

RAPPORT DE STAGE DE DEUXIÈME ANNÉE DE MASTER

---

COMPARAISON DE L'ÉVALUATION DE L'INDEX GÉNOMIQUE  
EN MÉTHODOLOGIE AGILENT/SUREPRINT G3 VERSUS  
AFFYMETRIX/ONCOSCAN CNV

---

*Auteur :*

BORDRON Elie

Etudiant en Master 2 de bio-informatique  
Université de Bordeaux

*Directrices de stage :*

LARMONIER Claire, PhD

DARBO Elodie, PhD

19 mai 2022

# Sommaire

<b>Introduction</b>	<b>1</b>
0.1 Contexte sur la cancérologie . . . . .	1
0.2 La cancérologie à l'institut Bergonié . . . . .	1
0.3 Le calcul de l'index génomique par CGH . . . . .	1
0.4 Transposer le calcul du GI d'Agilent à OncoScan . . . . .	2
0.5 Communication Bioinformatique-Biologie . . . . .	2
<b>1 Etat de l'art</b>	<b>4</b>
1.1 Calcul du GI par Agilent . . . . .	4
1.2 . . . . .	4
<b>2 Matériel et Méthodes</b>	<b>9</b>
2.1 Matériel . . . . .	9
2.2 Méthodes . . . . .	9
<b>3 Résultats et Discussion</b>	<b>11</b>
<b>Bibliography</b>	<b>12</b>

# Tables des figures

1	Vue d'ensemble des technologies utilisées dans l'unité de pathologie moléculaire à l'institut Bergonié . . . . .	2
1.1	Le Pipeline du logiciel ChAS qui détermine les altérations de nombre de copies. . . . .	5
1.2	Comparaison des apports de chaque outil dans le calcul du GI . . . . .	5
1.3	Le Pipeline d'OncoscanR qui détermine les altérations de nombre de copies. .	6
1.4	Le Pipeline de rCGH qui détermine les altérations de nombre de copies. .	6
1.5	Pipeline de détermination du nombre de copies du package CGHcall . . .	7
1.6	Pipeline d'ASCAT aboutissant au nombre d'altérations de nombre de copies	8

# Introduction

## 0.1 Contexte sur la cancérologie

Dans le traitement de lésions cancéreuses, le diagnostic et le pronostic se basent sur des critères cliniques, anatomiques et pathologiques. Quand de tels critères ne suffisent pas à caractériser un cas difficile, des technologies de screening moléculaire sont disponibles, permettant l'identification d'altérations moléculaires spécifiques ou caractéristiques du diagnostic. Ces données moléculaires peuvent permettre également d'orienter les choix thérapeutiques, en particulier vers des thérapies ciblées.

Au-delà du diagnostic, il existe actuellement un intérêt croissant pour l'identification de signatures moléculaires d'intérêt pronostique et/ou prédictif, dans l'optique d'une prise de décision thérapeutique éclairée et d'un suivi adapté. C'est le cas d'un certain nombre de signatures d'expression génique par exemple dans le cancer du sein [1], actuellement réalisées en routine et d'autres signatures en cours d'évaluation dans des essais cliniques [2].

## 0.2 La cancérologie à l'institut Bergonié

À l'institut Bergonié, différentes technologies sont utilisées en routine pour répondre à ces problématiques dans l'unité de pathologie moléculaire (fig.1). Parmi elles, on peut citer l'hybridation in situ en fluorescence (FISH), qui permet de détecter des anomalies au niveau de séquences d'ADN cible, ou le séquençage qui, en analysant la séquence d'un gène, permet d'évaluer le risque d'une maladie génétique. La CGH-array est utilisée pour analyser les variations de nombre de copies d'un génome. Les altérations détectées caractérisent le gain ou la perte d'ADN, ce qui permet de connaître le statut de gènes liés au développement de tumeurs et d'adapter la prise en charge thérapeutique. Plus précisément, deux méthodologies sont utilisées : SurePrint G3 d'Agilent et OncoScan CNV d'Affymetrix.

