

# Séquençage à haut débit : outil de diagnostic des déficits immunitaires héréditaires

STÉPHANIE DROZ-GEORGET, PhD<sup>a</sup>, ORBICIA RICCIO, PhD<sup>a</sup>, BÉRYL ROYER-BERTRAND, PhD<sup>b</sup>, Prs ANDREA SUPERTI-FURGA<sup>b</sup> et FABIO CANDOTTI<sup>a</sup>

Rev Med Suisse 2017; 13: 763-6

**Etablir un diagnostic précis est souvent difficile dans le cas des déficits immunitaires héréditaires (DIH) car les manifestations cliniques sont hétérogènes et parfois atypiques ou communes à des maladies distinctes. Le séquençage à haut débit, aussi appelé next generation sequencing (NGS), permet d'identifier des variants génétiques responsables de caractéristiques cliniques. Le recours au NGS pour découvrir les gènes mutés dans les DIH ou connus pour être impliqués dans le développement, la différenciation et la régulation du système immunitaire, permet de cibler des centaines de gènes d'intérêt sur des patients suspectés de DIH pour lesquels le tableau clinique est documenté. Cette approche répond à la fois à des besoins de diagnostic et de recherche pour comprendre les mécanismes qui régissent les DIH et découvrir de nouveaux outils ou cibles thérapeutiques.**

## Next generation sequencing: a diagnostic tool for inherited immune defects

*Establishing the definitive diagnosis in the case of inherited immune defects (IID) is often challenging because the clinical features can be heterogeneous, atypical and overlapping different disease entities. The next generation sequencing technology (NGS) allows identifying genetic variants that are responsible for the observed clinical presentations. The use of NGS applied to the genes mutated in IIDs or known to be involved in the development, differentiation and regulation of the immune system allows to target hundreds of relevant genes in well characterized patients suspected of carrying inherited immune defects. This approach answers both diagnostic and research needs, facilitates the understanding of the mechanisms that underlie IIDs, and ultimately leads to the discovery of new therapeutic targets.*

## DIAGNOSTIC DES DÉFICITS IMMUNITAIRES HÉRÉDITAIRES

Les déficits immunitaires héréditaires (DIH) sont génétiquement déterminés et se caractérisent par une altération du nombre, de la fonction ou de la régulation des cellules du système immunitaire. Ils peuvent se présenter avec une sévérité allant de légère à létale, et provoquer une sensibilité accrue aux infections ou des réactions inflammatoires chroniques dont la fréquence, la localisation ou l'intensité sont anor-

males. Alors que la majorité des DIH se manifeste chez l'enfant, de plus en plus de cas sont diagnostiqués à l'âge adulte. De plus, les DIH peuvent être associés à des phénotypes autoimmuns, autoinflammatoires, allergiques et tumoraux.

Il existe plus de 200 formes cliniquement différentes de DIH. Il y a clairement plusieurs raisons pour lesquelles il est très important de poser un diagnostic précis, y compris les renseignements sur la sévérité de la maladie et l'anticipation de son évolution, la prescription de traitement curatif ou prophylactique approprié et l'indication des options thérapeutiques associées à un risque significatif, comme la transplantation de cellules souches hématopoïétiques ou l'enrôlement dans des essais cliniques (par exemple, thérapie génique). Avoir un diagnostic permet également d'offrir un conseil génétique aux patients et à leur famille, incluant l'évaluation du risque de transmettre la maladie en cas de grossesse, la possibilité de tester d'autres membres de la famille et de procéder à un diagnostic prénatal.

Plus de 300 gènes sont associés aux DIH, dont le mode de transmission peut être autosomal ou lié au chromosome X, récessif ou dominant.<sup>1</sup> Dans de nombreuses situations, les manifestations cliniques sont aisément attribuables à des mutations dans des gènes connus, associés à la maladie. Cependant, le tableau clinique est souvent hétérogène et peut aussi être atypique, ce qui rend le diagnostic difficile et retarde la mise en place d'une prise en charge optimale. Par exemple, des patients ayant la même mutation homozygote dans le gène *RAG2* peuvent avoir différents phénotypes, comme le syndrome d'Omenn (OMIM # 603554) ou le syndrome hyper-IgM (OMIM # 308230).<sup>2</sup> A l'inverse, des patients qui présentent des manifestations cliniques similaires, même typiques de syndromes bien décrits, comme le syndrome lymphoprolifératif autoimmun (ALPS), peuvent être porteurs de mutations dans des gènes différents (*FAS*, *CTLA4* ou *PIK3CD*).<sup>3-5</sup> Pour ces raisons, l'identification précise de la mutation génétique présente chez les patients reste de première importance.

