



**HAL**  
open science

# Diagnostiquer les maladies infectieuses à l'ère de la génomique

Philippe Pérot

► **To cite this version:**

Philippe Pérot. Diagnostiquer les maladies infectieuses à l'ère de la génomique. [Rapport de recherche] Institut Pasteur Paris; Université Pierre & Marie Curie - Paris 6. 2018. pasteur-01926297

**HAL Id: pasteur-01926297**

**<https://pasteur.hal.science/pasteur-01926297>**

Submitted on 19 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License



---

# DIAGNOSTIQUER LES MALADIES INFECTIEUSES A L'ERE DE LA GENOMIQUE

---

Philippe Pérot



Diplôme Universitaire de Génie Biologique et Médical  
Promotion 2018



## Table des matières

<b>Introduction</b> .....	2
<b>I/ Le marché et l'industrie du DIV</b> .....	2
Présentation du secteur .....	2
La place du DIV dans le parcours de soins .....	4
Les méthodes diagnostiques historiques et l'essor des approches moléculaires.....	4
<b>II/ L'apport du séquençage à haut débit (NGS)</b> .....	6
Séquençage à haut débit et science génomique.....	6
Utilisation du NGS pour l'identification de pathogènes connus et défis méthodologiques.....	7
Apport du NGS pour l'identification de nouveaux pathogènes .....	9
Usages du NGS pour un diagnostic de terrain ou au lit du patient.....	9
<b>III/ La bataille du « NextGen Dx »</b> .....	11
Les sociétés qui fabriquent et commercialisent les séquenceurs et les solutions d'analyses .....	11
Les utilisateurs du NGS en DIV .....	12
Un modèle français possible alliant médecins, scientifiques et industriels.....	13
<b>Conclusion</b> .....	14
<b>Références</b> .....	15

## Introduction

Les maladies infectieuses sont responsables dans le monde de millions de morts chaque année. Les infections des voies respiratoires, le VIH, les maladies diarrhéiques, la tuberculose, le paludisme et la rougeole représentaient au début du 21<sup>ème</sup> siècle plus de 90% des décès par infection selon l'OMS. Les progrès considérables dus à l'introduction de l'hygiène pasteurienne au siècle dernier et l'usage des antibiotiques et de la vaccination ont fait reculer la mortalité dans les pays industrialisés, mais de nouveaux facteurs d'émergence inquiètent aujourd'hui les autorités sanitaires. Dans un monde globalisé, les transports terrestres, maritimes et aériens, les guerres et les migrations qu'elles provoquent, la pauvreté, la densification des contacts entre l'homme et les animaux d'élevage, les modifications des habitats naturels par la déforestation ou les changements climatiques rapides, forment une combinaison complexe de facteurs environnementaux et socio-économiques qui exposent les êtres humains à de nouveaux risques infectieux. L'augmentation des traitements anticancéreux et des greffes d'organes s'accompagnent d'immunodéficiences induites favorisant les infections par des pathogènes variés habituellement sans conséquence chez les individus immunocompétents. Dans le même temps, l'antibiorésistance est devenue un sujet d'inquiétude majeur. En France, la défiance à l'égard de la vaccination se traduit actuellement par la réémergence de foyers de rougeole. Dans le monde, les épidémies récentes dues aux virus Chikungunya, Ebola, Zika, West Nile, ou encore celui du MERS ont illustré le caractère à la fois difficilement prévisible des émergences et le rôle central de certaines espèces vecteurs dans la propagation des maladies.

Etre en mesure de détecter efficacement des agents infectieux par des méthodes de Diagnostic *In Vitro* occupe donc une place centrale à la fois dans certains parcours de soins et dans une perspective de surveillance épidémiologique plus large. Les avancées technologiques récentes en matière de séquençage de l'ADN et les nouvelles approches de génomique qui en découlent fournissent des outils puissants pour la détection de pathogènes connus ou nouveaux. Cela est rendu possible aussi par un effort au niveau mondial de mise en commun de l'information génétique sur les microorganismes ainsi que par les technologies toujours en évolution de l'informatique et du numérique. Après une brève présentation du secteur du Diagnostic *In Vitro*, ce rapport expliquera et illustrera les apports de la génomique au diagnostic des maladies infectieuses.

## I/ Le marché et l'industrie du DIV

### Présentation du secteur

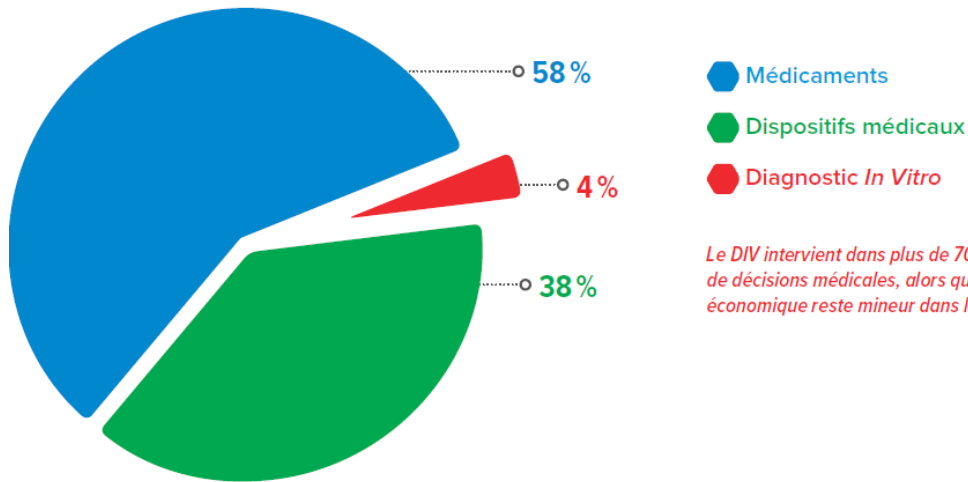
L'industrie du Diagnostic *In Vitro* (DIV) est comparativement moins connue du grand public que l'industrie du médicament ou des dispositifs médicaux. Les tests de DIV jouent pourtant un rôle essentiel dans le parcours de soins, en permettant au médecin d'orienter ses décisions thérapeutiques en fonction de résultats d'analyse obtenus à partir de prélèvements du patient (sang, urine, biopsies ou autre).

Comme le rappelle le livre blanc 2017 du Syndicat des Industries du Diagnostic *In Vitro* (SIDIV)<sup>1</sup>, le DIV est utilisé dans environ 70% des prises de décision médicale en médecine de ville et dans plus de 80% à l'hôpital. En France, le secteur est composée d'une centaine d'entreprises dont près de 90% de PME

et TPE, représentant environ 12000 emplois directs sur le territoire. Au niveau européen, la France est le 2<sup>ème</sup> marché derrière l'Allemagne et devant l'Italie.

