

Intégration de données hétérogènes : imagerie, clinique, génomique

Olivier Colliot, Directeur de Recherche CNRS

Responsable de l'équipe ARAMIS – www.aramislab.fr

Chaire de l'Institut 3IA PRAIRIE



**ARAMIS
LAB**
BRAIN DATA SCIENCE



inria
informatics mathematics

Inserm
Institut national
de la santé et de la recherche médicale

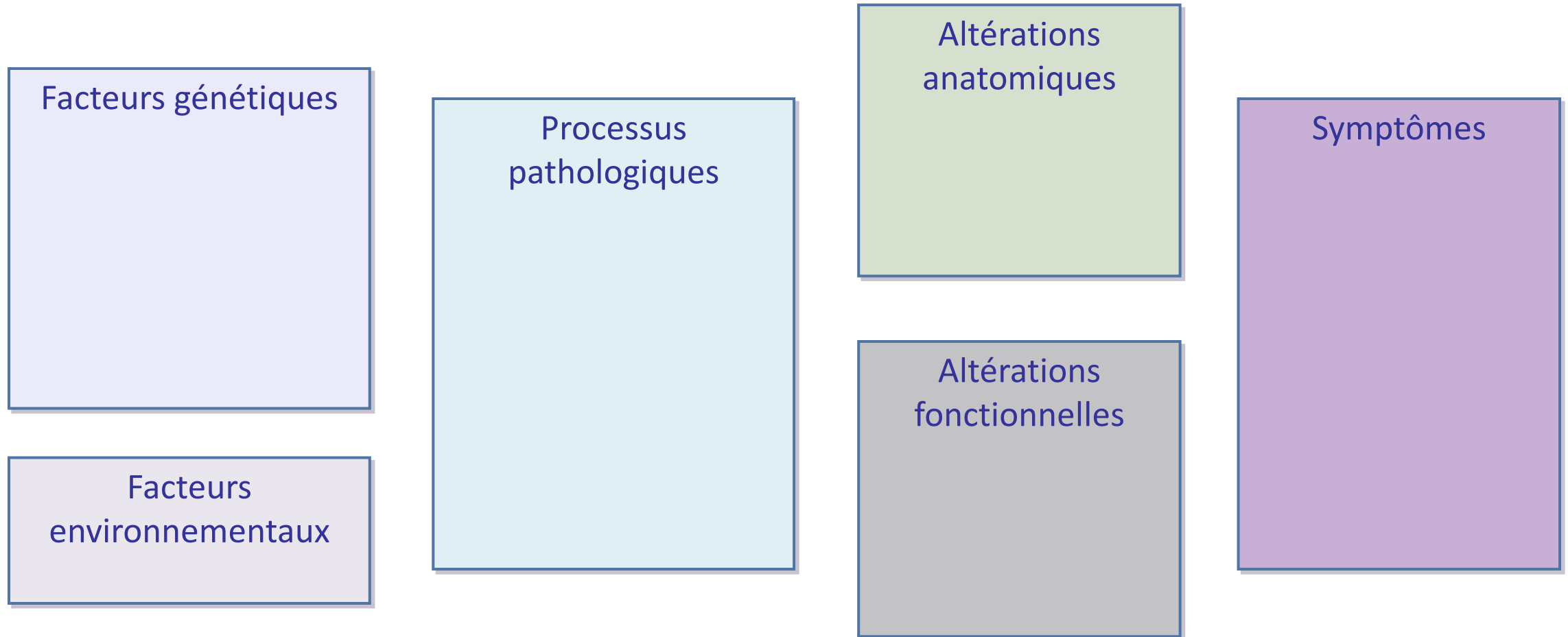


**SORBONNE
UNIVERSITÉ**
CRÉATEURS DE FUTURS
DEPUIS 1257

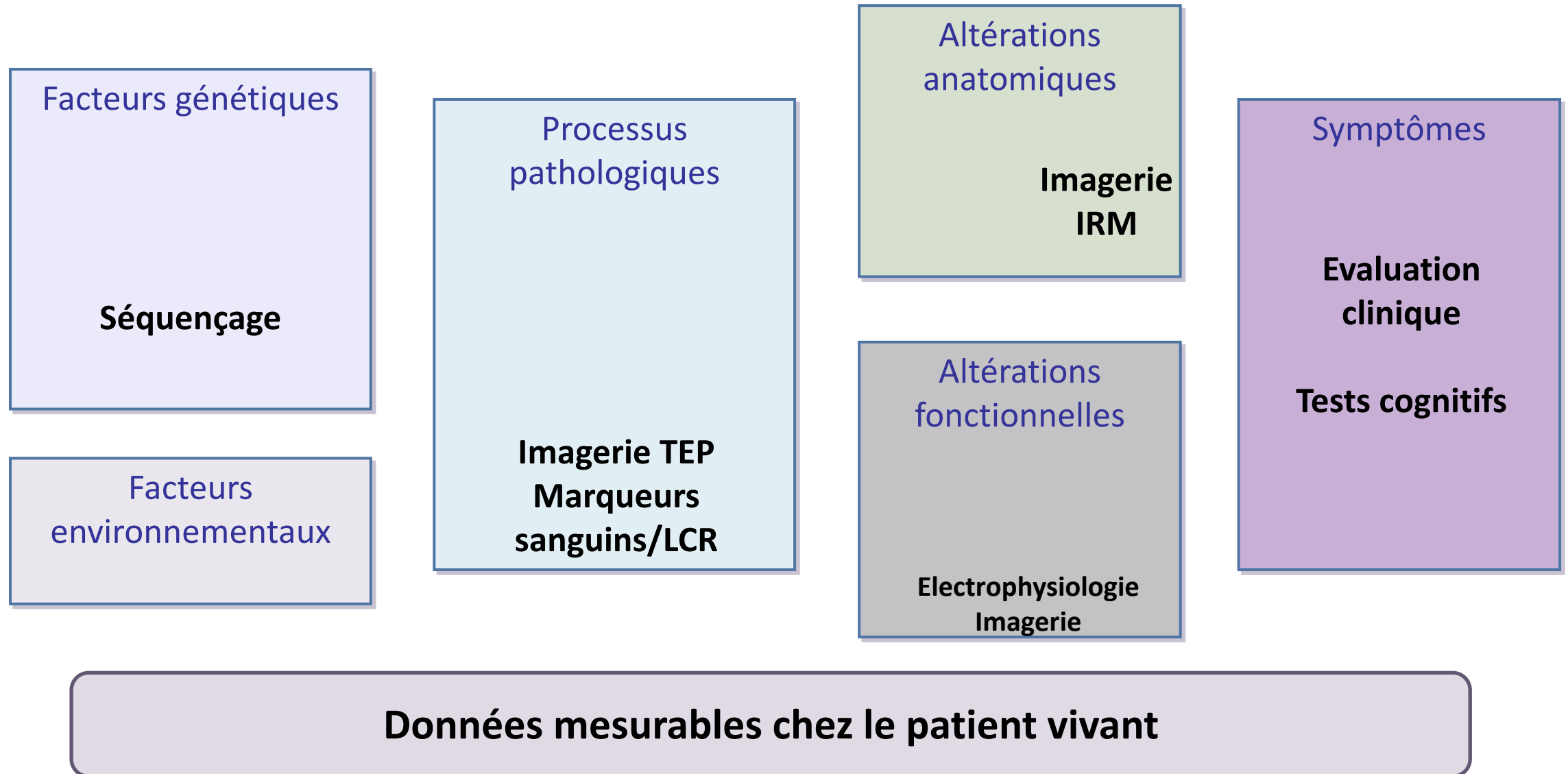


- **Maladie d'Alzheimer**
 - **Maladie de Parkinson**
 - **Sclérose en plaques**
-

