|  |
| --- |
| HEIG-VD |
| BBC – Projet |
| Test miARN pour diagnostiquer le cancer de la prostate à partir de l’urine |

|  |
| --- |
| Thibault Schowing & Jan Purro  [Date] |

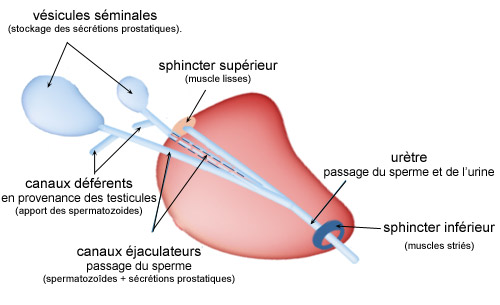
# Introduction

Le but de ce travail est de trouver un moyen informatique de diagnostic du cancer de la prostate à partir des urines. Dans le cadre du cours de Bio-informatique et Biologie Computationnelle, nous effectuerons des recherches sur le cancer de la prostate et les moyens de détection actuels, puis nous orienterons la recherche sur les relations de la miARN (microARN) avec la mécanique cellulaire.

Bla bla statistiques analyse dataset etc

## La prostate

La prostate est une glande de l’appareil génital masculin dont la principale fonction est de sécréter et stocker une partie du liquide séminal qui constitue le sperme. Elle se trouve à l’intersection des voies urinaires et spermatiques, l’urètre et les canaux déférents (via les vésicules séminales).



## Le cancer de la prostate

Le cancer se développe à partir des tissus de la prostate. Une mutation des cellules provoque une multiplication incontrôlée des cellules qui peuvent ensuite se métastaser dans d’autres organes, en particulier les os et les ganglions lymphatiques. Dans une immense majorité des cas, il s’agit de tumeurs malignes pouvant se propager (métastases).

Il existe aujourd’hui deux principaux moyens de détection. Le toucher rectal, afin de détecter une croissance anormale de la prostate, et l’analyse du taux de PSA dans le sang (Antigène Prostatique Spécifique, c’est une protéine fabriquée exclusivement par la prostate). Des symptômes extérieurs apparaissent généralement lorsque le cancer est à un stade avancé.

## L’urine

L’urine, composée à 95% d’eau, sert à évacuer les déchets métaboliques issus de la filtration du sang par les reins. L’urine contient des éléments tels que l’urée, issus de la dégradation de certains acides aminés, et l’acide urique, issus de la dégradation des acides nucléiques.

Afin d’être évacué, l’urine passe des reins, à la vessie puis, via la prostate chez l’homme, dans l’urètre.

## Micro ARN

Les micro-ARN sont de courts acides ribonucléiques simples-brins d’environ 21-24 nucléotides. Ce sont des régulateurs post-transcriptionnel permettant l’extinction de l’expression d’un gène. Ils s’apparient à une séquence complémentaire d’ARNm et empêche la traduction de l’ARNm ou déclenche sa dégradation. Plus de 1000 gènes seraient à l’origine de transcription de miARN chez l’humain. Ils sont présents dans un très grand nombre de cellules.

Ce micro-ARN a une fonction de spécialisation des cellules. Toutes les cellules contenant la totalité du génome, il faut empêcher, par exemple, les cellules de peau de fabriquer de l’acide gastrique, et les cellules de l’estomac de fabriquer de la mélanine. C’est ici qu’intervient la répression de la miARN et ciblant quels gènes doivent s’exprimer et les quels ne le doivent pas.

### miARN et cancer prostatique

Le cancer prostatique est le plus répandu chez l’homme aux États-Unis, il affecte dans de plus grosses proportions les afro-américains que les américains caucasiens. Malgré cela, aucun marqueur n’a été affecté à la sévérité ou aux différences ethniques de la maladie.