## L'industrie du DIV en France

Répartition du CA des industries de santé



*Le DIV intervient dans plus de 70% des prises de décisions médicales, alors que son poids économique reste mineur dans le marché de la santé.*

Chiffre d'Affaires global : 1,8 Mds €

- dont 70 % auprès des laboratoires de biologie médicale (publics et privés) et 22 % pour l'autodiagnostic

Nombre d'entreprises : 80

- dont 90% de PME
- dont 33 % produisant en France, lesquelles réinvestissent 12 % en R&D et réalisent 85 % de leur CA à l'export

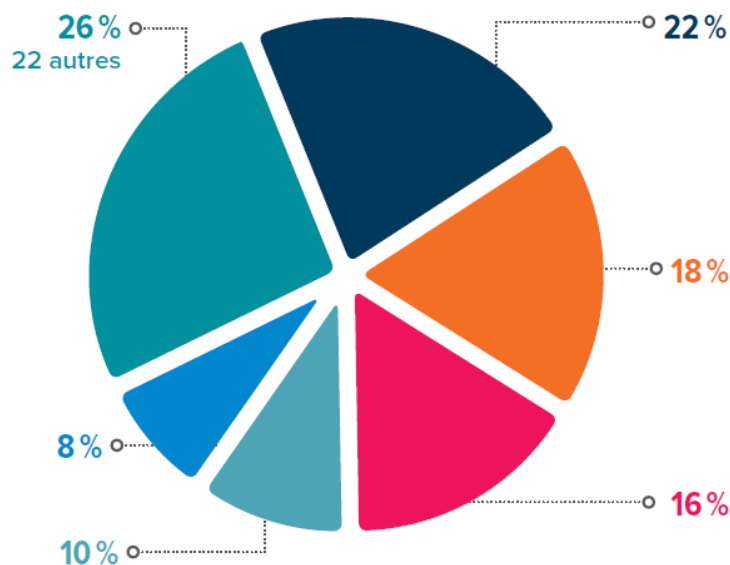
Nombre d'emplois directs : 12 000

- dont 46 % en R&D (16 %) et en production (30 %)
- dont 70 % au sein des entreprises productrices en France

Source : SIDIV 2017

## Le marché du DIV en Europe

Principaux marchés européens



L'industrie du DIV en Europe :

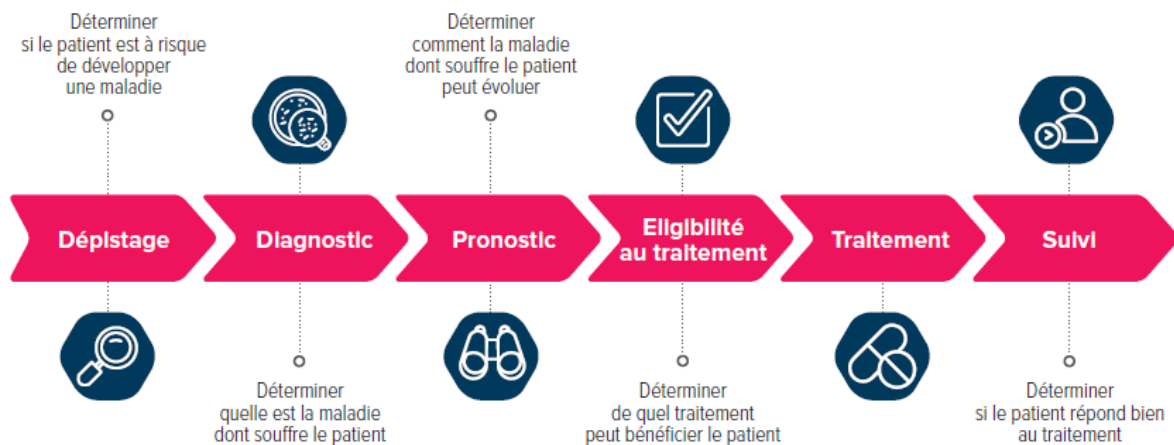
- EU (27+EFTA) : 10,638 Mds €
- 5 marchés principaux :
  - Allemagne : 2,189 Mds €
  - France : 1,785 Mds €
  - Italie : 1,656 Mds €
  - Espagne : 0,972 Mds €
  - GB : 0,812 Mds €
- La France occupe la 2<sup>e</sup> place

Source : SIDIV 2014

## La place du DIV dans le parcours de soins

Il ne faut pas réduire l'usage du DIV au seul besoin de diagnostiquer la maladie dont souffre un patient. Le Diagnostic *In Vitro* est utilisé tout au long d'un parcours de soins, en permettant aussi de dépister ou prévenir un risque, de faire le pronostic de son évolution, de sélectionner un traitement adapté ou au contraire d'éviter certains choix thérapeutiques inutiles ou à risque pour l'individu, et enfin d'évaluer et surveiller la réponse à un traitement. Il peut donc intervenir avant, accompagner, ou se faire après un diagnostic.

### *Le Diagnostic In Vitro dans le parcours de soins*



Les instruments, réactifs ou services de l'industrie du DIV interviennent à toutes les étapes du parcours de soins du patient. (Source : SIDIV 2017)

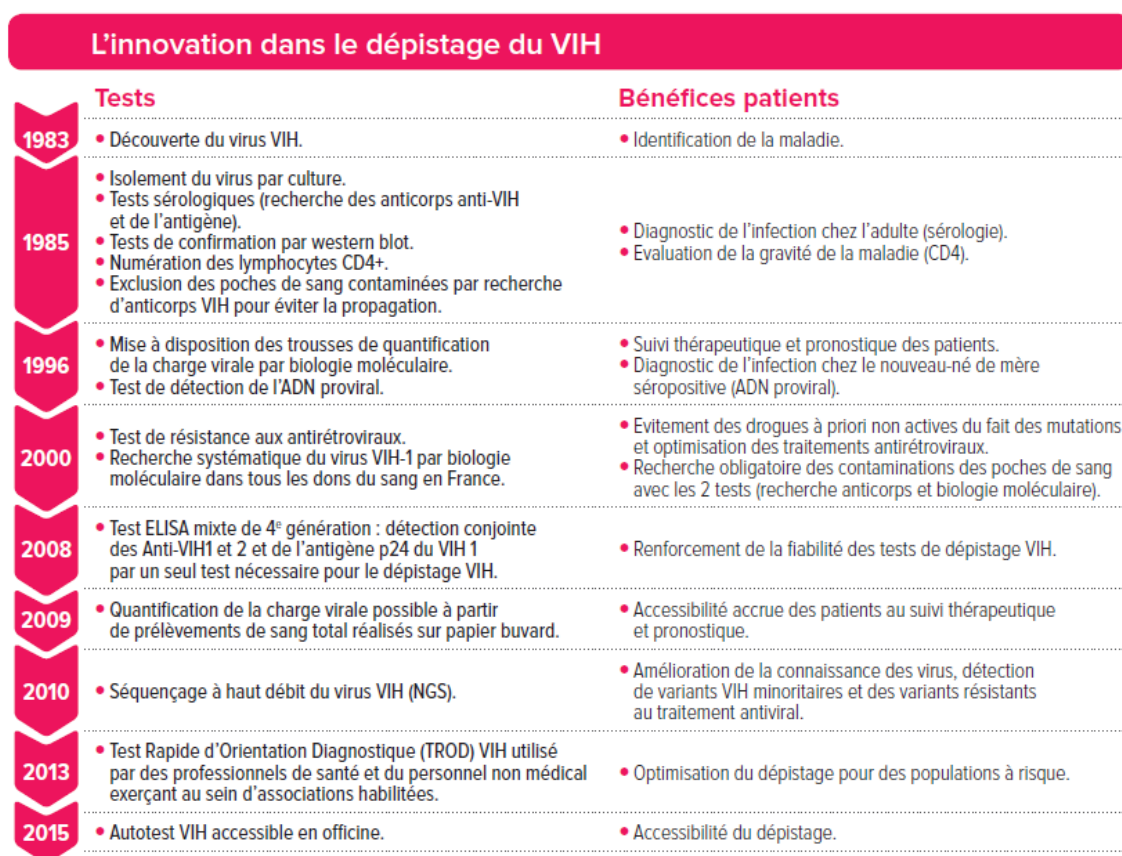
Cette trame générale qui fait référence à des applications de DIV dans des domaines variés tels que le diabète, les maladies cardio-vasculaires ou les pathologies cancéreuses, reste tout à fait valable lorsque l'on considère plus particulièrement le champ des maladies infectieuses. En effet, dépister la présence d'un pathogène avant l'apparition de signes cliniques améliorera la prévention d'une maladie transmissible ; réaliser un diagnostic d'infection dans un contexte symptomatique établira la cause et la source d'une contamination ; détecter la présence d'un marqueur bactérien ou viral aidera au pronostic du malade ; élire un traitement antibiotique, par exemple, se fera sur la confirmation de la nature bactérienne d'une infection ; le choix thérapeutique dépendra évidemment de l'agent infectieux identifié ; et des indicateurs de contrôle de l'infection dans le temps contribueront au suivi d'un patient.

