

**Séminaire d'Animation de l'Axe 1 du
Cancéropôle Nord-Ouest :**

*« Le Séquençage de Nouvelle Génération dans le Cancéropôle
Nord-Ouest : Evolution, Résultats, Projets »*

INTRODUCTION

Pierre FORMSTECHE

Le Président du Cancéropôle Nord-Ouest ouvre la séance. Il rappelle qu'une première réunion sur le séquençage de nouvelle génération avait eu lieu à Deauville et avait entraîné des débats passionnés sur ces nouvelles techniques de séquençage qui révolutionnent la recherche et le transfert au diagnostic. La génomique est une des forces du Cancéropôle Nord-Ouest. Le séquençage de nouvelle génération et la médecine personnalisée sont des priorités du plan Cancer III, lancé par le Président de la République le 4 décembre 2012.

Développement du NGS en recherche et diagnostic en 2012 : situation internationale, succès et contraintes

Thierry FREBOURG, Inserm U1079 et Fédération de Génétique, Rouen.

Thierry FREBOURG indique que le NGS est une révolution conceptuelle et technologique.

Dans l'histoire de la génétique, l'ère actuelle est celle de la génomique. Cette entrée en génomique s'est effectuée par vagues successives avec l'entrée de la CGH. Les grandes publications historiques sur le séquençage de nouvelle génération datent de la fin de l'année 2009 et du début de l'année 2010. Le NGS permet d'avoir accès à l'intégralité des exomes. Il est fascinant d'avoir accès, pour la première fois dans l'histoire de l'humanité, à l'intimité de notre patrimoine génétique. En 2011, de nouvelles machines ont fait leur entrée dans l'univers du diagnostic. Cette révolution conceptuelle est d'abord une révolution technologique.

Les trois grands leaders technologiques utilisent l'amplification clonale, sectorisée par deux processus : l'émulsion et l'amplification sur phase solide. La première inventivité du séquençage de nouvelle génération est l'amplification en très grande quantité. La seconde inventivité du NGS relève de la chimie. Certaines technologies utilisent de nouveaux types d'analogues régressifs.

Il est possible de classer les appareils existant sur le marché selon leur débit. Les appareils disponibles ont des puissances différentes. Le format de certains appareils est davantage compatible avec un secteur diagnostic. Certains appareils sont plutôt dédiés à la recherche.

Deux types d'analyse sont réalisés par les appareils :

- Une analyse ciblée Capture / amplicons (altérations exoniques de gènes connus) (technologie Sanger). Le but est de rechercher des altérations exoniques de gènes connus ou de marqueurs tumoraux.
- Une analyse globale d'exomes ou de génomes et transcriptomes (altérations privées et nouvelles) : cette démarche requiert des appareils de grande puissance et de haute technologie.

Depuis un an, on assiste à une relative régression des publications sur le génome. L'exome, qui représente 1% du génome, est déjà extraordinairement complexe à interpréter.

L'analyse globale vise l'exploration de nouveaux types d'altération comme des inversions complexes.

La scission entre les types d'appareils est en train de disparaître. La logique du NGS est une logique essentiellement ciblée. La recherche est essentiellement ciblée sur une logique exomique. Néanmoins, on constate une entrée prudente et sectorisée de l'exome dans le diagnostic.

En 2012, le NGS a eu de nombreux impacts.

La découverte insoupçonnée de la variabilité du génome humain est un choc philosophique. Il existe environ 15 000 « Single Nucleotide Variations » (SNV). 50% des variations sont non synonymes. 1500 SNV sont non répertoriées dans les bases de données. La clé du NGS est donc l'interprétation. Le temps de « l'innocence génétique » est révolu. On sait aujourd'hui que les maladies génétiques sont le prix à payer de l'adaptabilité de l'homo sapiens. Les bioinformaticiens ont un rôle important à jouer.

En termes de maladies, on note une révolution dans la caractérisation des bases moléculaires. Des publications ont démontré qu'à partir d'un exome d'un individu présentant un phénotype extrême, la soustraction des exomes des parents permettait d'identifier les mutations de novo à l'origine de la maladie. Il est possible à partir d'un sujet d'une famille d'identifier les bases moléculaires qui constituent des portes d'entrées physiopathologiques. Plus que jamais, l'évaluation du phénotype se révèle cruciale pour que les données du NGS puissent être interprétées.

Trois stratégies se sont avérées extrêmement payantes grâce au NGS pour identifier les bases de maladies inconnues avant le NGS. Le maître mot est la notion d'exomes comparatifs. Tous les travaux ne peuvent se faire sans un travail bioinformatique conséquent. Rouen est un centre de référence pour les formes précoces de la maladie d'Alzheimer. Les logiciels permettent d'identifier les gènes les plus mutés. Des filtrations ont lieu selon le principe des stratégies de récurrence. Le gène SORL1 a été identifié. Il est fascinant d'être capable d'identifier une variation parmi 15 000 autres.

La seconde stratégie concerne les exomes comparatifs entre des apparentés distants. La démarche vise à identifier ce que les apparentés ont en commun génétiquement.

La troisième stratégie est celle des phénotypes extrêmes. Elle consiste en l'analyse comparative d'un sujet. Cette approche est particulièrement utile pour les cancers précoces.

L'optimisation du diagnostic moléculaire est un impact important du NGS. Ce chantier vise l'intégration dans les structures de diagnostic, très soumises aux contrôles de qualité au niveau national et européen. La formation du personnel est très importante. Le but est de

travailler mieux, plus vite, avec des techniques ciblées. Le site de Rouen a travaillé sur le diagnostic des formes héréditaires de cancer colorectal. Il convient d'analyser plus de patients plus rapidement. Dans un second temps, on peut analyser des régions non couvertes par le Sanger. La question de la sécurité et de la qualité de l'analyse est primordiale. Pour un sujet présentant une forme héréditaire de la maladie, des fichiers Excel permettent de présenter toutes les étiquettes des gènes. La localisation, les mutations, l'impact, les qualités de profondeur et l'impact biologique sont précisés. Des outils permettent une lecture globale en termes de profondeur.

Le dernier Congrès International de génétique a eu lieu à San Francisco du 6 au 11 novembre 2012.

549 communications concernaient les exomes. On constate l'entrée de l'exome au titre du diagnostic. Dans le cas de pathologies extrêmes, des exomes comparatifs sont envisageables, pour identifier des formes d'épilepsie néonatale par exemple. Néanmoins, il est important de casser l'idée de l'analyse de l'exome de tout sujet. Cette démarche est proposée aux Etats-Unis. Elle n'a pas de sens scientifiquement. Il n'existe aucune

certitude de la conformité diagnostic en termes de profondeur de capture. Dans l'immensité des cas, les variations sont inconnues sans interprétation possible. Des catastrophes interprétatives apparaissent lorsque l'on ne tient pas compte des effets individuels, psychologiques et transgénérationnels.

