



# Guide technique d'accréditation de la technologie de séquençage à haut débit (NGS)

SH GTA 16 - Révision 00

LA VERSION ELECTRONIQUE FAIT FOI





## SOMMAIRE

1	OBJET .....	4
2	REFERENCE ET DEFINITIONS .....	4
2.1.	REFERENCES .....	4
2.2.	ABREVIATIONS ET DEFINITIONS .....	5
2.2.1.	DEFINITIONS .....	5
2.2.2.	ABREVIATIONS .....	7
4	DOMAINE D'APPLICATION .....	7
5	MODALITES D'APPLICATION .....	7
6	CONTEXTE NORMATIF ET REGLEMENTAIRE .....	8
7	L'ANALYSE PAR TECHNIQUE NGS : ETAPE PAR ETAPE .....	8
8	GUIDE DE LECTURE DES EXIGENCES NORMATIVES ET RECOMMANDATIONS .....	9
4.1	Responsabilités en matière d'organisation et de management .....	9
4.3	Maîtrise des documents .....	9
4.5	Examens transmis à des structures sous-traitantes .....	10
4.6	Services externes et approvisionnement .....	10
4.7	Prestation de conseils .....	10
4.9	Identification et maîtrise des non-conformités .....	11
4.13	Maitrise des enregistrements .....	11
5.1	Personnel .....	12
5.2	Locaux et conditions environnementales .....	13
5.3	Equipements .....	13
5.4	Processus préanalytique .....	18
5.5	Processus analytique .....	19
5.6	Garantie de qualité des résultats .....	20
5.7	Processus post-analytique .....	21
5.8	Compte-rendu des résultats .....	22
5.10	Gestion des informations de la structure .....	23
9	MODALITES D'EVALUATION DES LABORATOIRES DE BIOLOGIE MEDICALE ET STRUCTURES D'ACP POUR LES ANALYSES .....	24
	Annexe 1 : Qualification du pipeline bioinformatique au travers de la maîtrise des risques .....	25
	Cas n°1 : La structure réalise toute l'analyse en interne .....	25
	Cas n°2 : Analyse fournisseur « Clef en main » .....	26



Cas n°3 : La structure fait appel à un service support informatique appartenant à la même entité juridique .....27

Annexe 2 : DESIGN, DEVELOPPEMENT ET MAITRISE DU PIPELINE .....28

Annexe 3 : Recommandations relatives au contenu du dossier de qualification .....29

Annexe 4 : Exemple d'étapes de l'analyse bioinformatique à maîtriser pour la détection de variants .....31

Annexe 5 : Exemple d'étapes de l'analyse bioinformatique à maîtriser pour la détection de microorganismes .....33

LA VERSION ELECTRONIQUE FAIT FOI



# PREAMBULE

La technologie de séquençage à haut débit est de plus en plus présente dans les structures. Les informations contenues dans ce guide sont le fruit de la réflexion collégiale de biologistes médicaux et bioinformaticiens issus de structures privés et publics, de membres du comité de section et de la commission d'accréditation de la section Santé humaine, de représentants des sociétés savantes et de l'Agence de la Biomédecine, d'évaluateurs et d'experts techniques du Cofrac, spécialisés dans la technologie NGS et dans différents domaines comme la génétique humaine, l'immunologie cellulaire spécialisée et l'histocompatibilité, la microbiologie.

## 1 OBJET

La norme NF EN ISO 15189 définit les exigences particulières concernant la qualité et la compétence des laboratoires de biologie médicale (LBM) et des structures d'Anatomie et de Cytologie Pathologiques (ACP).

Ce guide ne se substitue pas aux exigences et/ou normes applicables au sein de la structure. Les recommandations qu'il contient et que la structure est libre d'appliquer, sont celles reconnues par le COFRAC comme étant appropriées pour répondre aux exigences de la norme NF EN ISO 15189, ainsi qu'au document COFRAC SH REF 02 correspondant. Dans tous les cas, la structure doit démontrer que les dispositions prises permettent de satisfaire pleinement aux exigences.

Ce document est destiné à fournir des recommandations pour répondre aux exigences du référentiel d'accréditation pour réaliser les examens par séquençage à haut-débit, et s'assurer de la maîtrise de la phase d'analyse bioinformatique. Le présent guide explicite notamment les principales exigences de la norme NF EN ISO 15189 permettant d'assurer la qualité des prestations fournies par les structures. Pour répondre aux exigences de cette norme et du document d'exigences SH REF 02, ce guide s'appuie sur des publications et des recommandations issues des sociétés savantes basées sur les bonnes pratiques professionnelles admises dans le domaine. De plus, il précise les modalités d'évaluations appliquées à chacune des organisations des structures identifiées.

## 2 REFERENCE ET DEFINITIONS

### 2.1. REFERENCES

Ce document prend en compte les documents suivants :

1. Norme NF EN ISO 15189
2. Document SH REF 02 « Exigences pour l'accréditation selon les normes NF EN ISO 15189 et NF EN ISO 22870 »
3. Document SH REF 08 « Expression et évaluation des portées d'accréditation »
4. Document SH GTA 01 « Guide technique d'accréditation en biologie médicale »
5. Document SH GTA 04 « Guide technique d'accréditation de vérification (portée A)/validation (portée B) des méthodes en biologie médicale »



6. *Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN*: Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn* 2017, 19(3):341-365
7. *Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, Leon A, Pullambhatla M, Temple-Smolkin RL, Voelkerding KV, Wang C, Carter AB*: Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn* 2018, 20(1):4-27
8. Conception de logiciels pour le diagnostic clinique par séquençage haut-débit, collection Outils pour la pratique, INCa, février 2018.

## 2.2. ABREVIATIONS ET DEFINITIONS

### 2.2.1. DEFINITIONS

**Bases de données** (*database* en anglais) : Structures informatiques (ex sql, prostgre, mangoDB...voire excel, csv,...) permettant de stocker et de retrouver l'intégralité de données brutes ou d'informations en rapport avec un thème ou une activité ; celles-ci peuvent être de natures différentes et plus ou moins reliées entre elles. Les bases de données peuvent être incluses dans le pipeline et entrer dans le processus analytique.