## 0.3 Le calcul de l'index génomique par CGH

L'intérêt diagnostique et pronostique de l'index génomique (GI) par CGH-array dans différents types de sarcomes comme les tumeurs stromales gastro-intestinales (GIST)[3,5 gironde], les synoviosarcomes [4 gironde] et les tumeurs musculaires lisses de l'utérus [6 gironde] a été précédemment démontré. Cette « signature génomique » ou « genomic index » est le reflet direct du degré de complexité moléculaire et d'instabilité génomique

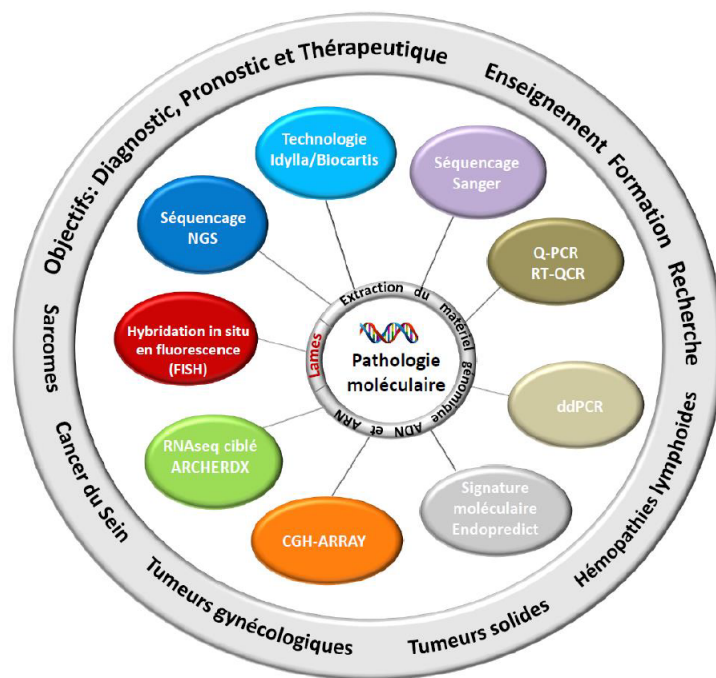


FIGURE 1 – Vue d’ensemble des technologies utilisées dans l’unité de pathologie moléculaire à l’institut Bergonié

de la tumeur et s’avère être un puissant prédicteur de l’agressivité tumorale et de la rechute métastatique des tumeurs. Ce score est utilisé en routine dans l’unité de pathologie moléculaire. Afin d’étendre l’utilisation de l’évaluation de ce score à la méthodologie OncoScan CNV, plus récente et plus résolutive, on se propose de transposer le calcul de l’index génomique validé sur puces AGILENT/SurePrint G3 8x60 K.

## 0.4 Transposer le calcul du GI d’Agilent à OncoScan

La couverture du génome étant différente entre les deux technologies, il est nécessaire de les comparer sur une cohorte de cas précédemment caractérisés et sur lesquels toutes les données sont disponibles. Pour mettre en place cette comparaison, on va identifier un outil bioinformatique qui permet de calculer le GI à partir des données OncoScan. On répondra alors à la question suivante : Peut-on établir la correspondance du calcul de l’index génomique et la détermination du seuil de classification des tumeurs de la technologie Agilent à la technologie Affymetrix pour pouvoir en faire bénéficier les patients analysés au sein du laboratoire dans le cadre diagnostic ou thérapeutique ?

## 0.5 Communication Bioinformatique-Biologie

Ce travail s’inscrit dans le projet GIRONDE, qui s’étend au-delà du stage. Un échange biologie-bioinformatique a été entretenu pour que les biologistes puissent suivre les avancées du projet, de sorte à ce que l’équipe ait la meilleure vision possible du travail effectué

et choisisse l'outil le plus pertinent. Cet échange a été facilité par l'immersion dans le côté biologique du projet, car j'ai suivi la technique à partir de l'ADN extrait jusqu'à la production des fichiers de données. Un travail de revue d'outils bioinformatiques a fait l'objet d'une vulgarisation auprès de l'équipe pour présenter les options disponibles.