La nouvelle génétique « sérieuse » se concentre toujours sur le phénotype, une analyse exomique, la filtration bioinformatique, la filtration in silico, et l'analyse biologique des mutations. Le NGS va renforcer les interactions entre les cliniciens et les généticiens.

Dans le domaine de la génétique somatique, l'avenir réside sans doute dans la combinaison de l'approche exomique et la quête de ciblage thérapeutique : l'exome soustractif. Le NGS va révolutionner la chimiothérapie car elle sera ciblée.

Diagnostic des formes héréditaires de cancer du sein et de l'ovaire par NGS, bilan des 549 premières familles

Dominique Vaur, Plateforme SéSAME CRLCC François Baclesse, Caen.

Il est rappelé que 5 à 10% des cancers du sein ou de l'ovaire surviennent dans un contexte de prédisposition héréditaire. Ne sont rendues en diagnostic que les mutations touchant les gènes BRCA1 et BRCA2. Toutes les autres variations sont aujourd'hui mises de côté.

D'une année sur l'autre depuis 2003, le pourcentage global de mutations ponctuelles identifiées chez les cas index reste stable, aux alentours de 10%. Dans 90% des cas, aucune réponse ne peut être donnée aux familles et aux oncogénéticiens. La question est de savoir comment améliorer le diagnostic qui est insuffisant. Deux hypothèses existent: la première est qu'il existe des mécanismes actuellement inconnus impactant les gènes BRCA1 et BRCA2, avec des événements introniques profonds impactant les transcrits ; l'autre est qu'il existe des mutations dans de nombreux autres gènes.

Monsieur VAUR présente le Workflow d'analyse diagnostic du laboratoire. Le Sampleprep a été robotisé grâce à deux robots à haut débit. Une extraction est réalisée à partir du sang total. Une normalisation est réalisée pour obtenir des ADN à concentration identique. Le robot réalise la ligation des adaptateurs pour des séries de 80 patients + 8 témoins en un seul Run. Après contrôle sur bioanalyseur, la PCR est réalisée. Une purification est ensuite réalisée sur un autre robot. Le robot réalise ensuite une normalisation/pooling. 88 échantillons sont traités en même temps. Dans chaque ligne du NGS, 10 échantillons à séquencer sont « poolés » en amont de la capture avec un échantillon témoin différent correspondant au panel des mutations les plus difficiles à mettre en évidence. Aujourd'hui, le workflow est totalement robotisé. Il permet la réalisation de diagnostics moléculaires à haut débit dans une qualité sécurisée.

En 2013, l'enjeu est d'utiliser des codes-barres moléculaires pour identifier les échantillons. Les bibliothèques capturées sont séquencées en « Paires end 2x76 dp ». Actuellement la difficulté n'est pas de séquencer mais d'assurer en amont un processus suffisamment solide pour avoir des données reproductibles et permettre une bonne analyse bioinformatique en aval. L'analyse bioinformatique est réalisée selon deux pipelines parallèles. Les deux pipelines permettent d'obtenir des tableaux Excel de résultats destinés aux biologistes. Les données issues du pipeline CASAVA sont également traitées par un script destiné à identifier les CNV : CNVseq.

Toute région capturée pour laquelle au moins une base n'a pas la couverture requise est retraitée par une technique standard. Les tableaux Excel fournis aux biologistes sont préannotés à partir de bases de données du laboratoire. Pour une série de 88 patients, il

faut compter environ une heure d'interprétation et une trentaine de Sanger en aval. Pour les CNV, des documents papier présentant les réarrangements génomiques sont fournis.

Les NGS sont des machines extraordinaires mais d'une mise en œuvre complexe et d'un coût très élevé. Les principales difficultés rencontrées lors des premiers mois d'utilisation sont les suivantes :

- Le changement de version du logiciel du GAIIX
- Concernant le Kit SureSelect Agilent : un flacon de tampon d'élution était vide, un tampon d'élution était défectueux, un set de capture dégradé.
- Concernant les réactifs Illumina, l'oubli d'un réactif à la reconstitution pour read2.

L'une des bonnes surprises à été la mise en évidence de la mutation C.156_157insAlu dans l'exon 3 de BRCA2 qui nécessitait auparavant la mise en œuvre de techniques spécifiques.

549 familles adressées au laboratoire en 2012 ont été séquencées sur GAIIX, avec 3 sets de capture différents. Les analyses sur les régions exoniques sont effectuées à +/- 50pb. Les variations sur les autres gènes que BRCA1 et BRCA2 sont pour la plupart non confirmées par Sanger à ce jour. Leur caractère délétère n'a pas encore été étudié. Actuellement aucune recherche de CNV n'a été effectuée sur les autres gènes. L'interprétation biologique dans les autres gènes n'est pas réalisée. Il a été nécessaire de créer une base de données pour extraire les données sur l'ensemble des gènes. L'analyse présentée est qualitative et non supervisée. Aucune interprétation sur la significativité des variants n'est effectuée. Les gènes sont anonymisés sur les graphiques.

Les données descriptives pour les 549 familles séquencées sont filtrées sur les scores de qualité et sur les fréquences ESP. Les mutations « tronquantes » sur l'ensemble des gènes séquencés sont les suivantes : 21 mutations non-sens, 45 mutations frame-shift exoniques, 23 mutations « d'épissage ». Environ un tiers des mutations non sens sont présentes dans d'autres gènes que BRCA1 et BRCA2.

80% des variants d'épissage seraient liés à d'autres gènes que BRCA1 et BRCA2. La moitié des mutations trouvées ne sont pas dans BRCA1 et BRCA2.

En conclusion, on peut dire qu'il se passe des événements moléculaires dans les gènes autres que BRCA1 et BRCA2 de manière extrêmement significative. Ce constat valide la pertinence du séquençage à haut débit dans une démarche diagnostique.

En guise de conclusion, Monsieur VAUR cite l'avocate Anne-Laure MORIN concernant « l'épineuse question des informations incidentes » : « ... la loi relative aux recherches impliquant la personne humaine du 5 mars 2012 (dite loi Jardé) prévoit la possibilité, sous conditions, d'une utilisation secondaire des échantillons biologiques pour un examen des caractéristiques génétiques. Cette loi crée l'obligation de révéler à la personne, lorsqu'elle est recontactée, le diagnostic d'une anomalie génétique grave. »

Détection des mutations de novo par analyses de trios dans les formes précoces de cancer

Isabelle Tournier, Inserm U1079-Plateforme rouennaise de génomique.

