

# Présentation

BA Alpha Oumar Master 2 IMB

15/02/2022-15/08/2022

Définition :Données Transcriptomiques ou données d'express  
Analyse Exploratoire de notre Jeux de données  
Normalisations des données  
Visualisation(ACP,T-SNE,UMAP)  
Construction de notre modèle Auto-encodeur  
Visualisation t-SNE



Encadrée Par Misbah Razzaq (chercheuse associée à INRAE)

- 1 Définition :Données Transcriptomiques ou données d'expression des gènes
  - Définition
  - Jeux de données
- 2 Analyse Exploratoire de notre Jeux de données
  - Figure et Valeurs aberrantes
  - Corrélation des données
  - Visualisations des données
- 3 Normalisations des données
  - Méthodes CPM
    - Exemple
  - Méthodes MinMax
- 4 Visualisation(ACP,T-SNE,UMAP)
  - Méthodes ACP
  - Méthodes T-SNE

## 5 Construction de notre modèle Auto-encodeur

- Auto-encodeur
- Modèle

## 6 Visualisation t-SNE

les données transcriptomiques sont issues de technologies variées (Polymerase Chain Reaction–PCR en temps réel, biopuces constituées d'acides désoxyribonucléique–ADN–complémentaires déposés sur membrane de nylon ou sur une lame de verre), chacune nécessitant des pré traitements spécifiques : Analyse d'image, quantification et normalisation.

	C01_FC_NS	C02_FC_S	C03_FB1_NS	C04_FB1_S	C05_FB2_NS	C06_FB2_S	C19_FC_NS	C20_FC_S	C21_FB1_NS	C22_FB1_S	...	C42_FB2_S	B1a
0	34.81	37.97	35.30	37.23	37.59	38.26	36.92	40.29	36.11	39.09	...	27.20	37.91
1	36.73	38.98	41.54	35.71	36.76	35.10	31.63	33.15	36.44	32.62	...	40.63	36.21
2	34.18	35.63	33.06	37.59	36.00	36.25	33.19	32.72	35.03	37.03	...	39.07	37.55
3	38.40	37.28	35.47	38.75	33.93	39.65	38.19	35.25	38.83	41.51	...	37.02	37.46
4	37.48	41.67	42.32	38.22	39.72	36.94	34.61	32.98	39.62	38.24	...	41.51	36.83

Figure – Jeux de données

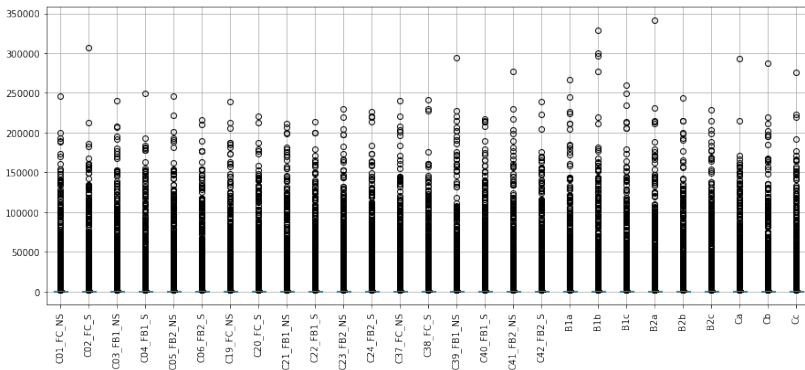


Figure – Boxplot des valeurs aberrantes

Soit  $X_1, X_2, \dots, X_n$  les données ordonnées dans l'ordre croissant, les données  $X_1$  et  $X_n$  sont respectivement l'observation extrême inférieur et supérieur. On cherche les valeurs aberrantes en dehors de

$$[\bar{X} - 1.5(Q_3 - Q_1); \bar{X} + 1.5(Q_3 - Q_1)]$$

## Analyse Exploratoire de notre Jeux de données

Normalisations des données

Visualisation(ACP,T-SNE,UMAP)

Construction de notre modèle Auto-encodeur

Visualisation t-SNE

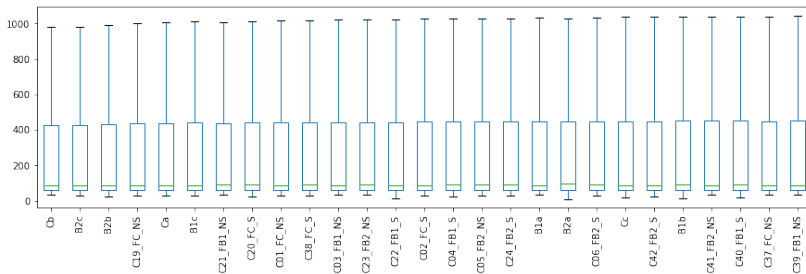


Figure – Boxplot de l'ensemble des observations



Définition :Données Transcriptomiques ou données d'express

**Analyse Exploratoire de notre Jeux de données**

Normalisations des données

Visualisation(ACP,T-SNE,UMAP)

