

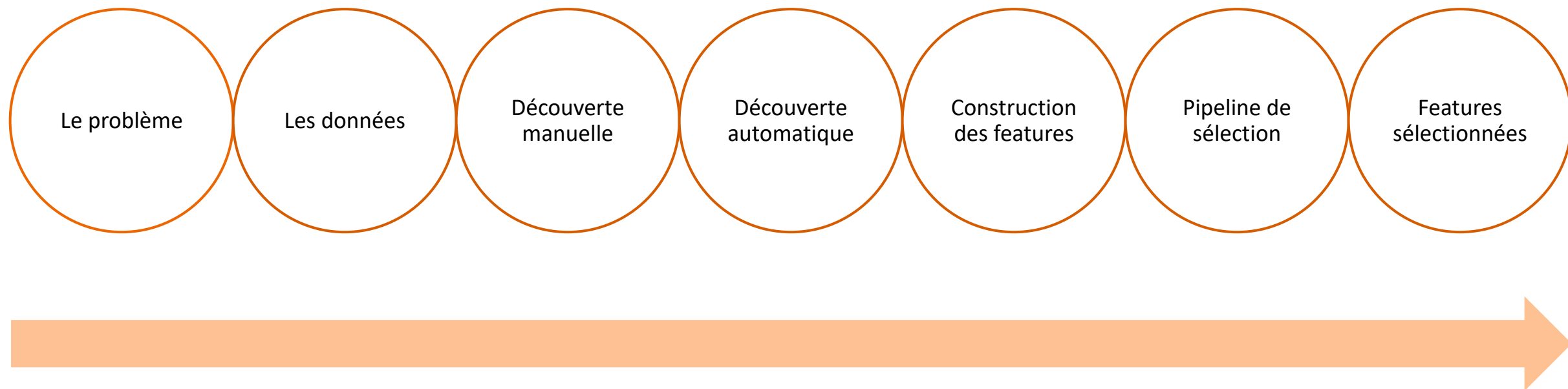
# Biomarqueurs corps entier en imagerie TEP/TDM pour les patients atteints de lymphome

Louis Rebaud

23/05/2023

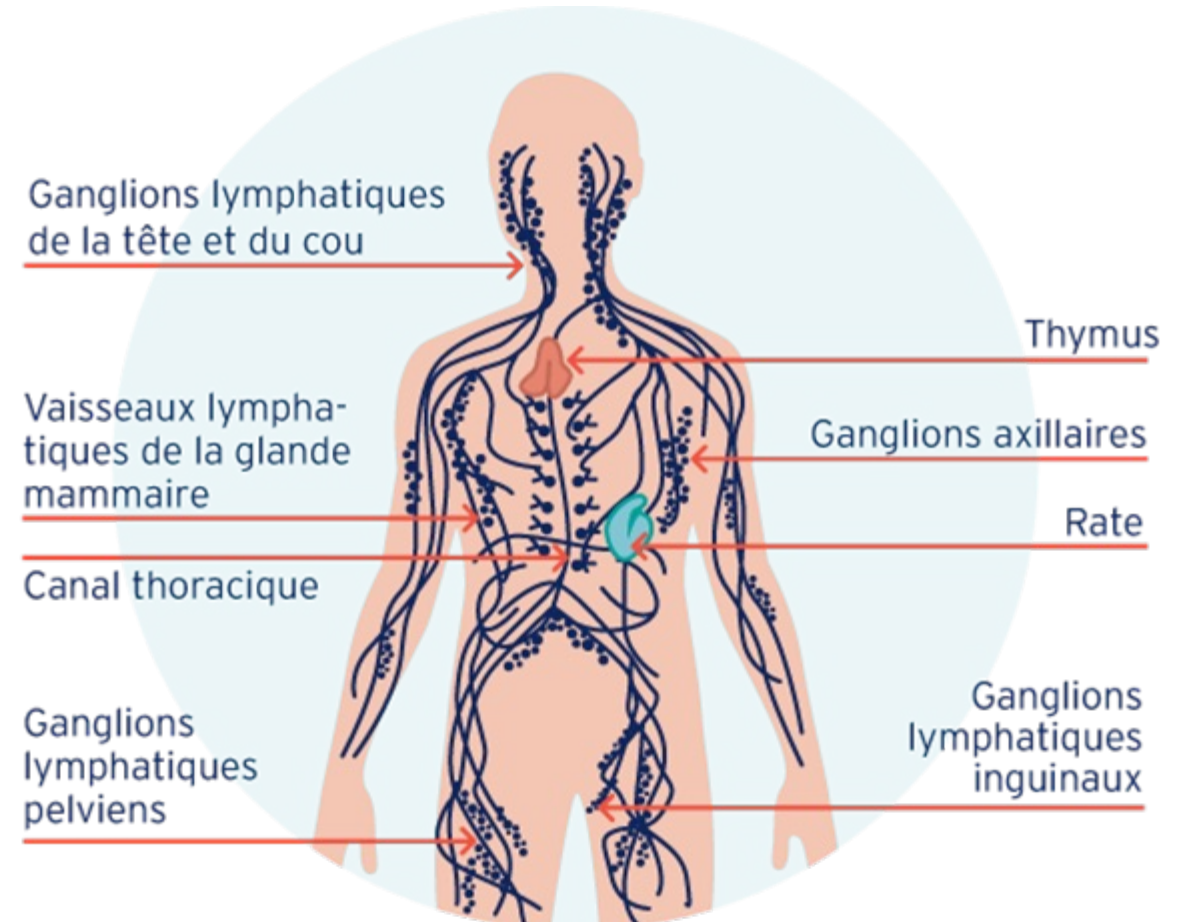


# Sommaire



# Les lymphomes

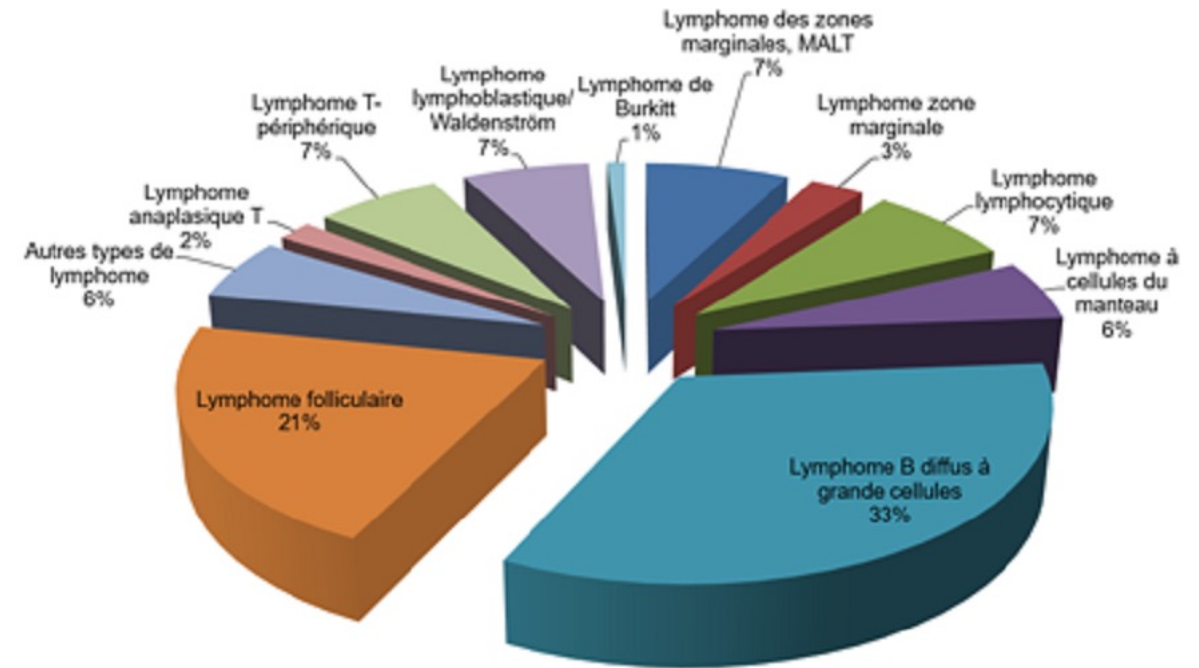
- Cancer du système lymphatique
- Symptômes:
  - ganglions lymphatiques enflés
  - fièvre
  - perte de poids involontaire
  - fatigue
  - sueurs nocturnes
- Traitement: combinaison de chimiothérapie, de radiothérapie et d'immunothérapie
- Multiples lésions dans différentes localisations



# Les lymphomes

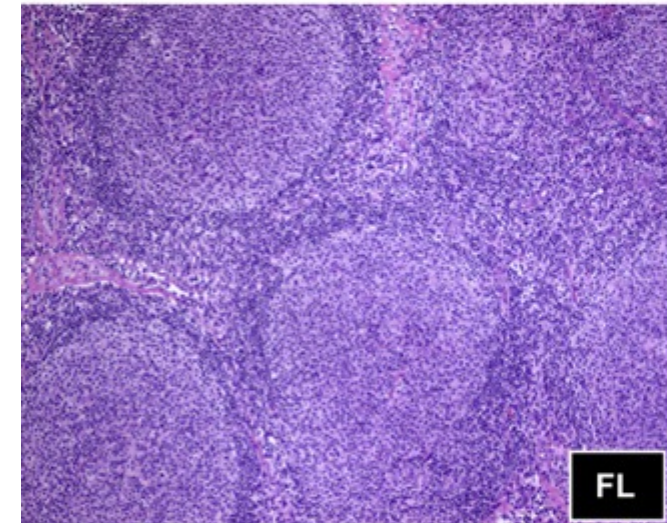
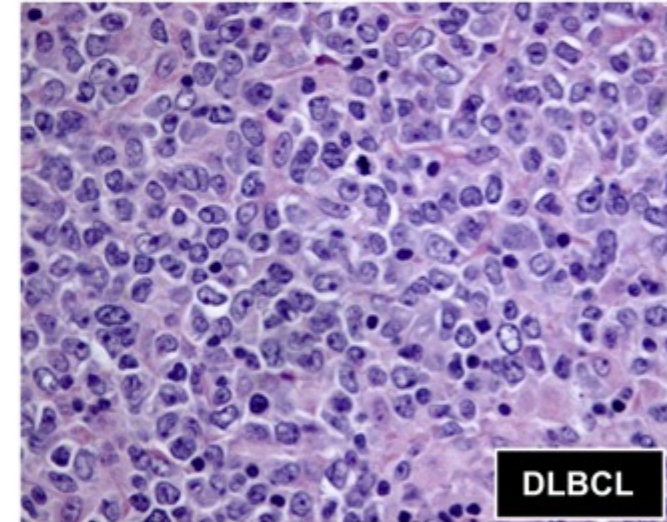
- 2 types de lymphomes :
  - Lymphome de **Hodgkin** :
    - relativement rare
    - un des cancers que l'on soigne le mieux (95% de guérisons)
  - Lymphomes **Non Hodgkiniens** :
    - 5e cancer en termes de fréquence (11 000 nouveaux/an en France)
    - Plus de 27 sous-types

