



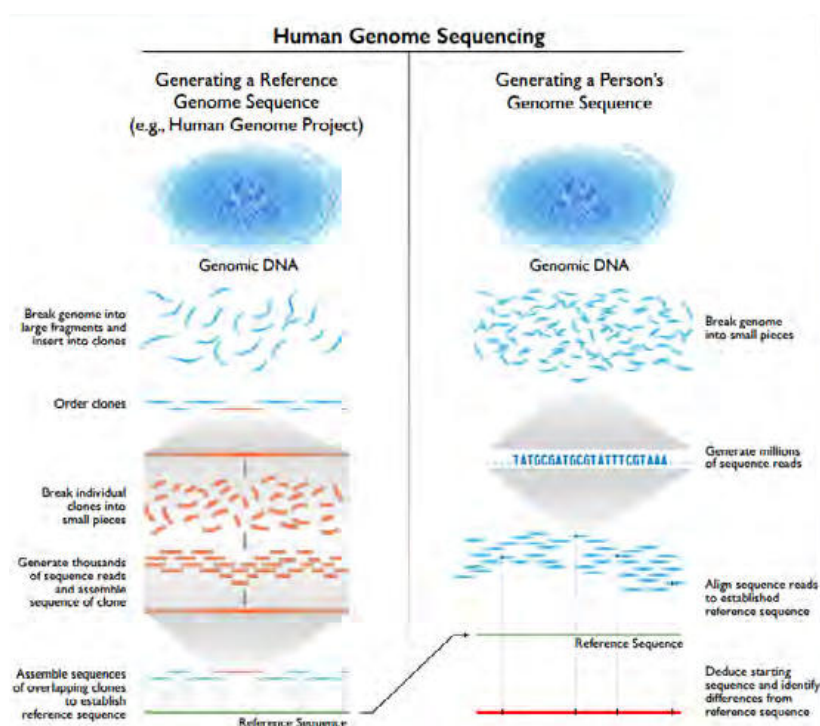
Pr Michel VIDAUD

→ PLAN FRANCE GÉNOMIQUE 2025 et Plateformes de séquençage à très haut débit (SeqIOA et AURAGEN)

PR MICHEL VIDAUD - SERVICE DE GÉNÉTIQUE ET BIOLOGIE MOLÉCULAIRES, HÔPITAL COCHIN, AP-HP
INSTITUT COCHIN - INSERM U1016 - UNIVERSITÉ PARIS DESCARTES

Bonjour,

Le principe fondateur de tout ce que je vais vous dire, est le séquençage du génome humain arrivé dans sa première version en 2001, sa deuxième version en 2004, ça a duré environ 13 ans et ça a été compliqué.



Quand j'ai commencé, je séquenciais péniblement un millier de paires de base et aujourd'hui, je séquence des milliards et des milliards de nucléotides, ce qui conduit à déplacer le problème vers le big data et vers l'interprétation de ces données.

Je rappellerai que grâce à tous ces progrès, on a pu mieux cerner ce qu'était un génome, on va faire les malins en séquençant tout le génome, mais il faut rappeler qu'on ne sait toujours interpréter que les gènes codant les protéines qui ne représentent à peu près que 19 901 (sur 58 000) dans la dernière version. Rappelons-nous pour le futur qu'il y a plus de gènes qui produisent des ARN et qui s'arrêtent là, que de gènes qui produisent des protéines mais la compréhension des variations de séquences dans ces ARN non codant est très complexe ; il y a également

des pseudo-gènes qui peuvent être très compliqués à interpréter.

Je vais juste vous rappeler une donnée importante, c'est la variabilité de ce génome qui est considérable puisque maintenant on est capable de faire des médianes de variation. Un papier est paru sur la variation du génome, on a étudié différentes ethnies et quand on compare un européen avec un autre européen, ils ont 3,5 millions de différences nucléotidiques, ils ont plus de 500 000 insertions-délétions de quelques nucléotides et ils ont plus d'un millier de délétions plus ou moins grandes dépassant les 50 nucléotides.

