



SECTION GENETIQUE (GEN)

Chef

Dr Paul Brennan

Groupe Epidémiologie génétique (GEP)

Chef

Dr Paul Brennan

Chercheurs

Dr Devasena Anantharaman
(jusqu'en juillet 2016)

Dr Estelle Chanudet-van den Brink

Dr Mattias Johansson

Dr Ghislaine Scélo

Assistants techniques

Valérie Gaborieau

Hélène Renard

Technicien de laboratoire

Priscilia Chopard

Assistant de projet

Laurène Bouvard

Secrétariat

Charlotte Volatier

Chercheurs extérieurs

Dr Dana Hashim

(jusqu'en février 2017)

Dr Hooman Khademi Kohnehshahri
(jusqu'en septembre 2016)

Dr Peng Li (jusqu'en août 2016)

Dr Brent Richards

Boursiers postdoctoraux

Dr Renata Abrahão

Dr Alice Billot-Grasset

(jusqu'en mars 2016)

Dr Robert Carreras Torres

Dr Szilvia Ecsedi

(jusqu'en décembre 2017)

Dr Anouar Fanidi

(jusqu'en avril 2016)

Dr Aida Ferreira-Iglesias

Dr Florence Guida

Dr Tricia Larose

Dr Ruhina S. Laskar

Dr Corina Lesseur Perez

(jusqu'en décembre 2016)

Dr Dariush Nasrollahzadeh Nesheli

Dr Sandra Perdomo Velasquez

(jusqu'en février 2016)

Dr Carolina Santamaria Ulloa

(jusqu'en mars 2016)

Dr Mahdi Sheikh

Dr Karl Smith Byrne

Dr Chanida Vinayanuwattikun

(jusqu'en mai 2016)

Etudiant en Master

Sandrine Magat (jusqu'en août 2017)

Stagiaire

Linda Kachuri

(jusqu'en septembre 2016)

Groupe Prédispositions génétiques au cancer (GCS)

Chef

Dr James McKay

Chercheurs

Dr Behnoush Abedi-Ardekani

Dr Lynnette Fernandez-Cuesta

Dr Matthieu Foll

Dr Florence Le Calvez-Kelm

Dr Maria Zvereva

(jusqu'en août 2017)

Techniciens de laboratoire

Amélie Chabrier

Geoffroy Durand

Nathalie Forey

Bioinformaticien

Catherine Voegele

Secrétariat

Isabelle Rondy

Andreea Spanu (jusqu'en juillet 2017)

Boursiers postdoctoraux

Dr Nicolas Alcalá

Dr Patrice Avogbe

Dr Md Ismail Hosen

Dr Rim Khlifi

Dr Dariush Nasrollahzadeh Nesheli

Etudiants

Salia Bamba

(jusqu'en septembre 2017)

Tiffany Delhomme

Jules Derks

(jusqu'en décembre 2016)

Pauline François (jusqu'en septembre 2016)

Aurélie Gabriel

Théo Giffon (jusqu'en octobre 2017)

Imen Hemissi

Noémie Leblay

Olesia Lole (jusqu'en mai 2017)

Gabriel Roberti de Oliveira

(jusqu'en février 2017)

La Section Génétique (GEN) réunit les Groupes Epidémiologie génétique (GEP) et Prédispositions génétiques au cancer (GCS). Ses travaux combinent de vastes études en population avec des compétences de laboratoire et en biostatistiques, pour identifier des gènes et des profils génétiques particuliers participant à la cancérogenèse et élucider de quelle façon ils exercent leurs effets en interaction avec des facteurs environnementaux. La Section GEN cherche également à identifier les individus qui présentent un risque de cancer suffisamment élevé pour tirer profit des stratégies possibles de dépistage.