**Bioinformaticien** : Personne, formée et qualifiée, en charge de la bioinformatique qui présente des compétences bioinformatiques en fonction du contexte et de l'organisation de la structure et qui :

- est le garant du cahier des charges fixant les objectifs (performances et finalités) de l'environnement bioinformatique infrastructure et pipeline (cf définition ci-dessous), et participe à la définition du cahier des charges,
- s'assure, en fonction de l'organisation des procédures bioinformatiques de la structure, de la création, de la mise en place, la mise en œuvre et/ou la maintenance du pipeline pour l'analyse NGS,
- peut être l'interlocuteur avec les services spécialisés ou les fournisseurs, le cas échéant.

**Infrastructure** : Schéma général du flux et du traitement des données dans l'environnement de production. La structure doit définir l'ensemble des outils informatiques utilisés dans le processus analytique.

**Interface utilisateur de validation** : Interface graphique destinée à des utilisateurs (biologistes, techniciens) pour leur permettre la gestion des analyses et la validation des résultats. Elle peut intégrer à différents niveaux de l'analyse : la gestion des runs, le lancement des pipelines, la réalisation des contrôles qualité, permettre le tri de variants, leur annotation voire la validation des résultats.

**Librairies** : Production de molécules d'acides nucléiques à séquençer.

**Métagénomique** : Méthode d'étude du contenu génétique d'échantillons issus d'environnements complexes. Elle permet l'étude de séquences de génomes de plusieurs individus d'espèces différentes dans un milieu donné. La métagénomique pan pathogène correspond à une technique dans laquelle les bactéries, les champignons et les virus sont recherchés.



**Métrique qualité** : Donnée mesurable, enregistrée à l'entrée, en des points intermédiaires, à la sortie du pipeline, permettant d'évaluer ses performances à l'initial et en continu, pour chaque run et pour chaque échantillon.

**Panel** : Ensemble de régions ou cibles d'intérêt séquencées de façon simultanée selon une méthode définie.

**Pipeline** : Chaîne de traitement informatique automatisée qui est constituée d'un ensemble de logiciels, succession de programmes permettant d'analyser les données brutes du séquenceur, de les assembler et de les traduire en séquences interprétables. Ces logiciels ou programmes peuvent être développés en interne ou être d'origine externe (commerciale ou opensource).

**Qualification** : La qualification d'un matériel consiste à vérifier dans l'environnement de production de la structure la conformité aux spécifications attendues de cet équipement.

**Recherche ciblée** : Analyse de séquençage haut-débit spécifique de gènes ou régions d'ADN connus.

**Recherche non ciblée** : Le séquençage haut-débit lorsqu'il est utilisé pour de la recherche non ciblée correspond :

- à l'identification d'une séquence connue mais non cible (en génétique avec la détection d'une variation incidente ou avec le séquençage de l'exome ou du génome entier d'un patient ; en microbiologie avec la détection d'un pathogène non cible) ou
- à l'identification de séquences non connues (en génétique avec l'identification de répertoires immunitaires ; en microbiologie avec l'identification d'un pathogène non décrit mais présentant une forte homologie de séquence avec un pathogène de même famille, ...)

**Run** : passage d'un test composé d'un ensemble d'échantillons séquencés en parallèle sur le séquenceur NGS.

**Script** : programme chargé d'exécuter une action pré-définie. Il s'agit d'une suite de commandes simple permettant l'automatisation de certaines tâches successives dans un ordre donné.

**Séquençage haut-débit** : Cette méthode est plus proprement appelée séquençage massif en parallèle (ou Massive Parallel Sequencing : MPS). Bien que moins approprié, l'acronyme le plus couramment utilisé est NGS (Next Generation Sequencing). Par commodité d'usage, ce dernier sera donc utilisé dans le corps du document pour désigner le séquençage haut-débit.

Le principe de cette méthode repose sur l'analyse simultanée de plusieurs dizaines de kilobases à plusieurs centaines de gigabases aboutissant à la génération d'un très grand nombre de données brutes informatiques, traduites en séquence d'ADN, alignées à des génomes de référence.

**Wiki** : Application web qui permet la création, la modification et l'illustration collaborative de pages à l'intérieur d'un site web.



## 2.2.2. ABREVIATIONS

CNV : Copy Number Variation  
DVM : Dossier de vérification/validation de méthode  
ESHG : European Society for Human Genetics  
ET : Evalueur Technique  
LBM : Laboratoire de Biologie Médicale  
LP : Ligne de portée  
NGS : Next Generation Sequencing  
SMQ : Système de Management de la Qualité  
SNV : Single Nucleotide Variant  
SV : Structural Variation

## 4 DOMAINE D'APPLICATION

Ce document s'applique :

- aux laboratoires de biologie médicale (LBM) et aux structures d'Anatomie et de Cytologie Pathologiques (ACP), accrédités ou en démarche d'accréditation selon la norme NF EN ISO 15189 ;
- aux évaluateurs techniques et experts techniques du COFRAC : il constitue à ce titre une base d'harmonisation à leur usage ;
- aux membres des instances du COFRAC (Comité de Section, Commission d'Accréditation Santé Humaine) ;
- aux membres de la structure permanente du COFRAC ;
- aux clients des structures accréditées ;
- aux structures exécutant des examens/analyses relevant du champ d'application de la section Santé Humaine (dont les domaines d'analyses médico-légales, de biologie humaine, ...) et qui le jugeraient utiles.

## 5 MODALITES D'APPLICATION

Ce document est applicable à compter du 1<sup>er</sup> octobre 2019.

Au jour de son approbation, ce guide technique d'accréditation reflète l'état d'avancement des connaissances en termes de séquençage à haut débit et de traitement des données bioinformatiques et au regard des organisations existantes des structures. Le terme structure utilisé tout au long du document s'applique au laboratoire de biologie médicale et à la structure d'Anatomie et de Cytologie Pathologiques (ACP).

