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

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

Table des matières

[Introduction 2](#_Toc450289001)

[Planning 2](#_Toc450289002)

[La prostate 2](#_Toc450289003)

[Le cancer de la prostate 3](#_Toc450289004)

[L’urine 3](#_Toc450289005)

[Micro ARN 3](#_Toc450289006)

[miARN et cancer prostatique 3](#_Toc450289007)

[Les Biopuces et microarray 5](#_Toc450289008)

[Biomarqueurs 7](#_Toc450289009)

[Biological Pathways 8](#_Toc450289010)

[miARN 8](#_Toc450289011)

[Extraction de la miARN 9](#_Toc450289012)

[Relation avec le cancer 9](#_Toc450289013)

[Outils 9](#_Toc450289014)

[Etat de l’art 9](#_Toc450289015)

[Démarche d’analyse des données 11](#_Toc450289016)

[Méthodologie classique 11](#_Toc450289017)

[Limite des données 11](#_Toc450289018)

[Machine learning 11](#_Toc450289019)

[Conception 12](#_Toc450289020)

[Set de données 12](#_Toc450289021)

# Introduction

Le but de ce travail est de trouver un moyen informatique de diagnostic du cancer de la prostate à partir des micro-ARN présents dans les 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.

Une fois les différentes bases sur le sujet acquises, nous étudierons un set de données prélevées sur des personnes atteintes ou non du cancer, puis avec les outils d’analyse à notre disposition, nous développerons un outil de diagnostic.

## 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 |

## 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).

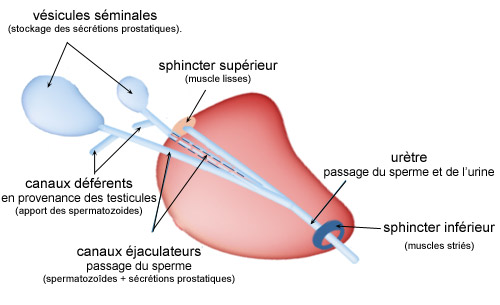


Figure Schéma de la prostate

## 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.



Figure Photo d'une bio-puce

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.

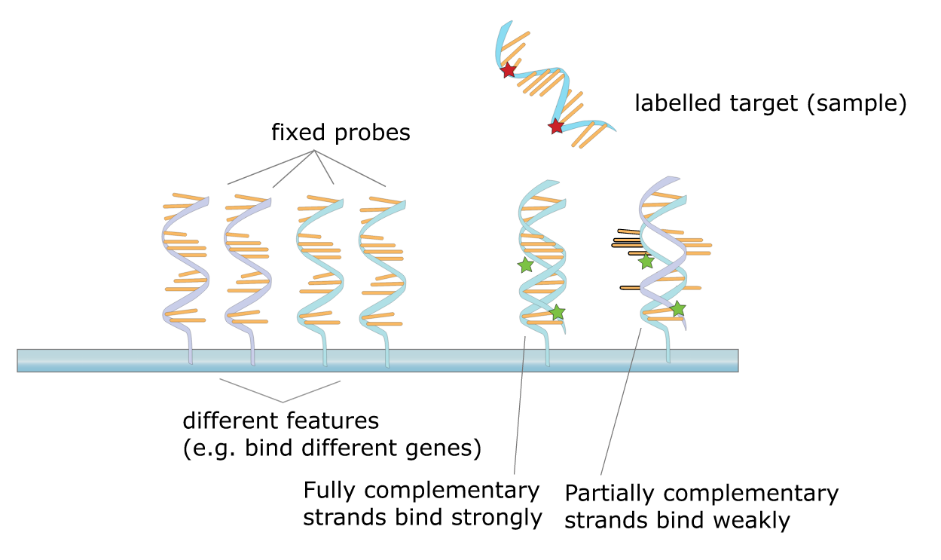


Figure Fonctionnement d’une bio-puce

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.

