

Jamais seuls face au génome ! Produire un diagnostic génétique à l'ère du séquençage à très haut débit

REVUE MÉDECINE ET PHILOSOPHIE

Catherine Bourgain¹, Quentin Lade²

¹ Université de Paris, CNRS, INSERM, CERMES3, F-94800 Villejuif, France,

² Centre Léon Bérard, Département SHS, Chaire Inca, F-69001, France

RÉSUMÉ

A partir de l'étude d'une plateforme centralisée de séquençage à très haut débit (STHD) pour le soin, cet article analyse les pratiques développées pour maintenir les conditions de production de diagnostics dont la qualité est jugée satisfaisante par les acteurs. Il décrit un système sociotechnique complexe, encore peu stabilisé, dans lequel les professionnels organisent une gestion prudente de l'accès au STHD. Le travail diagnostique est déployé le long d'une chaîne dont les étapes constituent autant de points de contrôle des flux. La nouvelle division du travail opérée notamment entre cliniciens, biologistes et bioinformaticiens, permet d'assurer les conditions de délibérations collectives expertes et, avec, de faire face aux nombreuses incertitudes qui caractérisent une démarche encore largement exploratoire. Au-delà des capacités des technologies et des infrastructures numériques mobilisées, la valeur épistémique des diagnostics biocliniques produits repose fortement sur le travail d'une diversité croissante d'acteurs, et dans lequel la bioinformatique occupe aujourd'hui une place centrale.

MOTS-CLÉS : Séquençage du génome entier, Plateforme, Régulation professionnelle, Incertitude.

DOI : 10.51328/220503

Introduction

La génétique en contexte de soin est une spécialité médicale iconique de la biomédecine (Gaudillière, 2002). Les technologies statistiques, moléculaires et aujourd'hui bioinformatiques ont pris une place croissante dans ce qui constitue un marqueur central de cette activité clinique – la capacité à produire un diagnostic (Rosenberg, 2002), un pronostic (Christakis, 1999) ou une prédiction (Bourgain et Beaudevin, 2019).

En France, c'est une pratique très régulée par la loi mais également par les professionnels eux-mêmes (Cassier et Gaudillière, 2000 ; Inserm, 2008 ; Lowy et Gaudillière, 2008). Cette régulation professionnelle prend

différentes formes, notamment un travail pour encadrer les tests génétiques dans la clinique et conditionner ainsi les analyses moléculaires aux données cliniques (Turini et Bourgain, 2021), ou encore l'implication dans le développement de réseaux d'expertise nationaux (Bourret, 2005). Ces dispositifs visent à maintenir un lien fort entre les étapes de production de l'information moléculaire et celles qui permettent d'en proposer une interprétation clinique. Pour les professionnels, la qualité du diagnostic (mais aussi du pronostic ou de la prédiction) réside dans la capacité à maîtriser ces deux étapes et à les articuler (Cassier et Gaudillière, 2000), de façon à évaluer le niveau d'incertitude toujours important dans cette

forme de biomédecine (Eisinger, 1998).

Le développement des technologies de séquençage à très haut débit (STHD) constitue un facteur de déstabilisation fort de ce modèle historique. D'une part la capacité accrue à produire des informations génétiques moléculaires, à moindre coût, a favorisé le déploiement du marché des tests génétiques dans différents domaines. Peu à peu, une figure de la donnée génétique semble avoir émergé, de nature strictement moléculaire, susceptible de circuler seule et de n'être augmentée qu'ultérieurement par d'autres données (cliniques, statistiques, populationnelles) pour produire des informations d'intérêt sur la santé ou les origines biogéographiques. Cette figure véhicule ainsi un modèle où étapes d'analyse moléculaire et étapes d'interprétation sont totalement déconnectées. D'une certaine façon, cette figure de la donnée génétique « nue » est contraire à l'idéal historiquement construit par les professionnels de la génétique clinique en France. D'autre part, en contexte clinique, le déluge d'informations moléculaires nouvelles a fragilisé l'organisation professionnelle mise en place. Le STHD permet de séquencer plus de patients pour des variations moléculaires connues, mais aussi de faire émerger des variations nouvelles à un rythme bien supérieur aux pratiques antérieures. Ce faisant, c'est la façon de faire un diagnostic de qualité mobilisant l'analyse génétique et la capacité à le faire qui se retrouvent mis en cause¹.

Monopole historique des médecins (Freidson, 1970), le diagnostic est l'étape d'interprétation des différentes données disponibles, en vue d'associer la situation particulière d'un patient à une catégorie diagnostique définie, validée par la profession et reconnue par la société. Le moment du diagnostic cristallise ainsi le savoir médical produit tout au long de la chaîne de prise en charge des patients : il en concentre la valeur épistémique.

Dans une série de travaux fondés sur des ethnographies notamment hospitalières, le sociologue Stephen Barley (1996) distingue la figure du technicien, intermédiaire entre l'empirie et les concepts de celle de l'expert, dont le travail consiste précisément à conduire les interprétations de signes empiriques pour produire les concepts. Dans le contexte de la biomédecine, et particulièrement avec le recours au STHD, le travail d'interprétation et de production du diagnostic est néanmoins très largement conditionné par le processus de fabrication des informations (cliniques, biologiques, moléculaires...) sur lesquels il repose. Ainsi, si la valeur épistémique reste concentrée dans l'étape d'interprétation finale réalisée par les médecins biologistes, elle est également dépendante de ces étapes qui la rendent possible, et qui mobilisent d'autres acteurs, des ingénieurs et techniciens notamment. Le travail de ces derniers occupe ainsi une place plus importante qu'auparavant, bien que relativement peu visible dans la production du diagnostic.

