Analyse de « Genomics and big data in biomedicine »

Kilian Debraux

# Thème

L’utilisation massive de données génétiques pour améliorer la rapidité et la véracité des diagnostiques.

# Problématique

Comment mettre en place un contexte efficace permettant l’utilisation de données génétiques pour améliorer la rapidité et la véracité des diagnostiques tout en respectant la déontologie et l’éthique inhérentes à ce projet ?

# Thèse

L’auteur présente plusieurs points clés dans la conception d’un tel contexte :

* Il faut d’abord une meilleure *compréhension* de la génétique et de son lien avec les autres phénomènes biologiques afin d’éviter que cette masse de données n’entraîne de fausses interprétations
* Ensuite, pour éviter de mauvais diagnostiques, il faut des données *complètes* et *justes*. Pour cela, l’auteur préconise une infrastructure solide et fiable qui permet de stocker l’entièreté des données et de multiplier les sources d’information.
* De plus, l’auteur met l’accent sur la *circulation* et le *partage* des informations et des connaissances. C’est une étape essentielle dans la quête de données complètes et fiables en réutilisant des connaissances préalablement acquises pour compléter ses propres données
* Enfin, l’auteur relève les limites éthiques de ce contexte comme la *vente* ou le *piratage* de potentiels informations sensibles

# Concepts

**Big Data**: Le big data est un domaine de l’analyse de données concernant un volume massif de données. Généralement, un ensemble de données est considéré comme du big data lorsque cet ensemble contient un très haut *volume* de données, que ces données sont sont intégrées et mises à jour avec une importante *vélocité* et que ces données sont *variées*. Ces 3V sont souvent les critères qui permettent de définir le domaine du Big Data.

**Base de données**: Une base de données est un outil permettant le stockage de données sous de multiples formes. Les bases de données peuvent être vues comme des entrepôts dans lesquelles on peut stocker des informations, puis les récupérer en tant voulu lors d’analyses. Les bases de données modernes permettent aussi de faire automatiquement des liens entre les données, ce qui peut être utile dans le cas d’analyses médicales.

**Médecine de précision** : « La médecine de précision est une nouvelle discipline médicale qui utilise les données biologiques, environnementales ou sociales spécifiques au patient pour mieux connaître la probabilité de développer des maladies. Elle vise à proposer des mesures préventives ciblées, affiner les diagnostics et personnaliser les traitements. ». Source : https://www.chuv.ch/fr/medecine-precision/accueil/patients-et-familles/medecine-de-precision

**Génétique**: La génétique est un domaine de la biologie visant l’étude des gènes et des phénomènes adjacents (ADN, Phénotypes...).

**Génome**: On désigne par le terme de génome l’ensemble de l’information génétique codée par l’ADN (ou l’ARN chez les virus à ARN) ([https://fac.umc.edu.dz/snv/faculte/tc/2020/COURS%20GENETIQUE%20PDF.pdf](https://fac.umc.edu.dz/snv/faculte/tc/2020/COURS GENETIQUE PDF.pdf) p.14).

# Arguments

* Aujourd’hui, les données médicales et génétiques sont récupérées depuis une multitude de sources comme des services « direct-to-customers », c-à-d des applications où le patient entre ses données lui même, des études cliniques ou bien des analyses médicales d’hôpitaux ou de médecins « Big health data include a vast variety of data types […] such as hospitals and individual GP practice » (p.25). Ces différentes sources fournissent des données dans des formats différents, n’ont pas le même niveau de rigueur et peuvent donc renseigner des informations contradictoires. Il n’est pas possible d’avoir des données *complètes* et de les *partager* avec une telle infrastructure.
* En oncologie, une infrastructure plus robuste a déjà été déployée sous le nom de « The Cancer Genome Atlas » qui permet d’alimenter une base de donnée appelée « Catalogue of Somatic Mutations in Cancer ». Cependant, il est difficile de créer une telle infrastructure pour des domaines moins étudiés et moins connus que l’oncologie.
* Dans un système de partage d’information, il y a un risque de fuite d’information par le *piratage* ou la *vente* d’informations, qui peut se produire dans des projets ayant pour motivation première le profit et ayant un rapport controversé à l’éthique comme avec l’entreprise Human Logetivity Inc. : « data can be mixed, leaked and disseminated in sometimes irrelevant, controversial and inappropriate ways, which is particularly concerning in the case of genomic data featuring in ethically controversial projects (Regalado 2015) ». C’est pourquoi, d’après l’auteur, il faudrait que l’organisme dirigeant la création d’un tel projet n’ait comme objectif que l’amélioration de la santé publique.

# Discussion

Ce texte présente un point de vue intéressant et des limites pertinentes mais qui peuvent être triviales comme le risque de piratage des informations.

Il est pertinent de mettre en avant la difficulté de construire une infrastructure de récupération de données en présentant les contraintes auxquelles cette infrastructure serait confrontée. Il est aussi intéressant d’avoir vu qu’un tel projet est néanmoins possible pour des domaines très étudiés comme l’oncologie.

Enfin, l’auteur présente les limites éthiques d’une telle infrastructure qui partagerait de l’information telles que le piratage ou la vente de données. Il présente aussi une solution : avoir une organisation à but non lucratif qui soit en charge de construire ce projet. Cependant, cette partie semble quelque peu superficielle et la solution proposée est légère. Premièrement, avoir une Organisation à but non lucratif à la tête de ce projet peut protéger un temps de la vente de données, mais ne protège pas plus du piratage qu’une entreprise privée. Deuxièmement, il y a des exemples précédents d’organisations à but non lucratifs finissant par vendre leur projet à des entreprises privées, ou bien à changer de statut pour se tourner vers le profit comme OpenAI (https://techcrunch.com/2019/03/11/openai-shifts-from-nonprofit-to-capped-profit-to-attract-capital/?guccounter=1&guce\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\_referrer\_sig=AQAAAH9ABI2S7W1P3lIwc5Ur\_YTp8n1BgRJQzqWjfAfWAciARUvEHhiAb7t89Sist9qyC2NMDSsyVgSHuT3q-30ADXqdftaXnEQ\_NLhCwJWwqQlcItCmlv0VUxj1kYu1\_TOp-MTzz7MNgzm9t7-4FIxiQMdwcyKi7nQW97pc6OM0H8VE)