

Conférence X-Biotech du 6 décembre 2012 aux Mines de Paris :

Biologie et « Big Data » : Comprendre la complexité du vivant

Nous vivons actuellement une époque charnière, une transition entre la médecine classique et la médecine personnalisée. D'un côté, une pression forte au changement : la nécessité de maîtriser l'explosion des coûts de la santé et d'optimiser le rapport bénéfice-risque-coût ; de l'autre, deux révolutions technologiques majeures : l'avènement de la génomique haut débit et la consolidation des données de santé.

L'avènement de la génomique haut débit, qui multiplie par un facteur 10^9 le nombre de données disponibles pour chaque patient, porte la promesse d'un diagnostic et d'un choix de traitement personnalisé exploitant des informations du domaine de la génétique et de la biologie moléculaire d'une richesse extraordinaire. Ces technologies ne cessent d'étendre leur champ : après le génotypage, l'arrivée des RNA assays permettant l'étude du transcriptome, voici maintenant la perspective du génome à 1000 €, le développement de la protéomique haut débit, l'étude de l'épigénome, du microbiome, du métabolome, etc.

En parallèle un mouvement de fond est en cours : la consolidation des données de santé, porté par la dématérialisation, le développement de la télémédecine, l'harmonisation des standards et de l'interopérabilité, et surtout la volonté de plus en plus partagée de lever les obstacles juridiques et pratiques à la mutualisation des données. Cette consolidation va permettre un meilleur suivi longitudinal des données de santé, dans l'objectif d'améliorer la connaissance de l'impact des différents facteurs, et d'optimiser la prise en charge des patients dans une démarche intégrative, de la prévention au suivi long terme : c'est le disease management.

Le frein majeur au progrès dans la recherche médicale ne sera donc bientôt plus le manque de données à analyser, mais bien plutôt le trop-plein, et le manque de capacités d'analyse à la mesure de cette quantité gigantesque de données complexes. L'enjeu stratégique devient donc la capacité à identifier les causes des phénomènes et à déchiffrer la complexité biologique sous-jacente. Les outils statistiques (modèles de régression, analyse discriminante...) sont en effet très insuffisants dans leur capacité d'explication des phénomènes, et les algorithmes d'apprentissage classiques (arbres de décisions, réseaux de neurones, etc.) ont également de nombreuses limites.

Pour surmonter ces limites, des start-ups et laboratoires académiques développent des approches innovantes et différentes. Parmi elles, le logiciel HyperCube, inventé par A. Huret (X90), et fondé sur des mathématiques non-euclidiennes et non-statistiques, qui trouve des applications dans de nombreux domaines, notamment dans la santé.

Yann Gaston-Mathé (X90), Directeur Life Sciences chez HyperCube, travaille sur ces sujets complexes et passionnants. Il brossera un tableau de l'état de la Science dans le domaine et nous donnera quelques exemples concrets.