Dans ce cadre, la miARN représente une nouvelle classe prometteuse de biomarqueurs grâce à sa stabilité et sa résistance naturelle. Une étude de 2013, «  MicroRNA Profiling in Prostate Cancer - The Diagnostic Potential of Urinary miR-205 and miR-214 », a été menée sur une quarantaine d’américains. Différents prélèvements ont été effectués dans les tissus cancéreux et les tissus sains pour comparaison et une différence des taux de miARN a été observée sur plusieurs d’entre elles. Ces analyses de miARN ont aussi été effectuées dans les urines, afin de déterminer s’il serait possible de s’en servir comme une technique de dépistage non-invasive.

Huit miARN ont été sélectionnées pour l’analyse dans des microarray : miR-205, mir-214, miR-221 et miR-99b étaient en quantité insuffisante dans les tissus cancéreux. Le miR-99b est présent en quantités encore moindre chez les afro-américains par rapport aux caucasiens, ce qui pourrait lier ce miARN à l’agressivité plus importante du cancer dans cette population.

Dans l’urine, miR-205 et miR-214 sont en quantités anormalement basses chez les patients cancéreux et peut permettre de distinguer un individu sain d’un individu cancéreux avec 89% de sensibilité et 80% de spécificité (vrais/faux, positif/négatif).

Source : <http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0076994>

## Les Biopuces et microarray

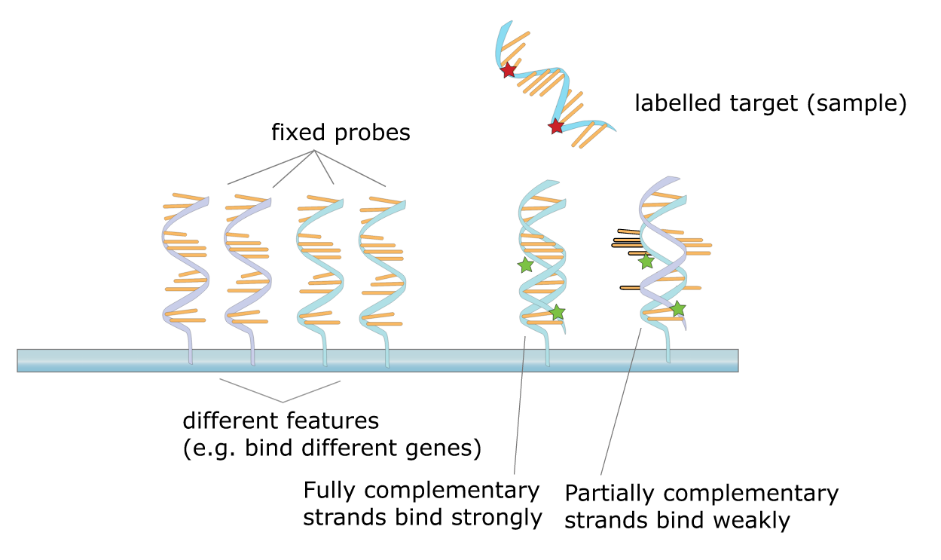
Les biopuces, aussi appelées puces à ADN ou micromatrices d'ADN (DNA microarrays en anglais), sont principalement utilisées afin d'analyser le niveau d'expressions des gènes transcrits dans un milieu (cellules, tissu ou autres) donné.

[photo puce]

<https://upload.wikimedia.org/wikipedia/commons/2/22/Affymetrix-microarray.jpg>

Lors d'une expérience utilisant des biospuces, les ARNm (ARN messagers, issu de la transcription d'une partie d'ADN) sont extraits du milieu. Ils sont ensuite retranscrits en ADNc (ADN complémentaire, qui correspond à la partie codante de l'ADN que l'ARNm avait transcrit à la base), on parle de transcription inverse. Ces brins d'ADNc sont ensuite marqués des molécules fluorescente. Les plus utilisées sont fluorochromes la Cyanine 3 et la Cyanine 5 qui fluorescent dans le vert respectivement le rouge. Ces ADNc marqués sont ensuite mis en contact avec une biopuce.