L'exemple des maladies neurodégénératives

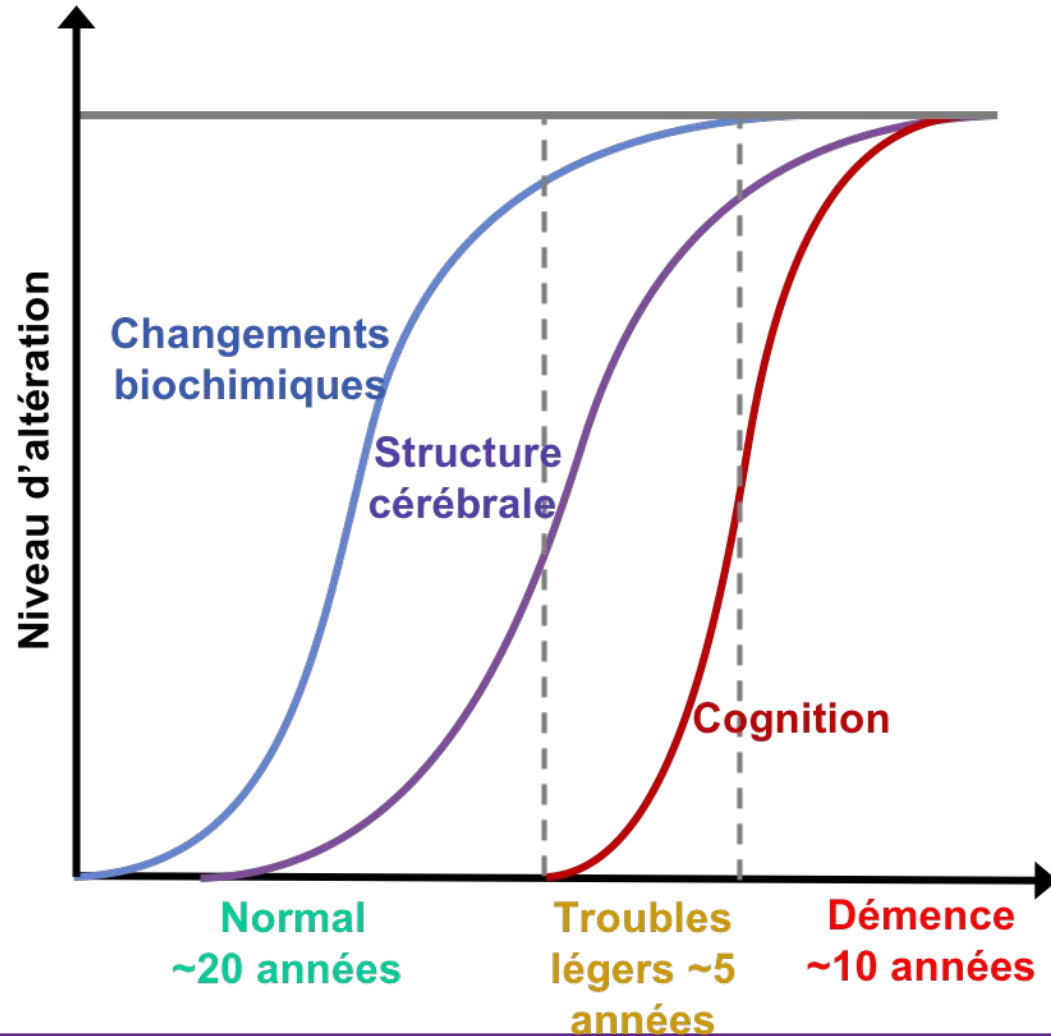


L'exemple des maladies neurodégénératives



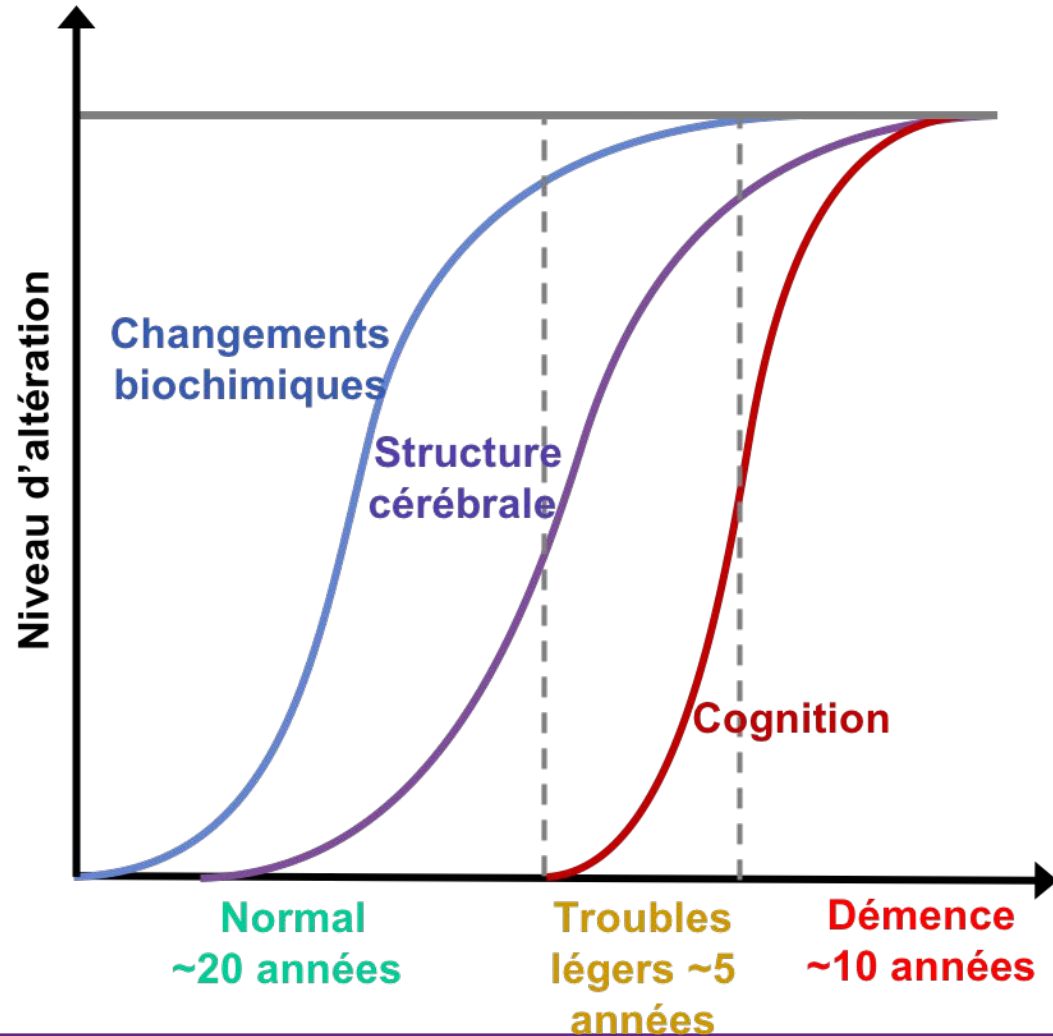
- **Dynamique temporelle**

- Evolution sur plusieurs décennies, longue phase silencieuse asymptomatique



L'exemple des maladies neurodégénératives

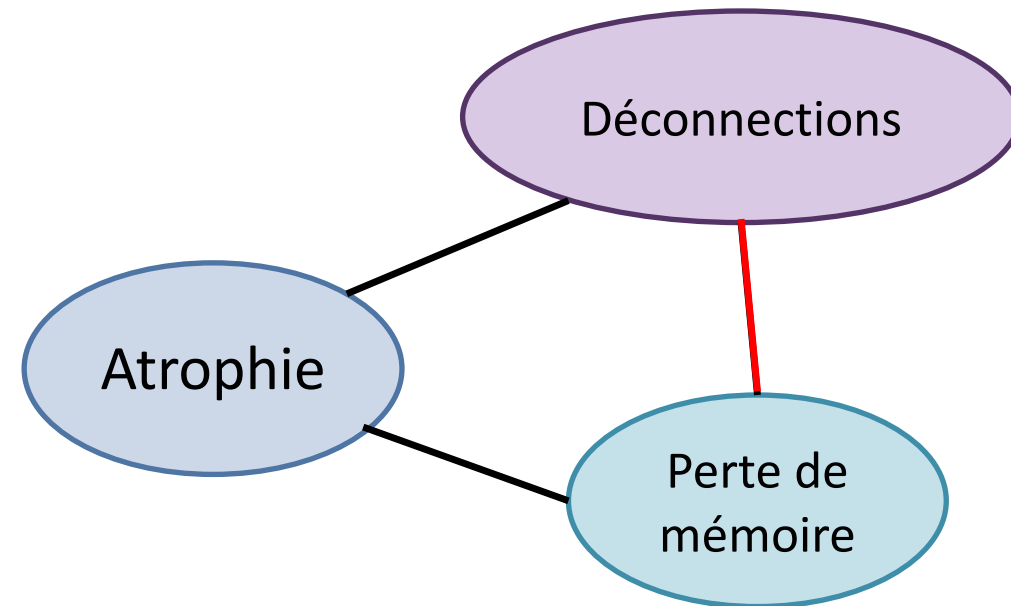
- **Dynamique spatio-temporelle**



- **Quelles sont les relations entre ces mesures ?**
 - **Quelles mesures sont nécessaires à la caractérisation de la maladie ?**
 - Les mesures réalisées sont-elles complémentaires ou redondantes ?
 - **Quelle est la dynamique temporelle des altérations ?**
 - Quelles altérations apparaissent en premier ?
 - Quelles régions du cerveaux sont les premières touchées ?
 - **Peut-on combiner ces informations pour prédire et prévenir la maladie ?**
-

Etudes des relations entre les mesures

- **Différentes altérations sont reliées**
 - Atrophie, hypométabolisme, déconnexions, déclin cognitif...
- **Corrélations**
 - Graphe de corrélation peu informatif car presque toutes les mesures sont reliées (éventuellement via des variables intermédiaires)
- **Corrélation conditionnelle**
 - Supprime l'effet des variables intermédiaires



Etude des relations entre mesures

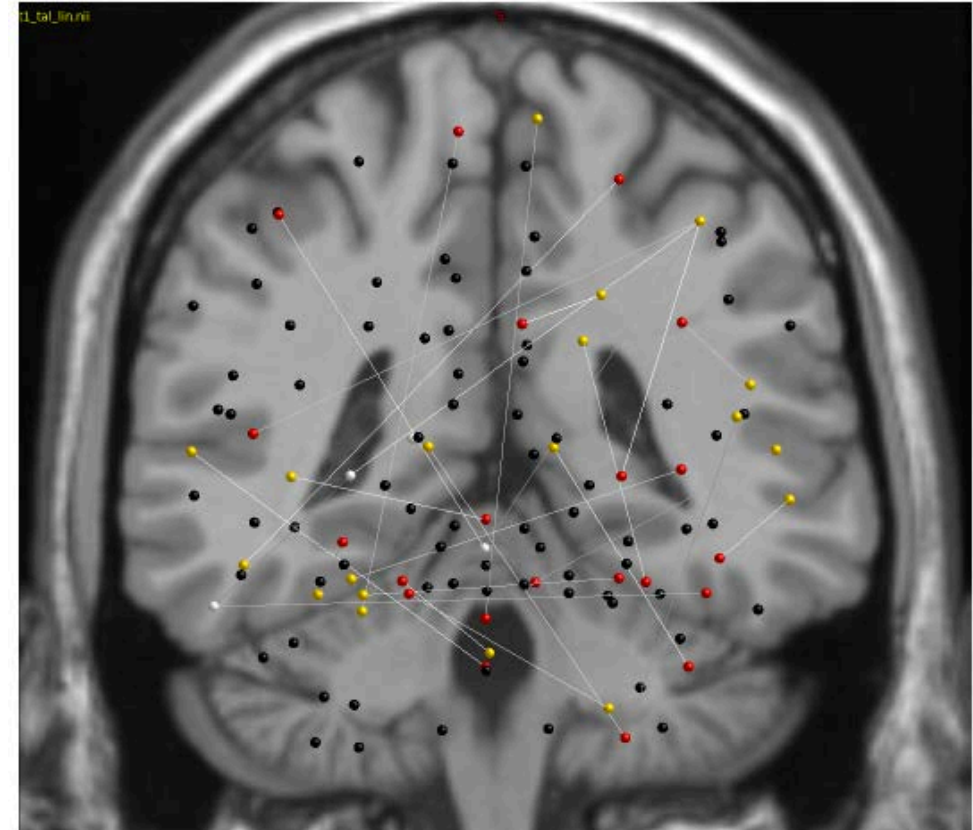
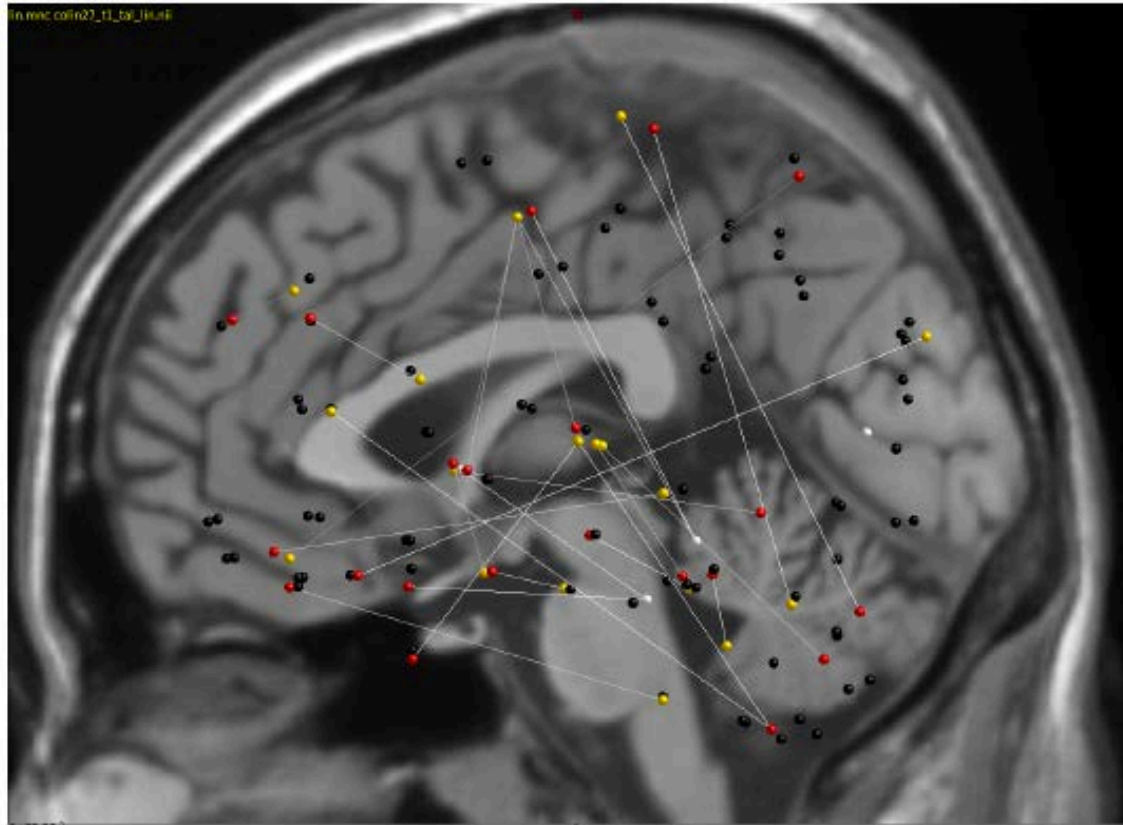
- **Grande dimension**