Ces variations de séquences, elles vont siéger dans nos fameux exons, parties qui vont persister dans l'ARN messager mature destiné à être traduit en protéine.

Depuis qu'on a séquencé le premier génome, séquencer de nouveaux génomes aujourd'hui est extrêmement simple puisqu'on a la copie d'origine. Il suffit de séquencer des petits morceaux d'ADN et les réappairier au génome d'origine pour les animaux.

C'est ce qui explique que pour le premier génome, on a mis 13 ans et que ça a coûté 2,7 milliards de dollars et qu'aujourd'hui, en 2018, ça va prendre quelques jours. Il y a même un Guinness des records qui est actuellement en cours pour savoir quel est celui qui va être capable de séquencer le plus rapidement le génome et l'interpréter. On en est à moins de 24 heures et ça peut descendre sur des tarifs de quelques milliers d'euros.

Ce changement de paradigme amène bien évidemment à un changement de prise en charge des patients.

Puisqu'il y en a 20 000, chaque fois qu'on séquencera tous les exomes d'un patient et qu'on le comparera au génome de référence, il y aura 20 000 différences et ces différences peuvent siéger dans les introns et dans toutes les séquences de régulation.

NOVASEQ 6000 Illumina



Donc les méthodes qu'on va utiliser pour faire ces grands projets génome sont de deux natures :

- Soit je séquence les 220 000 exons du génome, à peu près 50 millions de paires de base et je m'expose à avoir 20 000 variations de séquences
- Soit je séquence tout le génome et je vais donc séquencer ces fameux 3 milliards et je vais être exposé à 3, 4, 5 millions de différences qu'il va falloir ensuite interpréter.

Si vous souhaitez en savoir un peu plus, il y a un très bon papier, sorti il y a 15 jours dans le New England Journal of Medicine, qui fait un peu le point sur toutes ces approches de séquençage et qui conduit à se poser la question de la médecine génomique en France, fort de l'expérience internationale à l'époque où le plan avait été imaginé. Il y avait 14 initiatives nationales et la fameuse initiative de Génome England qui était certainement l'initiative européenne la plus avancée.

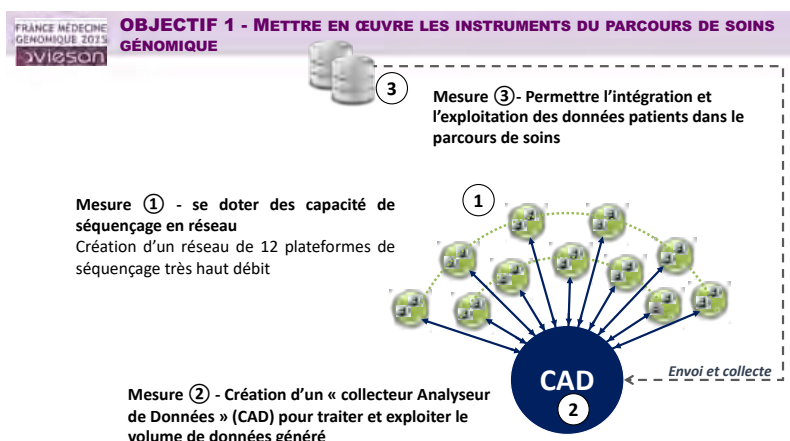
C'est Manuel Vals qui avait posé la question au Directeur Général de l'INSERM par l'intermédiaire de Marisol Touraine avec Thierry Mandon : « **En France, quelle organisation mettre en place pour introduire la médecine génomique dans le parcours de soin ?** ».

La caractéristique de ces 4 personnes étant qu'elles ne sont plus en position. Donc, lettre au premier ministre, du premier ministre à Yves Lévy qui réunit des experts qui ambitionnent un plan. Le plan est d'intégrer le séquençage dans un parcours de soin génomique et développer une

filrière nationale de la médecine génomique si possible industrielle, faire l'Airbus européen des génomiques. Ce plan existe, il est piloté par Franck LETHIMONNIER (Directeur ITMO Technologies pour la santé) puisque c'est AVIESAN qui pilote ce plan. Comme tout plan, il est assez complexe : 3 objectifs et 14 mesures ; il faut reconnaître que ces mesures ont été bien pensées parce qu'on y retrouve la façon de décliner le plan.