Les projets de la Section GEN nécessitent généralement un important travail de terrain en collaboration avec des chercheurs extérieurs, afin de mettre

en œuvre des études épidémiologiques de grande envergure qui collectent des données cliniques et d'exposition adéquates, ainsi que des échantillons biologiques. C'est l'approche adoptée par le Groupe GEP dont les analyses génétiques englobent l'étude de gènes candidats, des études de génotypage pangénomique et d'importants travaux de séquençage. Les recherches du Groupe GEP consistent également à évaluer les expositions aux facteurs non génétiques, en partie à cause de leur impact important sur l'incidence des cancers, mais aussi pour permettre une évaluation précise des interactions gène-environnement. A l'inverse, le Groupe GCS axe davantage ses recherches sur l'identification de variants génétiques rares ou peu courants, susceptibles d'exercer un effet plus important que les polymorphismes nucléotidiques courants, mais dont la

faible fréquence ne permet pas leur détection par les techniques actuelles de génotypage pangénomique. Pour étudier ces variants rares, le Groupe GCS utilise les techniques de génomique et de bioinformatique, en complément d'approches plus traditionnelles. Il s'appuie également sur la génomique pour étudier la façon dont ces variants peuvent conférer une prédisposition génétique au cancer. Le programme de recherche du Groupe GCS vient ainsi compléter celui du Groupe GEP. Il offre par ailleurs une plateforme de techniques génomiques haut débit et des services de bioinformatique, permettant la réalisation des projets d'épidémiologie moléculaire à grande échelle de la Section GEN et d'autres études génomiques conduites au CIRC.

GROUPE EPIDEMIOLOGIE GENETIQUE (GEP)

L'objectif général du Groupe Epidémiologie génétique (GEP) vise à une meilleure compréhension des causes de cancer en étudiant les variants génétiques de prédisposition de différentes localisations anatomiques de cancer et les spectres de mutations génétiques observées dans les tumeurs. Un autre objectif consiste à élaborer des modèles précis de prédiction du risque, qui tiennent compte à la fois des données démographiques (âge et sexe) et des biomarqueurs (génétiques et non génétiques). Les recherches du Groupe GEP concernent notamment les cancers associés à la consommation de tabac et d'alcool (cancers du poumon et des voies digestives supérieures) et les cancers associés à l'obésité (cancers du rein, du pancréas et du côlon-rectum). A cette fin, il consacre d'importantes ressources au travail sur le terrain pour recruter de grandes séries de cas et de témoins, accompagnées d'informations détaillées (questionnaires) et d'échantillons biologiques. Généralement, l'analyse génétique de la prédisposition héréditaire débute par une approche pangénomique, suivie d'études similaires, coordonnées à grande échelle dans différentes populations. Ce dernier

point nécessite la mise en place de consortiums internationaux, étape au cours de laquelle le Groupe GEP joue un rôle majeur. Une fois confirmés, les locus de prédisposition sont étudiés plus en détail, grâce à toute une variété de techniques, notamment des études *in silico* de séquençage et d'expression, souvent réalisées en collaboration avec d'autres Groupes du CIRC. Ces grandes études pangénomiques comportent également une approche de randomisation mendélienne afin d'examiner la façon dont les facteurs liés au mode de vie influencent l'apparition du cancer.

Par ailleurs, le Groupe GEP a entrepris une vaste étude internationale sur l'étiologie du cancer qui consiste à analyser les spectres de mutations (ou signatures mutationnelles) dans les génomes de tumeurs. La plupart des travaux réalisés par le Groupe dans ce domaine font partie du projet *Mutographs* visant à élucider les causes de cinq différents types de cancer sur cinq continents (voir encadré).

Parallèlement à ces recherches sur les facteurs génétiques, le Groupe GEP

poursuit un large éventail d'études portant sur les facteurs non génétiques. Il s'agit notamment d'étudier des biomarqueurs circulants, tels que les anticorps dirigés contre le virus du papillome humain (VPH) pour les cancers de la tête et du cou, et toute une série de protéines et autres biomarqueurs pour le cancer du poumon. L'objectif général de ces études consiste à identifier des individus dont le risque de cancer est suffisamment élevé pour justifier un dépistage et une détection précoce.

Quelques travaux marquants du Groupe au cours la période biennale 2016–2017 sont décrits ci-dessous.

