

RECHERCHE SUR LE CANCER : mégadonnées et santé de précision

Le présent Bulletin du PPP explique le rôle des mégadonnées pour aider à faciliter la santé de précision.

Principaux concepts

- Mégadonnées
- Cycle « des données aux connaissances à l'action »
- Défis liés aux mégadonnées
- Principales initiatives

Bulletins du PPP connexes

- Essais cliniques : médecine de précision et essais cliniques
- Recherche sur le cancer : financement de la recherche sur le cancer

La recherche sur le cancer a toujours été fondée sur les données, mais dans les deux dernières décennies, une surabondance de nouvelles méthodes de collecte, de stockage et d'analyse de données a complètement changé la donne et permis aux chercheurs d'améliorer grandement leur compréhension à la fois de la biologie moléculaire du cancer et des particularités de chaque personne qui contracte cette maladie.

Le terme « mégadonnées » fait référence non seulement à la quantité accrue de données auxquelles les chercheurs ont désormais accès, mais également aux façons dont les énormes ensembles de données peuvent être combinés afin de révéler des liens inattendus, des facteurs de risque cachés et des interventions adaptées à des sous-groupes et même à des personnes en particulier.

Seules, les mégadonnées ne sont pas en fait très utiles. Pour les mettre à profit, il faut utiliser des méthodes d'analyse perfectionnées qui dépassent souvent les capacités humaines et nécessitent le recours à l'intelligence artificielle pour extraire des renseignements. Ces renseignements favorisent l'émergence de nouvelles générations de « santé de précision » en transformant les données en connaissances et en actions (voir l'illustration à la page suivante).

Les mégadonnées permettent de « saisir les occasions et de relever les défis associés à l'obtention, à la gestion, à l'analyse et à l'intégration de l'information au sein de divers ensembles de données [...] À l'heure actuelle, ces ensembles de données sont au-delà des capacités des méthodes classiques de gestion de données¹. »

¹Tiré du site : <https://cancercontrol.cancer.gov/brp/priority-areas/big-data.html> (en anglais seulement).

Les mégadonnées sont caractérisées par² :

- le volume et le rythme auxquels elles sont produites;
- des sources et des formats de données multiples, puisqu'elles sont obtenues dans différents emplacements, à différents moments, à différents intervalles et à l'aide de différentes méthodes;
- l'inclusion de sources de données qui peuvent être incomplètes et inexactes;
- la possibilité d'utiliser les données à la fois pour avoir une « vue d'ensemble » de la recherche sur le cancer et pour examiner jusqu'au niveau granulaire chaque personne touchée.

Les chercheurs utilisent des algorithmes informatiques pour analyser ces vastes ensembles de données diverses. Par exemple, un chercheur pourrait vouloir prévoir les effets à long terme de traitements chimiothérapeutiques particuliers sur des patients cancéreux. Le bon algorithme pourrait traiter des millions de dossiers de santé électroniques — contenant chacun des centaines de points de données — pour déterminer comment les caractéristiques propres à chaque patient sont associées à des résultats particuliers.

Dans le contexte des soins de santé, les sources de mégadonnées peuvent comprendre :

- les cabinets de médecin;
- les hôpitaux;
- les pharmacies;
- les laboratoires et les centres d'analyses (y compris les résultats d'analyse génétique, d'imagerie, de pathologie et autres);
- les fournisseurs d'assurance maladie;
- les moniteurs, les détecteurs et les dispositifs portables, y compris les téléphones intelligents;
- les médias sociaux;
- les renseignements et les modèles de recherche sur Internet;
- les dossiers gouvernementaux (recensement, statistiques démographiques, registres de maladie, surveillance);
- les études de recherche, y compris les études de cohorte, les essais cliniques et les grandes études de séquençage.

²Willems, S. M. et coll. (2019). The potential use of big data in oncology. *Oral Oncology*, 98, 8-12.

DES DONNÉES AUX CONNAISSANCES À L'ACTION

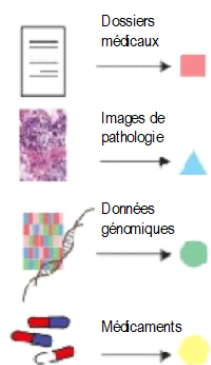
La médecine de précision utilise des algorithmes d'apprentissage automatique pour intégrer de nombreux fils de données et aider les médecins à prédire ce qui arrivera au patient et à choisir le meilleur traitement qui soit.