• 1^{er} objectif : Mettre en œuvre les instruments du parcours de soins génomiques.

- 1.1 C'est se doter de plateformes qui vont séquencer. L'objectif à terme, qui sera maintenu je l'espère par nos collègues des régions, est de créer 12 plateformes de séquençage à très haut débit.
- 1.2 Ensuite permettre l'intégration de ce parcours génomique dans le parcours de soin, et on verra que c'est probablement la mesure la plus pertinente puisqu'elle va obliger à faire des réunions de concertation pluridisciplinaires de médecins pour l'amont et l'aval.
- 1.3 Créer un objet pour l'instant mal identifié mais qui va probablement exister, un collecteur analyseur de données (CAD) qui serait capable de récupérer l'intégralité des données génomiques avec l'intégralité des données cliniques et qui serait amené ensuite à être étudié avec des mathématiciens, des statisticiens, des bio-informaticiens ad hoc, ... pour en tirer toute la substance intellectuelle.



- 2^{ème} objectif : Assurer la mise en œuvre opérationnelle et la montée en puissance. C'est le fameux CAD, je n'insisterai pas sur ce point.
- 3^{ème} objectif : Mettre en œuvre les outils de suivi et de pilotage. Une filière industrielle pourrait éventuellement se greffer là-dessus parce qu'il faut un peu d'argent.

LES 30 ANS

On va maintenant descendre au plus proche de ce projet. Il y a des mots clés, ce sont « les verrous » et les verrous, il faut les faire sauter. Il y a la pince monseigneur et la clé et on ne sait pas trop avec quel système on va faire sauter ces verrous.

Le plan avait initié 4 projets pilotes. Le problème est qu'ils n'ont toujours pas commencé. Ils devaient lever les verrous avant de permettre de mettre en place les projets de plateformes de soins. On espère qu'ils vont commencer au premier trimestre 2019.

- **un projet pilote maladies rares autour de la déficience intellectuelle (projet DEFIDIAG)** qui va séquencer le génome complet.
- **un projet pilote cancers qui va s'occuper du sarcome des tissus mous et du cancer colorectal (MULTIPLI)** et qui utilisera du séquençage exonique et du séquençage de type RNA Seq.
- un projet en cours de discussion et de constitution qui est **un projet pour les maladies communes**. Il a été sélectionné le **diabète atypique** (projet **GLUCOGEN**)
- **un projet en population générale** (projet **POPGEN**) et qui a commencé par les bretons.

Donc tout le séquençage sera fait au Centre National de Recherche en Génomique Humaine CNRGH Evry par Jean François DELEUZE. On peut avoir un peu de sens par exemple sur la déficience intellectuelle où on s'aperçoit que très souvent ces déficiences sont liées à des mutations De Novo, donc le papa et la maman n'ont rien, l'enfant est atteint ; s'il est atteint c'est qu'il a une mutation qui n'existe ni chez son père, ni chez sa mère. Il faut la trouver et on sait que par différentes méthodes, on est capable de trouver 10% des cas et si on va au bout du génome, on est capable d'identifier 60% des causes de ces déficiences intellectuelles ce qui justifie que ce soit le premier projet pilote maladies rares.

Le deuxième axe a été de constituer un appel d'offre pour des plateformes de séquençage à très haut débit auquel il fallait répondre pour le 9 mars 2017.

10 régions y ont répondu, 2 ont été sélectionnées : SeqOIA, imaginé par Martin Hirsch lui-même **et AURAGEN** qui est la plateforme de la région Auvergne Rhône Alpes.

