

# Analyse et Intégration de données du projet « Milieu Intérieur » coordonné par l’Institut Pasteur

Violaine Saint-André – Institut Pasteur  
HUB-C3BI  
[violaine.saint-andre@pasteur.fr](mailto:violaine.saint-andre@pasteur.fr)

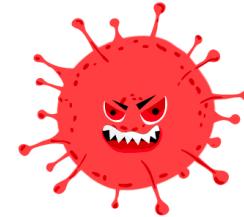
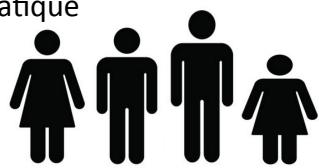


# La réponse immunitaire diffère entre les individus

Population diversifiée (age, sexe, poids, sérologies...) évoluant dans des environnements différents (ville, famille...), avec des habitudes différentes (tabac, heures de sommeil...)

Stimulation  
immunitaire

Individu  
asymptomatique



Réponses immunitaires  
variables chez les individus

Cette variabilité inter-individuelle n'est généralement pas prise en compte pour les traitements et vaccinations

Il est nécessaire de décrire les facteurs associés à la variabilité de la réponse afin de contribuer au développement d'approches médicales de précision

# La réponse immunitaire diffère entre les individus

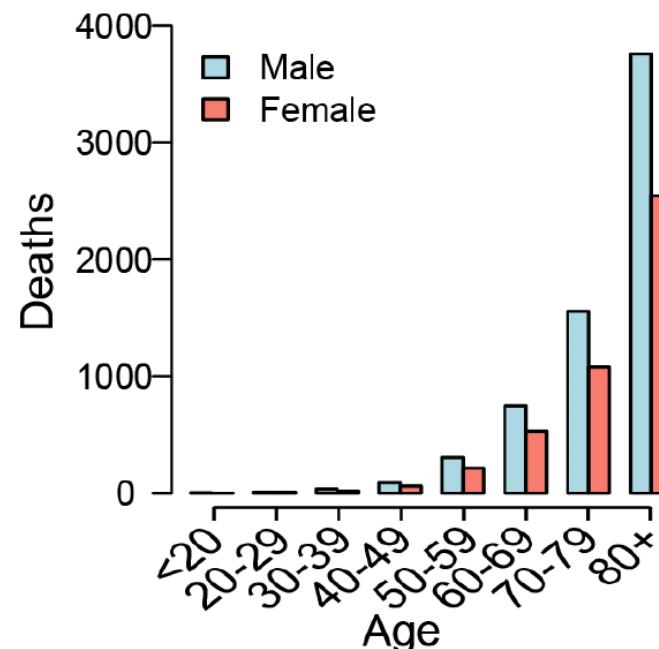
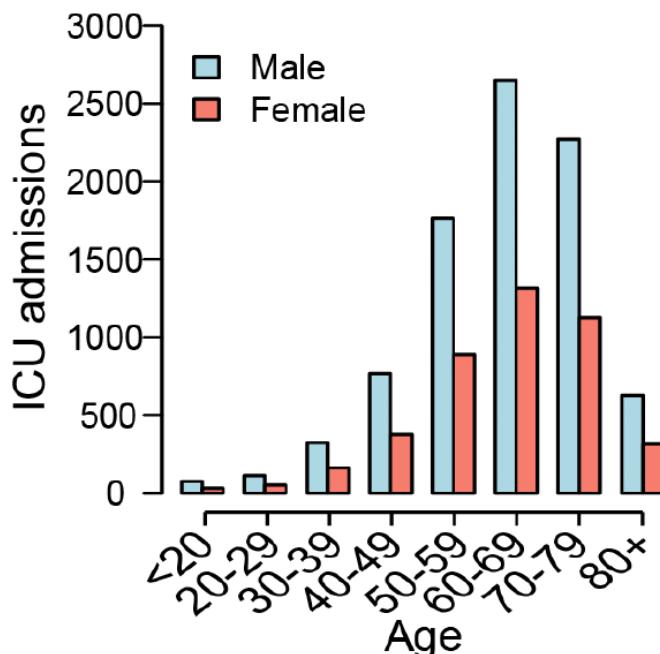
Science

REPORTS

Cite as: H. Salje *et al.*, *Science* 10.1126/science.abc3517 (2020).

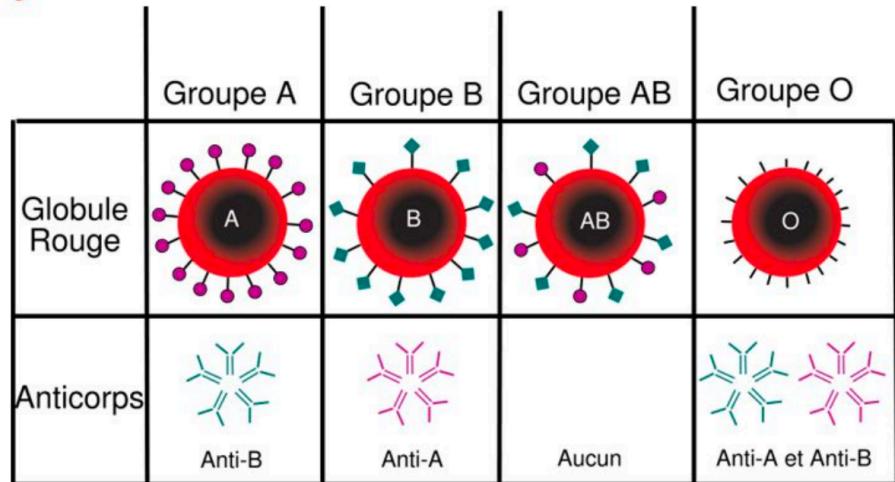
## Estimating the burden of SARS-CoV-2 in France

Henrik Salje<sup>1,2,3\*</sup>, Cécile Tran Kiem<sup>1,4\*</sup>, Noémie Lefrancq<sup>1</sup>, Noémie Courtejoie<sup>5</sup>, Paolo Bosetti<sup>1</sup>, Juliette Paireau<sup>1,6</sup>, Alessio Andronico<sup>1</sup>, Nathanaël Hozé<sup>1</sup>, Jehanne Richet<sup>5</sup>, Claire-Lise Dubost<sup>5</sup>, Yann Le Strat<sup>6</sup>, Justin Lessler<sup>3</sup>, Daniel Levy-Bruhl<sup>6</sup>, Arnaud Fontanet<sup>7,8</sup>, Lulla Opatowski<sup>9,10</sup>, Pierre-Yves Boelle<sup>11</sup>, Simon Cauchemez<sup>1†</sup>

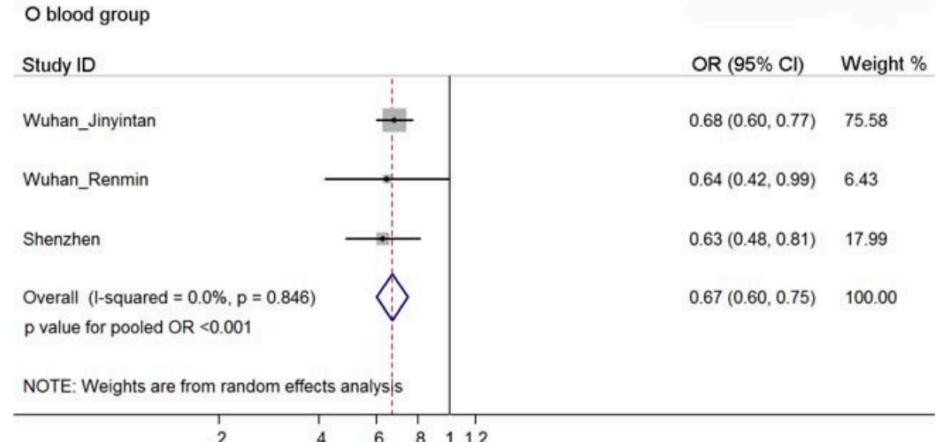
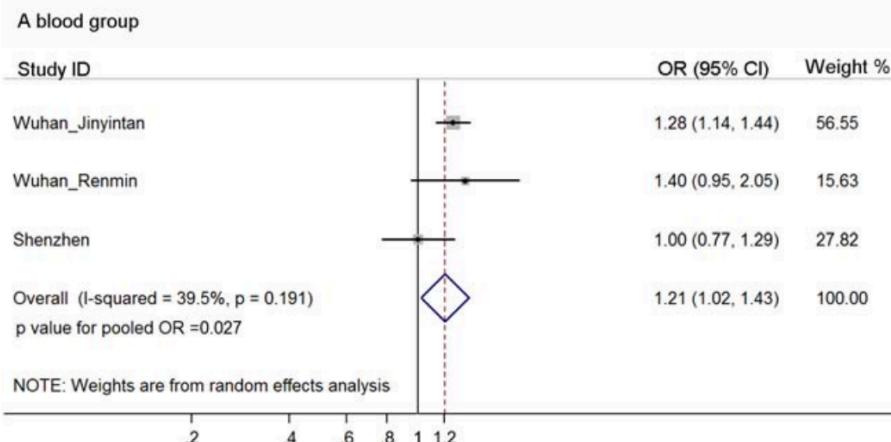