## Les méthodes diagnostiques historiques et l'essor des approches moléculaires

Face aux maladies infectieuses, l'industrie du DIV a considérablement évolué notamment depuis l'apparition du VIH au début des années 1980. L'exemple du VIH servira donc à illustrer un ensemble de méthodes couramment utilisées pour le Diagnostic *In Vitro* de maladies infectieuses.

Historiquement, l'observation d'une bactérie ou d'un virus est rendue possible par sa mise en culture et son isolement, puis par l'observation en microscopie ou en microscopie électronique. L'observation directe d'un pathogène dans sa forme native et entière est cependant peu sensible, lourde à mettre en place et donc inadaptée à une pratique clinique en routine. Des techniques de détection directes de constituants d'agents infectieux, réalisables à partir d'échantillons de patients, ont donc pris le relais avec la détection d'antigènes par Western blot. Elles ont été complétées par le développement de tests indirects sérologiques de type ELISA qui détectent la présence d'anticorps produits par le système immunitaire du patient en réponse à l'infection. Dans le cadre du VIH, les tests ELISA de 4<sup>ème</sup> génération combinent une détection d'antigènes du virus et d'anticorps du patient. C'est sur le principe de détection d'anticorps que reposent les tests rapides et les autotests VIH disponibles en pharmacie.

Un virage important a été pris dans les années 1990 avec l'arrivée de la biologie moléculaire. En permettant la détection de l'ADN ou de l'ARN d'un microorganisme, la PCR (et ses variantes RT-PCR, PCR quantitative, etc) a permis des gains de sensibilité très importants tout en améliorant souvent la spécificité du signal et avec un rendu des résultats en quelques heures. Il est par exemple possible aujourd'hui de détecter et de quantifier une charge virale du VIH à des seuils inférieurs à 100 copies d'ARN/mL de sang. Ces approches moléculaires ne sont toutefois rendues possibles que par la connaissance préalable de la séquence du génome du pathogène recherché, et deviennent inadaptées à la détection d'agents infectieux nouveaux ou inattendus.

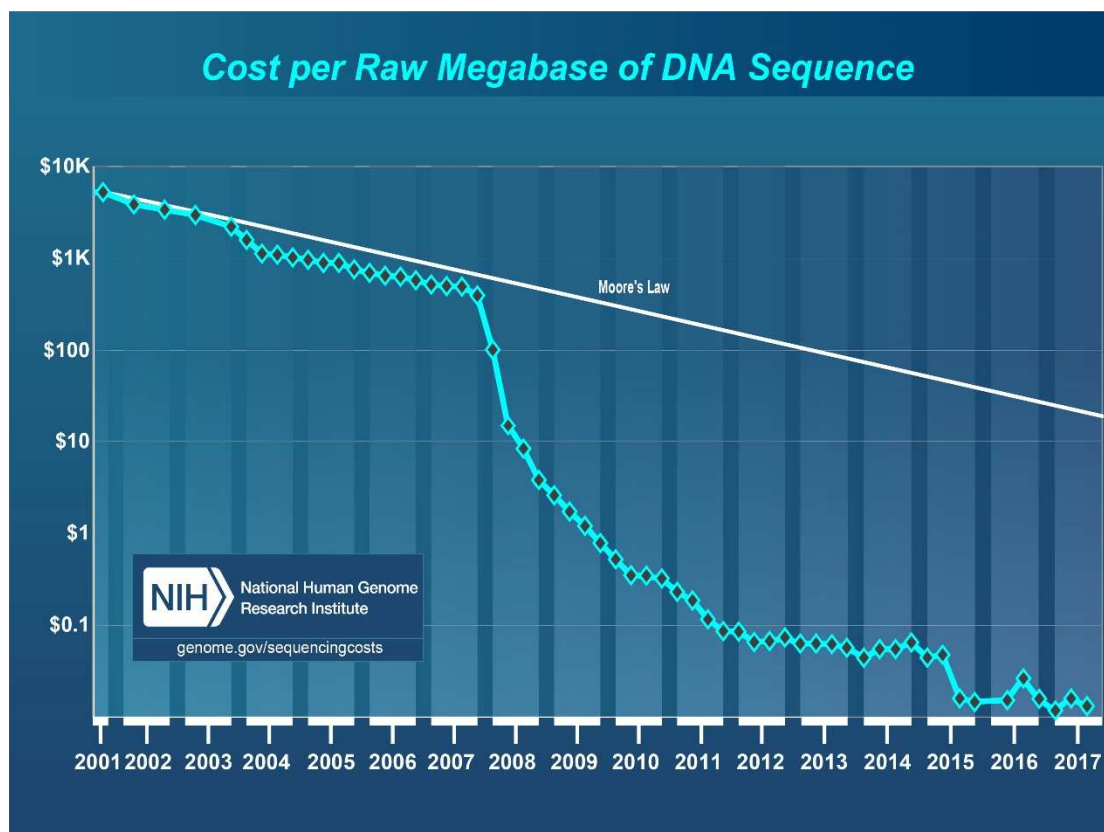


De la connaissance du statut sérologique de l'individu à la numération des lymphocytes T et aux tests de biologie moléculaire, les tests de Diagnostic *In Vitro* ont rendu possible la détection du virus en moyenne 15 jours après contamination, ainsi que la quantification de la charge virale circulante. (Source : SIDIV 2017)