# Chapitre 1

## Etat de l'art

### 1.1 Calcul du GI par Agilent

besoin du nombre d'altérations de nombre de copies et du nombre de chromosomes les portant. La détermination des altérations se base sur le log Ratio (pour le calcul, voir l'aide de ChAS :  $\log_2(\text{echantillon}) - \log_2(\text{ref})$ ). Le GI est défini ainsi :  $A^2/C$ , où A est le nombre d'altérations et C le nombre de chromosomes qui en porte. Pour calculer le GI,

### 1.2

Dans le but de calculer le GI à partir des données OncoScan, on cherche l'outil le plus adapté. Ce que l'outil devra faire :

chaque outil devra déterminer les altérations de nombre de copies à partir du log ratio. calculer le GI

donc prendre en input les données OncoScan — > présenter les données OncoScan : log ratio, BAF... et déterminer les segments d'altération à partir de ces données.

(être utilisable en routine)

Voici les outils disponibles :

§ ChAS :

Outil propriétaire En quoi est-ce utile

- Permet de visualiser les altérations 1.1
- donc calculer le GI
- ainsi que de recentrer le profil, fusionner des segments, pour être sûr de prendre les vraies altérations

En quoi est-ce limité

- le recentrage et la fusion des segments, si ils ont lieu d'être effectués, le sont manuellement.
- le calcul du GI aussi est manuel.
- plus nombreuses sont les interventions humaines, plus grand est le risque d'erreur humaine.
- boîte noire

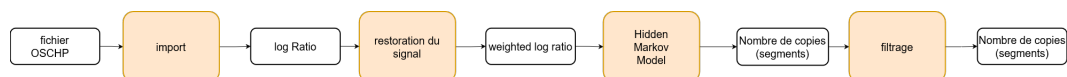


FIGURE 1.1 – Le Pipeline du logiciel ChAS qui détermine les altérations de nombre de copies.

Outil	oncoscanR	rCGH	CGHcall	ASCAT
input	Fichier Segments	Fichier Sondes	Fichier Sondes	Fichier puce
nettoyage des données	oui	oui	deux	non
Normalisation/recentrage	non	recentrage	deux	non
Segmentation	non	CBS	CBS	ASPCF
Calling	non	non (estimation)	Modèle de mélange	ASCAT
autre	calcul des altérations de bras calcul de scores	Visualisation interactive	non	estimation ploidie et cellularité
output permettant le calcul du GI	altérations de bras scores HRD	Nombre de copies par segment	Nombre de copies par sonde	Nombre de copies par segment

FIGURE 1.2 – Comparaison des apports de chaque outil dans le calcul du GI

- pas de possibilité de calculer \*directement\* ce score ni d'autres
- On n'a jamais validé les résultats de ce logiciel à l'aide d'une autre technologie donc on ne connaît pas ses biais (mais c'est le cas pour la plupart des logiciels utilisés dans le domaine du diagnostic cancer)
- pas open source — > reproductibilité des résultats

On regarde ce que d'autres outils permettent. Les outils ne font pas tous les mêmes étapes pour déterminer le nombre d'altérations (fig.1.2).

Outils open source :

§ oncoscanR :

Presentation globale outil et l'article dont il découle : point principal de l'article -> détermination des ALA. L'outil en lui-même : détailler le pipeline 1.3 Préciser comment on peut calculer le GI à partir de cet outil.

§ rCGH

pipeline 1.4 Presentation globale outil et l'article dont il découle : point principal de l'article -> prise de décision thérapeutique -> déterminer statut gènes exploitable -> aCGH -> défis techniques :

centralisation : a un impact important car définit la baseline (état neutre de 2 copies) -> rCGH implémente une approche qui a fait ses preuves [3]

segmentation : l'algo CBS [4] est parmi les plus performants [5], mais utiliser les paramètres par défaut peut être limitant dans l'étude de cas hautement remaniés. rCGH estime la valeur d'un paramètre clé de la segmentation DNACopy à partir du bruit des données. il n'y a pas consensus sur les seuils à utiliser, donc on doit pouvoir manipuler et visualiser le profil -> visualisation interactive