# La réponse immunitaire diffère entre les individus

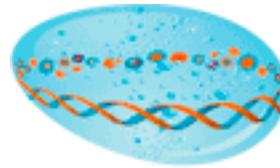
Système ABO



## Relationship between the ABO Blood Group and the COVID-19 Susceptibility



# Projet « Milieu Intérieur » coordonné par l’Institut Pasteur

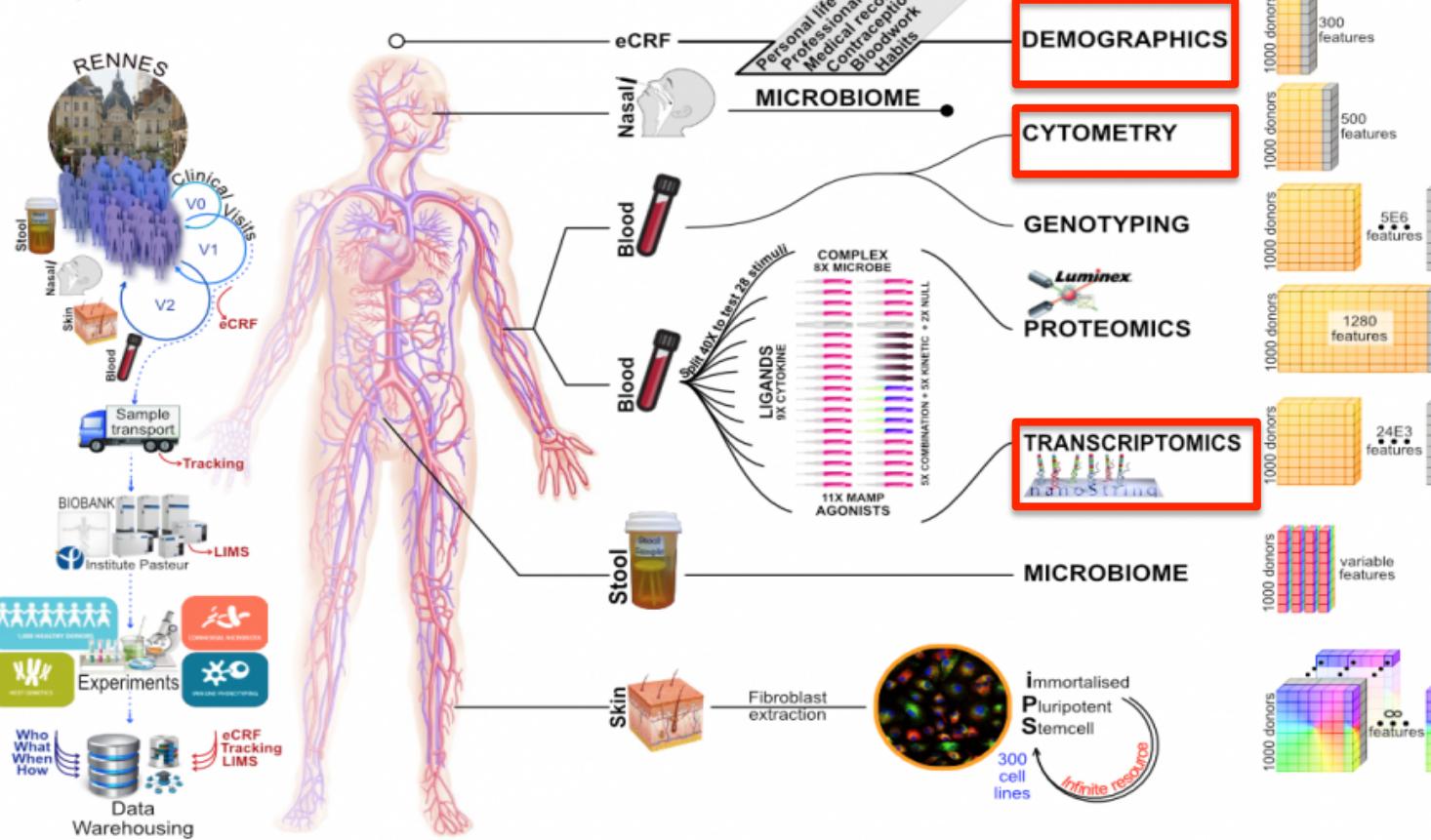


Milieu Intérieur  
Vers une médecine personnalisée

Etablissement d'une cohorte de 1000 individus en bonne santé

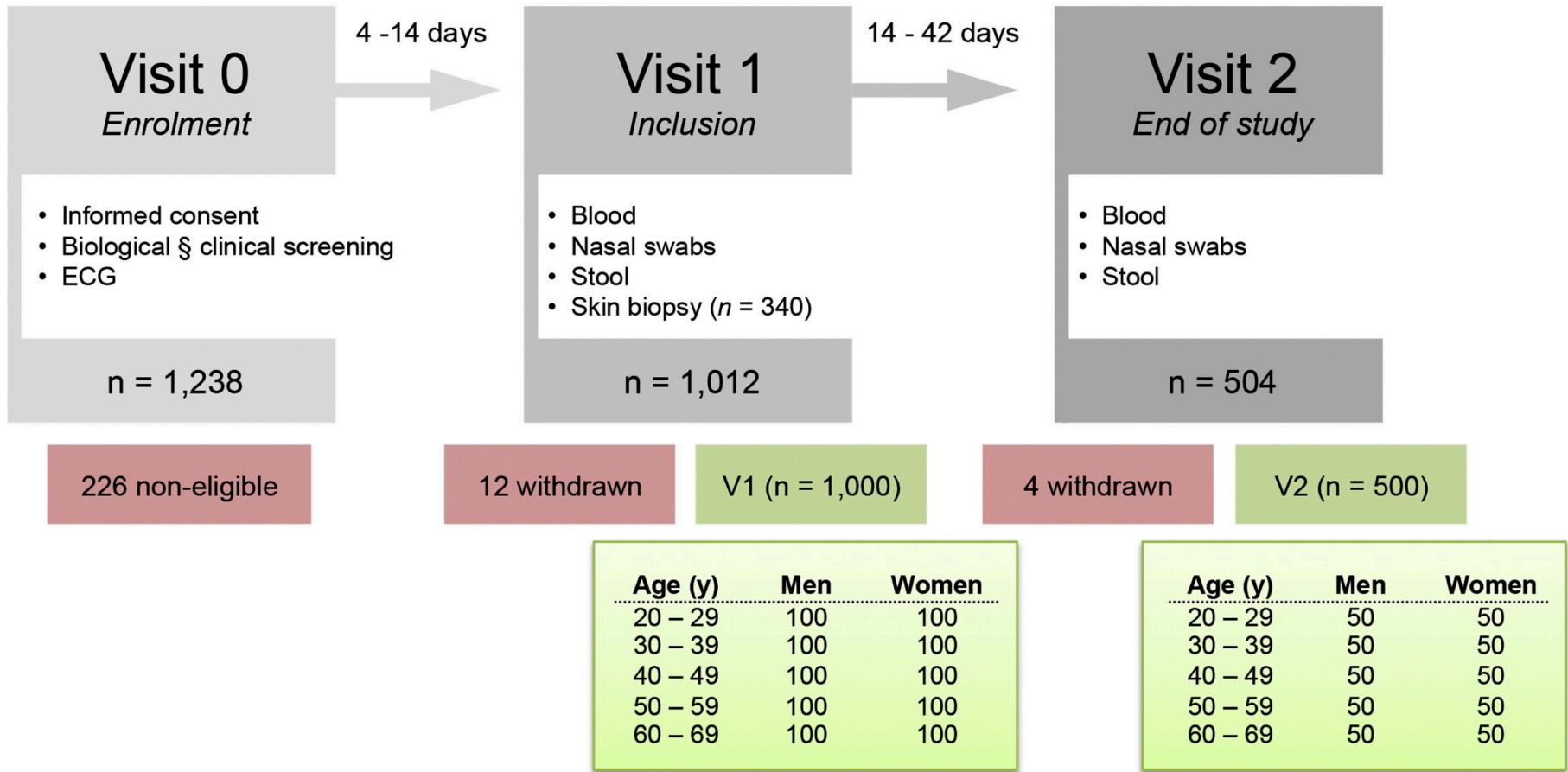
## The Milieu Intérieur Project

Hosted by Institute Pasteur on behalf of the Milieu Intérieur Consortium



Thomas et al., Clinical Immunol 2015;  
<http://www.milieointerieur.fr>

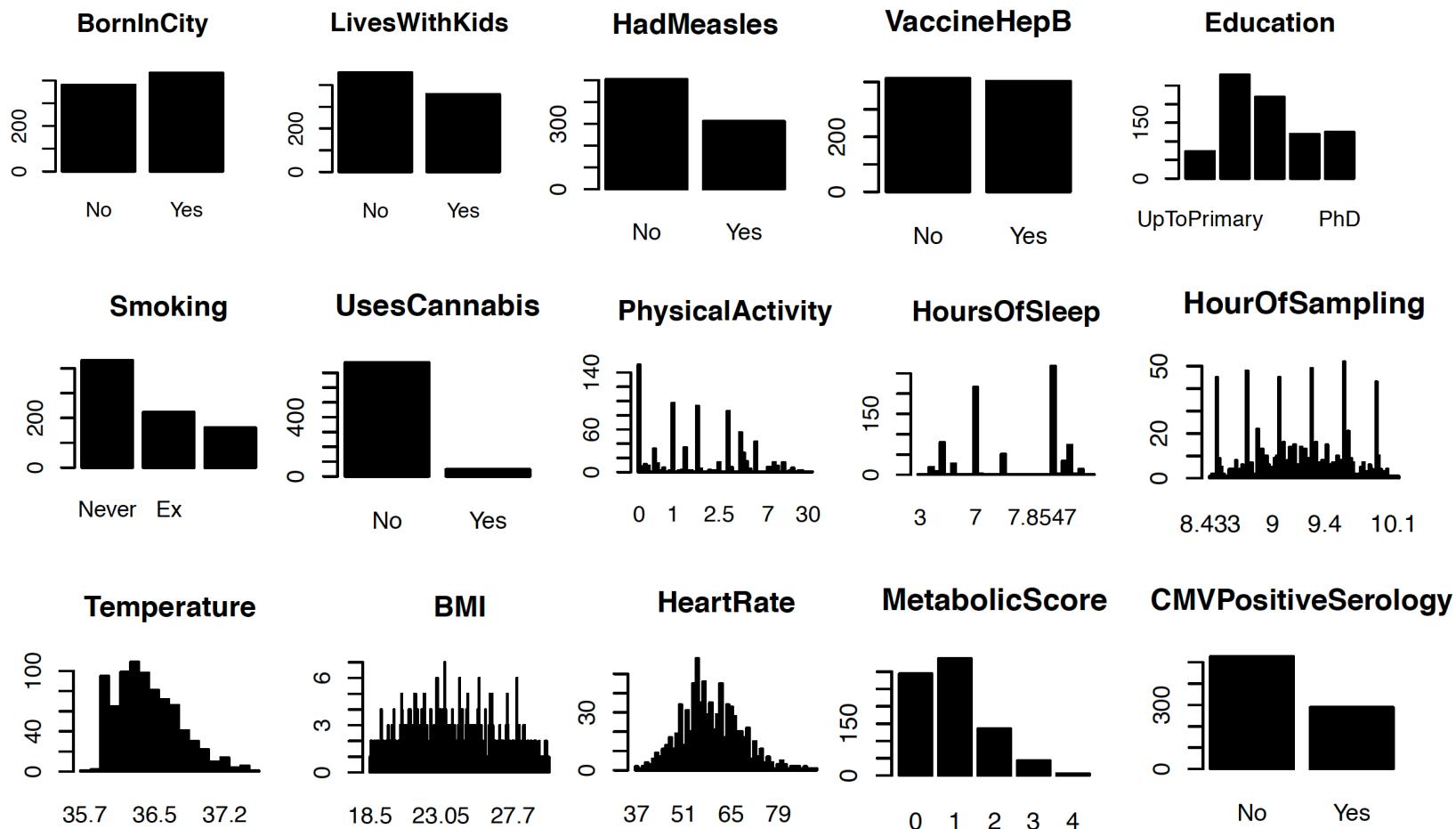
# Recrutement des donneurs



Etude clinique approuvée par le Comité de Protection des Personnes – Ouest 6 le 13 Juin 2012 et l'Agence Nationale Française de Sécurité du Médicament le 22 Juin 2012  
Le protocole est enregistré dans ClinicalTrials.gov (study#NCT01699893)

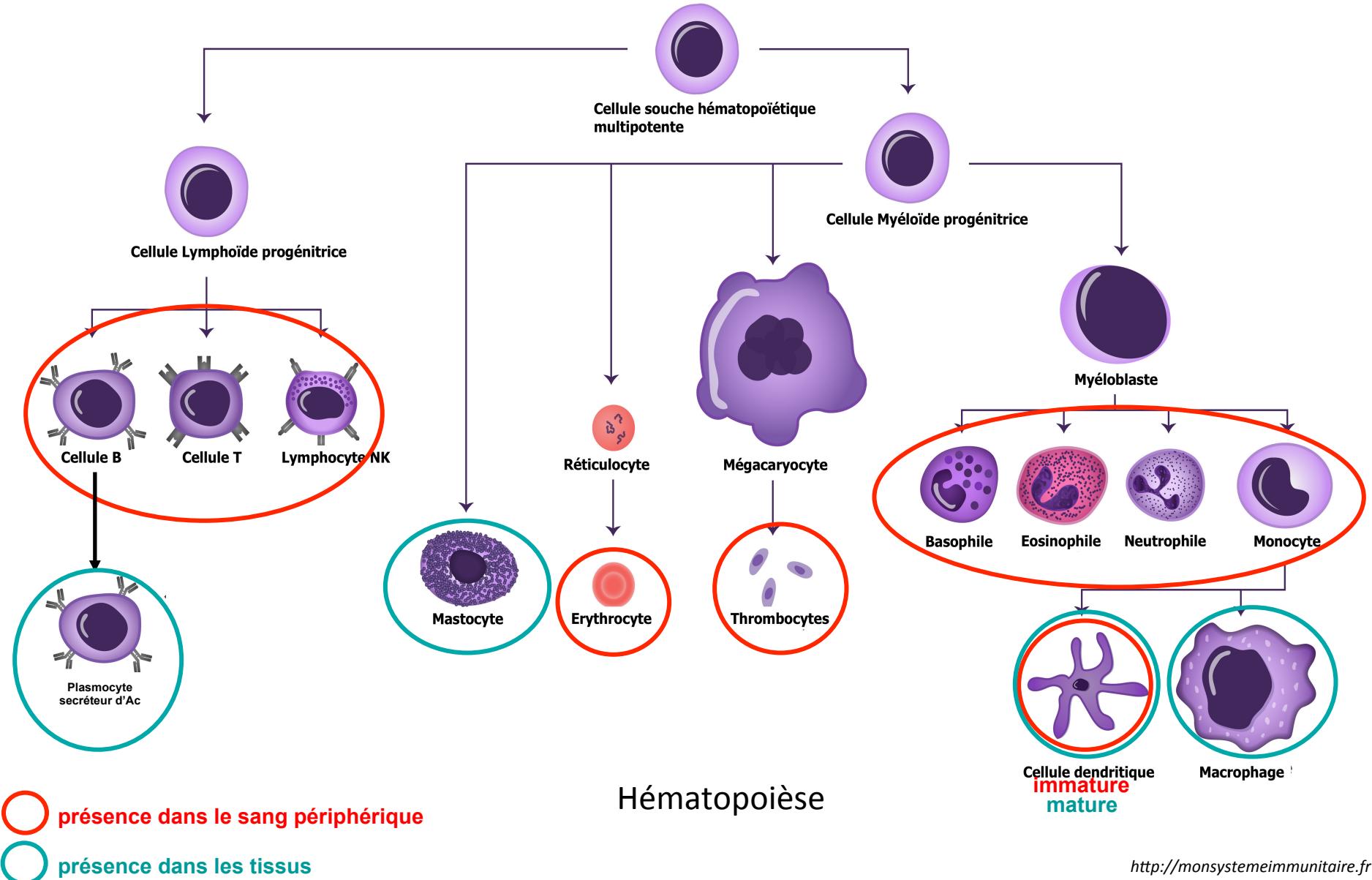
Thomas et al., Clinical Immunol 2015

# electronic Case Report Form (eCRF)



Mélange de variables numériques et catégorielles  
Imputation des données manquantes (package R missForest)

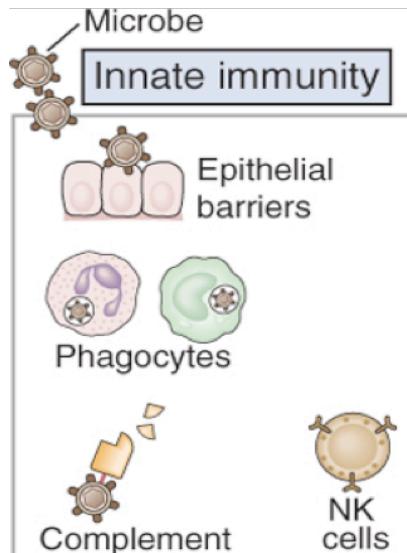
# Les cellules du système immunitaire



# Immunité innée et immunité acquise (adaptative)

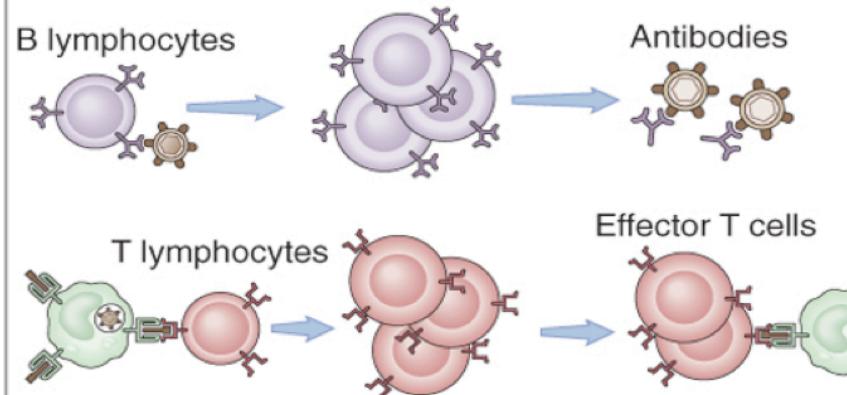
2 systèmes complémentaires et coopératifs

Réponse immunitaire innée:



Réponse immunitaire acquise:

Adaptive immunity



Réponse à  
médiation  
humorale

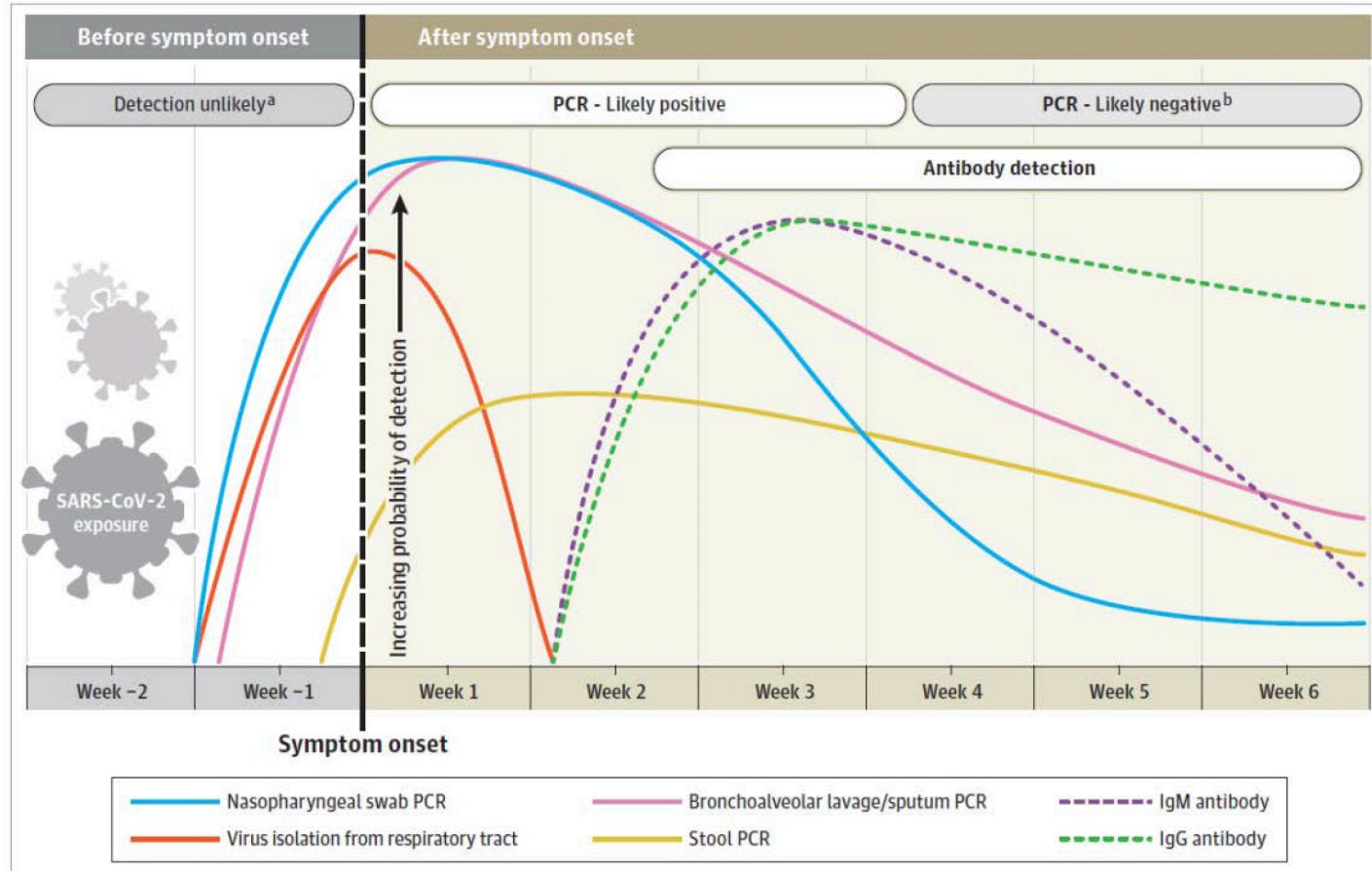
Réponse à  
médiation  
cellulaire



Rapide  
Générale  
Arrête l'infection  
Pas de mémoire

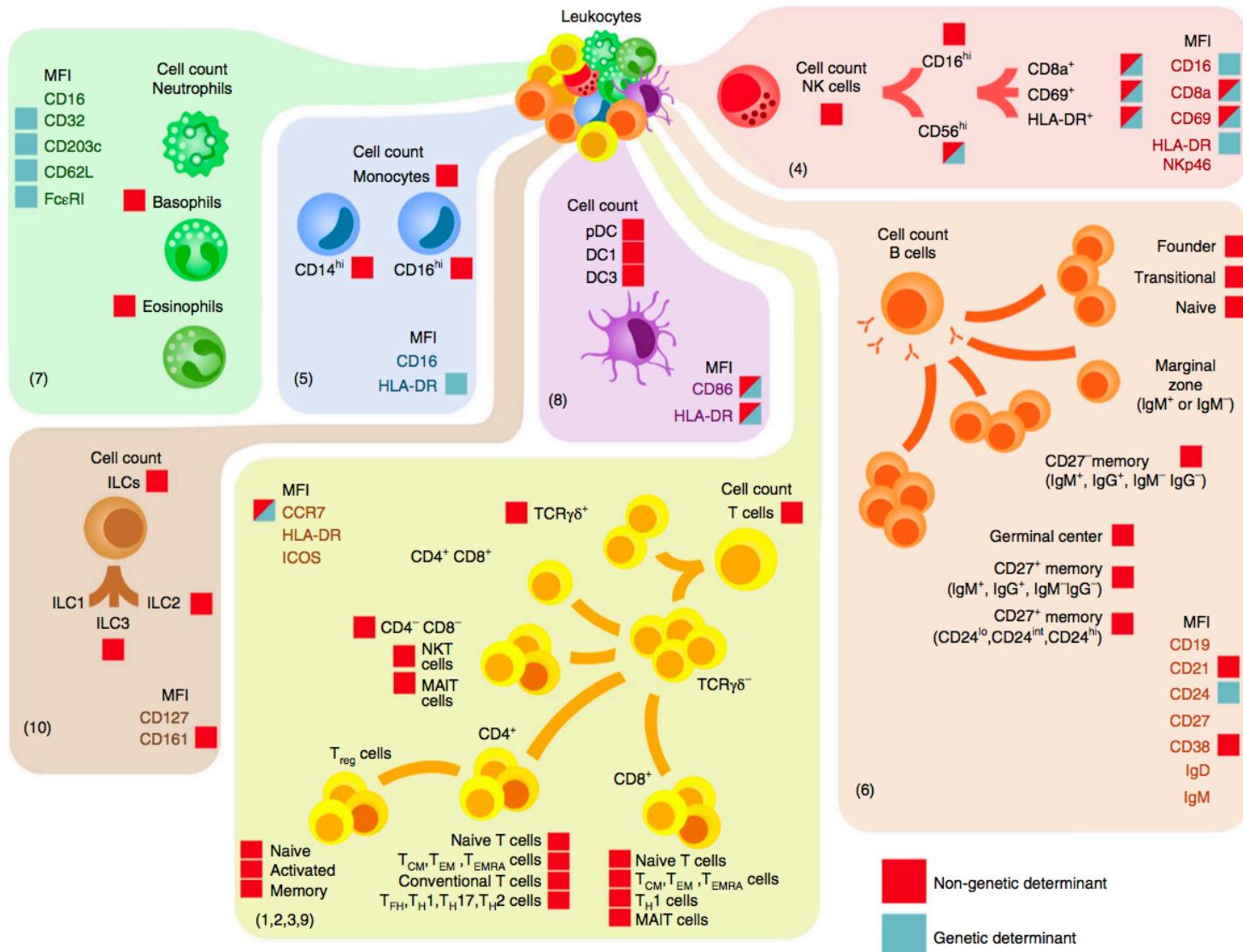
Plus lente  
Spécifique  
Nettoie l'infection  
Mémoire

# La réponse immunitaire primaire ou secondaire



La réponse immunitaire secondaire est plus rapide et plus efficace  
Les immunoglobulines produites lors de la réponse mémoire sont majoritairement des IgG

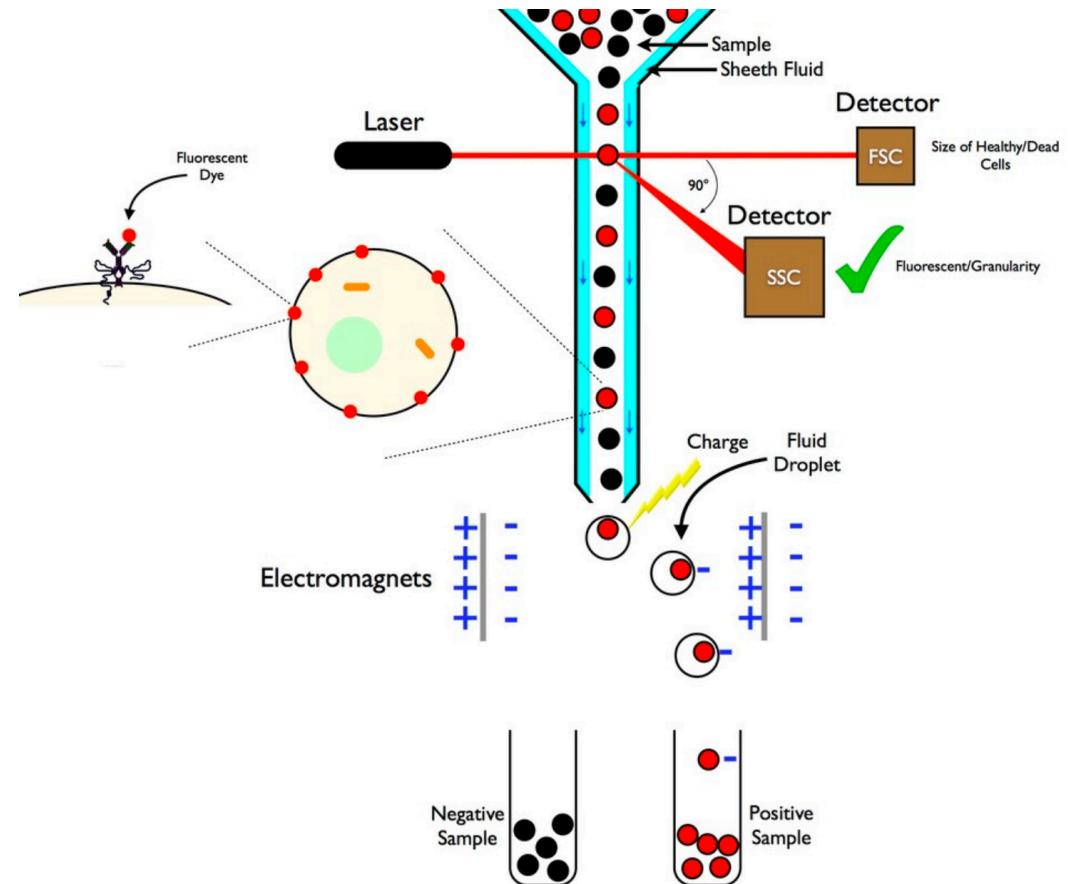
# Quantification de 76 populations cellulaires



# Quantification des cellules immunes par cytométrie de flux

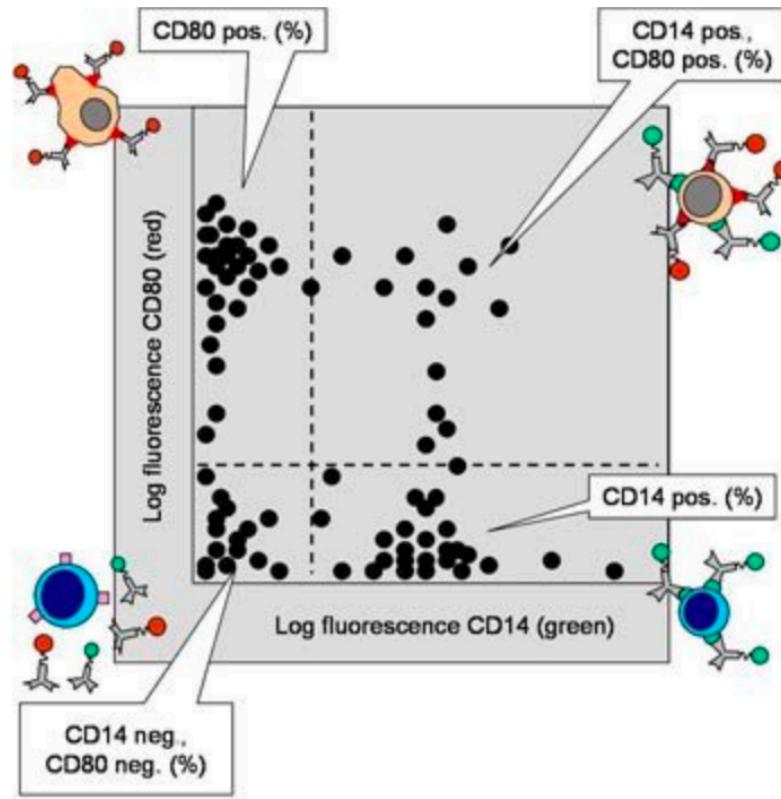
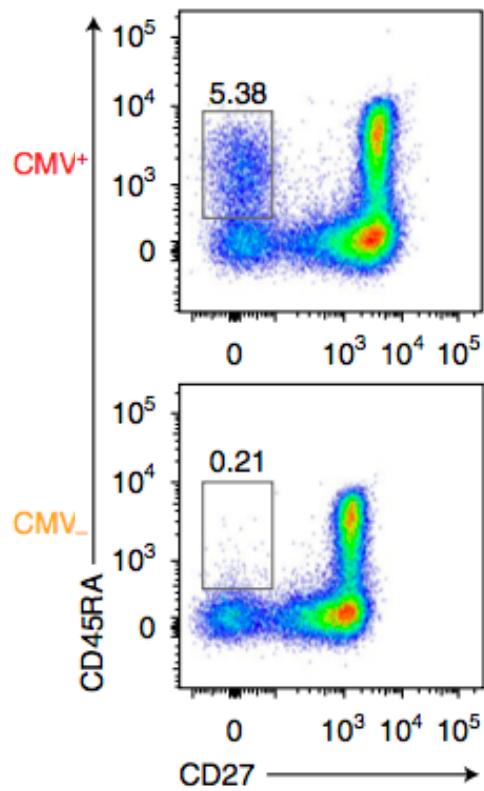
La composition en cellules immunitaires varie en fonction de l'âge et du sexe