## II/ L'apport du séquençage à haut débit (NGS)

### Séquençage à haut débit et science génomique

Le séquençage de l'ADN n'est pas une technique nouvelle. Différentes méthodes chimiques de lecture des bases ont été inventées aux Etats-Unis et au Royaume-Uni dans les années 1970, dont celle de Frederick Sanger basée sur la synthèse enzymatique qui reste encore aujourd'hui très utilisée. Le séquençage du premier génome humain (publié en 2001) reposait encore en partie sur des principes introduits par Sanger. Mais à partir des années 2005, une avancée technologique majeure a été réalisée avec la parallélisation des réactions chimiques de séquençage, permettant de révéler simultanément toute la diversité des molécules d'ADN présentes dans un échantillon. Cette seconde génération de méthodes, désignée par Next Generation Sequencing (NGS) ou encore High Throughput Sequencing (HTS) pour souligner le caractère de synthèse massivement parallèle, a ouvert la voie à la production de millions de séquences d'ADN à faible coût. C'est à partir de cette période qu'est née véritablement une science génomique, et que la recherche sur les génomes s'est popularisée en même temps qu'elle a basculé de problématiques strictement physico-chimiques ou biologiques vers de nouveaux enjeux liés à la production et la gestion de données numériques.



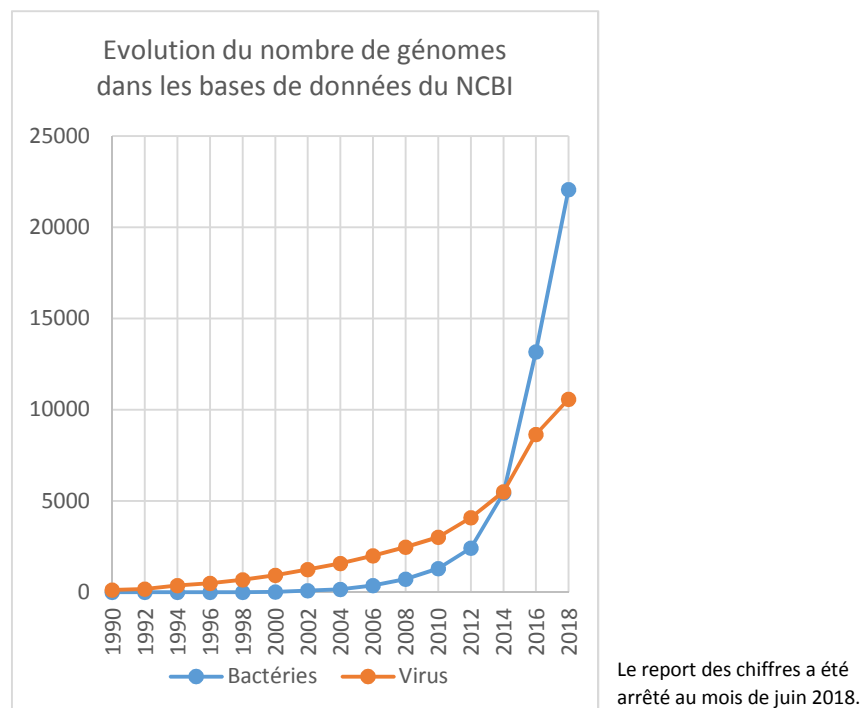
Le premier génome bactérien à avoir été complètement obtenu, celui de la bactérie *Haemophilus influenzae*, a nécessité un travail de plus d'un an pour un coût global estimé à 1 million de dollars<sup>2</sup>. Vingt ans plus tard, le même génome peut être obtenu en moins d'un jour pour moins de 100 dollars par un laboratoire spécialisé. Il en va bien sûr de la même tendance pour le génome humain et l'on en profitera pour souligner à ce sujet que certaines sociétés américaines, par exemple 23andMe, proposent aujourd'hui des analyses génomiques partielles au grand public pour quelques centaines de dollars, à partir d'un prélèvement de salive, avec envoi des résultats personnalisés par e-mail sous



quelques semaines. Les nombreuses applications et problématiques de l'utilisation de la génomique en matière de génétique humaine font actuellement l'objet de débats en France dans le cadre des états généraux 2018 de la bioéthique.

### Utilisation du NGS pour l'identification de pathogènes connus et défis méthodologiques

Si le séquençage du génome humain a reçu un fort écho médiatique, de nombreux génomes d'autres espèces et en particulier de microorganismes avaient été obtenus avant 2001 et ont continué à l'être depuis. L'arrivée du NGS a en particulier considérablement accéléré la production de données génomiques d'espèces virales et bactériennes à partir de prélèvements cliniques et environnementaux. Ces données de génomique microbienne, une fois publiées, deviennent d'accès public et gratuit au travers de grands portails internet de partage d'informations scientifiques expérimentales tels que le NCBI (côté américain) ou l'EMBL-EBI (côté européen). A titre d'illustration voici l'évolution du nombre de génomes de bactéries et de virus rendus accessibles sur le portail du NCBI depuis 1990 :



Cet ensemble d'informations, alimenté en permanence par la communauté scientifique mondiale, constitue une ressource publique de grande valeur pour le développement d'outils de diagnostic moléculaire. Par l'utilisation d'approches informatiques adaptées, il devient en effet possible de comparer les séquences générées par NGS à partir de prélèvements biologiques de malades (biopsies, sang, LCR ou autre) avec ces bases de données de séquences microbiennes, et d'établir ainsi la présence de microorganismes dans des échantillons cliniques, d'une manière totalement agnostique, c'est-à-dire ne reposant pas sur des hypothèses infectieuses prédéterminées.

Au-delà du seul aspect d'identification d'un pathogène, un diagnostic infectieux basé sur l'utilisation du NGS peut permettre une caractérisation moléculaire plus directe et complète du microorganisme incriminé, et rendre possible par exemple la prédiction d'un phénotype de résistance à certains

antibiotiques dans le cas d'une infection bactérienne, ou l'identification d'un génotype particulier dans le cas d'une infection virale. Cela peut également aider à retracer l'histoire d'une épidémie à différentes échelles, comme dans des infections nosocomiales ou lors de l'épidémie d'Ebola de 2014<sup>3,4</sup>. Cela peut aussi aider à anticiper des menaces infectieuses émergentes lorsque des pathogènes nouveaux, et donc encore non intégrés dans les tests courants, sont identifiés par une approche non ciblée. La quantification totale d'une flore microbienne complexe devient accessible au travers d'expériences de métagénomique, ce qui peut contribuer au diagnostic de certaines maladies digestives ou respiratoires par exemple. Enfin, un rendu de résultat d'analyse permettant d'exclure une cause infectieuse peut conforter dans la mise en place d'un traitement à base d'anti-inflammatoires dont l'impact immunodépresseur serait sinon nuisible au patient en cas d'infection.