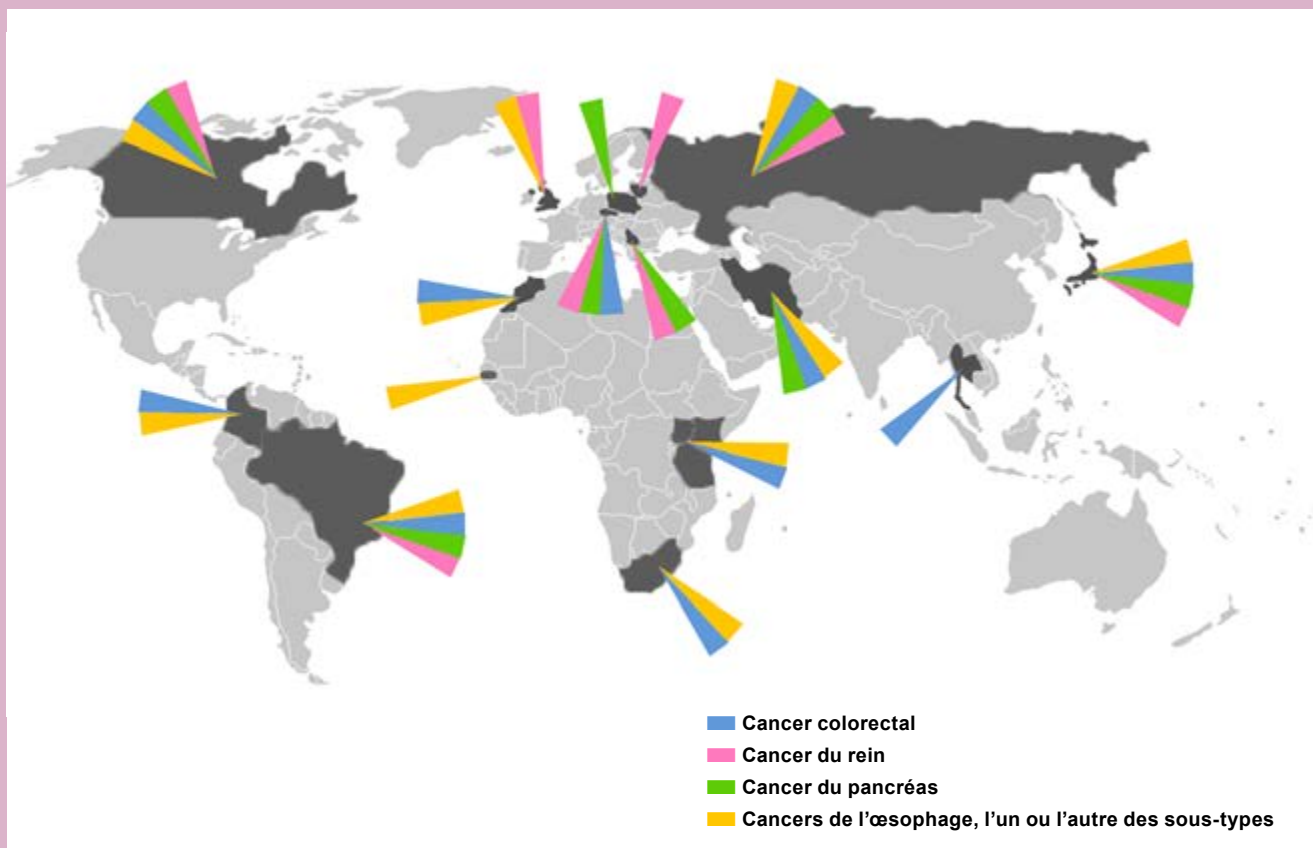
ELUCIDER LE ROLE ETIOLOGIQUE DE L'OBESITE ET DES FACTEURS DE RISQUE CONNEXES DANS PLUSIEURS TYPES DE CANCER – UNE APPROCHE DE RANDOMISATION MENDELIIENNE

Le Groupe GEP a étudié plusieurs types de cancer associés à un indice de masse corporelle (IMC) élevé et à des facteurs de risque liés à l'obésité. Ces facteurs étant fortement interdépendants, les études épidémiologiques traditionnelles

Un nouveau projet majeur de la Section vise à comprendre l'étiologie du cancer en établissant des profils de signature mutationnelle à partir des données de séquençage pangénomique. Cette étude est réalisée grâce à l'obtention d'une importante subvention du programme mondial *Grand Challenge* du Centre de recherche sur le cancer au Royaume-Uni (CRUK pour *Cancer Research UK*) – l'une des subventions de recherche sur le cancer les plus ambitieuses au monde. Le Dr Paul Brennan co-dirige ce projet avec le chercheur principal, le Professeur Sir Mike Stratton du *Sanger Institute* (Cambridge, Royaume-Uni) et quatre autres co-chercheurs. Le projet s'intitule « Comprendre les causes du cancer grâce à l'étude des signatures mutationnelles – Mutographes » (*Understanding of the Causes of Cancer through Studies of Mutational Signatures – Mutographs*).