§ CGHcall :

pipeline 1.5

§ ASCAT :

pipeline 1.6



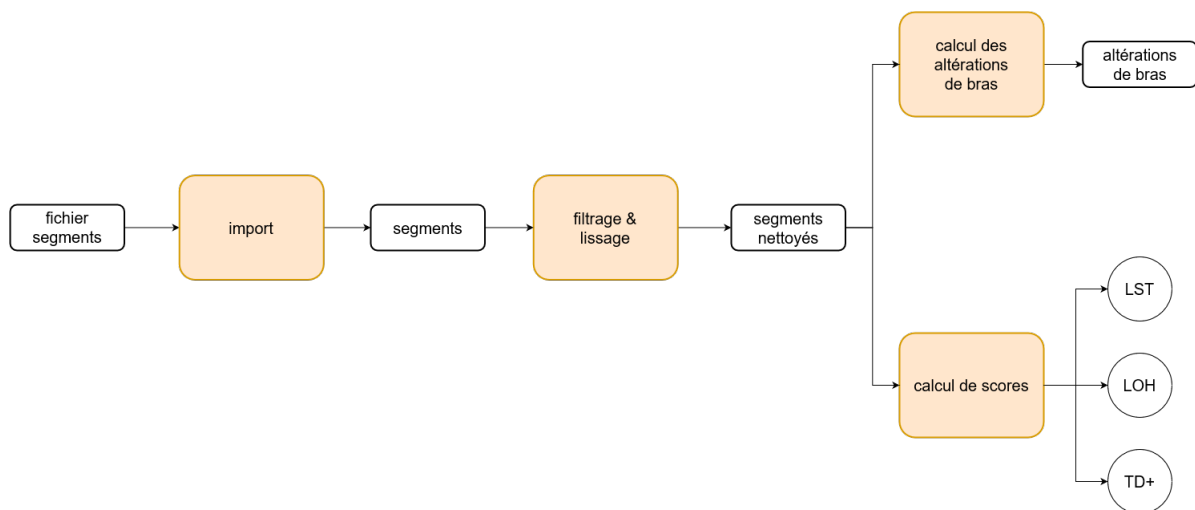


FIGURE 1.3 – Le Pipeline d'OncoscanR qui détermine les altérations de nombre de copies.

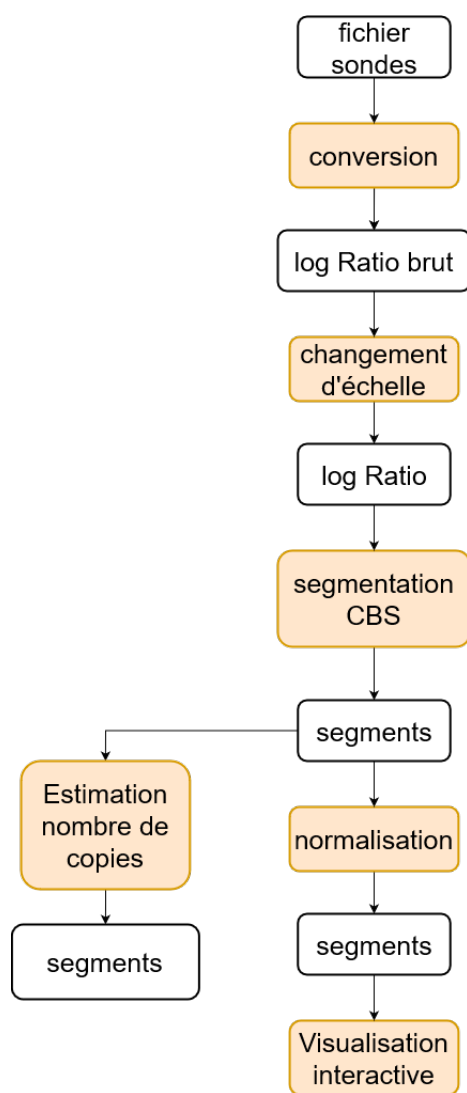


FIGURE 1.4 – Le Pipeline de rCGH qui détermine les altérations de nombre de copies.