Dans cet article, nous nous intéressons à la valeur épistémique du diagnostic génétique fondé sur une analyse du génome entier, aux formes du travail requis pour la produire et aux dispositifs sociotechniques déployés pour assurer la qualité qui sous-tend la valeur. Nous nous appuyons, pour ce faire, sur une étude du dispositif actuellement en développement, en France, dans le cadre du plan national France Médecine Génomique 2025

(FMG2025) (Aviesan, 2016). Commandé par le Premier Ministre, ce plan est assorti d'un investissement financier remarquable au regard du contexte général de limitation des dépenses de santé (Bourgain, 2018). Il organise, sur le territoire national, l'accès au séquençage du génome entier (*Whole Genome Sequencing*, WGS) pour tous les patients pour lesquels ce séquençage est jugé susceptible d'améliorer la prise en charge. Le cœur de ce dispositif est l'installation de deux plateformes de séquençage à très haut débit, organisées par des institutions publiques de soin, et couvrant respectivement les besoins du Nord et du Sud du pays, pour les maladies rares, les cancers et à terme toutes les maladies chroniques.

Pour suivre les conditions de mise en place de ce séquençage du génome entier et analyser ses effets sur la production du diagnostic génétique, nous mobilisons ici des données collectées dans le cadre d'une enquête sociologique en cours depuis 2019 (« Faire sens du cancer à l'ère de la génomique », financement INCa, avis favorable du Comité d'Évaluation Ethique de l'Inserm, IRB 0003888). De façon plus spécifique, l'article s'intéresse à l'une des deux plateformes du plan, à partir d'éléments collectés lors d'entretiens semi-directifs avec des acteurs impliqués dans la mise en œuvre concrète de la plateforme STHD (biologistes, bioinformaticiens, généticiens, médecins, directeurs scientifique et opérationnel) et lors d'observations ethnographiques (réunions d'organisation de la plateforme, de concertation pluridisciplinaire et de formation de professionnels aux outils développés spécifiquement par cette plateforme notamment). Il mobilise également un corpus de littérature grise (rapports, recommandations...) et de littérature scientifique, constitué dans cette enquête. L'approche monographique adoptée privilégie l'étude d'une plateforme sur les deux mises en place. Elle ne permet donc pas de rapporter la variété des configurations de pratiques, en lien avec l'importance des ressources et contextes locaux. Nous faisons toutefois l'hypothèse qu'elle permet de dégager des éléments significatifs des transformations en cours dans la production du diagnostic génétique. Conformément aux normes communes en sciences sociales, chaque entretien fait l'objet d'un accord oral avec l'enquêté sur les conditions d'usage de sa parole dans les travaux académiques, fondées sur le souci d'anonymisation (et parfois, en pratique de pseudonymisation) : les acteurs sont désignés par des attributs choisis pour étayer l'analyse tout en limitant la possibilité d'identification immédiate (e.g. profession, spécialité, fonction...).

Nous montrons dans cet article que les professionnels ont organisé les conditions d'une gestion prudente de l'accès au séquençage entier du génome dans les parcours de soin. La production du diagnostic est déployée le long d'une chaîne dont les étapes sont plus nombreuses et localisées dans des espaces géographiques différents. Le centralisme de la plateforme de séquençage conduit ainsi à une forme d'éclatement du diagnostic. L'organisation de ces étapes de production incarne également un souci de contrôle des incertitudes et des flux, par les professionnels. Il s'agit d'optimiser la proportion de diagnostics posés en bout de chaîne parmi les cas effectivement soumis à l'analyse du génome entier. C'est, en conséquence, le travail mis en œuvre pour contrôler l'intégralité de cette chaîne éclatée qui assure la faisabilité, la robustesse et la qualité des diagnostics informés par l'analyse du génome

¹ Voir les articles de L. Faivre, et de C. Binquet et C. Lejeune dans ce numéro.

entier, et qui concourt à produire la valeur épistémique.

La bioinformatique tient une place centrale dans le travail d'articulation requis pour faire tenir cette chaîne. Elle organise le circuit logistique des échantillons et des données ; permet la coordination du travail des acteurs mobilisés aux différentes étapes de production ; assure le contrôle qualité des données de séquençage ; organise enfin les conditions du travail conduit par les biologistes pour confronter les données moléculaires à la diversité de données complémentaires disponibles et en proposer, le cas échéant, une interprétation clinique. Les modalités de fonctionnement adoptées permettent ainsi aux professionnels de maintenir une connexion forte entre les étapes de production des données moléculaires et les étapes d'interprétation clinique. Elles supposent également de pouvoir compter sur des réseaux de professionnels experts pour faire face à la complexité et aux incertitudes générées par l'usage à grande échelle des technologies de STHD en contexte clinique. Cette capacité à conserver l'esprit et les formes de la régulation professionnelle historique s'opère toutefois dans un contexte de dépendance accrue aux acteurs et outils de la bioinformatique, qui impacte le cœur de l'activité diagnostique, les compétences professionnelles requises, et les conditions de production de la valeur épistémique.