Différents facteurs de risque de cancer, tant génétiques qu'environnementaux et liés au mode de vie, induisent différents spectres de mutations somatiques dont beaucoup ne sont pas connues. Dans le cadre du projet *Mutographs*, le Groupe GEP coordonne le recrutement de 5000 cas de cancer (cancer colorectal, cancer du rein, cancer du pancréas, adénocarcinome de l'œsophage ou carcinome épidermoïde de l'œsophage) sur cinq continents, afin de savoir si différentes signatures mutationnelles peuvent expliquer les fortes variations d'incidence. Des échantillons biologiques, ainsi que des données démographiques, histologiques, cliniques et tirées de questionnaires, sont recueillis grâce à tout un réseau de collaborateurs. Le *Sanger Institute* assure le séquençage pangénomique des paires « tumeur-ADN germinale ». Les mutations somatiques détectées sont ensuite corrélées avec les données relatives aux facteurs de risque.

En apportant une meilleure compréhension de l'étiologie du cancer, ce travail sans précédent devrait permettre d'identifier les facteurs de risque modifiables, conduire à de nouvelles approches de prévention du cancer, et offrir des opportunités pour renforcer la détection précoce, préciser les groupes à haut risque et contribuer au développement de techniques thérapeutiques.



Pays participant au recrutement du projet *Mutographs*.

n'ont pas réussi à démêler ceux qui ont une influence déterminante et ceux qui sont simplement corrélés avec un facteur déterminant sous-jacent.

En exploitant les données des études d'association pangénomique sur des dizaines de milliers de cas de cancer et de témoins, conduites par le Groupe GEP ou auxquelles il a participé, nous avons réalisé une série d'analyses concernant la pertinence du rôle étiologique de plusieurs facteurs de risque liés à l'obésité dans différents cancers. Ces analyses s'appuyant sur des outils génétiques ne sont pas influencées par la causalité inverse et sont moins sensibles aux facteurs de confusion que celles qui s'appuient sur des mesures directes d'exposition. Nous avons ainsi obtenu des résultats édifiants pour les cancers

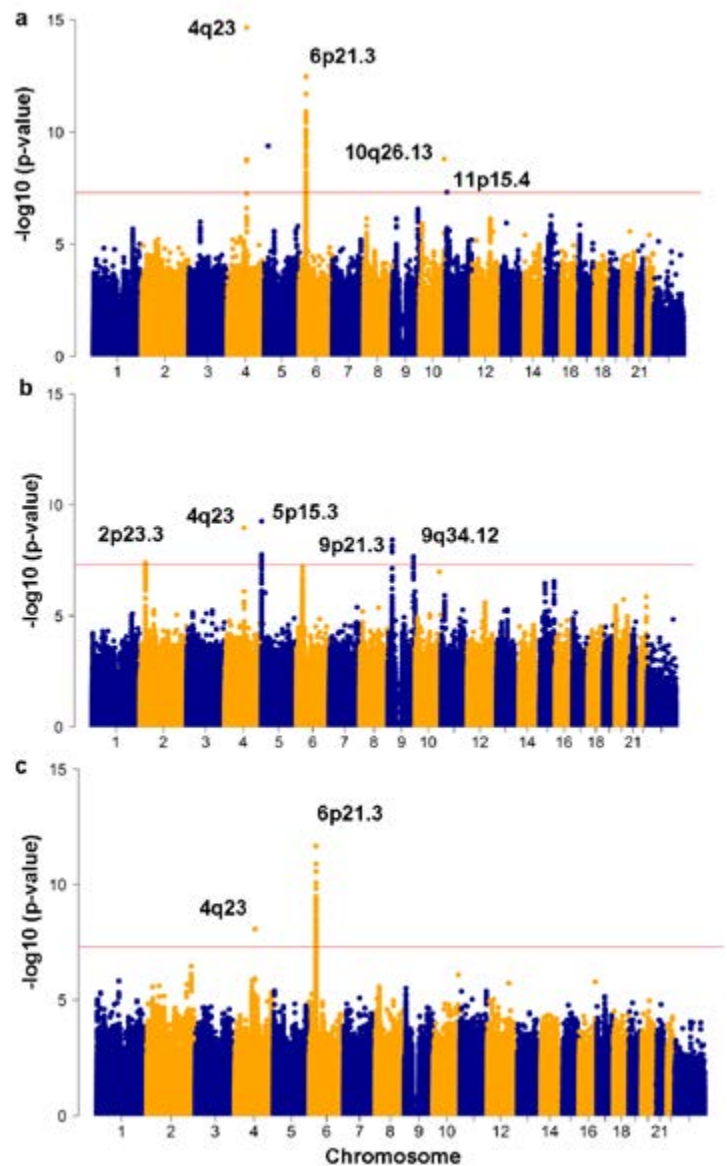
du rein et du pancréas, pas seulement parce qu'ils confirment le rôle étiologique d'un IMC élevé dans ces cancers, mais surtout parce qu'ils mettent en lumière le rôle médiateur majeur de l'insuline dans ce risque accru associé à un IMC plus élevé. Ces résultats permettent de mieux comprendre l'importance de l'obésité dans l'étiologie des cancers du rein et du pancréas. Par ailleurs, ils suggèrent un rôle potentiellement important de l'obésité et de la résistance à l'insuline dans les cancers du poumon (Carreras-Torres et coll., 2017a, 2017b).