En génétique des cancers, on considère deux grandes catégories de cancers : les cancers de présentation sporadique, résultant de l'accumulation séquentielle au cours de la vie de mutations somatiques, et les cancers de présentation familiale, résultant de mutations

constitutionnelles suivies par des mutations somatiques. Cette deuxième catégorie de tumeurs se caractérise par un âge de survenue précoce.

Les études récentes menées sur les génomes et les exomes ont révélé chez l'homme un taux de mutation de novo très élevé. Lors de la réalisation d'un exome, on s'attend à trouver entre 0 et 1 mutation de novo potentiellement à effet biologique. Ces mutations de novo peuvent intervenir à un niveau prézygotique ou post-zygotique après la fécondation. Face à ce taux élevé de mutations, une hypothèse a été formulée : des mutations de novo rares pourraient expliquer une fraction des cancers de survenue précoce, de présentation sporadique. Ces mutations rares pourraient intervenir sur des gènes différents de ceux impliqués dans les pathologies.

Le projet développé s'articule autour de trois points. La première étape consiste en la sélection clinique des phénotypes extrêmes : il s'agit de tumeurs très précoces de présentation sporadique (cancers précoces de l'ovaire ou du sein, cancers précoces du colon, tumeurs pédiatriques sans mutation détectée dans les gènes connus). La deuxième étape consiste en la constitution de trios enfant/parents indemnes. La troisième étape est la réalisation d'une analyse exomique du trio par soustraction des variants hérités et identification des mutations de novo.

Le cancer de l'ovaire a été choisi comme paradigme pour ce projet. Ce cancer touche une femme sur 10 000 par an en France. Le plus souvent, il s'agit d'une tumeur épithéliale. Le pic de fréquence du cancer est atteint entre 60 et 70 ans. De ce fait, la présence d'un cancer épithélial chez la femme jeune est très fortement évocateur d'un événement génétique majeur chez la patiente.

La patiente suivie pour cette étude a présenté à l'âge de 21 ans un cystadénocarcinome séreux de bas grade, compliqué de nombreuses métastases ayant conduit à une opération et d'une chimiothérapie. Les parents de la patiente, âgés de 52 et 56 ans, indemnes de cancer, ne présentaient pas d'histoire familiale avec des cancers.

Un criblage extensif a été réalisé sur les gènes BRCA1, BRCA2 et TP53. Aucune mutation délétère n'a pu être mise en évidence. Au laboratoire, un criblage en CGH-array haute résolution a été réalisé et a donné des résultats négatifs. L'hypothèse d'une mutation de novo a été formulée. Une analyse exomique soustractive a été réalisée sur le trio d'individus. Les liens de parentés ont été confirmés par une analyse de microsatellites. Chez la patiente, le séquençage exomique a été effectué sur du tissu ovarien non tumoral afin de permettre une détection de mutations de novo intervenus après la fécondation.

Pour le séquençage exomique, le kit SureSelect Human All Exons 50 mb a été utilisé. La librairie d'ADN est ensuite séquencée sur un séquenceur Illumina Paired-end 2x75pb GAIIX. Après l'analyse des images, les reads sont alignés. Les variants sont étudiés avec le logiciel CASAVA 1.8.2 puis annotés avec le logiciel ANOVAR Variant Effect Predictor.

8 Gb ont été séquencés chez la patiente. 87% de ces bases ont été séquencés avec un score de qualité élevé. La profondeur moyenne des bases chez la patiente a été de 77x. 88% des bases ont été lues avec une profondeur minimale de 10x.

L'analyse des variants de la patiente a été réalisée à partir du logiciel EVA (Exome Variation Analyzer). 15 715 variants ont été détectés chez la patiente, dont 14 446 variants connus (dbSNP131) et 1269 variants non connus. À l'aide du logiciel EVA, des filtres ont été appliqués sur les variants pour identifier les mutations de novo. Le filtrage a permis de passer de 15 715 à 307 variants. Un nouveau filtrage est appliqué à partir des variants présents chez les parents. On passe alors de 307 à 47 variants, potentiellement de novo. Ces 47 variants sont visualisés sur les fichiers BAM de la patiente et des parents grâce aux logiciels Alamut et IGV. De manière surprenante, 37

variants sur 47 étaient présents sur les BAM des parents mais n'avaient pas été détectés lors du variant calling. Ces variants pouvaient correspondre à des faux négatifs chez les parents. En réalité ces variants sont des faux positifs chez l'enfant. Il s'agit d'artefacts de séquences. Les variants n'étaient lus que sur un des deux reads réalisés et ne passaient pas l'étape du variant calling. Certains de ces faux positifs étaient répertoriés dans les bases de données et étaient associés à un numéro RS. Les variants détectés sur moins de 20% des reads ont été éliminés. En filtrant les variants restants sur la base de donnée interne des 72 exomes réalisés pour d'autres pathologies, une mutation de novo a pu être identifiée chez la patiente avec un Qscore très élevé, présente sur 50% des reads. L'existence de la mutation a été confirmée par Sanger.

La mutation identifiée est une mutation hétérozygote faux-sens qui affecte le gène Inhibin Beta A. La mutation est probablement intervenue chez la patiente au niveau prézygotique, sur une des gamètes des parents.

Le gène INHBA code pour l'Inhibin Beta A qui fait partie de la famille du TGF-Beta. L'Inhibin Beta A est sécrétée par les cellules de la granulosa de l'ovaire. Les Inhibin ont un rôle clef dans le développement ovarien.

La mutation identifiée chez la patiente intervient sur un nucléotide très conservé de ce gène et touche un acide aminé conservé jusqu'au poulet. Cet acide aminé est présent dans le domaine mature de l'Inhibin de la chaîne Inhibin Beta A. Dans le but d'évaluer l'impact fonctionnel de cette mutation, l'équipe de Madame TOURNIER a collaboré avec le Docteur Craig HARRISSON, spécialiste australien de ces protéines. Les analyses ont montré que cette mutation n'avait pas d'effet sur le folding et le processing de l'Inhibin A ou l'Activin A. En revanche, N386S favorise la production d'Inhibine A et réduit la production d'Activine A. Les Activines sont de puissants suppresseurs de croissance des cellules épithéliales ovariennes tumorales.

