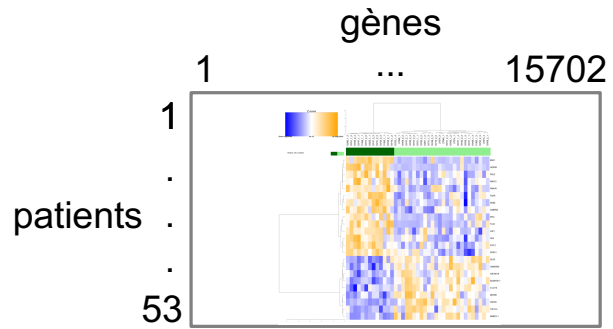
?



emplacements



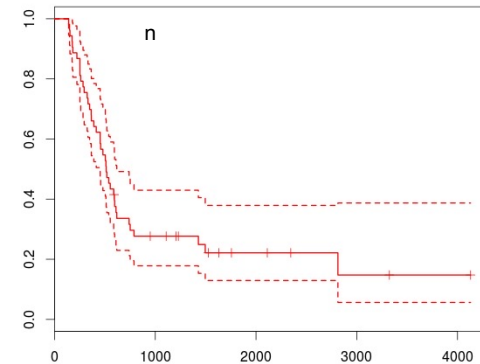
# Le jeu de données pHGG



$y_{surv}$

t	1	1
t	0	53

patients



# Questions

- **Emplacement de la tumeur** (hémisphères, ligne médiane, tronc cérébral)
  - Quels sont les gènes et les segments caractéristiques
  - de chaque localisation, qui pourraient nous renseigner
    - sur la cellule d'origine des tumeurs ?

# Questions

- **Emplacement de la tumeur** (hémisphères, ligne médiane, tronc cérébral)
  - Quels sont les gènes et les segments caractéristiques
  - de chaque localisation, qui pourraient nous renseigner
    - sur la cellule d'origine des tumeurs ?
- **Survie globale**
  - L'expression des gènes et les aberrations du nombre de copies ont-elles une influence sur la survie globale ? si oui, laquelle ?

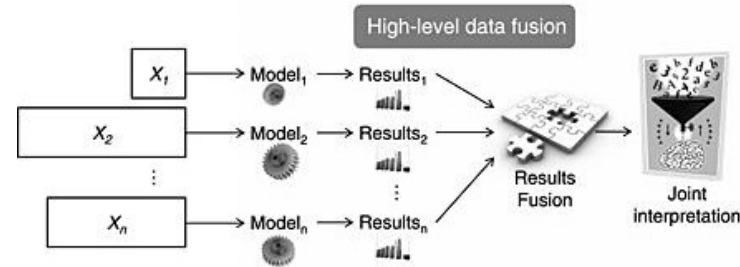
# Questions

- **Emplacement de la tumeur** (hémisphères, ligne médiane, tronc cérébral)
  - Quels sont les gènes et les segments caractéristiques
  - de chaque localisation, qui pourraient nous renseigner
    - sur la cellule d'origine des tumeurs ?
- **Survie globale**
  - L'expression des gènes et les aberrations du nombre de copies ont-elles une influence sur la survie globale ? si oui, laquelle ?

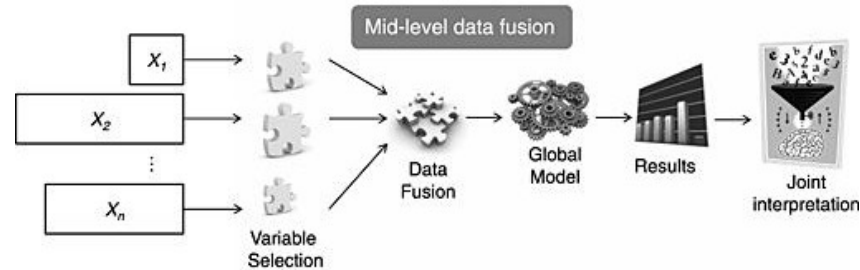
**Améliorer la classification  
des gliomes malins pédiatriques,  
notamment son pouvoir pronostique.**

# Méthodes d'intégration de données

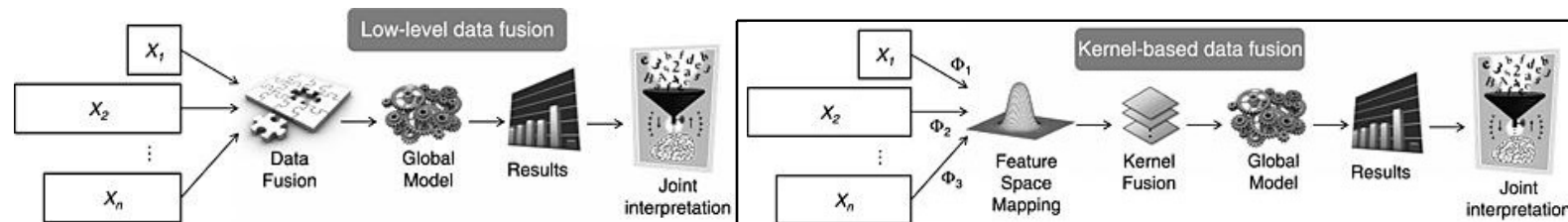
D'après Boccard & Rudaz  
(2014)



Sturm *et al* (2012)

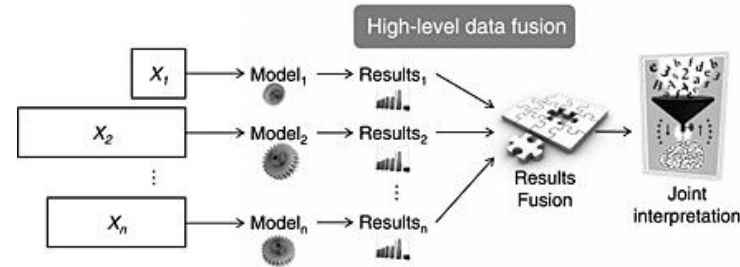


Le Floch *et al* (2012)

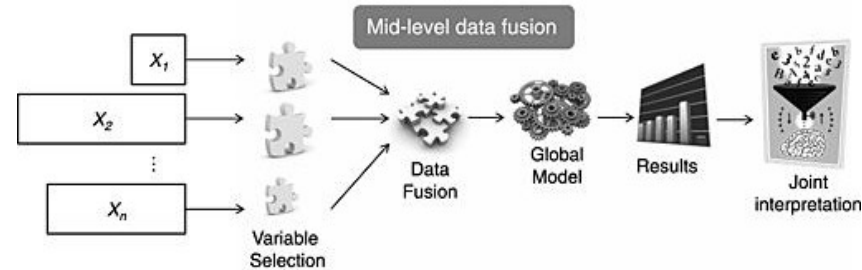


# Méthodes d'intégration de données

D'après Boccard & Rudaz  
(2014)



Sturm *et al* (2012)



Le Floch *et al* (2012)

