

Séquençage génétique : l'indispensable collaboration DSI/bio-informatique

Afin d'accompagner le développement de la médecine génomique et les programmes de santé publique visant à intégrer le séquençage dans les parcours de soins, les laboratoires hospitaliers ont commencé à s'équiper de séquenceurs à haut débit et à recruter des bio-informaticiens. Côté DSI, cette nouvelle activité constitue un axe de développement et d'innovation qui doit être construit en amont, dès la conception d'un projet d'implémentation dans l'établissement.

Depuis une dizaine d'années, la médecine génomique s'est construite autour d'évolutions technologiques vertigineuses, en particulier autour du séquençage de l'ADN. Il devient ainsi possible d'obtenir, pour un grand nombre de patients, un profil de plus en plus exhaustif de leurs gènes. On peut d'ores et déjà envisager le séquençage complet de l'intégralité du génome d'un patient. Ces nouvelles techniques regroupées sous le terme de séquençage à haut débit (ou HTS pour *high-throughput sequencing*, ou encore NGS pour *next-generation sequencing*) définissent un ensemble de méthodes apparues à partir de 2005 et capables de séquencer massivement l'ADN à faible coût. Elles se caractérisent par l'utilisation d'approches hautement parallélisées qui ont permis de diviser le coût du séquençage d'un génome par 1 000 en dix ans.

Convergence des TIC et des sciences de la vie

En France, une première phase de déploiement de ces techniques dans les laboratoires hospitaliers a débuté dès 2012 sous l'impulsion du plan national Maladies rares 2 et du plan Cancer 3. Ces initiatives ont permis aux laboratoires hospitaliers de s'équiper de séquenceurs à haut débit et de recruter des bio-informaticiens. Le transfert de l'activité des laboratoires vers le NGS nécessite toutefois de profondes

réorganisations dans les établissements. Elles concernent les laboratoires, mais également leur environnement dans les hôpitaux, en particulier les aspects informatiques nécessaires pour stocker, analyser et interpréter les grandes quantités de données générées. Ainsi, les sciences et technologies de l'information et de la communication sont appelées à converger avec les sciences de la vie et de la santé.

Le nouveau métier de bio-informaticien médical

Différents chaînons du parcours diagnostique sont appelés à évoluer par l'acquisition de nouvelles compétences relevant de plusieurs champs disciplinaires allant de la génétique à l'informatique. Ainsi, l'intégration de nouveaux métiers dans les centres hospitaliers, tel que celui de bio-informaticien, s'avère essentielle pour relever le défi de la génération, de l'exploitation et de l'interprétation des données de NGS. Les bio-informaticiens garantissent la maîtrise et la compréhension des processus d'analyse du NGS et sont des interlocuteurs privilégiés, à l'interface entre les généticiens, les biologistes moléculaires et les informaticiens des hôpitaux.

L'analyse des données de séquençage dans un but diagnostique vise à reconstruire des séquences d'ADN à partir des données brutes produites par les machines de séquençage, à visualiser et à interpréter ces informations. Ainsi, les bio-informaticiens médicaux développent et implémentent des suites d'outils sous la forme de pipelines d'analyse afin de traiter les données. Ils constituent également des

bases de données biologiques afin d'intégrer ces informations dans le flux diagnostique. Il devient donc nécessaire, au sein des hôpitaux, d'associer à la production massive de données de NGS un dimensionnement des structures d'analyse et de calcul *ad hoc*.

Une bonne définition des besoins

La DSI est garante des infrastructures informatiques hospitalières mises à la disposition des utilisateurs. Ces moyens incluent un réseau informatique qui permet la circulation efficace et la sécurisation des flux de données, la puissance de calcul nécessaire, le stockage et l'archivage des données.

Les capacités et les besoins actuels des analyses bio-informatiques liées au NGS vont souvent bien au-delà des modèles usuels déployés dans les DSI hospitalières. Ainsi cette activité constitue-t-elle un axe de développement et d'innovation pour les DSI, qui doit être construit en amont, dès la conception d'un projet d'implémentation du NGS dans l'établissement. Une bonne définition des besoins permettra un dimensionnement optimal et une évolution aisée du système. Elle permettra aussi de définir à la fois le mode d'organisation le plus adapté pour l'établissement et le choix entre une installation locale ou le recours à un hébergeur de données de santé externalisé.

Deux environnements distincts

Par ailleurs, le développement informatique dans un système de soins nécessite

la mise en place par la DSI de deux environnements distincts. D'une part, un environnement de développement, d'intégration, afin d'installer et d'évaluer de nouvelles solutions, de modifier des paramètres, de faire la mise à jour du système et de tester la charge informatique des protocoles. D'autre part, un environnement de production qui permet de réaliser les analyses de routine en garantissant la stabilité du traitement à partir des versions robustes et éprouvées des protocoles bio-informatiques. Le passage d'une nouvelle version de l'environnement d'intégration à celui de la production des protocoles se fait par versioning, après validation systématique par des tests de non-régression réalisés en amont.