Ce document contient à la fois des exigences et des recommandations.

Le terme « **doit** » est utilisé pour exprimer une exigence. Les exigences correspondent à la retranscription des exigences de la norme d'accréditation, du prescripteur ou de la réglementation, ou relèvent des règles d'évaluation et d'accréditation du Cofrac.

Le terme « **devrait** » est utilisé pour exprimer une recommandation. Les recommandations relèvent de bonnes pratiques visant à satisfaire les exigences. L'organisme est libre de ne pas suivre la recommandation, s'il peut démontrer que les dispositions alternatives qu'il met en œuvre satisfont l'exigence d'accréditation correspondante.



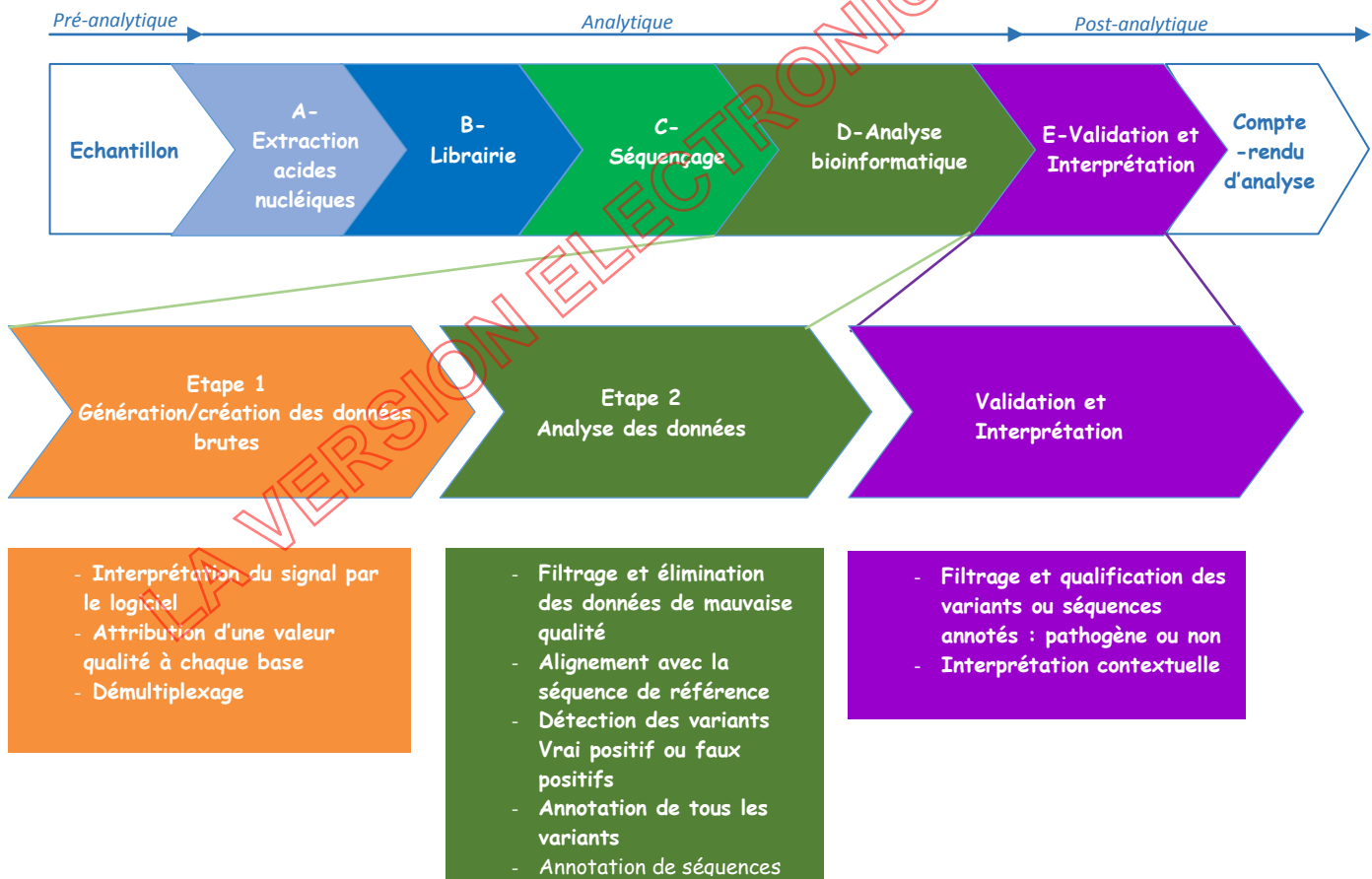
## 6 CONTEXTE NORMATIF ET REGLEMENTAIRE

Pour la réalisation de l'analyse NGS sous accréditation, les structures doivent répondre à l'ensemble des exigences de la norme NF EN ISO 15189 et du document d'exigences SH REF 02. Cependant, certaines de ces exigences nécessitent d'être explicitées au regard des spécificités de la technologie NGS et de la maîtrise des différentes phases de l'analyse. Les recommandations relatives à ces exigences spécifiques sont reprises dans la suite du document.

**Les paragraphes de la norme NF EN ISO 15189 non cités dans le présent document sont considérés comme explicites et ne nécessitent pas d'indication complémentaire par rapport au texte de la norme NF EN ISO 15189 et du document SH REF 02.**

## 7 L'ANALYSE PAR TECHNIQUE NGS : ETAPE PAR ETAPE

L'analyse par technique NGS peut être découpée en différentes étapes qui sont représentées dans le schéma suivant et, en particulier, celles de l'analyse bioinformatique.



Les recommandations présentées dans ce guide s'appuient sur l'enchaînement des étapes précitées  
NB : cet exemple ne présente pas une liste d'étapes exhaustive ou obligatoire.





