



UNIVERSITÉ D'ANGERS
UFR INFORMATIQUE

RAPPORT DE STAGE
MASTER 1 2016-2017



UMR INSERM 1232
-EQUIPE IMMUNITÉ INNÉE ET IMMUNOTHÉRAPIE

ANALYSE TRANSCRIPTOMIQUE

Présenté par :
RASOLONIAINA
MARLINO

Tuteur : M. Le TUTEUR
Chef d'équipe : M. Chef
D'ÉQUIPE

12 Avril 2017 — 20 Juin 2017

Table des matières

1 INTRODUCTION :	2
2 ACQUISITION DES DONNÉES :	2
2.1 Les puces à ADN :	2
2.2 La technologie Illumina :	2
2.3 Plan expérimental :	3
2.4 Les biais possibles :	3
3 DESCRIPTION DES DONNÉES :	3
3.1 Decryptage et lecture :	3
3.2 Données brutes :	3
3.3 Contrôle et qualité :	4
4 PRÉTRAITEMENT DES DONNÉES :	4
4.1 Transformation :	4
4.2 Normalisation :	4
4.3 Filtrage :	4
5 ANALYSE DES DONNÉES DE TRANSCRIPTOME :	4
5.1 Gènes différentiellement exprimés :	4
5.2 Gènes co-exprimés :	4
6 INTERPÉTATION :	4

1 INTRODUCTION :

2 ACQUISITION DES DONNÉES :

2.1 Les puces à ADN :

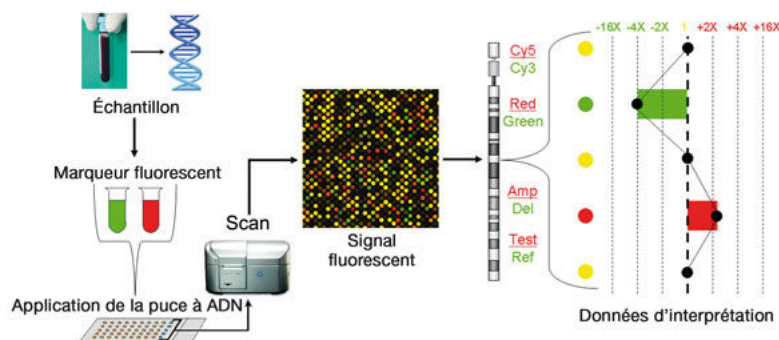
Une puce à ADN est constituée d'un support physique (le plus souvent une lame de verre) sur lequel sont déposées des molécules d'ADN correspondant à de petits fragments du génome (jusqu'à 40 000 dépôts différents par puce). On recouvre la puce de la solution contenant la population d'ARN à étudier. Les ARN s'hybrident sur les fragments d'ADN complémentaires. La quantité d'ARN fixée reflète la concentration de cet ARN dans la solution.

Pour des raisons pratiques, on utilise des ADNc plutôt que directement les ARN. Les ADNc sont marqués par un nucléotide radioactif ou un fluorochrome. Il est possible d'étudier simultanément plusieurs populations d'ADNc sur une même puce en utilisant des fluorochromes différents. La meilleure façon d'utiliser cette possibilité est de marquer l'ADN génomique avec un fluorochrome, toujours le même. On obtient ainsi une référence stable au cours des années qui permet de mettre toutes les puces à la même échelle, quelle que soit leur origine.

Un scanner mesure l'intensité du signal émis par l'ADNc hybridé au niveau de chaque dépôt. Parmi les valeurs que proposent les logiciels pour cette intensité, la plus fiable est la médiane de l'intensité des pixels car elle est moins sensible aux défauts de l'image (pixels sur-brillants par exemple).

Les puces comportent généralement plusieurs dépôts identiques pour chaque gène. Cela simplifie le travail lorsqu'il faut repérer les aberrations dans la lecture des intensités puisqu'il suffit d'examiner les cas où les valeurs diffèrent beaucoup d'un dépôt à l'autre. Il s'agit le plus souvent d'un défaut physique sur la puce et il est facile d'éliminer la valeur aberrante. Dans le doute, on conserve la médiane des différentes mesures.