Les cellules avec des protéines transmembranaires spécifiques peuvent être triées par Cytométrie de flux avec un « Fluorescence-Activated Cell Sorting » (FACS)



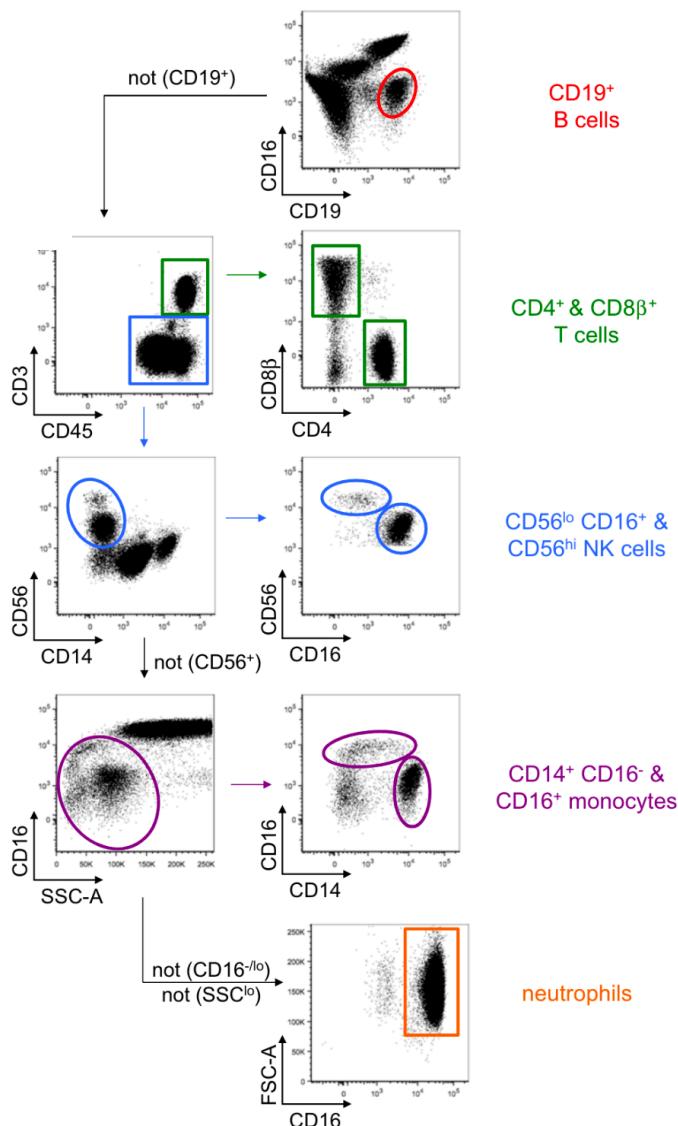
# Quantification des cellules immunes par cytométrie de flux

## Quantification des cellules à partir des résultats de FACS



# Quantification des cellules immunes par cytométrie de flux

Stratégie de sélection des populations cellulaires



Immunophénotypage cytométrique 8 couleurs

# Test de l'association des critères de l'eCRF avec les quantifications de populations cellulaires

---

```
# Pour chaque population cellulaire régression linéaire avec le critère d'intérêt et les co-variables age et sexe
```

```
fitm1 <- lm(DF[[i]] ~ DF[[crit]] + DF[['AGE.V0']] + DF[['SEX']], na.action=na.omit)
```

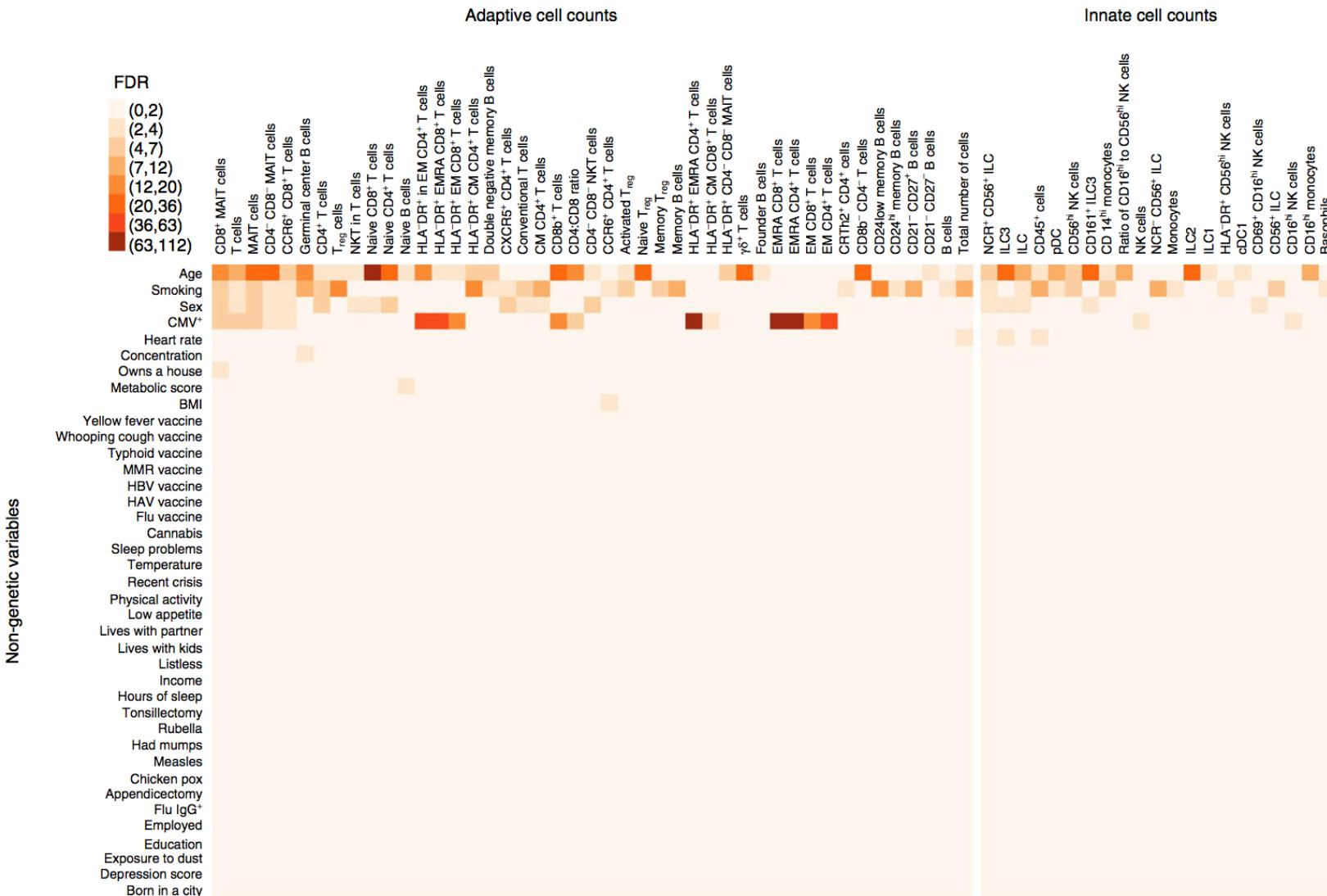
```
# Effectuer une régression linéaire sans le critère d'intérêt
```

```
fitm2 <- lm(DF2[[i]] ~ DF2[['AGE.V0']] + DF2[['SEX']], na.action=na.omit)
```

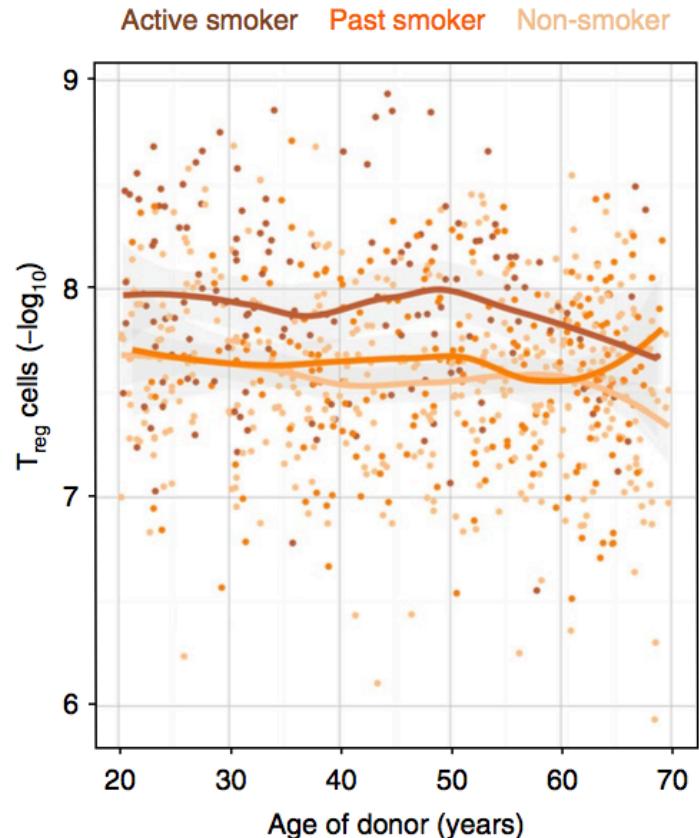
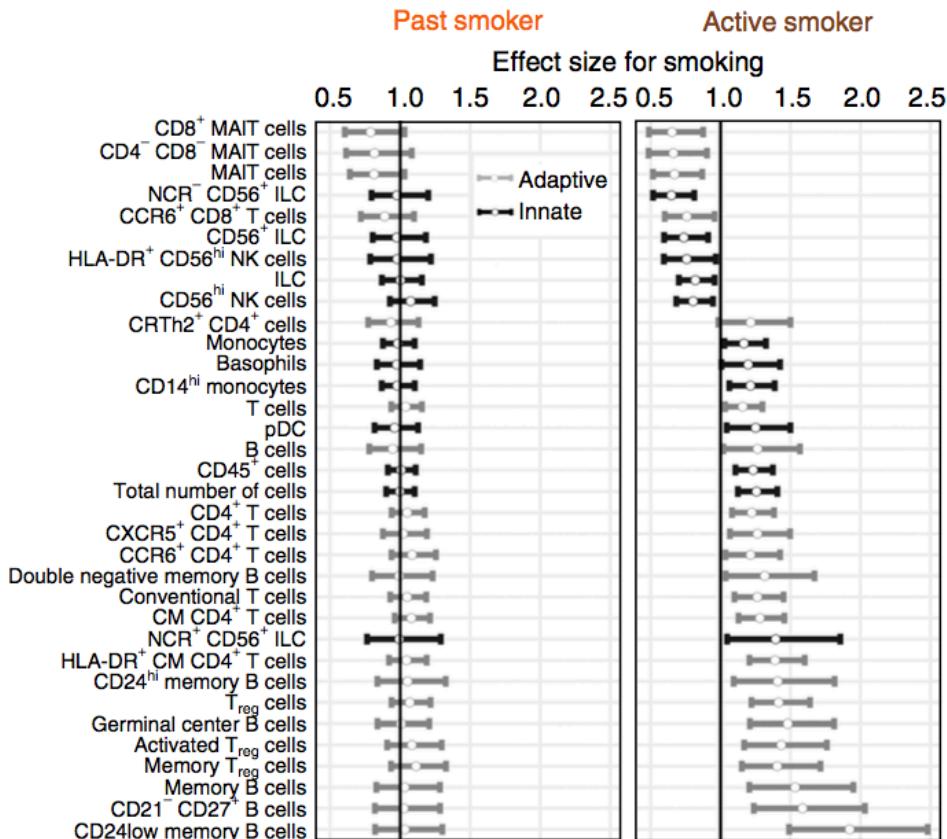
```
# tester si l'ajout du critère dans la régression entraîne une différence significative dans l'explication de quantification cellulaire
```

```
fit <- anova(fitm1, fitm2, test = 'Chisq')
```

Mise en évidence de l'association de l'âge, du sexe, du tabac et de l'infection au CMV avec les populations cellulaires sanguines