## 8 GUIDE DE LECTURE DES EXIGENCES NORMATIVES ET RECOMMANDATIONS

Le plan de ce chapitre et sa numérotation sont identiques à ceux de la norme NF EN ISO 15189. Sauf précisions contraires, la référence des paragraphes renvoie à ceux de la norme NF EN ISO 15189 ou du document SH REF 02.

Les encadrés de couleur insérés dans le texte ont pour but d'expliquer ou d'illustrer certaines exigences pour les domaines spécialisés de la

Génétique

ou de la Microbiologie.

### Exigences relatives au management

#### 4.1 Responsabilités en matière d'organisation et de management

Le directeur de la structure est responsable des prestations proposées par la structure. Il doit s'assurer de la disponibilité et de l'adéquation des ressources pour le bon déroulement de toutes les phases de l'analyse. Le directeur de la structure peut déléguer certaines tâches ou responsabilités pour lesquelles il ne serait pas compétent, à des personnes compétentes/spécialistes de sa structure. Cependant, il doit rester responsable des prestations finales et des résultats rendus quelle que soit l'organisation mise en place par la structure (exemples : dans le cas d'une organisation faisant appel à un service spécialisé appartenant à la même entité juridique que la structure, dans le cas d'une plateforme mutualisée au sein d'un établissement de santé...).

Au regard des exigences du § 4.1.2, dans le cas d'une structure qui fait appel à un service support spécialisé (en informatique par exemple) du groupement auquel il appartient, ce service doit être géré par la structure comme un prestataire externe en s'assurant de la qualité de ces prestations au travers d'accords négociés (cf §4.6). Pour s'assurer de la conformité de la situation, le service spécialisé est évalué en même temps que la structure dans le cas où elle fait partie de l'entité juridique accréditée et que les informations ne sont pas disponibles au sein de la structure.

Au travers des accords négociés avec le service support ou au sein de la structure, la direction doit s'assurer d'avoir les ressources nécessaires pour réaliser les examens (cf §4.1.2.1 h et i) pour répondre aux utilisateurs (§4.1.2.2).

#### 4.3 Maîtrise des documents

La structure dispose de modalités décrivant toutes les phases de l'analyse. Les documents associés à la gestion de l'analyse NGS, y compris la phase bioinformatique, doivent être intégrés à la gestion documentaire de la structure. La gestion documentaire pour la phase bioinformatique peut s'appuyer sur des outils adaptés à l'informatique qui offrent des options de suivi, des wiki, de la documentation du code, et qui permettent le dépôt des sources et le suivi de leurs versions.

Dans ce cadre, la structure devrait documenter :

- l'architecture générale de circulations des données NGS au sein de la structure et le cas échéant, d'interactions entre les différents outils,
- la définition des différents outils bioinformatiques ainsi que leur « versionnage » et la documentation fournisseur associée.



#### 4.5 Examens transmis à des structures sous-traitantes

La transmission d'un échantillon biologique à une autre structure pour la réalisation de l'analyse jusqu'à l'interprétation des résultats associés (matérialisée par un compte-rendu validé et signé par un biologiste) doit être considérée comme une situation de sous-traitance. La structure doit sélectionner et évaluer la structure sous-traitante selon des critères définis en fonction de ses besoins (cf tableau ci-dessous) et lui permettant de s'assurer que le sous-traitant est en mesure de réaliser les examens et de satisfaire les exigences du client. Dans certains secteurs d'activités, la structure « transmetteur » doit être attentive à respecter les exigences réglementaires en vigueur.

<b>Exemples de critères de sélection et d'évaluation de structure sous-traitante</b>
<i>Structure accréditée pour cet examen</i>
<i>Performances analytiques, performances bioinformatiques</i>
<i>Modalités de gestion des données bioinformatiques, confidentialité, prise en compte des problèmes et délai de résolution</i>
<i>Mise en place d'un système conforme aux exigences de la norme NF EN ISO 15189</i>
<i>Suivi des recommandations des sociétés savantes pour les informations du compte-rendu et modalités d'interprétation</i>
<i>Définition des dispositions en cas d'anomalie</i>

#### 4.6 Services externes et approvisionnement

La structure doit sélectionner, évaluer et approuver les fournisseurs de matériels tels que les équipements (matériels informatiques, serveurs, etc...), les programmes (logiciels de traitement des données, d'interprétation des données...), et les automates (séquenceurs, robots de pipetage, ...) selon des critères définis (cf §5.3). Pour rappel, dans le cas d'une structure qui fait appel à un service support spécialisé (en bioinformatique par exemple) du groupement auquel il appartient, ce service doit être géré par la structure comme un prestataire externe en s'assurant de la qualité de ces prestations au travers d'accords négociés (cf §4.1.2). La structure devrait s'assurer de la compétence du service support périodiquement, par exemple au travers d'audits internes.

La structure doit s'assurer que la confidentialité est maintenue notamment pour le traitement de données patients.

#### 4.7 Prestation de conseils

La structure doit s'assurer de la compétence des biologistes en matière de prestation de conseils sur ces examens et des modalités de communication des résultats (cf § 5.1 et 5.8) notamment dans le cas d'examens spécifiques (exemples : exome, génome complet, examen spécialisé de génétique pour l'HPV circulant en ACP, ...).

La structure doit informer le prescripteur des techniques utilisées et de leurs limites. Dans certains secteurs d'activités, la structure doit être attentive à respecter les exigences réglementaires en vigueur, notamment concernant les données incidentes et secondaires.



## 4.9 Identification et maîtrise des non-conformités

La structure doit s'assurer de la qualité et la maîtrise des prestations fournies par le biais de la gestion des non-conformités (exemples : évaluation du service support en bioinformatique, ...). Lorsqu'une non-conformité peut se reproduire, la structure doit prendre des mesures visant à identifier, documenter et éliminer la ou les causes profondes.