- Typiquement plusieurs centaines de mesures pour chaque modalité
- Nouvelle approche pour l'estimation de modèles graphique gaussiens en grande dimension

- **Application à la maladie d'Alzheimer**

- Etude des relations entre atrophie (IRM anatomique), métabolisme (TEP) et disconnections (IRM de diffusion)
 - 92 patients de la base ADNI

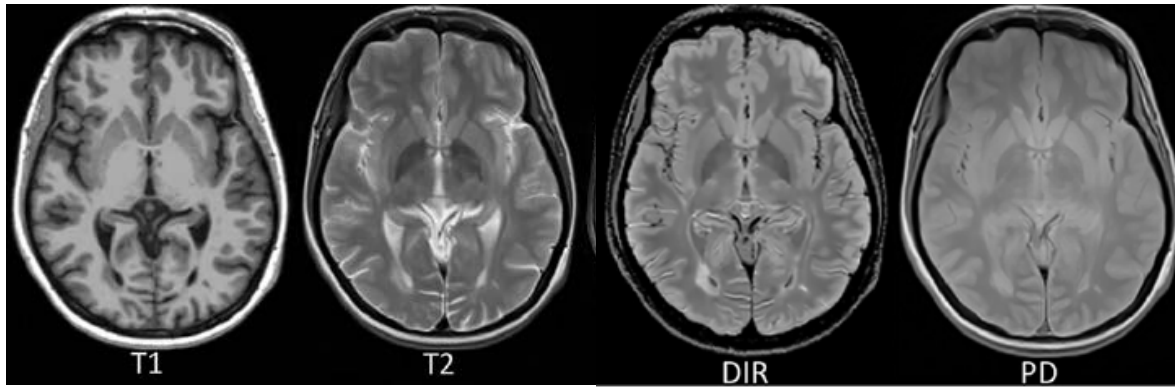
*Collaboration: S. Allasonnière
(Université Paris-Descartes)*



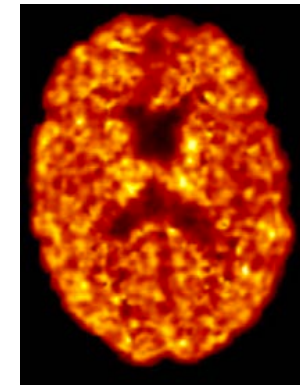
Graphe des relations entre mesures IRM (atrophie) (en rouge)
et TEP (metabolisme) (en jaune)

Redondance des mesures:
synthétiser un type de donnée à
partir d'un autre

- **Multiple examens** pour caractériser une pathologie donnée
 - Temps d'acquisition, coût, contraintes pour le patient



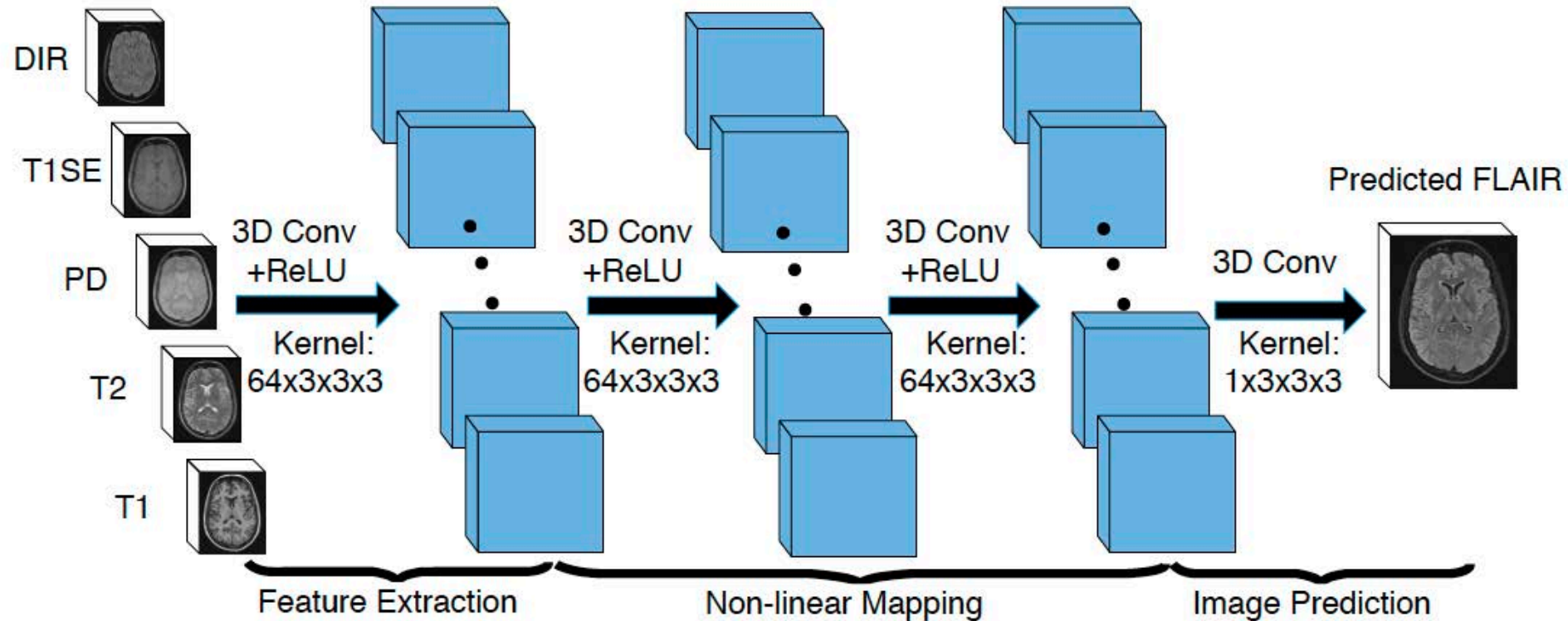
Différentes séquences d'un examen IRM



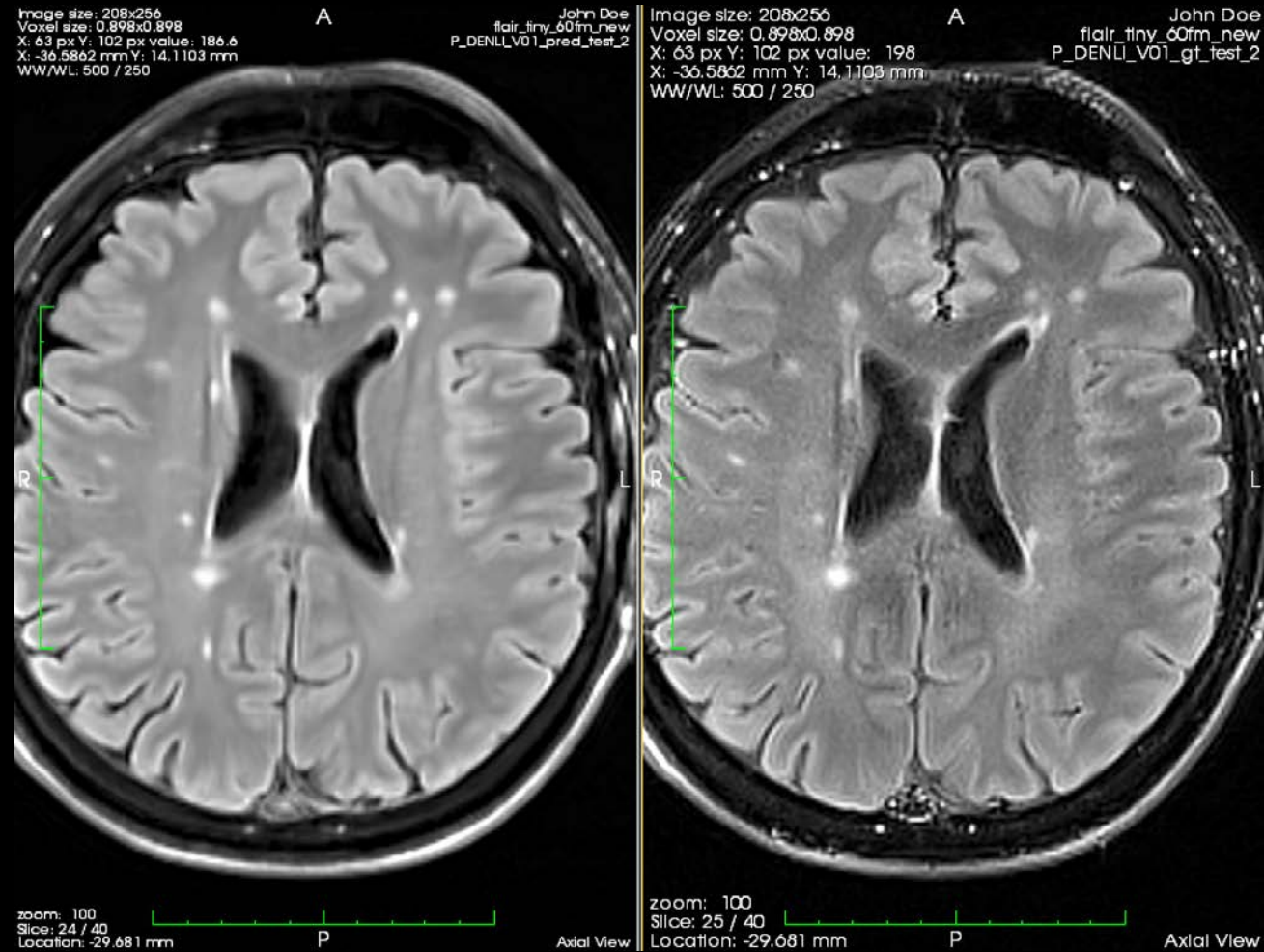
Imagerie TEP

- Les informations contenues dans ces examens sont-elles **complémentaires ou redondantes** ?
 - Est-il possible de **synthétiser un examen à partir d'autres** ?

➤ Synthèse d'IRM FLAIR à partir d'autres séquences IRM

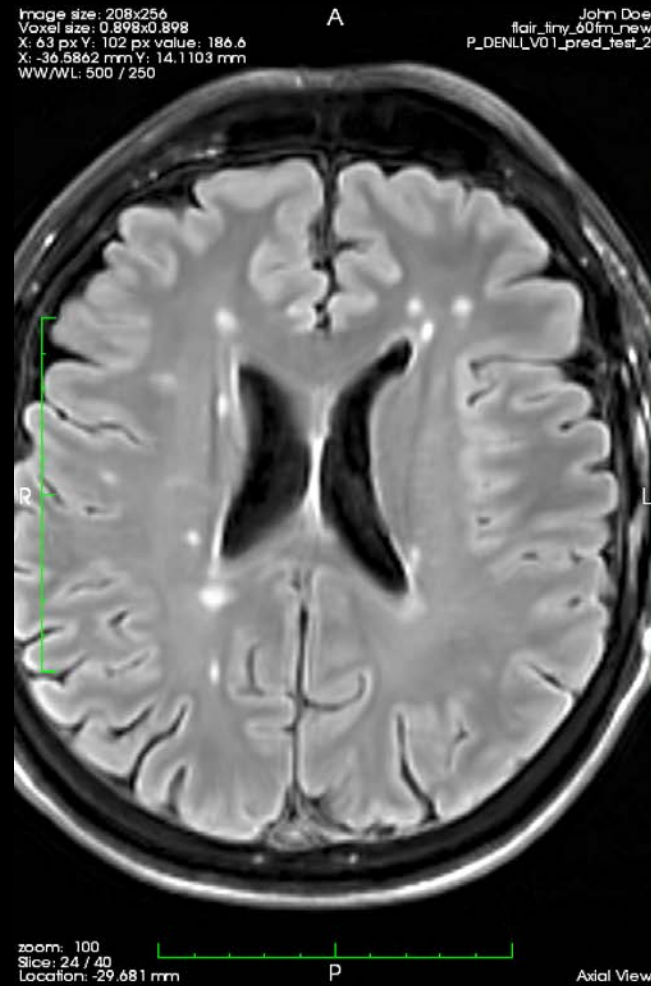


Laquelle de ces images est synthétique?

