



# Génomique des cancers, outils et état des lieux

## OUVERTURE

---

***Pierre FORMSTECHE***

*Président du Cancéropôle Nord-Ouest*

Pierre Formstecher souhaite la bienvenue aux participants à cette session organisée sous l'égide du Cancéropôle Nord-Ouest sur le thème de « la génomique des cancers, outils et état des lieux ». Ce sujet est d'autant plus d'actualité que des progrès fulgurants ont été accomplis en matière de séquençage du génome au cours des précédentes années, ce qui a des conséquences significatives en matière d'exploration médicale.

Le Cancéropôle Nord-Ouest est justement doté d'une plate-forme de génomique parmi les mieux équipées de France, grâce au travail de longue haleine menée par Jean-Pierre Kerckaert et au soutien de ses partenaires ; l'Université, l'INSERM et l'Institut de Recherche contre le Cancer. Sous l'égide de Martin Figeac, l'équipe du plateau s'est renouvelée et enrichie en intégrant de nouvelles compétences, notamment en matière de bio informatique et d'exploitation des résultats.

Cette plate-forme de génomique travaille en étroite collaboration avec son homologue de Rouen au sein du Cancéropôle Nord-Ouest. Elle bénéficie du label IBiSA au sein de la plate-forme Ligand. Cette dernière est spécialisée en génomique multi factorielle tandis que la plate-forme lilloise est davantage tournée vers des applications médicales. Le cancer occupe une place très importante au sein de cette plate-forme. Ainsi, cet outil a-t-il permis de mener les premiers essais pilote du SOLiD3 et des études sur les patients, notamment grâce au financement du GEFLUC sur le séquençage visant à rechercher des anomalies génétiques dans des cas de syndromes de Lynch en l'absence de mutation.

Pierre Formstecher souligne le coût élevé de ces nouveaux outils, sans compter celui des réactifs et des analyses. A cet effet, en collaboration avec le Cancéropôle, animé par Véronique Pancré et Sophie Broutin, il souhaite réfléchir à la manière de stimuler les équipes afin qu'elles tirent le meilleur parti de ces nouveaux outils. Les technologies en question sont développées par des partenaires industriels, qui sont impliqués au premier chef lors de l'implantation des outils et de leur valorisation scientifique, et qui sponsorisent cette manifestation (Affymetrix, Agilent, Applied Biosystems et Sigma-Aldrich).

En guise de conclusion, Pierre Formstecher présente le programme de cette journée, orientée autour de deux volets, l'un technologique et l'autre médical (les thumorothèques, les registres et l'hémopathie). Cette journée a été à l'initiative de l'axe 2 du Cancéropôle (« initiation et évolution tumorales des hémopathies malignes »), mais l'intérêt suscité par ce projet a incité le Cancéropôle à élargir la réflexion au-delà de l'onco hématologie. Cette réunion scientifique suscite également la participation de jeunes chercheurs, à travers des posters, qui seront récompensés.

## PREMIERS RESULTATS DU SOLID3, APPLIED BIOSYSTEMS

---

**Martin FIGEAC**

*Responsable de la plateforme de génomique fonctionnelle et structurale, IRCL, Lille*

Martin FIGEAC présente la plate-forme de génomique fonctionnelle, service commun de l'Université de Lille II, rattaché à l'Institut de médecine prédictive et de recherche thérapeutique, et partenaire de l'IRC de Lille. Cette plate-forme est organisée en quatre plateaux :

- trois plateaux technologiques répondant à des demandes précises :
  - le plateau Agilent permettant de réaliser des études sur la variation du nombre de copies au niveau de l'ADN génomique : des travaux récents ont permis de détecter des aberrations de l'ordre de 3 kilobases, sur des échantillons partiellement dégradés, par exemple.
  - le plateau Affymetrix, qui utilise notamment la technologie Snib6 qui permet de mettre sur une puce 1,8 millions de sondes afin de faire du génotypage sur la moitié et d'étudier le nombre de copies sur le restant (soit un espace médian entre deux sondes de 700 paires de bases !)
  - le plateau SOLiD, qui permet de faire du séquençage haut débit de courts fragments
- un plateau de bio informatique et de bio statistique, qui permet de répondre aux demandes générées par ces trois plateaux technologiques, ainsi qu'à des questions posées dans le cadre de projets annexes.

Cette plate-forme comporte 11 personnes, notamment Jean-Pierre Karcar, consultant de marque, dont Martin Figeac salue l'arrivée récente au sein de cette équipe.

Le séquençage haut débit permet d'étudier simultanément, en une expérimentation, 400 millions de fragments, dont chacun représente 50 paires de base, soit une sortie brute de 20 gigabases pour chaque expérimentation. Ces 20 GB peuvent être utilisés conjointement pour l'étude d'un seul échantillon, avec une définition extrême, ou être multiplexés sur plusieurs échantillons, ce qui permet d'analyser jusqu'à 160 échantillons par expérimentation. Le SOLiD permet également de réaliser deux expérimentations en parallèle, ce qui permet de gagner du temps et de réaliser l'analyse de 320 patients.

Martin Figeac souligne l'importance du système de séquençage par ligation, ce qui permet de détecter de manière très efficace les mutations par rapport aux erreurs de séquençage. Les applications découlant du séquençage haut débit sont extrêmement nombreuses, mais dans son exposé, Martin Figeac s'intéresse plus particulièrement :

- au reséquençage ciblé d'une région d'ADN génomique ou de multiples régions d'ADN génomique, à l'étude par séquençage du transcriptome ;
- à l'étude par séquençage des microARN ;
- à l'étude de l'immuno-précipitation de la chromatine ;
- l'étude de la méthylation par séquençage ;
- aux études de DGE et de SEDGE.

## I. Le séquençage du transcriptome : RNA-seq

### 1. L'intérêt de la méthode SOLiD

Martin Figeac présente son travail de séquençage des ARN messagers avec SOLiD. Les 20 GB génèrent un certain nombre de fragments qui vont permettre d'obtenir tous les différents transcrits ainsi qu'une information quantitative. En effet, le nombre de fragments séquencés va correspondre au nombre de transcrits exprimés au sein des cellules. A l'occasion du séquençage des ARN messagers, certains des fragments vont « mapper » les jonctions exon-exon, ce qui permettra de reconstruire partiellement les transcrits. De même, il sera possible de détecter les fusions de gènes, puisque certains fragments vont recouper deux gènes distincts.

Martin Figeac rappelle que les séquences des transcrits permettent de détecter les mutations, sans les *a priori* induits par l'utilisation d'une puce à ADN qui ne permet d'étudier que l'expression d'un certain nombre de gènes connus. Le RNA-seq supprime cette idée *a priori* en permettant de séquencer les transcrits de manière brute. Ceci permet de découvrir de nouveaux exons, de nouveaux transcrits et de nouveaux gènes, ce qui revêt une importance primordiale, notamment dans le cadre de l'application sur les miRNA, qui ne sont pas forcément connus pour toutes les espèces.

### 2. LAL : une étude quantitative des transcrits

Martin Figeac se propose de présenter un projet portant sur les Leucémies Aigues Lymphoblastiques (LAL) de l'enfant, et plus particulièrement le séquençage des transcrits après déplétion des ARN ribosomiaux. Deux lignées ont été chacune rendues résistantes ou sensibles à un run : il en résulte quatre échantillons, qui sont comparés par RNA-seq.

Martin Figeac présente un exemple de résultat obtenu à travers un histogramme de la couverture (nombre de fragments qui couvrent chaque paire de bases). L'interloking 23 n'est pas du tout couverte. Le transcrit n'est pas du tout exprimé au sein des cellules. En revanche, pour l'interloking 12 bêta 2, les exons sont présents au sein des cellules. Au-delà des aspects quantitatifs, ce résultat permet de constater que ce gène est transcrit à partir du brin positif. Aucun fragment séquencé ne mappe sur la SERBP1, contrairement aux compléments des fragments. On sait que ce gène est transcrit à partir du brin négatif.

### 3. La détection des transcrits très faiblement exprimés

Les mêmes échantillons ont été comparés sur RNA-seq et sur les puces Affymetrix. La corrélation entre l'intensité moyenne mesurée pour chaque gène et le nombre de fragments moyens qui mappent dans le même gène est très forte, de l'ordre de 0,81. Au total, 95 % des sondes sont marquées comme « absentes » par la puce Affymetrix (son intensité mesurée est au niveau du bruit de fond). En revanche, les résultats d'analyse par le SOLiD restent dynamiques. L'étude du transcriptome par séquençage permet donc de réaliser les transcrits très faiblement exprimés.

Martin Figeac donne également l'exemple des :

- **Topoisomerase II beta**

Les informations obtenues pour chaque exon permettent de déterminer quels sont les transcrits qui sont réellement exprimés au sein des cellules. Seul le grand transcrit est effectivement présent au sein des cellules.

- **Adam 15**

Adam 15 est connu pour son rôle dans la progression métastatique. En fonction des différents transcrits présents au niveau des cellules, il aura un rôle plus ou moins important au niveau de la progression. Tout est « mappé » sur le brin positif, ce qui concorde avec les données de l'UCFC.

Si on tente de mapper chaque pic avec l'un des exons, on se rend compte de l'existence de données légèrement incohérentes. Un exon présent dans quasiment tous les transcrits n'est pas présent dans les cellules. Il s'agit donc d'un épissage alternatif, non référencé par les bases de données. Ceci permet d'étudier le transcriptome avec une résolution jamais atteinte auparavant.

En étudiant les fragments qui font la jonction entre les exons, on peut obtenir une information encore plus précise (exemple du DDT-D dopachrome tautomerase). Des fragments mappent à l'intérieur des exons tandis que d'autres réalisent les jonctions. Ces informations permettent de reconstruire entièrement le transcrit.

#### 4. La détection de mutations au sein des transcrits

SOLiD présente la particularité de permettre le séquençage de couples de bases. Le fragment séquencé est attaché à un adaptateur, le P1.

Un premier primer se pose et reconnaît l'adaptateur, ce qui permet de séquencer par fragments de cinq paires de bases. Les quatre fluorochromes permettent de marquer les deux premières bases. Par exemple, le couple complémentaire AT-TA va être marqué en rouge.

Le séquençage intervient donc par ligation, deux bases toutes les cinq paires de bases. Avec le premier primer, on va séquencer les positions 1 et 2 avec une couleur, puis les positions 6 et 7 avec une autre couleur et ainsi de suite, dix fois de suite, pour couvrir les 50 paires de bases. On recommence avec un deuxième primer, qui sera décalé d'une position, et ainsi de suite avec cinq primers qui seront chacun décalés d'une position.

Ceci permet de séquencer intégralement les cinquante paires de bases. Chaque base (ou position) a été lue à deux reprises (une fois avec sa voisine de gauche et une autre avec sa voisine de droite) ce qui permet d'établir de bonnes statistiques sur le séquençage. L'intérêt de ce système de séquençage réside dans la détection de mutations.

On séquence des suites de couleurs qui vont correspondre à des bases. Par exemple, pour séquencer cCt, on séquence du bleu suivi de jaune (Ct). Une mutation ponctuelle va donner lieu à un changement de deux couleurs consécutives. Si une couleur va correspondre à une erreur séquençage, deux couleurs correspondront plus vraisemblablement à une mutation.

Cette méthode permet d'obtenir une statistique fiable : chaque base a été étudiée à deux reprises et parmi les seize couples possibles de dix bases, trois seulement vont correspondre à une mutation valide. Ceci permet de distinguer de manière très efficace une erreur de séquençage d'une mutation et d'établir des statistiques fiables.

Martin Figeac cite également l'exemple de SETMAR dans le cadre de la détection des mutations dans les transcrits. SETMAR est connu pour son implication dans les leucémies et au niveau du cycle cellulaire et de la réplication.

## **II. L'analyse des mutations à partir d'ADN génomique**

Martin Figeac présente le projet d'identification des nouvelles mutations responsables du syndrome de Lynch à l'aide de la technologie SOLiD, mené par Henri-Pierre Busine, en collaboration avec Sophie Lejeune et Sylvie Manolivier au sein de la plate-forme de génomique.

### **1. Présentation de la problématique**

Le syndrome de Lynch est la principale cause de cancer colorectal et de cancer de l'utérus héréditaires. Il s'agit d'une pathologie à transmission autosomique dominante. Ce syndrome est diagnostiqué suivant les critères d'Amsterdam : au moins trois cas dans une même famille ayant développé un cancer colorectal ou du spectre, liés au premier degré et dont l'un des cancers a été développé avant l'âge de 50 ans. Il est lié la survenue précoce de cancers, qui est liée à la mutation des gènes de réparation de mésappariement de l'ADN.

### **2. Principes de l'étude**

La présente étude s'intéresse plus particulièrement aux protéines MSH2 et MSH6, même si PMS1 et PMS2, entre autres, sont également impliquées. Les porteurs des mutations sont soumis à un risque très élevé de développer des cancers précoces. La détection de ces mutations permet d'identifier au sein de la famille les porteurs de ladite mutation, ce qui permet d'intensifier la surveillance sur les personnes concernées et de rassurer celles qui ne sont pas porteuses de la mutation au sein de la famille et qui ne la transmettront pas à leur descendance. Pendant longtemps, cette mutation était détectée dans 50 % des cas, mais ce taux est passé à 70 % depuis 2004.