## IDENTIFICATION DES MUTATIONS RESPONSABLES DES DÉFICITS IMMUNITAIRES HÉRÉDITAIRES

Les données cliniques, de laboratoire et l'histoire familiale permettent de suspecter un DIH dont les gènes responsables sont connus. Le diagnostic peut alors être confirmé en séquençant ce(s) gène(s) par la méthode classique de Sanger.<sup>6</sup> Cependant, cette méthode n'est pas suffisamment efficiente

<sup>a</sup> Laboratoire des déficits immunitaires héréditaires, Service d'immunologie et allergie, <sup>b</sup> Service de médecine génétique, CHUV et Université de Lausanne, 1011 Lausanne  
stephanie.droz-georget@chuv.ch | orbicia.riccio@chuv.ch | beryl.royer@chuv.ch  
andrea.superti-furga@chuv.ch | fabio.candotti@chuv.ch

pour investiguer de nombreux gènes simultanément. Si la cause génétique suspectée est exclue par cette méthode, l'approche peut s'élargir à des catalogues de gènes à haute valeur prédictive, groupés dans des panels correspondant à des catégories de maladies (par exemple, les immunodéficiences combinées sévères (IDCS), les DIH autoinflammatoires, etc.). Cette approche est appelée séquençage ciblé et a recours aux techniques plus puissantes de séquençage à haut débit (next generation sequencing (NGS)), qui permettent de lire des millions de segments ADN en parallèle. Si le choix de la catégorie clinique n'est pas évident ou est limité, le champ de la recherche de la mutation devrait s'élargir additionnellement vers le séquençage non biaisé des régions codantes du génome (c'est-à-dire l'exome) ou du génome entier, stratégie qui utilise aussi le NGS. Alors que le séquençage de l'exome est souvent approprié (85% des variants pathogéniques sont situés dans les régions codantes),<sup>7</sup> seule l'analyse du génome entier permet d'investiguer aussi les introns et régions promotrices qui peuvent être également impactés par des mutations pathogéniques. Ces deux dernières approches génèrent une grande quantité de données qui doivent être analysées et interprétées en considérant le tableau clinique. La technologie NGS a prouvé son efficacité dans le diagnostic des DIH et, dans le courant de cette décennie, a conduit à la découverte de 10 à 15 nouvelles anomalies génétiques associées aux DIH par année.<sup>1</sup>

## ANALYSE DE SÉQUENÇAGE À HAUT DÉBIT PAR SEMI-CONDUCTION

Plusieurs plateformes technologiques utilisent le NGS.<sup>8</sup> Le séquençage par semi-conduction, décrit ci-après, est l'un des plus fréquents. L'analyse du séquençage à haut débit se fait en plusieurs étapes (tableau 1). L'ADN du patient est extrait à partir d'un échantillon de sang ou d'autres tissus, et une banque de fragments ADN est générée en deux étapes. Tout d'abord, les régions génomiques d'intérêt sont ampli-

TABLEAU 1		Etapas principales du séquençage à haut débit	
PCR: réaction de polymérase en chaîne.			
<b>En laboratoire</b>			
Obtention de l'ADN	Extraction de l'ADN génomique du sang ou d'autres tissus du patient		
Génération des fragments génomiques à séquençer	Amplification des régions génomiques d'intérêt (gènes ciblés ou génome entier) par PCR		
Séquençage	Lecture du code génétique et génération des séquences (10 à 100 Gpb)		
<b>Analyse bioinformatique</b>			
Génération des fichiers de données	Alignement des séquences sur le génome de référence (p. ex. GRCh37/hg19) et génération de fichiers contenant les variants identifiés		
Analyses des données	Application de filtres sur les variants selon des paramètres prédéfinis (p. ex. qualité, fréquence allélique, mode d'héritabilité)		
<b>Interprétation des résultats</b>			
Discussion interdisciplinaire	Analyse de chaque variant relevant pour le diagnostic entre cliniciens, biologistes et bioinformaticiens		

PCR: réaction de polymérase en chaîne.