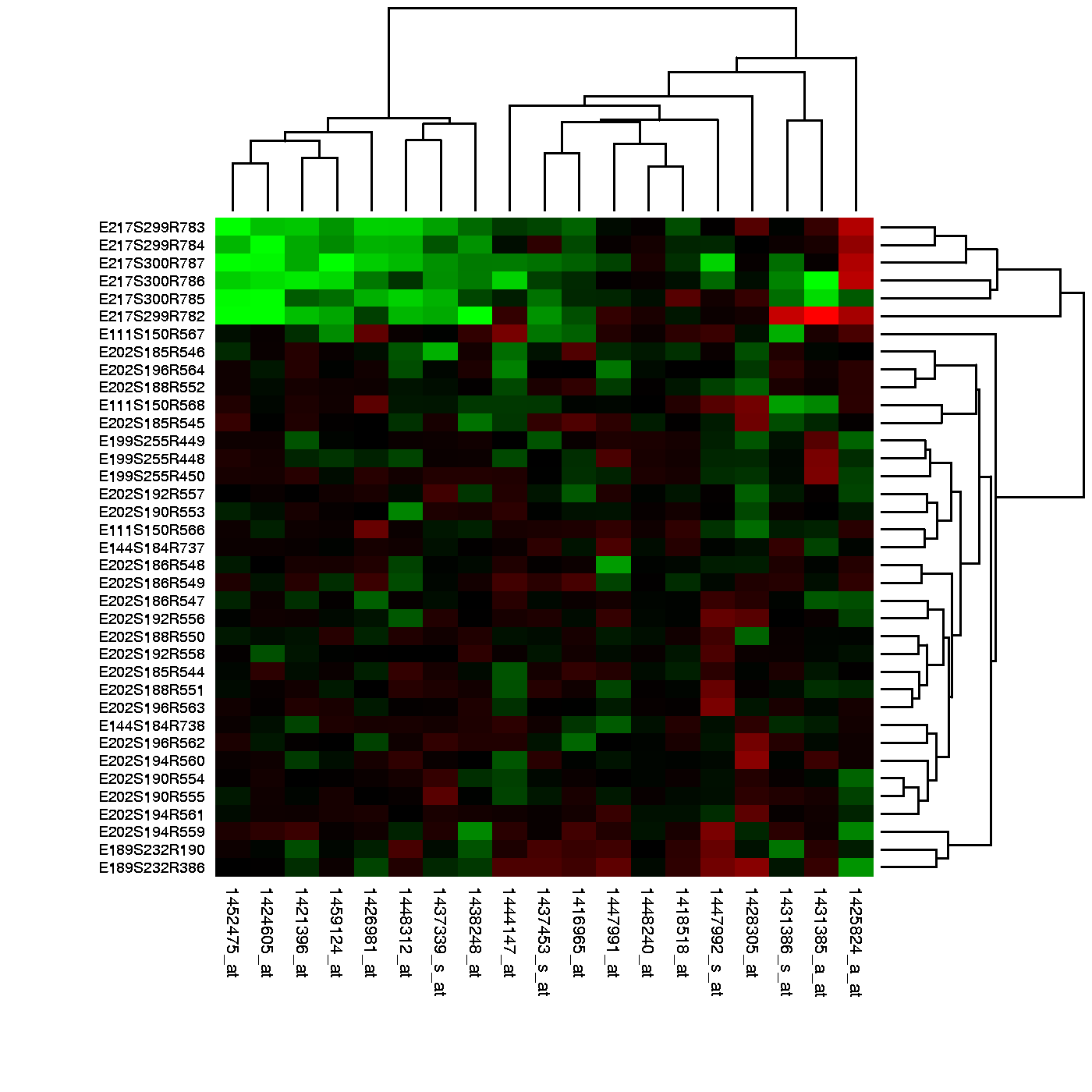


Figure Scan d'une bio-puce

## 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.

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.