DONNÉES

Informatiser les données cliniques

Distiller les données

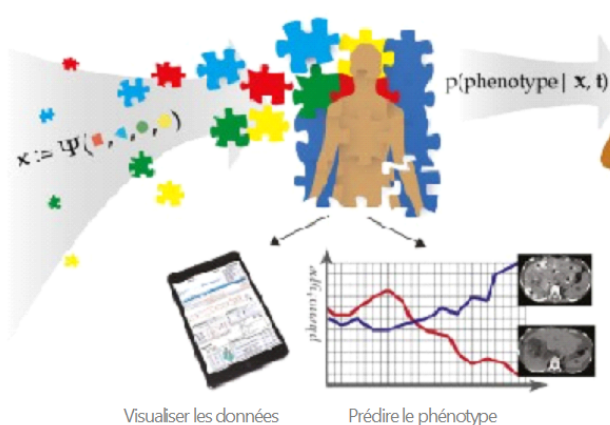
Intégrer les données



CONNAISSANCES

Prédire les phénotypes cliniques

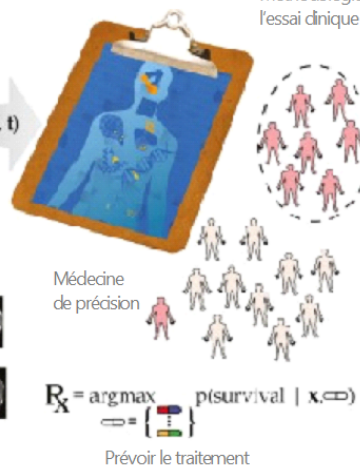
Relier les données



ACTION

Choisir le traitement optimal

Méthodologie de l'essai clinique



Source : Andre Kahles, Gunnar Ratsch, Chris Sander. Tiré de : <https://www.nature.com/articles/d42473-019-00035-5>

S'il peut être difficile d'extraire de l'information significative et pertinente à partir de ce mélange varié de sources de données, pour y arriver, les chercheurs ont accès à des outils en constante amélioration.

Les mégadonnées ont le potentiel de permettre :

- la détection précoce d'un cancer, et donc une intervention précoce;
- l'évaluation des facteurs de risque uniques à chaque sous-groupe et à chaque personne;
- la personnalisation du traitement en fonction du génome, des antécédents médicaux et des autres facteurs personnels de chaque patient;
- la découverte de liens inattendus entre des facteurs génétiques, environnementaux, socioéconomiques et autres qui ont une incidence sur le risque et l'apparition du cancer et les interventions requises pour prévenir et réduire les cas de cette maladie.

Défis liés aux mégadonnées

Normes de données

De nombreuses sources actuelles de données médicales ont été créées et stockées par certains établissements qui les ont organisées sans tenir compte de la façon dont elles pourraient « communiquer » avec d'autres sources de données.

La consolidation et l'harmonisation de ces ensembles de données demeurent l'un des plus grands défis de la communauté de la recherche sur le cancer.

L'élaboration d'un plus grand nombre de normes universelles de données requiert l'expertise de scientifiques de nombreuses disciplines, de même que de fournisseurs de soins de santé, d'experts en science des données et en informatique, d'ingénieurs en logiciels, d'éthiciens, de décideurs et de chercheurs de l'industrie³. Elle comporte également des défis logistiques comme l'élaboration de normes applicables au-delà des frontières territoriales.

Capacité de l'infrastructure

L'infrastructure des mégadonnées comprend les outils, les gens et les systèmes permettant d'assurer la collecte, le stockage, la sécurisation, la récupération, le partage et l'analyse des données. Au-delà des défis technologiques, il faut tenir compte de facteurs comme la confidentialité et la sécurité, la propriété intellectuelle, la légalité et les aspects éthiques.

L'utilisation de mégadonnées nécessite et stimule de nombreuses innovations technologiques. Les plateformes infonuagiques, les algorithmes de compression, l'indexation des données, les interfaces utilisateur et les outils de visualisation permettent tous de faciliter la gestion des mégadonnées, d'en améliorer l'accès et d'en augmenter l'utilité. Les ressources à code source libre dont la documentation est solide permettent d'en assurer la démocratisation de façon que tous puissent tirer profit des renseignements qu'elles contiennent⁴.

Gérance

Au Canada, l'« écosystème » des mégadonnées fait en sorte que les gouvernements fédéral et provinciaux doivent adopter des approches réfléchies pour élaborer des politiques unifiées concernant les données médicales et liées à la santé. Les gouvernements doivent maintenir la sécurité et la confidentialité des données médicales personnelles et s'assurer que les nouvelles recherches réduisent les disparités en santé au lieu de les amplifier⁵. Bien que ce type de recherche puisse sembler très technique, elle doit tout de même placer les gens au centre des politiques et des pratiques. L'autonomisation signifie qu'il faut donner aux participants les options et les outils nécessaires pour leur permettre de comprendre les risques et les avantages potentiels.

Principales initiatives

Comme les données deviennent de plus en plus complexes, le développement des capacités est essentiel. **Bioinformatics.ca** offre des ateliers de formation avancée sur la bio-informatique, la génomique, la protéomique et les technologies liées à la bio-informatique. Pour en savoir plus, consultez le site <https://bioinformatics.ca/> (en anglais seulement).



³Hulslen, T. et coll. (2019). From big data to precision medicine. *Frontiers in Medicine*, 6(34).
<https://www.frontiersin.org/articles/10.3389/fmed.2019.00034/full>

⁴Hinkson, I. V. et coll. (2017). A comprehensive infrastructure for big data in cancer research: accelerating cancer research and precision medicine. *Frontiers in Cell Development and Biology*, 5(83). <http://dx.doi.org/10.3389/fcell.2017.00083>.