Dans le but d'identifier de nouvelles mutations de ce gène chez d'autres femmes ayant développé des cancers épithéliaux à un âge précoce, une cohorte de 54 patientes ayant eu un cancer épithélial de l'ovaire avant l'âge de 40 ans a été sélectionnée, avec un criblage BRCA1/BRCA2 négatif. 19 patientes étaient françaises et 35 américaines. Aucune mutation du gène INHBA n'a été détectée chez ces patientes. Dans un deuxième temps, un séquençage Sanger des gènes INHBA et INHBB a été effectué. Aucune mutation n'a été détectée sur le gène INHBB. En revanche, une mutation hétérozygote « c.179G>T p.R60L » a été identifiée sur le gène INHBA. Cette mutation n'était pas répertoriée dans les bases de données. Cette mutation a été détectée chez une patiente ayant présenté un cystadénome séreux papillaire à l'âge de 29 ans. Ses deux parents étaient indemnes de tumeur. Cependant, une histoire familiale existait des deux côtés. La mutation affecte un site de clivage secondaire pour la furine, la pro-convertase, responsable de la maturation des Inhibines. Des mutations dans les sites de clivage des furines ont déjà été mises en évidence dans d'autres gènes de la famille du TGF-Beta. Ces mutations étaient associées à des maladies génétiques. Cette mutation pourrait être délétère et à l'origine de la pathologie observée.

En Conclusion, Il a été identifié chez deux patientes ayant développé des tumeurs précoces de l'ovaire deux mutations de la voie Activine / Inhibine. Ces deux mutations pourraient affecter le ratio Activine/Inhibine. Ces résultats suggèrent un défaut dans la communication entre cellules de la granulosa et les cellules épithéliales. Il pourrait s'agir d'un mécanisme oncogénique. Sur la cohorte séquencée, on a détecté dans 4% des cas une mutation de la voie des Inhibine. Il est important de confirmer ces résultats dans d'autres cohortes.

Ce travail renforce l'idée selon laquelle les mutations de novo rares, intervenant dans les gènes non décrits, pourraient être à l'origine des formes précoces de cancer de présentation sporadique.

Recherche de mutations du gène TET2 par séquençage à haut débit sur plateforme Roche

Olivier Nibourel, CHRU Lille.

Le gène TET2 (Ten-Eleven Translocation gene 2) est localisé en 4q24. Il appartient à une famille de 3 gènes homologues (TET1, TET2, TET3) et a été particulièrement conservé lors de l'évolution. Cette enzyme est oxygénase dépendante du 2-OG et du Fe (II). Sa fonction est la conversion 5-méthylcytosine (5mC) en 5-hydroxyméthylcytosine (5hmC). TET2 hydroxyde les méthylcytosines ce qui les rend invisibles au gène DNNT1, responsable de la maintenance de la méthylation au cours des réplifications. Le gène TET2 participe donc à une déméthylation passive lors des réplifications successives. Les différentes mutations de TET2 aboutissent à une perte de fonction de la protéine ce qui aboutit à une hyperméthylation de certains sites.

Les mutations de TET2 ont été rapportées dans la leucémie myéloïde. Certains travaux ont rapporté, au sein des leucémies aiguës myéloplastiques à cariotype normal, un impact pronostic négatif de ces mutations. Ces résultats restent controversés au regard d'études réalisées sur de larges cohortes.

Les mutations TET2 sont présentes dans d'autres pathologies myéloïdes : les néoplasmes myéloprolifératifs. Ces mutations ont également été rapportées dans les SMD avec des fréquences variables. Les mutations de TET2 auraient un impact sur la réponse au traitement. La dernière pathologie liée aux mutations de TET2 est la CMML.

En raison de la présence de TET2 dans de nombreuses pathologies, les demandes de recherche de mutations de TET2 ont augmenté. Le gène TET2 contient 11 exons. Ce gène est lourd à séquencer en Sanger (17 PCR). L'objectif de l'étude est de mesurer la possibilité d'utiliser le séquençage à haut débit pour rechercher les mutations de TET2. Par rapport aux techniques classiques, le NGS permet une augmentation du nombre de séquences réalisables. La sensibilité de détection est augmentée. Le NGS permet également une standardisation du séquençage. Ce projet a été réalisé en collaboration avec le laboratoire MLL de Munich.

68 patients atteints de pathologies variables ont été étudiés. Pour réaliser cette analyse, la librairie automatisée par PCR simplex en plaque a été préparée. Un séquençage a été effectué sur le séquenceur Roche GS junior. Les séquences ont été validées par séquençage direct.

Pour répondre aux exigences du séquençage sur le Roche GS Junior, le gène a été découpé en 27 amplicons.

L'approche avec les amplicons permet une couverture très homogène. Cette approche a permis d'obtenir une couverture moyenne supérieure à 300 séquences. La sensibilité théorique atteinte est de 7% environ.

Une mutation stop a été identifiée sur l'échantillon n°61 au niveau de l'arginine 1216x. La fréquence allélique a été évaluée à 17%. Parmi toutes les variations identifiées, 3 variations n'ont pas été confirmées par le Sanger. Une validation chez tous les patients permet de mieux connaître les séquences. Au total, les mutations retenues sont celles qui ont été vérifiées en séquençage direct et qui n'étaient pas répertoriées dans les bases SNP. 32 événements ont été identifiés : 11 frameshift, 14 stop, 7 substitutions. 27 patients étaient concernés, donc 5 avaient 2 mutations.

En conclusion, 32 événements distincts ont été détectés chez 27 patients, dont 5 avec une atteinte bi-allélique. Une bonne concordance avec le séquençage direct est soulignée. La stratégie PCR simplex a permis une couverture suffisante et homogène, une augmentation du débit par rapport au séquençage direct. Néanmoins, la construction de la librairie est fastidieuse en dépit de l'automatisation.

Ce projet ouvre plusieurs perspectives : l'extension du panel de gènes avec une transition vers une autres plateforme de séquençage et l'utilisation des approches de PCR multiplex pour la construction des librairies.

Détection de mutations somatiques par NGS sur Ion Torrent

Martin Figeac, Université Lille II

L'ion torrent permet une très grande souplesse et une très grande rapidité d'utilisation. Ce séquenceur est très réactif. La phase d'acquisition d'images a été totalement éliminée du séquençage. Cette technologie est plus simple. Elle permet un séquençage de 400MB en 2 heures 30.

Le Ion Torrent a été utilisé pour identifier le Janus Kinase 2 mais est adaptable à toutes les mutations. La mutation recherchée se situe dans 0,1% des cellules. Son identification permet d'adapter le traitement des patients à l'aide d'interférents. La sensibilité dépend de la couverture choisie. Une mesure relative est réalisée. L'objectif était d'obtenir une quantification absolue, sans gamme. Une dizaine de patients sont traités sur une puce 314 à 70 euros. 350 000 reads ont été réalisés. 3 reads sont positifs. La limite de la méthode est relative au seuil à détecter.

La méthode peut être employée sur tout type de substitution. Des adaptations sont à prévoir lorsque l'on cherche à travailler sur des délétions, des insertions ou sur des régions contenant plusieurs hotspots. Il conviendrait de réaliser les séquençages sur d'autres séquenceurs afin de vérifier que les mêmes reads sont identifiés. Le temps de réalisation pourrait être optimisé. Par ailleurs, les homopolymers posent problème.