Construction de notre modèle Auto-encodeur

Visualisation t-SNE



représentations des observations

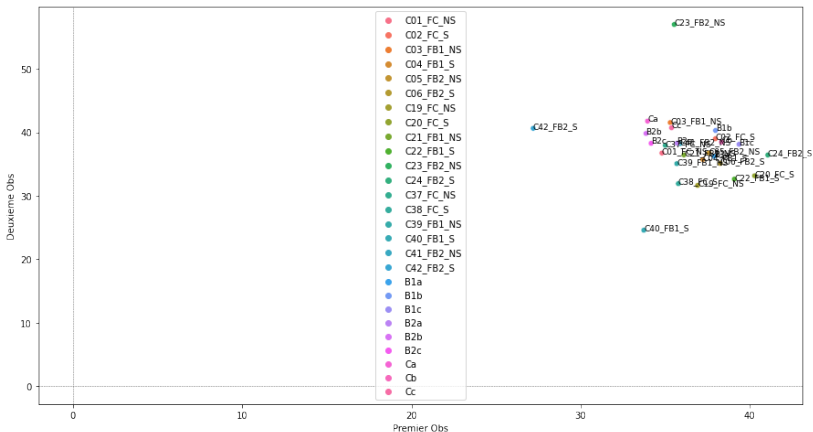


Figure – nuages de points des observations

Nous séquençons une bibliothèque avec 5M de lectures.Parmi eux, un total de 4M appariées à la séquences du génome et 5000 lectures appariées à un gène donnée.

$$CPM = \frac{5000 * 10^6}{4 * 10^6} = 1250$$

$$X_{normalise} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

avec :

- $X_{min}$  :la plus petite valeur observée pour la feature X
- $X_{max}$  :la plus grande valeur observée pour la feature X
- $X$  : la valeur du feature qu'on cherche à normaliser.

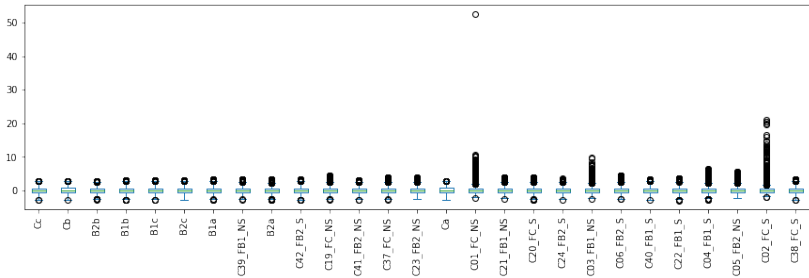
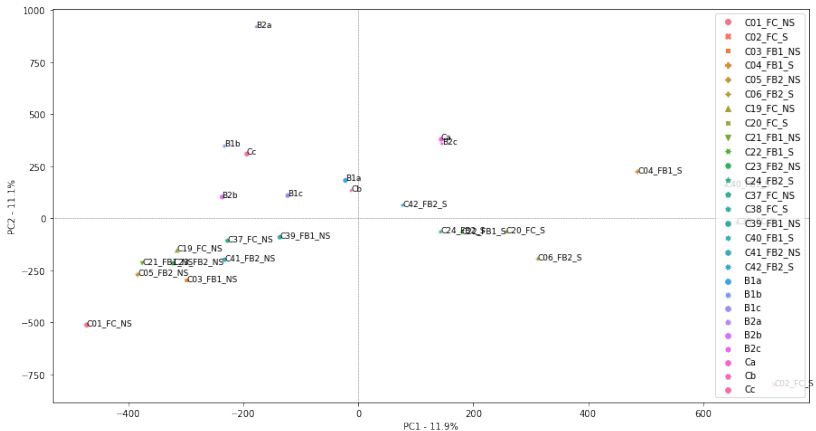


Figure – Après Normalisation

Analyse Exploratoire de notre Jeux de données  
 Normalisations des données  
 Visualisation(ACP,T-SNE,UMAP)  
 Construction de notre modèle Auto-encodeur  
 Visualisation t-SNE

Premier plan factoriel (23,029%)



(a) méthode ACP

Définition :Données Transcriptomiques ou données d'express

Analyse Exploratoire de notre Jeux de données

Normalisations des données

**Visualisation(ACP,T-SNE,UMAP)**

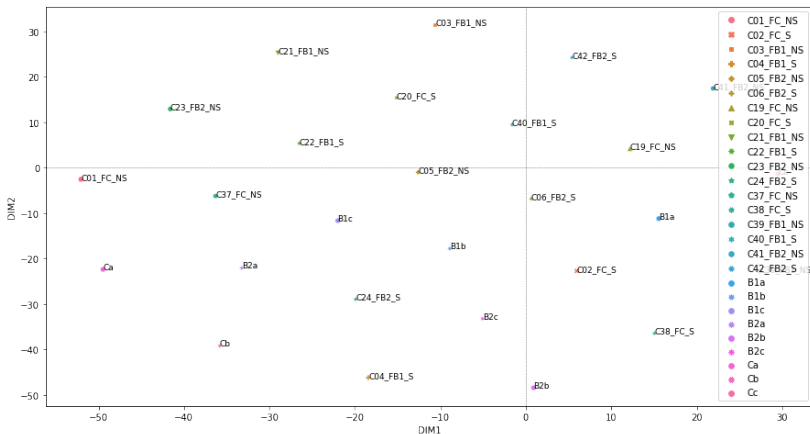
Construction de notre modèle Auto-encodeur