⁵Vayena, E. et coll. (2018). Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, 96(1), 66-68.
<http://dx.doi.org/10.2471/BLT.17.197426>

Le **Cancer Genome Collaboratory** est un laboratoire infonuagique de recherche universitaire mis sur pied par l'Institut ontarien de recherche sur le cancer (IORC). Ce collaboratoire conserve les données du Consortium international sur la génomique du cancer. Les chercheurs peuvent utiliser les installations du collaboratoire pour effectuer des opérations d'exploration et d'analyse de données complexes au sein d'un vaste répertoire de séquences génomiques de cancer et des renseignements cliniques sur le donneur qui y est associé. Pour en savoir plus, consultez le site <https://www.cancercollaboratory.org/> (en anglais seulement).



Calcul Canada offre les services et l'infrastructure de calcul informatique de pointe indispensables aux chercheurs canadiens et à leurs collaborateurs dans les domaines universitaire et industriel. En collaboration avec des organismes régionaux, Calcul Canada aide les chercheurs à utiliser les mégadonnées pour favoriser l'accélération de la recherche et de l'innovation. Pour en savoir plus, consultez le site <https://www.computecanada.ca/?lang=fr/>.



La **Global Alliance for Genomics and Health (GA4GH)** est une alliance internationale sans but lucratif qui vise à accélérer les progrès en recherche et en médecine pour faire avancer la santé humaine. Réunissant plus de 500 organismes de soins de santé, de recherche, de défense des intérêts des patients, de sciences de la vie et de technologies de l'information provenant de plus de 90 pays, l'alliance GA4GH élabore des cadres de travail et des normes pour permettre le partage responsable, volontaire et sécurisé de données génomiques et médicales. Pour en savoir plus, consultez le site <https://www.ga4gh.org/> (en anglais seulement).



ICES est une société indépendante sans but lucratif qui utilise l'informatique de la santé pour mener des recherches sur les services de santé et les résultats de santé obtenus au sein de la population générale en Ontario. Son répertoire de données est composé de dossiers médicaux de patients anonymisés et fiables recueillis par le système public de santé provincial. Par des partenariats, ce répertoire permet également de relier en toute sécurité les données provenant de divers registres et enquêtes de santé. En combinant les différents ensembles de données à l'aide d'identifiants numériques uniques anonymes, les scientifiques d'ICES peuvent suivre, au fil du temps, différents aspects de l'utilisation des services de santé et des résultats des patients dans toute la province. Pour en savoir plus, consultez le site <https://www.ices.on.ca/> (en anglais seulement). [Il est à noter que les autres provinces ont également des données liées et des procédés pour avoir accès aux données de santé à l'échelle de la population.]



Pour en savoir plus sur les mégadonnées, visionnez les courtes vidéos suivantes :

- Aviesan. *Vivre la révolution de la médecine génomique*. (YouTube) 20 janvier 2020 [3 :41 minutes]
<https://www.youtube.com/watch?v=rhcWJb1ei10>
- Bayer France. *La médecine de précision : l'interview du Pr Jean-Yves Blay, Directeur du CLB*. (YouTube) 25 sept 2019 [3 :18 minutes] <https://www.youtube.com/watch?v=PC5lhyrni9U>
- Inserm. *La notion de réseaux en biologie, biologie des systèmes, OMICS*. (YouTube) 19 décembre 2017 [5 :41 minutes]
<https://www.youtube.com/watch?v=gVb6NCV9bNk>
- Dalhousie FCS. *Big Data Institute*. (YouTube) 30 mars 2016 [4:16 minutes]
https://www.youtube.com/watch?time_continue=96&v=ibtR1dHkebM (en anglais seulement)
- École Polytechnique. *How data science can help doctors fight cancer*. (YouTube) 29 juin 2018 [2:18 minutes]
<https://www.youtube.com/watch?v=TRkKnnYgfkS> (en anglais seulement)
- GA4GH. *GA4GH: The Global Standards Organization for Genomics*. (YouTube) 2019 [2:09 minutes]
https://www.youtube.com/channel/UCmCg1AcAY_qHXfOhePAFAeQ (en anglais seulement)
- HuffPost. *How Big Data Could Transform the Health Care Industry*. (YouTube) 2 février 2017 [3:48 minutes]
https://www.youtube.com/watch?v=_mXrZEIpNMw (en anglais seulement)
- ICES Ontario. *Institute for Clinical Evaluative Sciences (ICES): The power of data to improve health*. (YouTube) 5 octobre 2016 [2:39 minutes] <https://www.youtube.com/watch?v=5pZhRSM1cyk&feature=youtu.be> (en anglais seulement)
- UHN Toronto. *Lillian Siu talks 'big data' in cancer prevention and care*. (YouTube) 1er septembre 2016 [2:54 minutes]
https://www.youtube.com/watch?v=1JhSGPQE3_0 (en anglais seulement)