Face au génome, un processus diagnostic éclaté tenu par une articulation bioinformatique

Une gestion prudente de l'accès au génome : un travail organisé en étapes le long d'une chaîne de production du diagnostic

Lors du lancement du plan, un constat est largement partagé parmi les généticiens des différentes spécialités médicales concernées : le diagnostic fondé sur le génome entier n'est pas une pratique suffisamment mûre pour être inscrite strictement dans le soin. Elle relève, au mieux, d'un entre deux soin/recherche. Une généticienne clinicienne impliquée dans le FMG2025 avance ainsi, à propos du séquençage du génome entier que « *La frontière entre diagnostic et recherche dans les maladies rares n'est pas claire : on fait tout le temps des aller-retours.* », quand une oncogénéticienne s'interroge « *est-on dans le diagnostic ou dans la recherche, il y a là une ambiguïté fondamentale* » (notes de terrain, 2019).

Le plan est toutefois clair sur le contexte d'accès au séquençage du génome entier, qui doit être intégré dans les parcours de soin des patients. « *Nous on ne fait que du soin ! J'insiste, c'est important, et on ne bougera pas de cette position-là* », rappelle ainsi un responsable de la plateforme étudiée (et désigné RP dans la suite du texte) devant un auditoire de cliniciens concernés (notes de terrain, 2019). Cette attention à la dimension encore fortement expérimentale se retrouve toutefois dans l'organisation retenue. Le dispositif mis en place peut être analysé comme une succession d'épisodes associant productions de données et tri des cas orientés vers l'analyse du génome entier, faisant l'objet d'ajustements successifs, comme le précise, en entretien, un biologiste impliqué dans la plateforme « *tout doit être fait de A à Z, donc clairement avant que ce soit vraiment une machine bien huilée ça va prendre des années et des années (...)* Il y a déjà des choses qui fonctionnent mais il y a beaucoup de choses en construction, qui vont être modifiées, ça c'est sûr. » (entretien biologiste 1)

Le circuit commence par l'étape de prescription qui prend désormais la forme d'une e-prescription sur un logiciel (E-PRES). Elle vise à collecter des données cliniques, rapportées de façon structurée par signes cliniques, complétées le cas échéant d'informations sur l'histoire familiale, et accompagnées du consentement spécifiquement développé dans le cadre du plan. Ces éléments sont ensuite évalués par des réunions de concertation pluridisciplinaire (dite RCP d'amont), spécifiques aux différentes indications pathologiques déclarées éligibles à la prise en charge par le plan, dites pré-indications. Il s'agit pour les professionnels de se prononcer, collectivement, sur l'adéquation du cas avec les critères retenus pour circonscrire les préindications, en évaluant le dossier clinique, l'évolution de la maladie et de sa prise en charge, et le cas échéant, les examens biologiques et génétiques déjà réalisés. La validation par cette RCP déclenche le circuit logistique qui permet la collecte du matériel biologique sur le lieu de son prélèvement et son acheminement vers la plateforme de séquençage. Dans le cas des échantillons tumoraux, une étape supplémentaire de tri dans un laboratoire de biologie expert, permet d'évaluer leur qualité au regard des standards requis pour l'analyse du génome entier.

Arrivés sur la plateforme, les échantillons sont pris en charge en vue du séquençage de l'ADN. Ce travail dont les étapes varient selon le type d'échantillons (extraction de l'ADN nécessaire ou déjà faite ; préparation des librairies ; séquençage par les séquenceurs à très haut débit) est réalisé sur place, par une entreprise prestataire, qui emploie pour ce faire des ingénieurs, des techniciens et un gestionnaire. Les données numériques de séquences générées par les machines reviennent dans le périmètre de la plateforme. Elles sont en effet prises en charge par le groupe de bioinformatique, intégré à la plateforme, qui effectue une série de contrôles qualité, sur la base de mesures définies pour partie avec le prestataire de séquençage (pour la régulation de son activité propre) et pour partie par la plateforme elle-même, selon des standards de qualité répondant aux besoins des traitements ultérieurs.

Les étapes suivantes se déploient par l'interface d'un logiciel (INTER) qui organise les conditions du travail d'articulation, réalisé par des biologistes interpréteurs, entre les données moléculaires, les données cliniques entrées dans E-PRES et les données disponibles dans des bases de données externes. INTER permet en effet de réaliser trois tâches fortement interdépendantes : (1) le repérage de l'ensemble des variants présents sur l'ADN analysé (dit *variant calling*), (2) l'annotation de ces variants – c'est-à-dire la collecte d'informations sur ces variants, produites lors du séquençage ou extraites de bases de données externes-, (3) l'organisation de tris pour guider le repérage de variants pertinents au regard des données d'annotation et des données cliniques disponibles sur le cas.