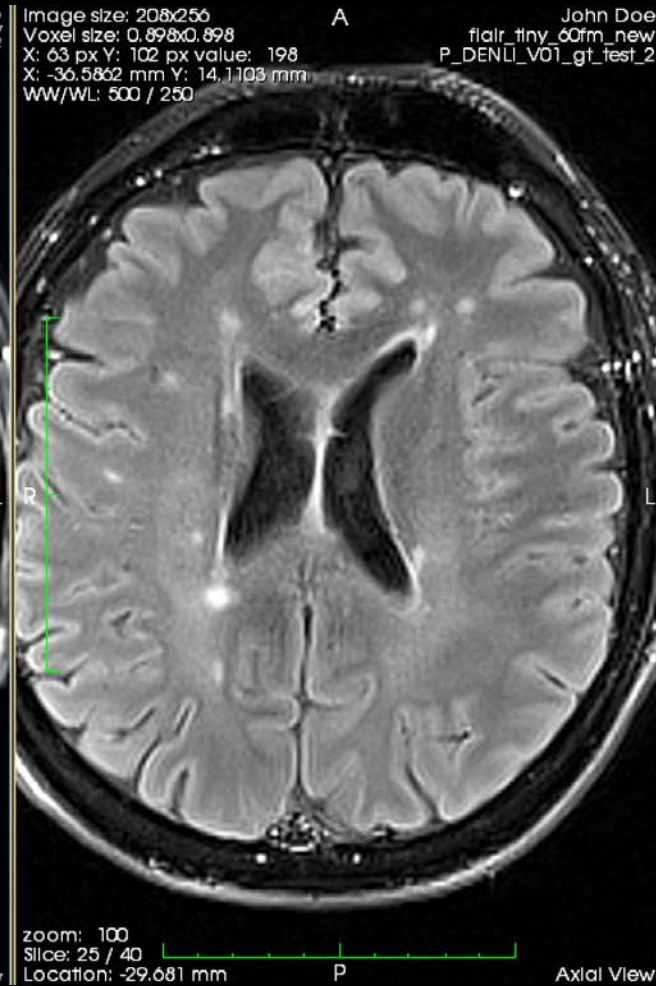


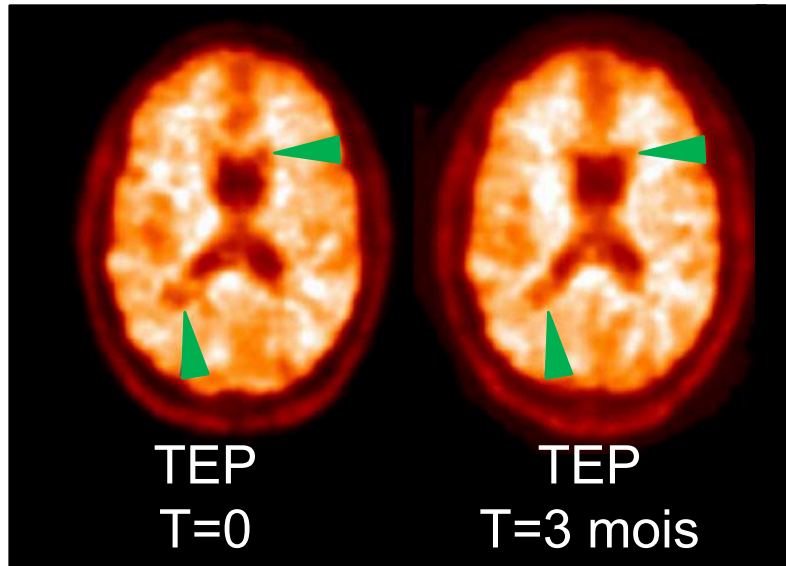
Laquelle de ces images est synthétique?

Synthétique

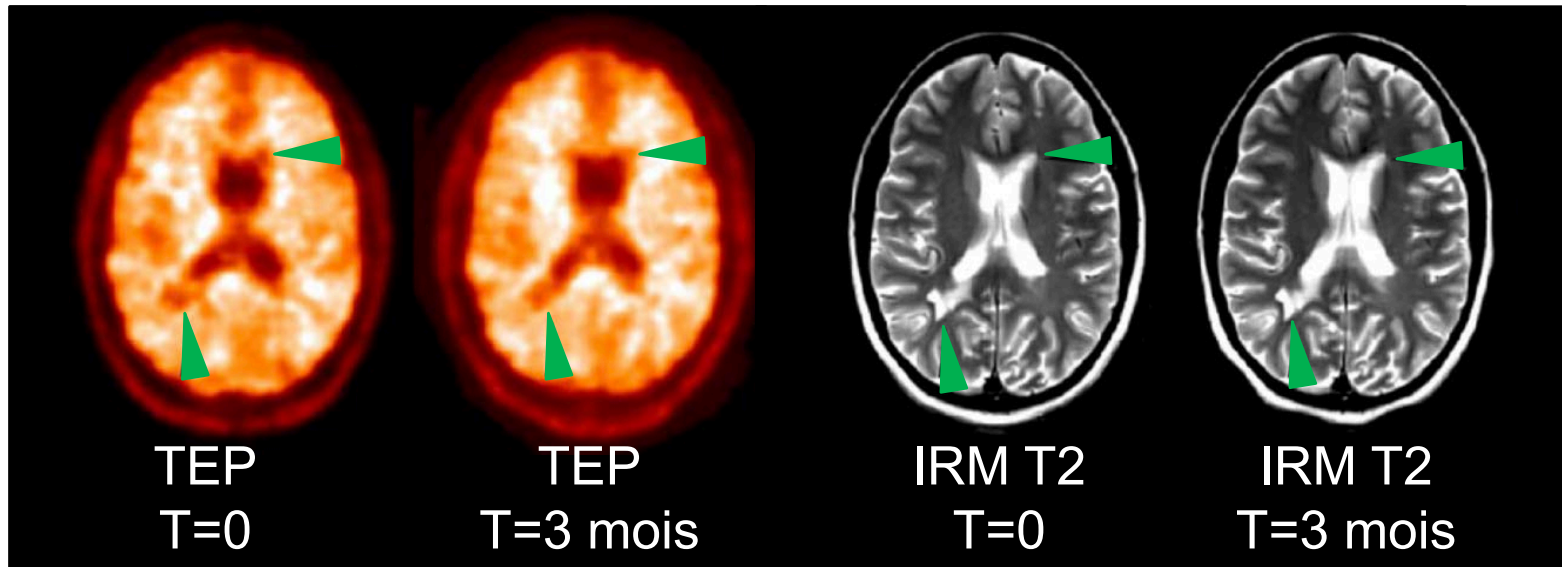


Vraie

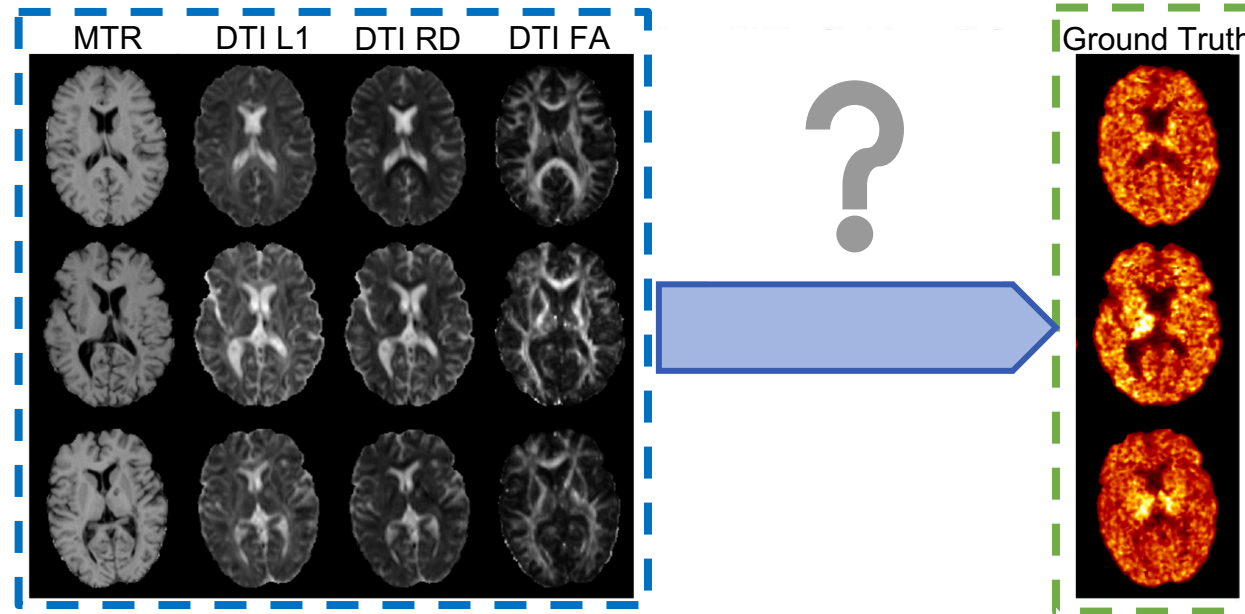




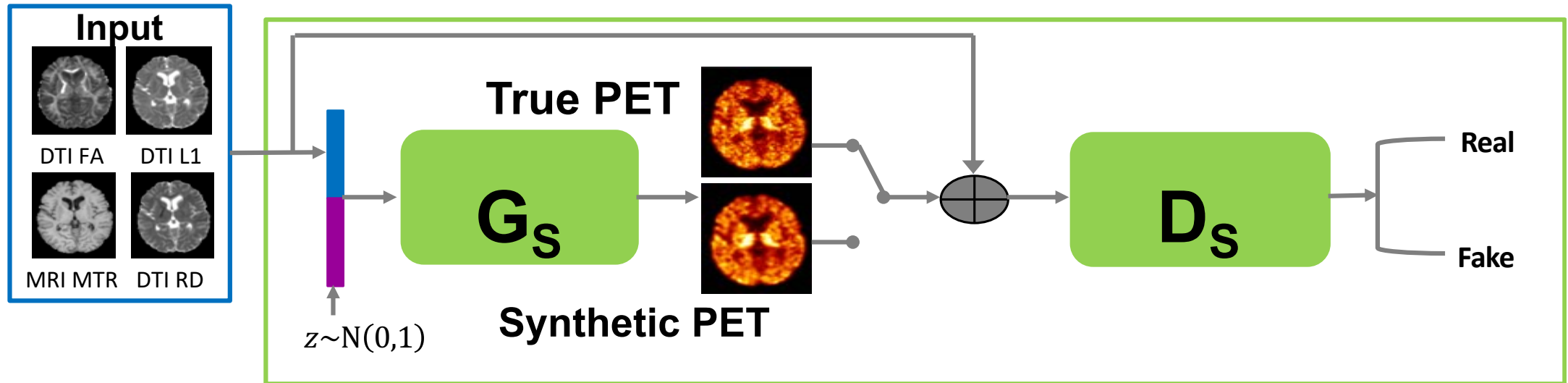
- **L'imagerie TEP peut mesurer la démyélinisation** dans la sclérose en plaques (avec le traceur $[^{11}\text{C}]$ PIB).
- **Mais : examen coûteux et invasif** (traceur radioactif)