ANALYSE PANGÉNOMIQUE DES CANCERS ASSOCIÉS À LA CONSOMMATION DE TABAC

Le Groupe GEP a coordonné une grande analyse OncoArray sur plus de 7000 cas de cancer de la cavité buccale ou de

l'oropharynx et sur un nombre identique de témoins. Un résultat frappant de cette étude révèle le rôle majeur de la région HLA dans le cancer oropharyngé. Une analyse plus poussée de ce locus a permis d'identifier un haplotype HLA également associé au cancer du col utérin, ce qui suggère une importante interaction avec cet haplotype particulier et le virus du papillome (Lesseur et coll., 2016) (Figure 1). Par ailleurs, dans le cadre de la vaste étude d'association pangénomique, menée par le Groupe GEP en partenariat avec le *National Cancer Institute* des Etats-Unis, sur plus de 10 000 cas de cancer du rein et 20 000 témoins, sept nouveaux loci de prédisposition pour ce cancer ont été identifiés et viennent s'ajouter aux six loci précédemment découverts (Scelo et coll., 2017).

Figure 1. Résultats de la méta-analyse des études d'association pangénomique. Les lignes rouges correspondent à $P = 5 \times 10^{-8}$. Sur les axes verticaux figure l'échelle logarithmique des valeurs p [$-\log_{10}(\text{valeur-}p)$]. a) Analyse globale des cancers de la cavité buccale et de l'oropharynx portant sur 6034 cas et 6585 témoins. b) Analyse du cancer de la cavité buccale portant sur 2990 cas et 6585 témoins. c) Analyse du cancer de l'oropharynx portant sur 2641 cas et 6585 témoins. La localisation génomique est précisée pour les loci présentant d'importants polymorphismes nucléotidiques sur l'ensemble du génome et validés au plan technique. D'après Lesseur et coll. (2016) avec l'autorisation de Macmillan Publishers Ltd, copyright 2016.



GROUPE PREDISPOSITIONS GENETIQUES AU CANCER (GCS)

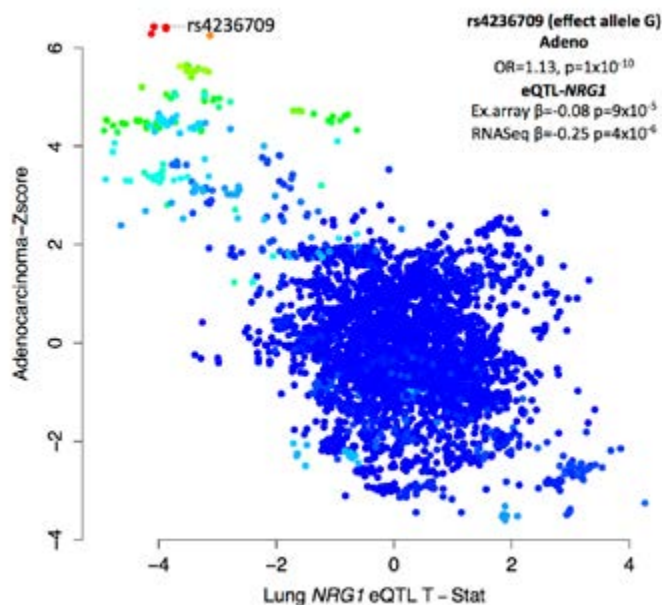
Le Groupe Prédilections génétiques au cancer (GCS) est formé d'une équipe scientifique multidisciplinaire disposant des compétences en génétique, génomique, bioinformatique et en pathologie. Cette multidisciplinarité permet d'entreprendre des études génétiques et génomiques pour identifier des gènes associés au cancer et explorer leurs mécanismes d'action, afin de mieux comprendre l'étiologie de la maladie et pouvoir ainsi améliorer sa détection précoce et sa prévention.