Dans le logiciel INTER, les biologistes interpréteurs, accrédités par la plateforme, réalisent le travail d'articulation entre les différentes informations moléculaires et cliniques collectées, pour produire une interprétation clinico-biologiques. Le logiciel permet un travail individuel, mais l'interface collaborative disponible permet également un travail collectif et des prises de décisions partagées. L'interprétation clinico-biologique se matéri-

alise alors par un compte-rendu d'analyse, signé selon les conditions de sa production par un ou plusieurs biologistes, et déposé dans le logiciel E-PRES. Ce compte-rendu est considéré comme concluant, lorsqu'un ou plusieurs variants jugés « causaux » sont identifiés, non conclusif dans le cas contraire. Selon les pathologies, ce compte-rendu peut, soit être directement transmis au médecin prescripteur, soit faire l'objet d'une réunion pluridisciplinaire visant à préciser l'interprétation clinico-biologique qui peut en être faite et/ou les recommandations de prise en charge à envisager.

Tout au long du circuit allant de la décision initiale de prescription jusqu'à l'interprétation clinico-biologique finale, les étapes sont ainsi conditionnées les unes par les autres. Seuls les cas ayant été retenus à toutes les étapes font l'objet d'un compte-rendu d'analyse, qui lorsqu'il est concluant permettra de faire un diagnostic, de proposer un pronostic ou de fonder une orientation théranostique. L'organisation en place vise à optimiser la proportion de diagnostics posés en bout de chaîne parmi les cas effectivement soumis à l'analyse du génome entier, et non parmi ceux qui ont fait l'objet d'une prescription initiale par le clinicien traitant. En ce sens, toutes ces étapes de tri des cas participent à la production du diagnostic. Ce dernier ne se réduit pas, en dépit de son importance, au travail réalisé lors des étapes finales d'interprétation des données clinico-biologiques. C'est le travail mis en œuvre pour contrôler l'intégralité de cette chaîne qui assure la faisabilité, la robustesse et la qualité des diagnostics informés par l'analyse du génome entier. Il se caractérise par une série de dispositifs permettant l'enchaînement des étapes, au sein desquelles la bioinformatique tient une place tout à fait centrale.

Un diagnostic qui requiert un travail important d'articulation, pour lequel la bioinformatique est centrale

Parmi ces dispositifs, la logistique du circuit des échantillons biologiques est importante. La plateforme a en effet la charge d'analyser les échantillons provenant de la moitié du territoire national, dont bon nombre d'hôpitaux avec lesquels il n'existe pas de routine préétablie pour les circulations d'échantillons, ni d'outils de gestion numérique partagé. Une société privée a été retenue pour organiser les aspects matériels et numériques du recueil des échantillons dans les hôpitaux partenaires et de leur transport jusqu'aux sites centralisés de traitement organisés par la plateforme. Comme le reconnaît en entretien, un responsable de la plateforme, « *tout ça a pu être sous-estimé, dans la conception du plan. Les aspects pratiques* ». Deux aspects sont importants dans ces étapes, la maîtrise des temps d'acheminement qui doivent rester compatibles avec les impératifs des prises en charge des patients, et la qualité des échantillons et des ADN qui pourront être extraits. En effet, pour les analyses du génome entier, « *les critères de qualité de l'ADN sont très supérieurs à ceux qui sont utilisés quand on fait des panels* » (entretien RP), c'est-à-dire lorsqu'un nombre défini de variations localisées dans des gènes choisis sont recherchées.

De façon plus générale, la possibilité même de mener à bien un diagnostic résultant d'un travail réparti en autant d'étapes successives et conditionnées les unes aux autres, repose sur une infrastructure bioinformatique centralisée. Celle-ci a été mise en place sans trop de controverses, dans la mesure où il s'agit d'une plateforme informatique nou-

velle, sans précédent important déjà en place. L'ingénieur responsable de la plateforme (désigné IRB dans la suite du texte) a ainsi été chargé de sa mise en place « *à partir de zéro* », selon ses termes.

Ce dernier avait alors une dizaine d'années d'expérience en milieu hospitalier. Rodé aux interactions avec les biologistes et les cliniciens dans ses fonctions antérieures, il défend dès le départ le choix de développer localement des logiciels open source, plutôt que d'acheter des solutions commerciales. Il s'agit ainsi de modéliser ces logiciels en collaboration avec les biologistes interprétateurs, et dans une moindre mesure les cliniciens, au plus près de leurs pratiques et de leurs besoins.

L'équipe des bioinformaticiens est constituée d'une petite dizaine d'ingénieurs dont les profils reflètent l'organisation d'une plateforme orientée vers les exigences cliniques de production routinière des analyses de génome et assurant le développement interne des logiciels. Elle comprend notamment un ingénieur système capable de faire fonctionner l'infrastructure matérielle, en particulier les serveurs de calcul. Deux bioinformaticiens sont chargés de l'élaboration des algorithmes de détection et d'annotation des variants pour le cancer et les maladies rares et un « *intégrateur* » est chargé, à partir des outils d'analyse mis au point par les bioinformaticiens de « *prendre l'information de quels outils, dans quel ordre et avec quels paramètres et en faire une chaîne de traitements vraiment de production* » (entretien IRB). L'équipe est aussi composée de développeurs, pour coder les logiciels sous licence open source. Si bien que la bioinformatique au sein de la plateforme ne repose sur « *aucune licence logiciel utilisée et donc aucun achat de logiciel pour la partie traitement analytique* » (entretien IRB).