En face de ce potentiel immense, les défis méthodologiques d'un diagnostic reposant sur l'utilisation des techniques de séquençage à haut débit n'en sont pas moins très importants. Il faut d'abord mettre en avant la capacité à travailler à partir d'un prélèvement clinique qui contient le pathogène responsable de la maladie. Cependant, contrairement aux approches historiques de cultures bactériennes par exemple, la détection de séquences d'un pathogène par NGS ne nécessite pas d'avoir un prélèvement du microorganisme vivant ou entier, mais *a minima* d'avoir dans la prise d'essai des fragments de son génome ou de ses transcrits. Cela est le cas également pour d'autres techniques moléculaires de référence comme la PCR ou RT-PCR, mais se décline d'une manière très particulière dans le cas du NGS : en effet, la capacité utile du séquençage étant répartie sur toute la diversité des séquences d'ADN présentes dans un échantillon, la sensibilité de détection dépendra avant tout de l'abondance relative des séquences d'un microorganisme par rapport aux séquences du génome de l'individu prélevé, consubstantielles au prélèvement. En d'autres termes il faut appliquer dans le choix des échantillons puis dans les modes de préparation des banques NGS un critère de rapport cible/bruit de fond qui n'est pas indispensable pour des approches de type PCR. L'accès à des prélèvements très ciblés, anatomiquement et temporellement, est donc un premier enjeu majeur.

La prévention des risques de contaminations d'un prélèvement clinique par de l'acide nucléique de l'environnement ou du manipulateur lui-même est tout aussi essentielle, dans la mesure où la moindre trace d'acide nucléique présente ou introduite dans un échantillon sera détectée, et pourra être interprétée comme une cause de la maladie. Cela implique la mise en place et le respect d'un certain nombre de règles rigoureuses dans l'organisation des laboratoires.

Il faut ensuite être en capacité de concevoir, exploiter, maintenir et améliorer des méthodes d'analyses bio-informatiques performantes permettant de traiter des volumes de données importants et en inflation. A titre d'exemple, une expérience NGS sur un échantillon clinique peut générer plusieurs dizaines de millions de séquences, qu'il faut pouvoir comparer aux dizaines de milliers de séquences des bases de données présentées plus haut. Cela requiert une ingénierie logicielle ainsi que des capacités matérielles de calcul, qui restent à l'heure actuelle un frein pour beaucoup d'infrastructures hospitalières ou universitaires.

Enfin, la question du temps de rendu des résultats, dont découle directement la capacité à choisir une option thérapeutique, et celle des coûts de la technologie, sont posées. Aujourd'hui et en situation optimale, il est possible de rendre un résultat d'analyse NGS en 48h à partir de la mise à disposition de l'échantillon. Cela peut donc être très profitable dans certaines circonstances, mais reste bien plus long que les quelques heures nécessaires à un rendu de résultat de PCR, et trop long par exemple pour les

besoins des unités de soins intensifs. Le coût des réactifs pour traiter un échantillon unique reste élevé, autour de 1500 euros à ce jour pour atteindre une profondeur de séquençage suffisante, mais il est possible de grouper plusieurs échantillons sur un même séquenceur pour réduire le coût par échantillon. Ces estimations ne reflètent toutefois pas le coût englobé qui inclue aussi l'environnement de travail et les salaires.

### **Apport du NGS pour l'identification de nouveaux pathogènes**

Le potentiel de détection d'un microorganisme par NGS dépend de la documentation préexistante de ce microorganisme dans les bases de données de séquences. Une question qui se pose est donc celle du degré de divergence au-delà duquel un microorganisme présent dans un échantillon biologique, mais éloigné par rapport aux bases de données, ne sera plus détecté faute de représentant connu suffisamment proche d'un point de vue phylogénétique. Il est en pratique possible de faire varier très largement les paramètres informatiques de comparaison de séquences afin d'autoriser un niveau de divergence plus ou moins grand, mais il faut certainement distinguer deux cas d'application :

D'une part, un diagnostic infectieux établi dans le cadre d'un rendu de résultat clinique, pour lequel il y a une forte légitimité à la fois pratique et réglementaire à se limiter à l'identification de pathogènes connus et bien caractérisés à l'aide de méthodes ayant fait l'objet d'une validation clinique.

D'autre part, une recherche de pathogènes plus distants du connu dans un contexte de recherche, qui peut venir en appui d'un diagnostic médical mais ne s'y substitue pas, et qui permet à la fois d'inclure la découverte de nouveaux microorganismes tout en permettant de faire évoluer les méthodes.

Ces deux approches ne sont en fait pas complètement exclusives, et un intérêt majeur du NGS est de pouvoir ré-analyser un jeu de données généré à un instant  $t$  à la lumière d'une nouvelle version de base de données ou d'un nouveau procédé informatique validé quelque temps plus tard, sans avoir à séquencer de nouveau l'échantillon. Cela offre l'avantage sur d'autres méthodes comme la PCR (qui nécessite par ailleurs de réutiliser une partie de l'échantillon clinique souvent précieux) de pouvoir se raccrocher à tout moment aux nouvelles ramifications de l'arbre de vivant.

### **Usages du NGS pour un diagnostic de terrain ou au lit du patient**

Pour compléter ce tour d'horizon rapide des enjeux liés à l'application des nouvelles techniques de séquençage dans un cadre de Diagnostic *In Vitro*, il faut souligner le virage vers la médecine ambulatoire pris par les systèmes hospitaliers depuis plusieurs années et la tendance générale qui va vers une médecine délocalisée et personnalisée, soit au lit du patient, soit sur le terrain en contexte de surveillance épidémiologique ou de réponse à des situations épidémiques. La transférabilité du modèle de diagnostic par NGS au plus près du malade a fait récemment l'objet d'une revue de référence très complète<sup>5</sup> dont on peut retenir ces quelques points importants :

Un premier besoin évident est de disposer de séquenceurs portables de petite taille pouvant être transportés et utilisés *in situ*. Depuis plusieurs années la tendance est à la réduction de la taille des séquenceurs par les fabricants, avec des modèles dits « benchtop » tels que les systèmes NextSeq, MiSeq, MiniSeq ou iSeq d'Illumina ou les systèmes Ion PGM et Ion S5 de ThermoFisher (voir plus bas

pour un comparatif des plateformes). Ces équipements ont un coût d'achat inférieur à 150k€ qui peut permettre à certaines structures locales d'utiliser la technologie NGS.

En rupture avec les formats existants, la société Oxford Nanopore Technologies a lancé en 2014 la commercialisation du MinION, un séquenceur de troisième génération de la taille d'une clé USB. Ce séquenceur se branche directement sur un ordinateur portable et son modèle économique diffère des séquenceurs benchtop dans la mesure où le hardware est en général gratuit et où l'utilisateur paie simplement les consommables (le prix par expérience restant malgré tout à ce jour assez élevé, entre 500 et 1000 euros, pour une profondeur de séquençage assez faible). Le MinION a par exemple été utilisé sur le terrain en Afrique durant l'épidémie d'Ebola<sup>6</sup> ou dans un laboratoire ambulatoire au Brésil dans le cadre du projet ZiBRA<sup>7</sup>. Ce séquenceur a aussi pu être testé avec succès dans des environnements extrêmes comme par exemple au sein de la Station Spatiale Internationale<sup>8</sup>.