Visualisation t-SNE



Définition :Données Transcriptomiques ou données d'express  
Analyse Exploratoire de notre Jeux de données  
Normalisations des données  
**Visualisation(ACP,T-SNE,UMAP)**  
Construction de notre modèle Auto-encodeur  
Visualisation t-SNE

Visualisation avec T-SNE



(b) méthode T-SNE

Définition :Données Transcriptomiques ou données d'express  
Analyse Exploratoire de notre Jeux de données  
Normalisations des données  
Visualisation (ACP, T-SNE, UMAP)  
**Construction de notre modèle Auto-encodeur**  
Visualisation t-SNE

Méthodes ACP  
Méthodes ACP

Figure – Différentes méthodes de réductions de dimensions

La variable "auto-encoder" représente le réseau dans son ensemble (modèle), commençant par encoder et finissant par decoder.Nous en définissons les caractéristiques d'apprentissage en spécifiant l'algorithme d'optimisation(Optimizer="adam").

Définition :Données Transcriptomiques ou données d'express

Analyse Exploratoire de notre Jeux de données

Normalisations des données

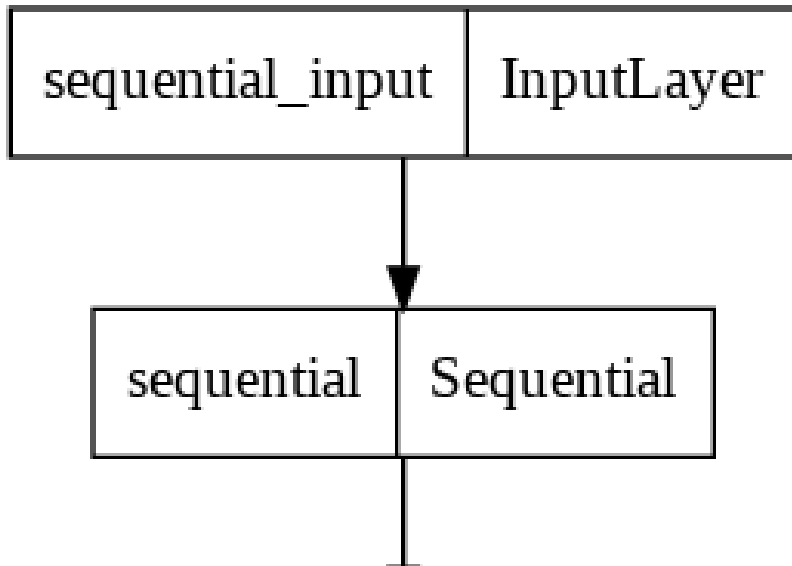
Visualisation (ACP, T-SNE, UMAP)

**Construction de notre modèle Auto-encodeur**

Visualisation t-SNE

Méthodes ACP

Méthodes ACP



Définition :Données Transcriptomiques ou données d'express

Analyse Exploratoire de notre Jeux de données

Normalisations des données

Visualisation(ACP,T-SNE,UMAP)

**Construction de notre modèle Auto-encodeur**

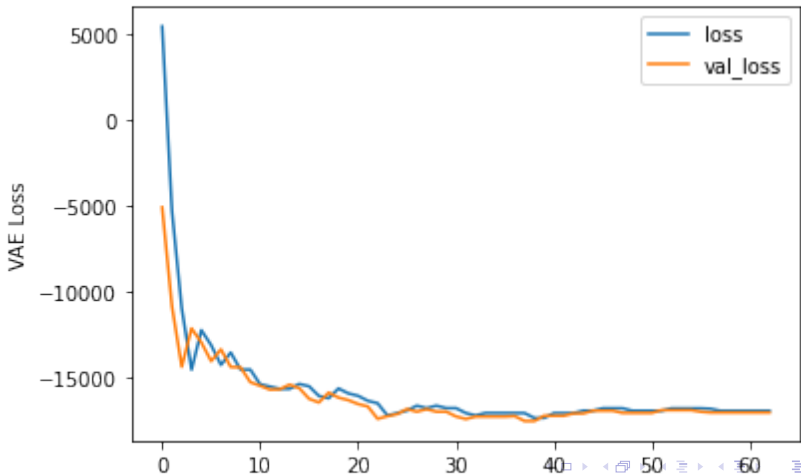
Visualisation t-SNE

Méthodes ACP

Méthodes ACP

## Figure – Auto-encoder

Nous pouvons représenter la décroissance de la fonction perte en utilisant l'objet 'historique' renvoyer par fit.



Définition :Données Transcriptomiques ou données d'express  
 Analyse Explorative de notre Jeux de données  
 Normalisations des données  
 Visualisation(ACP,T-SNE,UMAP)  
 Construction de notre modèle Auto-encodeur  
**Visualisation t-SNE**

