

Validation de méthode : Validation initiale approfondie de la méthode de séquençage haut débit pour la recherche de mutations constitutionnelles (analyse qualitative)

Date de rédaction : 09/05/2016

Date de révision :

Version : 1

Ce document a été préparé par les membres du GT4 « qualité et séquençage haut débit »

Groupe de travail : Cécile Aquaviva (Lyon), Christine Bellane-Chantelot (Paris), Stéphane Beziau (Nantes), Eric Bieth (Toulouse), Nelly Burnichon (Paris), Laurent Castera (Caen), Bénédicte Gérard (Strasbourg), Claude Houdayer (Paris), Eulalie Lasseaux (Bordeaux), Jean Muller (Strasbourg), Adrien Pagin (Lille), Nicolas Sevenet (Bordeaux), Cécile Saint-Martin (Paris), Dominique Vaur (Caen).

Sommaire

1	Présentation de la technique, de l'appareillage et du mode opératoire ; domaine d'application et but de la validation de méthode.....	2
1.1	Description de la méthode.....	2
1.2	Présentation de l'appareillage et du mode opératoire (enrichissement).....	3
1.3	Domaine d'application et but de la validation de méthode.....	3
2	Liste des critères de performance à vérifier et leurs limites acceptables choisies.....	4
2.1	Spécificité et Sensibilité analytiques : reflet de l'exactitude.....	4
2.2	Robustesse.....	4
2.3	Limite de détection.....	4
2.4	Contamination entre échantillons.....	4
2.5	Stabilité des réactifs.....	4
2.6	Comparaison avec une méthode de référence ou avec une méthode déjà utilisée au laboratoire.....	5
2.7	Confirmation des performances en pratique quotidienne : CIQ et EEQ.....	5
3	Références bibliographiques.....	6
4	Etude de risques.....	6
5	Plan d'expérience et mise en œuvre dans le laboratoire.....	7
5.1	Spécificité et Sensibilité.....	7
5.2	Robustesse.....	7
5.3	Contamination entre échantillons.....	7
5.4	Comparaison avec une autre méthode.....	7
6	Compilation et traitement statistique des données obtenues.....	7
7	Conclusion et décision quant à la validation opérationnelle de la technique, au regard des spécifications (limites acceptables) initialement fixées.....	7
8	Gestion de la portée flexible.....	8
8.1	La définition des régions analysées.....	8
8.2	Les performances attendues.....	8
8.3	La vérification de la sélectivité du test.....	8

1 Présentation de la technique, de l'appareillage et du mode opératoire ; domaine d'application et but de la validation de méthode

Ce dossier de validation initiale se base sur les recommandations du groupe de validation de méthode ANPGM (Association Nationale des Praticiens de Génétique Moléculaire) en ce qui concerne les spécifications attendues. Ce dossier ne s'intéresse qu'à la détection de mutations constitutionnelles (analyse qualitative).

Pour la détection de variations en mosaïque, il est nécessaire d'effectuer des tests complémentaires : répétabilité, fidélité intermédiaire, limite de détection. Il est nécessaire également d'évaluer dans ce cas le bruit de fond de la méthode analytique (taux d'erreur généré).

Pour chaque grand processus analytique, c'est à dire méthode d'enrichissement (ie PCR multiplex, capture ...) et chimie du séquenceur, un dossier de validation de méthode initiale est réalisé.

La qualification des automates de séquence (capacité à produire des séquences selon les spécifications du constructeur) doit être réalisée préalablement et les conditions environnementales nécessaires doivent être définies.

Le pipeline bioinformatique doit être maîtrisé.

Chaque test diagnostique utilisant cette méthode analytique fait ensuite l'objet d'un rapport spécifique sur un format de type SH form 43. Ce dossier de validation se base alors sur la vérification que le test répond aux exigences établies dans ce document, en fonction de son application en diagnostic.