Cette étude porte sur sept familles répondant aux critères d'Amsterdam, qui présentent une perte d'expression des protéines MSH2 et MSH6 au sein de la tumeur. Après analyse, il n'y aucune mutation au niveau des séquences des exons, et des régions flanquantes de ces exons et des régions promotrices. Les CGH array de haute densité, avec une résolution de 100 paires de bases, sur la région MSH2 et MSH6, n'a permis de détecter aucune aberration.

A l'heure actuelle, les chercheurs restent incapables de détecter la mutation au niveau génomique. C'est pourquoi l'équipe de la plate-forme a utilisé la technologie SOLiD pour séquencer une région de 521 kb autour de MSH2 et MSH6 pour détecter des mutations dans les régions non codantes de cette région. On dispose de sept cas index et de leurs parents non atteints et également, dans deux familles, de deux cas supplémentaires atteints, ce qui permet de combiner plusieurs stratégies. On dispose d'un pool de porteurs des mutations, d'un pool de cas sains (avec deux allèles sains) et d'individus porteurs au sein de la même famille. En combinant ces trois approches et en comparant les résultats obtenus avec les bases de données et des données de séquençage externes, on va pouvoir déterminer un ensemble de mutations qu'on testera sur davantage d'échantillons.

Ce ciblage sur la région de 500 kb permet d'étudier plusieurs échantillons en cours d'expérimentation. Pour réaliser l'enrichissement de l'ADN afin de cibler la région à séquencer, il existe plusieurs techniques :

- l'enrichissement par PCR, qui permet de cibler des régions de centaines de paires de bases ;
- la réalisation de long ranges PCR, qui permet de cibler des régions de dizaines, voire de centaines, de kilobases ;
- la mise en œuvre d'une capture grâce à une puce à ADN, qui permet de capturer une région d'1 MB (mais cela nécessite une quantité importante d'ADN) ;
- le recours à la technologie SureSelect de la société Agilent, qui part du même principe que précédemment avec des oligos de 120 paires de bases qui reconnaîtront les séquences et seront capturés par les billes magnétiques (avec une quantité d'ADN tout à fait raisonnable, de l'ordre de 1 à 3 microgrammes).

### 3. Ciblage des oligo nucléotides

L'équipe de la plate-forme a donc opté pour la solution proposée par Agilent. Celle-ci nécessite au préalable de désigner les oligos qui devront reconnaître l'ADN de la région concernée. La société Agilent dispose d'une base de données avec des oligos qui ont déjà été désignés. Il n'y a pas d'oligo par région marquée par repeat masker. Pour la région 500 kb, dans la base de données proposée par Agilent, il y a 7 364 oligos nucléotides de 120 paires de bases qui permettent de réaliser la capture, ce qui permet de couvrir seulement 42 % de la région étudiée (58 % de la région est située dans des régions de faible complexité).

Il était donc nécessaire de procéder d'une autre manière. Un premier design a été réalisé sans repeat masker, en prenant tous les oligos disponibles dans la base de données Agilent, soit environ 22 000 oligos nucléotides. Chacun d'entre eux a été mappé sur le génome humain, ce qui a permis d'obtenir 17 000 oligos qui mappaient plus ou moins de manière unique. Ceci permet d'avoir 11 % seulement des régions non couvertes par des oligos (il s'agit de régions répétées). Or, l'objectif étant la découverte de mutations, capturer de l'ADN à plusieurs endroits du génome ne représente que peu d'intérêt.

Martin Figeac présente un exemple d'échantillon séquencé, avec un taux de couverture de 92 %. Tous les échantillons séquencés atteignent au moins un taux de 90 % (86 % si chaque base doit être couverte par au moins 10 fragments, 78 % pour au moins 20 fragments). Le second exemple porte sur les séquences qui mappent de manière unique sur le génome : le taux de couverture moyen s'élève à 143x et peut aller jusqu'à 800x.

Certaines régions correspondent à une perte brutale de la couverture, ce qui correspond peut-être à un biais de séquençage ou de capture. Sur ce point, l'analyse doit être poursuivie. Néanmoins, Martin Figeac conclut que ces données, toutes récentes, reflètent des résultats très satisfaisants en matière de capture et de séquençage, qui permettent de détecter les mutations de manière très efficace.

## APPROCHE MIRNA SUR SOLID3

---

**Pascal BARBRY**

Directeur de l'IPMC, Nice

Pascal Barbry explique qu'initialement, l'intérêt des micro-ARN résidait dans leur nombre limité, ce qui en faisait un territoire *a priori* beaucoup plus simple à explorer. La réalité s'est avérée beaucoup plus complexe puisqu'il est nécessaire de développer des stratégies expérimentales élaborées pour essayer de comprendre comment ces micro-ARN fonctionnent.

Dans ce cadre, le développement d'une plate-forme IBISA au cours des dix dernières années a été caractérisées par :

- le passage de certaines technologies, comme Affymetrix, dans le domaine clinique.
- une évolution technologique importante à travers laquelle le séquençage allait devenir la méthode de référence.

De ce fait, Pascal Barbry et son équipe sont en train de transférer tous leurs protocoles des technologies de puces vers les technologies de séquençage, en utilisant au maximum la technique du multiplexage, afin d'obtenir des coûts comparables à ceux d'une étude de puces classique.

### III. Principe de fonctionnement des micro-ARN

Pascal Barbry rappelle que la relative stabilité du nombre de gènes codant pour des protéines à travers le règne animal ne correspond pas exactement à la taille des génomes. Le ratio entre les quantités d'ADN non codant et la quantité d'ADN transcrit s'inverse selon qu'il s'agisse de prokaryotes ou de vertébrés. Ceci traduit l'apparition d'un nouveau système de molécules connu depuis des années par l'intermédiaire de mécanismes biologiques complexes (activation du chromosome X). Ces exemples biologiques établissent parfaitement le fait qu'on n'a pas besoin d'être une protéine pour avoir un rôle biologique important.

Depuis 1993 et les premiers travaux sur Lin-4, il a été montré qu'une séquence courte d'une vingtaine de nucléotides peut inter agir à trois primes non codants d'un ARN (à savoir Lin-14 dans le cadre de Lin-4). Ce système est resté largement ignoré pendant plusieurs années jusqu'à ce que dans les années 2000, Slack & Ruvkun identifient la conservation de la séquence Let-7, ce qui a conduit à l'identification d'un millier de nouvelles séquences chez l'homme. Le micro-ARN ayant la capacité d'interagir sur ses cibles, principalement par l'intermédiaire d'une région de six à huit nucléotides (le seed, en général stabilisé par les séquences 3' de ce micro-ARN). La reconnaissance s'effectuant à partir d'un nombre limité de bases ( $4^7$  combinaisons possibles), une grande partie du génome peut être ciblée par ces micro-ARN. De ce fait, une fraction importante du transcriptome pourrait être effectivement régulée par les micro-ARN.

Selon le modèle initial, un micro-ARN (Lin-4) interagit avec l'ARN Lin-14, et un phénotype est observé. Cette situation est intermédiaire : certains micro-ARN interagissent probablement avec un seul micro-ARN messenger, tandis d'autres micro-ARN en ciblent plusieurs. Pascal Barbry souligne que ce point doit encore être précisé. De la même façon, il faut imaginer

différents mécanismes faisant suite à la mise en place de ce mécanisme de régulation post-transcriptionnel. Différents modèles ont été proposés :

- la deadenylation ;
- l'inhibition de la traduction ;
- le clivage de l'ARN messager ;
- voire la relocalisation du complexe micro-ARN mRNA dans les P-bodies, qui pourraient servir de zones de stockage jusqu'à réutiliser l'ARN messager lorsque les conditions redeviennent adéquates.

#### IV. MiR-34

MiR-34 apparaît comme un effecteur important de p53, ce qui pourrait expliquer un certain nombre de mécanismes d'inhibition portés par p53, miR-34 étant capable de « silencer » un certain nombre de gènes après son activation par p53.

Au sein du département « hématologie » de l'hôpital de Nice, Pascal Braby et sa collègue Sophie Raynaud ont observé qu'une délétion en 11q était souvent associée à une forme régressive de CLL. Depuis des années, cette délétion est associée à ATM. Or, ces deux chercheurs ont constaté qu'à proximité immédiate d'ATM, se trouvaient miR-34b et miR-34c (les séquences de miR-34 existent souvent trois formes, a, b et c). En outre, ils ont constaté que les délétions observées pour ATM associaient dans tous les cas sauf un une délétion de miR-34b et c, au point qu'ils se sont demandés si cela ne pouvait pas participer à l'aggravation de la pathologie. Parmi les candidats, cibles potentielles pour miR-34, la molécule TCL1 est exprimée de façon plus importante dans des formes sévères de CLL.

A travers cette expérience, Pascal Barbry est parvenu à démontrer une interaction spécifique entre les différents miR-34 et le 3' non codant de TCL1, à travers un seul des trois micro-ARN, qui correspond à la forme de miR-34b. Or, miR-34b a une séquence légèrement différente des autres : la séquence s'aligne parfaitement avec a et c, mais est décalée d'une base au niveau du seed, ce qui peut modifier la nature des séquences qui sont reconnues. Il est possible de s'en rendre compte en mesurant l'activité de TCL ou en faisant des expériences dans lesquelles on va sur-exprimer miR-34 a, b et c dans des cellules, auquel cas le profil des gènes qui sont réprimés transcriptionnellement sera différent entre miR-34 a, b et c. En regardant quelles sont les séquences qui pourraient être spécifiquement enrichies dans les différents transcrits inhibés, on retrouve effectivement certaines séquences, les séquences de seed de ces molécules, avec notamment la séquence qui correspond à miR-34 b.

Pascal Barbry en conclut que :

- il existe une correspondance entre cette représentation dans laquelle miR-34b pourrait jouer un rôle légèrement différent que miR-34a et c ;
- la délétion de miR-34 b pourrait avoir un impact fonctionnel (et participer à l'agressivité de ces CLL)

#### V. Comment l'expression de ces micro-ARN est-elle contrôlée ?

La cellule dépense une énergie considérable pour générer ces molécules. Les micro-ARN sont transcrits à partir d'un système classique. Au niveau du noyau, le complexe microprocesseur, reconnaît des Rpin au niveau des ARN messagers. Une fois dans le cytoplasme, une 2<sup>ème</sup> coupure intervient. Dicer, en association avec d'autres molécules comme TRBP, peut générer une seconde coupure. A ce moment-là, deux brins correspondent aux deux parties de la tige « boucle ». Une des deux formes va être prise en charge par le complexe « risque », qui

comprend notamment les protéines argonautes. Ce complexe interagit avec les ARN messagers.

Pascal Barbry souligne que les interventions se multiplient dans la littérature scientifique s'agissant de l'un ou l'autre mécanisme de régulation des micro-ARN et de contrôle de cette régulation. Ces études portent notamment sur les interactions entre les microprocesseurs et certains complexes nucléaires. Des systèmes de régulation peuvent cibler plus particulièrement certains micro-ARN et modulent le niveau de production du précurseur. Il existe des systèmes d'édition, qui peuvent toucher d'autres ARN comme 942 (ce qui permet de contrôler le précurseur au niveau nucléaire). S'agissant du cytoplasme, d'autres mécanismes ont été décrits, notamment l'expression de Dicer dans les cellules cancéreuses. Pour Pascal Barbry, d'autres mécanismes pourraient s'avérer intéressants à explorer, par exemple Dead End 1 pourrait interagir spécifiquement avec certains micro-ARN et les stabiliser ou les déstabiliser. De même, TRIM32 interagirait avec certains micro-ARN.

De ce fait, on aurait pu imaginer que les mécanismes qui contrôlaient les niveaux d'expression des micro-ARN au final n'avaient rien à avoir avec la transcription. Or, Pascal Barbry a collecté des données qui suggéreraient plutôt le contraire. Il a réalisé des expériences de transcriptome à l'aide d'une puce couvrant toutes les zones de micro-ARN, afin d'analyser les régions de ces micro-ARN après hybridation d'ARN nucléaire et d'ARN cytoplasmique. Dans ces conditions, il s'est révélé possible d'identifier le pool d'ARN cytoplasmiques et le pool d'ARN nucléaires. En règle générale, les chercheurs s'étaient jusqu'alors contentés de procéder à une QPCR sur de l'ARN total, ce qui n'est pas la situation la plus favorable pour regarder ce qui se passe au niveau des précurseurs.

Pascal Barbry salue les performances de cette puce, comme en témoignent les courbes de signal exonique et antronique pour un ARN cytoplasmique ou les courbes des signaux cytosoliques et nucléaires (cf « *Nuclear signal for primitiv -181a-* », diapositive 12). Dans les jurkats, le signal nucléaire est particulièrement prononcé et couvre une zone étendue (100 kb). Il est possible de superposer d'autres marquages et de s'assurer de la spécificité en regardant le brin « plus » et le brin « moins ». Pascal Barbry souligne l'excellente corrélation entre les niveaux de signal dans différents types cellulaires et les niveaux obtenus. Or, il a obtenu cette corrélation à de nombreuses reprises.