Au sein du consortium international d'études cas-témoin sur le cancer du poumon ILCCO (*International Lung Cancer Case-Control*) et du consortium OncoArray de l'Initiative GAME-ON (*Genetic Associations and Mechanisms in Oncology*) du *National Cancer Institute* des Etats-Unis, le Groupe GCS a entrepris une étude d'association

pangénomique sur près de 30 000 cas de cancer pulmonaire et 57 000 témoins. Cette étude a permis d'identifier 18 loci de prédisposition, dont 10 nouveaux. Ces allèles de prédisposition ont été examinés en combinant les données supplémentaires de plus de 250 000 personnes, notamment des données concernant l'expression des gènes dans les tissus pulmonaires et autres tissus, la propension au tabagisme, la spirométrie (capacité vitale forcée [CVF]/volume expiratoire maximal pendant la première seconde [VEMS]) et la longueur des télomères des leucocytes (McKay et coll., 2017a). L'analyse de l'expression génique dans l'épithélium pulmonaire a mis en évidence des gènes non impliqués précédemment dans l'étiologie du cancer du poumon, tels que *RNASET2* ou *SECISBP2L*, ainsi que des gènes tels que *NRG1*, un gène faisant rarement l'objet d'une translocation somatique

dans les adénocarcinomes pulmonaires (Figure 2). Par ailleurs, nous avons identifié au locus 8p12, des variants de prédisposition au cancer pulmonaire qui influencent les niveaux d'expression de *CHRNA2* codant pour un récepteur nicotinique cholinergique. Contrairement à nos précédentes observations sur les variants génétiques dans les tumeurs pulmonaires et les gènes codants pour des récepteurs nicotiniques cholinergiques, ces variants 8p12 n'étaient pas associés à la quantité de tabac fumée, mais à l'âge de la première cigarette. Ils semblaient également liés aux niveaux d'expression de *CHRNA2*, notamment dans le cervelet, renforçant les preuves naissantes indiquant que cette région du cerveau pourrait jouer un rôle dans certains aspects de la dépendance. Par ailleurs, le Groupe GCS a obtenu une subvention (accordée par France Génomique et l'Institut national

Figure 2. Diagramme de dispersion comparant 5417 variants au locus de prédisposition 8p12 et leur association avec l'adénocarcinome pulmonaire (axe vertical) et les loci de caractères quantitatifs d'expression cis pulmonaire (eQTL pour *expression quantitative trait loci*), d'après les données issues du projet GTEx (pour *Genotype-Tissue Expression*) (axe horizontal). La couleur de chaque variant est fonction du degré de déséquilibre de liaison (R^2) avec un variant sentinelle pour le cancer du poumon (étiqueté rs4236709) à ce locus (rouge pour degré élevé, bleu pour degré faible). Tableau en haut à droite : association entre le variant sentinelle et l'adénocarcinome pulmonaire ainsi qu'avec les eQTL dans l'épithélium pulmonaire au sein de cinq cohortes, niveaux d'expression de l'ARN déterminés grâce aux techniques de puces à ADN et de séquençage RNASeq. Les variants associés à l'adénocarcinome pulmonaire ont tendance à être ceux qui sont des eQTL-*cis* pulmonaire pour *NRG1*. De façon surprenante, alors que les mutations somatiques par translocation sont généralement associées à l'activation ectopique de *NRG1* et aux non-fumeurs, le risque génétique germinal était corrélé avec une diminution de l'expression de *NRG1* et il était présent dans les adénocarcinomes pulmonaires des fumeurs et des non-fumeurs. © CIRC.