Ce dossier s'appuie sur les dossiers de validation des extractions d'ADN, sur les analyses de risques du séquençage haut débit selon la méthode analytique d'enrichissement (AMDEC et 5M).

1.1 Description de la méthode

Ce dossier de validation couvre les étapes suivantes :

Donnée d'entrée : ADN de qualité suffisante

1. Etape de préparation de la librairie
2. Etape d'amplification clonale
3. Etape de séquençage proprement dit
4. Analyse des données brutes
5. Mise en évidence des mutations

Donnée de sortie : fichiers de séquences avec données d'intensité et de QV et liste de variants

Le séquençage est réalisé par la **méthode de séquençage haut débit (SHD)**. Le principe de ce séquençage repose sur l'analyse en parallèle de molécules d'ADN fixées individuellement sur support solide. Après fixation, une étape de multiplication clonale est réalisée puis une élongation d'amorce permet la détermination de la séquence de chaque clone individuellement. Les données enregistrées par le séquenceur (données de fluorescence ou de pH) sont ensuite interprétées puis traduites en base de l'ADN (*base calling*).

Les séquences déduites sont ensuite alignées sur le génome, et les variations par rapport au génome de référence sont identifiées (*variant calling*).

La recherche d'une mutation par séquençage haut débit est considérée comme relevant d'une méthode **qualitative basée sur des données quantitatives** : il s'agit de détecter la présence d'une variation qualitative de séquence nucléotidique par rapport à une séquence de référence.

La détermination du caractère homozygote, hémizyote ou hétérozygote pour une variation de séquence se fait en fonction de la détection ou non de l'allèle normal ou pathologique et du mode de transmission (gène lié à l'X ou gène autosomique).

Base normale	Base variante	Résultat
+	-	Profil normal
+	+	hétérozygote
(-)	+	Homozygote muté ou hémizyote

Cette méthode analytique est adaptée à la détection des SNV (*single nucleotide variant*) et des petites indels (de taille variable selon la chimie de séquençage utilisée). La détection des variations de structures du génome (tels que les variations du nombre de copie ou CNV) taille fait l'objet d'une analyse bioinformatique séparée et d'un dossier de validation séparé.

1.2 Présentation de l'appareillage et du mode opératoire (enrichissement)

1.2.1 Présentation du séquenceur

Présentation succincte du PGM, Miseq ...

Le logiciel XXX va effectuer différentes étapes :

- soustraction du bruit de fond
- analyse des variations pH

Une valeur qualité (QV) est attribuée à chaque base séquencée (ratio signal spécifique/bruit de fond) ; elle est le reflet de la probabilité d'erreur d'assignation de la base à cette position.

L'appareil utilisé est qualifié pour une analyse de type qualitative.

Les performances opérationnelles du séquenceur sont vérifiées à chaque lancement de l'automate (*run*) par le suivi de paramètres établis :

Par exemple,

- Analyse d'un contrôle propre au séquenceur (*PhiX, contrôle spécifique séquenceur*)
- Analyse du nombre de variants moyens détecté pour chaque patient
- Nombre de bases lues
- Analyse d'un patient contrôle
- Analyse des QV30+ sur les élongations (*Q plot*)

1.2.2 Présentation des logiciels d'analyse

Les séquences sont ensuite analysées sur un pipeline bioinformatique capable :

- d'aligner la séquence produite sur une séquence de référence
- d'identifier les variants et de les nommer par rapport aux différentes isoformes des gènes cibles
- d'interpréter les variants de séquence en termes de variation nucléotidique, protéique et d'épissage

Ce pipeline doit être maîtrisé. Il fait l'objet d'un dossier de qualification séparé. Le numéro de version du pipeline utilisé pour ce dossier de validation de méthode doit être précisé.

1.2.3 Présentation des modes opératoires associés

Citer les MOPE

1.3 Domaine d'application et but de la validation de méthode

Ce dossier de validation concerne la recherche de variant inconnu en génétique constitutionnelle par séquençage selon la méthode analytique suivante : [*méthode d'enrichissement/chimie du séquenceur/séquenceur*].