Dans le cadre de l'analyse bioinformatique, le traitement d'erreurs dans une des phases de l'analyse doit être réalisé par le biais de la gestion des non-conformités. La structure doit mettre en place des actions permettant de les corriger et pouvant impliquer des modifications potentielles dans les scripts, les types de logiciels, leur paramétrage, les bases de données...

### 4.13 Maîtrise des enregistrements

L'utilisation du pipeline bioinformatique qualifié doit être réalisée en adéquation avec le SMQ de la structure selon la norme NF EN ISO 15189 (traçabilité des événements, non-conformité, paramètres du pipeline, manuel utilisateur, ...).

Pour répondre aux exigences, la structure devrait documenter :

- le dossier de qualification initiale et continue du pipeline bioinformatique (notes techniques des versions successives en développement et/ou suite à la mise en production, paramétrages définis, filtres, valeurs seuils des métriques, la méthode de sélection et le versionnage des bases de données, le versionnage des algorithmes, du génome de référence, ...),
- les enregistrements concernant la calibration, la maintenance des instruments, les contrôles de qualité, les comparaisons inter-laboratoires, les réactifs critiques, ...,
- les enregistrements des données informatiques telles que les fichiers FASTQ, BAM, VCF les comptes-rendus finaux,
- les enregistrements concernant les qualifications du personnel concepteur, utilisateur pour toutes les étapes critiques de l'analyse,
- les enregistrements relatifs à la gestion des risques liée à la conception et/ou à l'utilisation des outils bioinformatiques.

La durée de conservation des documents et enregistrements doit être conforme à la réglementation en vigueur. La structure doit conserver les enregistrements des données brutes et de toutes les informations nécessaires, lui permettant une exploitation ultérieure potentielle des résultats.

#### 4.14.5 Audit interne

La structure doit s'assurer que :

- l'ensemble de ses activités fait l'objet d'audits internes,
- le choix des auditeurs est en adéquation avec le domaine d'application. Il peut nécessiter des compétences en bioinformatique en fonction de l'organisation choisie.

#### 4.14.6 Gestion des risques

Une analyse doit être menée pour identifier les risques associés à chacune des étapes critiques de l'analyse, des phases pré, per et post-analytiques, et pour déterminer les actions à mettre en place pour éliminer ces risques.

Une analyse spécifique de l'étape bioinformatique et des différentes phases (exemples donnés en Annexes 4 et 5) qui la composent intégrant la circulation des fichiers NGS, leur traitement et leur



stockage, doit être conduite au préalable de la mise en production sur le sous-processus bioinformatique. La structure doit identifier et établir les moyens de maîtrise correspondants.

Cette analyse de risque associée à chaque examen fait partie du dossier de vérification/validation des méthodes (cf §5.5).

## Exigences techniques

### 5.1 Personnel

Le personnel (biologistes, techniciens, bioinformaticiens, ...) doit être formé et qualifié spécifiquement sur l'ensemble de la technologie NGS, du pré au post-analytique.

La structure doit s'assurer notamment de la compétence du personnel utilisateur pour :

- la réalisation du processus analytique (par exemple lors de l'utilisation de matériel partagé)
- l'interprétation des séquences (alignement, relecture/rejet, ambiguïtés de lecture si séquençage en simple brin,...),
- l'utilisation et/ou la gestion des données,
- la connaissance des logiciels/paramétrages logiciels,
- l'identification d'anomalie ou de situation anormale sur l'ensemble du processus.

Le biologiste compétent sur l'activité NGS doit posséder également les compétences bioinformatiques requises ou s'appuyer sur l'expertise d'un bioinformaticien interne ou au sein d'un service support, en fonction de l'organisation de la structure. La reconnaissance des compétences de ce biologiste devrait faire l'objet d'une qualification spécifique qui peut être établie sur la base de formation(s), diplôme(s), et/ou de l'expérience professionnelle (validation d'examens NGS, connaissance des outils bioinformatiques : logiciels entrant dans la composition des pipelines, paramétrage de ces logiciels, bases de données publiques ou commerciales, etc.).

Les missions bioinformatiques en particulier devraient comporter (liste indicative et non exhaustive) :

- la définition des besoins de la structure (cahier des charges, ...),
- l'utilisation des bases de données, des logiciels et paramètres associés,
- l'accompagnement des phases de qualification du pipeline bioinformatique et/ou de design & développement si réalisé en interne,
- l'analyse des risques,
- l'évaluation et la sélection des logiciels du pipeline,
- le paramétrage des logiciels et l'approbation des seuils retenus,
- le choix des échantillons de référence appropriés en regard du type de test envisagé en diagnostic,
- la sélection des métriques et critères de performance, l'approbation des seuils retenus,
- la qualification et l'approbation finale des outils utilisés,
- le maintien du système y compris en cas de modification du pipeline,
- la prise de contact avec le service support et/ou le fournisseur, le cas échéant, ainsi que la qualité des prestations fournies.

Au sein de la structure, le biologiste compétent sur cette activité doit interagir avec le personnel possédant une expertise spécialisée en bioinformatique (bioinformaticien). En collaboration avec le bioinformaticien, il doit définir les besoins de la structure, participer et superviser le processus de développement des outils bioinformatiques. Il doit approuver le processus de qualification des outils bioinformatiques, sélectionner et approuver les fournisseurs, superviser et approuver le design et les résultats de la validation de méthode, suivre les modifications des outils et évaluer leur impact sur les résultats, le cas échéant.



## 5.2 Locaux et conditions environnementales

L'accès aux locaux doit être contrôlé et réglementé. Ils ne doivent être accessibles qu'aux personnes autorisées. L'accès au poste d'analyse doit permettre le respect de la confidentialité et doit être protégé (par exemple par un mot de passe).