Pascal Barbry en conclut que le niveau de micro-ARN est d'abord déterminé par le niveau du précurseur. De ce fait, l'importance des régulations successives par différents mécanismes est plus faible qu'on aurait pu le penser.

## VI. Quels sont les apports du séquençage à l'étude des micro-ARN ?

Pascal Barbry présente les travaux réalisés par Chi et ses collaborateurs en 2009 en utilisant une technologie mise en place par le laboratoire de Parnell, qui sert à identifier des interactions spécifiques « protéines/ARN ». Ils ont réalisé une immuno-précipitation du complexe « risques » avec des anticorps dirigés contre argonaute 2 et obtenu un complexe dans lequel argonaute 2 et les autres protéines sont présentes, ainsi que les ARN messagers reconnus par le complexe et les micro-ARN. L'idée proposée par Darnell consiste à dégrader tout ce qui dépasse à coup de RNase et à séquencer ensuite avec un séquenceur à haut débit les séquences protégées.

Cette expérience fait revisiter la notion d'interaction entre le micro-ARN et le 3' non codant, car ce dernier ne représente que 40 % de toutes les interactions connues. Une grande partie se situe dans le CDS et très peu au niveau du 5' UTR. Au niveau des ARN classiques, il y a un tiers de CDS et un tiers de 3'. Le fait que le site de reconnaissance du micro-ARN ne se trouve

pas dans le 3' ne modifie pas ces données. L'émergence du modèle visant à exclure toute interaction dans le CDS est fondé sur le fait que :

- Lin-4 et Lin-14 interagissent *via* le 3' non codant de Lin-14 ;
- les bio informaticiens qui ont élaboré les programmes ont utilisé la conservation inter espèces qui faiblit fortement dans le 3' non codant, mais qui est biaisée par la conservation d'une séquence à partir du moment où l'on s'intéresse à une séquence codante.

## VII. Le protocole de profilage des micro-ARN

Pour profiler les petits ARN, Pascal Barbry utilise la machine SOLiD 3,5, avec le kit correspondant. Il a tenté de mettre en place le multiplexe assez rapidement afin d'obtenir des réductions de coûts sensibles.

Dans le cadre de manipulations normales, entre 5 et 20 % des résultats correspondent à des micro-ARN, sachant que Pascal Barbry et son équipe n'ont pas encore dressé de bilan exhaustif des séquences qui ne s'alignent pas avec le génome de référence. Ceci nécessite de délaissier une grande part des informations obtenues, sachant que les chercheurs en génomique ne sont pas les seuls à pouvoir se voir reprocher de ne pas valoriser suffisamment les informations issues de leurs expériences !

Quoi qu'il en soit, Pascal Barbry souligne la nécessité de structurer les données de la façon la plus rigoureuse possible. De ce point de vue, il salue l'effort consenti par les plateaux de transcriptomique à travers des systèmes comme Géo ou Eurexpress, qui permettent de rechercher des résultats d'expériences disponibles à travers le monde entier. Ces dispositifs consistent souvent un moyen assez efficace pour valoriser ses propres expériences. Du point de vue de notre intervenant, l'effort qui avait été consenti avec les puces peut être transféré assez facilement vers le séquençage. Ainsi, il sera possible de ré-analyser les données *a posteriori*, avec de nouveaux algorithmes.

Pascal Barbry est en train de travailler sur xénopuce levis. Le génome correspondant n'est pas disponible. Pour autant, il ne ressent aucune difficulté à réaliser d'ores et déjà des expériences et à vérifier la correspondance des deux ou trois séquences spécifiques qui l'intéressent. Lorsque ce génome sera enfin disponible, il aura la possibilité de croiser correctement ces séquences et d'approfondir son analyse. Ceci était totalement impossible lorsqu'il travaillait avec des puces, car cette méthode nécessitait de disposer d'une collection de référence, dont la qualité avait d'ailleurs un impact prédominant sur la qualité du résultat final.

## VIII. La sensibilité de la méthode utilisée

Le pipeline d'analyses utilisé par Pascal Barbry est essentiellement basé sur le programme RNA2MAP. Celui-ci procède à un certain nombre d'analyses sur différents génomes et listes de séquences. Il a obtenu des résultats de plusieurs ordres : le set de référence mirVana qui propose un mélange équimolaire d'ARN synthétiques qui correspondrait à tous les micro-ARN chez l'homme, la souris et le rat. Pascal Barbry et son équipe ont souhaité vérifier si les niveaux de détection étaient comparables pour tous ces micro-ARN. Il faudrait donc trouver le même nombre de copies pour tous les micro-ARN. Or, le nombre de read varie de 0 à 60 000 en fonction des séquences. Un certain nombre de séquences vont être correctement détectées, d'autres ne vont pas l'être du tout, et d'autres seront trop détectées, et ceci quelque soit l'adaptateur.

Ceci suscite tout de même une restriction. Compter le nombre de clones dans une banque, c'est évaluer l'abondance de ce micro-ARN dans ladite banque. Le recensement des séquences permet d'obtenir un résultat plus précis que le recensement de la fluorescence. Quelle que soit la méthode utilisée, toute surinterprétation sera sans doute tendancieuse. Néanmoins, ceci permet de constater que dans tous les échantillons, avec une trentaine de séquences de micro-ARN, on peut couvrir 70 à 80 % des séquences présentes dans cette banque. Cet élément est d'autant plus important puisque les micro-ARN ont un effet inhibiteur et que la titration des cibles s'effectuera d'autant mieux avec de grandes quantités. Pascal Barbry en conclut que seuls les micro-ARN fortement exprimés sont significatifs dans ce modèle expérimental.

Pascal Barbry signale qu'il est dans l'impossibilité d'identifier de nombreuses séquences, dont certains peuvent correspondre à des micro-ARN. On trouve notamment des séquences assez mystérieuses, comme des piARN (en proportion beaucoup plus faible que les micro-ARN), ainsi qu'un certain nombre de snoRNA (ce qui peut s'expliquer par le fait qu'un certain nombre de petits ARN peuvent être dérivés à partir de snoRNA).

## IX. Quels résultats ?

L'un des intérêts de cette approche réside dans la possibilité de mapper plus finement ce qui peut se passer pour un micro-ARN et de constater l'existence d'erreurs dans les bases de données. MiR-21, par exemple, joue un rôle extrêmement important dans le cancer. Or, grâce à cette approche, l'équipe de Pascal Barbry a constaté l'existence, dans les modèles utilisés, d'un décalage d'une base entre la séquence détectée par rapport au panel mirVana. Le séquençage permet d'effectuer la correction qui s'impose. Ce décalage se constate également pour miR-92. Cette analyse à la base près permet de dégager des informations intéressantes.

L'équipe de Pascal Barbry a également regardé un profil de micro-ARN au cours de la différenciation de kératinocyte, sur trois lignées différentes qui n'avaient rien à voir. Dans un certain nombre de cas, on parvient à reproduire sur les lignées immortalisées le profil qui était obtenu avec des cultures primaires provenant de différentes origines. Par exemple, miR-203 est un micro-ARN très important dans la différenciation des kératinocytes.

Mir-29a va être détectée de façon importante, à hauteur de 12 à 18 % sur un nombre total d'un million de selon les conditions de hit. La capacité de distinguer correctement les différents formes miR-29a, b et c, qui ne diffèrent que par une ou deux bases, est intéressante.

La comparaison avec les autres technologies par exemple Agilent est assez bonne lorsqu'on compare des logs ratios. Ces tableaux de données sont traités exactement comme nous le faisons avec les data puces, si ce n'est qu'il y a une information supplémentaire susceptible d'être utilisée.

En guise de conclusion, Pascal Barbry présente un résultat dont il a envisagé en premier lieu qu'il pouvait correspondre à une détection du catabolisme des micro-ARN : une cellule dans une situation a, lorsqu'elle passe en b, subit une modification extrêmement importante des profils de micro-ARN. Son équipe a donc essayé de mettre en évidence les mécanismes présidant à la diminution de la quantité de micro-ARN. En reprenant toutes les séquences de ces micro-ARN, elle a étudié la distribution de ces séquences et s'est rendu compte que les situations étaient extrêmement différentes selon qu'on considère la séquence 3' ou 5'. Chez 5', la distribution sur un très grand nombre de micro-ARN est relativement claire. Chez 3', la distribution est beaucoup plus étalée. Pascal Barbry attribue ce phénomène à un problème de fidélité de la coupure au moment du processing, qui est très bonne du côté 5' et moins bonne du côté 3'. Ainsi, les coupures de Drosha et de Dicer donnent-elles des effets inverses. Il est probable que l'activité enzymatique de chacun des domaines ne soit pas affectée, et que ce

phénomène résulte plutôt de la cinétique de l'opération. Pierre Barbry suppose que la coupure pourrait se faire d'abord du côté 5', puis du côté 3' et générer cette petite incertitude.

En conclusion, Pascal Barbry salue Rainer Waldmann et Laure-Emmanuelle Zaragozzi pour le développement sur la technologie du SOLiD, Kevin Le Brigand, qui prend en charge la bioinformatique, Julien Maurizio, qui entame un travail sur la statistique, l'ensemble du personnel de la plate-forme (Géraldine Rios, Virgine Magnone et Sandra Fourré), ainsi que Karine Robbe-Sermesant et Bruno Carinaud. qui ont travaillé sur le projet miR-34.

## PRESENTATION DE LA SOCIETE AGILENT

---

**Juliette GREGOIRE**

*Spécialiste « produits génériques »*

Juliette Grégoire se propose de présenter une solution d'enrichissement de cible en amont du séquençage haut débit, SureSelect.

### **X. SureSelect, une réponse aux défis suscités par la nouvelle génération de séquenceurs à haut débit**

Les séquenceurs à haut débit génèrent, lors du lancement d'un run de séquençage, une masse de données considérable. Or, Agilent propose une solution d'enrichissement qui permet de se focaliser sur la séquence d'intérêt et de réduire ainsi à la fois le coût de la séquence et le volume de données à analyser.

SureSelect se décline en deux versions :

- un enrichissement en phase liquide (sur la base d'une quantité de matériel de 1 à 3 microgrammes) ;
- une solution de capture sur puces (sur la base de 20 microgrammes d'ADN génomique).

Ces deux solutions sont complémentaires et sont fonction de la quantité d'ADN génomique disponible, de la taille de la cible et du débit correspondant au projet (à l'heure actuelle, on n'est pas capable de synthétiser des librairies pour la phase liquide pour moins de 10 échantillons). Ces solutions peuvent être utilisées avec SOLiD et Genom analyzer. Agilent a conclu un accord de co-marketing avec les sociétés en question, ce qui permet de travailler en partenariat et de développer des protocoles robustes.

### **XI. Présentation des deux solutions**

#### **1. Le protocole d'enrichissement en phase liquide**

L'utilisation du kit SureSelect intervient après la préparation de la librairie d'ADN générique selon le protocole préconisé par le constructeur du séquenceur. Un délai de 24 heures de mise en présence de la librairie d'ADN avec la librairie d'ARN est nécessaire avant de créer des hybrides, capturer les séquences d'intérêt, récupérer l'ADN visé, amplifier le matériel ainsi obtenu et lancer le run de séquençage.

Les principales caractéristiques de cette solution en phase liquide sont :

- une hybridation ADN/ARN, avec une interactivité plus forte qu'une hybridation ADN/ADN ;
- une efficacité maximale au niveau de la capture ;
- une longueur d'oligo de 120 mères ;
- la petite quantité de matériel qu'elle permet d'utiliser (3 microgrammes).

#### **2. Le protocole en phase solide**

La librairie préparée, fractionnée, avec les séquences adaptatrices, est déposée sur le microarray (244k et oligos de 60 bases) pour hybridation, lavage et déshybridation. L'ADN ainsi récupéré passe par une étape d'amplification avant le lancement du run de séquençage.

Cette solution est particulièrement adaptée aux petits projets. Dans le cadre de la phase liquide, on fournit un kit qui contient la totalité des réactifs, alors qu'en phase solide, Agilent ne fournit que les lames de microarray. Ceci rend personnalisable le protocole de capture en phase solide.

### 3. Les différentes étapes de la commande du kit d'enrichissement

Pour commander le kit d'enrichissement, il est nécessaire de procéder au design des régions cibles *via* l'interface web eArray (sachant qu'un kit catalogue correspond à l'exon humain), ce qui permet de procéder à la synthèse de la librairie sur puce (dans le cas de la phase liquide, les oligos sont rétro-transcrits). Le kit est conditionné avec les tampons et les autres réactifs nécessaires.

eArray est un serveur gratuit et libre d'accès. Il suffit de s'y inscrire. Deux nouveaux espaces applicatifs sont dédiés à la capture sur puces et à l'enrichissement en phase liquide. Ce site permet également aux équipes de partager leurs travaux respectifs.

## XII. Paramètres techniques

### 1. Adaptation du produit Agilent au design et au paramétrage de l'échantillon

Suivant le protocole de séquençage utilisé, Agilent va préconiser un design différent pour la librairie de capture. Dans le cadre du single read, Agilent recommande de couvrir chaque zone cible par deux oligos de capture au minimum (tagging 2x). Dans le cadre du paired-end, on obtient un taux de couverture satisfaisant sans que les oligos se chevauchent (end to end).