Un deuxième aspect des choses est de réussir à coupler les données génomiques diagnostiques ou épidémiologiques avec des plateformes digitales de rendu de résultat et de partage de l'information. Un des défis est donc par exemple de sécuriser un accès à une connexion internet à l'endroit de l'analyse, dans le but de partager l'information en temps réel et d'avoir un diagnostic qui peut se traduire efficacement en actions thérapeutiques ou sanitaires. On peut d'ailleurs aller plus loin en imaginant des réseaux de séquenceurs portables déployés en différents endroits, hôpitaux, laboratoires, ou zones d'émergences, qui collectent et partagent des informations de détection de pathogènes en temps réel et les mettent à profit de la communauté par des plateformes internet et des systèmes d'alertes.

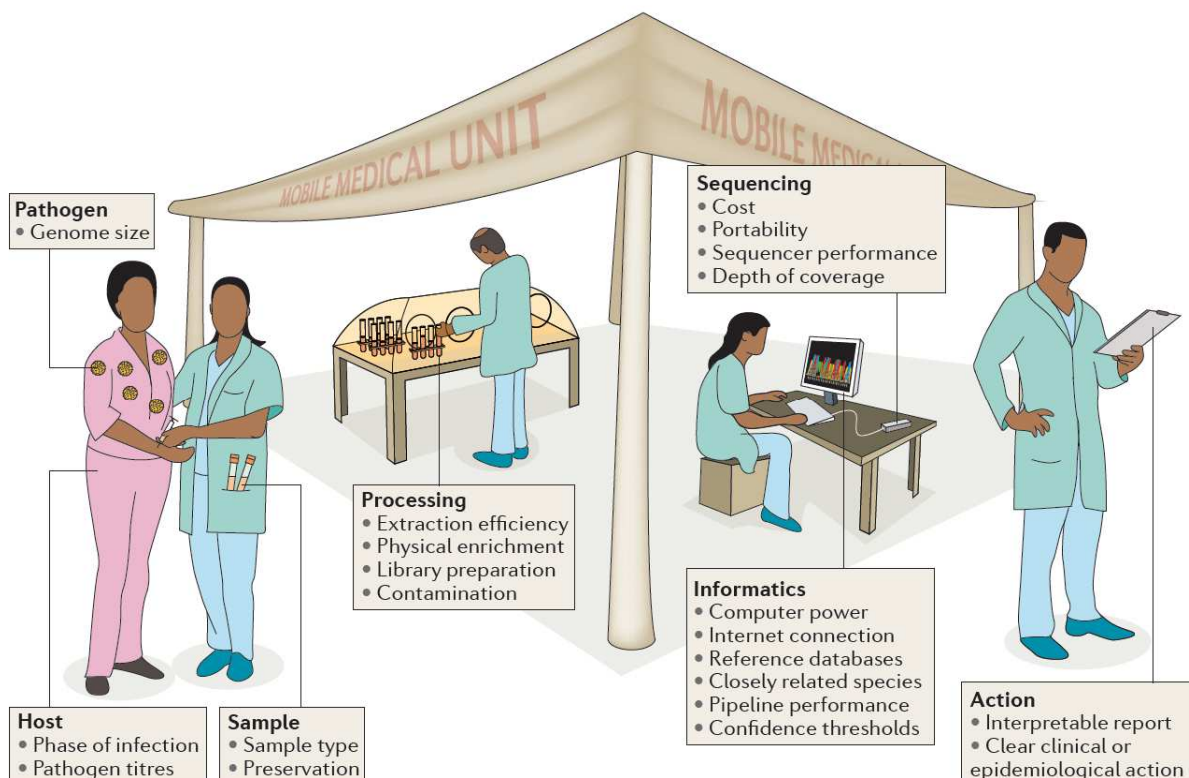


Illustration d'une unité médicale mobile ayant déployé un séquenceur portable à l'épicentre d'une épidémie. Différents enjeux se posent tout au long du processus de rendu de résultat diagnostique, allant du choix du prélèvement biologique initial au rendu de résultat des données. L'objectif final est de collecter des données numériques pouvant se transformer efficacement en recommandations thérapeutiques ou sanitaires. (Repris et adapté de la revue Towards a genomics-informed, real-time, global pathogen surveillance system<sup>5</sup>)

### III/ La bataille du « NextGen Dx »

Le rapport d'analyse Research and Markets "Genomics Market - Forecast to 2022"<sup>9</sup> prédit que le marché mondial de la génomique connaîtra une croissance annuelle de 10,2% pour passer de sa capitalisation actuelle de 14,71 milliards de dollars (en 2017) à 23,88 milliards de dollars en 2022. Cette croissance sera notamment structurée par une progression des activités de recherche en génomique, un nombre en augmentation de start-ups, et le développement d'applications de séquençage appliquées au diagnostic et à la médecine personnalisée. Il est anticipé que le segment de marché du diagnostic infectieux, avec celui des maladies génétiques, sera le plus important contributeur à cette croissance sur la période 2017-2022. Le contexte économique autour du diagnostic par NGS est donc extrêmement concurrentiel et nous allons présenter dans cette troisième partie les différents acteurs du domaine, des fabricants aux utilisateurs finaux, pour finir par une réflexion plus personnelle sur un modèle possible de service de métagénomique clinique en France.

#### Les sociétés qui fabriquent et commercialisent les séquenceurs et les solutions d'analyses

Le marché de la génomique est segmenté entre l'instrumentation et les systèmes d'analyses (séquenceurs, solutions logiciels), les réactifs (consommables chimiques et plastiques) et les services (prestations de séquençage, interprétation des résultats). Parmi les grandes sociétés positionnées sur l'un ou l'autre de ces segments on peut citer Illumina (USA), Thermo Fisher Scientific (USA), Pacific Biosciences of California (USA), Bio-Rad Laboratories (USA), Agilent Technologies (USA), Danaher (USA), Oxford Nanopore Technologies (UK), GE Healthcare (UK), Qiagen (Pays-Bas), Eurofins Scientific (Allemagne), Roche (Suisse) ou encore le Beijing Genomics Institute (Chine), parmi d'autres.

On compte à ce jour quatre sociétés leaders en matière d'instrumentation, qui couvrent l'essentiel du marché mondial. Elles proposent des séquenceurs de deuxième ou troisième génération dont les caractéristiques sont résumées dans le tableau suivant.

Company	Equipment	Output/run (Gb)	Maximum read length (bp)	Reads (x10 <sup>6</sup> )	Running time
Illumina	MiniSeq	0.6–7.5	2 × 150	25	4–24 h
Illumina	Miseq	0.3–15	2 × 300	25	5–55 h
Illumina	NextSeq	20–120	2 × 150	130/400	12–30 h
Illumina	HiSeq 3000	125–700	2 × 150	2500	<1–3.5 days
ThermoFisher	Ion PGM™	0.03–2	200–400	0.4–5.5	2–7 h
ThermoFisher	Ion 5S™	0.6–15	200–400	3–80	2.5–4 h
ThermoFisher	Ion 5S™ XL	0.6–15	200–400	3–80	<24 h
Oxford Nanopore	MinION	21–42	230,000–300,000	2.2–4.4	1 min–48 h
Pacific Biosciences	Sequel	0.75–1.25	>20,000	370,000	30 min–6 h
Pacific Biosciences	RSII	0.5–1	>20,000	55,000	30 min–4 h

Caractéristiques des plateformes NGS actuelles.