### En laboratoire

Obtention de l'ADN	Extraction de l'ADN génomique du sang ou d'autres tissus du patient
Génération des fragments génomiques à séquençer	Amplification des régions génomiques d'intérêt (gènes ciblés ou génome entier) par PCR
Séquençage	Lecture du code génétique et génération des séquences (10 à 100 Gpb)

### Analyse bioinformatique

Génération des fichiers de données	Alignement des séquences sur le génome de référence (p. ex. GRCh37/hg19) et génération de fichiers contenant les variants identifiés
Analyses des données	Application de filtres sur les variants selon des paramètres prédéfinis (p. ex. qualité, fréquence allélique, mode d'héritabilité)

### Interprétation des résultats

Discussion interdisciplinaire	Analyse de chaque variant relevant pour le diagnostic entre cliniciens, biologistes et bioinformaticiens
-------------------------------	--

fiées par réaction de polymérase en chaîne (PCR). Les fragments purifiés sont ensuite amplifiés une seconde fois avant d'être accrochés à des microsphères, chacune portant ainsi des millions de fois le même fragment. Ces microsphères sont réparties individuellement dans les puits d'une puce à séquençage ou « chip ». Une réaction de synthèse ADN a lieu ensuite dans chaque puits qui est rempli avec une solution contenant l'un des quatre nucléotides (A, T, C ou G) séquentiellement. L'incorporation d'un nucléotide par la polymérase se fait s'il est complémentaire à la séquence liée aux microsphères. Cela provoque la libération d'un proton et un changement de pH, information qui est traduite digitalement en une séquence.

Les séquences obtenues sont alignées sur le génome de référence (par exemple, GRCh37/hg19, <https://genome.ucsc.edu>), permettant l'identification de variants affectant un nucléotide unique ou des insertions/délétions de 2 à 16 paires de bases (pb) sur l'ADN du patient. Ces variants sont répertoriés sur des fichiers et analysés bioinformatiquement selon une série de paramètres incluant profondeur de séquençage, qualité du variant, fréquence allélique, impact prédit sur la fonction biologique, etc. Les variants retenus sont comparés à des bases de données permettant d'établir leur possible association à une pathologie connue (OMIM, ClinVar, CeGaT).

S'il y a une histoire familiale positive, il est opportun d'évaluer les données du patient en parallèle avec celles du ou des parent(s) sain(s) ou affecté(s), afin d'augmenter la sensibilité de l'analyse.

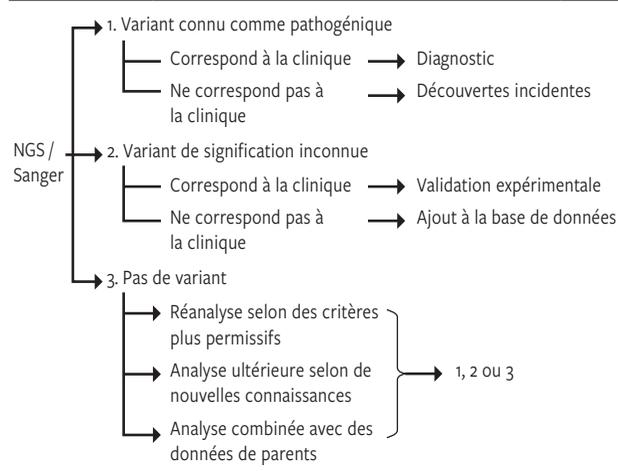
La pertinence des variants identifiés doit être interprétée par rapport aux informations cliniques du patient (figure 1). S'il s'agit d'un variant connu dans un gène impliqué dans une pathologie qui correspond au tableau clinique, le diagnostic peut être posé, après confirmation par une deuxième méthode de séquençage (Sanger). Si par contre, le tableau clinique ne correspond pas, le résultat est décrit comme découverte incidente.<sup>9</sup> Si le variant est nouveau, son impact sur la fonction biologique du gène doit être validé expérimentalement. Similairement, si le variant identifié est dans un gène pas encore associé à une pathologie, une validation expérimentale est nécessaire.

Dans le cas où l'analyse ne rend aucun variant relevant alors que les données techniques sont validées, il est possible de réanalyser les données avec des paramètres plus permissifs, ou de répéter l'analyse après que des nouvelles connaissances scientifiques et/ou bioinformatiques soient établies.