L'exemple du centre François-Baclesse à Caen

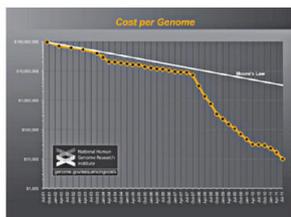
Ainsi, l'implémentation du NGS au sein du centre François-Baclesse de Caen s'est faite en étroite collaboration avec les équipes impliquées dès 2011. « Cette coopération a été très importante au départ pour le dimensionnement du système informatique et l'installation logicielle. Par la suite, nos relations se sont étendues à certaines tâches avancées d'administration conjointes, comme l'archivage des données sur bandes. Au quotidien, les contacts sont plus occasionnels puisque nous sommes indépendants dans le lancement des analyses, l'administration système et le développement de nos outils sur l'environnement physique dédié et distinct de ceux du reste de l'hôpital », rapporte Germain Paimparay, bio-informaticien au centre François-Baclesse.

L'analyse bio-informatique est très consommatrice en ressources système et produit de nombreuses données qui s'accumulent avec le temps. Les activités principales du laboratoire de génétique du cancer du centre François-Baclesse sont réalisées dans le cadre du diagnostic

moléculaire des prédispositions au cancer du sein ou de l'ovaire et du séquençage de tumeurs dans le contexte des thérapies ciblées. Une équipe de trois bio-informaticiens y assure l'analyse de deux à trois séries de séquençage hebdomadaire en plus des activités de recherche et développement – soit plus de 8 000 patients en génétique constitutionnelle depuis la mise en place de cette activité en 2012. « Les données "chaudes" générées sont temporaires, mais une dizaine d'analyses peuvent cohabiter et utilisent 2 To de stockage en permanence. Les données "froides" s'accumulent au fil des semaines et représentent environ 2 To par an sur la baie de stockage et 10 To sur bandes, pour une analyse aujourd'hui considérée de moyen débit pour 40 patients par semaine. Le type des données peut être très différent : par exemple, les données brutes sont constituées de très nombreux petits fichiers sollicités une seule fois au début de l'analyse. Les données de référence représentent de très gros fichiers lus et mis en mémoire cache pour chaque patient. Les fichiers générés lors de l'analyse sont de taille modeste pour finir sur une mise en base de milliers de mutations représentant 15 Mo par patient », explique Germain Paimparay.

Les recommandations des praticiens de génétique moléculaire

Il est donc important que le service informatique, en collaboration avec les bio-informaticiens et les laboratoires de génétique, mette en place une architecture informatique et des volumes de stockage appropriés. À cette fin, l'ANPGM (Association nationale des praticiens de génétique moléculaire) a élaboré des recommandations pour la gestion des données de séquençage à haut débit à l'intention des laboratoires. « Ce document disponible



Évolution du prix du séquençage

Source: Wikimedia Commons

auprès de l'ANPGM a été conçu pour aider les laboratoires de génétique et les DSI à évaluer correctement les besoins en ressources informatiques lors de la mise en place des technologies de séquençage. Il contient également un certain nombre de recommandations techniques pour la mise en œuvre de cette technologie », souligne Jean Muller, responsable du réseau NGS Diagnostic au sein de l'ANPGM.

Et maintenant, l'avenir...

Par ces premiers pas, de nombreux hôpitaux se sont dotés des moyens de proposer une meilleure prise en charge des patients à travers le séquençage haut débit. Au-delà de la technologie elle-même, c'est bien la mise en commun des spécialités et des compétences de tous les acteurs impliqués dans ce parcours de soins qui pousse à concrétiser ces innovations. Et ce sont ces interactions qui permettront de relever les défis de demain.

En effet, un changement d'échelle est nécessaire pour intégrer un nombre toujours plus important de patients tout en réalisant des analyses de plus en plus larges et sensibles. C'est dans ce cadre que le plan France Médecine génomique 2025 a été lancé par le ministère chargé de la Santé, de même qu'un appel à projets national pour la mise en place du séquençage à très haut débit à visée sanitaire.

Une toute nouvelle organisation du parcours de soins est indispensable, le challenge consistant alors à identifier et à lever les verrous scientifiques, technologiques, cliniques et réglementaires. Particulièrement, le lien entre la bio-informatique et la DSI devra imaginer les environnements et les infrastructures permettant le traitement des données et leur stockage à travers une vision « big data médical ».

■ Étienne Lonchamp

WWW.DSIH.FR | OCTOBRE 2017 79

Une équipe NGS

L'implémentation de tels systèmes sur des données de santé nécessite parfois de changer les habitudes. « Les nombreux échanges entre la DSI et l'équipe bio-informatique nous ont conduits à redéfinir nos modes de travail, en créant notamment une équipe NGS composée à la fois d'informaticiens et de bio-informaticiens afin de fluidifier la communication et de créer une infrastructure adaptée. Nous avons ainsi pu mettre en place une technologie innovante issue de la recherche, en y intégrant les processus de maîtrise des risques indispensables pour garantir des résultats diagnostiques précis et contrôlés », rapporte Antony Le Béhec, responsable de l'équipe bio-informatique du CHU de Strasbourg, Institut Régional du Cancer, Centre Paul Strauss.