- **L'imagerie IRM est largement disponible et non-invasive**
- **Mais** : les sequences IRM actuelles ne fournissent **pas d'indices spécifiques de la démyélinisation**



Synthétiser **les mesures issues de l'imagerie TEP** à partir de **données IRM multimodales** (MTR et DTI qui sont sensibles à la myéline)

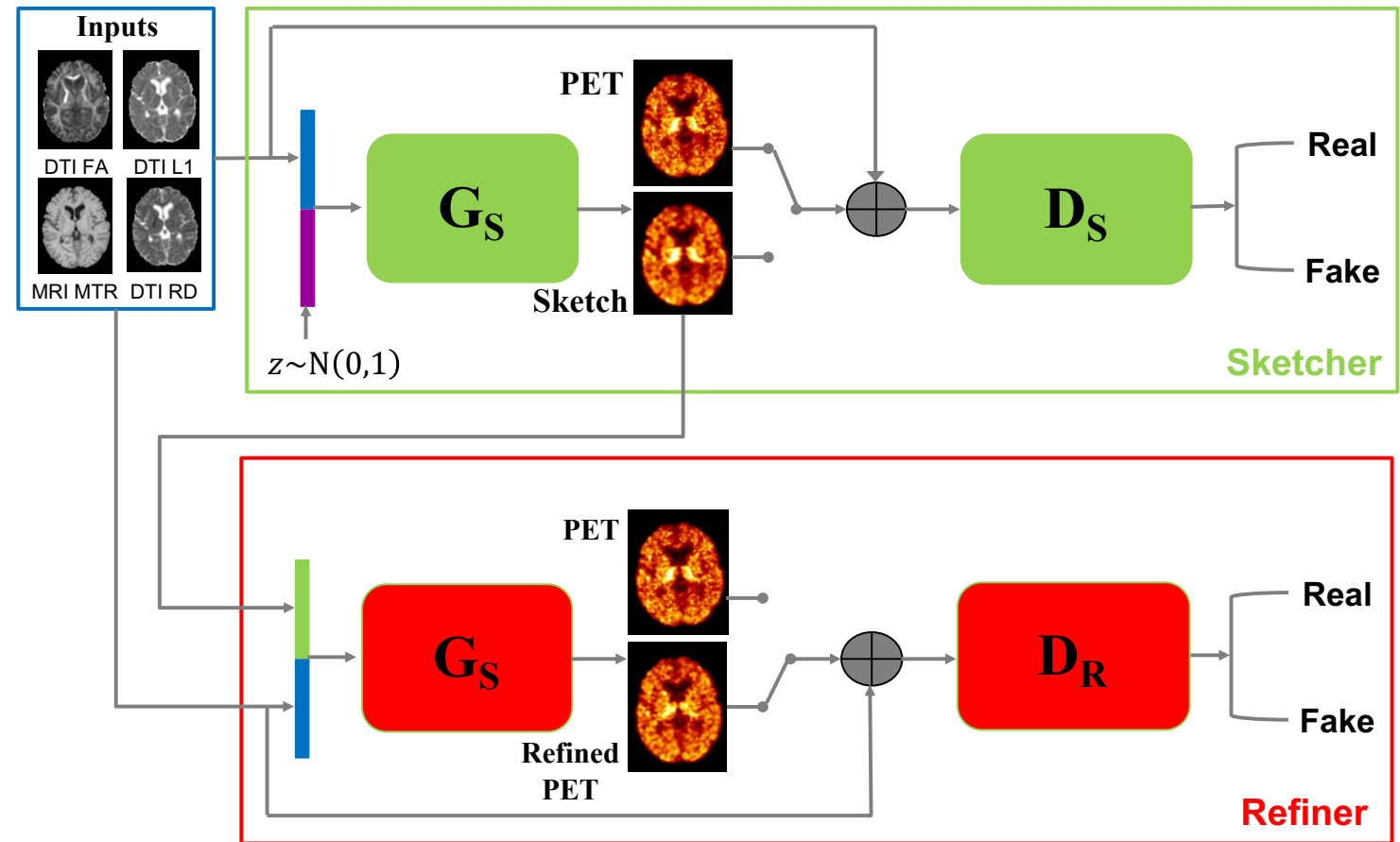


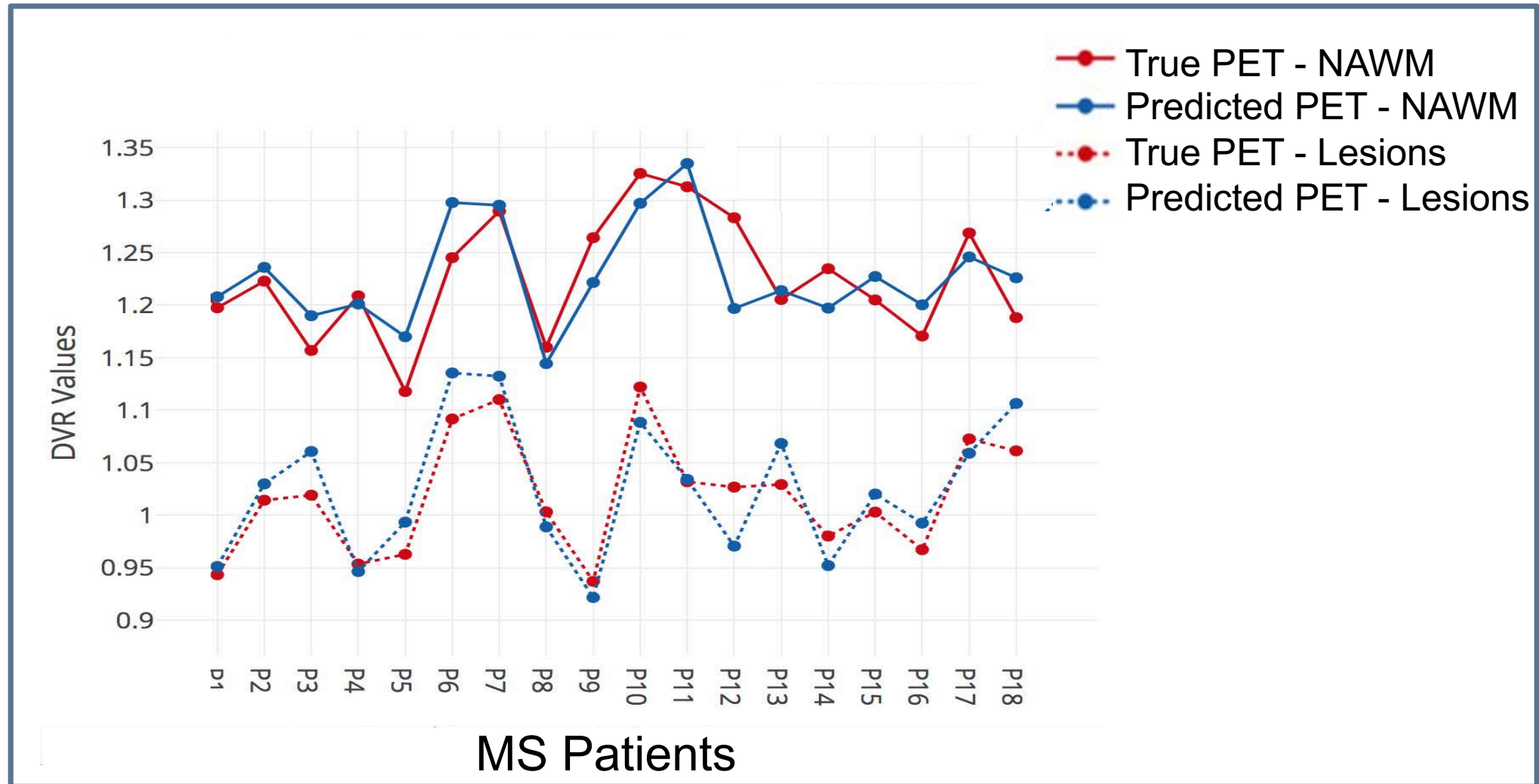
- **Problème**

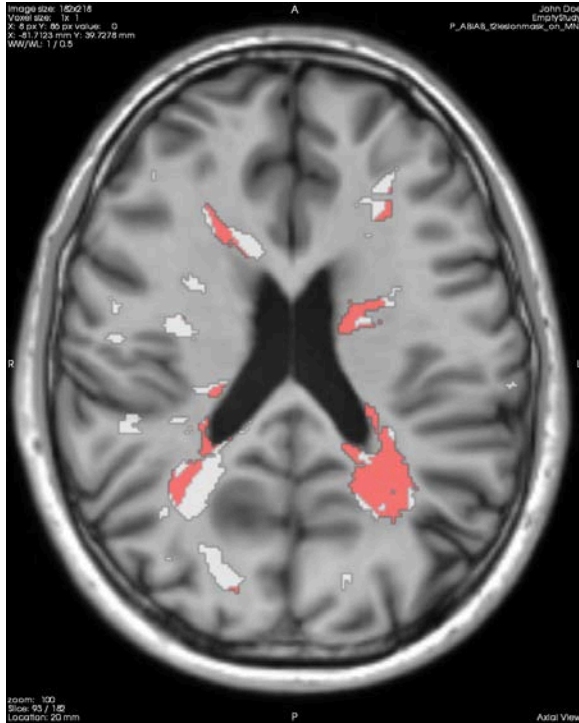
- Génère de façon adéquate l'information anatomique et les principales variations mais pas l'information dans les lésions

- **Approche proposée:
2 réseaux successifs**

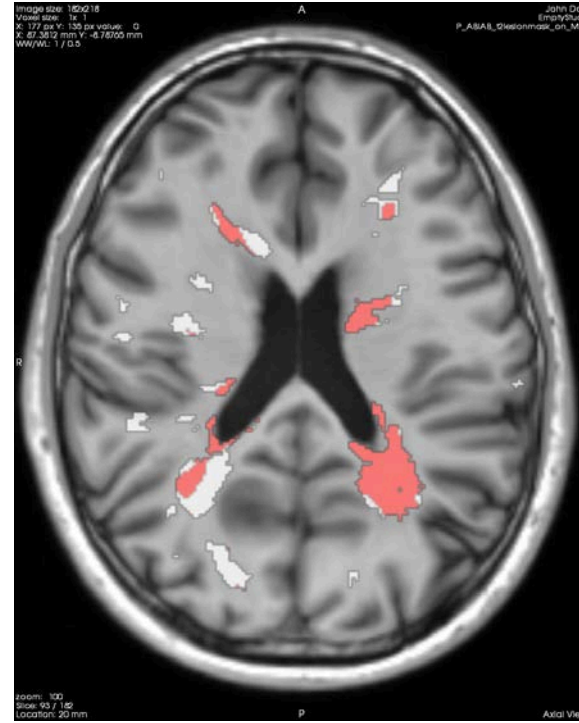
- 1^{er} réseau (ébauche) :
génère l'apparence globale
de l'image
- 2^{ème} réseau (affineur) :
fonction de coût spécifique,
focalisation sur les lésions