En 2012, plusieurs technologies permettent de réaliser des PCR multiplex ou des PCR isolés dans des microréacteurs. La technologie Haloplex présente un intérêt particulier : elle permet d'effectuer un important Tiling. Le Tiling d'Haloplex pose également problème. Les régions à fort Tiling allant jusqu'à 25 amplicons, seront enrichies 25 fois plus. Les régions ayant peu de restrictions seront moins séquencées. Agilent recommande de séquencer de manière plus large. Néanmoins, cette solution ne résout pas le problème de l'hétérogénéité de la couverture à l'intérieur des exons à séquencer. De plus, cette démarche entraîne le séquençage de séquences non intéressantes. Cette technologie est donc perfectible.

Thierry Frebourg note que l'intérêt du Cancéropôle est la mise en commun des technologies. Un regard de veille est effectué sur les différentes plateformes. L'ion Torrent est bien adapté à la détection d'une mutation. Il se demande si cette technologie est adaptée à la détection d'une série de mutations.

Martin Figeac estime que tout dépend du nombre de gènes étudiés. Au delà de 30 gènes, les gros séquenceurs sont plus économiques.

Détection de mutations somatiques par NGS sur GaIIX

Aude Lamy, Laboratoire de Génétique Somatique des Tumeurs, CHU de Rouen.

La médecine personnalisée telle qu'elle est pratiquée actuellement dans les laboratoires a pour porte d'entrée les pathologies. En fonction de la pathologie, on étudie un ou plusieurs biomarqueurs. On adapte ensuite le traitement des patients.

Le séquençage haut débit permet de dépasser l'étude de quelques mutations par des techniques ciblées, ou de quelques exons par séquençage de Sanger. Le NGS permet le séquençage en parallèle des exons de plusieurs gènes, le séquençage en parallèle de tous les exons du génome, le séquençage en parallèle de tout le génome. Le NGS permet donc un séquençage simultané de millions de fragments d'ADN.

Le schéma de l'étude réalisée est le suivant : 17 régions cibles de valeur prédictive de réponse à une thérapie ciblée ont été identifiées. L'ADNg extrait provenait de 12 patients étudiés pour une ou plusieurs régions cibles et présentait des altérations représentatives des régions étudiées (mutations ponctuelles, délétions, insertion / duplication).

Le séquençage haut débit a été réalisé à l'aide de la technologie Illumina. L'objectif de l'étude est d'évaluer la faisabilité du séquençage haut débit pour des échantillon FFPE tout venant et de comparer les résultats avec le génotype de référence.

Pour chaque patient, la méthodologie a été la suivante : l'amplification et la purification des régions cibles, le dosage et la réalisation de pools équimolaires (avec le Twinckle 470), la fragmentation par sonication (avec le Covaris S220), la fixation des adaptateurs et des index et la purification sur billes sans sélection de taille (avec le Primer de seq). Les amorces de l'amplification clonale ont été incorporées lors d'une PCR d'enrichissement. Pour chacun des 12 patients, 17 amplicons étaient concernés. Chaque pool a été marqué par un index différent. Les pools ont été mélangés de façon équimolaire et déposés sur une flowcell. Suite à l'immobilisation des fragments sur la flowcell et à l'amplification clonale, le séquençage a été réalisé sur un séquenceur GAIIX. Les séquences ont été analysées, démultiplexées et alignées sur le génome humain.

La visualisation des résultats s'effectue à l'aide des fichiers BAM grâce au logiciel Alamut. L'analyse a été réalisée sur les 17 régions ciblées. Le premier niveau d'analyse concerne la profondeur de lecture.

La profondeur de lecture augmente avec la taille de l'amplicon. On note une perte des fragments de petites tailles lors de la préparation de la librairie.

Le second niveau d'analyse est celui de l'analyse des reads. Une mutation KRAS c.35G>A a été détectée sur 25 des 87 lectures correspondant à cette position (29%). On note également une délétion au niveau de l'exon 19 de l'EFGFR c.2236_2257delinsATCT.

Sur 14 altérations recherchées (8 mutations ponctuelles, 2 insertions / duplications, 4 délétions), 13 altérations ont pu être identifiées par séquençage illumina. Tout un pannel de mutations a pu être identifié.

La faible représentation de la mutation ne peut pas s'expliquer par le pourcentage de cellules tumorales.

Les résultats préliminaires montrent qu'il est possible d'utiliser une méthode de séquençage haut débit pour séquencer plusieurs régions cibles d'un ADN extrait de blocs FFPE tout venant. Cette étude va se poursuivre. Il reste à effectuer le Variant Calling illumina. L'analyse paired-end par défaut du logiciel Casava s'est avérée non adaptée à la taille des amplicons. L'analyse sera relancée en single read ou avec un logiciel alternatif. Le second objectif de l'étude était de comparer deux technologies NGS : Illumina et Ion Torrent.

Les nouvelles technologies de séquençage haut débit ouvrent la porte à une médecine personnalisée pour le futur. La porte d'entrée des cas traités ne sera plus la pathologie mais le patient. Pour chaque patient, un panel de biomarqueurs pourra être recherché et permettre un traitement ciblé.

Thierry Frebourg note que l'intérêt d'avoir plusieurs technologies à disposition est de capitaliser les avantages et les limites des unes et des autres. La multiplication des technologies multiplie le travail de veille. Le NGS permettra de traiter plus de patients plus rapidement.

Etude par NGS des réarrangements VDJ comme marqueurs de la maladie résiduelle

Mikaël Salson, Université Lille I.

L'étude porte sur la leucémie aiguë lymphoblastique. Cette maladie concerne principalement les enfants. Il convient d'étudier le génome de chacun des lymphoblastes. Le NGS a été utilisé. Le but de la démarche est d'effectuer des paquets de séquences correspondant aux mêmes réarrangements.

Les principales perspectives de cette étude sont les suivantes :

- la finalisation et la diffusion de l'application permettant à l'hôpital de l'utiliser.
- Il conviendrait d'obtenir de nouvelles données de séquençage permettant de revoir la gamme et corriger le biais de PCR.
- Concernant l'évolution multiclonale, il faudrait comprendre si un clone a évolué ou non en plusieurs sous-clones. Il serait intéressant de savoir si le clone du diagnostic a évolué ou non vers une autre forme. La méthode des amorces spécifiques ne permet pas de répondre à ce problème.

Analyse bioinformatique à visée diagnostique

Antoine Rousselin et Baptiste Brault, Plateforme SésAME CRLCC François Baclesse, Caen.

Thierry Frebourg souligne le rôle central de la bioinformatique pour rendre intelligible les résultats du NGS à visée diagnostic et pour l'exploitation de données complexes concernant le resequençage ciblé.