C'est ainsi le groupe de bioinformatique qui développe le logiciel E-PRES. Cette e-prescription n'était pas encore en place dans le principal groupement d'hôpitaux de la plateforme, « *c'est un peu une irruption dans le monde de certains cliniciens qui étaient habitués à remplir des notes* » (entretien RP). Pour le responsable de la bioinformatique, ce développement est une condition importante pour parvenir à un diagnostic de qualité à partir du génome entier : « *C'est un des points forts qu'on a. C'est qu'il y a beaucoup de boîtes qui proposent des outils d'aide à l'interprétation à prix d'or. Le truc qu'ils n'ont pas c'est la partie prescription. Et franchement... c'est les deux faces d'une même pièce, en fait. La prescription c'est vraiment l'entrée et on peut recueillir un nombre d'informations et de data qui sont ultra importantes* » (entretien IRB). Avoir la main sur la structuration des données cliniques permet ainsi d'anticiper les besoins ultérieurs du diagnostic génétique.

S'il est qualifié de « *logiciel de e-prescription* », E-PRES assure en pratique d'autres fonctions importantes en plus du recueil des données cliniques. C'est de fait un outil de coordination de toute la chaîne de production du diagnostic. Il permet de stocker les données familiales lorsqu'elles sont pertinentes et les consentements. C'est le lieu où sont déposés les comptes rendus techniques relatifs à la qualité du séquençage et, en fin de chaîne, les comptes rendus de l'analyse génétique avec, le cas échéant, le diagnostic proposé. Enfin, il assure le suivi de l'avancement de la séquence de traitement, et rend cette information directement accessible aux prescripteurs et autres acteurs autorisés.

Les bioinformaticiens ont également mis au point l'autre interface logicielle au cœur du travail productif de la plateforme, INTER. Souvent qualifié de logiciel « d'aide à l'interprétation » des séquences génétiques à destination des biologistes, INTER assure lui aussi plusieurs fonctions parmi lesquelles : (1) contrôler la qualité et donner accès aux données de séquençage et (2) permettre le travail d'articulation entre données moléculaires, données cliniques et données issues de bases externes.

Ce logiciel constitue une nouveauté importante car les biologistes interpréteurs sont habitués à travailler sur des fichiers de données de séquençage, téléchargés sur leur ordinateur et ouverts dans des tableurs Excel. Ces tableurs sont bien adaptés à la démarche par filtres successifs qui caractérise les étapes de traitement de grande quantité de variants génétiques mais les données issues de séquences de génomes entiers sont désormais trop volumineuses pour être téléchargées et manipulées sur les ordinateurs de bureau. INTER est ainsi une interface en réseau donnant accès à distance à ces données, sur le serveur de la plateforme. À cette nécessité matérielle, les bio-informaticiens et un responsable de la plateforme ont ajouté un impératif complémentaire au développement de INTER : « que la personne qui interprète ait la visibilité sur tout ce qui se passe dans la boîte » (entretien RP). Cette volonté se traduit de deux manières dans le fonctionnement de l'interface. Les concepteurs ont choisi de donner accès à une série de mesures qui caractérisent la qualité technique du séquençage (qualité de l'alignement, taille des inserts séquencés, couverture moyenne, scores de qualité...). Ils ont également fait le choix de ne pas filtrer les très nombreux variants détectés par les algorithmes bio-informatiques, contrairement à la plupart des logiciels existants dans lesquels le choix des filtres est en partie automatisé. Tous les variants détectés sur la séquence sont ainsi accessibles, avec pour chacun d'eux le détail des informations disponibles.

Cette exhaustivité vise à laisser aux biologistes interpréteurs la liberté de sélectionner les informations dont ils ont besoin. De fait, comme le reconnaît le responsable de la bioinformatique : « On essaye de rentrer dans leur tête, voir comment ils fonctionnent et essayer d'avoir une méthodologie automatique mais le problème c'est qu'ils abordent chaque dossier de manière différente. Pour moi c'est incodable, impossible à remplacer » (entretien IRB). A propos du travail d'annotation, un biologiste précise ainsi que celui-ci requiert « de donner toutes les informations dont on a besoin pour interpréter un variant. Mais d'éviter de nous donner des informations dont on n'a pas besoin. Le but c'est qu'on ne soit pas noyé dans des informations inutiles » (entretien biologiste 1). Pour autant, deux biologistes peuvent avoir des approches différentes de ce travail de tri « Il y a des annotations qui sont un peu redondantes. Moi, je vais en choisir une et laisser tomber l'autre. Mais un autre biologiste peut très bien préférer l'autre. » (entretien biologiste 1).

Cette difficulté à développer des procédures automatisées devant la diversité et la complexité des situations ne concerne pas uniquement l'étape d'interprétation des variants. L'étape de repérage des variants (dite *variant calling*) est elle-même « vraiment très compliquée parce qu'il y a beaucoup de types de variants différents » (entretien biologiste 1), et justifie de pouvoir accéder à toutes les données disponibles.