En phase solide, les paramètres ouverts dans l'interface eArray sont les suivants : 244K, oligo de 60 paires de base et un *tilling* de trois paires de base. Ces paramètres, très stricts, vont être conduits à évoluer rapidement en vue de plus de flexibilité dans le choix du tilling. Agilent propose des protocoles en paired-end et en single-end pour Illumina, en fragment, pour Applied. Le kit Full Exon recommandé en paired-end sur Illumina et en fragment sur Applied.

Dans le cadre du séquençage haut-débit, un outil se révèle particulièrement important et est recommandé par les constructeurs de séquenceurs et de bio analyseurs : le kit ADN haute sensibilité, qui permet de détecter des concentrations d'ADN jusqu'à 6mg/microlitre pour 50 à 7 000 paires de bases.

### 2. Les performances du kit SureSelect

Différents enrichissements ont été réalisés sur des cibles de différentes tailles de 3,1 à 0,2 mégabases. Le taux d'enrichissement peut atteindre jusqu'à 80 %, avec une variabilité en fonction des régions du génome.

Des erreurs peuvent être introduites par les séquenceurs de nouvelle génération. La préparation d'échantillons ne doit donc pas introduire de nouveaux biais. Sur un échantillon qui a été enrichi en phase liquide, 90 % des bases sont au moins à une profondeur de 20X.

Cette profondeur permet d'avoir la quantité donnée qui, statistiquement, permettra d'assurer une bonne homogénéité.

La reproductibilité du kit SureSelect est excellente, à hauteur de 95 %.

### **Présentation du kit All Exon**

Le kit All Exon permet de cibler une région de 38Mb (180 000 exons issus de la base de données CCDS) dont 1000 ncRNA (région non codantes). Ce kit nécessite 3 microgrammes d'ADN génomique. Ce kit présente un pourcentage de read on target de l'ordre de 70 % et une très bonne reproductibilité.

Juliette Grégoire conclut que le kit est disponible pour Illumina et pour Applied, tout en précisant que les protocoles sont également adaptables sur Roche. Agilent propose une solution « customs », (jusqu'à 3,3 Mb) pour un projet à façon, ainsi qu'une solution de capture sur puces pour des projets spécifiques. Agilent travaille sur un protocole d'indexage pour permettre le multiplexage, l'élargissement de son offre de kits « catalogue » et la création d'un support pour Roche.

## PRESENTATION DE LA SOCIETE AFFYMETRIX

---

**Yann LEBEAU**

*Responsable commercial Lille*

Yann Lebeau présente Affymetrix, qui fournit des puces à ADN permettant d'analyser les mutations ou l'expression des gènes. Cette société a développé un programme intitulé Integrated Genomics, qui vise à intégrer les données au niveau de l'ADN et du transcriptome pour faciliter l'identification de cibles potentielles à visées diagnostiques.

### **XIII. L'offre d'Affymetrix en matière d'expression des gènes**

Les microarrays permettent d'utiliser un très grand nombre d'échantillons sur un très grand nombre de gènes. Affymetrix a développé plusieurs plates-formes, GCS 3000, Titan et GeneAtlas. Le coût des puces et plates formes diminue, ce qui permet de faciliter l'accès à cette technologie au maximum de laboratoires.

#### **1. Les puces 3'**

Historiquement, toutes les puces développées pour l'étude de l'expression des gènes utilisent une chimie qui cible la partie 3' du transcrit. Dorénavant, la technologie WT (Whole Transcript) permet de cibler la totalité du gène ou la totalité des exons.

La partie 3' ayant fait l'objet de plus de 12 000 publications, la puce 833 Affymetrix est dorénavant validée par la FDA pour faire du diagnostic clinique en oncologie. Même si cette puce est dorénavant un peu ancienne, elle très bien maîtrisée et fournit des données très robustes (cf projet MAQC).

#### **2. La technologie WT**

Les puces à expression WT, les gene arrays et les exon arrays, ont pour particularité de cibler la totalité du transcrit, avec une couverture plus ou moins importante des exons. Au total, 25 000 gènes ont été identifiés et 500 000 protéines.

Les premières puces utilisées étaient principalement focalisées sur la partie 3' et rataient de ce fait un certain nombre d'évènements. Le gène Bcl-x, en fonction de son splicing, inhibe ou active l'apoptose. Devant un cas de cancer, il est intéressant d'identifier le splicing différentiel et l'expression des exons

En gene array, les sondes de l'ARN messenger sont réparties sur la totalité du transcrit et les exons ne sont pas forcément ciblés dans leur totalité. En exon array, on cherche à identifier de manière précise tous les transcrits présents dans le génome (la puce permet d'avoir en moyenne quatre sondes pour chaque transcrit).

La puce exon possède plus de 6 millions de sondes qui ciblent plus d'un million d'exons (exons qualifiés et exons putatifs), ce qui représente de nombreuses données disponibles. Quant à la puce gene array, elle se focalise sur 29 000 gènes identifiés avec 26 sondes par gène réparties

sur la totalité du transcrit et une couverture de l'ordre de 95 % des exons annotés. Affymetrix recommande donc les puces exon dans le cadre des études les plus poussées.

L'analyse des puces exon peut effrayer certains laboratoires, c'est pourquoi Affymetrix a développé des partenariats avec des sociétés fournissant des logiciels « clé en main » et des prestataires de services.

#### **XIV. Les solutions ADN**

L'analyse exonique peut être couplée avec des analyses au niveau de l'ADN. En 2008, Kwann et ses collaborateurs se sont aperçus qu'une mutation de l'ADN ne générât une variation au niveau de l'expression du transcrit que dans 39 % des cas (au niveau de l'initialisation, de la terminaison ou du splicing). Dans 55 % des cas, la modification du transcrit ne modifiait pas forcément la valeur d'expression du transcrit. Les modifications observées dans cet article étaient soit des splicing différentiels, soit au niveau du terminal, soit une dégradation du transcrit, avec la partie 3' largement plus représentée que la partie 5'.

##### **1. Présentation de la puce SNP6**

La puce SNP6 est la plus utilisée, et combine des marqueurs SNP et SNV. Dans le cadre de l'étude du nombre de copies d'ADN, on trouve une corrélation entre l'intensité lumineuse du SNP et le nombre de copies du génome de l'individu. Ces puces peuvent donc être utilisées dans le cadre d'études du remaniement chromosomique. Il est donc possible de détecter des événements de type « perte d'hétérozygotie », même si le nombre de copies reste le même. Les puces développées par Affymetrix combinent des marqueurs SNP et des marqueurs CNV choisis pour leur espacement sur le génome. La puce SNP6 possède plus de 1,8 millions de marqueurs, soit une très grande densité, qui permet une très grande résolution. Cette puce est beaucoup utilisée par les hématologues afin d'étudier le remaniement des cellules cancéreuses.

Une application Notes a été développée au sein d'Affymetrix pour montrer l'intérêt des LOH dans le cadre du cancer. En effet, les informations sur la variation du nombre de copies peuvent être d'autant plus intéressantes combinées à des informations sur la perte d'hétérozygotie. Plusieurs publications sont parues, qui allaient dans ce sens. Par exemple, George et al ont réalisé une étude du 22 échantillons dans le cadre d'un neuroblastome. Il s'est aperçu de pertes d'hétérozygotie assez importantes au niveau du chromosome 11. Pour 15 tumeurs, cette perte intervient sur le bras Q du chromosome. En outre, 14 de ces tumeurs présentent une perte d'hétérozygotie « copy neutral ». Le chromosome 11 d'une de ces tumeurs est en LOH. Ce phénomène est assez intéressant à investiguer dans le cadre cette étude. Une étude similaire a été publiée par Ross et Al en 2007 sur un autre cas de cancer et a mis en évidence une perte d'hétérozygotie, avec un événement « copy neutral ».

Yann Lebeau signale que la SNP6 peut également être utilisée sur des échantillons de type FFPE en adaptant le protocole.

##### **2. Les évolutions récentes**

La puce 2.7M dédiée à la cytogénétique a été développée comme une solution intégrée afin qu'elle soit utilisée à terme en diagnostic. Elle est validée pour des échantillons cytogénétique classiques pré ou post natal. Les onco-hématologues qui l'utilisent également obtiennent des résultats assez prometteurs.

Outre le fait que cette puce possède 2,7 millions de marqueurs, son protocole est très simple à mettre en œuvre par le laboratoire. Elle est assortie de son propre logiciel, mais qui n'est pas encore adapté dans le cadre du cancer. Ce développement doit intervenir dans le courant de l'année.

### 3. Des résultats

En intégrant les données provenant des puces exons et SNP6, quelques publications ont déjà permis d'extrapoler des gènes potentiellement intéressants dans le cadre de certaines maladies.

En 2007, l'équipe Mao & al a été capable, en combinant des études de microarrays et de FISH, d'identifier des points de cassure et notamment des gènes potentiellement intéressants. Sur les 96 gènes identifiés au niveau des points de cassure, 29 étaient sous-exprimées, 16 surexprimés et 28 avaient une expression différentielle au niveau des exons.

En 2007, dans le cadre d'un neuroblastome, l'étude combinée par microarrays de l'ADN des remaniements chromosomiques a permis à l'équipe de La Tosca de cibler uniquement les gènes susceptibles d'être influencés par les points de remaniement et d'identifier 84 gènes d'intérêt. Ils se sont focalisés dessus pour continuer leur étude.

Yann Lebeau conclut que les outils proposés par Affymetrix permettent de combiner les études au niveau de l'ARN et de l'ADN, en utilisant les deux types de puces.

## WACA ET ASSOCIES

---

***Frédéric LEPRETRE***  
*Bioinformatique, IRCL, Lille*

En guise d'introduction, Frédéric Leprêtre rappelle que la cytogénétique classique a rapidement dérivé vers la FISH, avec le développement de la fluorescence, la multiFISH, puis la CGH array, grâce à l'informatique.

Malgré la précision croissante de ces outils, l'examen du profil ne permet toujours pas de distinguer une aberration d'un bruit. Or, sur différents pathologies et profils, Frédéric Leprêtre et son équipe ont observé un profil « vagué » sur tout le génome. En examinant la bibliographie, ils ont vérifié que ce phénomène se répétait partout ailleurs :

- au Japon, avec Komura (avec SNP d'Affymetrix) ;
- en Angleterre, avec Marioni (avec des puces Affymetrix) ;
- aux Pays-Bas, chez Cardoso (avec des puces BAC d'Agilent) ;
- en Angleterre, chez Hurles (avec des puces SNP Affymetrix) ;
- en Espagne, chez Rueda et Diaz (avec des puces BAC d'Agilent).

A chaque fois, le même profil « vagué » est observé sur les chromosomes. Frédéric Leprêtre et son équipe se sont donc demandés s'il s'agissait d'un biais de fragmentation, de marquage, d'hybridation, de lecture ou biologique. Après avoir opté pour l'hypothèse biologique, restait à savoir s'il s'agissait d'une différence de pureté entre deux échantillons, d'un problème de dégradation, de protéine-histones, d'une modification épigénétique ou de composition GC.

### **XV. La création de WACA**

Or, les publications de Cardoso en 2004 et de Marioni en 2007 montraient une forte implication du GC sur les vagues. En reprenant les données de l'UCSC, Frédéric Leprêtre a observé une bonne concordance entre un profil et les compositions même du GC sur le génome. En partant de cet exemple, son équipe s'est focalisée sur le GC et a constaté le bon alignement entre le pourcentage du GC sur les génomes et les profils qui se dégageaient. Ce phénomène pouvait également correspondre chez Hurles.

Il aurait été possible de s'arrêter au GC des sondes, mais l'équipe de Frédéric Leprêtre a préféré affiner son étude jusqu'à un million de paires de bases. Elle a observé une modification du LogRatio par le GC sur une fenêtre autour de la sonde davantage que sur le GC de la sonde. Sur un certain nombre d'individus, différents paramètres ont été testés avec une modification de l'ordre d'apparition de ces derniers. L'algorithme ainsi obtenu permet de corriger cette modification du LogRatio par le GC en prenant en compte une fenêtre de 500 kilobases et de 150 kilobases, le GC des sondes, le GC des fragments ainsi que la taille de ces fragments. Cet algorithme, baptisé WACA (Waves aCGH Correction Algorithm) permet pour les échantillons A et B d'obtenir un profil un peu plus correct, ce qui permet de mieux voir apparaître toutes les aberrations.

## XVI. Tests sur WACA

Lors des différentes publications, l'équipe de Frédéric Leprêtre s'est vu reprocher le manque d'introduction d'aberrations artificielles lors des tests. Elles ont donc été introduites sur les chromosomes 1, 2, 3 et 4 sur deux échantillons, avec ou sans utilisation de WACA.

On constate que :

- Sur le profil brut, les grandes aberrations sont conservées.
- Sur le profil « aberration sans WACA », les petites aberrations sont également conservées.
- Sur une aberration de 9 sondes introduite artificiellement, on constate soit une restriction de de l'aberration aux 9 sondes modifiées par le logiciel CBS sur le premier échantillon, soit que l'aberration était illisible par le logiciel sans l'utilisation de WACA (celui-ci permet donc de restaurer un profil artificiellement aberrant sur le second échantillon).  
Cet algorithme a été utilisé sur différentes pathologies (glioblastomes, mélanomes, LDGC, LLC et TALL). L'utilisation de WACA a permis de mieux former un dendrogramme par clustering.