Une biopuce est une surface solide (verre, silicium ou plastique) sur laquelle des molécules d'ADN (appelées sondes dans ce contexte) sont disposées en rangs. Lorsque les ADNc entrent en contact avec des sondes qui possèdent des bases nucléiques complémentaires, elles forment des ponts hydrogènes. Plus il y a de bases complémentaires entre l'ADNc et la sonde, plus la liaison sera forte, une correspondance parfaite donnant lieu à des doubles hélices très fortement liées.



[schéma d'une biopuce]

<https://upload.wikimedia.org/wikipedia/commons/a/a8/NA_hybrid.svg>

La biopuce est ensuite lavée pour enlevés les brins d'ADNc qui ne se sont pas ou mal hybridés, puis scannées à la longueur d'onde d'excitation des fluorochromes utilisés. L'analyse de l'intensité lumineuse dans les différentes parties de cette image permet de déterminer la quantité de sondes qui ont été hybridées. L'analyse est toutefois différente suivant le type de biopuce utilisée.

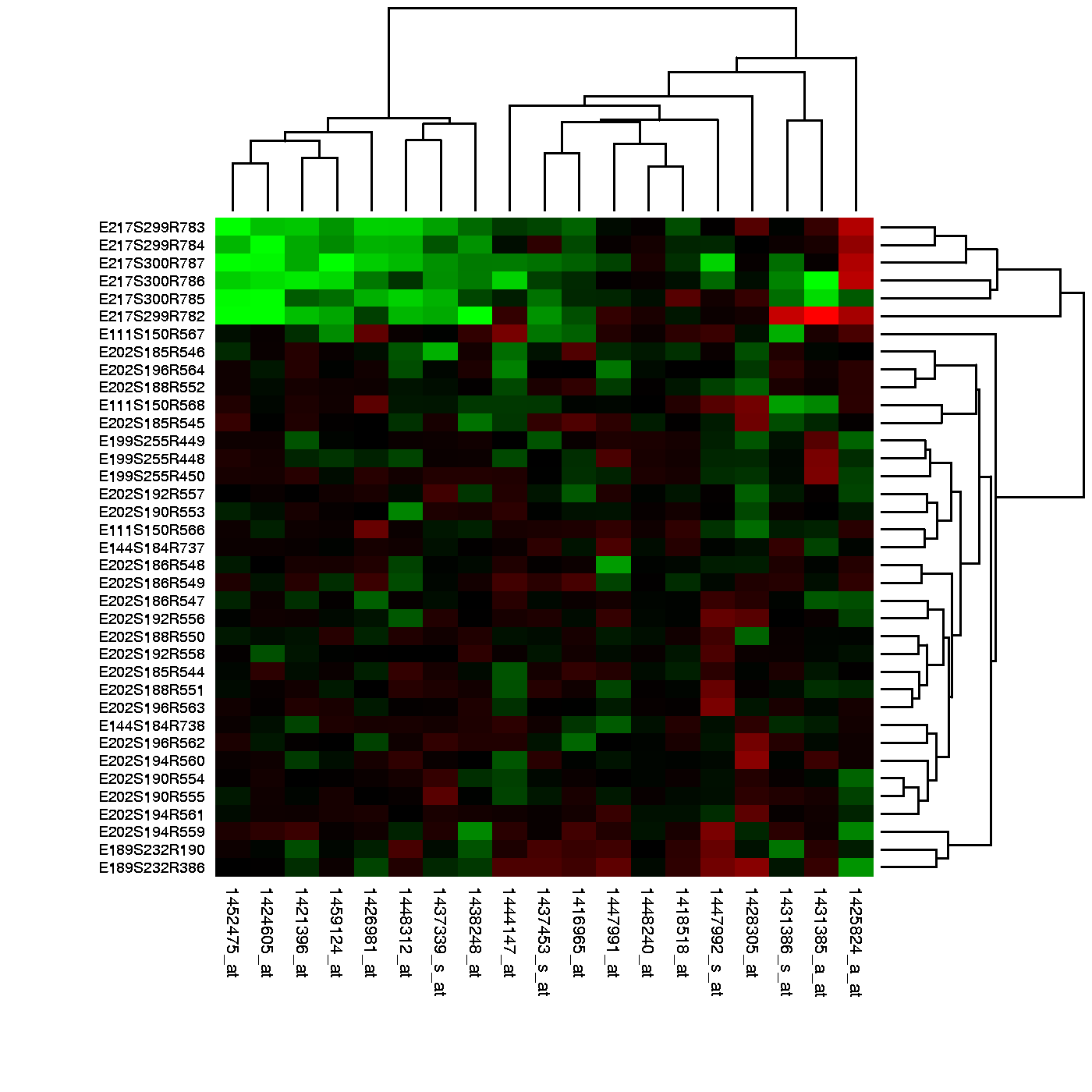
On distingue en effet deux types de biopuces. Les biopuces à un canal et les biopuces à deux canaux.