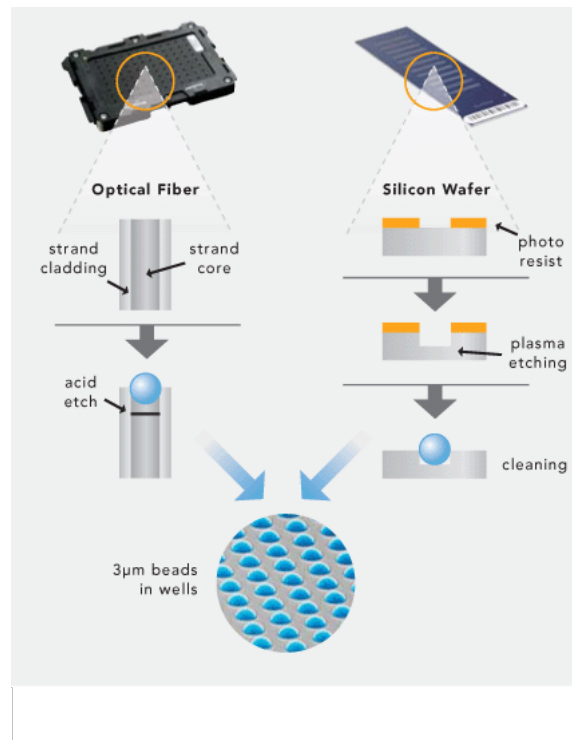
Plusieurs types de puces à ADN existent selon le support, la nature des fragments fixés à la surface, le mode de fabrication, la densité, le mode de marquage des cibles et les méthodes d'hybridation.



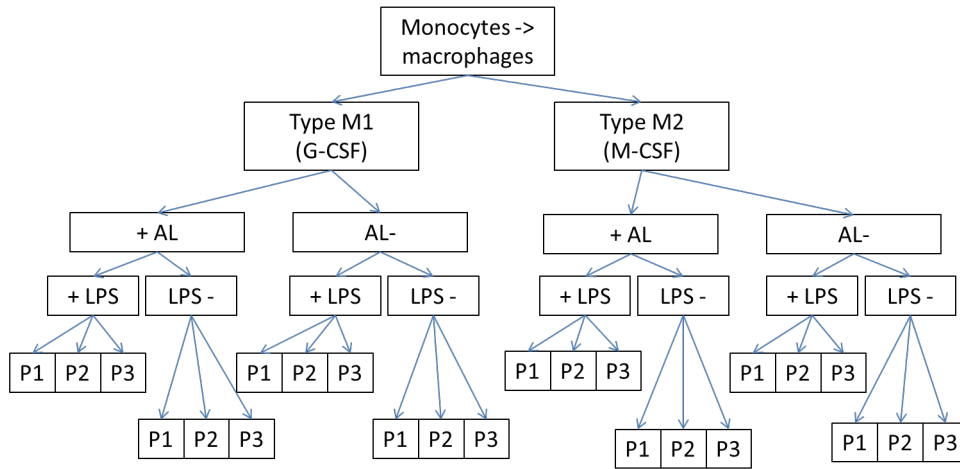
2.2 La technologie Illumina :

Illumina, Inc. est une société américaine qui fabrique et commercialise des systèmes intégrés pour l'analyse de la variation génétique et la fonction biologique notamment des gammes de produits et services qui servent les marchés du séquençage, génotypage et expression génétique.

Une de ces récentes fabrications, la puce "BeadArray technologie".



2.3 Plan expérimental :



N=24 prélèvements
8 conditions expérimentales différentes

L'expression des gènes des macrophages a été analysée par des puces d'expression génique de technologie Illumina avec 48210 sondes pour chaque prélèvement. Ce qui nous donne une matrice de données de dimension (48210 x 24) avec les 24 prélèvements.

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

2.4 Les biais possibles :

Il peut exister des biais systématiques dus à d'autres facteurs, tels que l'affinité des séquences ou l'efficacité du marquage.

3 DESCRIPTION DES DONNÉES :

3.1 Decryptage et lecture :

File	Description
(Serial Number).txt	un fichier qui stocke la positions,l'identité et l'intensité de chaque spots
Metrics.txt	un pour chaque BeadChip et contient des informations récapitulatives sur l'intensité des signals, la quantité de saturation, la mise au point et l'enregistrement sur l'image (s) de chaque section
Effective.cfg	fichier de configuration des paramètres du scanner
(Serial Number).sdf	fichier de description des échantillons d'Illumina utilisé pour déterminer les propriétés (positions) physique d'une section et savoir les sections liées sur chaque échantillons
*.idat	contiennent la moyenne des intensités du signal de chaque spots

- HumanHT-12_V4_0_R2_15002873_B.bgx
- idat
- bgx
- sdf
- cfg
- Metrics.txt

3.2 Données brutes :

- Barcode
- Section
- ChipType
- RunInfo
- Quants
 - MeanBinData
 - TrimmedMeanBinData
 - DevBinData
 - MedianBinData
 - BackgroundBinData
 - BackgroundDevBinData
 - CodesBinData
 - NumBeadsBinData
 - NumGoodBeadsBinData
 - IllumicodeBinData

3.3 Contrôle et qualité :

4 PRÉTRAITEMENT DES DONNÉES :

4.1 Transformation :

4.2 Normalisation :

4.3 Filtrage :

5 ANALYSE DES DONNÉES DE TRANSCRIPTION :

5.1 Gènes différentiellement exprimés :

5.2 Gènes co-exprimés :

6 INTERPRÉTATION :

Caractérisation d'un ensemble de gènes

Résumé