Répartition des différents types de lymphomes



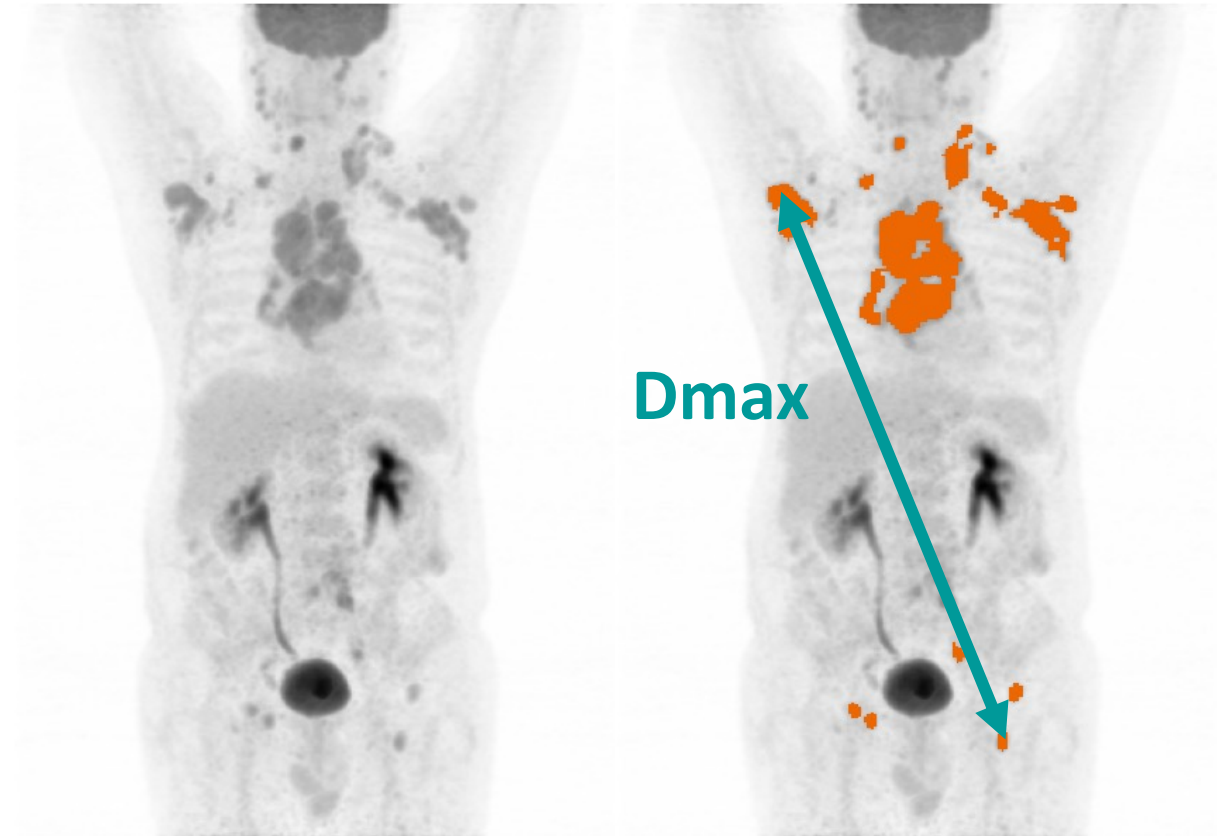
# Les lymphomes

- Lymphome B diffus à grandes cellules (DLBCL)
  - Agressif
  - 65% de survie à 5 ans
  - Lymphomes Non Hodgkiniens le plus courant
- Lymphome folliculaire (FL)
  - Indolent
  - 90% de survie à 5 ans, 50% à 20 ans
  - Guérison complète rare
  - Peut muter en DLBCL



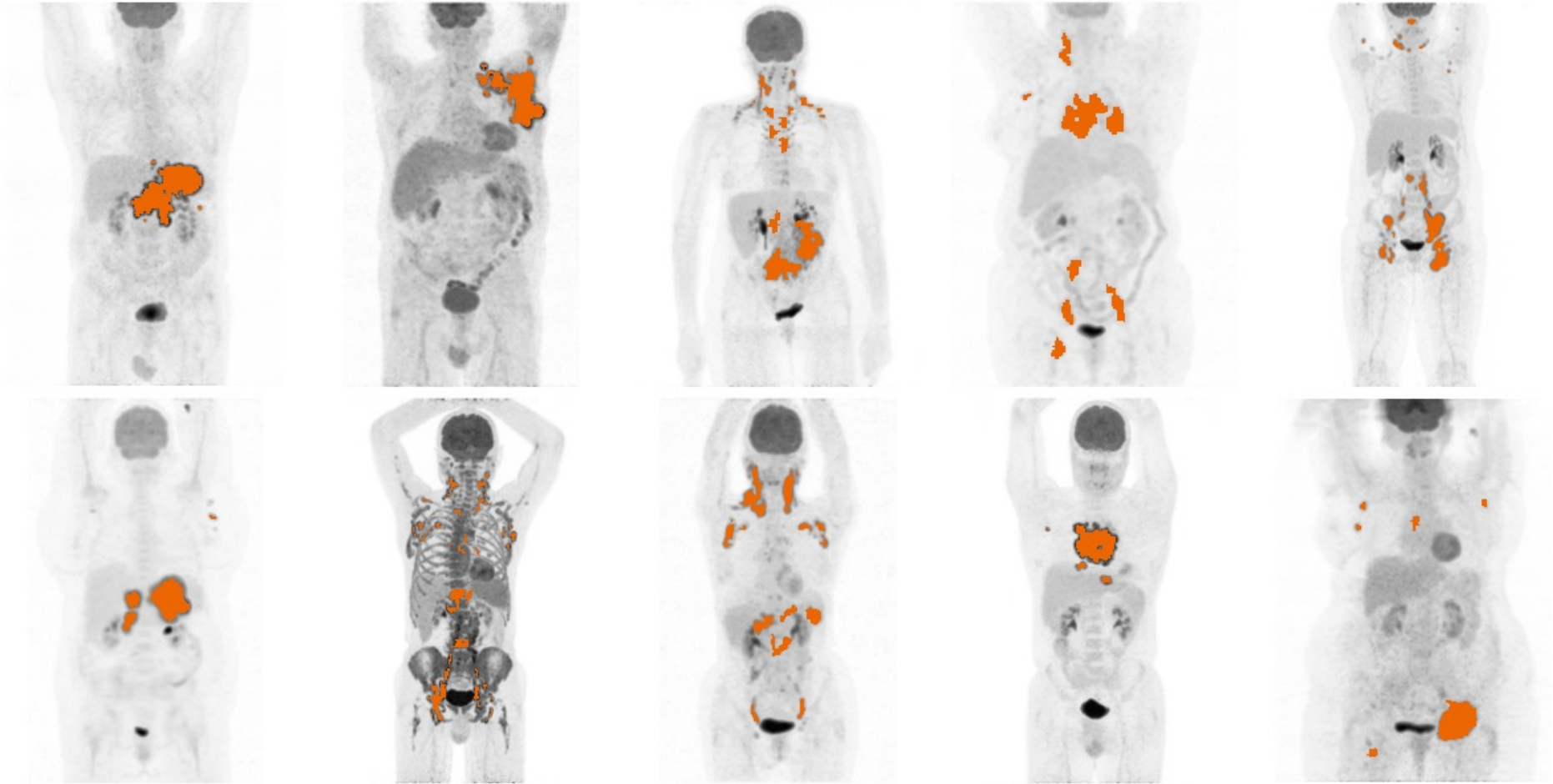
# Les lymphomes

- TEP 18F-FDG/TDM
  - Diagnostic
  - Détection tumeurs
  - Evaluation de la réponse au traitement
- Volume métabolique tumoral total (TMTV): seul facteur pronostique mesuré dans l'image utilisé en clinique
- Le signal dans les régions saines n'est pas interprété.