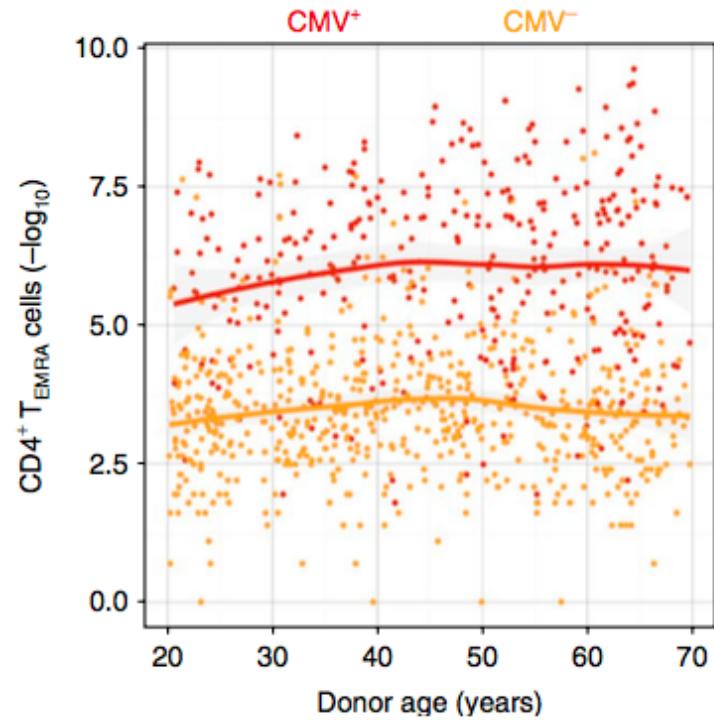
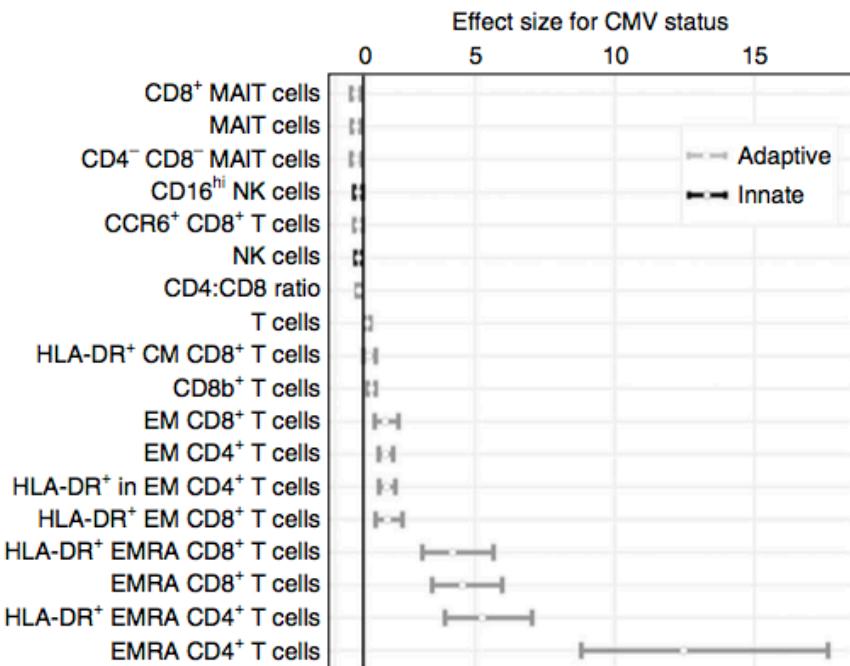


# Mise en évidence de l'effet du tabac sur la composition cellulaire sanguine



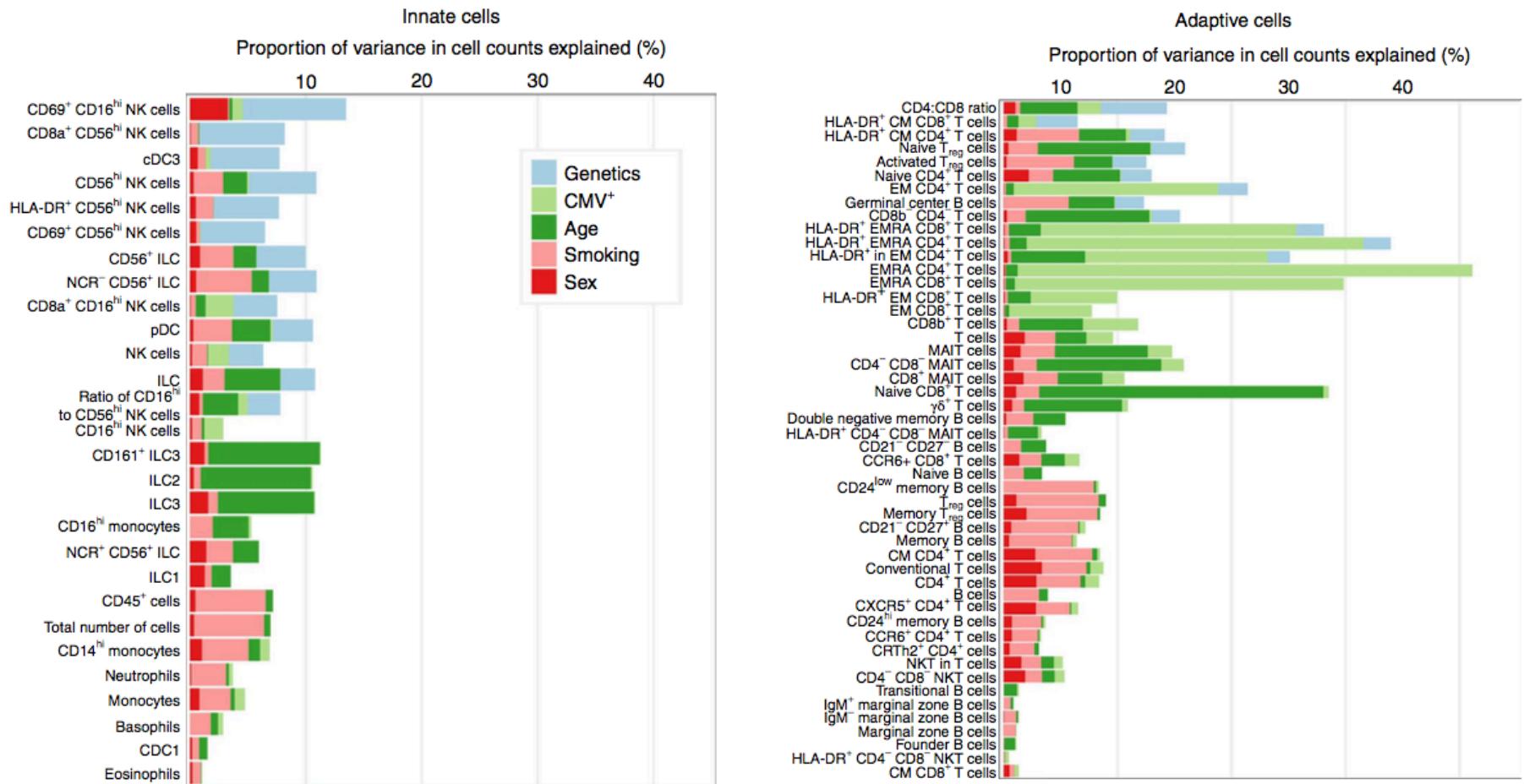
Le tabagisme actif a un effet sur l'immunité similaire en magnitude à celui de l'âge et il affecte à la fois les cellules innées et adaptatives

# Mise en évidence de l'effet de la sérologie positive au cytomegalovirus (CMV) sur la composition cellulaire sanguine



L'infection latente avec le CMV induit un changement majeur dans le nombre de cellules T mémoires indépendant de l'effet de l'âge

# Contribution des différents facteurs associés avec la composition en populations de cellules sanguines



La variation des nombres de cellules de l'immunité innée est surtout corrélée à la génétique  
 La variation des nombres de cellules de l'immunité acquise est préférentiellement liée aux expositions environnementales

# Stimulation immune du sang des 1000 donneurs

---

## TruCulture whole blood assay

Échantillon de sang des 1000 donneurs



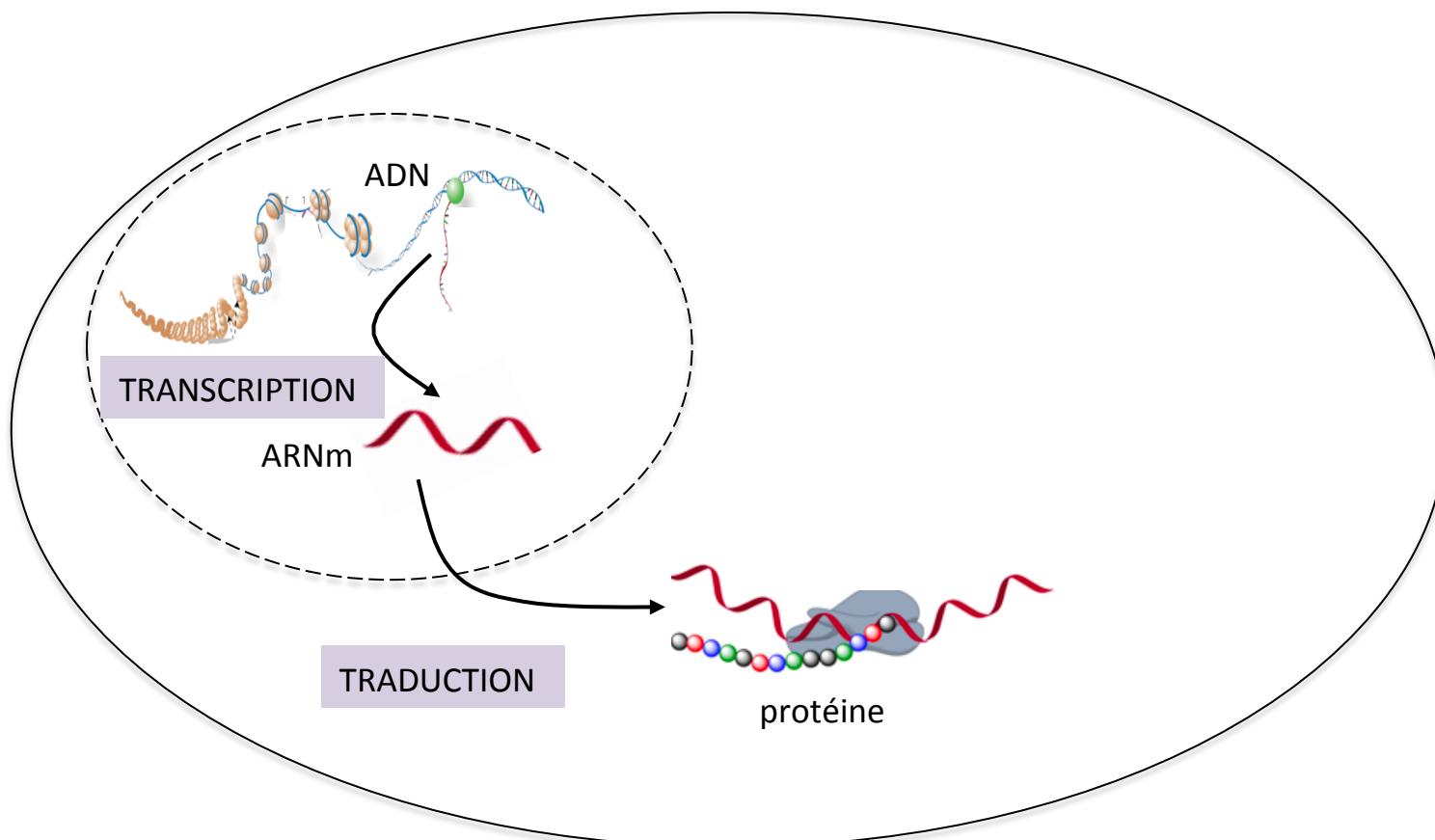
### Stimulations:



- Bactériens (*E. coli*, *S. aureus* et BCG)
- Fongique (*C. albicans*)
- Viral (Influenza)
- Activateurs des cellules T (SEB)

# Quantification des niveaux d'ARNm

Chaque population cellulaire dans une condition de stimulation donnée possède un profil d'expression spécifique qui peut être étudié par la quantification des niveaux d'ARNm présents dans la cellule

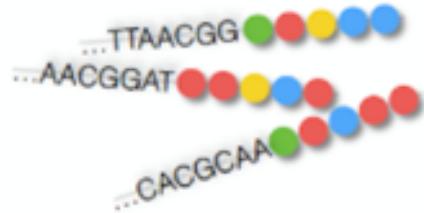


# Quantification des niveaux d'ARNm de 560 gènes de l'immunité par la technologie Nanostring

---



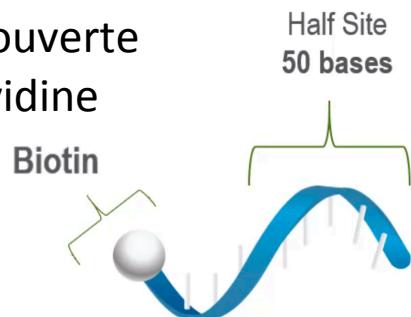
Expression levels of  
560 immune-related genes



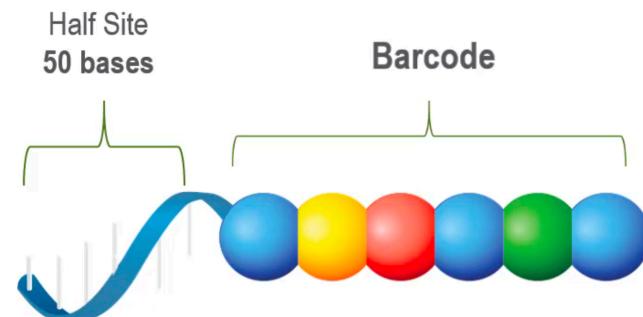
# Quantification des niveaux d'ARNm

## Technologie Nanostring

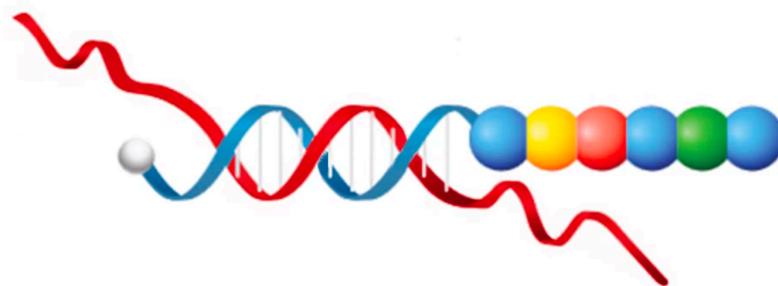
Pour liaison à la plaque recouverte de streptavidine