Le but de ce dossier de validation est de démontrer que les modes opératoires mis en place et les appareillages utilisés génèrent des données de qualité suffisante pour pouvoir par la suite appliquer cette méthode en diagnostic, c'est à dire permettre la détection de mutations inconnues à l'état homozygote, hémizygote ou hétérozygote.

Test qualitatif ; **Type D selon Mattocks et coll.**

La recherche d'une mutation par séquençage est considérée comme relevant d'une méthode **qualitative** : il s'agit ici de détecter la présence d'une variation qualitative de séquence nucléotidique par rapport à une séquence de référence.

Il est conseillé d'analyser des mutations pathogènes difficiles (stretch de polyA, grande del ou ins, régions répétées) dans le cadre de l'évaluation de la robustesse de la méthode mais non pour l'évaluation de la sensibilité, spécificité de la méthode. A adapter en fonction du panel de gènes étudiés et du profil mutationnel de ces gènes.

2 Liste des critères de performance à vérifier et leurs limites acceptables choisies

2.1 Spécificité et Sensibilité analytiques : reflet de l'exactitude

Les faux positifs ou faux négatifs en séquençage haut débit (SHD) peuvent être liés soit à la phase analytique proprement dite (wet lab), soit au pipeline bioinformatique (dry lab). Cette seconde partie est traitée dans le dossier de qualification du pipeline bioinformatique.

Les spécificité et sensibilité analytiques du SHD sont évaluées dans des conditions optimales de séquençage, c'est à dire des conditions compatibles avec un rendu diagnostique. Ainsi, seules sont retenues pour l'analyse les variations couvertes au moins 30X (valeur à fixer par le laboratoire en fonction de son panel). Sont exclus les artefacts techniques récurrents.

Analyse d'un nombre suffisant de variations différentes (>30) de type SNV et de type indel à partir d'au moins un échantillon de génotype connu (par exemple, analysé en Sanger, ADN de référence analysé par exome ou génome ...).

Critères retenus :

Critères d'acceptabilité : la concordance doit être de 100 % pour les SNV ou les indel (pas de faux positif ou faux négatif analytique détecté), pour des variations couvertes au moins 30 X (cf supra)

2.2 Robustesse

La robustesse de la technique peut être évaluée sur les paramètres suivants :

- **ADN : qualité (dégradation), quantité, type d'extraction, tissu initial...**

Critères retenus : Pourcentage de bases couvertes >30X comparable à celui obtenu avec un ADN de bonne qualité

- **Mutations pathogènes connues et « difficiles » (régions répétées, riche en CG, insertion Alu ...)**

Critères retenus : Détection correcte de la mutation.

Rq : ces mutations difficiles peuvent être suivies en CQI

2.3 Limite de détection

Non applicable sauf si la détection de mutation en mosaïque est garantie. Dans ce cas, des analyses de répétabilité, reproductibilité, limite de détection de la méthode doivent être effectuées.

Par exemple, réalisation d'un mélange (ex ratios 9.1) de 2 ADN de génotypes connus et différents et analyse avec recherche de détection des variants minoritaires (si le pipeline bioinformatique le permet).

2.4 Contamination entre échantillons

Une étude de contamination doit être réalisée au moins une fois au moment de la validation initiale en réalisant un blanc suivant tout le parcours de l'échantillon.

Critères retenus : Analyse du nombre total de lecture (reads ou séquence) du témoin négatif :

- le nombre de reads doit être < au seuil désigné par le laboratoire,
- la taille moyenne des reads doit être <75-100pb (valeur à définir en fonction des plateformes, des panels et techniques).

2.5 Stabilité des réactifs

Des recommandations générales ont été mises en place dans le laboratoire pour éviter une dégradation de la qualité des séquences réalisées liée à un problème de stabilité des réactifs. Ainsi, la température et la durée de conservation des produits intermédiaires sont définies dans les modes opératoires.