(Repris et adapté de « Application of next generation sequencing in clinical microbiology and infection prevention »<sup>10</sup>)

A côté de ces acteurs incontournables, il faut noter en France l'existence d'un petit nombre de start-up qui développent des séquenceurs ou des méthodes d'analyses génomiques pouvant s'appliquer à la détection de pathogènes et reposant sur des principes innovants. Par exemple, la société Genomic Vision, basée à Bagneux et fondée en 2004 par un chercheur de l'Institut Pasteur, associe une méthode de peignage moléculaire de l'ADN à un principe de lecture de code morse génomique. Genomic Vision

est cotée en bourse sur Euronext Paris depuis 2015 (valeur GV). A Paris, la société Depixus (anciennement Picoseq) décode l'information génétique et épigénétique à partir de molécules uniques d'ADN ou d'ARN en utilisant une méthode développée par des chercheurs du laboratoire de physique statistique de l'Ecole Normale Supérieure. Depixus a remporté la phase 3 du concours mondial de l'innovation en 2017, ce qui signifie qu'elle va avoir accès à des financements importants pour l'industrialisation de ses produits et que l'Etat français peut entrer au capital de la société.

## Les utilisateurs du NGS en DIV

Les utilisateurs des nouvelles approches de génomique se répartissent entre des centres de recherche, des institutions académiques ou gouvernementales, des hôpitaux, des sociétés pharmaceutiques ou biotechnologiques et certaines organisations non gouvernementales. En 2017 les hôpitaux représentaient le marché le plus important au niveau mondial, porté surtout par une demande en pharmacogénomique et en tests génétiques<sup>9</sup>.

Le diagnostic de maladies infectieuses par NGS en situation syndromique est à l'heure actuelle essentiellement pratiqué dans un contexte de recherche, en lien souvent avec un réseau hospitalier, et orienté prioritairement sur des cas cliniques graves ou en impasse thérapeutique compte tenu des coûts et des contraintes organisationnelles associées. Le bénéfice médical est donc souvent indirect, mais des exemples remarquables ont été publiés ces dernières années et permettent d'alimenter les réflexions pour une diffusion plus large de ces technologies en clinique.

On peut citer à ce titre la détection de pathogènes nouveaux ou inattendus dans des cas d'encéphalites humaines d'origine inconnue, qui ont conduit par exemple à l'identification d'infections par une leptospire<sup>11</sup>, par une amibe (*Balamuthia mandrillaris*)<sup>12</sup>, par un nouvel astrovirus neurotrope<sup>13-16</sup>, par un bunyavirus décrit comme une arbovirose<sup>17</sup> ou par un nouveau bornavirus transmis au contact d'écureuils d'élevage<sup>18</sup>. Dans le cas de la leptospirose, 38 tests diagnostiques différents avaient été réalisés et avaient échoué à identifier le pathogène avant que l'approche par NGS ne soit appliquée. Sur une thématique médicale un peu différente on peut retenir l'identification inattendue par NGS de la souche vaccinale du virus de la rubéole et de sa persistance dans des granulomes cutanés et viscéraux chez des enfants atteints d'immunodéficits primaires, résultat d'abord obtenu sur une cohorte française<sup>19,20</sup> puis confirmé par la suite aux Etats-Unis sur une cohorte indépendante<sup>21</sup>. D'autres études ont utilisé des approches de métagénomique clinique dans des cas de sepsis<sup>22</sup>, de pneumonies<sup>23</sup>, d'infections urinaires<sup>24</sup> ou d'infections intraoculaires<sup>25</sup>. Beaucoup de ces travaux ont été réalisés soit à l'Université de Californie à San-Francisco (groupes de Charles Chiu et de Joseph DeRisi), soit à l'Université Columbia de New York (groupe de Ian Lipkin), soit à l'Institut Pasteur de Paris (groupe de Marc Eloit), en lien étroit avec des hôpitaux partenaires rendant possible l'accès aux échantillons des malades.

A côté ou issues de ces groupes de recherche, des sociétés, notamment américaines, se développent en proposant des tests de détection de pathogènes par NGS. La société Karius par exemple propose un test sanguin reposant sur une base de données de plus 1000 pathogènes (bactéries, champignons, virus à ADN et parasites) dont les performances ont été évaluées lors d'une étude clinique sur 350 patients suspectés de sepsis<sup>26</sup>. Dans la version actuelle du test de Karius il faut tout de même noter

l'impossibilité de détecter des virus à ARN qui comprennent pourtant de nombreux pathogènes majeurs pour l'homme. Egalement implantée sur la baie de San Francisco, IDbyDNA, en lien avec les laboratoires ARUP, développe un test respiratoire (Explify) pour l'identification de pathogènes dans les infections pulmonaires, qui repose sur le séquençage d'ADN et d'ARN extraits de lavages broncho-alvéolaires, couplé à un outil d'analyse métagénomique original appelé Taxonomer<sup>27</sup>. En France, différents acteurs de l'industrie ou du service proposent des prestations d'identification de pathogènes par NGS s'appuyant soit entièrement sur une expertise interne, soit sur un modèle de délégation auprès de laboratoires partenaires. Ainsi, la société parisienne PathoQuest a développé son propre test NGS d'identification de pathogènes à partir de prélèvements sanguins (iDTECT Dx) qu'elle destine à certaines catégories de patients immunodéprimés, et qui est le premier test NGS de diagnostic de maladies infectieuses marqué CE. A Lyon, ViroScan 3D et ProfilXpert, ainsi que Bioaster, offrent des services de production de données génomiques et de découverte de pathogènes à des clients académiques ou industriels de la région. On peut mentionner aussi que bioMérieux a établi en 2016 un partenariat avec Illumina pour un service d'analyses NGS visant à améliorer la gestion des épidémies d'infections associées aux soins, en faisant réaliser l'étape de re-séquençage d'isolats à des laboratoires équipés de séquenceurs MiSeq (bioMérieux EpiSeq).

### **Un modèle français possible alliant médecins, scientifiques et industriels**

Il devient possible dans ce contexte de réfléchir à une offre de services ambitieuse articulant l'excellence scientifique, médicale et industrielle française.

Rappelons que la France, comme l'ont fait plusieurs grands pays avant elle, s'inscrit dorénavant dans un plan « Médecine France génomique 2025 », piloté par AVIESAN et soutenu par l'Etat, et dont le rapport ainsi que l'annonce des objectifs peuvent être consultés sur le site du gouvernement<sup>28</sup>. Dans ses grandes lignes le plan concerne (i) le déploiement d'un réseau de plateformes de séquençage à très haut débit, (ii) la mise en place d'un centre national d'analyse de données qui collectera l'ensemble des analyses réalisées sur les plateformes et (iii) la création d'un centre d'expertise et de veille technologique qui permettra d'harmoniser le fonctionnement du système et son évolution.

