Le Centre national de Génotypage du CEA et la Société Biofacet annoncent le démarrage de la seconde phase d'un développement logiciel de traitement de données de séquençage à haut débit (NGS : Next Generation Sequencing). Faisant suite à la validation d'une étude pilote menée en 2015 sur des exomes*, cette seconde phase vise à produire une plateforme logicielle de stockage, d'organisation et d'interrogation de variants de séquences issus de données NGS sur des génomes complets.

Si les programmes de séquençage se développent de manière soutenue, force est de constater que la masse des données accumulée reste difficilement exploitable pour la communauté scientifique. L'explosion de cette production, couplée à la spécificité du domaine, empêche les technologies classiques de base de données d'opérer de manière efficace. Paradoxalement, la masse de connaissance accumulée sur les myriades de projets nationaux ou internationaux génère le propre frein à son exploitation. Ainsi, il n'existe pas de système capable d'exploiter finement et/ou de croiser massivement des données de variants** de séquences. Repérer les variants d'un génome, recouper à grande échelle des séquences pour déceler des profils de variants communs constitue un outil de recherche puissant, ainsi qu'une aide au diagnostic et à la prise en charge des patients.

Face aux limites vite atteintes des systèmes actuels, le CEA et Biofacet ont développé les spécifications d'un système de gestion de base de données capable de stocker et d'interroger à très grande échelle des banques de données nationales ou internationales de variants provenant de séquençage d'exomes ou de génomes complets (WGS : Whole genome Sequencing). La technologie résultante, implantée dans le logiciel Biofacet™, permet d'agréger et d'interroger des études sur des milliers d'échantillons. Par un couplage optimisé entre données numériques et phénotypiques, la technologie développée permet plus précisément:

- l'interrogation « profonde » des données, i.e. la possibilité d'établir des requêtes sur la totalité des valeurs produites par les SNP-callers et ce pour chaque position du génome (3 milliards de nucléotides pour le génome humain) ;
- l'interrogation mixant les génotypes et les phénotypes, permettant ainsi de mieux cibler les causes ou effets potentiels des variants de séquences liés à des maladies ;
- l'ajout incrémental d'échantillons.

Pour Jean-François Deleuze, directeur du CNG : « Le développement de la médecine de précision ne pourra se concrétiser sans des outils capables d'analyser rigoureusement les données de variants à très grande échelle. De par sa couverture massive, le séquençage Whole Genome introduit une rupture technologique, non couverte par les outils classiques du « Big Data ». Produisant et analysant ces données en routine au CNG depuis des années, nous connaissons la difficulté de gestion de ces données. Nous sommes heureux de collaborer avec une start-up française dans ce domaine à haute valeur ajoutée ».

Jean-Jacques Codani, Président de Biofacet SAS : « Bien que le logiciel Biofacet™ ait déjà obtenu une certification CLIA outre-Atlantique pour le diagnostic clinique, le défi posé par le CNG est tout autre. Nous l'avons relevé parce que nous avons trouvé au CNG les deux composantes indispensables au déploiement d'une telle technologie : tout d'abord un savoir-faire en production de données NGS et des compétences scientifiques indiscutables, ainsi qu'un environnement et une expertise technique d'excellence pour le calcul haute performance ».

L'application pilote ayant été déployée avec succès au CEA, les partenaires envisagent une montée en puissance progressive pour le traitement des milliers de génomes WGS en production au CNG. Ce faisant, ils valideront un composant logiciel apte à répondre aux défis d'analyse posés par l'avènement de la médecine génomique, et plus généralement par l'étude des variations génétiques d'intérêt pour l'industrie du vivant.

* Les exomes regroupent l'ensemble des gènes fonctionnels de l'organisme (1,5 % du génome)
** Les variants sont des mutations pouvant contribuer au développement d'une pathologie

À propos du CEA

Le CEA est un organisme public de recherche qui intervient dans quatre domaines : la défense et la sécurité, les énergies nucléaire et renouvelables, la recherche technologique pour l'industrie et la recherche fondamentale.

S'appuyant sur une capacité d'expertise reconnue, le CEA participe à la mise en place de projets de collaboration avec de nombreux partenaires académiques et industriels. Fort de ses 16 000 chercheurs et collaborateurs, il est un acteur majeur de l'espace européen de la recherche et exerce une présence croissante à l'international.

En savoir plus : www.cea.fr

À propos de Biofacet

Biofacet SAS est une société d'édition de logiciels spécialisée en Bioinformatique.

Biofacet SAS développe et commercialise le logiciel Biofacet™, cœur technologique mis au point par ses fondateurs depuis 1998, au sein de la société GQ Life Sciences Inc.

Disposant d'une algorithmique conçue pour résoudre les problèmes posés par l'analyse de données génomiques à grande échelle, Biofacet accompagne ses clients par sa longue expertise et son expérience du marché international.

En savoir plus : http://www.biofacet.com/