Les locaux et les conditions environnementales mises en place sont similaires à ceux de la biologie moléculaire. Une vigilance importante doit être mise en place afin d'éviter les risques de contamination des échantillons. La séparation entre les phases pré-PCR et post-PCR (paillasse d'extraction, de pré-PCR, de post-PCR) est toutefois définie au regard du processus NGS, des automates et analyseurs utilisés. En fonction des techniques et équipements utilisés, la structure devrait définir les dispositions de cette séparation sur la base d'une analyse de risques détaillée pour s'assurer de la qualité des prestations fournies (par exemple : automatisation fermée, maîtrise des contaminations, ...). Pour les séquenceurs et les autres matériels informatiques utilisés pour réaliser les analyses, la structure doit s'assurer du respect des conditions environnementales au regard des spécifications fournisseurs.

L'infrastructure informatique doit être sécurisée (cf document SH GTA 02 et référentiels existants dans le domaine). De plus, les serveurs de calcul et tous les outils informatiques doivent être intégrés dans une infrastructure maîtrisée en termes de confidentialité et d'accessibilité, permettant un suivi environnemental, si besoin, et leur maintenance.

## 5.3 Equipements

Les équipements de séquençage haut-débit doivent répondre aux exigences de la norme en termes d'installation, de qualification, de maintenance, y compris dans le cas de matériel partagé. En fonction des séquenceurs utilisés, la structure doit mettre en place les qualifications, maintenances et raccordements métrologiques nécessaires en adéquation avec son statut d'automate ouvert ou fermé et en fonction des recommandations des fournisseurs.

L'étape de traitement bioinformatique de l'analyse NGS nécessite pour les structures de développer ou d'adopter différents outils informatiques tels que les équipements informatiques, le pipeline bioinformatique, les bases de données de variants, les logiciels, etc., qui sont considérés comme du matériel au sens de la norme. De ce fait, ces équipements sont soumis aux exigences du §5.3 et notamment le § 5.3.1.2.

De même, le pipeline bioinformatique nécessite une qualification initiale avant la réalisation des étapes de vérification/validation des méthodes.

### 5.3.1.2 Essai d'acceptation de l'équipement

Concernant la phase bioinformatique de l'analyse, cet item de la norme correspond à la qualification du pipeline bioinformatique.

Un **pipeline bioinformatique** est une chaîne de traitement de données qui admet des fichiers de séquence (bcl, fastq...) et émet des fichiers de données annotées (vcf, csv, xls...). Les différents outils informatiques utilisés (logiciels et matériels) dans le cadre de la phase de traitement et de préparation à l'interprétation des données font partie intégrante de la phase analytique (cf. Logigrammes 1 et 2 : exemples de pipelines pour la génétique ou la microbiologie). La validation de méthode doit se baser sur la qualification au préalable du ou des pipelines bioinformatiques.

Dans la suite du document, la notion de qualification sera appliquée plus spécifiquement au pipeline bioinformatique (tel que défini ci-dessus). La structure devrait définir les outils utilisés dans le pipeline

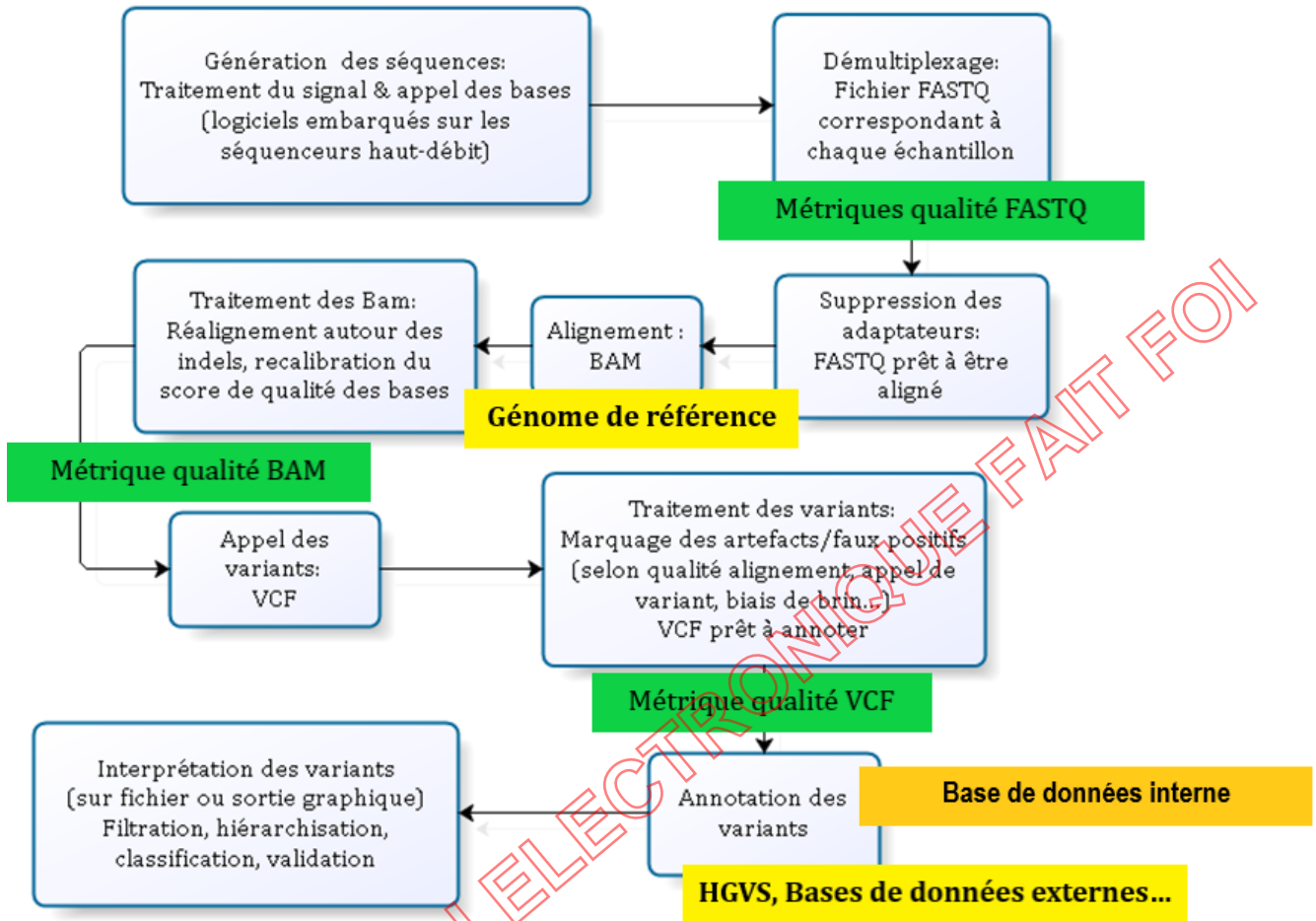


bioinformatique et les qualifier. Il s'agit de réaliser la qualification de la chaîne de traitement entière. Il n'est pas nécessaire de qualifier chaque outil indépendamment ainsi que chaque liaison entre les outils.