Carte de démyélinisation
issue du **vrai TEP**

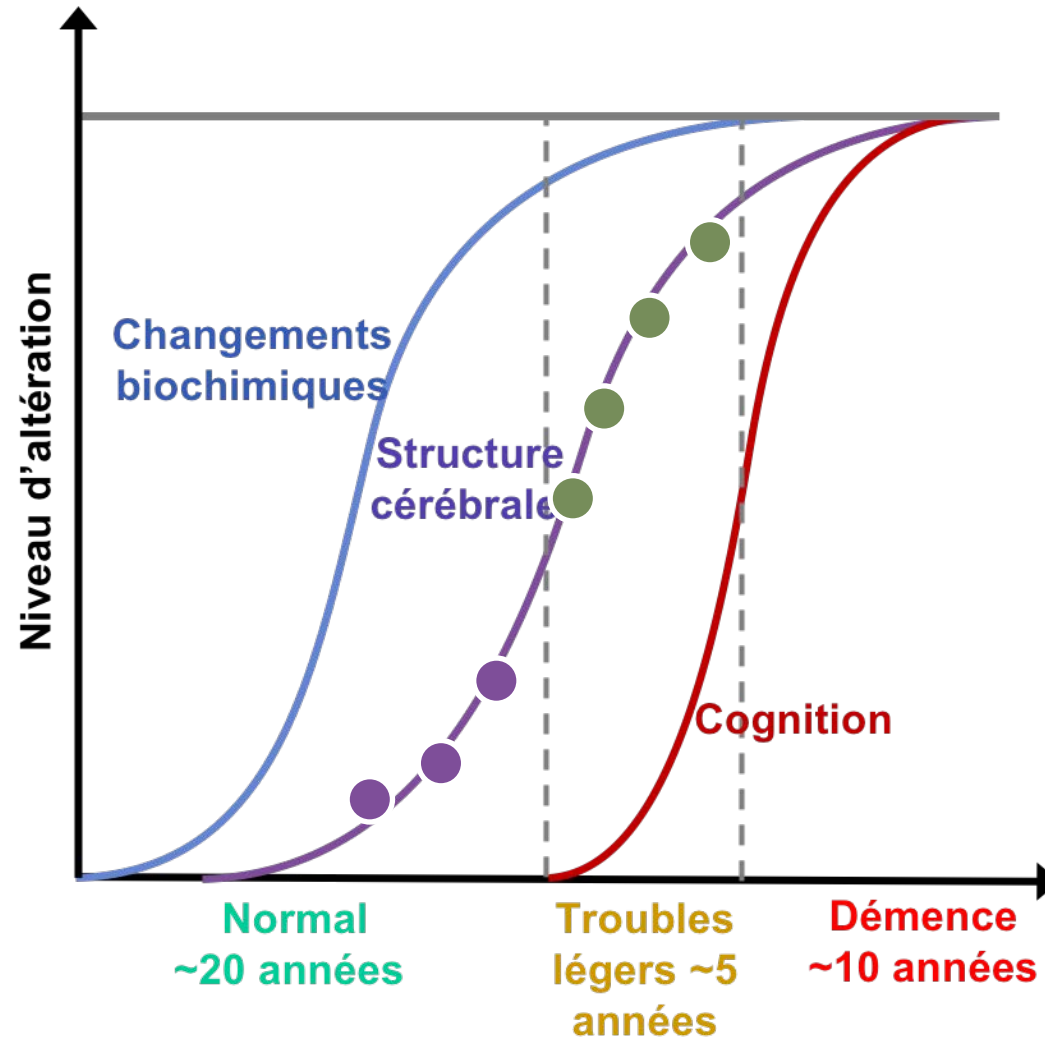


Carte de démyélinisation
issue du **TEP synthétique**

Recouvrement moyen
entre les cartes de
démyélinisation : **0.83**
(indice de Dice)

Modélisation de la dynamique temporelle

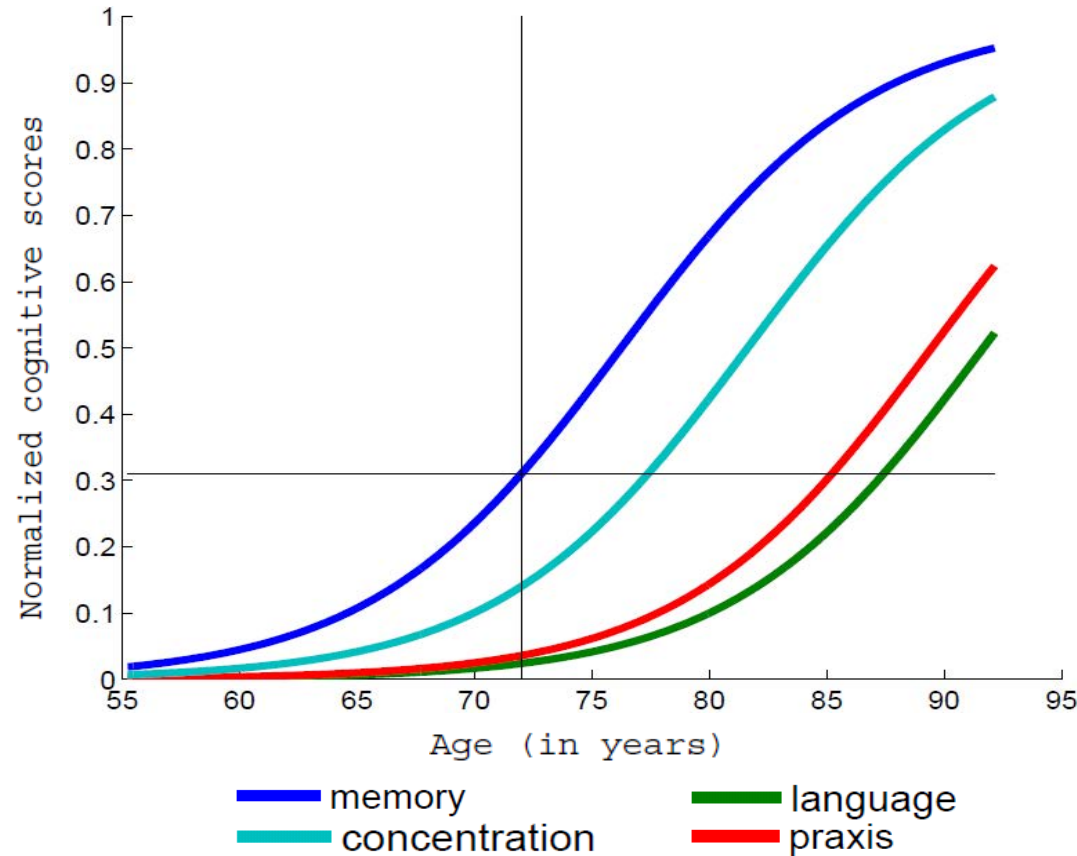
- Evolution sur plusieurs décennies, longue phase silencieuse asymptotique



A partir d'observations répétées
(intervalles typiques: 6-18 mois)