Target-specific **Capture** Probe



Target-specific **Reporter** Probe

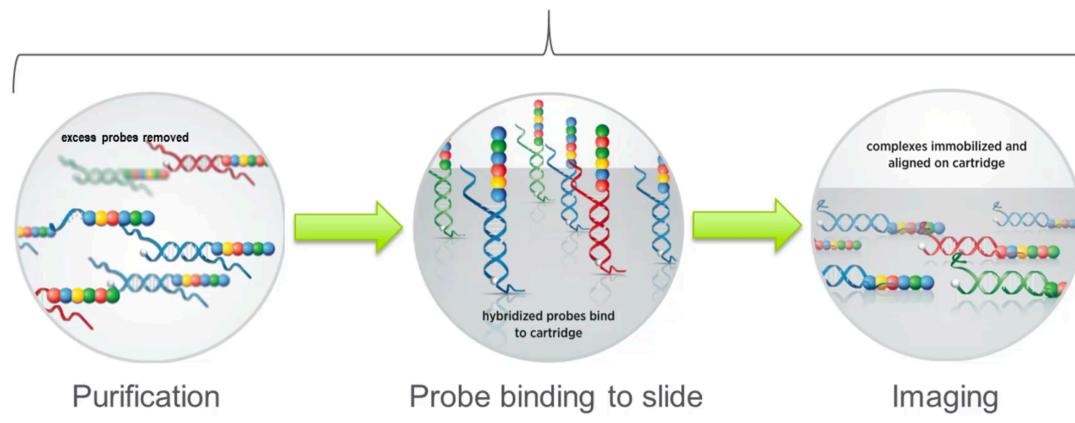


Barcode	Identity
● ● ● ● ●	XLSA
● ● ● ● ●	FOX5
● ● ● ● ●	PDCD1

Code barre moléculaire pour l'identification des ARNm:  
combinaison unique de 4 couleurs x 6 marqueurs fluorescents

# Quantification des niveaux d'ARNm

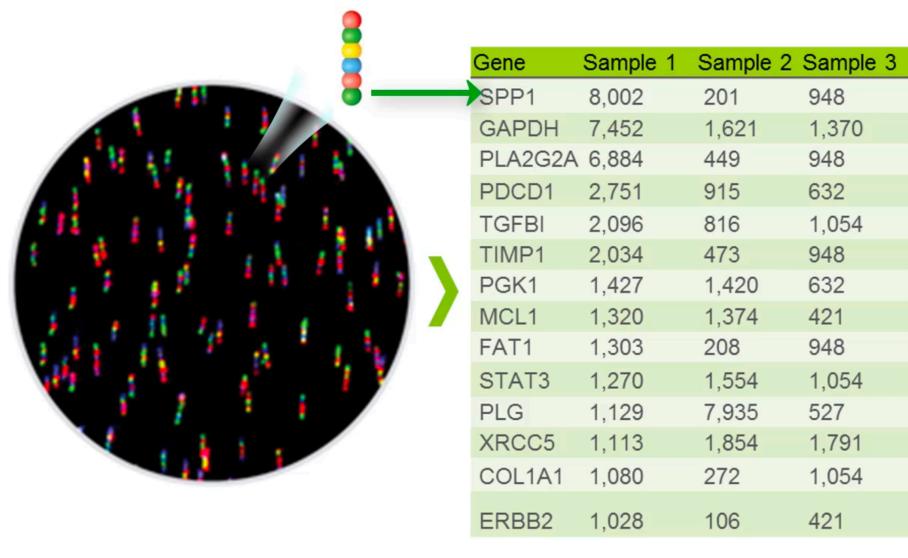
## Automated instrumentation



Pas besoin d'amplification PCR

Application d'un courant électrique pour aligner les molécules

Compter le nombre de fois que le code bar spécifique de SPP1 est présent sur la plaque



## nSolver software:

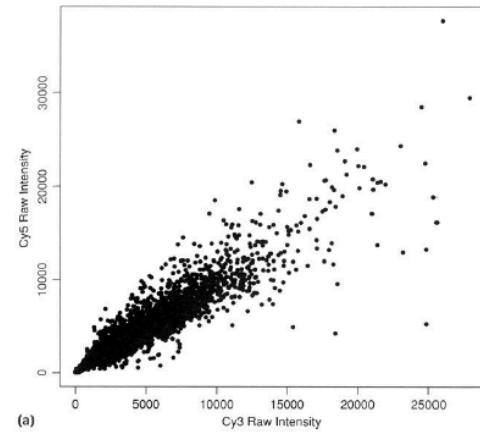
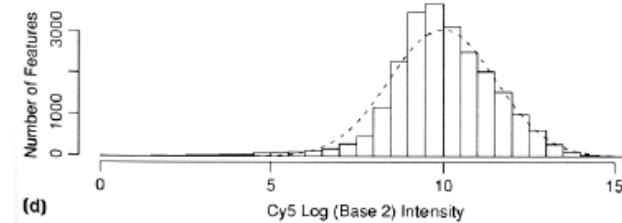
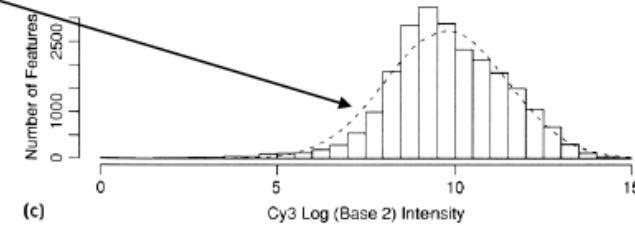
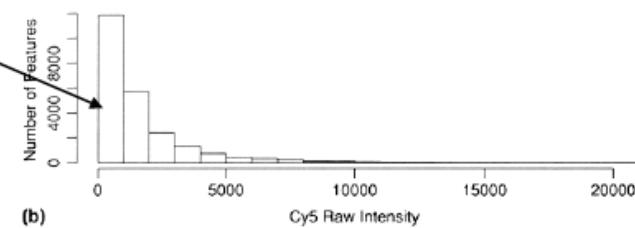
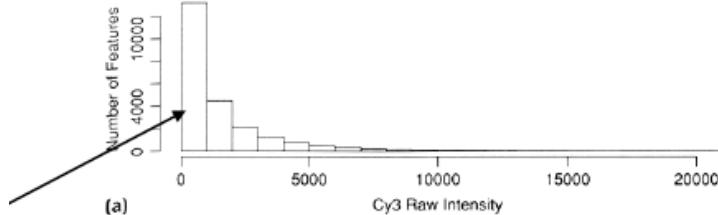
Filtre sur la qualité  
normalisation par rapport à des controles internes et un groupe de gènes de référence dont l'expression ne change pas en fonction de la stimulation ("house keeping genes")  
Comptes absolus → comptes relatifs entre échantillons

# Quantification des niveaux d'ARNm

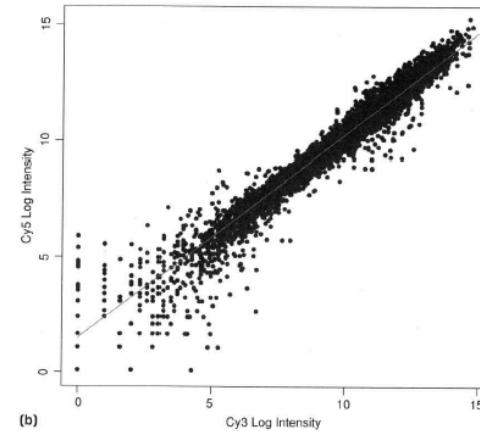
## Transformation log2 des données

La plupart  
des  
intensités  
sont faibles

Log2  
normalisation



Augmentation de la  
variabilité avec l'intensité

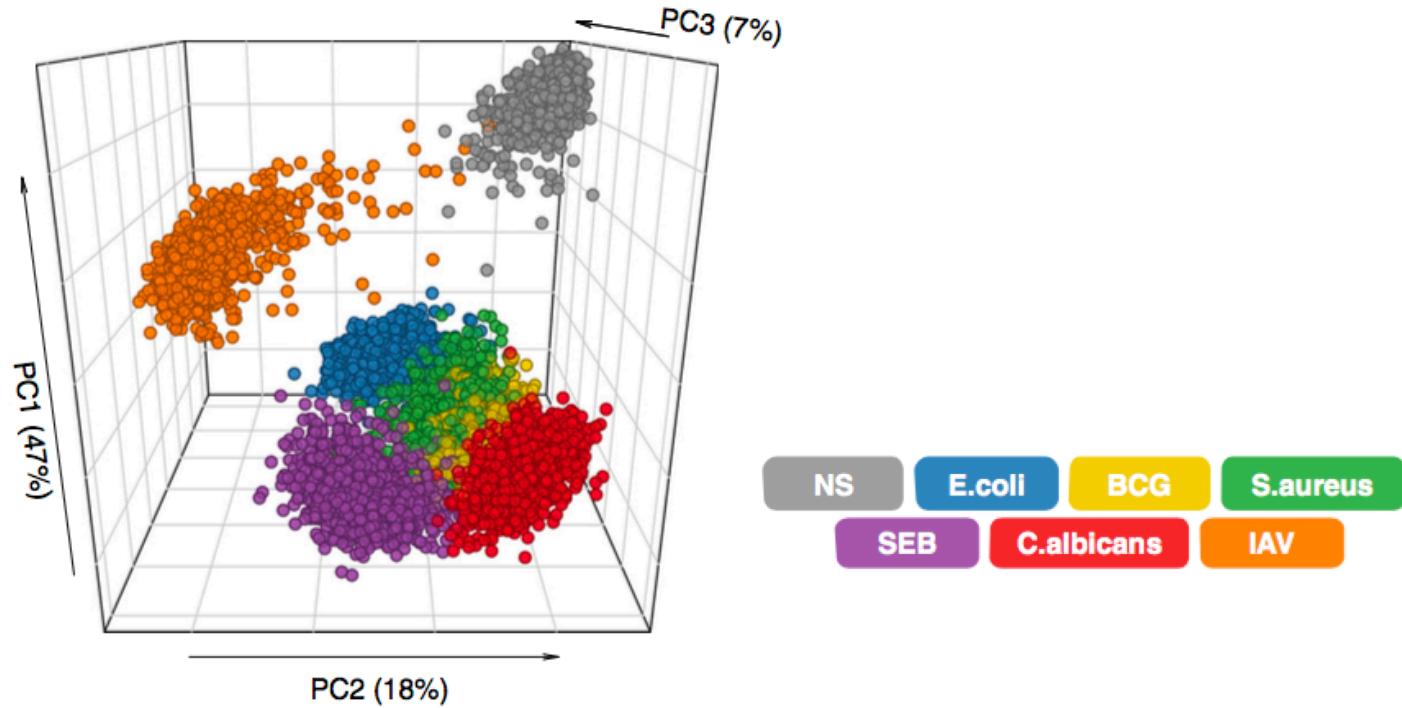


Variabilité constante

# Réponses transcriptionnelles différentes en fonction du type de stimulation

Les stimulations ont modifié les niveaux d'ARN d'environ 90% des gènes testés, la plupart étant “up-régulés” en condition de stimulation

Augmentation  
des niveaux  
d'ARNm dans  
le plupart des  
stimulations

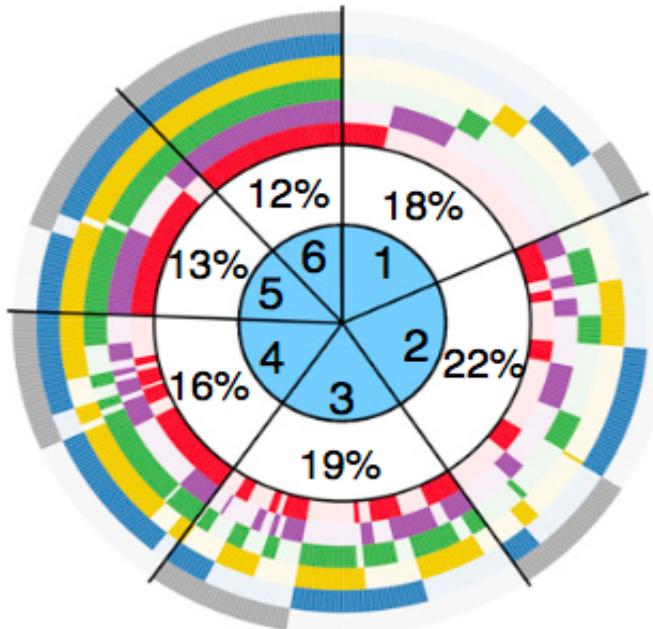


Signature plus spécifique  
des gènes induits par une  
des stimulations

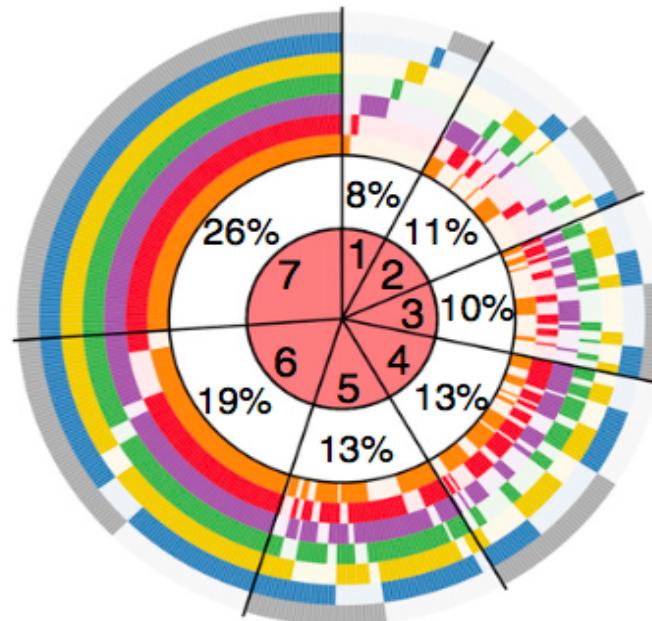
# Effet de l'age et du sexe des individus sur la variation des niveaux d'ARNm

L'age et le sexe affectent les niveaux d'ARNm de la plupart des gènes de l'immunité  
Les effets de l'age sont plus spécifiques du stimulus que ceux du sexe