Les biopuces à un canal sont en fait monochromatiques. C'est-à-dire qu'elles permettent uniquement de testé une seule condition initiale. La comparaison entre deux milieux différents ne peut s'effectuer qu'avec deux expériences sur deux biopuces distinctes.

Les biopuces à deux canaux permettent de tester deux conditions initiales différentes, et afficheront donc deux couleurs, ce qui permet de comparer le niveau d'expression des gènes dans deux milieux différent en une seule expérience.

On peut utiliser les biopuces pour comparer le niveau d'expression de certains gènes dans différentes situation. Par exemple, comparer des cellules saines avec des cellules atteintes d'une pathologie, comme un cancer, par exemple. Elles permettent également de détecter les variations du nombre de copies d'ADN dans le temps, typique des cancers.

[Représentation du scan d'une biopuce à deux canaux] <https://upload.wikimedia.org/wikipedia/commons/4/48/Heatmap.png>



Prix d’une puce :



## Planning

|  |  |
| --- | --- |
| Semaine | Travail à effectuer |
| 1 | - Intro théorique générale |
| 2 | - Intro théorique générale |
| 3 | - Intro théorique générale |
| 4 | Analyse du problème et recherche |
| 5 | Analyse du problème et recherche |
| 6 – 05.04.2016 | Analyse du problème et recherche |
| 7 – 12.04.2016 | Contexte technique et scientifique, état de l’art |
| 8 – 19.04.2016 | Contexte technique et scientifique, état de l’art |
| 9 – 26.04.2016 | Décomposition du problème, conception de la solution |
| 10 – 03.05.2016 | Décomposition du problème, conception de la solution |
| 11 – 10.05.2016 | Décomposition du problème, conception de la solution |
| 12 – 17.05.2016 | Réalisation, implémentation et tests |
| 13 – 24.05.2016 | Réalisation, implémentation et tests |
| 14 – 31.05.2016 | Réalisation, implémentation et tests |
| 15 – 07.06.2016 | Analyse de résultats, documentation et présentation |
| 16 – 14.06.2016 | Analyse de résultats, documentation et présentation |

## Biomarqueurs

Un biomarqueur est une caractéristique mesurable liée à un processus normal ou non et peut être utilisé pour le dépistage, le diagnostic ou la réponse à un traitement. Le plus souvent, ce sera une protéine dosable dans le sang ou une molécule présente dans l’urine.