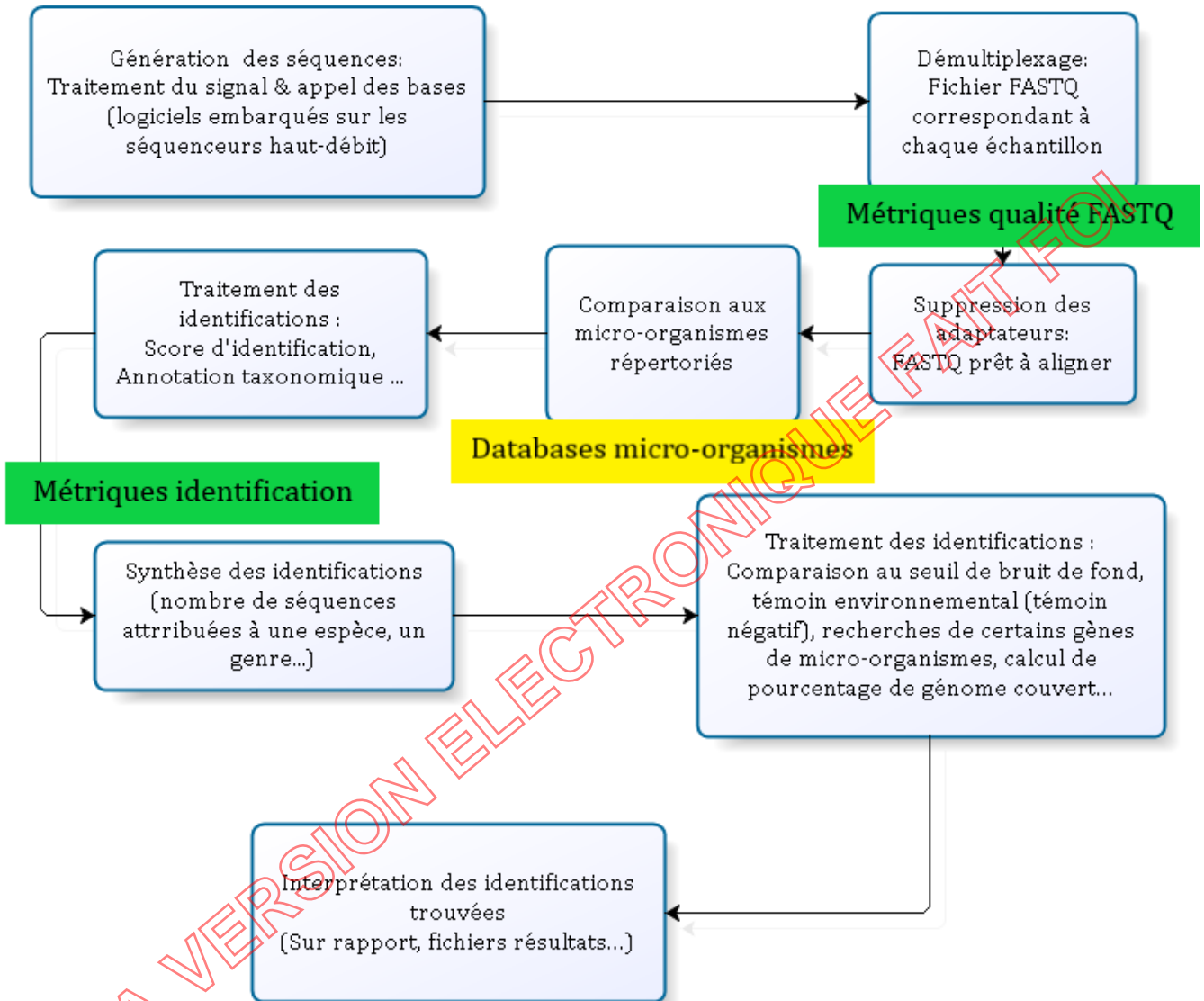
La qualification d'un pipeline doit permettre de s'assurer que l'ensemble des fonctionnalités souhaitées du pipeline sont opérationnelles, sans perte de donnée, dans l'environnement de production et à des niveaux de performances acceptables. La structure devrait définir le ou les pipelines, les bases de données associées ou tout autre processus pertinent entrant dans la phase analytique. Pour les bases de données, la structure devrait s'assurer de l'utilisation des bases en adéquation avec son champ d'expertise et du suivi des données.

La structure devrait **définir la stratégie la plus pertinente de qualification et l'argumenter** dans un **dossier de qualification de ses outils bioinformatiques NGS**.

LA VERSION ELECTRONIQUE FAIT FOI



**Logigramme n° 1 :** Exemple de pipeline de la génération des séquences jusqu'à l'annotation des variants. Ce schéma est un exemple à adapter à chaque domaine d'activité.



**Logigramme n° 2 :** Exemple de pipeline présentant les étapes de la génération des séquences jusqu'à l'identification de micro-organismes dans le cadre de la Métagénomique diagnostique. Ce schéma est un exemple à adapter à chaque domaine d'activité.