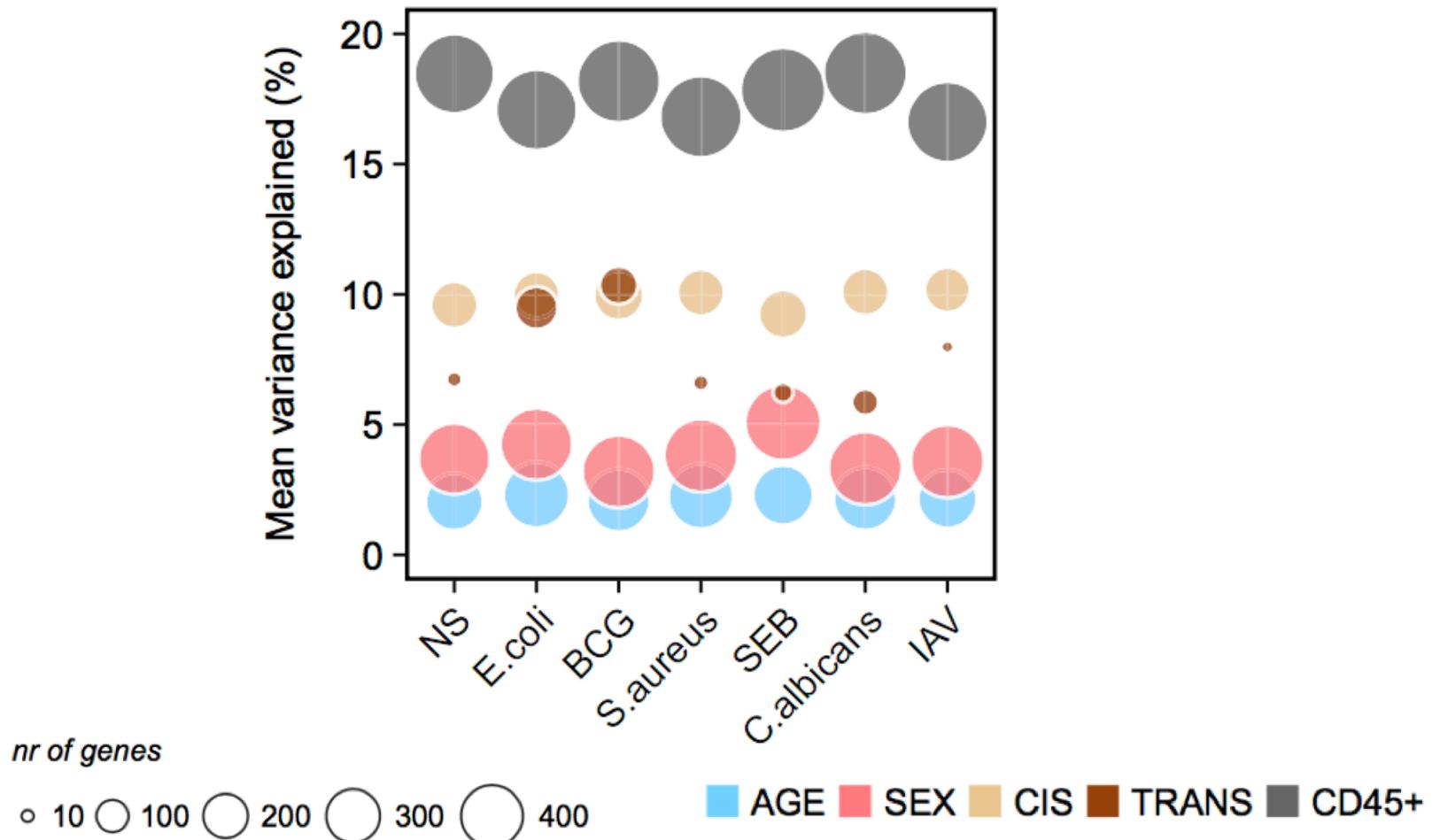
**Spécificité de l'effet de l'age**



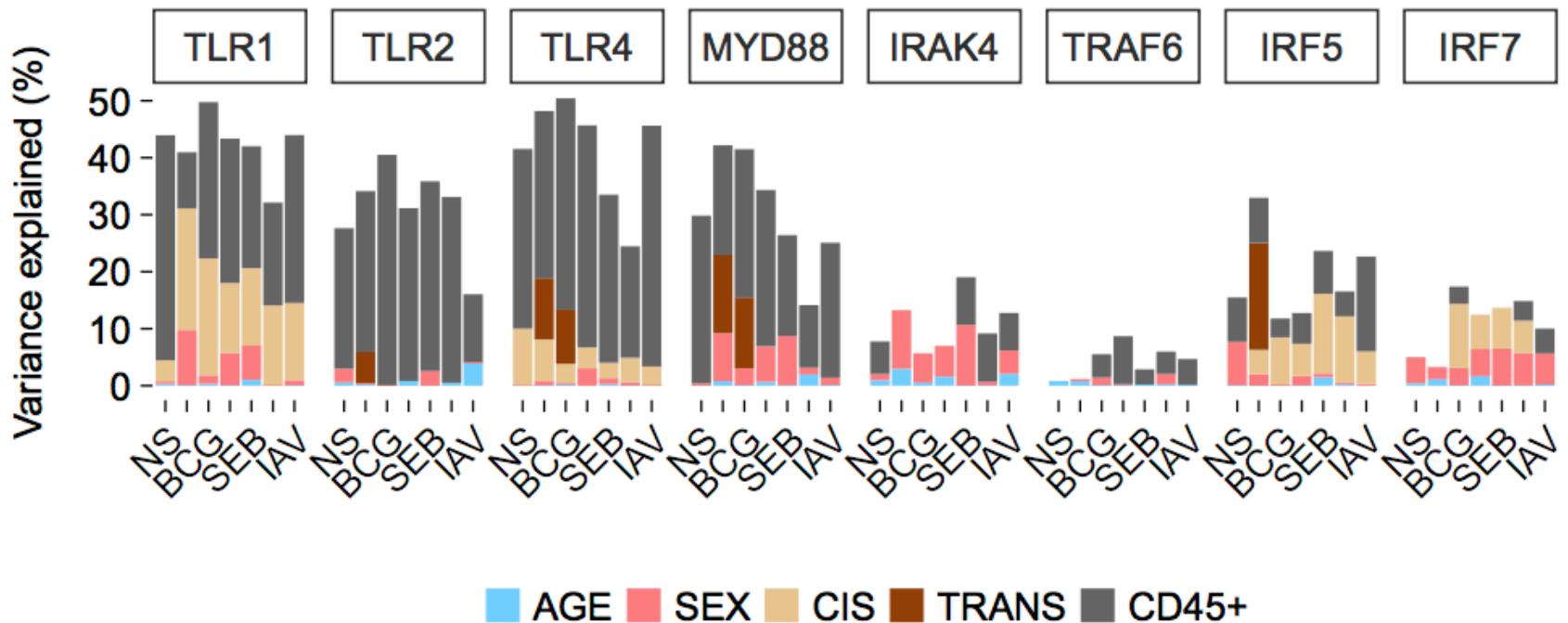
**Spécificité de l'effet du sexe**



# Variance des niveaux d'ARNm expliquée par l'âge, le sexe et la composition en cellules sanguines



# Variance des niveaux d'ARNm expliquée par l'age, le sexe et la composition en cellules sanguines



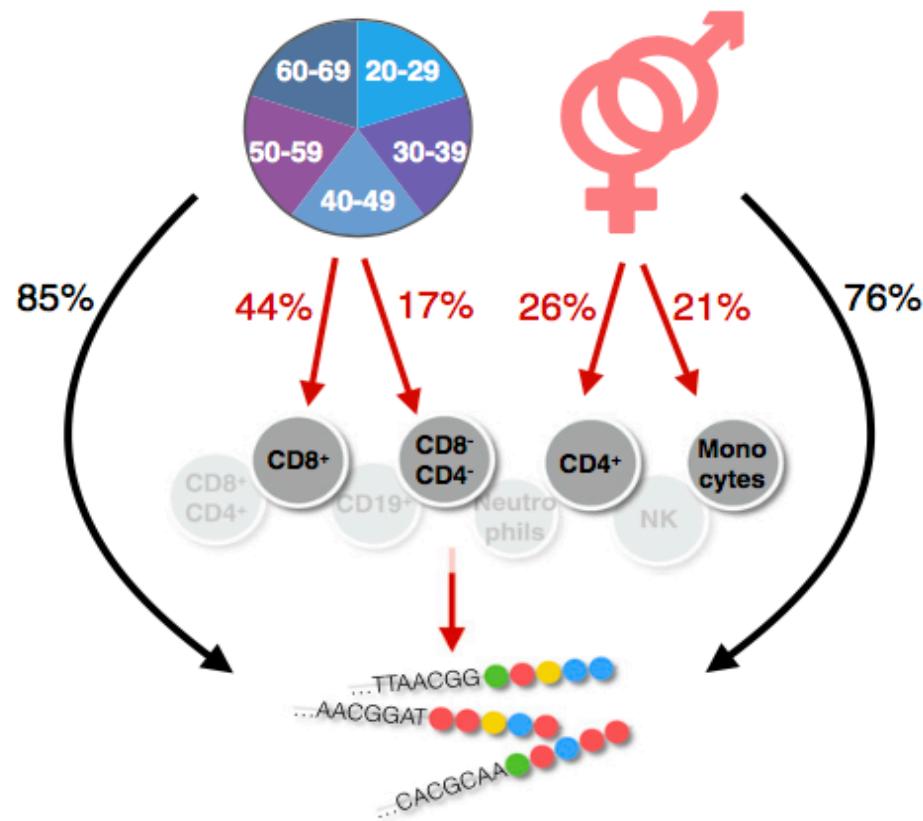
La composition cellulaire a un impact fort sur la variation des niveaux d'ARNm des récepteurs comme les TLR 1/2/4

L'effet de l'âge ou du sexe est plus important pour des ARNm situés plus en aval dans les voies de régulation

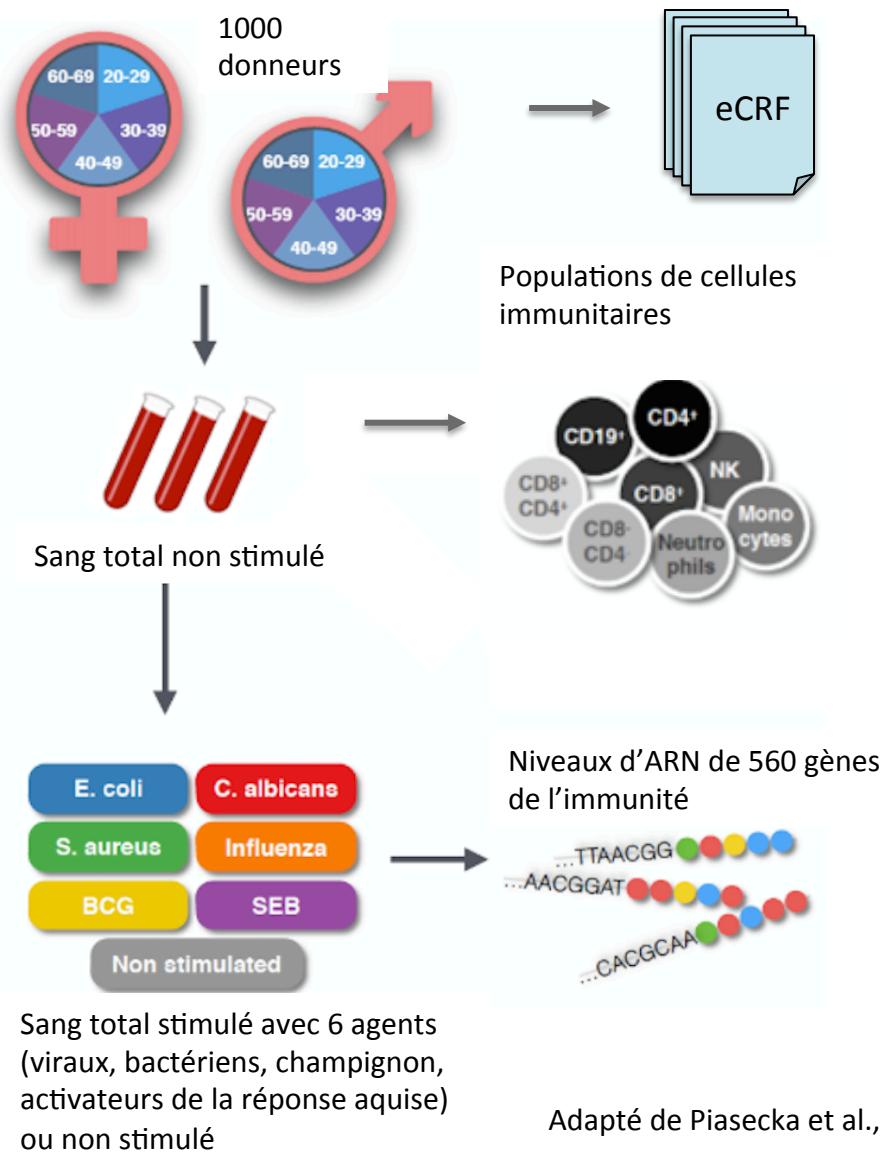
# Médiation des effets de l'âge et du sexe sur les niveaux d'ARNm par la composition en cellules sanguines

La variation des niveaux d'ARNm en fonction de l'âge et du sexe est-elle due aux changements de proportions cellulaires induites par la stimulation?

Pour chaque gène dont les niveaux d'ARNm sont associés à l'âge, construction d'un modèle avec 8 effets indirects et un effet direct de l'âge avec le sexe et les variables techniques en co-variables. Pour chaque gène les effets directs et indirects peuvent ainsi être observés en sortie du modèle.