Dès 2018, trois projets pilotes seront lancés pour permettre l'accès au diagnostic génomique à des patients souffrant de cancers graves, de maladies rares ou de déficience intellectuelle, ainsi qu'aux diabétiques atypiques. Si ces premiers projets bénéficieront donc surtout à la cancérologie et à la génétique humaine, l'Etat soutiendra avec des moyens très significatifs l'ensemble des mesures du plan pendant les cinq prochaines années, avec pas moins de 400 millions d'euros investis. Une place pour le diagnostic des maladies infectieuses est donc certainement à construire.

Le dispositif du plan autour du réseau de plateformes de séquençage a retenu en 2017 deux premiers lauréats : le projet SEQOIA, porté par l'Assistance Publique – Hôpitaux de Paris, l'Institut Curie, l'Institut Gustave Roussy et l'Institut IMAGINE, avec pour partenaire industriel la société IntegraGen ; et le projet AURAGEN, porté par les Hospices Civils de Lyon, le CHU de Grenoble, le CHU de Saint-Etienne, le CHU de Clermont-Ferrand, le Centre Léon Bérard, le Centre Jean Perrin, l'Institut de cancérologie de la Loire et avec pour partenaire industriel le laboratoire Eurofins Biomnis.

Inscrire dans cette dynamique la détection de pathogènes en situations syndromiques ou de veille épidémiologique pourrait alors se décliner de la manière suivante :

Dans un premier temps, en s'appuyant sur le réseau hospitalier, identifier en situation de soin courant ou chez des personnes dont le pronostic vital est engagé des syndromes considérés comme infectieux mais d'étiologie inconnue. En cas de résultats non concluants obtenus par des méthodes diagnostiques standards, faire réaliser le séquençage à haut débit des échantillons des malades par les infrastructures en place et rendre un résultat de recherche pouvant venir en support au diagnostic médical. Lorsqu'un pathogène nouveau ou inattendu est identifié, poursuivre par une stratégie de valorisation avec les industriels du DIV dans le but d'améliorer ou de compléter les tests diagnostiques existants, à des fins d'utilisations futures en clinique.

Dans une perspective de plus long terme et en tablant sur une réduction des coûts du NGS et sur le développement des outils numériques, faire entrer cette pratique en première ligne diagnostique.

Ce modèle translationnel, à la fois Bed-to-Bench (du patient au laboratoire) et Bench-to-Bed (du laboratoire au patient), présente aussi l'avantage de partenariats public/privé possibles et effectifs dans la durée. Il pourrait se développer en capitalisant sur l'expérience de nos groupes de recherche nationaux et se structurer en intégrant par exemple les recommandations de la Food and Drug Administration en matière de diagnostic NGS des maladies infectieuses<sup>29</sup>.

## **Conclusion**

Les nouvelles méthodes de génomique sont une formidable source d'innovation dans le domaine du Diagnostic *In Vitro*. La question aujourd'hui n'est probablement plus de savoir si le séquençage à haut débit sera un jour utilisé en routine pour diagnostiquer les maladies infectieuses, mais plutôt d'arriver à anticiper la manière dont il le sera et de mettre en place dès à présent les structures de la médecine génomique de demain. Cela nécessitera la coordination d'acteurs du monde médical, scientifique et industriel, face à des problématiques d'un monde globalisé, dans un contexte de très forte compétition internationale dans lequel la France a de nombreux atouts à faire valoir.



## Références

1. Le Livre Blanc du SIDIV - 2017.
2. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
3. Faye, O. *et al.* Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect. Dis.* **15**, 320–326 (2015).
4. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
5. Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018).
6. Hoenen, T. *et al.* Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg. Infect. Dis.* **22**, 331–334 (2016).
7. Faria, N. R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* **8**, 97 (2016).
8. Castro-Wallace, S. L. *et al.* Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci. Rep.* **7**, 18022 (2017).
9. Genomics Market by Product and Services (Instruments/Systems, Consumables, Services), Technology (Sequencing, Microarray, PCR), Process (Library Preparation, Sequencing, Data Analysis), Application (Diagnostics) - Forecast to 2022.
10. Deurenberg, R. H. *et al.* Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* **243**, 16–24 (2017).
11. Wilson, M. R. *et al.* Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **370**, 2408–2417 (2014).
12. Greninger, A. L. *et al.* Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. *Genome Med.* **7**, 113 (2015).
13. Naccache, S. N. *et al.* Diagnosis of Neuroinvasive Astrovirus Infection in an Immunocompromised Adult With Encephalitis by Unbiased Next-Generation Sequencing. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2015).
14. Frémond, M.-L. *et al.* Next-Generation Sequencing for Diagnosis and Tailored Therapy: A Case Report of Astrovirus-Associated Progressive Encephalitis. *J. Pediatr. Infect. Dis. Soc.* **4**, e53-57 (2015).
15. Quan, P.-L. *et al.* Astrovirus Encephalitis in Boy with X-linked Agammaglobulinemia. *Emerg. Infect. Dis.* **16**, 918–925 (2010).
16. Brown, J. R. *et al.* Astrovirus VA1/HMO-C: An Increasingly Recognized Neurotropic Pathogen in Immunocompromised Patients. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2015).
17. Wilson, M. R. *et al.* A novel cause of chronic viral meningoencephalitis: Cache Valley virus. *Ann. Neurol.* **82**, 105–114 (2017).
18. Hoffmann, B. *et al.* A Variegated Squirrel Bornavirus Associated with Fatal Human Encephalitis. *N. Engl. J. Med.* **373**, 154–162 (2015).
19. Bodemer, C. *et al.* Live rubella virus vaccine long-term persistence as an antigenic trigger of cutaneous granulomas in patients with primary immunodeficiency. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **20**, O656-663 (2014).

20. Neven, B. *et al.* Cutaneous and Visceral Chronic Granulomatous Disease Triggered by a Rubella Virus Vaccine Strain in Children With Primary Immunodeficiencies. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **64**, 83–86 (2017).
21. Perelygina, L. *et al.* Rubella persistence in epidermal keratinocytes and granuloma M2 macrophages in patients with primary immunodeficiencies. *J. Allergy Clin. Immunol.* **138**, 1436–1439.e11 (2016).
22. Grumaz, S. *et al.* Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Med.* **8**, 73 (2016).
23. Pendleton, K. M. *et al.* Rapid Pathogen Identification in Bacterial Pneumonia Using Real-Time Metagenomics. *Am. J. Respir. Crit. Care Med.* **196**, 1610–1612 (2017).
24. Schmidt, K. *et al.* Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* **72**, 104–114 (2017).
25. Doan, T. *et al.* Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens. *Genome Med.* **8**, 90 (2016).
26. Blauwkamp, T. A. *et al.* Analytical Validation of a Microbial Cell-Free DNA Sequencing Test for Infectious Disease. Poster presented at ASM Microbe (2018).
27. Flygare, S. *et al.* Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* **17**, 111 (2016).
28. Plan France médecine génomique 2025 : discours d'Édouard Philippe. *Gouvernement.fr* <http://www.gouvernement.fr/partage/9345-plan-france-medecine-genomique-2025-discours-d-edouard-philippe>.
29. Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers. Draft Guidance for Industry and Food and Drug Administration Staff.