Enfin, INTER a été développé avec le souci de rendre

possible un travail collaboratif des biologistes interpréteurs, à distance puisque les biologistes sont répartis sur une zone géographique étendue. Mais, pour le responsable de la bioinformatique, au-delà d'une réponse à un problème logistique de distance entre les professionnels, ce « côté communautaire » est essentiel avec le génome entier : « Vous avez le droit de regarder les dossiers, mais vous savez que peut-être vous allez buter sur certains aspects. Si vous avez besoin d'un coup de main, n'hésitez pas à appeler à l'aide. Arrêtez de travailler dans votre coin et de penser que vous savez tout et tout faire tout seul. Vous êtes face au génome. Face au génome, il ne faut pas rester tout seul. Ça va mal se passer sinon. » (entretien IRB)

Face au génome, le travail diagnostic se transforme. Il est alors réparti le long d'une chaîne de production aux étapes plus nombreuses, diverses, et géographiquement éloignées. Il implique un nombre grandissant d'acteurs et nécessite un travail d'articulation entre les étapes de la chaîne, rendu possible par les développements de la bioinformatique. Dans la partie qui suit, nous analysons comment les modalités de régulation des pratiques de génétique dans le soin en sont elles-mêmes transformées.

Une capacité à conserver une régulation professionnelle sur le diagnostic génome entier, au prix d'une dépendance accrue aux acteurs et outils de la bioinformatique

Une organisation qui permet aux généticiens de conserver les formes de la régulation professionnelle

Alors que le plan FMG2025 est une initiative politique nationale, qui aurait pu incarner une stricte montée en puissance d'une régulation administrative sur cette activité clinique, le dispositif décrit ici permet aux généticiens de conserver l'esprit et les formes de leur régulation professionnelle.

Dans le cas présenté, il apparaît en effet que les professionnels ont organisé les conditions pour que les étapes de production des données moléculaires restent intimement connectées aux étapes d'interprétation clinique, et pour qu'ils en gardent le contrôle. Avec l'analyse du génome entier, les incertitudes associées à l'identification des variants génétiques apparaissent encore plus fortes, ainsi que la nécessité de pouvoir tenir compte de ces incertitudes dans le travail d'interprétation clinique. Le logiciel INTER donne accès, pour ceux d'entre eux qui sont formés et accrédités par la plateforme, aux mesures de qualité des séquences, à l'ensemble des variants détectés et aux annotations disponibles. Il permet ainsi d'intégrer l'analyse critique de l'étape de production des données moléculaires et d'en faire une étape conditionnant le travail d'interprétation. Mais maintenir la possibilité de cette articulation a un coût. C'est une activité difficile, qui impose un effort important aux biologistes, pour se former et prendre en main INTER. L'enjeu est aussi de construire les conditions d'une confiance de ces généticiens biologistes, dans les données issues du séquençage, sur la production desquelles ils n'ont plus directement la main – puisqu'elle est assurée par l'entreprise prestataire avec la plateforme de bioinformatique. Un ensemble de mesures assorties d'un seuil de confiance sont ainsi proposées avec l'objectif de venir renforcer la qualité du travail d'interprétation.

De la même façon, l'importance des réseaux d'expertise professionnelle est renforcée pour faire face

à la complexité du travail d'interprétation. Le choix d'organiser les étapes d'analyse et d'interprétation en favorisant les démarches collaboratives, permet aux biologistes de travailler de façon collective. L'expertise professionnelle s'est historiquement structurée autour de spécialistes reconnus pour leur connaissance pointue de quelques gènes impliqués dans une pathologie donnée et leurs variants. Mais, avec les analyses sur le génome entier, un généticien peut « *trouver des variants qui ne sont plus dans des gènes qu'il connaît. Et souvent, c'est des gènes qui sont connus dans d'autres pathologies. Donc c'est bien de pouvoir discuter avec les biologistes qui sont plus experts de ces gènes-là* » (entretien IRB). La diversité d'expertise disponible dans le réseau devient ici une ressource précieuse pour faire face à la complexité. La mobilisation du réseau pour produire le compte-rendu diagnostique devient dans certains cas une nécessité et de façon générale une garantie de qualité. La possibilité réglementaire de cosigner à plusieurs biologistes ces diagnostics permet, en outre, une reconnaissance et une valorisation, y compris financière de ces expertises.

Au prix d'une centralité accrue de la bioinformatique et d'effets transformants importants

Le maintien de cette régulation professionnelle s'opère toutefois dans un contexte de dépendance accrue aux acteurs et aux outils de la bioinformatique. Un responsable de la plateforme, le reconnaît d'ailleurs sans hésitation. « *Nous avons dû dématérialiser tout le processus et nous n'aurions pas pu le faire sans une informatique du feu de Dieu* » (entretien RP). Cette informatique sert à organiser la chaîne de production du diagnostic, à gérer la masse des données produites et à organiser les conditions de leur traitement. Dans ce contexte, il convient d'être attentif aux effets transformants de cette situation nouvelle.