Le suivi de CIQ permet de détecter une anomalie de réactifs.

Un test d'acceptabilité doit être réalisé pour toute nouvelle réception de réactif.

Par exemple,

- *Le suivi des valeurs Phred (Q plot)*
- *Le nombre de variant détecté en moyenne par patient*
- *Le pourcentage de couverture moyen > 30X par patient*
- *L'analyse de SNV connus chez le patient (cf. identitovigilance)*
- *Analyse d'un patient contrôle ajouté dans chaque série*
- *Analyse du contrôle interne du séquenceur (PGM, Miseq ...)*

2.6 Comparaison avec une méthode de référence ou avec une méthode déjà utilisée au laboratoire

Dans le dossier initial, une analyse par comparaison avec une méthode de référence (ADN analysé par Sanger, ou par séquençage haut débit par un laboratoire déjà accrédité...) doit être effectuée sur un nombre suffisant de SNV et petites indel (>30). L'analyse des discordances doit être faite le cas échéant.

Critères retenus : La concordance doit être de 100 % pour les SNV ou les indel. En particulier, les régions difficiles sont à définir et à documenter (ex : homopolymères, riche en GC, etc.)

2.7 Confirmation des performances en pratique quotidienne : CIQ et EEQ

2.7.1 Description et suivi des CIQ

Par exemple,

- *Vérification de la performance opérationnelle du séquenceur (passage d'un CIQ séquenceur « fabricant » ; suivi du nombre total de reads générés ...)*
- *Vérification de la stabilité des réactifs*
 - o *Identification correcte des SNV connus chez un patient contrôle (contrôle positif)*
 - o *Identification correcte des SNV connus chez les patients testés (test d'identitovigilance)*
- *Vérification de l'absence de contamination intra et interéchantillons*

L'absence de contamination doit être suivie par une méthode appropriée. Par exemple,

- o *Passage d'un blanc à chaque série (contrôle négatif)*
- o *En cas de patient contrôle (contrôle positif), obtention du bon génotype*
- o *La proportion des reads non attribués aux index utilisés (autres index et index non identifiés) doit être inférieure à 5% du total des reads.*
- o *Vérification de l'absence de contamination sur les SNV homozygotes*
- o *Vérification de l'absence de contamination sur du gène SRY chez des patients de sexe féminin*
- o ...

Définir la conduite en cas d'anomalie observée sur le CIQ.

2.7.2 Suivi des EEQ

Un EEQ NGS est disponible (EMQN)

Mise en place d'échanges interlaboratoires

3 Références bibliographiques

- Practice guidelines for Sanger Analysis and Interpretation (*Clinical Molecular Genetics Society* 2009)
- Norme NF EN ISO 15189
- Guide technique d'accréditation de vérification (portée A) / validation (portée B) des méthodes en biologie médicale – SH GTA 04
- Mattijs et coll. Eurogenetest 2014 Guidelines for diagnostic next generation sequencing
- Mattocks CJ et coll. EuroGentest Validation Group. *Eur J Hum Genet.* 2010 Dec;18(12):1276-88. A standardized framework for the validation and verification of clinical molecular genetic tests.
- SH form43 : fiche type de vérification (portée A)/Validation de method (portée B) d'une method de biologie médicale

4 Etude de risques

Les **faux positifs** analytiques (génération physique d'une mutation qui n'existe pas initialement dans l'échantillon d'ADN) peuvent être liés à :

- *Erreur de l'ADN polymérase au moment de la PCR initiale (erreur pouvant être présente sur plusieurs clusters)*
- *Erreur de l'ADN polymérase au moment de l'amplification clonale (théoriquement limitée à quelques molécules d'un seul cluster et donc éliminée par le filtre initial mais possiblement majoritaire dans le cluster)*
- *Erreur de l'ADN polymérase lors de la phase de séquençage (théoriquement aléatoire et donc non retenue pour l'analyse bioinformatique)*
- *Amplification ou capture artéfactuelle d'une séquence homologue (pseudogène par exemple)*

Ces faux positifs peuvent être liés à des compositions en base particulières et ainsi conduire à des artefacts techniques récurrents.