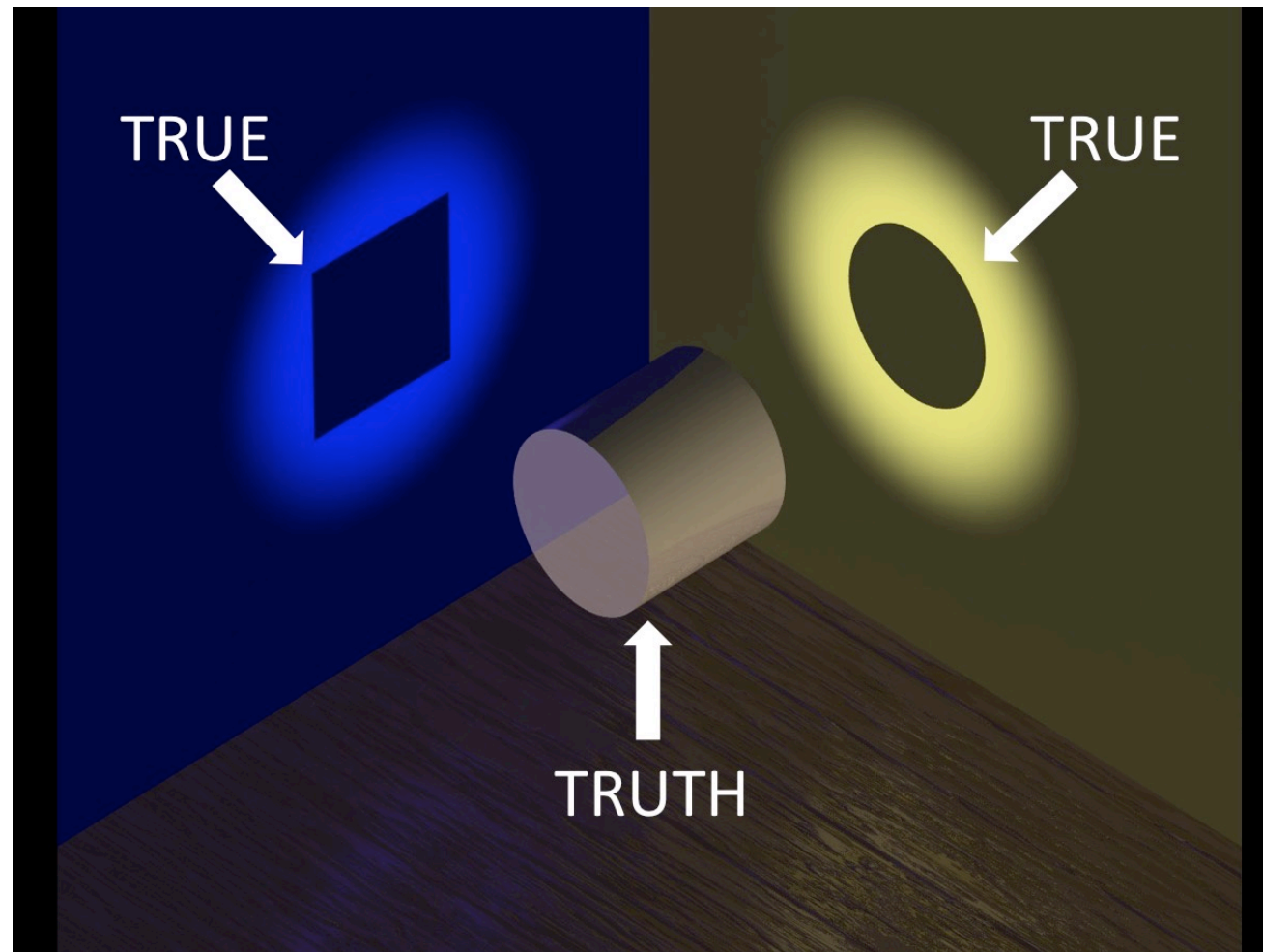




Le problème

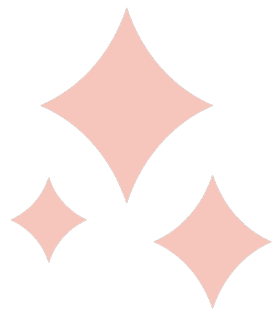


## Le problème

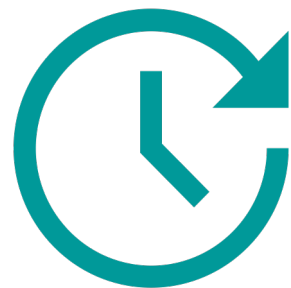




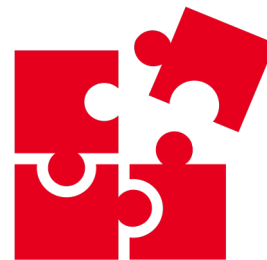
## Le problème – Propriétés souhaitées



Nouveaux



Prédictifs



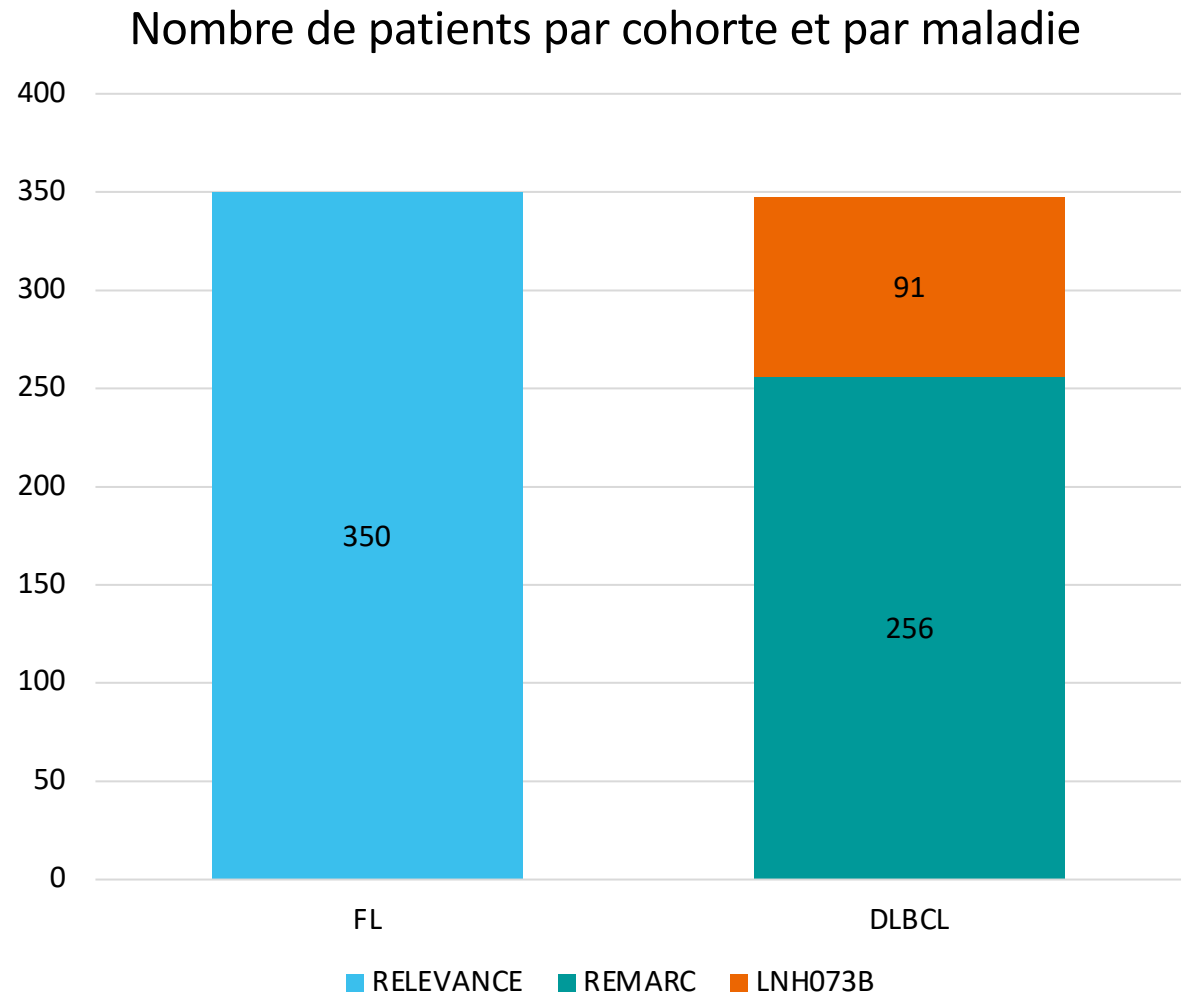
Complètent ce qu'on  
sait déjà



(si possible)  
explicables

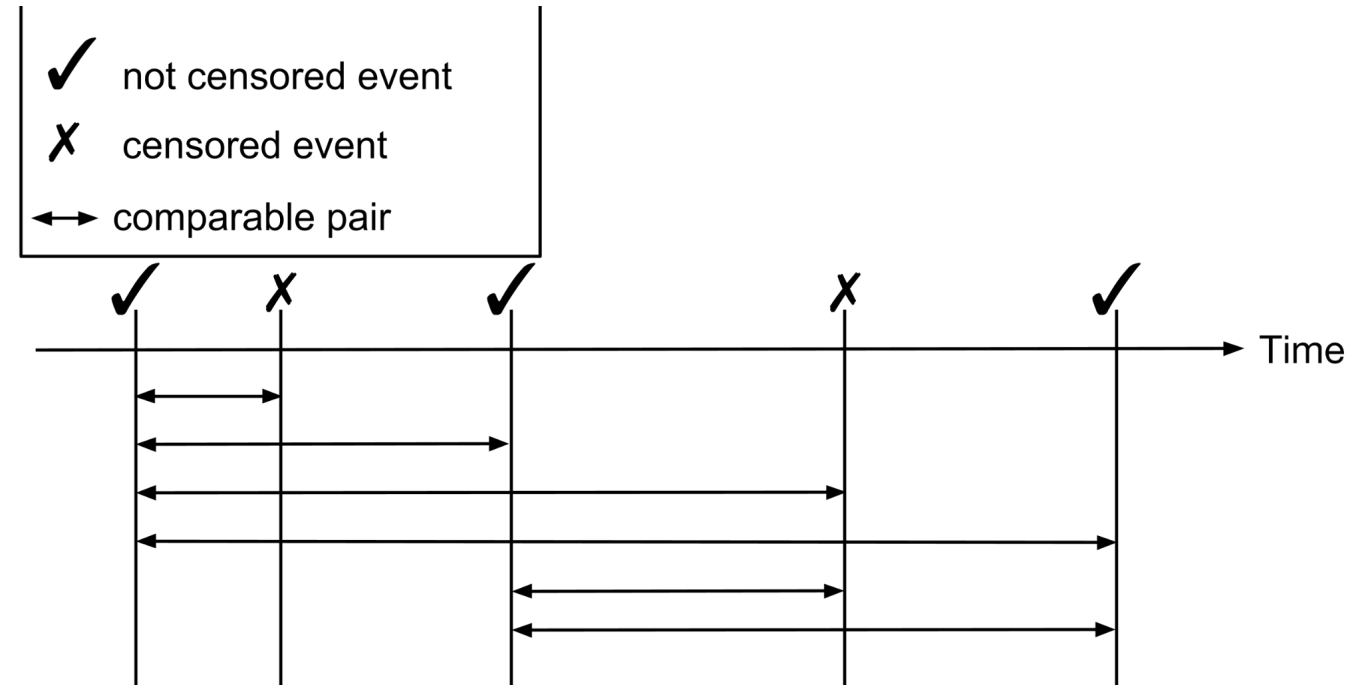
# Les données

- 2 cohortes :
  - DLBCL (REMARC + LNH073B)
  - FL (RELEVANCE)
- TEP 18F-FDG/TDM baseline pour chaque patient
- Survie sans progression (PFS) et survie globale (OS) disponibles



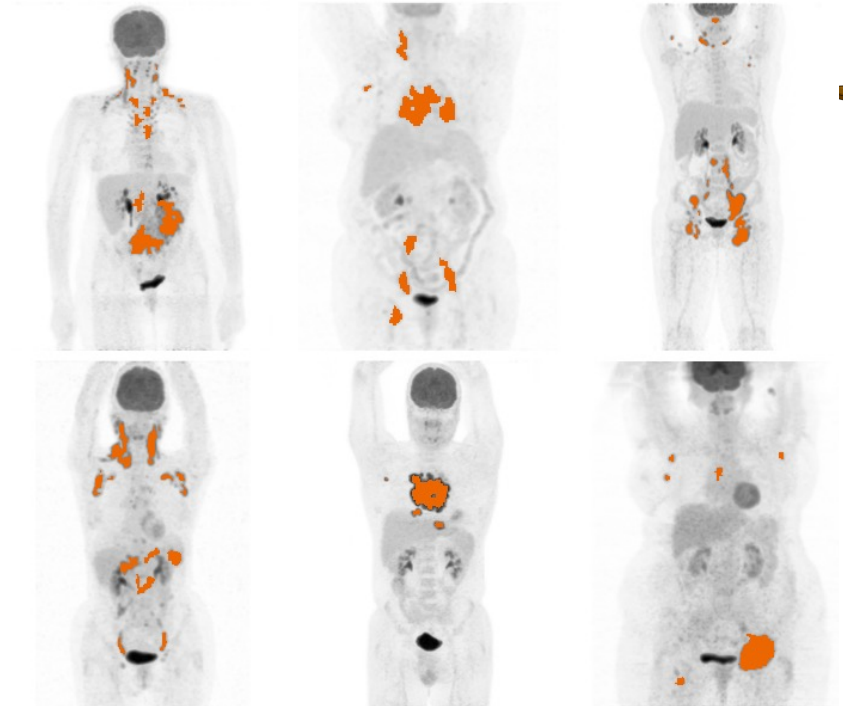
# Concordance index (C-index)

- Proportion de paires comparables correctement ordonnées
- 0,5 : prédictions aléatoires
- 1 : prédictions parfaitement concordantes
- -1 : prédictions parfaitement anti-concordantes



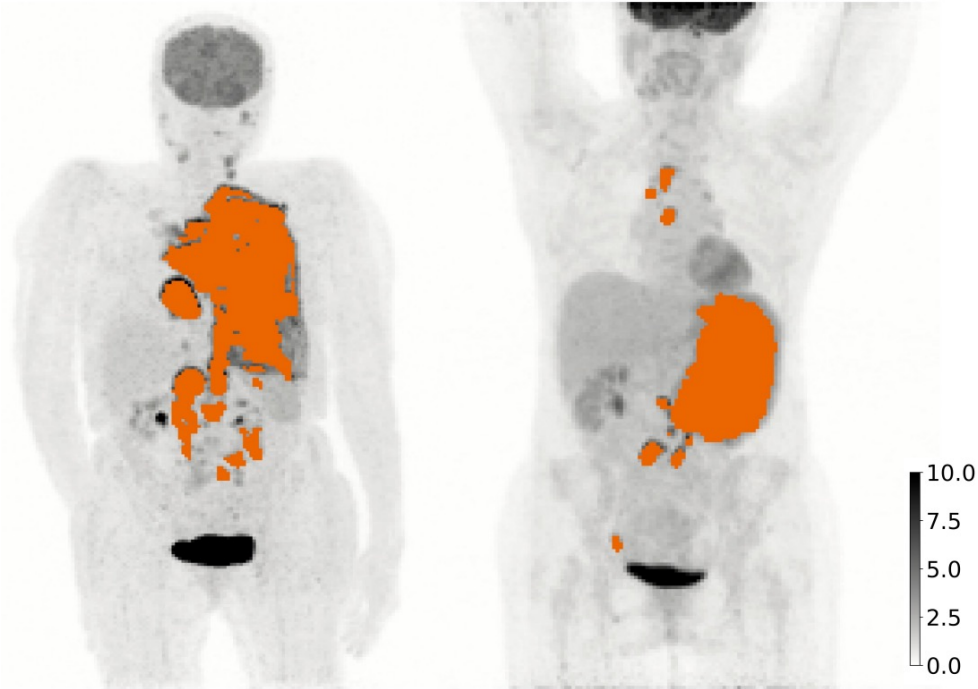
# Recherche manuelle

- Regarder les images et ce qui différencie les patients
- Construire manuellement un biomarqueur qui quantifie une nouvelle information
- Tester sa prédictivité



# Recherche manuelle - Résultats

- Seul résultat : fragmentation tumorale (TF)
- Plus la masse tumorale est fragmentée, meilleur est le pronostic
- Redécouverte (trouvée par Decazes et al.)
- Performances faibles par rapport au TMTV et Dmax



	Univariate mean time dependent ROC-AUC	
	PFS	OS
TMTV	0.70	0.70
SDmax	0.69	0.66
TF	0.58	0.57

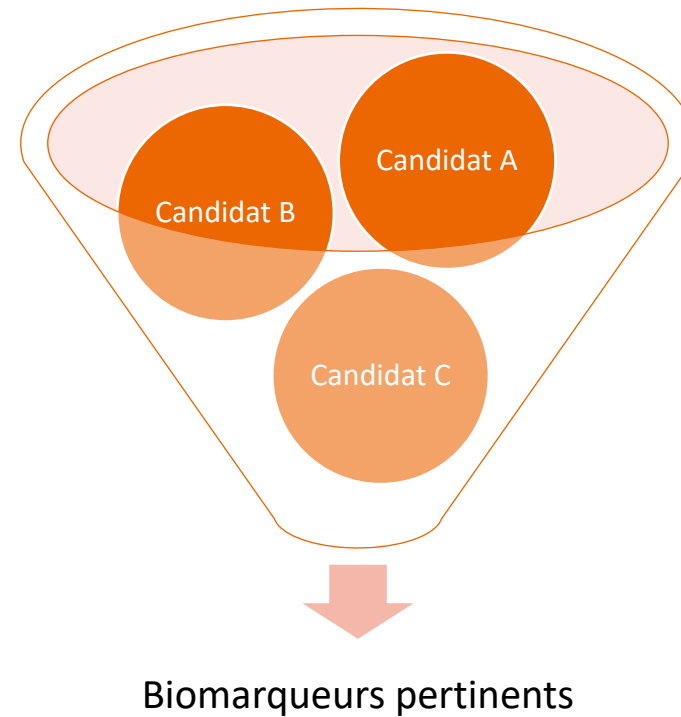
## Recherche manuelle - Problèmes

- Lent
- Peu exhaustif
- Sujette aux faux positifs
- Biaisé



# Recherche semi-automatique

- Construire automatiquement plusieurs biomarqueurs candidats
- Les tester un-à-un
- Ne garder que ceux qui ont les propriétés désirées
- Interpréter/simplifier les biomarqueurs sélectionnés