Par ailleurs, l'équipe de Frédéric Leprêtre a testé les performances de WACA avec des statistiques sur une LLC (44k Agilent, ADN très dégradé, de qualité très faible car lame très bruitée) pour laquelle nous avons le karyotype (dup12, del13q14). Avant l'utilisation de WACA, il était impossible de déterminer une duplication du chromosome 12, ni la délétion en 13q14. Par contre après l'utilisation de WACA, on rétablit la cytogénétique avec une bonne visualisation des aberrations.

Les performances de WACA ont été comparées avec des algorithmes déjà utilisés (CNAG, MA2C et NoWaves) : WACA affiche le meilleur pourcentage de correction du LogRatio sur le GC. L'utilisation de NoWaves et de WACA est assez proche, mais on note une amélioration des standards de qualité par WACA.

Frédéric Leprêtre montre un exemple dans lequel, malgré l'utilisation de WACA, on observait des sondes qui n'avaient pas un comportement normal par rapport aux autres. Elles ont été désignées en « flag ». En éliminant ce bruit récurrent, on obtient un profil moins vagué et moins bruité, et une élimination de CNV faussement détectés par les algorithmes.

Par ailleurs, Frédéric Leprêtre a été récemment été interpellé par un article paru dans Nature Methods qui démontrait des performances d'amplification d'une Taq polymérase sur l'ADN. Les auteurs avaient repris les données HapMap, les avaient passés en Affimetrix 500k et en Illumina 650k et observé les effets de ces amplifications sur le génome. Frédéric Leprêtre a observé une corrélation entre les vagues décrites par l'article et le GC que son équipe avait décrit. Cette corrélation est d'autant plus visible si l'on observe le chromosome 16.

Frédéric Leprêtre conclut son intervention en soulignant la grande influence exercée par le GC sur la CGH et les techniques SNP, et peut-être également sur le chip on chip et l'expression.

## BIOINFORMATIQUE ET GENOMIQUE

---

***Philippe DESSEN***

*Responsable bioinformatique à l'IGR, Paris*

### **XVII. Présentation de plate-forme de génomique de l'Institut de Cancérologie Gustave Roussy**

Philippe Dessen travaille en collaboration avec Vladimir Lazar au sein de la plate-forme de Villejuif, qui s'inscrit dans une enceinte hospitalière. Cette situation géographique conduit à un certain nombre de développements associés soit à la clinique, soit à la recherche fondamentale. Cette plate-forme a été créée en 2002 et s'est développée essentiellement sur la base de la technologie Agilent.

La plate-forme génomique revêt un caractère expérimental. L'équipe, qui utilise un certain nombre de technologies classiques de génomique, est constituée de six personnes chargées de l'expérimentation et de cinq personnes chargées de l'analyse. Si l'activité de l'équipe est centrée sur l'Institut Gustave Roussy, elle nourrit également de nombreux projets sur la région Ile-de-France et a noué des partenariats européens ou bilatéraux avec d'autres institutions (échantillons ou analyses).

Philippe Dessen rappelle que la génomique comporte une partie structurale et une partie fonctionnelle (expression à travers des puces d'expression). La partie miRNA s'est également développée de manière notable, à travers la technologie Agilent.

Dès leurs débuts, les membres de la plate-forme ont souhaité éviter de faire ce qui se faisait sur d'autres plates-formes en s'orientant directement vers des technologies commerciales de qualité bien meilleure, avec des résultats intégrés dans une base de données, ce qui a permis d'obtenir des résultats probants rapidement. A ce jour, plus de 9 000 arrays ont été réalisés sur la plate forme, ce qui représente 400 projets. La certification ISO 9 001 de la plate-forme, obtenue à l'été 2009, permettra à celle-ci de nouer des collaborations industrielles.

Cette plate-forme évolue sur des protocoles assez simples : sur la base d'échantillons qualifiés, on procède à l'extraction d'ARN, l'amplification, l'hybridation et l'analyse, avec des contrôles de qualité à différents niveaux qui permettent d'évaluer différentes étapes et d'éviter de procéder à des hybridations onéreuses sur des échantillons dont la qualité est insuffisante.

### **XVIII. Les infrastructures bioinformatiques**

L'activité bioinformatique de la plate-forme s'est développée petit à petit en intégrant tous les logiciels usuels, qui sont régulièrement remis à jour et intégrés.

De nouveaux développements sont en cours autour du séquençage. La plate-forme travaille avec un consortium réunissant plusieurs clusters, notamment l'Université Paris VII, autour d'un protocole MobyLe, qui permet d'intégrer différents logiciels autour d'une interface commune.

Par ailleurs, des interfaces de type « blast » permettent de localiser directement toute séquence, sachant que l'équipe de Philippe Dessen a intégré toutes les listes de microarrays

directement sur blast. A partir de n'importe quelle séquence, on peut connaître directement les résultats dans différentes typologies d'arrêt.

Enfin, un système de gestion de projets, WORMS (WORKflow Management System), a été développé en interne.

Un certain nombre de données sont régulièrement mises à jour toutes les semaines. Au cours du temps, la qualité des puces a augmenté vers une haute densité. La plate-forme est dotée d'un *pipe line* d'analyse (Resolver, etc) qui permet d'analyser les données d'expression et de passer à une analyse biologique.

## XIX. Les améliorations méthodologiques

### 1. Développements sur CGH

Philippe Dessen explique que depuis trois ans, avec la haute densité, un certain nombre d'analyses ont été effectuées, pour maîtriser l'outil CGH. L'analyse classique se base sur les méthodes de normalisation et de segmentation disponibles dans les logiciels. La plate-forme a développé :

- un ensemble d'outils post-analyse permettant d'analyser les aberrations et de les relier à la localisation des gènes ;
- une méthodologie sur les aspects « normalisation » et les corrections, en particulier liées aux variations de GC.

Philippe Dessen présente le *pipe line* d'analyse :

- il comporte l'acquisition, la normalisation, la segmentation sur la base de profils multiples qui permettent de parvenir à un consensus sur les gains et les pertes ;
- il prévoit l'utilisation des packages comme RCGH qui permettent de faire de la clusterisation et de la représentation, et d'autres packages permettent de voir les régions minimum ;
- il intègre ces divers résultats ;
- il permet de procéder à une analyse fonctionnelle au cas par cas.

L'équipe de Philippe Dessen présente la particularité de procéder à une normalisation un peu différente pour la CGH (en particulier des tumeurs). Dans l'exemple présenté, la ligne de base n'est pas du tout positionnée là où elle devrait être. Or, il est possible de corriger cette situation avec une autre méthode de normalisation.

En utilisant la distribution des ratios sur les profils, on passe d'une courbe relativement simple si la fenêtre est très large à des courbes plus résolues si la fenêtre est étroite. La normalisation n'est donc pas du tout satisfaisante. Dans ce type de situation, l'équipe de Philippe Dessen a choisi de prendre comme point de référence le pic qui a la plus grande aire et sa médiane, ce qui permet de rectifier le profil vers davantage de normalisation. Ce résultat n'est pas flagrant sur de grandes anomalies. En revanche, avec un pourcentage de cellules tumorales relativement faible, le niveau de détail peut s'avérer critique.

Dans une étude en *dye swap* réalisée sur des 44k ou des 244k, on s'aperçoit de l'existence d'un certain nombre de probes erratiques selon les profils. La plate-forme de Villejuif a tenté de travailler pour savoir d'où provenait ce phénomène et de le supprimer. La sonde de 60 bases a été désignée par Agilent et est calculée pour présenter un Tm correct. Lorsqu'on réalise ce type de design, on oublie l'environnement du fragment : celui-ci a été coupé par des

enzymes de restriction et comporte donc une partie 5' ou 3' de quelques centaines de bases et des compositions qui ne sont pas le reflet de la sonde.

En regardant l'intensité et non les ratios des probes qui sont le long d'une puce, sur un full set 244k, on constate l'existence d'une intensité médiane. En prenant des probes aberrantes, présentant des ratios différents dans un sens ou l'autre, la distribution évolue, avec une très forte concentration de probes ayant une très faible intensité. Elle vient essentiellement du fait qu'il y a compétition entre l'hybridation de votre probe dans une forme ouverte (forte intensité) et une hybridation en structure secondaire (très faible intensité).

Philippe Dessen en conclut que l'origine des faibles intensités réside dans le fait que les fragments découpés présentent des structures secondaires différentes. L'élimination de ces probes et cette re-normalisation conduisent à améliorer sensiblement la corrélation entre les profils effectués en *dye swap* de différents échantillons.

## 2. Développements sur microarrays exon

En transcriptome, on utilise comme tout le monde une puce classique de 44K qui a un design relativement ancien et un nombre de probes relativement limité (sachant qu'environ 20 000 gènes sont représentés par une ou plusieurs probes). Une partie de ces sondes sont obsolètes.

Dans le cadre d'un projet européen sur le poumon, la plate-forme a développé un design spécifique pour une puce exon, qui, si elle n'en est qu'au marquage 3', permet d'allumer un ensemble de probes. Parallèlement, l'équipe a réalisé des études sur les exons, qui ont été publiées. Il s'agissait de l'analyse de deux cohortes avec des cancers du sein bénins et malins. Ces cohortes permettent de distinguer à 100 % (avec un millions de probes) les deux états de développement tumoraux. Cette signature a été brevetée.

En collaboration avec Didier Goidin (Agilent), l'équipe de Philippe Dessen a désigné sur une puce 244k un peu moins de 200 000 probes correspondant à 95 % de l'intégralité des exons du hg18 UCSC RefSeq. Alors qu'environ 305 000 exons RefSeq sont localisés sur le génome, elle est parvenue à en designer 295 000. La logique qui a présidé au design est très simple puisqu'elle consiste à prendre l'intégralité des 310 000 exons et à les clusteriser, non pas en séquence, mais par un biais de checksum. La mesure d'un checksum donne une valeur de 16 caractères qui sont tous indépendants. Un cluster peut partager plusieurs exons, soit du même transcrit, soit de plusieurs transcrits différents du même gène, soit de plusieurs gènes. Sur cette base, il est possible de dresser un tableau relativement précis : par exemple, pour TP53 qui a 7 transcrits différents et 15 types d'exon différents. Ce type de tableau permet de relier chaque exon « virtuel » avec sa position dans les épissages alternatifs. On peut positionner de la même façon les probes qui ont été désignés sur ces différentes séquences alternatives. On dispose ainsi d'informations relativement simples pour réaliser l'analyse de chaque transcrit de chaque gène.

## 3. Exemple d'intégration : projet de recherche sur une étude du mélanome

Philippe Dessen présente ce travail est basé essentiellement sur des lignées cellulaires, pour trouver celles qui sont les plus proches pour travailler sur le mélanome normal. Il porte sur 23 lignées de mélanomes humains à travers la CGH, le miRNA et le transcriptome.

D'après cette étude, la CGH n'apporte pas d'information sur les différences entre les trois groupes (sujet sain, métastases et tumeurs). En revanche, ces différences sont beaucoup plus marquées au niveau du transcriptome. On commence à avoir des choses associées à ces

phénotypes. Le miRNA n'est pas extrêmement pertinent, même si les groupes « tumeurs » et « métastases » ont davantage de points communs que le troisième groupe.

Philippe Dessen souligne la difficulté de corréler plusieurs technologies entre elles, notamment la CGH et l'expression miRNA. Si on tente de procéder à une corrélation le long du génome sur les trois milliards de bases entre l'expression sous forme de ratio (positif ou négatif) et le gain ou la perte d'aberrations CGH, on constate l'existence de corrélations positives et négatives. Dans certaines régions, on va trouver des gènes corrélés entre le gain de CGH et le gain d'expression.

Philippe Dessen présente une tentative de représentation, pour un gène donné, des corrélations entre l'expression et la CGH : l'expression se retrouve sous forme rectangulaire, tandis que le gain de gènes prendra une forme circulaire, avec un diamètre proportionnel. Ce mode de représentation reste partiel.

Sur le chromosome 19, un ensemble de clusters de mRNA existent, dont on peut utiliser les prédictions de target pour les croiser avec les résultats du transcriptome et repérer un certain nombre de candidats intéressants en termes de régulation mRNA/expression.

Philippe Dessen souligne l'intérêt de ce projet qui intègre la dimension biologique (phénotype), les différentes technologies (CGH, mRNA et transcriptomes) et la liaison avec les données disponibles sur les tumeurs humaines.

Philippe DESSEN salue ses collègues de la plate-forme, notamment Bastien Job, qui est à l'origine des analyses de CGH, qui devraient être publiées rapidement.

## NOUVEAUTES SUR LE SOLiD3

---

**Alain RICO**

*Ingénieur application NGS Europe-Life Technologies, Milan*

Alain Rico présente la version 3+ de SOLiD. Ce système permet d'étudier les systèmes biologiques dans leur intégrité (il ne s'agit pas simplement d'un séquenceur d'ADN ou d'un système d'analyse d'expression). Cet outil est d'autant plus important qu'un changement intervient au sein du milieu scientifique : l'internationalisation de la recherche, la difficulté d'obtenir des financements, l'intérêt socioculturel (les patients préférant être beaucoup mieux soignés) et les évolutions techniques.