Modéliser la dynamique temporelle

- Application à des scores cognitifs

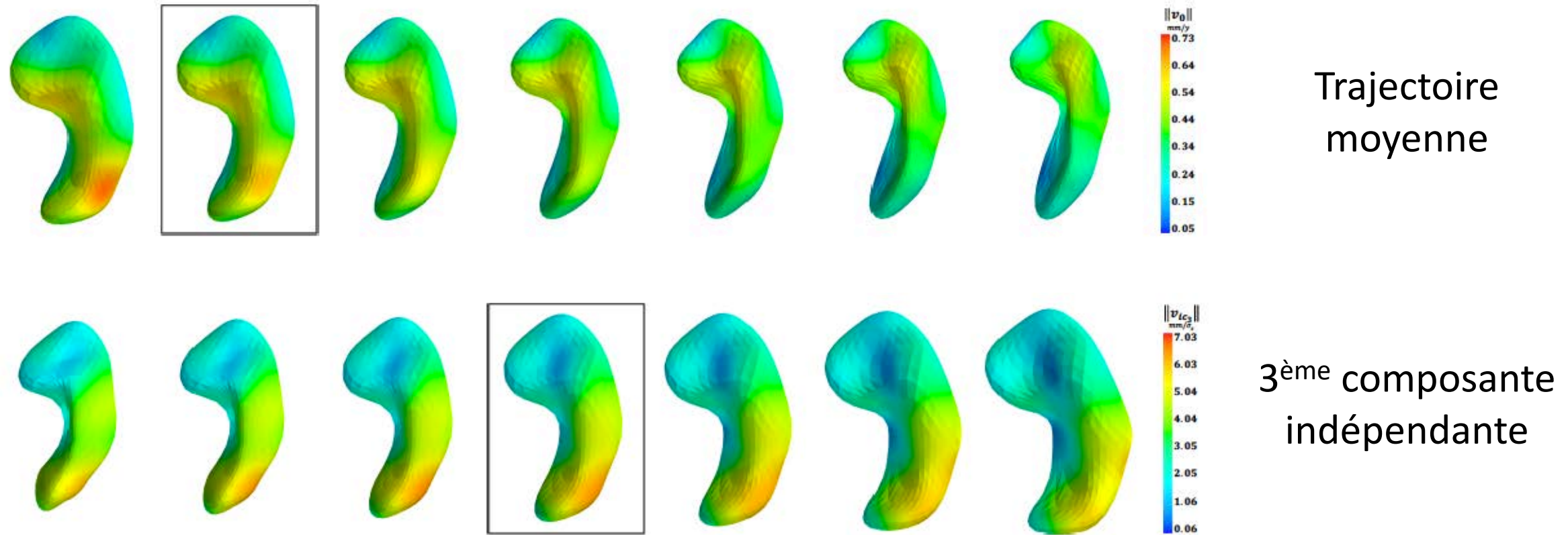


Tests cognitifs

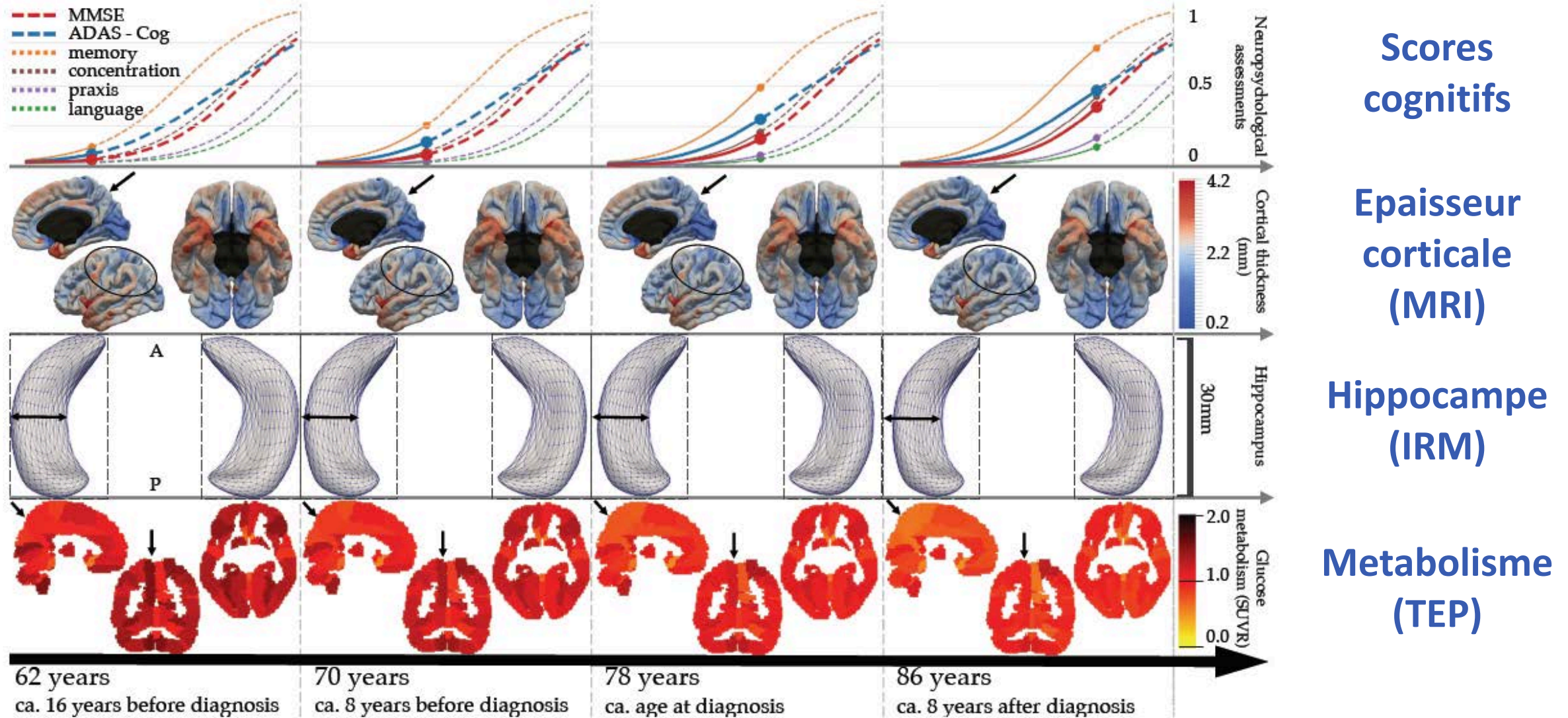
248 patients de la base ADNI

6 points temporels par patient (min 3, max 11)

- Application à des formes anatomiques



Bône et al, *CVPR*, 2018; *International Journal of Computer Vision* (Minor revision)



Aide à la décision à partir de données génétiques et d'imagerie

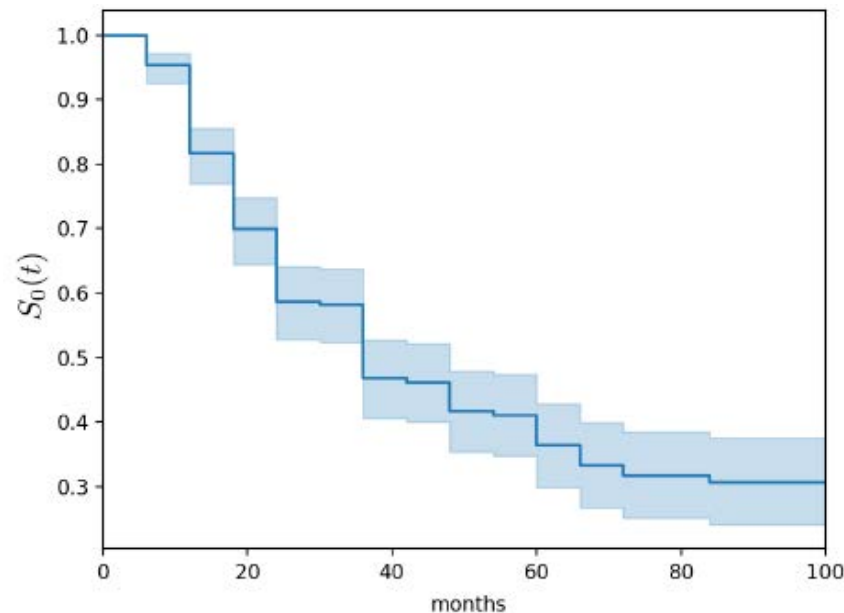
- **Des informations complémentaires**
 - **Génétique**
 - Module le risque et la vitesse de progression
 - Information statique
 - **Imagerie**
 - Fournit un instantané des lésions cérébrales
- **Les défis associés à leur combinaison**
 - Grande dimension ($\sim 10^6$ - 10^9)
 - Combinaison linéaire sous-optimale
 - Elimination de la modalité la plus faible
 - Néglige les interactions



- **Prédiction de la maladie d'Alzheimer**

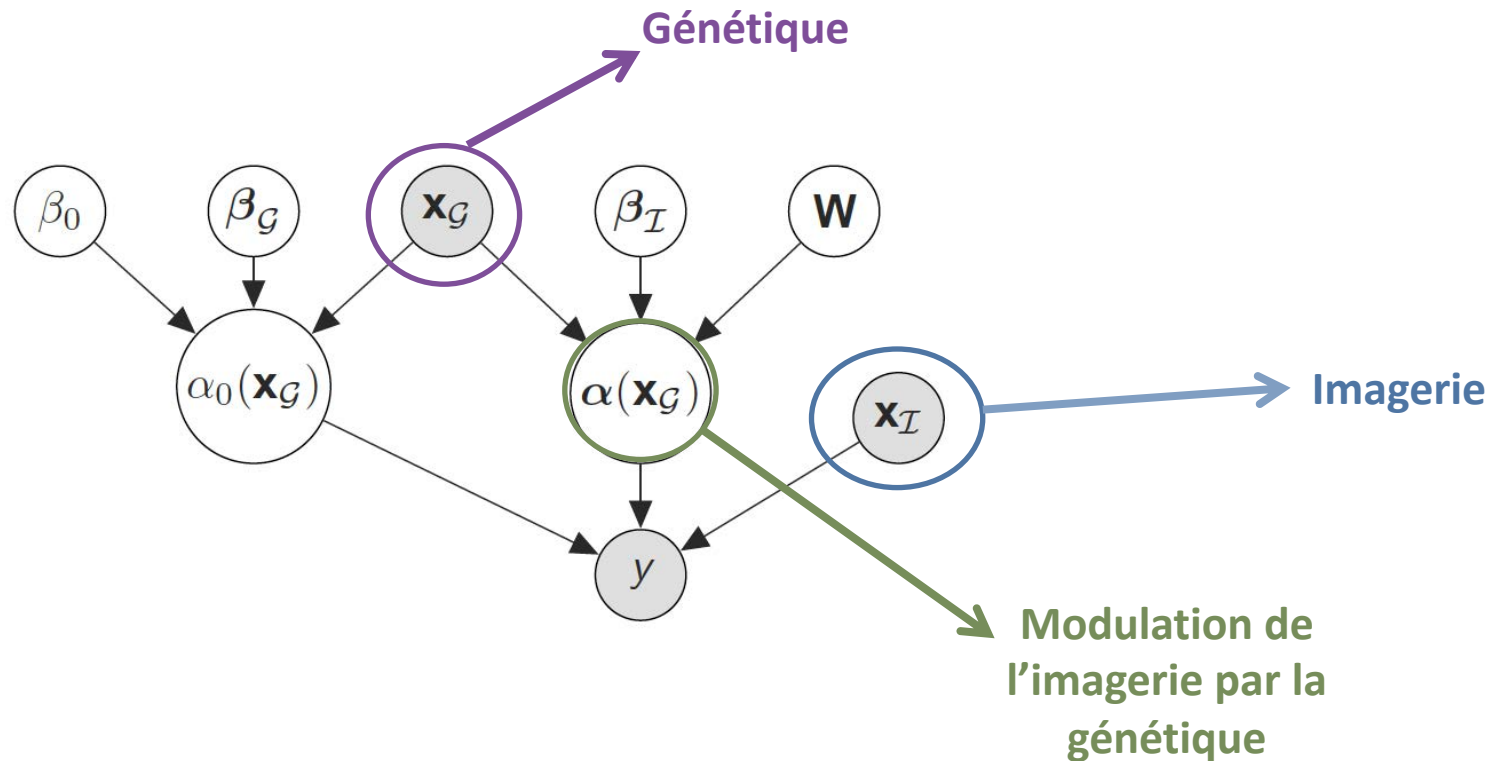
- Patients avec trouble cognitifs légers à l'inclusion
- Prédire à quelle date le patient risque de perdre son autonomie (démence)

- **Modèle de survie**



Proportion de patients sans démence en fonction du temps (estimateur de Kaplan-Meier)