Pour le développement d’un biomarqueur, on préconise un processus en 5 étapes :

1. Phase exploratoire de recherche de biomarqueurs candidats lors d’une étude sur quelques dizaines de personnes saines/malades.
2. Développement d’un test clinique reproductible sur un échantillon représentatif de la population ciblée.
3. Étude permettant de valider la capacité du biomarqueur à détecter la maladie avant l’apparition des symptômes
4. Étude clinique à long terme sur la population ciblée pour déterminer l’utilité du biomarqueur
5. Étude à long terme pour valider le biomarqueur

La validation est un processus long et complexe. Pour faciliter l’utilisation des biomarqueurs encore en développement, ils ont été classés en fonction de leurs niveaux de preuve :

1. (Plus haut niveau de validation) : biomarqueur validé par une étude clinique prospective randomisée sur la population et sur le long terme.
2. biomarqueur validé par une étude prospective sur un échantillon de la population.
3. biomarqueur validé par une étude rétrospective sur un échantillon représentatif de la population.
4. biomarqueur validé par une étude rétrospective sur un échantillon non représentatif de la population.
5. (Plus faible niveau de validation) : biomarqueur validé en laboratoire uniquement.

Source : <https://fr.wikipedia.org/wiki/Biomarqueur>

Pour détecter le cancer de la prostate nous étudierons les miARN présents dans les urines d’un groupe de personne test regroupant des personnes atteintes et des personnes saines. Nous pourrons ainsi définir si une des miARN présente dans les urines peut avoir une relation avec un cancer de la prostate.

## Biological Pathways

Un « biological pathway », représente une série d’action parmi les molécules dans une cellule qui mène à un produit ou un changement dans cette cellule. Cela peut déclencher l’assemblage d’une nouvelle molécule comme des graisses ou des protéines, ou activer ou désactiver des gènes ou même stimuler une cellule pour la faire bouger.

Connaitre les pathways, du gène à l’action qu’il mène une fois traduit, permet de construire un réseau métabolique et donc de comprendre quelles actions ont ces éléments entre eux et comment la cellule fonctionne.

Si certaines défaillances viennent à apparaitre dans un niveau ou un autre du pathway, par exemple un problème au niveau du gène qui cause la défaillance d’une protéine ou un problème de répression, la connaissance du chemin métabolique va permettre de prédire les conséquences et ainsi de trouver certains éléments qui peuvent indiquer le problème en question. Par exemple la répression miARN va influencer les miARN pour réguler la production d’une certaine protéine. En cas d’absence de répression, la présence élevée de la protéine dans la cellule ou ailleurs pourra indiquer où est la défaillance.

Par exemple, pour un cancer, certaines protéines seront plus ou moins présentes dans les cellules ou dans les sécrétions de ces cellules et pourront être utilisées pour détecter le cancer.

Qu’est-ce que c’est ? conséquences de la répression miARN ? Biomarqueur (intro)

Thibault

## miARN

Les miARNs servent à **réguler** l'expression d'un gène suite à une transcription. En d'autre terme les miARNs servent à rendre les ARNm issus de la transcription moins efficaces, voir totalement silencieux.

Les miARNs atteignent ce but en se fixant sur les ARNm qu'ils ciblent. Après s'être fixés les miARNs peuvent vont neutraliser l'ARNm de trois manières différentes :

- Couper l'ARNm en deux parties distinctes.

- Réduire la queue poly A de l'ARNm (augmentant ainsi la rapidité de sa dégradation et réduisant son taux de traduction).

- Réduire l'efficacité de la traduction des ARNm par les ribosomes.

Les miARNs, chez les animaux, peuvent cibler plusieurs ARNm différent (et, de manière similaire, un ARNm peut être ciblé par plusieurs miARNs).

En effet, les miARNs peuvent reconnaître leur cible à un nombre restreint de nucléotides (6-8). Ceci fait qu'un miARN peut cibler un grand nombre de ARNm différents.

En résumer, un miARN peut cibler plusieurs type de ARNm (plusieurs type de protéines donc). Il le fait dans le but de réduire l'expression de ces ARNm.

### Extraction de la miARN

Pour faire très simple, les cellules sont lysées[[1]](#footnote-1) pour former un substrat. Ensuite, en se servant de différents procédés chimiques et physiques, afin de précipiter certains éléments ou d’en éliminer d’autres (la RNase par exemple), l’ARN est extraite du substrat. Il existe différentes méthodes mais la plus commune aujourd’hui (d’après Wikipédia) est l’extraction « "thiocyanate de guanidinium-phénol-chloroforme» qui a l’avantage de permettre l’extraction des plus petits brins de nucléotides comme la miARN, que d’autres méthodes ne permettent pas.