Les nouvelles plates-formes de séquençage ont introduit une rupture technologique, ce qui a conduit beaucoup de chercheurs à reconsidérer les possibilités qui s'offraient à eux. Dorénavant, il est possible, par exemple, d'analyser le transcriptome d'une seule cellule. Dans ce contexte évolutif, il est nécessaire de choisir les bonnes plates-formes, susceptibles d'évoluer dans le temps et assorties d'un maximum de supports applicatifs et analytiques. Ainsi, dans le cadre de la collaboration avec Philippe Dessen, NGS Europe-Life Technologies devra assurer l'analyse du nombre de copies (ce qui prendra davantage de temps que la génération des données en elle-même).

SOLiD s'inscrit dans un processus de découverte de l'information jusqu'à son exploitation en médecine personnalisée. A la base, les séquenceurs à électrophorèse capillaire sont toujours utilisés (une gamme purement diagnostique a d'ailleurs été développée). Parallèlement, NGS Europe-Life Technologies part régulièrement à la découverte de terrains vierges, avec un séquenceur nouvelle génération comme SOLiD, par exemple. Il propose une gamme de réactifs dédiés en collaboration avec toutes les autres sociétés de la corporation.

Au cours des années, les données générées par SOLiD ont augmenté sensiblement, parallèlement à la baisse sensible du coût de revient des analyses. Si le seuil mythique de 1 000 dollars n'est pas encore atteint, le séquençage d'un génome humain coûte dorénavant une dizaine de milliers de dollars. La version 3 de SOLiD permet de multiplier différents échantillons au sein d'une même analyse, afin de réduire les coûts et le temps d'accès au résultat. Les nouvelles chimies sont plus performantes et moins chères. Le système d'encodage par couleur est maintenu, mais les données peuvent également être générées en bases, ce qui assure la compatibilité des résultats avec les logiciels d'analyses annexes. Alain Rico signale également l'existence de kits prêts à l'emploi.

Ces 60 millions de paires de bases sont trois fois plus nombreuses que les données générées en 1991 lors du premier séquençage d'un génome humain. Ce matériel permet de procéder au reséquençage d'une très petite région, au reséquençage dirigé (pour cibler une région en particulier), etc... En outre, SOLiD n'est pas seulement un séquenceur, en ce qu'il autorise des analyses d'expression. L'expression digitale permet le recensement du nombre de séquences correspondant à telle région, tel transcrit ou tel exon en vue de les comparer. Les domaines concernés sont donc diversifiées : transcriptome, profilage de miRNA, CGH, méthylation et méta génomique. D'ailleurs, Alain Rico se demande si ce dispositif ne représente pas le futur des microarrays.

Alain Rico argue de la simplicité d'utilisation de SOLiD. Mis à part une préparation spécifique selon l'application, NGS Europe-Life Technologies propose des kits pour la plupart des applications, avec des protocoles bien définis. La préparation d'échantillons sous forme

d'émulsions en PCR est très robuste et dorénavant automatisée. La chimie de séquençage, l'imagerie et l'analyse primaire font partie du même tronc commun. L'analyse des résultats sera spécifique.

En matière de préparation des échantillons, il est possible de faire de la bibliothèque de fragments, séquencer des deux côtés du fragment (paired-end), faire du multiplexage... Il n'est pas forcément besoin des 60 millions de paire de bases disponibles pour traiter un échantillon, ce qui permet de multiplier les applications sur le même échantillon. Il est possible de mélanger jusqu'à 104 échantillons différents dans un même run, ce qui équivaut à 60 millions de paires de bases par échantillon.

Pour les études plus complexes, il est fait appel à des bibliothèques un petit peu plus complexes, qui permettent de séquencer en deux temps deux parties d'une cinquantaine de paires de bases qui sont physiquement sur le génome d'origine à plusieurs milliers de paires de bases. Ceci représente un impact réel en termes de découverte de SNP.

Le système d'encodage sur deux bases est à la base du système : chaque base de l'échantillon est systématiquement séquencée deux fois, ce qui réduit d'autant les probabilités d'erreur. Jusqu'à 50 paires de base, on est au-dessus du fred 30. Ceci n'induit aucun compromis en termes de quantité ou de qualité des données.

La densité de l'information ne cesse d'augmenter, ce qui permet de générer encore davantage de paires de bases ou de lecture selon les applications dont les chercheurs ont besoins. Sur le plan quantitatif, SOLiD permet de traiter 60 milliards de paires de base de séquence ou jusqu'à 3 milliards d'étiquettes à répartir entre les différents échantillons.

SOLiD est très flexible avec le barre codage ou la séparation physique des échantillons sur les deux lames disponibles du système (de deux lames entières pour un seul échantillon jusqu'à des 8<sup>ème</sup> de lames, chacune de ces proportions comportant entre 16 et 20 échantillons différents). Les coûts sont inférieurs à ce qu'ils étaient il y a quelques mois : la chimie a été améliorée, les sondes sont incorporées, les amorces s'hybrident, le temps de run et la quantité de réactif nécessaire diminuent.

SOLiD génère des quantités de matériel très considérables, d'où la nécessité d'effectuer un important travail en matière d'analyse. Les logiciels d'analyse d'Applied Biosystems sont en *open source*. Ceux de certains des prestataires également. Les données sont compatibles avec les logiciels d'analyse ou de visualisation existants sur le marché. Pour autant, pour rendre le dispositif encore plus simple, l'Entreprise développe des systèmes permettant à l'opérateur d'effectuer l'analyse par lui-même (par exemple, « trouver SNPs »). Les visualisateurs tentent d'intégrer toutes les informations sur un même écran. En mettant à la disposition des chercheurs un système qui permet de générer énormément de données, ceux-ci peuvent s'approprier plus globalement leur thématique de recherche.

Alain Rico rappelle que SOLiD demeure néanmoins un séquenceur. Les données de certaines publications permettent de réduire la quantité d'informations nécessaire pour obtenir une qualité de lecture équivalente.

Des sociétés tierces comme Agilent proposent des solutions de capture et d'enrichissement : SureSelect (par exemple le Kit All Exome dédié à SOLiD). HybSelection a la particularité de pouvoir multiplexer les échantillons au sein d'une analyse, pour réduire les coûts de séquençage et de capture, avec une très grande reproductibilité. On peut également citer Olink.

La société Applied Biosystems propose donc des données très ombreuses, des systèmes prêts à l'emploi de génération de zones enrichies, l'automatisation de la préparation des échantillons et des systèmes d'analyse beaucoup plus pratiques à utiliser.

L'ARN constitue également un domaine d'utilisation essentiel. Grâce à SOLiD a eu lieu la première publication du transcriptome global d'un mammifère, d'une bactérie et d'une seule cellule. La société propose des kits prêts à l'emploi pour le transcriptome global, le SAGE et les petits ARN (identification, compréhension des interactions et découverte des implications dans le domaine de recherche). Pour un même échantillon, on peut avoir besoin d'une lame entière (par exemple, pour chercher une explication très pointue sur un transcriptome) ou de 10 millions de lecture, ce qui permet d'augmenter le débit des recherches.

Par rapport aux autres méthodes, ce système présente l'avantage de ne pas introduire d'*a priori* : il séquence tout l'ARN présent dans l'échantillon avant de voir à quoi il correspond. SOLiD a montré de grandes diversités en 3' des micro RNA. En outre, cette technologie permet d'analyser spécifiquement tel ou tel brin (alène) de l'ARN pur. Les bibliothèques ne sont pas construites après avoir capturé l'ARN avec des adaptateurs spécifiques. Si on compare les deux brins, on trouve des différences très sensibles d'expression d'informations. Selon les brins, la tumeur ne va pas avoir le même profil, c'est pourquoi il est important d'avoir l'intégralité de l'information. Alain Rico donne l'exemple de neuf gènes dont les alènes sont exprimées en rouge : en cas de tumeurs, on assiste à un changement sensible des alènes prioritaires, qui correspondent à d'autres alènes.

Ceci fournit des informations sur les nouveaux points de fusion : c'est d'autant plus facile qu'on n'est pas obligé de scinder la lecture de 50 paires de bases en deux, mais qu'on peut lire des deux côtés de l'échantillon. La plupart ont été validés : on a trouvé une corrélation très nette entre le nombre de TAG, ce qui prouve l'aspect quantitatif. SOLiD génère des données de qualité, mais fournit par défaut les scripts d'analyse pour la plupart des applications sur le SOLiD Software Community. En outre, il possède beaucoup de partenaires en fonction du domaine concerné.

SOLiD est également doté d'un système de contrôle interne, d'où la possibilité d'ajouter des contrôles ERCC, avec différentes gammes dynamiques. Ceci permet de calibrer les résultats grâce à ces contrôles et affiner l'analyse de l'expression. Nous disposons la lecture sur un même brin des deux côtés.

## LES TUMOROTHEQUES : ORGANISATION ET FONCTIONNEMENT EN FRANCE

---

**Henri SEVESTRE**

*Chef du Service d'anatomie et de cytologie pathologiques du CHU d'Amiens*

Henri Sevestre énonce le principe régissant les tumorothèques, à savoir conserver des fragments tumoraux.

### XX. L'évolution des pratiques en matière de conservation des tissus

#### 1. L'évolution des techniques de conservation

La conservation des tissus est une pratique très ancienne : Madame Necker, qui était connue pour sa beauté, avait souhaité être conservée dans un bain d'alcool pur après sa mort. De la même façon, le cœur de Louis XVII a longtemps été conservé (il a d'ailleurs été identifié par un chercheur lillois). Il a fallu atteindre les progrès de l'histologie et notamment de l'inclusion en paraffine, vers 1890, pour pouvoir conserver de manière efficace des collections d'échantillons tissulaires. Ces collections d'échantillons se prêtent très bien à l'étude morphologique, et, en fonction des étapes préalables à l'inclusion en paraffine, plus ou moins bien à l'extraction de protéine ou d'acide nucléique.

A partir des années 30, les plaisirs qu'on donnés la conservation par congélation des prélèvements ont permis d'effectuer des progrès dans l'extraction des éléments constitutifs des tissus et cellules. C'est au physicien écossais James Dewar qu'on doit l'invention du thermos qui a permis d'utiliser efficacement l'azote liquide (fabriqué il y a 130 ans par des chimistes polonais qui sont parvenus à comprimer de l'azote jusqu'à le liquéfier et obtenu une température stable de  $-196^{\circ}\text{C}$ ). La conservation à  $-196$  a l'air de donner des résultats très satisfaisants, c'est pourquoi elle est largement employée depuis une cinquantaine d'années, notamment en recherche biomédicale.

#### 2. Les évolutions de la législation

Ces prélèvements, qui proviennent de patients, ont un statut qui rend leur usage incertain pour la recherche scientifique et médicale. En effet, le Code Civil inscrit dans la loi des principes très sévères d'inviolabilité, d'incessibilité et d'indisponibilité du corps humain. Il ne serait donc pas possible de prélever des fragments du corps humain, sauf pour des raisons médicales ou thérapeutiques. Il est encore moins question de les vendre et personne ne peut en disposer, même l'individu lui-même.

Ces obstacles juridiques, qui présentent la particularité d'être punissables en cas de manquement, ont été rapidement contournés face à la mise en place du don du sang, du don d'organes, et à la nécessité de faire appel à ces produits pour les progrès de la recherche médicale. Des autorisations d'usage de produits du corps humain ont été mises au point. Ce droit d'utilisation s'exerce sous réserve du consentement de l'intéressé.

En France, on rencontre de grandes difficultés à trouver une formule simple à ce consentement. Ce point de la révision des lois de bioéthique soulève, au-delà des aspects

pratiques, des questions philosophiques. Lorsque la justice s'en mêle, la situation devrait très complexe. A un moment donné, on a pensé que le formulaire signé par un patient revêtait une valeur indéniable, à partir du moment où l'intéressé donnait un consentement éclairé, devant témoins. Cependant, la justice a donné raison aux intéressés qui se sont plaints.

### **3. Une situation qui demeure incertaine**

La médecine utilise dorénavant un simple document d'information attestant que le patient a reçu une information en présence d'une tierce personne, qui est conservé dans le dossier du patient. Henri Sevestre préférerait toutefois qu'il s'agisse d'une marque de non opposition. Ce dispositif régissait autrefois les autopsies, qui étaient réalisées dans l'intérêt supérieur de la Science. Les sujets entrant à l'hôpital étaient supposés ne pas être opposés à la pratique de l'autopsie après leur décès. Ce mode de fonctionnement a permis pendant très longtemps de pratiquer l'autopsie de femmes enceintes. Une autre solution consisterait, à chaque fois qu'un patient entre à l'hôpital, à ce qu'il ne accorde ou pas un droit d'utilisation universelle pour l'ensemble des produits de son corps.

En France, le problème est loin d'être résolu : on ignore si les prélèvements humains pratiqués par les médecins dans un but thérapeutique doivent être considérés comme des éléments du dossier médical ou comme un dépôt qu'accepte de faire un individu dans un hôpital. Une interprétation de la réglementation spécifique que l'individu reste propriétaire de l'organe qui lui est enlevé et qu'il le confie à l'hôpital.