# Avantages

Manuel	Semi-automatique
Lent	Rapide (milliers de candidats testés)
Peu exhaustif	Exhaustif (nombreuses mesures dans plusieurs endroits du corps)
Sujette aux faux positifs	Nombre d'expériences prédéfini -> taux de fausses découvertes contrôlé
Biaisé	Moins biaisé

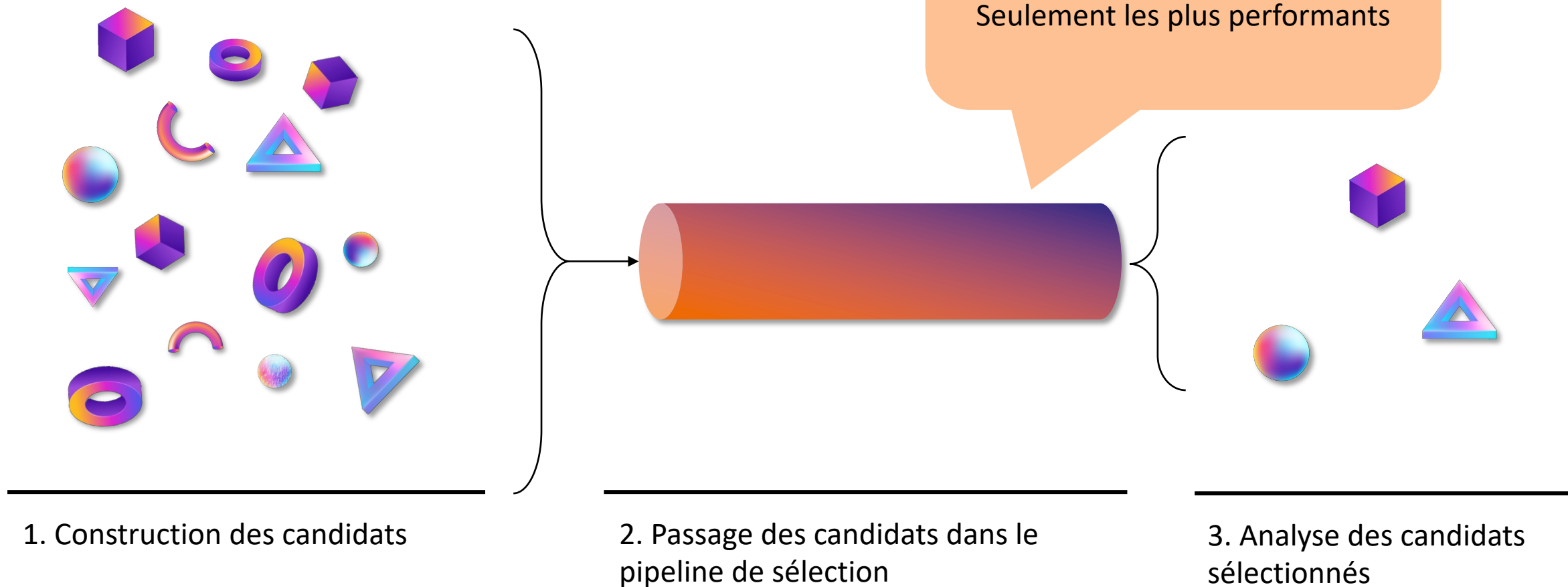
Recherche manuelle



Recherche semi-automatique

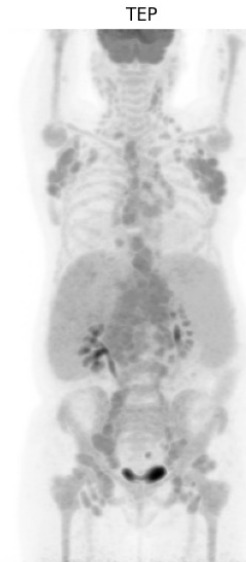


# Processus de la recherche semi-automatique



# Construction des candidats

- Candidats calculés sur la TEP et le scanner
- ROIs :
  - Lésions
  - Couronnes de tissus autour des lésions (shell)
  - Organes
- Organes segmentés par TotalSegmentator
- Features radiomiques extraites avec PyRadiomic pour chaque ROI dans chaque modalité
- Plusieurs agrégations à l'échelle du patient (min, max, median, mean, std)
- 7436 candidats créés



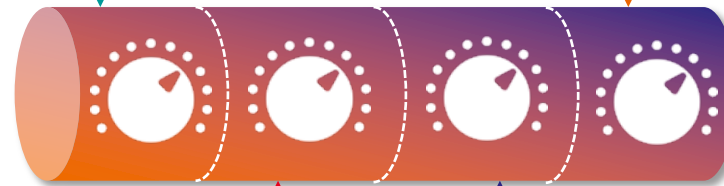
# Liste des tests

## Réduction du nombre de candidats

- Enlève les candidats trop corrélés à d'autres candidats

## Complète ce que l'on sait déjà

- Test multivarié sur la **PFS** avec des biomarqueurs déjà connus
- Test multivarié sur l'**OS** avec des biomarqueurs déjà connus



## Nouveaux

- Enlève les candidats trop corrélés à des biomarqueurs déjà connus

## Prédictifs

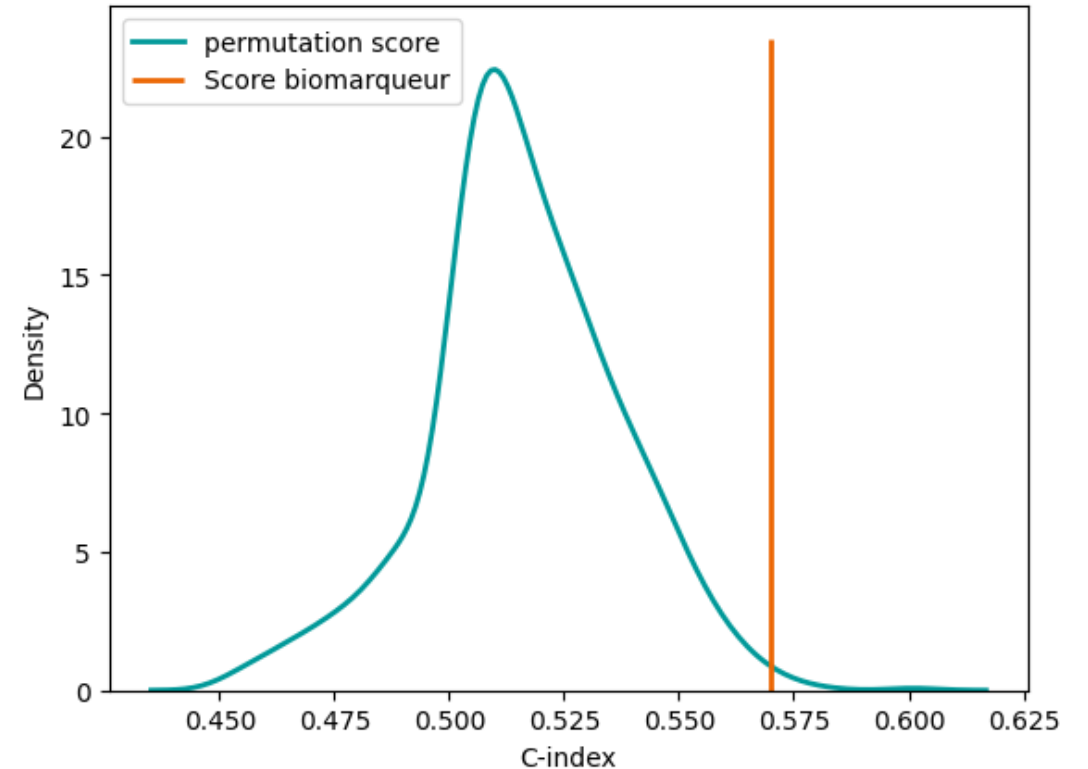
- Test univarié sur la **PFS**
- Test univarié sur l'**OS**



# Tests univariés

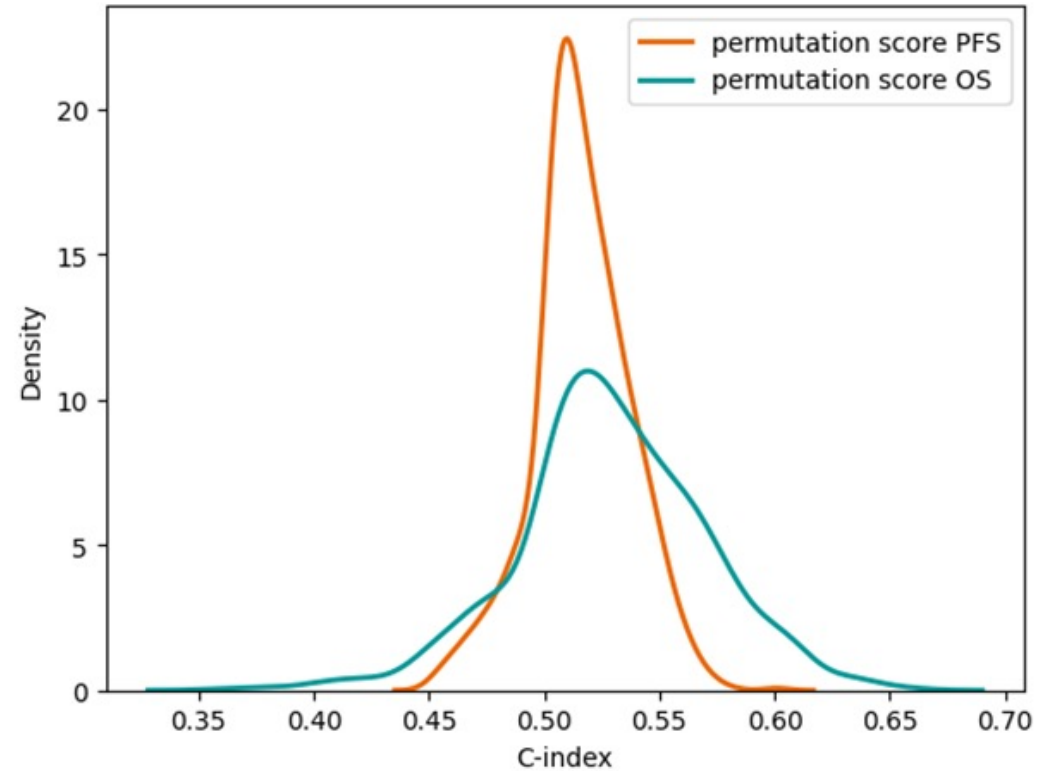
- Tests réalisés avec des tests de permutations
- On génère N features aléatoires
- On mesure le C-index de chaque feature aléatoire
- On mesure le C-index de la feature candidate
- La p-valeur est définie par:

$$pval = \frac{nb\ de\ C - index\ aléatoires\ >\ C - index\ candidat}{N}$$



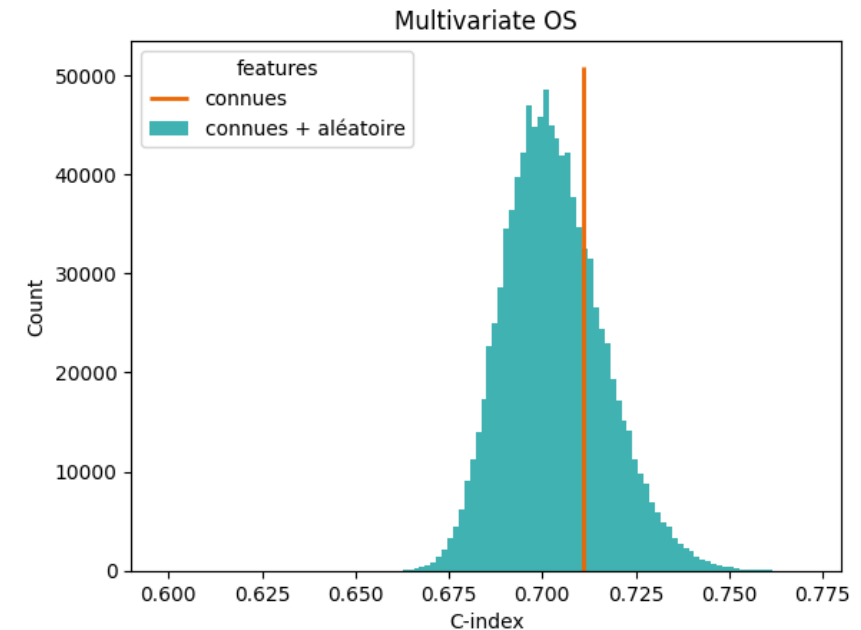
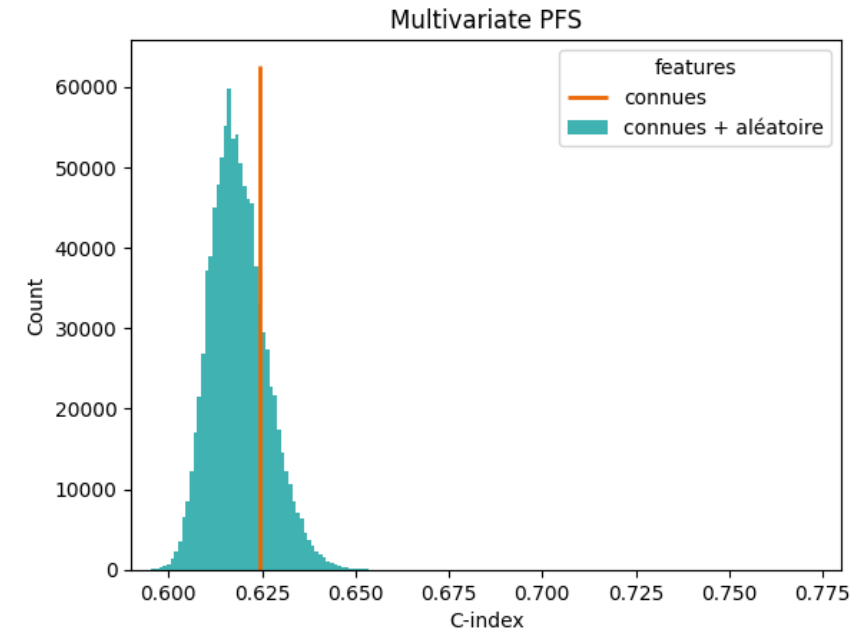
# Tests univariés

- Les C-index aléatoires varient selon la cohorte et selon la cible
- Un même seuil sur le C-index ne donnera pas le même taux de fausses découvertes sur deux jeux de données différents
- Le test de permutation permet de gommer cet effet



# Tests multivariés

- On ne sélectionne que les biomarqueurs qui améliorent un modèle multivarié utilisant les features déjà connues :
- $\text{Model( TMTV, IPI )} = 0,62$
- $\text{Model( TMTV, IPI, biomA )} = 0,65$
- $\text{Model( TMTV, IPI, biomB )} = 0,57$
- On sélectionnera le biomarqueur A dans ce cas là.
- On peut également utiliser un test de permutation



# Tests multivariés

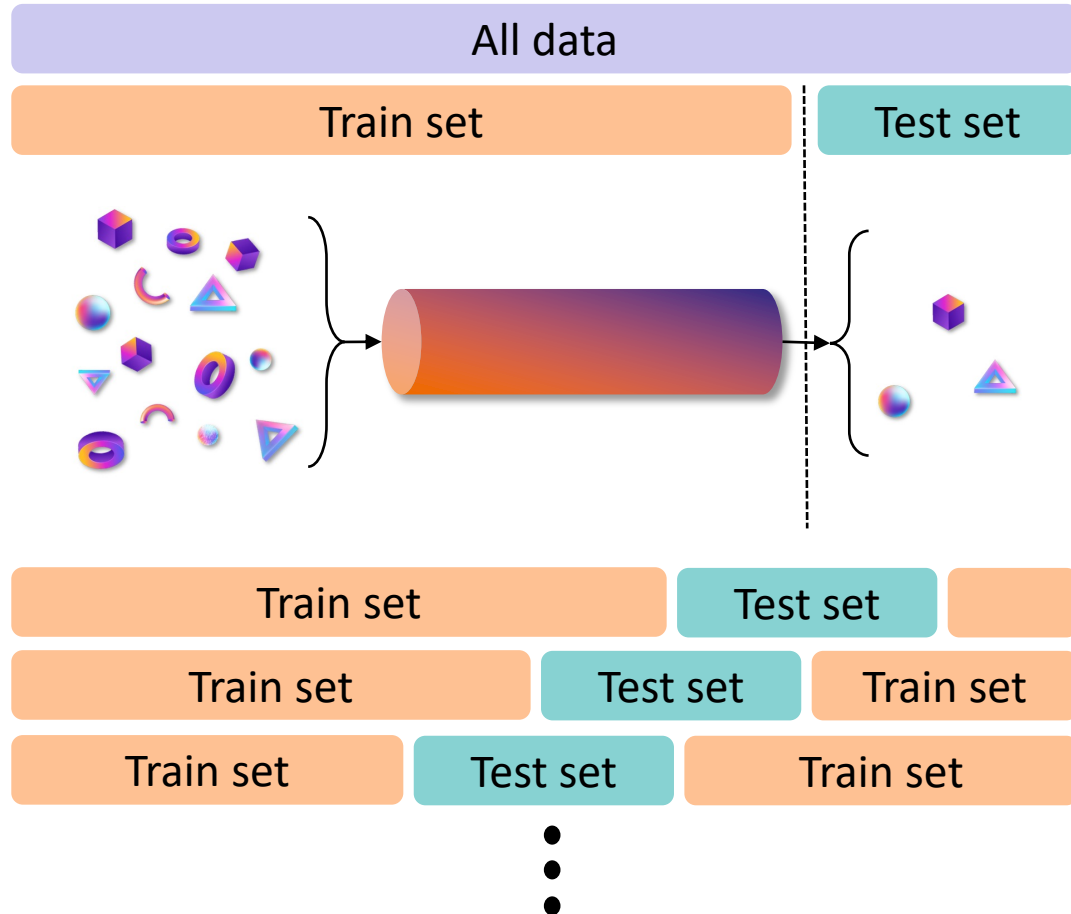
- ICARE
  - Simple
  - Robuste
  - Parallélisable
- Leave-Pair out Cross Validation (LPOCV)
  - Plus précis
  - Plus reproductible

$$c_i = C_{index}(x_i, y) \quad s_i = \begin{cases} +1, & \text{if } c_i \geq 0.5 \\ -1, & \text{otherwise} \end{cases}$$
$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad \hat{y} = \frac{1}{M} \sum_i^M s_i \times z_i$$



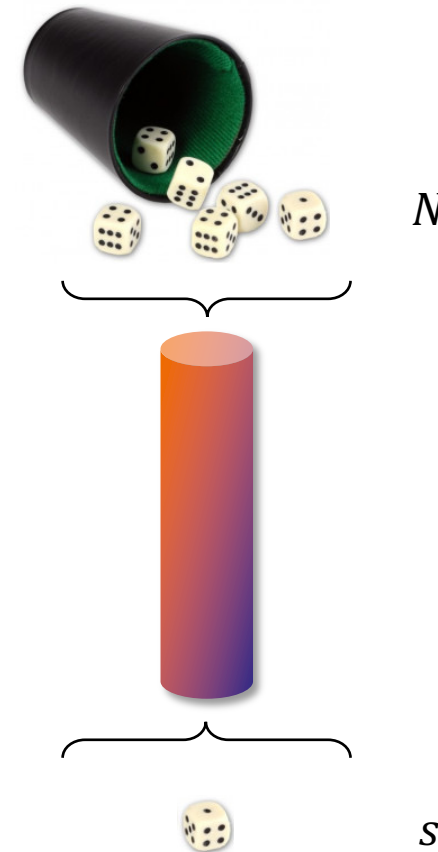
# Validation et ajustement du pipeline

Leave-Pair Out Cross-Validation :



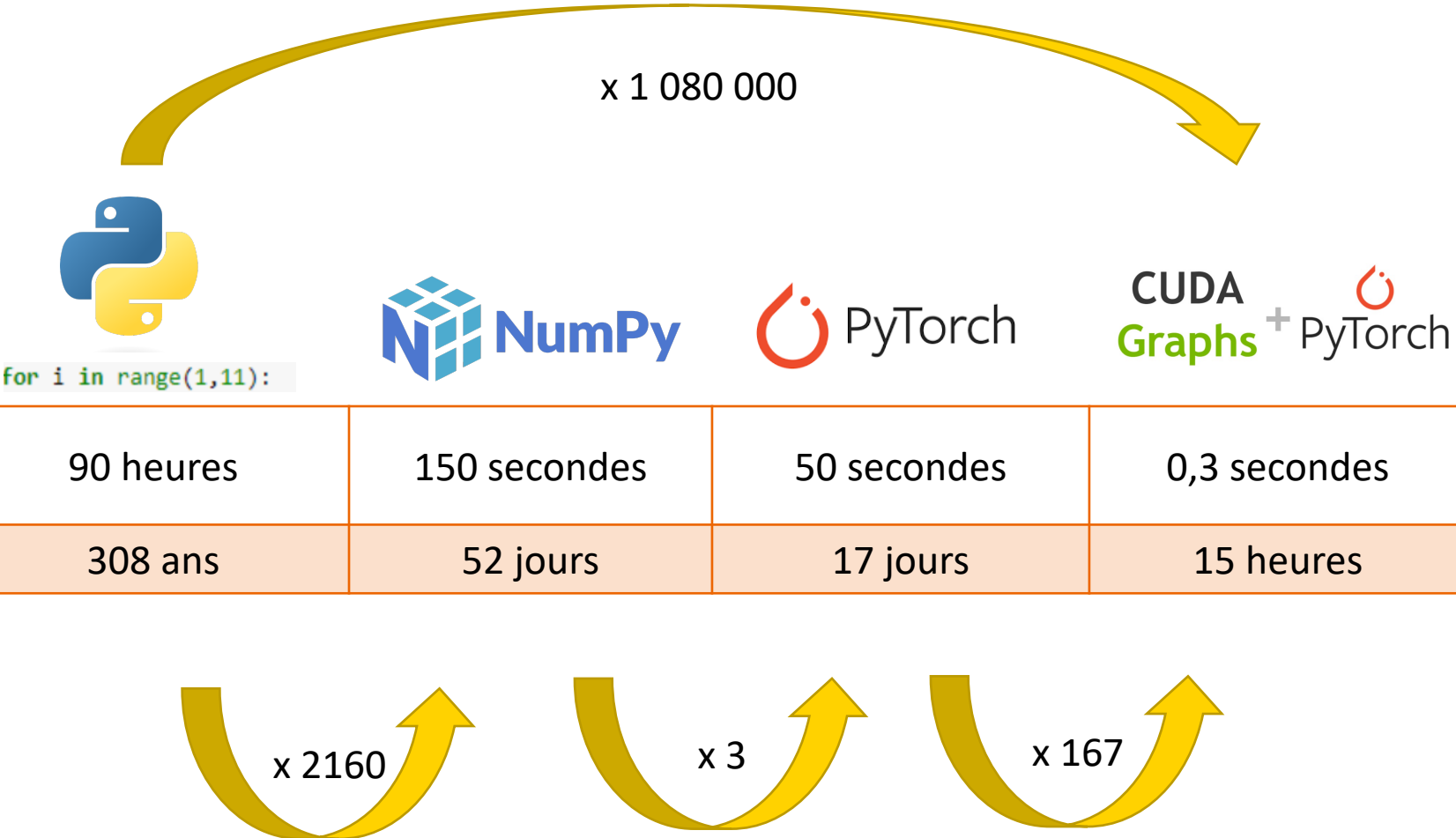
Louis Rebaud

Taux de fausses découvertes (FDR):



$$FDR = \frac{S}{N}$$

# Accélération du pipeline





# Résultats LPOCV

Cohorte	uni_PFS	uni_OS	multi_PFS	multi_OS	FDR
FL	0,56	0,59	0,625 (+0.001)	0,719 (+0.009)	0,0010
DLBCL	0,59	0,58	0,624 (+0.012)	0,646 (+0.002)	0,0041