Par centrifugation du substrat de cellules lysées dans une solution de phénol et de chloroforme, on obtient deux phases dont une contient l’ARN. On utilise en suite le Thiocyanate de guanidinium, très utilisé pour le lysage, qui a aussi la particularité de dénaturer les enzymes comme la DNase ou la RNase qui peuvent endommager l’extrait.

Une fois ces miARN extraites puis placées sur des chips (puce), on étudie la luminosité dégagée par les molécules selon leur position sur la puce.

### Relation avec le cancer

Comme les miARNs servent à réguler l'expression de certains gènes, et donc de certaines protéines, une augmentation ou une réduction du nombre d'un certain type de miARNs, par rapport à la normale, peut indiquer une réduction respectivement une augmentation de l'expression d'une protéine. La présence de certaines protéines peut être liée à certains cancers, tout comme l'absence d'autres protéines.

Les miARNs peuvent ainsi aider au diagnostic d'un cancer, et également aider à déterminer la gravité de celui-ci ou l'efficacité d'un traitement.

### Outils

Il existe plusieurs outils qui permettent de prédire les cibles des miARNs. On peut notamment citer

* RNA22 qui permet des sites cibles dans une séquence de un ou plusieurs miARN On peut directement lui passer des miARNs et une séquence que l'on souhaite étudié. Il est aussi possible d'utiliser des résultats déjà connus.
* TargetScan qui cherche à prédire les cibles des miARNs en recherchant la présence de sites ciblés par les miARNs.

# Etat de l’art

Qu’est-ce qui a été fait et comment ?

Recherches déjà faites (voir intro)

Autres moyens de diagnostic (pour le m’eme cancer, et/ou avec les miRNA), PRISE DE SANG (pour PCa)

Quelles techniques utilisées dans ces travaux

Quels autres moyens de diagnostic du cancer

THIBAULT

# Démarche d’analyse des données

Méthodologies actuelles état de l’art

Méthodes d’analyse array actuelles

Limites des données

Contexte, antécédents etude miARN

## Méthodologie classique

Prélévement -> extraction de données exploitable

Méthod actuel étude miARN, procédures, différents moyens etc -> cours (chips)

Repérer miARN importantes -> pathway -> quel est leur action

THIBAULT

Aujourd’hui, quelles sont les différentes méthodes pour extraire et analyser les miARN et déduire leurs actions ?

### Limite des données

Quelle est la limitation des données utilisées -> microarray

JAN

## Outils statistiques

Méthode de calcul

Pvaleurs

JAN

### Machine learning

Dans le cadre de notre recherche le machine learning semble être intéressant dans la phase d'analyse des données, car il peut nous permettre d'apprendre à un programme à établir un diagnostic selon des données reçues.

En disposant de données adéquates, c'est-à-dire parmi lesquelles on connaît quelles sont les personnes saines et quelles sont les personnes cancéreuses, nous pouvons espérer apprendre à distinguer les personnes malades des personnes saines à partir de données de bases. En mettant à jour les données régulièrement, on réduira les marges d’erreurs.

Une autre utilité serait de distinguer quels miARNs sont surreprésentés, ou sous-représentés, entre les malades et les cancéreux, ce qui peut fournir des informations précieuses sur d'éventuels moyen de prévenir ou guérir la maladie.

# Conception

Décomposition du problème et conception de la solution

Sélection des variables (réduction)

Choix des méthodes d’analyse (Plusieurs ???)

Analyse des données reçues (quantité, qualité, tri nécessaires ?, analyses déjà faites (p-value, score), etc)

## Set de données

1. Désintégration de la membrane cellulaire [↑](#footnote-ref-1)