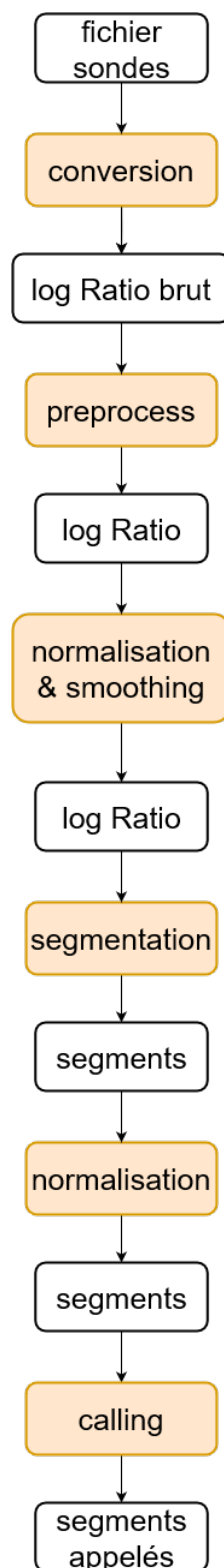


FIGURE 1.5 – Pipeline de détermination du nombre de copies du package CGHcall

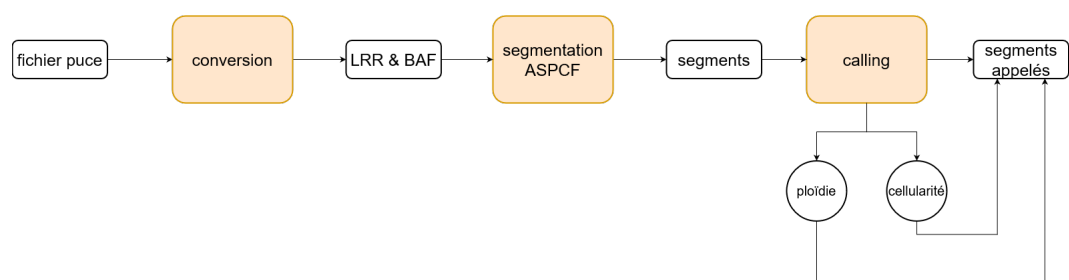


FIGURE 1.6 – Pipeline d’ASCAT aboutissant au nombre d’altérations de nombre de copies

# Chapitre 2

## Matériel et Méthodes

### 2.1 Matériel

Données utilisées

préciser que je travaille sur des données brutes/traitées, quels fichiers. ADN tumoral, FFPE, tumeurs GIST (biopsie), patients... quelques phrases pour tout ça

Logiciels utilisés : *tableau comparatif par étape : l'outil fait cette étape ou non*. ChAS version X -> voir notes prises pendant formation ChAS Rstudio version X R version X packages R : oncoscanR version X - > article... CGHcall version X - > article... ASCAT version X rCGH version X

### 2.2 Méthodes

Nativement, chaque outil a des + et des -. -> détailler exhaustivement, sélectionner les figures les plus pertinentes de chaque ppt, puis-je en fusionner 2 ensembles, ...

oncoscanR les +

filtre et lisse les données OncoScan - > l'intérêt

détermine les altérations de bras chromosomique. Les altérations de nombre de copies sont généralement classées, selon la longueur du segment altéré, en altérations focales ou de bras [parler des différences entre les 2, cf. [6]]. La définition d'une altération de bras ne fait pas consensus au sein de la littérature : généralement, on considère qu'elle correspond à une unique altération couvrant une grande part du bras[7], mais le seuil qui détermine à quel pourcentage d'altération le bras est considéré comme altéré varie selon les études [de 70 à 98, citer ça] et dépend du contexte. La procédure validée par Christinat et al. décrit une définition qui se veut applicable au contexte clinique au cas par cas.

L'intérêt d'une telle définition pour le calcul du GI : pour trancher sur un cas difficile, calculer le GI sur des altérations de bras permet de passer outre des altérations de petite taille qui n'ont pas forcément de d'utilité pour l'aide au diagnostic.