Baptiste Brault indique que le laboratoire travaille sur la prédisposition du cancer du sein et de l'ovaire.

Le séquençage Haut débit est utilisé pour le diagnostic sur les gènes BRCA1 et BRCA2. Il permet un rendu des résultats plus rapide un cas index en 3 mois maximum. Le NGS permet également de rechercher une des mutations ponctuelles et un des réarrangements de grandes tailles. Il est adapté à l'étude sur les variants dans les autres gènes.

L'objectif du laboratoire, après validation de la technique est de réaliser un Run par mois. Un Run représente 80 patients et 8 témoins. 960 patients sont donc traités annuellement. Ce contexte nécessite l'automatisation de l'analyse bioinformatique.

Pour le séquençage, un set de capture est utilisé. Actuellement, la version 3 du set de capture (CASO) permet de traiter 28 gènes, avec une amélioration de la capture sur certaines zones. Pour le rendu diagnostic, seuls les données issues des gènes BRCA1 et BRCA2 sont utilisées.

Avant le mois de mai, l'ensemble des analyses bioinformatiques était exécuté manuellement. A la fin du séquençage, la présence d'un bioinformaticien était nécessaire pour lancer l'analyse. CASAVA, NextGENe et CNV-seq sont les logiciels utilisés pour l'analyse.

CASAVA est utilisé pour les étapes suivantes : le démultiplexage (la conversion des fichiers basecalling en FASTQ), l'alignement (sur le génome humain version hg19), le Variant Calling (la recherche de variants). Après conversion de format, les fichiers sont utilisés par le logiciel commercial NextGENe (alignement, détection de variants et annotation). Les fichiers Bam issues de CASAVA sont interprétés par le logiciel NextGENe pour l'annotation des variants.

NextGENe où chaque patient (88 échantillons) doit être saisi manuellement. Après démultiplexage par CASAVA, les fichiers FASTQ sont convertis en FASTA, puis alignement et Variant Calling avec en sortie différents rapports.

CNV-seq est utilisé à partir des fichiers bam du chromosome à analyser. Création d'un fichier d'entrée (fichiers hits) Puis analyse par CNV-seq (étude des réarrangements de grandes tailles), les fichiers générés par cette analyse sont ensuite traités par un script R qui permet de générer le rapport sur la zone d'intérêt. Les rapports sont mis en forme au format Excel afin d'être intelligibles pour les biologistes.

Du mois de mai au mois de septembre, un pipeline semi-automatique a été mis en place. Les fichiers de séquençage sont présents sur un serveur. Lors du démultiplexage, les fichiers FASTQ sont envoyés sur le serveur d'analyse pour CASAVA. Après démultiplexage, les fichiers FASTQ sont envoyés sur un autre serveur d'analyse pour NextGENe. A l'issue de ces deux pipelines, les rapports contenant les résultats sont envoyés sur un autre serveur. Durant cette période à la fin du séquençage, le démultiplexage, et les différentes étapes de CASAVA sont lancés automatiquement. La présence d'un bioinformaticien n'était plus nécessaire. Seul le pipeline NextGENe restait manuel.

Les premiers pas de l'automatisation ont entraîné quelques déconvenues : il n'existait pas de gestion des erreurs, le code était désorganisé, la maintenance et l'évolution étaient difficiles.

De plus, les résultats du Variant Calling de CASAVA n'étaient pas utilisés. L'analyse des fichiers bam de CASAVA été réalisées par NextGENe qui génère de mauvais résultats. Dans NextGENe, de nombreuses tâches sont manuelles, ajout des 88 patients pour l'analyse, l'importation des fichiers bam de CASAVA et la génération des rapports depuis l'interface de NextGENe.

Depuis le mois de septembre, la phase d'automatisation complète a été mise en place. La nouvelle version de NextGENe (2.30) a permis avec de nouvelles fonctionnalités, dont la génération automatique des rapports lors de l'analyse pour rendre ce pipeline automatique. Alamut Ht a été utilisé pour l'annotation de variants de CASAVA. CASAVA2VCF est utilisé pour convertir les résultats du variant calling de CASAVA (fichiers txt en VCF). Samtools pileup est utilisé pour l'analyse de la couverture. La mise en place du nouveau pipeline a permis le refactoring du code existant, de créer des tests unitaires et fonctionnels pour chaque étape, la création d'un pipeline unique DiagCaso (Diagnostic Cancer Sein Ovaire), le choix du programme à exécuter, la gestion des erreurs, la reprise sur erreurs, la vérification des pré-requis, la génération d'un fichier de configuration, un monitoring complet du pipeline, une analyse bioinformatique pour un rendu en diagnostic complètement automatisé.

Différents fichiers de log sont créés lors de l'analyse. La fin de l'analyse est visualisée par la création d'un fichier `nom_de_l_analyse.completed`. Chaque pipeline permet un contrôle de la qualité de la couverture. Un rapport général regroupant tous les patients permet de visualiser les trous de couverture, le seuil minimal par base à atteindre pouvant être modifié facilement. Un rapport de qualité est généré pour chaque patient. Ce rapport présente plusieurs onglets : BRCA, Exons de la capture, l'ensemble de la capture.

Pour identifier les variants dans BRCA1 et BRCA2, une base de données locale est interrogée. Pour les annotations délétères et neutres, les données utilisées sont celles du groupe génétique et cancer. Pour les SNP, l'expérience du laboratoire est sollicitée.

DiagCaso présente de nombreuses perspectives :

- De nouvelles fonctionnalités du pipeline peuvent être ajoutées avec une amélioration de la qualité comme la détermination des biais de PCR.
- L'étude des variants peut s'étendre à d'autres gènes. Une base de données pourrait être créée avec l'ensemble des variations sur les gènes de la capture. Les résultats pourront être insérés automatiquement.
- Une nouvelle interface de validation des séries pourra être créée avec une sécurisation des données. Cette interface permettra également d'éviter la multiplication des fichiers.

Thierry Frebourg souligne l'obsession de transformer les fichiers avec un grade clinique. Ce travail est la condition nécessaire des normes de qualité. La création de postes en bioinformatique dans des laboratoires de diagnostic sera importante pour le futur.

Evolution de l'outil EVA pour la recherche et le diagnostic

Sophie Coutant, Plateforme rouennaise de génomique.

L'un des grands défis du séquençage à haut débit est l'interprétation des variations. Un besoin important de filtration des variations est apparu. Depuis 2 ans, le logiciel EVA est développé. Il s'intégrait en aval d'un pipeline classique Illumina. La porte d'entrée des données pour intégrer la base de données EVA était les fichiers de variation annotés. Récemment, un module d'annotation permet d'entrer des fichiers PCF. EVA est intégré à davantage de pipelines que celui d'Illumina.