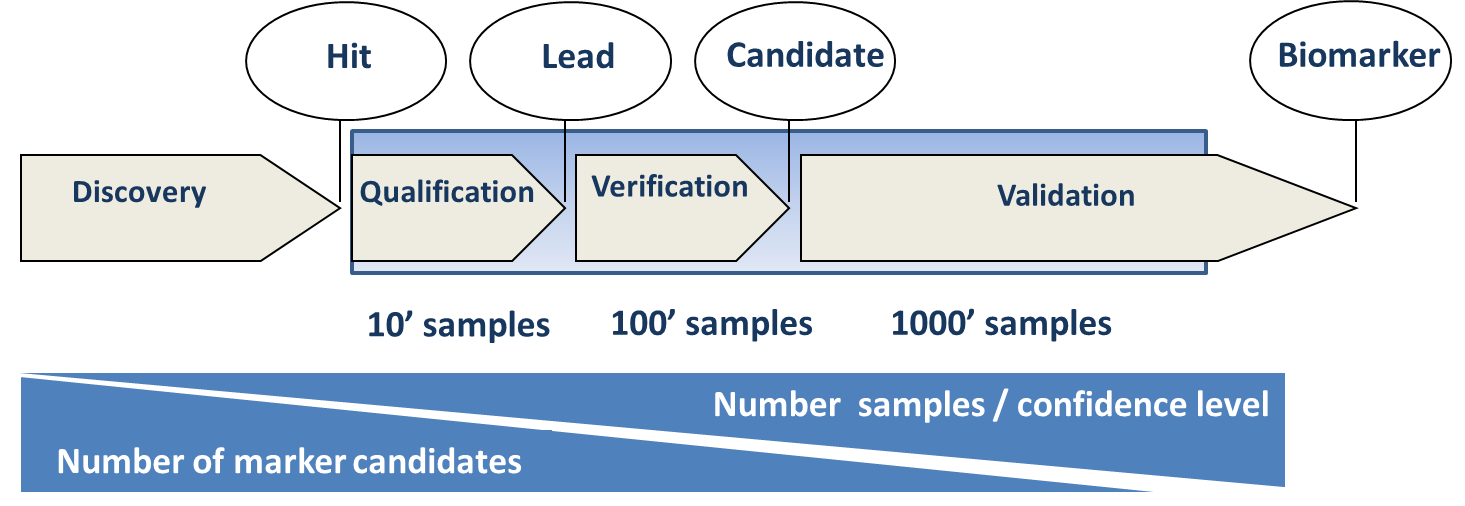


Figure Processus de développement d'un biomarqueur

## 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.

## 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.

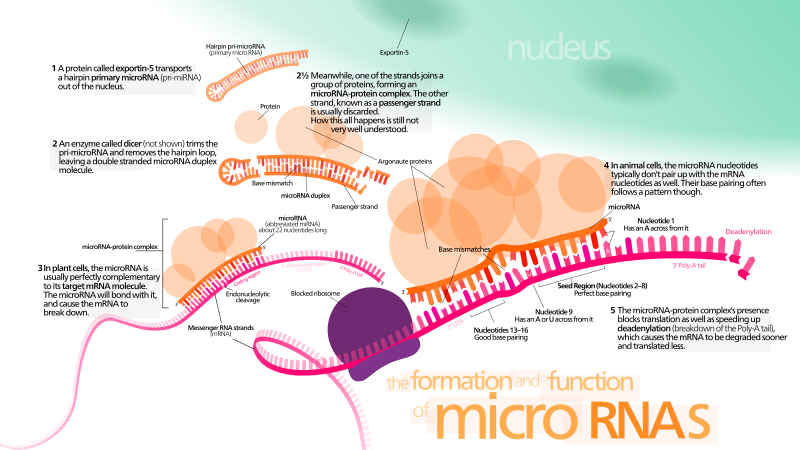


Figure Formation et fonction des miARN

### 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.

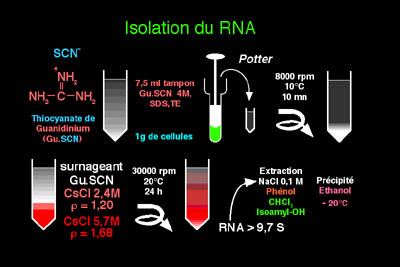


Figure Extraction miARN

### 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.

# http://www.docteurclic.com/galerie-photos/image_3053_m.jpgEtat de l’art

Figure Toucher rectal

Aujourd’hui, beaucoup de maladies se diagnostiques à l’aide d’un scanner, d’une biopsie ou de divers prélèvements ou autres méthodes pouvant être à risques. Non seulement beaucoup sont invasives, mais permettent seulement un diagnostic tardif. Pour le cas du cancer de la prostate, il existe des marqueurs présents dans le sang (Antigène Prostatique Spécifique PSA), mais on pratique encore beaucoup la biopsie, la radiologie ou le toucher rectal, pour connaitre l’état des cellules de l’organe, car le PSA peut être un marqueur pour d’autres pathologies. C’est à ce moment que le miARN intervient.

Le niveau d’expression des miARN est un moyen de diagnostiquer une pathologie ou un dérèglement avant même que les symptômes n’apparaissent, mais éventuellement aussi un futur moyen de traitement étant donné son influence sur l’expression du génome. Lors de différents traitement, ou autre diagnostics, l’étude du niveau d’expression des miARN peut permettre d’affiner le diagnostic et par exemple de faire un pronostic vital sur l’état de la pathologie.

Après sa découverte en 1993, puis d’autres travaux importants dans les années 2000, la micro ARN a été associée à une pathologie dans les années 2010, par exemple pour le cancer du poumon ou la leucémie lymphoïde chronique. Un dérèglement dans son niveau d’expression a été associé à la Leucémie Lymphoïde Chronique ce qui a permis développer un nouveau genre d’outils de diagnostic.

Aujourd’hui, l’étude du niveau d’expression des miARN, et donc de l’expression de certains gènes à travers les pathways, permet de développer des outils de diagnostic pour de plus en plus de pathologies.