La dématérialisation impose ainsi une standardisation des modalités de recueil de l'information clinique. Dans l'outil d'e-prescription, les cliniciens sont invités à décrire leurs patients en utilisant des codes internationaux HPO (*Human Phenotype Ontology*), catégories standardisées développées pour les phénotypes de maladies. Or, comme le pointe un responsable de la plateforme : « *On doit pousser les cliniciens à mieux décrire leurs patients sous des formes structurées, les HPO. Mais les codes HPO, il y a quand même 15 000 codes HPO. Il y a un enjeu de prescription* » (entretien RP). Cette standardisation de la prescription impose une nouvelle forme de travail autour des données cliniques. Il s'agit en effet d'effectuer une étape supplémentaire de traitement pour les traduire en HPO, autrement dit de faire des relectures analytiques des éléments cliniques disponibles et d'effectuer les arbitrages requis pour décider des codes les plus pertinents. La complexité de cette tâche chronophage et son importance dans la qualité voire la possibilité même de poser un diagnostic, a conduit les acteurs du plan FMG2025 à obtenir le recrutement d'assistants de prescription, acteurs supplémentaires de la chaîne diagnostique, pour aider à supporter la charge de ce fastidieux travail de traduction et de numérisation.

Un second aspect d'évolution à souligner tient aux compétences des généticiens biologistes interpréteurs, accrédités par la plateforme. A propos du petit groupe avec lesquels il collabore étroitement, le responsable bioinformatique précise ainsi : « *Le biologiste de demain, il est un peu bioinfo. On ne lui demande pas de savoir écrire du code.*

On lui demande de comprendre les algorithmes utilisés, pour sortir une information intelligible qui est rendue au patient ». La bioinformatique devient ainsi une compétence essentielle à la production du diagnostic. Comme le remarque ce bioinformaticien, « *Il suffit juste que je mette une erreur dans mon code pour produire des erreurs monumentales de diagnostic* » (entretien IRB). De fait, les responsables de la plateforme organisent de façon très régulière des sessions de formation à destination des généticiens biologistes, dont il s'agit de transformer les compétences, et qu'il faut « *mettre à niveau. Pour aligner tout le monde au niveau du whole genome* ».

Si, à propos de son travail le responsable de la bioinformatique avance « *Il faut que ça fonctionne tout le temps. Il ne faut pas qu'il y ait de trou. Et grosso modo, on passe notre temps à faire en sorte qu'on n'ait plus besoin de nous* » (entretien IRB), un biologiste interpréteur insiste lui sur le fait que « *La partie bio-informatique pour moi c'est le plus important sur la plateforme. [...] Quand on récupère le whole génome on a des millions de variants donc quand un biologiste se retrouve avec des millions de variants et incapable d'interpréter ça, c'est donc la bio-informatique qui permet de lui donner quelque chose qui est interprétable* » (entretien biologiste). Ce travail sur l'infrastructure, destiné à être invisibilisé, détermine en fait les conditions de possibilité des étapes dites d'interprétation.

Impliqués à la fois dans l'organisation logistique, le contrôle de la qualité des données, l'organisation des conditions pour leur interprétation et pour le fonctionnement des collectifs biocliniques, les bioinformaticiens occupent une position d'interface entre des fonctions relevant pour certaines d'un travail technique et pour d'autres d'un travail expert, pour reprendre la distinction de Barley (1996). Cette multipositionnalité contribue ainsi à brouiller les distinctions et à redéfinir les frontières entre les dimensions techniques et épistémiques de la chaîne de production diagnostique. En effet, en contribuant directement aux décisions qui organisent les choix auxquels les biologistes doivent faire face, et en organisant pratiquement les modalités des réponses possibles, les bioinformaticiens deviennent eux-mêmes, au-delà des compétences bioinformatiques que les généticiens peuvent acquérir, des coproducteurs de la valeur épistémique du diagnostic.

Conclusion

La production de diagnostics de qualité, adossés à l'analyse du génome entier, repose dans le cas que nous avons étudié sur un dispositif sociotechnique complexe. Celui-ci comprend une dimension matérielle importante qui inclut les échantillons biologiques nombreux et variés, les réactifs et technologies nécessaires à leur traitement pour en extraire l'ADN et pour le séquencer, l'infrastructure numérique sur laquelle repose toute la chaîne de traitement, de collecte et d'analyse des données (serveurs, logiciels, terminaux...). Mais il comprend également une très forte dimension de travail humain, requis pour mener à bien les différentes étapes d'évaluation et de tri des patients ou des échantillons éligibles à l'analyse moléculaire ; pour produire les données dont dépendent ces tris (recueil des consentements, dématérialisation par la standardisation des données cliniques) ; pour organiser la logistique des échantillons ; pour préparer ces échantillons, faire fonctionner et surveiller les séquenceurs et la qualité des séquences pro-

duites ; pour développer l'infrastructure bioinformatique et en assurer le fonctionnement et la maintenance au quotidien ; pour articuler les données moléculaires avec les autres données disponibles et en proposer une interprétation robuste.

Si ce travail est en partie similaire à celui qui est requis pour des configurations moins centralisées ou mobilisant des technologies moléculaires moins exhaustives, une série de transformations apparaît tout au long de la chaîne de production des tests génétiques. En particulier, une nouvelle division du travail est opérée, entre biologistes et bioinformaticiens mais aussi avec une montée en puissance des décisions concertées entre cliniciens et biologistes lors des RCP. De nouveaux acteurs, tels que les assistants de prescription, sont impliqués. Les compétences requises évoluent également, si bien que les généticiens biologistes interpréteurs sont formés aux outils bioinformatiques nouvellement créés.