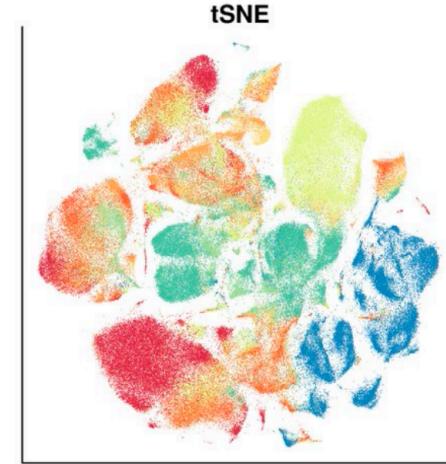
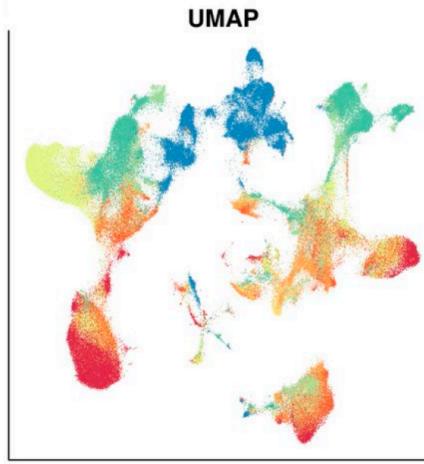
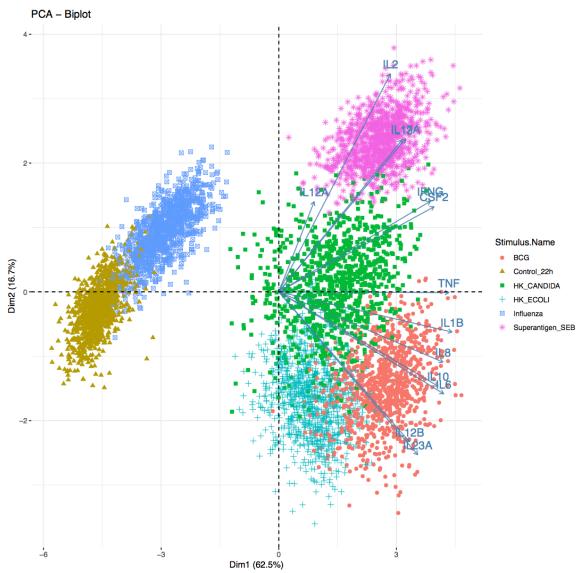


# Analyse et intégration de grands jeux de données



# Axe 1: Visualiser les données avec des méthodes de réduction de dimensions : ACP, tSNE, multibloc etc...

---



## code de départ:

Code R pour réaliser une ACP avec FactoMineR et factoextra sur les données de quantification d'ARNm dans les différentes conditions de stimulation

## pistes:

Comparer l'ACP avec celle publiée dans Piasecka et al.

Produire d'autres figures à partir de l'ACP et les interpréter.

Explorer d'autres méthodes de réduction de dimensions des données sur ce jeu de données.

# Axe 2: Visualiser la corrélation et le regroupement des variables

## code de départ:

Génération de dendogrammes et extraction de groupes.

Regroupement hiérarchique des ARNm dans une condition de stimulation avec coloration spécifique par une variable.

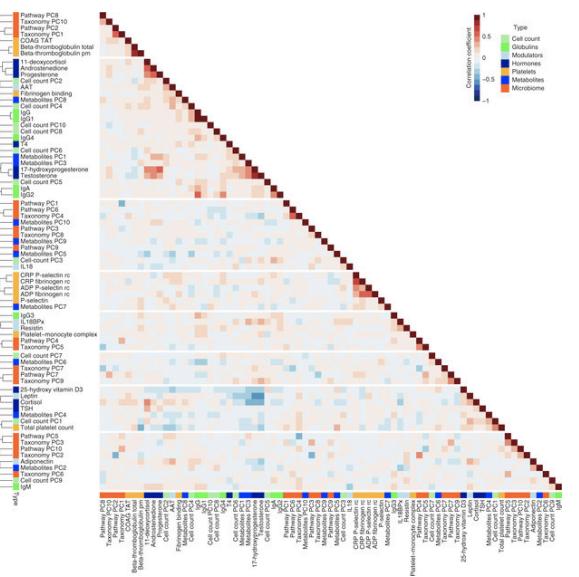
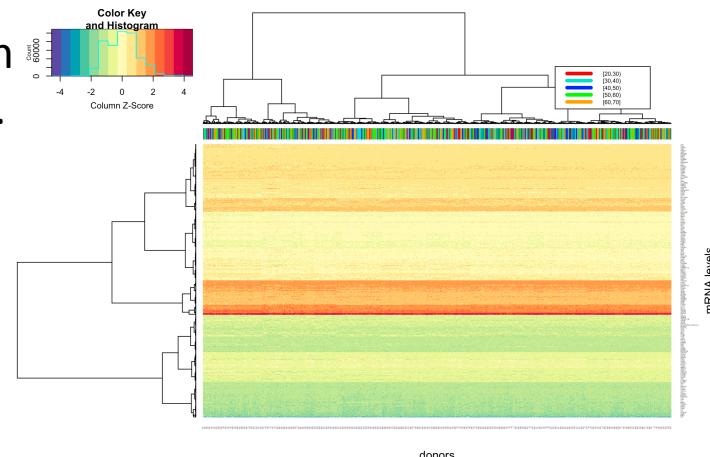
Génération d'une matrice de corrélation entre les niveaux d'ARNm dans deux conditions de stimulation

## pistes:

Réaliser des regroupements hiérarchiques des niveaux d'ARNm dans différentes conditions de stimulation.

Comparer les niveaux d'ARNm les plus variables dans différentes stimulations. Peut-on identifier des groupes d'individus qui ont des réactions immunes similaires dans plusieurs stimulations immunes ?

Générer des matrices de corrélation des variables de l'eCRF entre elles, entre les niveaux d'ARNm dans différentes conditions de stimulation, entre les variables de l'eCRF et les ARNm dans une des conditions de stimulation ou entre les niveaux des ARNm dans une condition de stimulation et les compositions cellulaires par exemples.



# Axe 3: Estimer les réseaux de régulation à partir des données de quantification d'ARNm et les comparer entre les différentes situations de stimulation

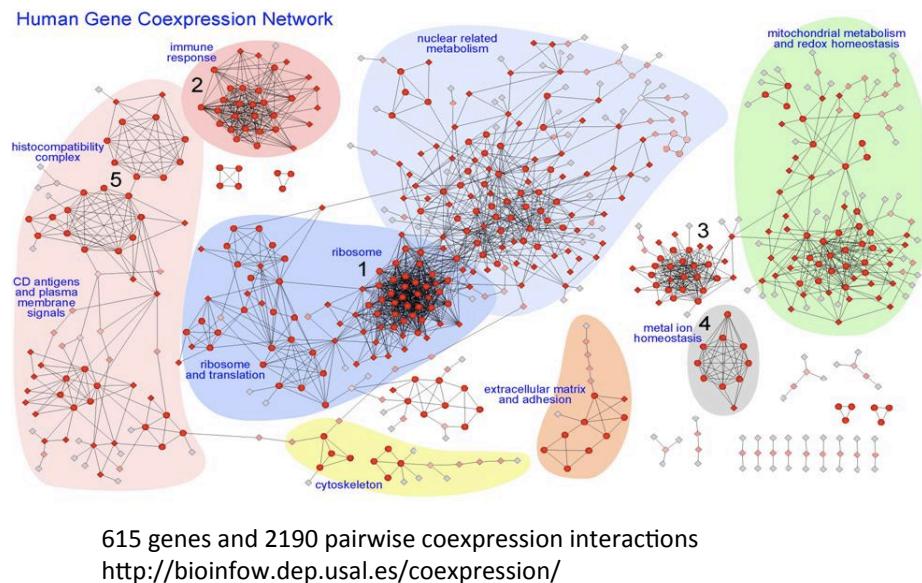
## code de départ:

Code pour construire des réseaux simples avec le package R igraph

## pistes:

Générer des réseaux d'expression géniques pour les différentes conditions de stimulation. Les comparer et produire des représentations graphiques pertinentes.

Utiliser des bases d'ontologies biologiques pour interpréter les résultats en terme de voies de régulations biologiques.



Axe 4: Tester l'association des niveaux d'ARNm avec les données de l'eCRF. Si des associations sont identifiées, tester si celles-ci semblent médiées par certaines populations cellulaires

## code de départ:

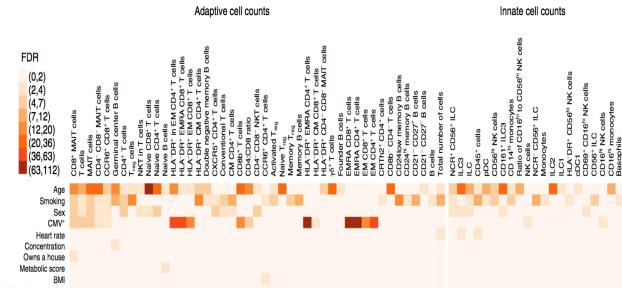
Code pour effectuer des regressions linéaires et anova pour tester les associations de variables numériques ou catégorielles du questionnaire avec les niveaux d'ARNm avec ou sans co-variables cellulaires.

Code pour utiliser l'algorithme de machine learning Random Forest qui permet de mettre en avant les variables qui contribuent le plus à la prédiction de la réponse.

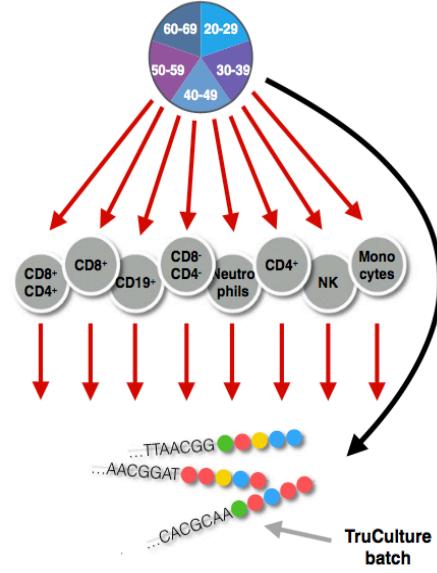
pistes:

Utiliser ces bases de code sur un grand nombre de variables et résumer les résultats de façon graphique et pertinente.

Une fois identifiées les variables les plus associées, tester une possible médiation de celles-ci par des populations cellulaires. Utiliser éventuellement d'autres méthodes pour faire le même type de tests.



Patin et al., Nat. Immunol. 2018



Piasecka et al., PNAS Supdata 2018

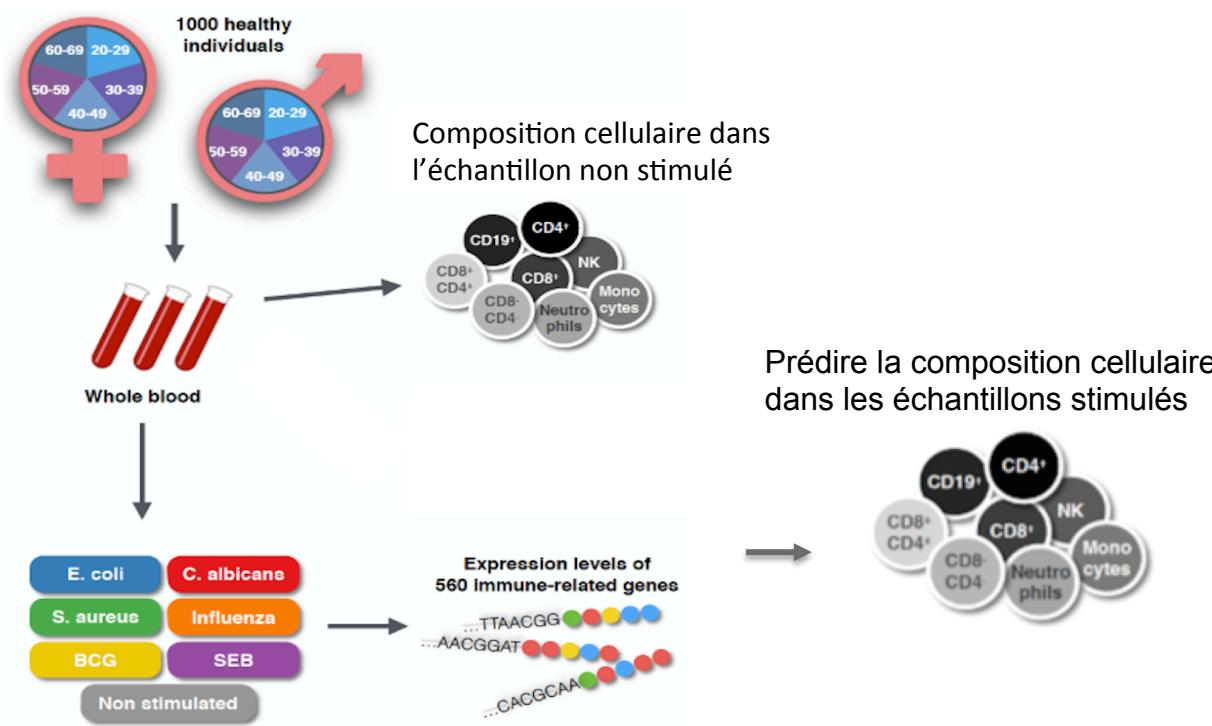
# Axe 5: Construire un modèle de prédiction de la composition cellulaire à partir des données de quantification d'ARNm de la condition non stimulée et utiliser ce modèle pour estimer la composition cellulaire dans les échantillons stimulés

code de départ:

Vignette et exemple de code avec le package `glmnet`

pistes:

Tester différents modèles de prédiction de composition cellulaire, comparer les résultats et discuter les méthodes.



# Analyse et Intégration de données du projet « Milieu Intérieur » coordonné par l’Institut Pasteur

Violaine Saint-André – Institut Pasteur  
HUB-C3BI  
[violaine.saint-andre@pasteur.fr](mailto:violaine.saint-andre@pasteur.fr)