Les **faux négatifs** analytiques (défaut de détection d'une mutation présente initialement dans l'échantillon d'ADN dans les régions analysées) peuvent être liés à :

- *« Allele drop out » lors de la PCR initiale (par ex, ADN en trop faible quantité, amplification d'une région difficile, SNV sur une amorce PCR)*
- *« Allele drop out » lors de l'élongation (par ex, polymorphisme en cis de la mutation créant une structure secondaire sur le simple brin). Ceci créé un biais de brin.*
- *Amplification ou capture artéfactuelle d'une séquence homologue (pseudogène par exemple)*

Ces faux négatifs peuvent être également lié à un défaut de couverture sur la base considérée (probabilité de détecter un hétérozygote = 94% si 4X, 99,9% si 10X) mais il ne s'agit pas ici d'un faux négatif analytique proprement dit.

En l'absence d'automatisation, le risque d'interversion étant important (étapes multiples de transvasements), il est recommandé de mettre en place un contrôle d'identité-vigilance (sonde taqman, SnapShot, analyse du trio, séquence Sanger).

Voir document Etude de risque 5M

Voir document AMDEC

5 Plan d'expérience et mise en œuvre dans le laboratoire

5.1 Spécificité et Sensibilité

Echantillons	
Méthodologie et critères retenus	
Résultats expérimentaux	
Conclusion / limites	

5.2 Robustesse

Echantillons	
Méthodologie et critères retenus	
Résultats expérimentaux	
Conclusion / limites	

5.3 Contamination entre échantillons

Echantillons	
Méthodologie et critères retenus	
Résultats expérimentaux	
Conclusion / limites	

5.4 Comparaison avec une autre méthode

Echantillons	
Méthodologie et critères retenus	
Résultats expérimentaux	
Conclusion / limites	

6 Compilation et traitement statistique des données obtenues

7 Conclusion et décision quant à la validation opérationnelle de la technique, au regard des spécifications (limites acceptables) initialement fixées

8 Gestion de la portée flexible

Chaque dossier de validation en NGS des panels diagnostiques doit contenir les items suivants :

8.1 La définition des régions analysées

Liste des gènes explorés avec leur l'identifiant de la séquence de référence (ex : NM_XXXXXX), 3' ou 5'UTR, exons +/- nb de bases introniques, promoteur...

Les mutations additionnelles détectées si sondes ou amorces spécifiques pour ces mutations.

La version du génome sur laquelle se base le design

La liste des gènes de classe A (couverture >30X garantie pour toutes les régions analysées), de classe B (95% de couverture garantie à 30X), de classe C (pas de couverture garantie) (Mattijs et coll. Eurogenetest 2014 Guidelines for diagnostic next generation sequencing)

8.2 Les performances attendues

Cette méthode analytique est adaptée à la détection des SNV et petites indels (*à préciser selon la chimie de séquençage utilisée*).

8.3 La vérification de la sélectivité du test

Pour les tests par amplicon :

- Blast sur l'ensemble du génome
- Pour les amorces PCR, vérification de l'absence de SNV fréquents (> 1%) au minimum sur les 5 dernières bases

Pour les tests par capture :

- Visualisation de la position des sondes gène/gène.
- Après un premier run, alignement sur génome complet pour voir les trous de couverture (0X) et les couvertures insuffisantes (seuil à déterminer par le laboratoire, par exemple 30X)
- Identification des régions « non analysables » en raison de trop fortes homologies et adaptation de l'analyse bioinformatique

-L'étude des couvertures moyennes et des absences de couverture (0X et <30X) sur au moins 3 échantillons normaux.

-Le cas échéant, la sensibilité et la spécificité sur des échantillons contrôles (EEQ, ADN de référence, échanges interlaboratoires, ADN analysés en Sanger).