Cette plateforme AURAGEN est dirigée par Damien SANLAVILLE et Jean-Yves BLAY. C'est un groupement de coopération sanitaire complexe dans sa mise en œuvre puisqu'il rassemble l'intégralité de l'assistance publique hôpitaux de Paris, Gustave Roussy, l'institut

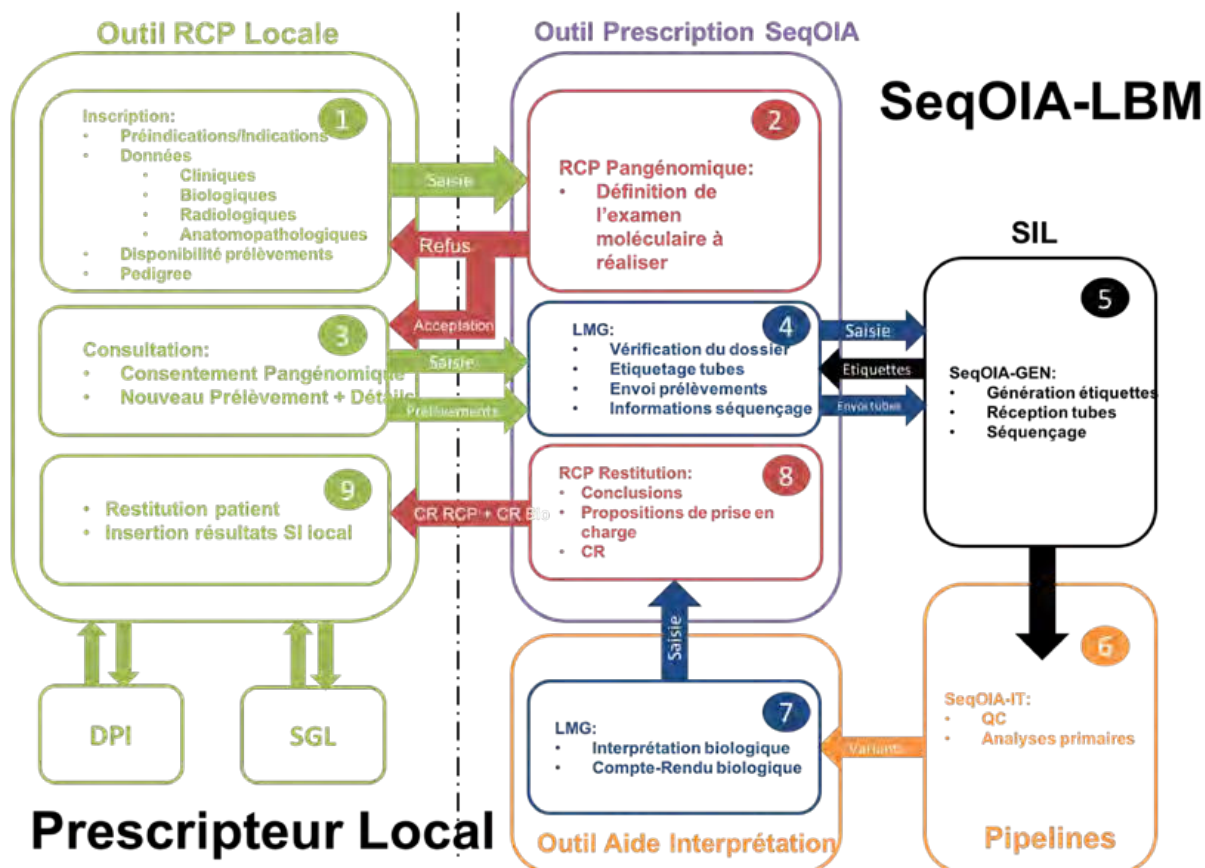


Curie et la fondation Imagine. On est donc en train de construire un LBM constitué de RCP d'amont et de RCP d'aval (Réunions de Concertation Pluridisciplinaires) au cours desquelles vont être instruits les dossiers destinés à aller sur ces plateformes de séquençage. Il y en a dans les domaines des maladies rares, de l'oncogénétique, des tumeurs solides, de l'onco-hématologie et de l'oncopédiatrie et il y aura des RCP d'aval de restitution au cours desquelles seront discutés les résultats obtenus. En Ile de France, le séquençage à très haut débit, qui n'a pas de très grande importance, a été délégué à un industriel INTEGRAGEN. La partie bio-informatique qui est absolument essentielle est prise en charge par le campus PICPUS mis en place par l'AP-HP. Les laboratoires d'amont et d'aval ont été identifiés avec des biologistes sélectionnés.