## Sélection DLBCL

feature	uni_PFS	uni_OS	multi_PFS	multi_OS
colon_PT_glcm_Correlation	0.592 (p <0.00)	0.600 (p <0.01)	0.633 (p <0.00)	0.668 (p <0.00)
shell_PT_firstorder_Minimum_range	0.603 (p <0.00)	0.605 (p <0.01)	0.629 (p <0.00)	0.651 (p <0.02)
esophagus_CT_glrIm_GrayLevelNonUniformity	0.581 (p <0.01)	0.607 (p <0.01)	0.632 (p <0.00)	0.666 (p <0.00)
pancreas_longestDistanceToTumor	0.615 (p <0.00)	0.578 (p <0.05)	0.635 (p <0.00)	0.655 (p <0.00)
shell_CT_firstorder_Skewness_range	0.625 (p <0.00)	0.604 (p <0.01)	0.630 (p <0.00)	0.655 (p <0.00)

## Sélection FL

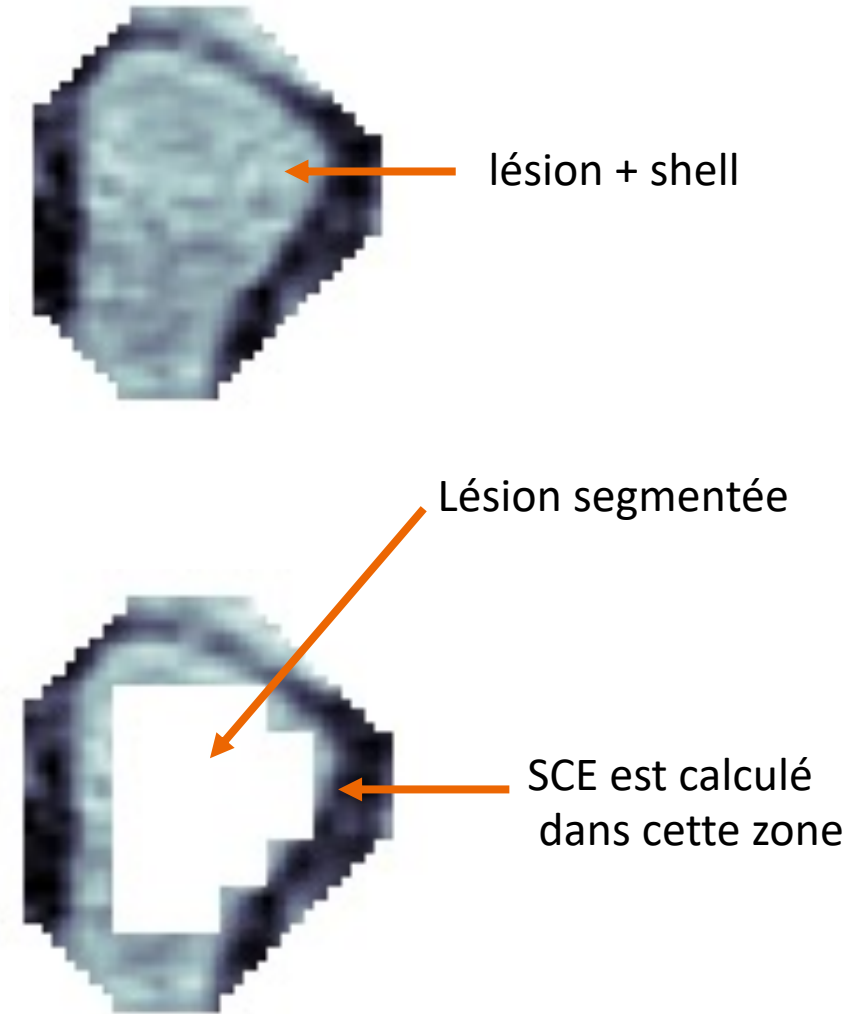
feature	uni_PFS	uni_OS	multi_PFS	multi_OS
lesion_CT_gldm_SmallDependenceLowGrayLevelEmphasis_range	0.570 (p <0.01)	0.618 (p <0.01)	0.639 (p <0.00)	0.730 (p <0.00)
shell_PT_gldm_SmallDependenceLowGrayLevelEmphasis_media	0.588 (p <0.00)	0.603 (p <0.03)	0.644 (p <0.00)	0.728 (p <0.00)
pancreas_PT_glszm_SizeZoneNonUniformityNormalized	0.579 (p <0.00)	0.613 (p <0.02)	0.642 (p <0.00)	0.731 (p <0.00)
shell_CT_gldm_DependenceEntropy_mini	0.606 (p <0.00)	0.640 (p <0.00)	0.643 (p <0.00)	0.738 (p <0.00)

## Sélection FL – avec score sur DLBCL

feature	uni_PFS	uni_OS	multi_PFS	multi_OS
lesion_CT_gldm_SmallDependenceLowGrayLevelEmphasis_range	<b>0.580 (p &lt;0.01)</b>	<b>0.589 (p &lt;0.03)</b>	<b>0.621 (p &lt;0.00)</b>	<b>0.652 (p &lt;0.01)</b>
shell_PT_gldm_SmallDependenceLowGrayLevelEmphasis_media	0.513 (p <0.63)	0.510 (p <0.74)	0.611 (p <0.22)	0.642 (p <0.36)
pancreas_PT_glszm_SizeZoneNonUniformityNormalized	0.493 (p <0.93)	0.525 (p <0.51)	0.613 (p <0.11)	0.622 (p <1.00)
shell_CT_gldm_DependenceEntropy_mini	<b>0.601 (p &lt;0.00)</b>	<b>0.616 (p &lt;0.00)</b>	<b>0.619 (p &lt;0.00)</b>	<b>0.653 (p &lt;0.01)</b>

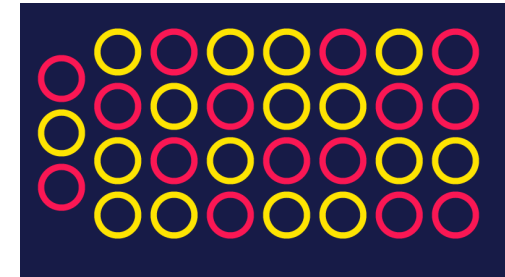
## Shell CT Dependence Entropy (SCE)

- `shell_CT_gldm_DependenceEntropy_mini`
- Nom raccourci: Shell CT dependence Entropy (SCE)
- Est calculé dans la shell de 8 mm d'épaisseur autour des ROIs
- Est calculé dans l'image CT

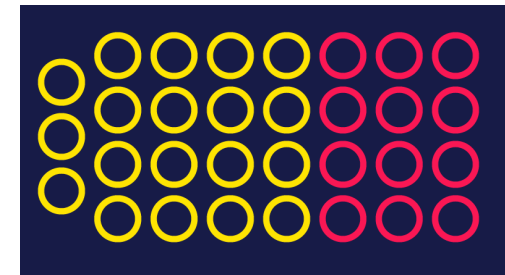


## Shell CT Dependence Entropy (SCE)

- SCE quantifie la quantité d'entropie dans le shell CT
- Des valeurs plus élevées sont associées à des zones CT moins bien définies
- SCE est calculé pour toutes les lésions du patient. Ces valeurs sont agrégées au niveau du patient en prenant la valeur minimale.
- Une **entropie plus faible** est associée à un **risque plus élevé**
- Une entropie plus élevée est associée à un **risque plus faible**



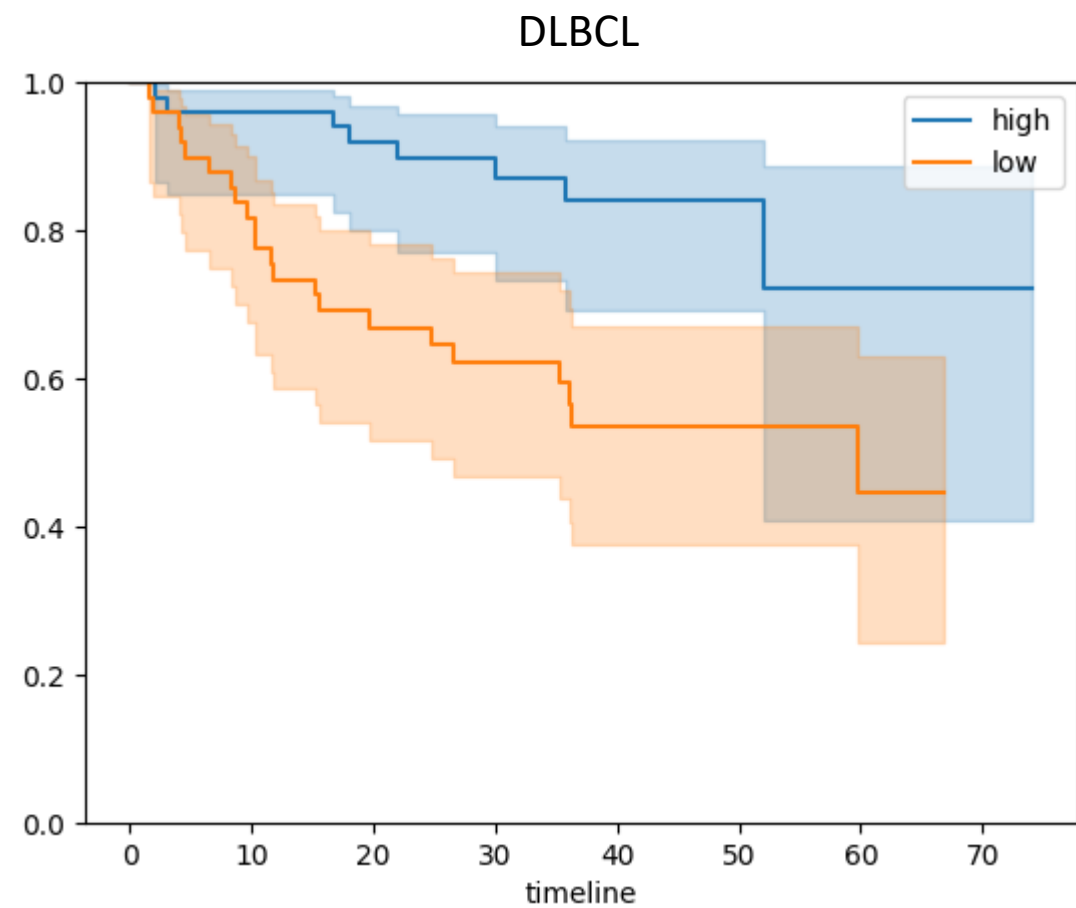
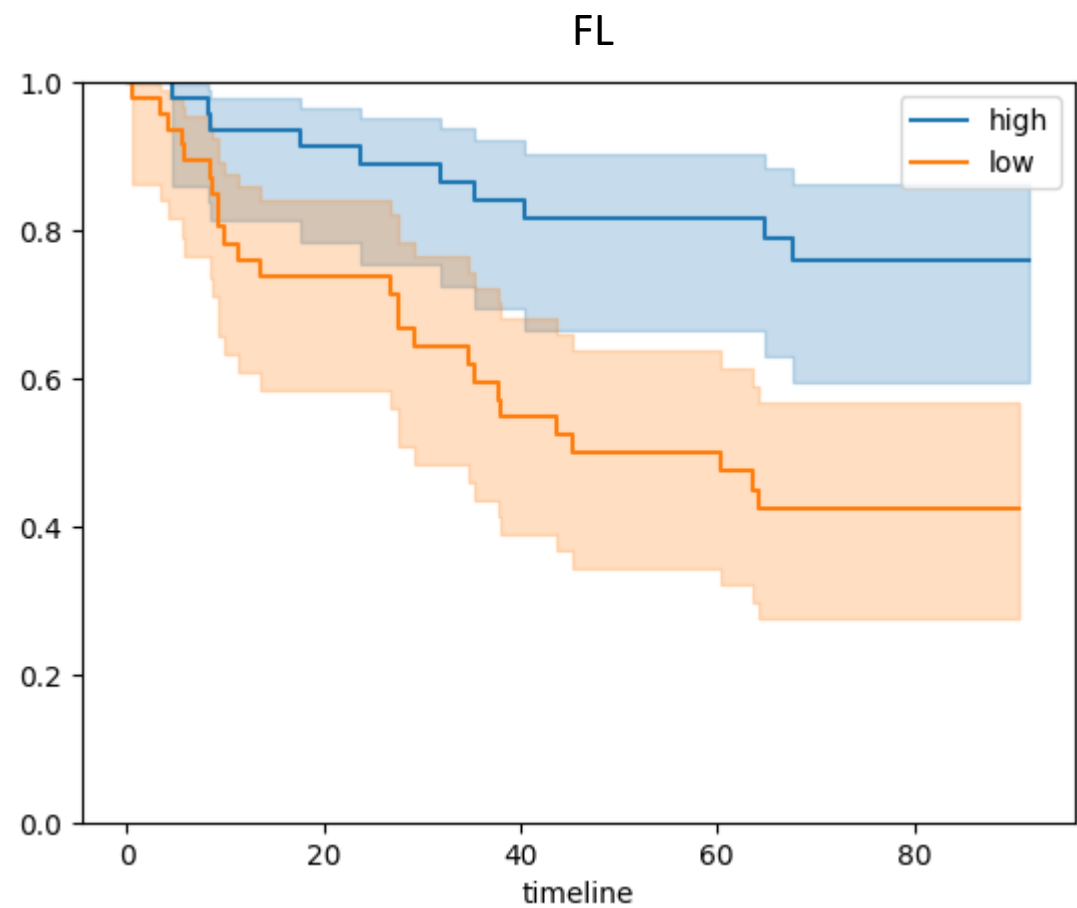
High  
entropy