Si l'approche utilisée a été le séquençage ciblé, il se peut également que les causes génétiques ne soient pas incluses dans les régions génomiques amplifiées, si par exemple elles sont localisées sur des introns ou des régions promotrices. Dans ce cas, un recours à une stratégie non biaisée du séquençage de l'exome, ou du génome entier, est envisageable. Cependant, il est connu que pour des cohortes de cas non sélectionnés, jusqu'à 60-70% des cas resteront non résolus;<sup>10</sup> ceci est dû aux limitations de la technique qui ne couvre pas la totalité du génome et qui ne détecte ni les variations de structures

**FIG 1** Résultats possibles et interprétation d'une analyse de séquençage à haut débit

NGS: next generation sequencing.



génomiques (par exemple, duplications, délétions ou translocations) ni la variabilité du nombre de copies des gènes (figure 1).

## SÉQUENÇAGE CIBLÉ POUR LE DIAGNOSTIC DES DÉFICITS IMMUNITAIRES HÉRÉDITAIRES

L'approche du séquençage à haut débit ciblé a l'avantage de générer des données plus fiables pour les gènes d'intérêt par rapport au séquençage de l'exome ou du génome entier. En effet, en limitant l'investigation à un nombre inférieur de gènes, la technique NGS génère une couverture et une profondeur de séquençage plus élevées, c'est-à-dire que le pourcentage de chaque gène qui sera effectivement séquençé et la fiabilité du résultat seront améliorés.

Pour effectuer un séquençage à haut débit ciblé, des panels incluant des gènes connus pour être mutés dans les DIH sont conçus. Ces panels peuvent être globaux et contenir tous ces gènes, ou être limités à des sous-groupes (par exemple, les gènes responsables pour IDCS ou les déficits en anticorps). Ces panels sont commercialement disponibles ou développés auprès d'institutions académiques comme celui que nous avons élaboré au sein du Service d'immunologie et allergie du CHUV. Notre panel permet précisément de séquencer les exons et régions d'épissage d'environ 240 gènes déjà impliqués dans la physiopathologie des DIH et couvre également environ 120 gènes immunologiques considérés comme des candidats plausibles (tableau 2).

Par exemple, en utilisant ce panel, nous avons investigué le cas d'une patiente de 38 ans avec un diagnostic clinique d'immunodéficience commune variable (IDCV) et une histoire de cytopénies autoimmunes (thrombopénie, anémie hémolytique) et de pneumopathie lymphoïde interstitielle. Le séquençage à haut débit identifiait 1282 variants comparés au génome de référence. Les variants présents dans les exons et les régions d'épissage (1184) étaient retenus et ensuite filtrés pour exclure ceux fréquemment présents (fréquence allélique de plus de 1%) dans la population générale. Cela

**TABLEAU 2** Panel de NGS ciblé pour les DIH

NGS: next generation sequencing; DIH: déficits immunitaires héréditaires.

Catégories de DIH	Nombre de gènes	Exemples de gènes
Immunodéficiences combinées sévères (IDCS)	13	ADA, JAK3, RAG1, RAG2
Immunodéficiences combinées	33	CD40L, CD8A, ICOS, ZAP70
Immunodéficiences combinées avec caractéristiques associées ou syndromiques	37	ATM, BLM, TBX1, WAS
Déficits en anticorps	22	AICDA, BTK, IGHM, PIK3R1
Dérégulations du système immunitaire	36	AIRE, CECR1, FOXP3, IL10
Déficits congénitaux de la fonction et/ou du nombre phagocytaire	27	CYBB, G6PC3, ELANE, RAC2
Déficits en immunité innée et intrinsèque	28	CXCR4, IFNGR1, MYD88, STAT2
Maladies autoinflammatoires	13	MEFV, MVK, NLRP3, NLRP12
Déficits du système du complément	31	CFB, CFD, CFH, CFI
Phénocopies des DIH	3	NRAS, KRAS
Indéfinies	118	TLR1, TLR2, TRAF1, TRAF2

réduisait le nombre de variants à 123, nombre ultérieurement diminué à 81 après exclusion des variants connus comme faux positifs dans nos analyses. Ensuite, après avoir éliminé les variants synonymes, l'analyse résultait en 15 candidats dont 2 dans le gène *TNFRSF13B* (NM\_012452.2: c.542C>A: p. Ala181Glu et c.310T>C: p. Cys104Arg) étaient retenus comme pertinents sur la base du tableau clinique. *TNFRSF13B* code pour TACI (Transmembrane Activator and Calcium-modulator and cytophilin ligand Interactor) et des mutations hétérozygotes composites de ce gène sont connues pour être associées au diagnostic d'IDCV.<sup>11</sup> En conséquence, le séquençage à haut débit a permis de définir le diagnostic moléculaire de cette patiente.