Après avoir récolté les données du niveau d’expression des miARN, les outils informatiques et statistiques prennent le relais afin de définir quels miARN sont les marqueurs les plus significatifs pour le diagnostic, ceci afin de réduire les couts en diminuant le nombre de détections nécessaires.

# Conception

## Décomposition du problème

### Données reçues

Les données que nous utiliserons, sont des donnés de micro-arrays déjà traitée (traitement d’image, normalisation afin de réduire le bruit statistique). Nous avons donc le niveau d’expression de miARN en grand nombre.

Notre jeu de données contient une soixantaine de prélèvement urinaire de patients souffrant ou non d’un cancer prostatique.

### Filtrage des données

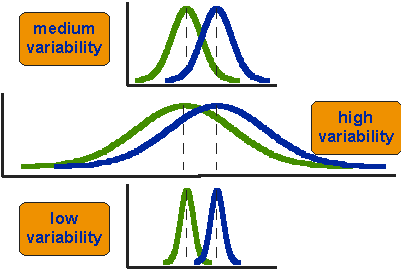
La quantité de données / miARN étant trop importante, il est nécessaire de réduire la quantité de données qui va être étudiée. Un premier filtre pourrait consister par exemple, à retirer les miARN dont le niveau d’expression n’est pas significativement différent entre les patients malades et les patients sains. Plusieurs méthodes existent afin d’affiner la quantité de gènes nécessaire à un diagnostic. On citera par exemple le test de Student (t-test) et la méthode SAM (Significance Analysis of Microarrays).

Figure t-test

### Machine Learning

Une fois les données pertinentes isolées, il faudra entrainer une méthode de machine learning afin de différencier les groupes de patients sains de malade. On pourra utiliser des méthodes telles que kNN (k nearest neighboors) ou SVM (Support Vector Machine) afin, d’avoir un outil de diagnostic entrainé.

Pour faire fonctionner ces outils de machine learning, nous séparerons les données en deux parties ; une pour entrainer le modèle et l’autre, plus petite, pour le tester après la phase d’entrainement.

### Analyse des résultats

Une fois les résultats obtenus, nous devrons analyser leur pertinence en explorant les pathways des différents miARN obtenus comme marqueurs afin de confirmer rapport avec la maladie (PCa).