L'outil EVA est constitué d'une base de donnée (exome DB), d'un explorateur, d'outils de recherche et de filtres.

Pour filtrer les données il existe plusieurs stratégies.

La stratégie de récurrence consiste à séquencer en exomes des individus non apparentés malades de la même pathologie, puis à rechercher les gènes les plus souvent altérés au sein de ces patients.

La stratégie familiale consiste à séquencer des individus atteints, d'une même famille, puis à rechercher les gènes affectés par des variations identiques au sein de cette famille.

La stratégie de novo consiste à séquencer le cas index malade et les deux parents sains afin de soustraire les variants hérités des parents chez le cas index pour retrouver les variations de novo.

Concernant la stratégie de récurrence, l'application EVA a été utilisée pour l'analyse de 14 exomes dans les formes précoces autosomiques dominantes de la maladie

d'Alzheimer sans mutation identifiée. Par individu, 15 000 variations sont répertoriées dans les bases de données (dbSNP) et 1500 variations sont non répertoriées dans ces bases de données. La première étape, de filtration des variations, est commune à toutes les stratégies. Avec la stratégie de récurrence, il convient de faire particulièrement attention à l'hétérogénéité génétique.

Avec la stratégie familiale, l'interface EVA permet de filtrer et sélectionner chez les individus apparentés les gènes affectés par des variations non filtrées identiques entre les individus. L'application EVA a été utilisée pour l'analyse de 2 exomes dans la maladie de Fahr. Il apparaît que 17 gènes sont affectés par des variations identiques chez les deux individus sélectionnés.

Avec la stratégie de novo, l'interface EVA permet de filtrer et sélectionner les gènes affectés par des variations présentes chez l'enfant et absentes chez les parents. L'application EVA a été utilisée pour l'analyse d'un trio d'exomes dans le cadre d'un cancer de l'ovaire précoce. 47 gènes affectés ont ainsi été identifiés.

En l'absence de variations de novo identifiées, il serait possible d'analyser la même famille sous l'hypothèse d'une transmission récessive. Les risques de pathogénicité peuvent être connus.

Un moteur de recherche est intégré dans EVA. Cela permet de trouver la fiche d'un gène dont on connaît le nom. Un tableau répertorie les détails de chaque variant sur les gènes identifiés. Pour une interprétation plus poussée, le logiciel Alamut est notamment utilisé. La dernière fonctionnalité ajoutée à EVA est le module de statistiques descriptives permettant pour un patient sur un gène ou pour l'ensemble des patients d'un projet de générer des statistiques sur les types de variants.

EVA est presque prêt à être disponible au téléchargement. L'interfaçage du module d'annotation est en cours de finalisation. EVA sera disponible au téléchargement pour un usage non commercial.

Un intervenant souligne la nécessité de promouvoir la bioinformatique au sein du Cancéropôle Nord-Ouest.

Thierry Frebourg note qu'avec la puissance de la filtration, un patient suffit à être informatif. Ce changement majeur n'est pas encore totalement appréhendé. Il évoque par ailleurs le problème des bases de données contaminées par des artefacts. Une des priorités actuelles est la mise en place de grades cliniques. Certaines variations peuvent avoir une valeur non médicale.

Alamut HT, un outil pour le NGS

André Blavier, Interactive Biosoftware.

Alamut HT a été mis au point en 2007. A l'origine, Alamut est un visualisateur avec des capacités d'aide à l'interprétation des variations. Il est possible d'importer un grand nombre de mutations et d'exporter les résultats. Le but est de permettre un rendu de variants annotés.

Les annotations fournies par le logiciel sont les suivantes :

- Gène(s), transcrit(s), protéine(s)
- Conservation

- Variant : type, localisation, effet
- Identification + fréq : dbSNP, ESP, HGMD
- Protéine/AA : domaines, conservation, scores physico-chimiques
- Faux-sens : Align GVDG, SIFT, MAPP
- Epissage : site naturel le plus proche, voisinage du variant

Les paramètres du logiciel sont les suivants :

- 1 ou n transcrits
- +ou- prédictions de faux-sens
- +ou- prédictions d'épissage
- Gènes d'intérêt
- Régions d'intérêt
- Intégration d'annotations externes
- Extraction d'annotations externes
- Extractions d'annotations VCF (ex : génotypes)

Alamut HT a été débarrassé de l'interface graphique du logiciel. Chaque variant sera analysé face à un ou plusieurs transcrits. Une version serveur existe pour Alamut HT. Une version « stand alone » de la base de données existe également. L'annotation de variants est hautement parallélisable. Un ordinateur pourrait être dédié à l'analyse d'un variant. Dans le futur, l'analyse du pipeline bioinformatique sera probablement possible en quelques secondes seulement.

Après le filtrage, la liste des variants d'intérêt peut être copiée dans Excel ou dans un logiciel équivalent.

Alamut API (Interface de programmation d'application) est un moyen pour des applications externes de contrôler Alamut. Ce dispositif a été intégré à plusieurs logiciels : ERIS (Integragen) a été mis en place. N'importe quel navigateur présentant les résultats de NGS peut être configuré pour que chaque variant puisse être ouvert en lien sur Alamut. Un partenariat a été lancé avec Cartagena. L'interface de contrôle entre logiciels sera étudiée au sein de ce partenariat.

Alamut HT a été commercialisé au printemps. Une douzaine de clients utilisent le logiciel : Le Centre François Baclesse de Caen, l'INSERM U674/CEPH de Paris, le Centre Jean Perrin de Clermont-Ferrand, l'IGBMC de Strasbourg, L'Université de Zurich, le Saint James University Hospital de Leeds, Le Partners HealthCare de Boston, le Transgenomic Labs d'Omaha, le Baylor College of Medicine de Houston, le Child and Family Research Institute de Vancouver, L'Université de Hong Kong, le Hong Kong Molecular Pathology Diagnostic Centre.

Thierry Frebourg souligne le privilège de pouvoir utiliser Alamut.

Evolution techniques et possibilités du RNAseq

Emmanuel Martin, IntegraGen.

Thierry Frebourg indique que l'étude du RNA est un enjeu considérable pour les projets exploratoires. Les développements dans les laboratoires académiques sont essentiels pour imbiber de manière pédagogique tous les acteurs, techniciens, ingénieurs, chercheurs, médecins, internes, jeunes en formation.

Integragen est une société basée à Evry qui existe depuis 10 ans. Elle dispose du matériel suivant : Illumina HiSeq2000 (x2), Illumina MiSeq, Agilent SureSelect Target

Enrichement kits, Fluidigm Access Array, Illumina IScan, Fluidigm BioMark, ABI 7900HT, ERIS.

