

## Compte rendu de la conférence de Mickael Guedj, 2 mai 2022

L'exposé de M. Guedj, *Vers une Médecine Computationnelle de Précision*, a porté plus sur les applications de la bioinformatique que sur cette nouvelle discipline comme telle, laquelle joue bien évidemment un rôle majeur dans cette « médecine de précision ».

MG fait remonter la discipline au début du 20<sup>e</sup> siècle avec le recours aux méthodes statistiques et les premières utilisations de tests statistiques  $\chi^2$  ; toutefois le vrai décollage se fait à l'occasion du projet *Human Genome*, dans les années 2000, dont il nous informe que le décodage complet, à 100%, vient d'être obtenu cette année, soit 20 ans de travail.

Il nous précise que la « santé » peut/doit se définir par un état d'équilibre entre les différents processus à l'œuvre dans l'organisme. Compte tenu de la complexité de cet état d'équilibre et des processus en interactions, il y a beaucoup de données à manipuler, dont celle du génome. Il faut souligner que seulement 2% du génome concernent les gènes et les protéines, les 98% restant étant qualifiés de « non codants », et leur fonction mal connue ; il est donc difficile de le décrire. Le principal problème est l'interprétation de ces données ; celle-ci nécessite beaucoup de travail de modélisation relevant des technologies dites *Big Data*, car un seul génome contient environ 3,4 Md de bases ATCG. Si on veut travailler sur une cohorte de 1.000 génomes, cela fait une très grosse base de données dont on sait depuis les années 1990 que les manipulations vont nécessiter une algorithmique ad hoc compte tenu des problèmes de complexité combinatoire à prendre en compte.

MG consacre une bonne partie de son exposé à présenter les grandes lignes des outils et méthodes actuellement disponibles : Data, algorithmique avec les *Computational Tool Box*, la modélisation, essentiellement statistique, avec les problèmes associés comme l'analyse dite « en grande dimension » qui reste un défi d'ordre mathématique en matière de « spatialisation » pour visualiser l'information ainsi collectée.

L'un des problèmes clé présentés par MG concerne l'analyse et la classification des interactions, d'où les problématiques de graphes de connaissances, de réseaux d'inférences, qui sont analysées avec les méthodes développées pour le *Deep Learning*, en signalant au passage certaines difficultés et/ou biais comme le sur-apprentissage. Toutes ces analyses sont faites grâce à des plates-formes intégrées – *All-in-one Computing platforms* – qui font partie de l'outillage standard actuellement disponibles.

La dernière partie de l'exposé de MG est consacrée aux applications comme l'analyse des cancers du sein, la rétro-analyse – *repurposed drugs* – qui consiste à repartir des gènes pour remonter aux maladies résultant des dysfonctionnements de ces gènes, par rétro-conception, et enfin l'analyse des patrimoines génétiques, pour aborder des problématiques plus globales comme celles du covid-19. Il ne traitera en détail que cette dernière application.

Dans sa conclusion, MG donnera 6 points d'amélioration imputables à ces nouveaux outils et méthodes : 1/ça va plus vite, 2/les coûts baissent, 3/la diversité moléculaire s'élargit, 4/les essais cliniques sont simplifiés, moins d'animaux de laboratoire, 5/moins de patients requis, et 6/meilleure efficacité globale. Sa dernière planche, ci-dessous, mérite une mention spéciale car elle illustre particulièrement bien les problématiques que nous souhaitons aborder dans notre colloque futur sur l'interdisciplinarité car derrière chacune des disciplines mentionnées

il y a des communautés dont il faut organiser les interactions, et pour cela trouver un langage commun.

### Questions/réponses

Plusieurs questions sont soulevées par l'auditoire in situ et à distance sur les thèmes :

- *Drug design*
- Interdisciplinarité et complexité
- Standardisation des données et des plates-formes pour faciliter les échanges, par analogie avec ce que fait le CERN depuis longtemps
- Modèles et *Big data* : comment calibrer tout cela en évitant les corrélations fallacieuses

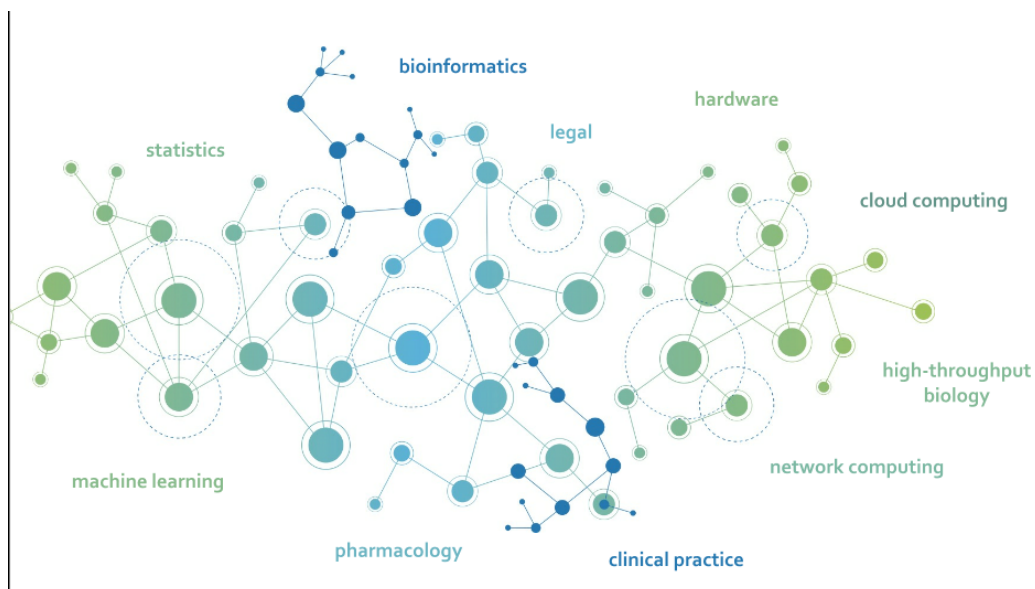
NB du rédacteur JPZ : sur ce dernier point, plusieurs interrogations sont soulevées concernant l'usage statistique des données en référence à un article de C.Anderson, dans la revue WIRED, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* dont je donne ici la référence précise :

[<https://www.wired.com/2008/06/pb-theory/?msclkid=f315ccfdcf6f11eca3f40678feaa0d8f>].

Cet article a eu et continue à avoir une audience importante dans le grand public peu informé de la méthode scientifique, bien que l'argument soit totalement faux ! il est donc important d'en avoir connaissance pour mieux le réfuter et expliquer la réalité de la modélisation.

### Dernière planche de MG :

Elle illustre particulièrement bien la nécessité de l'approche interdisciplinaire où les communautés en interactions doivent définir le langage commun – sujet considéré comme difficile – qui leur permet de se comprendre et d'agir de façon cohérente au sein des projets, un thème abordé dans une conférence précédente avec l'ingénierie système.



En complément, MG nous recommande son article, en coopération :

Philippe Moingeon, Melaine Kuenemann, Mickaël Guedj : *Artificial intelligence-enhanced drug design and development: Toward a computational precision medicine.*

Drug Discovery Today. 2022