Il en va différemment du dossier médical, qui est la propriété de l'hôpital. Sur ce point, les réglementations restent variées : dans le système libéral, les médecins doivent garder la preuve des éléments diagnostics pendant dix ans, alors que dans le système hospitalier, l'établissement doit les garder pendant vingt ans après le dernier passage du patient (voire 28 ans pour les enfants).

Les dispositions sont donc différentes selon que les prélèvements sont réalisés à visée sanitaire, de recherche ou de façon mixte. La réglementation est censée être différente si on envisage de procéder à des recherches génétiques permettant d'identifier un patient.

## **XXI. L'organisation des tumorothèques**

L'évolution de cette pratique de conservation des prélèvements a beaucoup changé. Au départ, elle était purement médicale, éventuellement médico-légale. Ensuite a émergé une réflexion sur l'utilité des prélèvements pour la recherche médicale. On s'oriente de plus en plus vers la mise à disposition d'échantillons pour la recherche scientifique. Cette évolution conceptuelle s'est accompagnée d'une évolution des produits conservés. Au départ, les organes ou les tissus entiers étaient conservés. Aujourd'hui, on conserve de manière différentielle, principalement des acides nucléiques isolés ou des protéines isolées.

### **1. L'émergence d'un réseau organisé**

Dans cette évolution à la fois conceptuelle et pratique, est apparue une espèce de jungle dans laquelle chacun agissait isolément. En France, le Ministre de la Santé créé un système de soutien aux tumorothèques existantes, en partenariat avec l'INSERM. L'INSERM, puis la direction de l'hospitalisation et de l'organisation des soins du Ministère de la Santé, ont lancé des appels à projets pour améliorer le fonctionnement des tumorothèques en fournissant essentiellement des moyens humains et en demandant aux hôpitaux publics de participer à l'acquisition de matériel pour mieux encadrer le recueil et la qualité de conservation de ces

collections à visée sanitaire et de recherche. En échange de ce soutien, le Ministère demandait la production d'un rapport régulier d'activité purement quantitatif.

Un deuxième appel à projets a été lancé en 2005, dirigé par l'INCa, pour former des réseaux régionaux autour des tumorothèques existantes afin de collecter à visée purement sanitaire des tumeurs particulières, soit rares, soit « intéressantes ».

## 2. Une situation différenciée

Ceci a abouti à un système très hétérogène qui reflète la taille, les capacités d'investissement et le poids politique des différents établissements hospitaliers en France. La carte des tumorothèques est à peu près homogène sur l'ensemble des régions, à l'exception de l'Ile-de-France et de PACA, plus richement dotées. Sous cette apparente unification, il existe de grandes variations dans les structures, le nombre de personnel employé, les capacités à fournir des produits dérivés, à stocker des échantillons et à systématiquement enrichir les collections par des annotations cliniques.

Certaines tumorothèques sont plutôt orientées vers la collection sanitaire que vers la recherche. Les collections ne sont pas organisées de la même façon. Si certaines tumorothèques se sont organisées de façon à répondre aux demandes de leurs interlocuteurs sans chercher à collectionner tout ce qui passait sur leur paillasse, d'autres, pour assurer un certain fonds de roulement, en l'absence d'interlocuteur, ont pris le parti de collectionner beaucoup. Il n'existe pas de règle en termes de démarche qualité. Au départ, les tumorothèques se sont engagées sur un certain nombre de points. Il n'y jamais eu de contrôle visant à évaluer la qualité des collections.

## 3. Etat des lieux

Au terme de dix ans d'existence des tumorothèques organisées, cette situation aboutit à une forme de désenchantement. Le meilleur encadrement des tumorothèques a certes évité les incidents (pannes de congélateurs). On s'interroge sur l'avenir des collections qui n'ont pas été organisées pour répondre à une demande s'inscrivant dans un protocole de recherche. Lorsque les collections sont organisées au fil de l'eau, les tumorothèques accumulent des masses de prélèvements et le rendement de ce type de collection est très faible. Même si les collections sont là, il y a peu de demandes : les tissus conservés n'intéressent pas la recherche dans la région géographique en question, ou parce que les étapes correspondantes de la recherche ont déjà été dépassées.

A Amiens, le taux d'utilisation des collections réalisées s'élève à 1 %, ce qui interroge les responsables de la tumorothèques à questionner la nécessité de collectionner « tout ce qui passe ». A l'échelle de la France, l'INCa, chargé de la collecte des informations, indique sur son site internet un taux d'utilisation des collections de 8 % en 2006, soit une multiplication par 8 depuis 2001. Il n'est donc pas impossible que la tumorothèque d'Amiens accuse simplement quelques années de retard en matière d'utilisation. Cependant, l'environnement scientifique français a plutôt tendance à se regrouper sous forme de grandes unités.

## **XXII. Comment remédier à ce désenchantement ?**

Henri Sevestre préconise de faire de la publicité pour sa tumorothèque ou d'intégrer des systèmes d'utilisation conjoints (ou de mise en ligne des ressources)

Les tumorothèques de la région PACA, organisées par le Professeur Hauffman à Nice et les équipes du centre Paoli Calmette à Marseille, ont pris le parti de promouvoir activement leurs installations. Ils ont mis en ligne leur catalogue, à la disposition des universités locales.

Le cancéropôle Nord-Ouest, qui a émergé en 2003, a lancé l'idée d'une tumorothèque virtuelle sur son territoire. Le projet est en train d'aboutir, à l'issue de cinq ans de pérégrinations administratives et techniques : il fonctionnera à l'aide d'un logiciel commun, HCCAP, sur un serveur hébergé à Lille. Ceci suppose en effet d'avoir un logiciel commun, des connexions qui fonctionnent bien ainsi que des formations à l'utilisation du logiciel (elles sont en cours). Le lancement de la tumorothèque virtuelle est prévu au 2<sup>ème</sup> trimestre 2010. Sur ce point, Henri Sevestre tient à rendre hommage à ceux qui ont appuyé ce dossier, notamment Messieurs Figeac et Burry.

Ces tumorothèques sont en train de changer de destination, car elles suscitent manifestement beaucoup d'intérêt. Elles avaient été lancées localement, par des médecins hospitaliers, puis à l'échelon national, par le Ministère, qui a souhaité clarifier leur fonctionnement et améliorer leur qualité. L'INCa, depuis 2005, remplit un rôle de gestionnaire pour vérifier que les tumorothèques fonctionnent, avec toute la lourdeur administrative d'une organisation centralisée (par exemple, on ne dispose que des données 2006). Enfin, les cancéropôles devraient stimuler l'utilisation de ces tumorothèques. De ce point de vue, la mise en commun des ressources des cinq tumorothèques du Nord-Ouest de la France devrait améliorer la connaissance de l'existence de ces installations par les partenaires médicaux et scientifiques.

## **INCIDENCE DES LEUCEMIES : EXPERIENCE DU REGISTRE REGIONAL DES HEMOPATHIES MALIGNES DE BASSE-NORMANDIE**

---

**Xavier TROUSSARD**

*Chef du service d'hématologie biologique du CHRU de Caen*

En guise d'introduction, Xavier Troussard rappelle quels doivent être les enjeux pragmatiques face au cancer en 2010 :

- réduire la mortalité ;
- faire reculer la fréquence des cancers ;
- améliorer la survie et la qualité de vie des patients ;
- favoriser l'équité face à la prévention des cancers, à l'accès au diagnostic et à l'accès aux traitements innovants et efficaces.

### **XXIII. La surveillance des cancers en Basse-Normandie**

Le système de surveillance des cancers est basé sur un réseau de registres. L'arrêt du 6 novembre 1995 définit le registre comme un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs éléments de santé dans une population géographiquement définie. De ce fait, il existe en France 21 registres qualifiés, couvrant 13 % de la population. La Basse Normandie bénéficie de l'existence d'un grand nombre de registres :

- deux registres généraux : le registre des cancers de la Manche et le registre des tumeurs du Calvados ;
- trois registres spécialisés : le registre des tumeurs digestives, le registre des mésothéliomes, le Registre Régional des Hémopathies Malignes de Basse Normandie, le RRHMBN (qui couvre les trois départements du Calvados, de la Manche et de l'Orne, soit 1,5 millions d'habitants)

La structure juridique du RRHMBN est basée au CHU de Caen. L'équipe comprend Xavier Troussard, directeur scientifique, Albert Colignon, médecin épidémiologiste, deux enquêteurs et plusieurs médecins bénévoles.

Le contexte du département de la Manche est particulier, de par la présence d'un certain nombre d'installations nucléaires incitant les scientifiques à étudier le rôle de la radioactivité sur l'incidence des hémopathies malignes. Notre expérience a débuté en 1997 et nous a permis, entre 1997 et 2004, de recenser 5 510 cas incidents d'hémopathies malignes, soit environ 800 cas par an. L'incidence est exprimée en taux standardisé à la population mondiale, qui lisse les caractéristiques démographiques de la population étudiée. Le RRHMBN est capable de donner les taux standardisés à la population mondiale de tous les différents types d'hémopathies malignes.

### **XXIV. Focus sur la Leucémie Lymphoïde Chronique (LLC)**

Les travaux d'Andy Rawstrom ont fait émerger le concept de lymphocytose B monoclonale (MBL), à savoir la présence chez des patients complètement asymptomatique, d'un clone de lymphocyte B ressemblant sur le plan phénotypique, à celui identifié dans la LLC. La MBL est identifiée chez 3,5 % de la population. Les individus avec une MBL présente les mêmes caractéristiques que les patients avec une LLC, ce qui laisse suggérer que la population ne corresponde pas à une expansion de cellules normales, mais plutôt à une expansion de cellules pathologiques probablement identiques à celles observées dans la LLC. Tout laisse à penser qu'en augmentant la sensibilité des techniques, il sera aussi possible de détecter la population anormale de façon beaucoup plus fréquente. La MBL existe quelle que soit l'origine géographique, en Chine, aux Etats-Unis, en Europe ou en France.

Ola Landgren a étudié aux Etats-Unis une cohorte de 150 000 sujets, dont les prélèvements ont été conservés de façon systématique. 129 patients ont développé une LLC, parmi lesquels 45 ont eu des cellules lymphoïdes sanguines cryo-préserverées avant la survenue de la LLC. Avant la survenue de la LLC, tous les patients présentaient une lymphocytose B monoclonale, laissant suggérer que la MBL est un état pré-leucémique. A cette occasion, Xavier Troussard insiste sur le besoin de développer les tumorothèques.

Les études biologiques réalisées par Dagklis *et al* montrent que le répertoire utilisé lorsqu'on a une MBL est un peu différent de celui observé lorsqu'un patient a une LLC, avec un certain nombre de sous ou sur représentations de certaines familles IGHV

Les études de registres permettent aussi d'étudier les facteurs de risque. Si la famille d'un individu comporte des antécédents de LLC, le risque de survenue de la maladie est multiplié par 8,5 par rapport à la population normale, sans compter une augmentation des risques de survenue de lymphome malin non hodgkinien.

La publication de Lindsay Morton dans *Blood* en 2007 montre que la LLC appartient aux lymphomes non hodgkiniens de type B. Dans la majorité des études, la LLC est combinée avec les patients avec un lymphome lymphocytaire. Ces deux entités représentent environ 17 % des lymphomes malin non hodgkiniens. Les études inter registres revêtent également leur importance. C'est sur cette base que Lindsay Morton a étudié le taux d'incidence de la LLC aux Etats-Unis (en taux standardisés par rapport à la population américaine). La LLC n'est pas très fréquente chez les indiens et est relativement moins fréquente chez les asiatiques. Au Japon, il n'existe quasiment aucun cas de LLC alors que les asiatiques vivant aux Etats-Unis peuvent très bien développer cette maladie, ce qui laisse suggérer qu'il existe des facteurs environnementaux, probablement alimentaires.

En France, la LLC représente 1 % de l'ensemble des cancers. Elle se situe en 19<sup>ème</sup> ou 20<sup>ème</sup> position parmi les cancers suivant qu'il s'agisse d'un homme ou d'une femme. Chez l'homme, entre 2 000 et 2005, l'incidence de la LLC reste relativement stable. Les mêmes chiffres sont observés chez la femme avec une augmentation des effectifs, mais une stabilité de l'incidence. Chez la femme et l'homme, le taux de survie est identique et de 94 % à un an. A trois et cinq ans, la survie relative est beaucoup moins bonne chez les hommes que les femmes.

L'étude de la survie relative sur les différentes périodes calendaire est assez décevante : la survie reste identique entre 1989 et 1997, ce qui est vrai pour quasiment toutes les hémopathies malignes. Ce résultat contraste avec les publications hospitalières qui annoncent une amélioration significative de la survie.

Plus l'âge du patient est élevé, moins grandes sont les chances de survie. Environ un tiers des hémopathies malignes survient après l'âge de 75 ans. Une étude américaine publiée en 2008 fait état de probables gains en matière de survie dans la période récente. Entre 1980-1984 et 2000-2004, une amélioration de la survie relative et de la survie absolue est notée à cinq ans

et à dix ans. La survie conditionnelle (à savoir l'étude de la survie lorsque les patients ont survécu dix ans après le diagnostic) diminue, ce qui montre que la LLC reste encore en 2010 une maladie hématologique incurable.