En termes d'applications, Integragen réalise du séquençage d'exomes, du séquençage ciblé, du séquençage de micro-ARN. Integragen fait le lien entre les applications et les plateformes en mettant en place des process industriels de travail.

L'utilisation des HiSeq2000 chez Integragen concerne plus de 317 projets de séquençage haut débit. Près de 2000 exomes ont été traités dont 1200 environ en 2012. 7 publications en co-auteurs ont été réalisées. Moins de 30 projets RNASeq ont été réalisés à ce jour.

Le RNASeq est intéressant pour plusieurs opérations : la quantification des transcrits, la détection, la découverte et la quantification de variants d'épissage, la détection de transcrits de fusion, la détection de SNP et la comparaison aux données de séquences ADN (exome), la visualisation via un browser.

Différents problèmes doivent être abordés :

- le choix du matériel : ARN total ou mRNA ? Quelle quantité de matériel ? Critères de qualification ?
- La réalisation de banques : quelle stratégie ?
- Le séquençage : pour quelle longueur des reads ? Pour quelle quantité de données ?
- L'analyse : quel pipeline ? Qui croire ? Comment annoter les résultats ?
- La qualité des ARN n'est pas choisie par l'opérateur
- La quantité est rarement un obstacle : 1 à 5 microgrammes sont en général faciles à obtenir.

La question de l'utilisation d'ARN Total ou de mRNA se pose. Avec l'ARN Total, il faut enlever l'ARN ribosomique. Différents kits permettent d'effectuer cette tâche (RIbozero, DSN).

25 à 45% des reads d'ARN Total sont des reads introniques. Le signal est dilué. Cette méthode nécessite un traitement à l'ADNase.

La capture mRNA génère un risque de biais 3' important. La réalisation de banques orientées est toujours possible.

L'intérêt du séquençage dirigé est de mettre en évidence des ARN anti-sens.

Le pipeline d'analyse est le suivant :

- Alignement des séquences sur le génome de référence et détection des régions de splices
- Assemblage, guidé par REFSEQ, des séquences en transcrits et estimation de leur abondance
- Annotation des transcrits assemblés par comparaison aux transcrits REFSEC
- Fusion de plusieurs assemblages
- Détection de changements significatifs d'expression dans les transcrits, les régions de splices et promoteurs.

Le RNA-seq permet la détection de gènes et transcrits connus et nouveaux, l'annotation des gènes et des transcrits, la quantification des gènes et transcrits, la détection de SNP. RNA-seq gère les transcrits de fusion, l'expression différentielle et la visualisation via IGV.

Le RNA-seq permet la quantification des transcrits (10 millions de reads ?), la détection, la découverte et la quantification de variants d'épissage jusqu'à 100 millions de reads, la

détection de transcrits de fusion (de 50 à 100 millions de reads). La question de la détection de SNP se pose, notamment s'agissant de la comparaison aux données de séquence ADN (exome). Un grand nombre de SNP sont situés hors des gènes connus, avec des misalignements ou des séquences parasites. La question des miRNA se pose également.

Thierry Frebourg souligne un important gâchis dans les analyses faites au niveau de l'ARN en raison de la très mauvaise qualité des prélèvements. La qualité des ARN est cruciale. Pour obtenir de l'ARN de grande qualité au niveau constitutionnel, il est important d'investir dans la collection de prélèvements.

Synthèse et Perspectives

Thierry Frebourg.

Thierry Frebourg remercie tous les participants.

Le premier objectif de la réunion de Deauville était de permettre un échange et une évaluation des plateformes NGS : Roche, Illumina, Ion Torrent. Certaines limites ont été pointées sur la plateforme Roche dès lors que l'on augmente la puissance exigée. Plusieurs plateformes européennes confirment cette appréciation. Au sein du Cancéropôle Nord-Ouest, les plateformes Illumina et Ion Torrent sont très appréciées. D'autres plateformes restent à évaluer : Miseq et PGM. Un état de veille technologique doit être maintenu.

On peut estimer que l'implémentation du NGS pour le diagnostic des formes héréditaires de cancer est réalisée, notamment pour le cancer du sein et de l'ovaire. La plateforme de Caen présente la plus grande richesse en termes de nombre de familles analysées. L'équipe de Rouen est très satisfaite des nouvelles technologies présentées. L'équipe de Lille est en cours de basculement vers ces technologies pour l'étude du cancer colorectal. L'interprétation des nouvelles mutations reste à faire.

Il est satisfaisant de constater que des outils bioinformatiques compatibles avec l'exigence diagnostique sont développés. Des filtres de qualité sont mis en place. La mise en place d'un outil commun aux différents sites pourrait s'avérer intéressante.

Le développement et l'évaluation du NGS pour la détection des altérations somatiques, des marqueurs pronostiques ou prédictifs sont réalisés. En 2013, il appartient aux laboratoires du Cancéropôle Nord-Ouest de passer en routine toutes ces technologies, en particulier dans le cadre du ciblage génétique des thérapeutiques ciblées, avec une certification diagnostique.

En termes de recherche, les exomes étaient déjà opérationnels dans leur interprétation il y a deux ou trois ans. Désormais l'analyse des exomes est réalisée au Cancéropôle Nord-Ouest. L'utilisation du logiciel EVA a permis un important développement et le traitement d'exomes comparatifs. La meilleure façon de répondre à la complexité de la variabilité génétique est de soustraire et de comparer. Il convient de ne pas faire abstraction du problème éthique de consentement. Le problème éthique est soulevé par la loi, en cours de réécriture : que faut-il faire des mutations secondaires ? L'exome comparatif permet de restreindre des variations dont on ne sait que faire. L'interprétation biologique reste à implémenter.

Il est probable que l'évolution technologique aboutisse à une évolution de la chimiothérapie avec la multiplication des thérapies ciblées par guidage génétique. Des exomes globaux seront réalisés chez des patients atteints de cancer au-delà de ressources thérapeutiques connus. L'identification des variations devrait permettre l'accès à des molécules.

Le grand défi est l'interprétation des données. Il conviendra sans doute d'aller plus loin dans l'utilisation des outils bioinformatiques. L'interprétation passera toujours par la confrontation avec les bases de grade clinique. Les bases de données cliniques seront essentielles pour interpréter les variations. L'évolution des connaissances en cancérologie permet de montrer que la mutabilité du génome est telle que l'on découvrira probablement de plus en plus de mutations privées délétères. Ces mutations nécessiteront une interprétation. L'interprétation de nouvelles mutations est un défi crucial pour les altérations constitutionnelles et somatiques. L'extension biologique est le grand défi des NGS. En 2013, le Cancéropôle Nord-Ouest devra réfléchir à une structure permettant de classer rapidement l'impact biologique des mutations.