On va rappeler les capacités de séquençage de ces plateformes : la 1^{ère} année, il va être possible de séquencer 3000 équivalents génomes, donc de prendre en charge 3000 dossiers de maladies rares et 3000 dossiers de cancers. Au fur et à mesure du temps, la capacité de séquençage va augmenter avec de nouveaux équipements de type NOVASEP pour arriver à séquencer 20 000 équivalents génomes par an.

Je ne commente pas ce diagramme mais je vous laisse le regarder.

C'est tout le diagramme qu'on va être amené à mettre en place pour faire que la prescription ait du sens et on va insister lourdement pour dire que s'il n'y a pas une prescription de qualité SeqOIA avec des données cliniques structurées, nous ne pourrons pas analyser le génome dans le cadre des maladies rares. Ceci impose d'avoir le développement d'un outil de prescription qui va être intégré à notre système informatique : on va utiliser un outil qui est déjà en place à l'institut Curie, qui sera rebaptisé TheSeqOIA. J'en profite pour dire qu'en France, on a la chance d'avoir Laurent Ménard qui a la particularité d'être PUPH en néphrologie (professeurs des universités-praticiens hospitaliers), de bien aimer l'informatique et d'avoir passé 2 ans à Cornell pour comprendre ce qu'est la génomique. Il a mis au point une maquette pour la prescription génétique des examens dans le contexte de la filière ORKID et ce concept peut parfaitement être décliné à l'ensemble des



filiales. Le problème est de rester à l'état de maquette, il faudrait maintenant l'amener à l'état opérationnel et c'est la petite guéguerre habituelle franco-française que l'on connaît très bien. C'est une filière qui le propose mais pourquoi les 22 autres prendraient ce système puisqu'elles-mêmes pourraient faire la même chose. Deuxième sujet, pas de chance, on a eu PNMR3 qui est sorti entre temps, qui est un autre plan, et qui a convenu effectivement qu'il faut s'occuper des RCP : on dit c'est le PNMR3 qui va s'en occuper mais comme c'est le PNMR3, on ne peut pas prendre une filière. On ne peut donc que faire du soutien au développement de cette maquette qui n'a aucune erreur, qui est parfaitement adaptée à l'ensemble des filières maladies rares et si on n'a pas cet outil, je pense qu'on ne va pas pouvoir interpréter beaucoup de génomes.

Florent SOUBRIER, consultant, a rejoint notre équipe et va avoir comme rôle dans le contexte de SeqOIA de travailler à la constitution de ces fameux RCP pangénomiques qui sont absolument indispensables et qui sont un outil du parcours de soin extrêmement nouveau et plus important que le séquençage lui-même. Séquencer tout le génome ouvre des perspectives éthiques réglementaires et sociétales. Il fallait qu'un groupe de travail s'en occupe et avec les associations de patients, il y a eu la création de consentements

spécifiques aux analyses en génomique et des consentements qui seront accompagnés de notices d'information puisque « malheureusement », nous séquençons tout le génome et en séquençant tout le génome, on s'expose à identifier les anomalies génétiques en lien avec la maladie mais on s'expose également à identifier des variations génétiques qui ne sont pas en lien direct avec la maladie que nous appellerons globalement « données incidentes » ; il a fallu beaucoup militer pour que l'agence de la biomédecine, que vous connaissez au niveau du rein, constitue un groupe de travail qui fasse des recommandations de bonnes pratiques pour ces notions de découvertes incidentes (qu'est-ce qu'on en fait, comment on les manipule, comment on les utilise, comment on les rend) sachant qu'actuellement pour le soin, nous ne sommes pas favorables à la production de données secondaires qui sont des données qui sont activement recherchées de façon systématique, ce que font les programmes américains de génétique de regarder les anomalies génétiques dans 59 gènes de façon systématique. On va plutôt s'orienter vers une analyse de données incidentes.