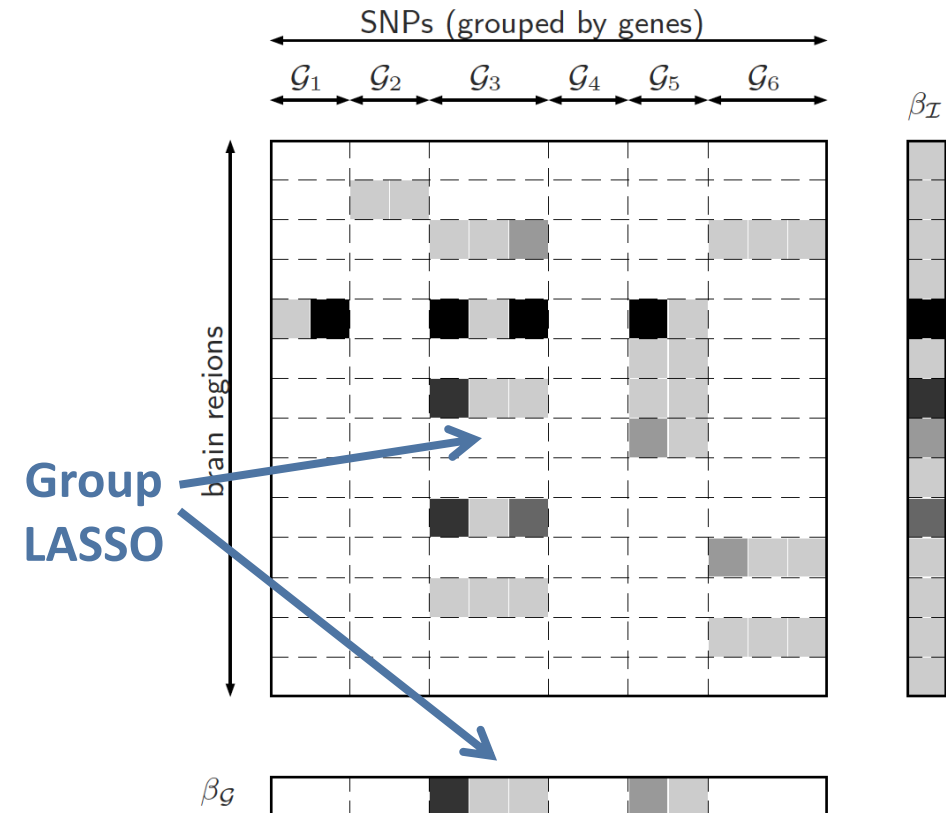
- Modèle multi-niveaux



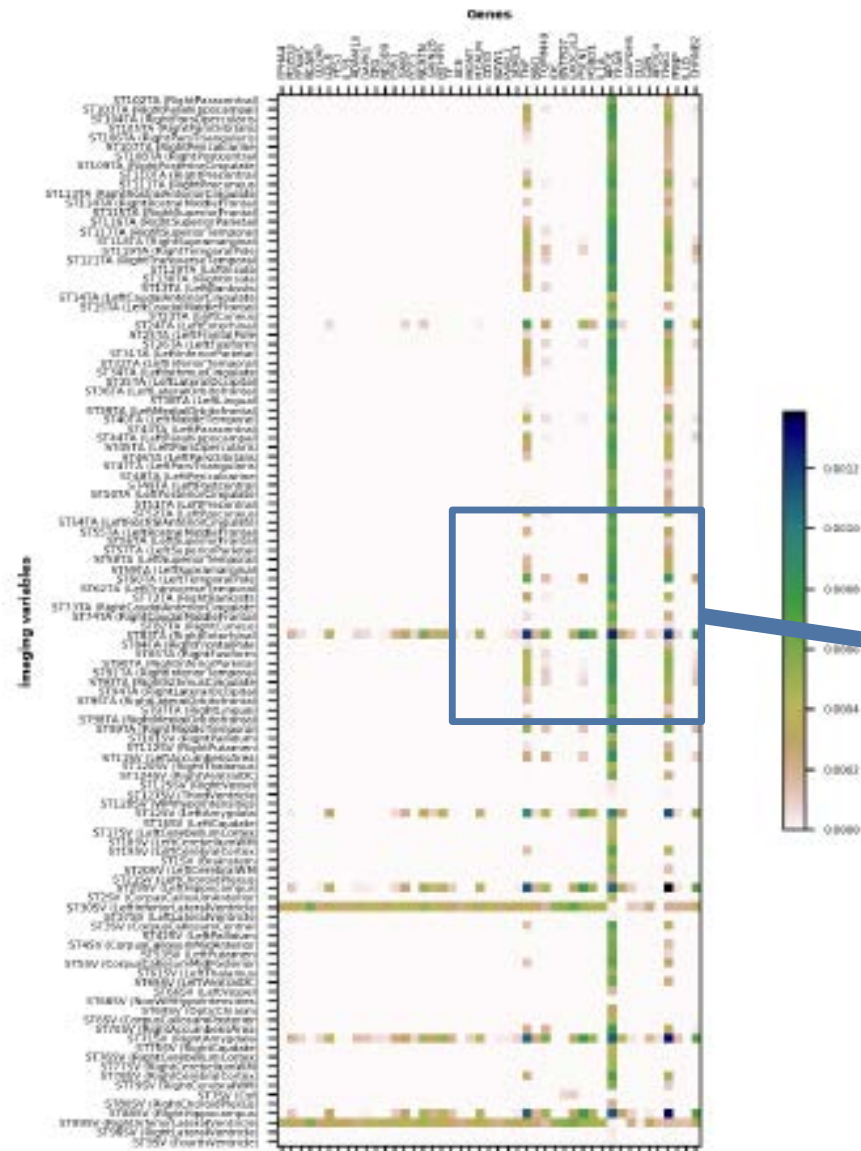
- Pénalités adaptées

- Group LASSO

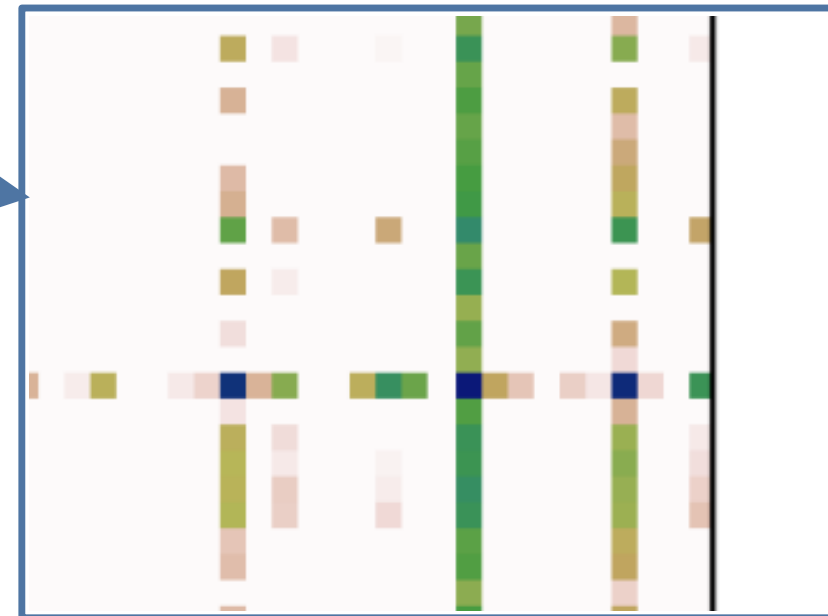
- Groupes: gènes x régions cérébrales



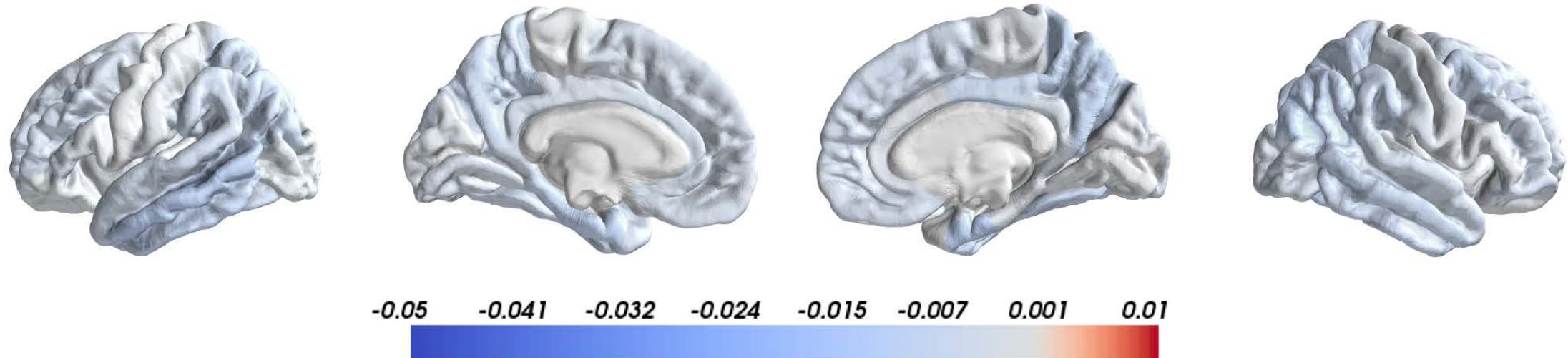
Modality	Method	C-index	Brier score	iAUC
SNPs only	Cox PH model (ℓ_1 penalty)	0.521 ± 0.040	0.166 ± 0.009	0.515 ± 0.031
MRI only	Cox PH model (no penalty)	0.636 ± 0.034	0.190 ± 0.017	0.636 ± 0.050
MRI only	Cox PH model (ℓ_2 penalty)	0.671 ± 0.022	0.149 ± 0.008	0.663 ± 0.044
All	Additive Cox PH model (ℓ_1 penalty)	0.677 ± 0.020	0.148 ± 0.006	0.680 ± 0.030
All	Multilevel model (ours)	0.681 ± 0.018	0.147 ± 0.006	0.686 ± 0.031

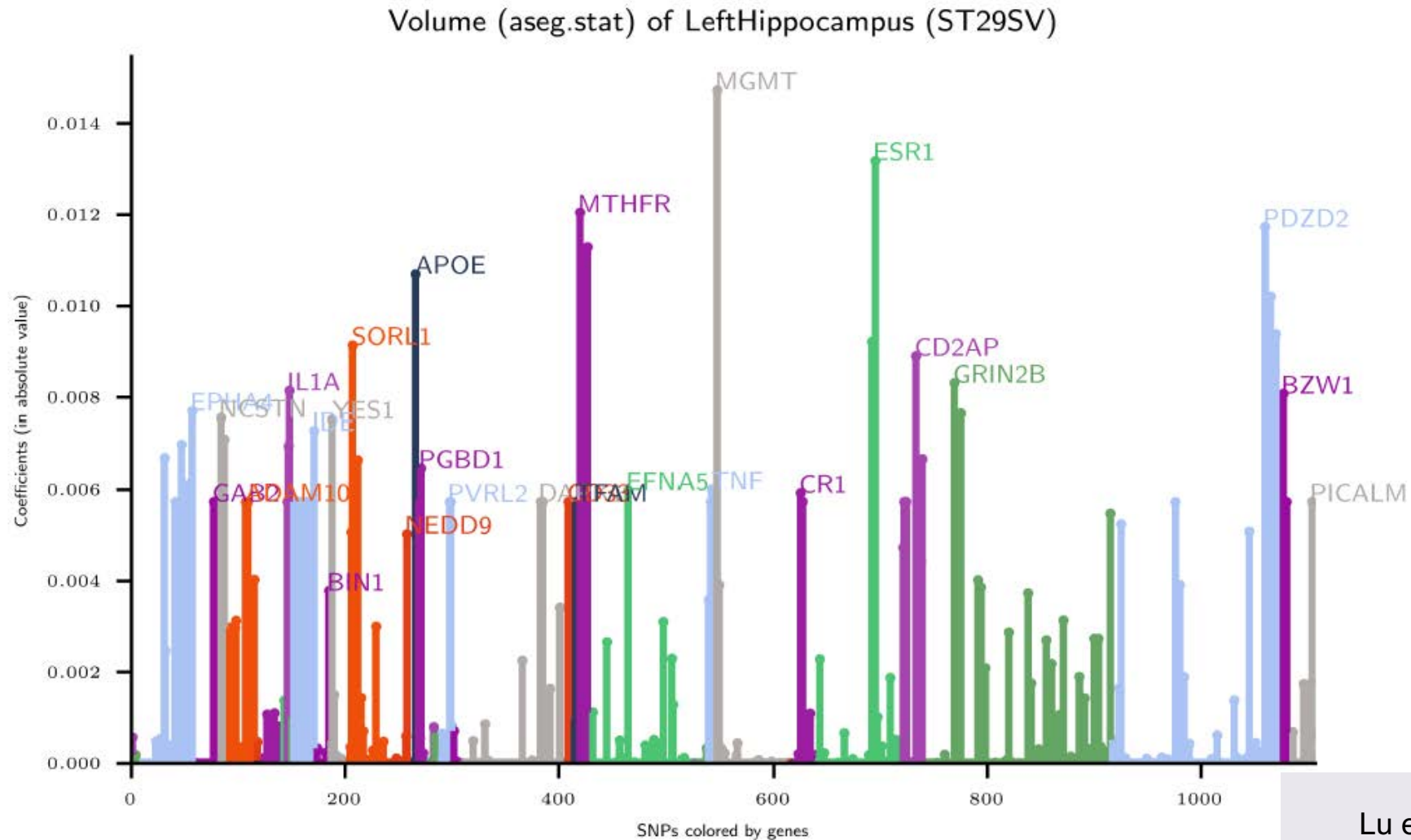


Matrice d'interactions entre gènes et régions cérébrales



Column "rs429358" (APOE) of W





Quels défis devant nous ?

- **Grande dimension reste un problème difficile**
 - Particulièrement dans le cas de la multimodalité
 - **Données faiblement annotées**
 - La plupart des annotations sont en texte libre
 - **Interprétabilité des modèles**
 - **Généralisation des modèles à des données de routine cliniques ?**
 - Qualité non contrôlée
 - Hétérogénéité des protocoles d'acquisition et des pratiques
-



**ARAMIS
LAB**
BRAIN DATA SCIENCE

Web: www.aramislab.fr

Twitter: [@AramisLabParis](https://twitter.com/AramisLabParis)