Le calcul de scores HRD et HRD-like est possible au sein de l'outil[détailler].Le score LST ... LOH... TD+... les -

Les données sont transformées : à partir des segments d'altération, on détermine un état binaire pour chaque bras chromosomique : altéré ou non. Le nombre d'altérations utilisé dans le calcul du GI correspond alors au nombre de bras altérés. un bras pré-

sentant 1 ou 8 segments d'altération peut ainsi donner le même GI. Sur les profils qui présentent de nombreux segments, le GI ainsi calculé sera donc limité par le nombre de bras chromosomiques.

rCGH les +

- normalisation
- segmentation
- estimation
- visualisation interactive des données

les -

- estimation simple, pas de calling
- L'interface graphique offre globalement les mêmes possibilités que ChAS

CGHcall

ASCAT

Comment compare-t-on les outils ? § GI comment il est calculé : à partir des segments d'altération ce qui a dû être adapté pour que les 4 outils fassent ce calcul Y a-t-il une proportionnalité similaire entre Agilent et OncoScan ? -> graphe de corrélation (ici, on annonce que cette question est posée dans la partie résultats). Discuter l'amplitude, le cutoff, a-t-on 2 ou 3 groupes...

GGally

§ performance vitesse précision courbes ROC

# Chapitre 3

## Résultats et Discussion

Sachant les plus et moins de chaque outil, on fait une comparaison des outils :

GI Y a-t-il une proportionnalité similaire entre Agilent et OncoScan ? -> graphe de corrélation (ici, on répond à cette question). Discuter l'amplitude, le cutoff, a-t-on 2 ou 3 groupes... Y a-t-il des outils qui n'en expriment pas ?

performance vitesse : quel est l'outil le plus rapide précision : quel est l'outil le plus précis ? (On prend aussi en compte la spécificité, et d'autres paramètres... ) courbes ROC Sur cette comparaison, quel est le meilleur compromis vitesse/précision ? Au vu de ce compromis, du GI obtenu et des points + et - de chaque outil vus dans l'état de l'art, quel est l'outil retenu ?

Conclusion sur la question

# Bibliographie

- [1] S. LAL, A. E. M. REED, X. M. de LUCA et P. T. SIMPSON, « Molecular signatures in breast cancer, » *Methods*, t. 131, p. 135-146, 2017.
- [2] F. CHIBON, P. LAGARDE, S. SALAS et al., « Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity, » *Nature medicine*, t. 16, n° 7, p. 781-787, 2010.
- [3] F. COMMO, C. FERTE, J. SORIA, S. FRIEND, F. ANDRE et J. GUINNEY, « Impact of centralization on aCGH-based genomic profiles for precision medicine in oncology, » *Annals of Oncology*, t. 26, n° 3, p. 582-588, 2015.
- [4] E. VENKATRAMAN et A. B. OLSHEN, « A faster circular binary segmentation algorithm for the analysis of array CGH data, » *Bioinformatics*, t. 23, n° 6, p. 657-663, 2007.
- [5] H. WILLENBROCK et J. FRIDLYAND, « A comparison study : applying segmentation to array CGH data for downstream analyses, » *Bioinformatics*, t. 21, n° 22, p. 4084-4091, 2005.
- [6] Y. CHRISTINAT, P. CHASKAR, S. CLÉMENT et al., « Automated Detection of Arm-Level Alterations for Individual Cancer Patients in the Clinical Setting, » *The Journal of Molecular Diagnostics*, t. 23, n° 12, p. 1722-1731, 2021.
- [7] D. M. ROY, L. A. WALSH, A. DESRICHARD et al., « Integrated Genomics for Pinpointing Survival Loci within Arm-Level Somatic Copy Number Alterations, » *Cancer Cell*, t. 29, n° 5, p. 737-750, 2016, ISSN : 1535-6108. DOI : <https://doi.org/10.1016/j.ccell.2016.03.025>. adresse : <https://www.sciencedirect.com/science/article/pii/S1535610816301088>.