Une deuxième notion a été de dire que les deux plateformes étaient présentes, les régions ne sont pas très contentes parce qu'elles n'ont pas été

LES 30 ANS

sélectionnées et donc, on nous demande d'avoir une vocation nationale. Donc je préfère cette carte ici par anamorphose qui permet de comprendre ce qu'est l'Île de France que la carte classique que vous pouvez trouver où l'Île-de-France est réduite à une portion territoriale extrêmement limitée. Du coup, la petite ligne noire, c'est la nouvelle ligne de démarcation entre SeqOIA et AURaGen, donc nous avons toute la Bretagne, les Pays de Loire, la Normandie et les Hauts de France, et AURaGen aura à traiter les demandes des patients venant des autres régions, de l'Outre-mer et de la Corse.

Il faudra donc également mettre en place un dispositif, axe 6 du plan qui est absolument essentiel, et il faudrait que votre Association ait un lobbying fort en la matière, puisqu'on a considéré que le plan ne pouvait pas commencer en disant, enfin les plateformes, en sélectionnant elles-mêmes les indications sur lesquelles elles seraient amenées à travailler par un système de lobbying classique que l'on connaît bien en France.

Donc, les préindications vont être définies par la Haute Autorité de Santé (HAS) et on est content de savoir depuis peu qu'il y a un groupe de travail en cours de constitution : ça fait 2 ans quand même que le plan tourne et la HAS vient de se réveiller. Il va être piloté par Cédric CARBONNEIL : il y aura un « délivrable » comme on dit dans le plan qui indique que le 19 décembre 2018, dernier jour de l'année presque, on aura des préindications sur lesquelles fonctionneront les deux plateformes. Il est absolument essentiel

que ces indications aient du sens et je ne vais pas me permettre de faire ce que va faire la HAS, c'est-à-dire des indicateurs qui permettront de choisir, mais il est essentiel de prioriser.

On les engage à regarder ce qu'ont fait nos collègues anglais puisque vous avez par ce lien internet (<https://www.england.nhs.uk/publication/national-genomic-test-directories/>) la connaissance qu'ont nos collègues anglais sur les indications qu'eux se reconnaissent sur le plan du wool génome, sur le plan du wool exome ou sur le plan des panels de gènes. Dernière notion, et ce sera ma conclusion, c'est le papier qui vient de sortir la semaine dernière dans le New England : on commence à pouvoir dire que ce n'est pas parce qu'on va séquencer un génome que le problème d'errance du diagnostic des maladies rares sera réglé. Voici le résultat de ce projet américain, qui a l'objectif de séquencer le génome de patients sans diagnostic :

- Sur 1519 patients qui sont référés et avec une annotation clinique extrêmement performante avec des codages PO et tout ce qu'il faut, il n'y en a que 601 qui sont acceptés pour l'évaluation. Parmi ces 601, il n'y en a que 382 pour lesquels on va faire finalement l'analyse, et parmi les 382 pour lesquels on fait l'analyse, on ne trouve la cause diagnostic sur le plan génétique que dans 132 cas, c'est-à-dire dans 1/3 des cas.

Donc, il faut qu'on s'attende dans cette première phase de mise en place complet du génome à accumuler beaucoup de données de séquençage mais que la causalité clinique du diagnostic de ces anomalies attende un petit peu ce qui justifie la mise en place d'un dernier outil du plan qu'on appelle le CAD (Collecteur Analyseur de Données).

Il est très probable que ce sont d'autres moyens très complexes d'analyse bio-informatique de ces données, en comparant plusieurs cas, en comparant avec des populations normales, en utilisant de nouveaux algorithmes, qui nous permettront peut-être d'augmenter ce chiffre mais donc le séquençage complet dans le domaine de l'errance diagnostic sur des cas indexés c'est-à-dire des situations isolées ne permet aujourd'hui que d'identifier environ que 1/3 des cas dans des indications qui étaient correctes.

Je vous remercie de votre attention. ■