## CONCLUSIONS ET PERSPECTIVES

Ces dix dernières années, le séquençage à haut débit a révolutionné l'approche diagnostique des maladies génétiques. Cette technologie est continuellement améliorée tant au niveau de la technique que de l'analyse bioinformatique. Par exemple, les connaissances acquises au cours des analyses des cas précédents sont utilisées pour affiner les paramètres analytiques des cas futurs. Cette approche systématique et rigoureuse reste toujours très dynamique et évolue dans le temps, ce qui peut permettre qu'une nouvelle analyse aboutisse au diagnostic d'un cas qui était auparavant non résolu.

Le NGS a prouvé son efficacité pour caractériser les bases génétiques des DIH et pour faciliter la définition des manifestations cliniques atypiques de ces maladies, ce qui permet la mise en place d'une prise en charge optimale.<sup>12</sup> Cette technologie promet aussi d'identifier des gènes responsables de maladies multifactorielles et de comprendre leurs mécanismes. Dans ce sens, le NGS stimule la recherche en immunologie en exigeant que la pertinence biologique de chaque variant

nouvellement identifié soit démontrée expérimentalement. Etant donné que la définition des mécanismes physiopathologiques promeut de manière fiable le développement de nouvelles options thérapeutiques, il est envisageable que l'utilisation du NGS continuera à faciliter les découvertes scientifiques dans le domaine des DIH et à soutenir les progrès dans leur gestion clinique.

**Remerciements:** Au Pr Pierre-Alexandre Bart pour les discussions cliniques. Ce projet est soutenu par la Gebert Rûf Stiftung (GRS 061/14).

**Conflit d'intérêts:** Les auteurs n'ont déclaré aucun conflit d'intérêts en relation avec cet article.

#### IMPLICATIONS PRATIQUES

- Poser un diagnostic moléculaire de déficits immunitaires héréditaires (DIH) est de première importance pour une prise en charge optimale et plusieurs facteurs pratiques pour le patient et sa famille (par exemple, planification familiale)
- Le séquençage à haut débit ciblé, de l'exome ou du génome entier, requiert moins de temps et de moyens financiers que le séquençage par la méthode classique de Sanger
- Le succès de l'approche est plus élevé dans les cas ayant une histoire familiale marquée
- Le séquençage à haut débit est en train de devenir un composant standard de l'approche diagnostique des DIH

1 \*\* Picard C, Al-Herz W, Bousfiha A, et al. Primary immunodeficiency diseases: an update on the classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. *J Clin Immunol* 2015;35:696-726.

2 Chou J, Hanna-Wakim R, Tirosh I, et al. A novel homozygous mutation in recombination activating gene 2 in 2 relatives with different clinical phenotypes: Omenn syndrome and hyper-IgM syndrome. *J Allergy Clin Immunol* 2012;130:1414-6.

3 Fisher GH, Rosenberg FJ, Straus SE, et al. Dominant interfering Fas gene mutations impair apoptosis in a human

autoimmune lymphoproliferative syndrome. *Cell* 1995;81:935-46.

4 Schubert D, Bode C, Kenebeck R, et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat Med* 2014;20:1410-6.

5 Lucas CL, Kuehn HS, Zhao F, et al. Dominant-activating germline mutations in the gene encoding the PI(3) K catalytic subunit p110delta result in T cell senescence and human immunodeficiency. *Nat Immunol* 2014;15:88-97.

6 Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463-7.

7 \*\* Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745-55.

8 \* Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics* 2011;38:95-109.

9 Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 2017;19:249-55.

10 Sawyer SL, Hartley T, Dymant DA, et al. Utility of whole-exome sequencing for

those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet* 2016;89:275-84.

11 Salzer U, Bacchelli C, Buckridge S, et al. Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. *Blood* 2009;113:1967-76.

12 Meyts I, Bosch B, Bolze A, et al. Exome and genome sequencing for inborn errors of immunity. *J Allergy Clin Immunol* 2016;138:957-69.

\* à lire  
\*\* à lire absolument