du Cancer, France) pour continuer à étudier les mécanismes d'action de ces variants de prédisposition génétique. Il a aussi initié des projets visant à étudier les événements génomiques dans des tumeurs thoraciques rares, telles que les tumeurs carcinoïdes pulmonaires (projet soutenu par la Ligue nationale contre le Cancer Rhône-Alpes, France ; la *Dutch Cancer Society*, Pays-Bas et le *National Cancer Institute*, Etats-Unis) et le mésothéliome pulmonaire (projet soutenu par l'Institut national du Cancer, France).

Le Groupe GCS étudie également la possibilité d'utiliser l'ADN tumoral circulant (ADNtc) comme biomarqueur pour la détection des cancers. Cette étude présente d'importantes difficultés techniques. En effet, la qualité de l'ADN et les fréquences alléliques des modifications génétiques d'origine tumorale dans l'ADN libre total circulant sont inférieures aux seuils de sensibilité des séquenceurs de nouvelle génération. De plus, si l'on veut utiliser les mutations comme biomarqueurs pour la détection précoce des cancers, il faut pouvoir les identifier sans connaître au préalable le génotype de la tumeur et sur n'importe

quelle région du gène. En combinant nos compétences de laboratoire et en bioinformatique, nous avons développé un pipeline d'analyse spécifiquement conçu pour l'ADNtc, baptisé Needlestack (<https://github.com/IARCbioinfo/needlestack> ; Figure 3). En appliquant cette approche à des études cas-témoins rétrospectives sur les cancers du poumon et du pancréas, nous avons démontré la présence d'ADNtc chez les cas de cancer et, plus important, chez les patients au stade précoce de la maladie (Fernandez-Cuesta et coll., 2016 ; Le Calvez-Kelm et coll., 2016). Alors que la présence d'ADNtc est détecté chez une bonne partie des patients, on a également noté des mutations généralement associées au cancer chez un pourcentage inattendu de témoins (~3–10 %). Ce dernier résultat quelque peu surprenant, en tenant compte des limites que cela implique, donne une idée des informations qui pourraient être obtenues grâce à l'application de ces techniques aux études d'épidémiologie moléculaire conduites par le CIRC. Dans cette perspective, nous examinons à présent leur application à l'étude d'autres cancers, notamment celui de la vessie (avec le soutien de l'Association pour

la Recherche en biologie moléculaire, France, et La Ligue nationale contre le Cancer Rhône-Alpes, France) et de l'œsophage (avec le soutien du *National Institute for Medical Research Development*, République islamique d'Iran).

Le Groupe GCS participe activement au développement des capacités du CIRC en matière de génomique. En effet, avec la contribution essentielle des autres Groupes, il facilite les interactions au sein du Centre dans le domaine de la génomique non seulement en participant au Comité directeur de biologie intégrative, bioinformatique et biostatistique et au Groupe associé de travail sur la bioinformatique, ainsi qu'au Comité directeur des laboratoires, mais aussi en mettant à disposition des techniques de laboratoire, des services d'expertise en pathologie et des ressources informatiques pour les recherches conduites au Centre, faisant appel aux approches génomiques. Ces progrès profitent également à l'ensemble de la communauté scientifique *via* un site GitHub: <https://github.com/IARCbioinfo/>.

Figure 3. Deux exemples de variants détectés en utilisant le modèle de régression de Needlestack permettant d'identifier une fraction allélique rare chez des individus aberrants (porteurs du variant). Chaque point représente un individu séquencé (deux points par échantillon) dont la couleur est fonction de sa valeur-*q* d'après l'échelle de scores « phred ». La courbe de régression en noir représente le taux estimé d'erreur de séquençage et l'intervalle de confiance à 99 % (courbes en pointillés noirs) des échantillons. Les courbes en pointillés de couleur correspondent aux limites des régions définies par différents seuils de signification de la valeur-*q*. Les deux dupliques techniques sont identifiés comme des valeurs aberrantes de la régression (en rouge) et sont par conséquent considérés comme porteurs de la mutation donnée. D'après Fernandez-Cuesta et coll. (2016). Copyright 2016, avec l'autorisation d'Elsevier.