Low  
entropy



# Courbes de survie des 50 SCE les plus basses et des 50 SCE les plus hautes



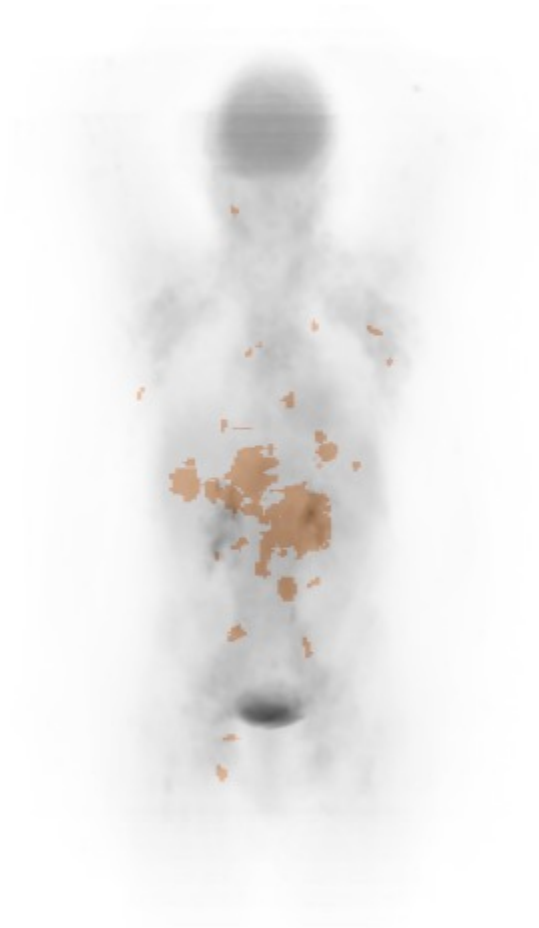
# Hypothèses

- Position de la lésion dans le corps (quel organe ?)
- Distance jusqu'à la frontière de l'organe hôte
- Contraintes mécaniques autour de la lésion
- Nature des tissus environnants
- Activation des systèmes de défense autour de la lésion
- Vitesse de croissance de la tumeur