Ces transformations s'inscrivent dans une tradition de pratiques professionnelles, organisées pour faire place aux nombreuses incertitudes de l'analyse génétique en clinique (Bourret et Rabeharisoa, 2008 ; Bourgain et Beaudevin, 2019). C'est notamment le cas de l'importance prise par les réseaux de professionnels experts, qui ont historiquement constitué des ressources pour maintenir une autonomie clinique et éviter les écueils d'approches standardisées (promues notamment par certains acteurs marchands), jugées insatisfaisantes du point de vue de la valeur épistémique du diagnostic. Ici, le réseau d'experts se retrouve mobilisé par une alliance nouvelle avec la bioinformatique et l'administration, qui assure les conditions pratiques et la reconnaissance y compris financière de formes collectives de travail, au cœur de l'activité diagnostique, et de la *production de la valeur épistémique*.

Ce dispositif, en cours d'élaboration, n'est toutefois pas stabilisé et reste fragile en raison de sa complexité logistique et organisationnelle notamment. Il demeure dépendant du réseau d'interconnaissances préexistant, pour décider par exemple des experts de confiance sur tel ou tel gène, et de la capacité à collaborer avec des acteurs en situation de potentielle concurrence : cliniciens, généticiens biologistes, bioinformaticiens ou acteurs administratifs. Au cours de notre enquête, nous n'avons pas identifié de conflits de juridiction entre généticiens biologistes interpréteurs et bioinformaticiens, ces derniers revendiquant une position au service des premiers. Il est néanmoins important de rester attentifs à la robustesse effective des dispositifs sociotechniques complexes mis en place pour faire *face au génome*. Si l'objectif de ce dispositif permet d'établir un nouveau standard de production des tests génétiques, il s'agit de tenir compte des configurations locales et des effets à moyen et long terme du modèle adopté - centralisé et largement digitalisé - sur la capacité effective à maintenir et transformer ces réseaux de compétences, inscrits sur les territoires, et qui sont essentiels à la maîtrise des technologies et à la production de la valeur épistémique du diagnostic génétique. Le suivi de l'évolution de ce dispositif sociotechnique dans le temps et la comparaison avec d'autres dispositifs similaires mis en place, notamment en France avec l'autre plateforme financée par le plan FMG2025, pourront permettre de poursuivre ces investigations.

Références

- Aviesan (2016). Plan France M.décine Genomique 2025. <https://www.inserm.fr/wp-content/uploads/2017-11/aviesan-planfrancemedecinegenomique-2025.pdf>
- Barley, S. (1996). Technicians in the Workplace: Ethnographic Evidence for Bringing Work into Organizational Studies. *Administrative Science Quarterly*. Vol. 41, pp.404- 441.
- Bourgain C. (2018). De la génétique clinique à la médecine génomique. Enjeux d'une « démocratisation » de l'accès aux technologies génomiques en contexte de soin. *Cahiers Droit, Sciences et Technologies*. Vol. 8, pp.15-29.
- Bourgain, C. et Beaudevin, C. (2019). Au-delà de la technologie : travailler à la fiabilité de la prédiction génétique, *Médecine et philosophie*, vol. 2, pp. 34-39.
- Bourret, P. (2005). BRCA Patients and clinical collectives : new configurations of action in cancer genetics practices. *Social Studies of Science*. vol. 35, pp. 41-68.
- Bourret, P., et Rabeharisoa, V. (2008). Décision et jugement médicaux en situation de forte incertitude: l'exemple de deux pratiques cliniques à l'épreuve de la génétique. *Sciences sociales et santé*, vol. 26, pp. 33-66.
- Cassier M., Gaudillière J.P. (2001). Recherche, médecine et marché : la génétique du cancer du sein. *Sciences sociales et santé*. Vol. 18, pp.29-51.
- Christakis, N. A. (1999). *Death Foretold: Prophecy and Prognosis in Medical Care*. Chicago and London: University of Chicago Press.
- Eisinger F. et al. (1998). Recommendations for Medical Management of Hereditary Breast and Ovarian Cancer: The French National Ad Hoc Committee, *Annals of Oncology*, vol. 9, pp. 939-50.
- Freidson, E. (1970). *Profession of Medicine*. New York: Dodd Meads.
- Gaudillière, J.P. (2002). *Inventer la biomédecine*. Paris : La Découverte.
- Inserm. (2008). Surveillance et régulation des tests génétiques. in *Tests génétiques, questions scientifiques, médicales et sociétales*. Paris : Les éditions Inserm.
- Löwy, I. et Gaudillière, J.P. (2008). Localizing the global: Testing for hereditary risks of breast cancer, *Sci Tech Hum Val*, vol. 33, pp. 299-325.
- Rosenberg, C. (2002). The Tyranny of Diagnosis: Specific Entities and Individual Experience, *Milbank Quarterly*, vol. 80, pp. 237-260.
- Timmermans, S. et Buchbinder, M. (2012). Expanded Newborn Screening: Articulating the Ontology of Diseases with Bridging Work in the Clinic. *Sociology of Health and Illness*, vol. 34, pp. 208-220.
- Turrini, M. et Bourgain, C. (2021), Gérer l'incertitude dans la prévention génétique d'une maladie fréquente : décision médicale et thrombophilie à l'hôpital. *Sciences Sociales et Santé*. Vol. 39, pp. 5-32.