Xavier Troussard conclut que la Basse Normandie représente un atout majeur en termes de registres, notamment à travers cette base de données, accessible pour l'ensemble du cancéropôle. En Basse-Normandie, un réseau régional a été mis en place, qui permet pour chaque patient ayant une hémopathie maligne d'avoir ses cellules cryo-préservées dans la tumorotheque de Caen Basse-Normandie (TCBN).

## INTERETS ET LIMITES DE L'ANALYSE GENOMIQUE DANS LA MALADIE DE WALDENSTROM

---

**Xavier LELEU**

*Service des maladies du sang du CHRU de Lille*

### **XXV. Présentation de la maladie de Waldenstrom**

Xavier Leleu et son équipe se sont lancés dans l'étude de la génomique de la maladie de Waldenstrom. Il s'agit d'un lymphome lympho-plasmocytaire (lymphoprolifération chronique de phénotype B) qui présente la particularité de sécréter de l'IgM dans le serum, décrite en 1944 par Jan Waldenstrom. Cette maladie est considérée comme rare, ce qui probablement inexact : en effet, c'est la forme symptomatique qui est vraisemblablement assez rare (à savoir un à 4 pour un million d'habitants). Cette maladie est probablement plus fréquente sous sa forme asymptomatique : des patients pouvant rester asymptomatiques pendant des années, voire toute leur vie. Certains patients pourraient rester non diagnostiqués. Cette maladie est le plus souvent découverte fortuitement, à l'occasion d'une électrophorèse lorsqu'on constate le pic d'IgM, ou si elle devient symptomatique.

La cellule tumorale de la maladie de Waldenstrom est post-germinative avec la capacité de réaliser le « switch » isotypique : les cellules, apparemment tout à fait capables de « switcher », ne le font pas, et sont bloquées au stade lympho-plasmocytaire. Une particularité morphologique de cette maladie est que le clone mélange fréquemment des cellules de type lymphocytaires, des lymphocytes, des plasmocytes et des lymphoplasmocytes. L'immunophénotype de la cellule tumorale est celui d'un lymphocyte B qui n'est pas différent du lymphocyte B normal, mais clonal. Cette maladie est assez difficile à étudier en laboratoire de part le peu de différence avec le lymphocyte B normal.

### **XXVI. Les avancées de la recherche**

Du point de vue de Xavier Leleu, des progrès significatifs n'ont été réalisés que depuis les années 1990 à 2000. Il y a peu d'équipes dédiées à cette maladie, en lien avec l'incidence rare des patients symptomatiques. Le regroupement des données clinico-biologiques, le travail en réseau et au niveau Européen ont permis de faire des avancées significatives tant dans la connaissance de la maladie que dans sa thérapeutique. Pourtant, la physiopathologie de la maladie de Waldenstrom reste inconnue.

#### **1. Résultats des études de cytogénétique conventionnelle et moléculaire**

La recherche a produit de toutes petites séries, françaises, espagnoles et américaines. Peu de récurrences ont pu être détectées : les patients ont des profils différents, ce qui complexifie l'étude de cette maladie. En cytogénétique conventionnelle, la majorité des patients présentent un profil normal. Deux grandes anomalies ont été détectées touchant 2 gènes différents, la trisomie 4 et la délétion 6q.

Résultats des études de transcriptomique : deux études ont été menées, aux Etats-Unis et en Espagne. L'étude américaine a suscité beaucoup d'espoir, puisque le profil des maladies de Waldenstrom était tout à fait différent de la LLC et du myélome. Les patients qui sont délétés pour l'anomalie 6q n'ayant cependant pas des profils différents des autres patients. La

deuxième étude de Gutierrez a confirmé que le profil transcriptomique des patients est très proche du lymphocyte B normal.

## 2. L'étude sur les miRNA de Roccaro

Xavier Leleu trouve les résultats de cette étude relativement intéressants, car les principaux miRNA ont été retrouvés et sont surexprimés, à l'exception du 9 qui est diminué. Ils revêtent tous un rôle pronostic qui est corrélé au score pronostic, l'ISS Waldenstrom (bas/intermédiaire/haut). Le rôle du Mir155 a été étudié plus particulièrement, dans la lignée BCWM1, développée chez un patient aux Etats-Unis. Le KD de ce miR bloque la pousse de la lignée BCWM1 induite par les cellules stromales. Ce miRNA est très intéressant pour l'étude des gènes impliqués dans l'adhésion, la migration, le contact entre la cellule stromales et la cellule tumorale. Xavier Leleu conclut que ces différentes manipulations prouvent que Mir 155 joue un rôle très intéressant dans la maladie de Waldenstrom.

## 3. L'étude pré-transcriptomique

Cette étude a été limitée par le fait que c'était la forme non phosphorylée donc non active de la polymérase 2 qui était étudiée. Cette analyse a permis de mettre en évidence de très nombreux gènes prêts à être transcrits répartis sur l'intégralité du génome. Pour trouver des anomalies spécifiques, il faudra donc considérer tout le génome, ce qui ne simplifie pas les études. On constate un profil d'expression de NFkB, et de tous ses régulateurs, positifs et négatifs, ce qui laisse à penser que la transcription des gènes impliqués dans NFkB est tout de même majeure.

## 4. L'étude en CGH

Cette étude en CGH a permis d'identifier de nouvelles anomalies non repérées en cytogénétique conventionnelle (gain sur le X, amplification 6p chez les 6q) sachant que ces résultats ont un faible niveau de récurrence. En outre, les échantillons provenaient de trois centres différents, avec des techniques d'extraction différentes, ce qui a sensiblement parasité l'analyse. Enfin, il est impossible d'étudier les anomalies équilibrées notamment les UPD avec la technique de la CGH ou de mettre en évidence un sous-clone.

Cette étude de CGH a été intéressante en ce qu'elle permet de repérer en 6q quatre zones particulières et un gène particulier retrouvé chez au moins quatre des cinq patients sur A20, un régulateur de la voie NFkB qui est porté par le chromosome 6. Les auteurs ont pu démontrer qu'une mutation sur ce gène expliquait sa perte d'expression. Cette découverte est assez intéressante : les auteurs sont dorénavant en train de chercher en quoi la perte de la régulation de ce gène affecte la voie NFkB.

## 5. Analyse des gènes impliqués dans la différenciation terminale du lymphocyte B

Dans la mesure où la cellule tumorale de la maladie de Waldenstrom est bloquée dans sa différenciation terminale : apparemment, la cellule a tout ce qu'il faut pour se différencier et elle ne se différencie pas. Les principaux gènes impliqués dans la différenciation terminale du lymphocyte B vers le plasmocyte, Pax5, BLIMP1, XBP1, IRE1 ont été tous étudiés : il s'avère que tous les patients expriment à peu près aussi bien ces gènes que les témoins normaux.

Aucune différence n'a été trouvée entre l'échantillon de patients et de donneurs concernant Pax5 et Blimp1. Xbp1 « splicé » présente des variations, ainsi que son enzyme Ire1, Xbp1 s'est

avéré être une cible relativement intéressante et surtout son enzyme. Aucune mutation acquise n'a été retrouvée : comme les chercheurs avaient systématiquement le matériel constitutionnel et le matériel tumoral, ils ont séquencé en parallèle les gènes et on n'a trouvé aucune anomalie acquise, mais un déséquilibre de fréquence important dans ces gènes. En outre, Stéphanie Poulain avait mis en évidence depuis quelques années que certains *snip* avaient un rôle pronostic pris un par un.

Pour Xavier Leleu, ceci démontre la nécessité d'aller beaucoup plus loin dans l'analyse, d'où l'émergence de l'idée de lancer des études génomiques et d'utiliser des puces arrays pour étudier la maladie de Waldenstrom.

## 6. L'étude protéomique

L'étude protéomique n'a rien apporté : 98% des protéines sont strictement inutiles en l'absence d'analyse de leurs formes active, souvent phosphorylées.

## XXVII. L'étude génomique

### 1. Bilan de connaissances sur la maladie de Waldenstrom

Xavier Leleu conclut qu'aucune voie de signalisation, anomalie génétique, oncogène ou gène suppresseur de tumeurs n'ont été mis en évidence par ces diverses expérimentations. Il ajoute que :

- la possibilité de l'existence de sous-clones est problématique ;
- il est impossible d'étudier la phosphorylation en l'absence de puce *ad hoc* sur le marché ;
- il est difficile d'aller au-delà de petites séries, même en regroupant des centres (40 séries maximum) ;
- le taux de récurrence est faible, ce qui est d'autant plus problématique avec de petites séries

### 2. La génomique, une nouvelle approche

L'approche par la génomique présente l'avantage de sa grande sensibilité, d'autant que Xavier Leleu et son équipe ont systématiquement comparé le matériel constitutionnel et tumoral. Les auteurs ont comparés des patients symptomatiques et asymptomatiques en posant l'hypothèse que leur présentation serait vraisemblablement différente. Par cette approche génomique, Xavier Leleu ambitionne de trouver une signature de la maladie de Waldenstrom.

### 3. Conditions de l'expérience

L'équipe a travaillé sur Affimetrix, avec 23 patients, dont 12 patients symptomatiques et 11 patients asymptomatiques (n'ayant subi aucune évolution au cours de l'année précédent le prélèvement). Les patients symptomatiques ont été traités très rapidement après le prélèvement. Tous les patients ont une cytogénétique conventionnelle et une étude FISH

Un couple tumoral/non tumoral a systématiquement été utilisé (lymphocyte T) et le bain de bouche a été récolté pour avoir les cellules buccales. La sélection cd19 s'effectue en négative et non en positive, alors tous les précédents travaux ont été réalisés en sélection positive, avec

le risque d'activer de nombreux gènes dont le BCR. Enfin, les deux anomalies les plus retrouvées dans maladie de Waldenstrom, la délétion 6q et la trisomie 4, ont été étudiées.

## 4. Résultats

### **CNV (copy number variation)**

Les anomalies en caryotype conventionnel et en FISH sont systématiquement retrouvées, ce qui confirme systématiquement la cytogénétique. Il y a beaucoup de CNV, qui sont surtout concentrés chez les patients symptomatiques. Pour Xavier Leleu, ceci laisse penser que les cellules tumorales des patients asymptomatiques sont très proches des lymphocytes B normaux et qu'en se focalisant sur les symptomatiques, on verrait des différences. Cette technique permet de récupérer des anomalies extrêmement petites, qui ne sont mesurées par aucune des autres techniques, CGH comme cytogénétique conventionnelle.

Le premier exemple porte sur un patient qui a une trisomie 4 et une délétion 6q, avec un gain mineur. Le deuxième patient affiche une perte encore plus petite sur le chromosome 13. Cette information est très intéressante, car c'est une anomalie que seule cette technique pouvait détecter. Si un seul patient présente cette anomalie, il est assez peu probable qu'il s'agisse du futur gène cible de la maladie de Waldenstrom.

S'il y a des sous-clones (deux caryotypes conventionnels différents), on atteint les limites de la technique. Cette technique a été utilisée pour étudier les LOH et pour voir si le mécanisme d'UPD était fréquemment retrouvé.

### **LOH**

On constate l'existence de beaucoup de LOH (67 %) et en grande majorité des UPD. Cette constatation permet d'initier un travail sur les gènes altérés par les mécanismes d'UPD. Ceci touche quasiment tous les chromosomes pour la majorité des patients, ce qui nécessite d'étudier de très nombreux gènes. Il n'y a pas beaucoup de récurrence, ce qui ne permettra pas d'avancer facilement dans le projet.

Lors du CGH, de nombreux patients ne présentaient pas d'anomalie. Ici, il n'y a que trois patients qui ne présentent pas d'anomalie. Parmi les 9 patients qui n'avaient aucune anomalie en cytogénétique conventionnelle ou/et en FISH, six présentent au moins une CNV ou une UPD et seulement trois n'ont rien. Ces anomalies se répartissent partout, sur tous les chromosomes. A ce stade, Xavier Leleu se propose d'étudier les distinctions asymptomatiques et symptomatiques et de vérifier les anomalies par Copie Number PRC temps réel chez Applied.

Xavier Leleu conclut que l'analyse génomique de la maladie de Waldenstrom ne révolutionnera pas la compréhension de la maladie. Ceci dit, cette technique est extrêmement sensible et dégage de nombreuses données.

Xavier Leleu remercie Stéphanie Poulain, Christophe Roumier, Bruno Quesnel, ainsi que les patients, notamment les personnes asymptomatiques qui ont accepté de donner leur moelle.

## REMISE DES PRIX ET CONCLUSION

---

**Frédéric Leprêtre**  
*Bioinformatique, IRCL*

Frédéric Leprêtre procède à la remise des prix :

- Franck Ladam, Gabriel Bidaux et Edouard Cornet remportent le 3<sup>ème</sup> prix de ce colloque. Frédéric Leprêtre leur remet le chèque de 300 euros correspondant au prix.
- Sami Joha remporte le 2<sup>ème</sup> prix. Frédéric Leprêtre lui remet le chèque de 500 euros correspondant au prix.
- Olivier Nibourel remporte le 1<sup>er</sup> prix et le chèque de 700 euros.

En guise de conclusion, Frédéric Leprêtre remercie les participants au colloque et les partenaires du Cancéropôle.