## MIP alignés - FL

low



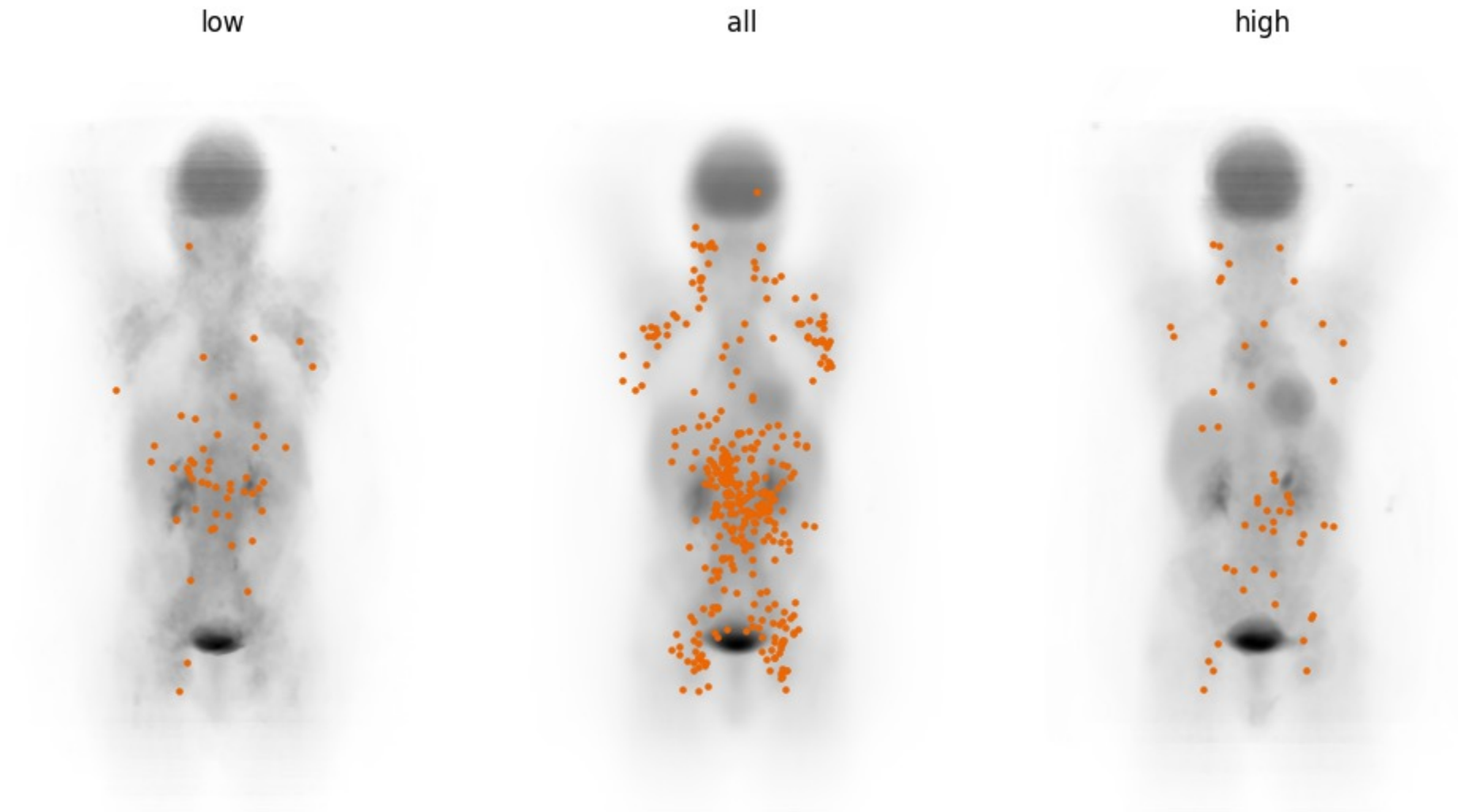
all



high

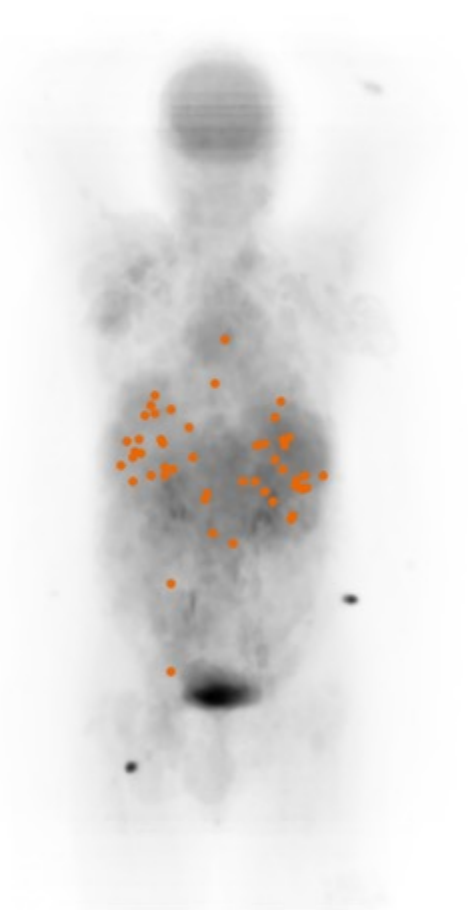


## MIP alignés - FL

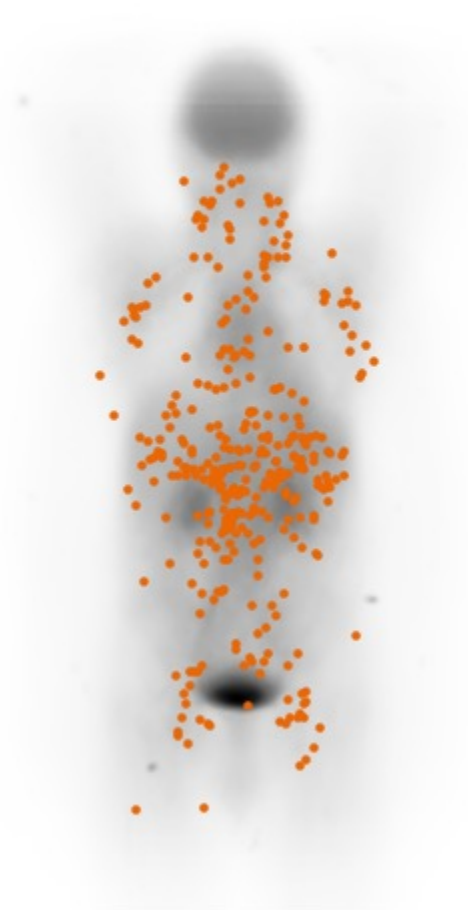


## MIP alignés - DLBCL

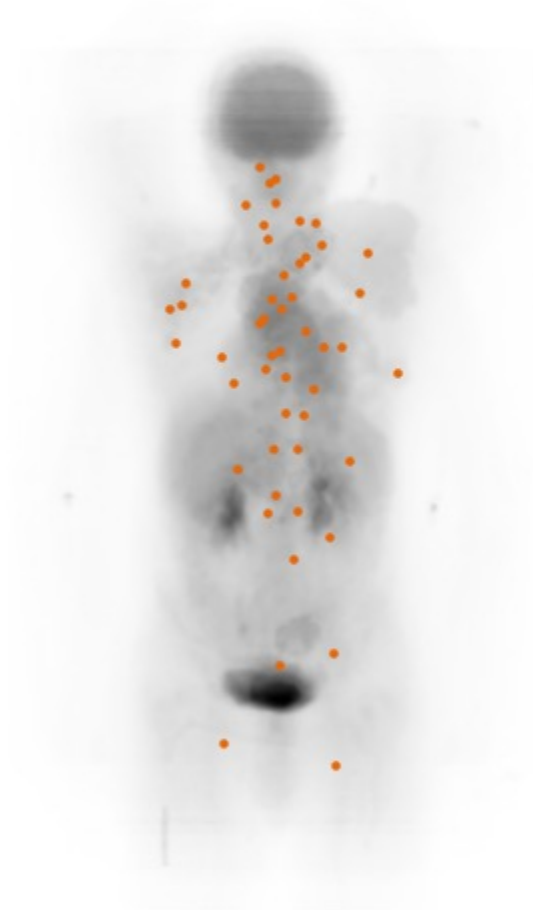
low



all

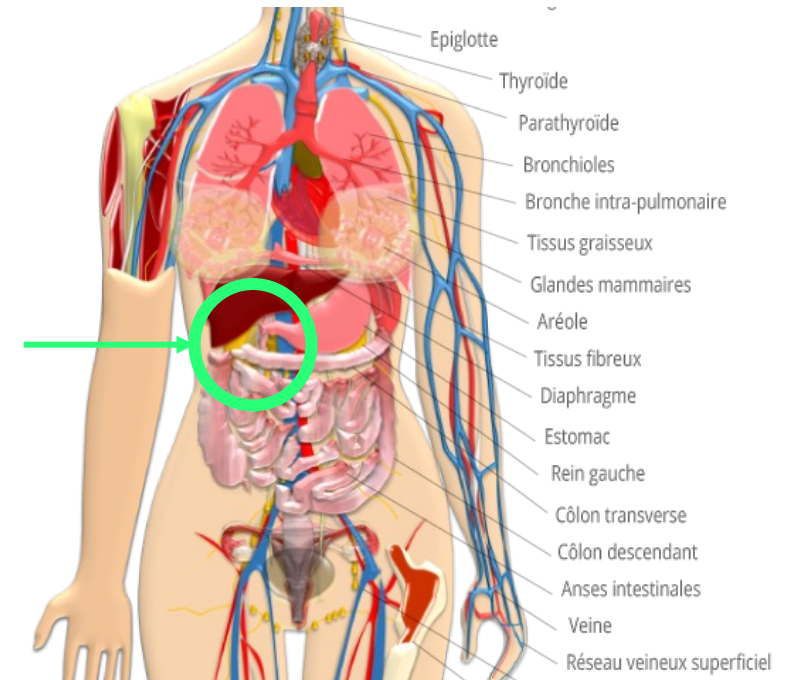
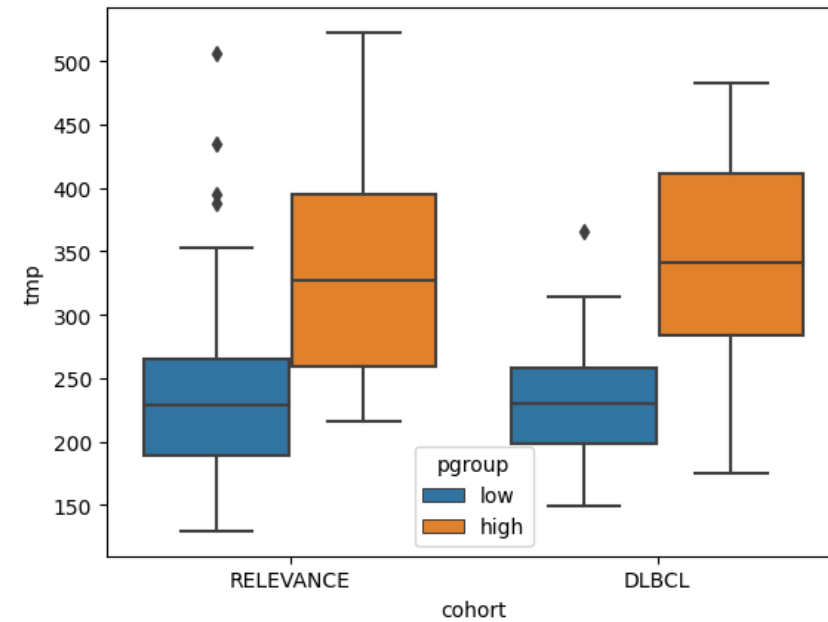


high



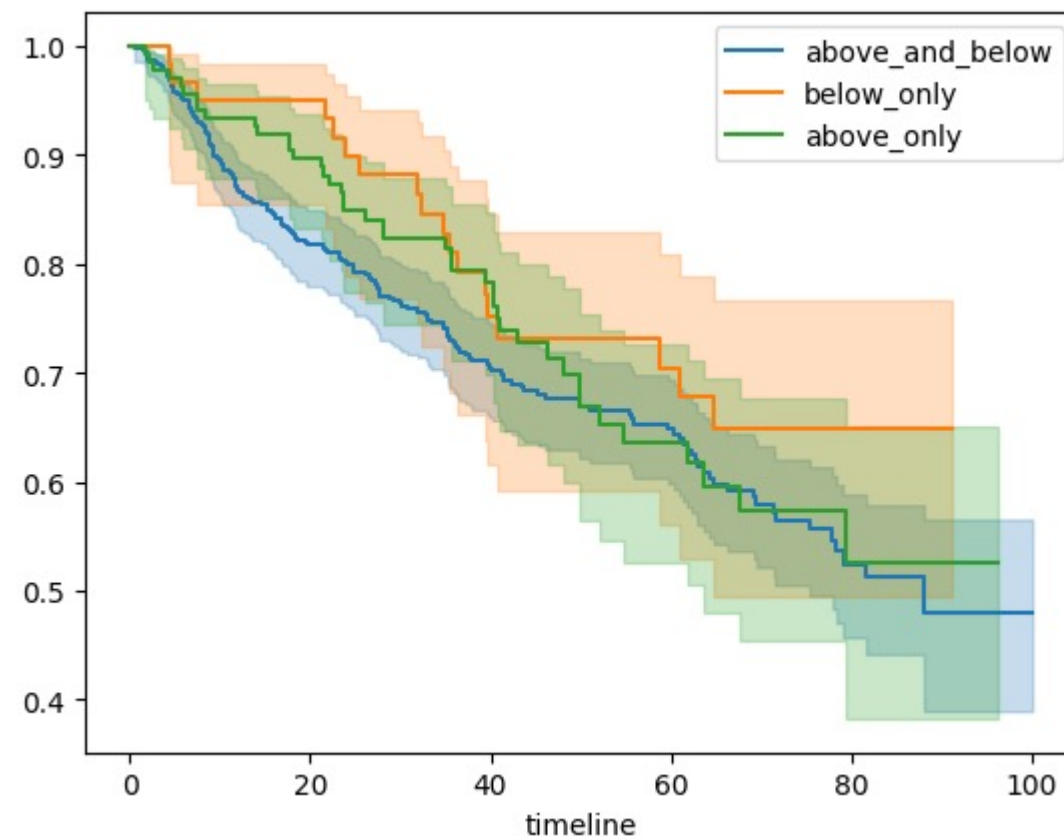
## Distance moyenne

- Distance moyenne à :
  - Vésicule biliaire
  - Foie
  - Rein droit
- SCE semble refléter la présence de lésions proches de cette zone

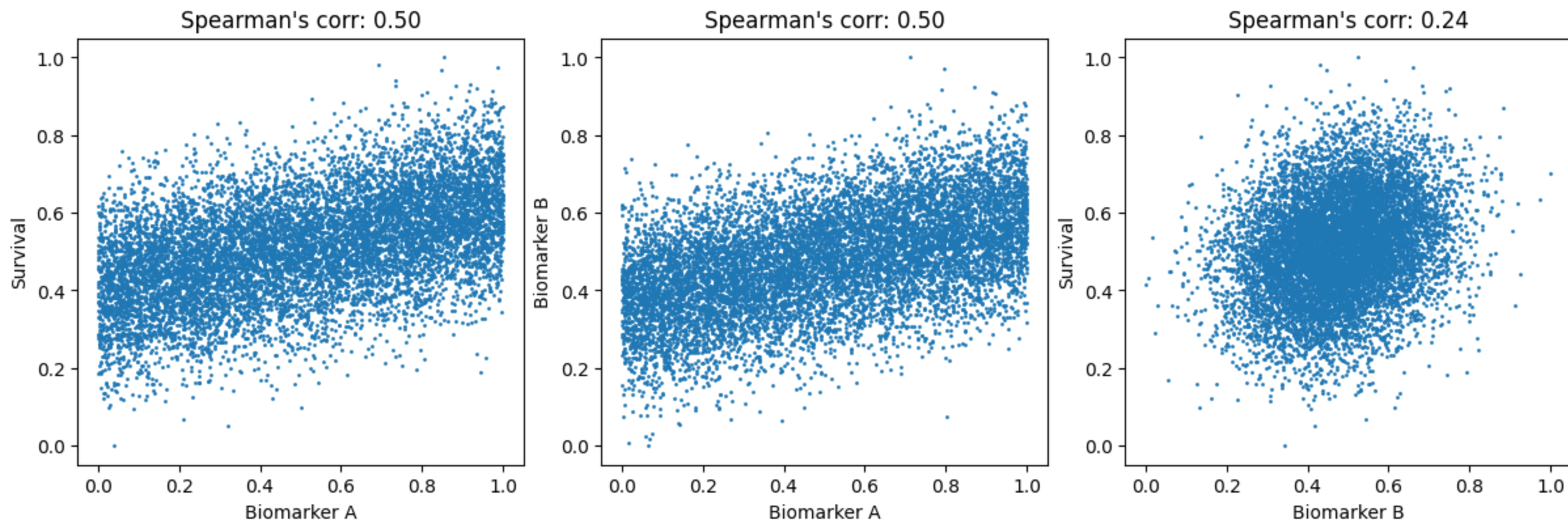


## Lésions au-dessus et en-dessous du diaphragme ?

- La présence de lésions au-dessus et/ou en dessous du diaphragme est un des paramètres du score Ann Arbor
- On n'observe pas de différences significatives en se basant sur ce critère



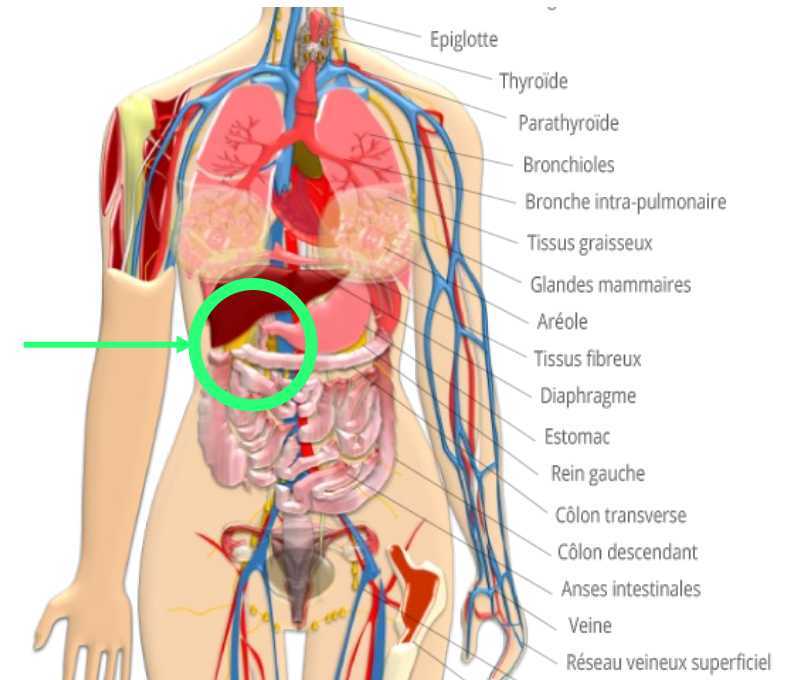
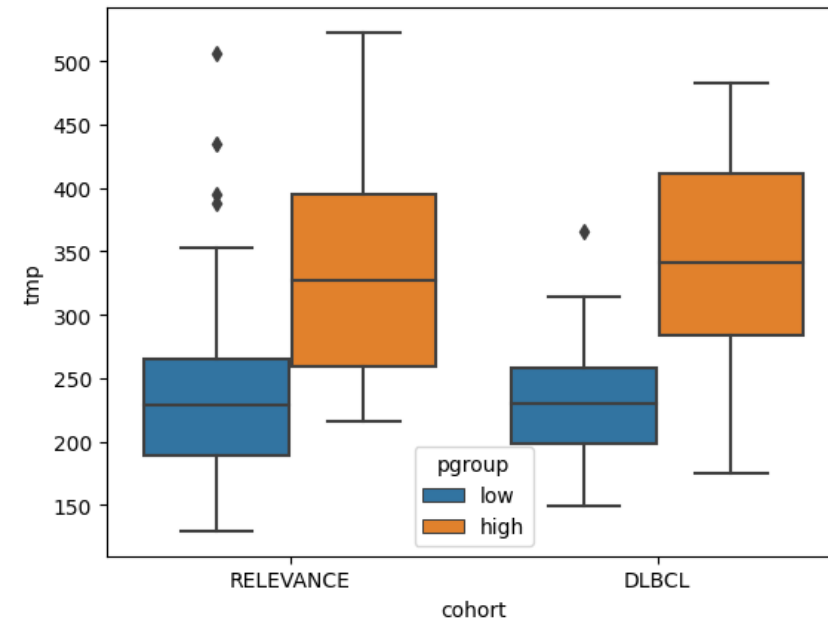
## La corrélation n'est pas transitive





## Trouver une alternative

- Il faut trouver un biomarqueur basé sur la présence de lésions dans cette zone et montrer qu'il est prédictif
- Ne pas trouver un tel biomarqueur serait une indication qu'une autre information que la position est prédictive dans SCE



## La suite

- Optimisation des paramètres du pipeline
- Interprétation des biomarqueurs sélectionnés
- Rédaction de deux articles (un par maladie)