## Outils utilisés

Nous utiliserons les outils vus en classe comme numpy et scipy et leurs différentes librairies ainsi que les outils de prédiction de cibles de miARN comme RNA22 par exemple.

# Sources

## Images

|  |  |
| --- | --- |
| Schéma prostate | <http://www.sante-sur-le-net.com/fiches-info/prostate/> |
| Photo bio-puce | <https://upload.wikimedia.org/wikipedia/commons/2/22/Affymetrix-microarray.jpg> |
| Schéma bio-puce | <https://upload.wikimedia.org/wikipedia/commons/a/a8/NA_hybrid.svg> |
| Scan bio-puce | <https://upload.wikimedia.org/wikipedia/commons/4/48/Heatmap.png> |
| Biomarqueur | <http://www.fournier-majoie.org/sites/default/files/landing/landing/visual/biomarkers%20stages_0.png> |
| Schéma miARN | <https://en.wikipedia.org/wiki/File:MiRNA.svg#/media/File:MiRNA-fr.svg> |
| Extract miARN | <http://www.chups.jussieu.fr/polys/biochimie/BGbioch/BG_49_PICT.jpg> |
| Toucher rectal | <http://www.docteurclic.com/galerie-photos/image_3053_m.jpg> |
| t-test | <http://www.socialresearchmethods.net/kb/Assets/images/stat_t2.gif> |

## Informations

|  |  |
| --- | --- |
| ARN, pathways, biomarqueurs, cancer, etc. | [www.wikipedia.org](http://www.wikipedia.org) |
| Cancer de la prostate | <http://www.creapharma.ch/cancer-de-la-prostate.htm> |
| British Journal of Cancer | <http://www.nature.com/bjc/journal/v100/n10/abs/6605058a.html> |
| Prostate | <http://www.em-consulte.com/article/281389/biologie-moleculaire-de-la-prostate-normale-et-pat> |
| miARN cancer vessie | <http://www.urologiconcology.org/article/S1078-1439(09)00031-3/abstract> |
| miARN et cancer | <http://www.arte.tv/magazine/futuremag/fr/lutte-contre-le-cancer-les-promesses-du-micro-arn-futuremag> |
| Prostate anatomie | <http://www.sante-sur-le-net.com/fiches-info/prostate/> |
| miARN study | <http://mon.univ-montp2.fr/claroline/backends/download.php?url=L0FubullMjAxMF8yMDExL0NvbXBfQ291cnMvTWljcm9fQVJOLnBkZg%3D%3D&cidReset=true&cidReq=BIO1> |
| Biomarqueur urine | <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3831115/> |
| miARN profiling for prostate cancer | <http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0076994> |
| Extraction miARN | <http://www.chups.jussieu.fr/polys/biochimie/BGbioch/POLY.Chp.7.4.html> |
| Expression miARN tumeur système nerveux central | <https://tel.archives-ouvertes.fr/file/index/docid/905298/filename/ThA_se_Elodie_Lages.pdf> |
| Cancer pancreas niARN | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680248/> |
| miARN in human cancer | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3704221/> |

# Table des illustrations

[Figure 1 Schéma de la prostate 3](#_Toc450298539)

[Figure 2 Photo d'une bio-puce 5](file:///C:\Users\thsch\Google%20Drive\HEIG-VD\Semestre6\BBC\BBCProjet\report\BBC2016_Projet_SchowingPuro.docx#_Toc450298540)

[Figure 3 Fonctionnement d’une bio-puce 5](file:///C:\Users\thsch\Google%20Drive\HEIG-VD\Semestre6\BBC\BBCProjet\report\BBC2016_Projet_SchowingPuro.docx#_Toc450298541)

[Figure 4 Scan d'une bio-puce 6](#_Toc450298542)

[Figure 5 Processus de développement d'un biomarqueur 7](#_Toc450298543)

[Figure 6 Formation et fonction des miARN 9](#_Toc450298544)

[Figure 7 Extraction miARN 10](#_Toc450298545)

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