La qualification des outils bioinformatiques est entendue comme la vérification sur site (*i.e.* dans l'environnement réel de production de la structure) du bon fonctionnement (conformité aux spécifications attendues) de ces outils, d'une manière équivalente à la qualification d'un automate de séquençage.

Une fois les outils définis, la structure peut adapter la qualification séquentielle des composants de son pipeline. Cette qualification peut se baser sur des tests intermédiaires et spécifiques de qualité définis par la structure.

La structure doit qualifier initialement l'utilisation du pipeline bioinformatique avec un jeu de données qu'elle peut réutiliser à une fréquence définie pour s'assurer du maintien de la qualification de son pipeline au cours du temps (test de non-régression). Chaque structure peut adapter ses jeux de données de test pour la qualification du pipeline bioinformatique et ceux pour la vérification/validation de méthode de l'examen en fonction de son cahier des charges, de la finalité du test et de son organisation (exemple : détection de variants constitutionnels de petite taille, SNV ou indels inférieurs à un seuil, détermination d'une espèce en microbiologie...).

**► Les recommandations relatives au déploiement de la qualification d'un pipeline bioinformatique en fonction de l'organisation de la structure et au contenu du dossier de qualification sont mentionnées en Annexes 1, 2 et 3.**

Les critères utilisés dans la phase préliminaire à la qualification (cf Annexe 3) doivent être suivis dans la phase de qualification puis dans la phase de validation. Ils correspondent à la fois à des métriques intermédiaires et à des critères de performance terminaux.

Lors du changement de version du pipeline, la structure doit s'assurer de la nécessité de réaliser une nouvelle étape de qualification (voire l'étape préliminaire à la phase de qualification) sur la base d'une analyse d'impact documentée (cf Gestion de portée flexible, §5.5).

### 5.3.1.3 Equipements – Mode d'emploi

Dans les situations particulières de partage de matériels, la structure doit s'assurer de l'information, de la formation et de l'autorisation du personnel qui utilise l'équipement ainsi que de l'utilisation, la manipulation, le transport, la sécurité, le stockage et la maintenance de l'équipement afin d'éviter toute détérioration ou contamination. Dans ce cadre, la structure doit être particulièrement attentive au maintien de la confidentialité sur les données patients.

### 5.3.1.5 Maintenance et réparation du matériel

La structure doit définir l'architecture informatique comprenant le mode de fonctionnement des postes d'analyses, le mode de stockage, le mode de sauvegarde... (Exemples de réseaux de stockage: Storage Area Network (SAN), Network Attached Storage (NAS), Direct Attached Storage (DAS)).

A l'instar des équipements de laboratoire, la structure doit définir un programme pour s'assurer de l'entretien du matériel informatique.

Dans le cas de matériel partagé, la structure doit définir les responsabilités en matière de maintenance et d'entretien des équipements et matériels.

Après maintenance ou réparation, la structure doit s'assurer du fonctionnement conforme des équipements et matériels par une nouvelle qualification du pipeline et/ou du séquenceur puis de l'ensemble du processus analytique NGS.



## 5.4 Processus préanalytique

La structure doit veiller aux étapes de préparation de l'échantillon en fonction de la nature de l'échantillon primaire. Celle-ci ayant un impact sur la qualité et la quantité d'ADN et sur les performances de la méthode (sensibilité et spécificité), la structure doit définir les critères de qualité de l'échantillon pour son acceptation ou son refus selon les spécifications de méthodes utilisées ou kit utilisé.

Lorsque les prélèvements et les traitements de l'échantillon sont réalisés hors de la structure, celle-ci doit formaliser ses besoins en termes de matrice en fonction de l'examen demandé et en termes de qualité des échantillons reçus et les communiquer à ses clients. De plus, elle doit s'assurer de la qualité des échantillons reçus en fonction de chaque matrice par des moyens adaptés (exemple critère de sélection : conditions de décalcifications des échantillons pour le NGS, ...).

Dans le cas de prélèvements précieux, la structure peut déroger aux critères d'acceptation des échantillons préalablement définis sous réserve de sa traçabilité et d'une communication à ses clients.

Définition d'un prélèvement précieux : un échantillon qu'on ne peut pas renouveler en raison d'une difficulté technique, du risque du geste (invasif, dangereux pour le patient, ...), du contexte thérapeutique... La structure établit au regard de cette définition ses prélèvements précieux (liquide amniotique, trophoblaste, ...).

En génétique constitutionnelle, la structure doit veiller au respect de la réglementation en vigueur sur l'information préalable du patient et son consentement écrit.

**Pour les applications du sous-domaine de la microbiologie et plus particulièrement lorsqu'une méthode de séquençage non ciblée est appliquée sur un échantillon biologique d'origine humaine, une part des séquences provient des cellules du patient sur lequel a été prélevé l'échantillon biologique. La structure doit donc veiller au respect de la réglementation en vigueur sur l'information préalable du patient et son consentement écrit si le génome du patient est analysé.**

De plus, dans le cas de la microbiologie, la structure doit veiller au respect des règles de sécurité microbiologique. Dans certains cas, il est possible que les prélèvements précieux ou des souches soient rétrocédés d'une autre structure pour un complément d'analyse par NGS. La structure doit spécifier les conditions d'acceptabilité de ces échantillons et tracer l'origine de provenance.

### 5.4.2 Informations pour les patients et les utilisateurs

Pour les examens de génétique et le consentement des patients, la structure doit s'assurer de respecter la réglementation en vigueur.

### 5.4.7 Manipulation préanalytique, préparation et entreposage

La structure doit établir des dispositions pour la conservation des échantillons primaires en cas de prescriptions d'examens supplémentaires (exemple : ré-interprétation des données bioinformatiques pour le NGS, ...).



## 5.5 Processus analytique

Toutes les étapes qui composent la phase analytique de l'analyse NGS doivent être conformes aux exigences concernant le processus analytique et notamment la vérification/validation des procédures analytiques (§5.5.1.2 ; 5.5.1.3, 5.5.3).

La structure peut se référer aux recommandations récentes des sociétés savantes des différents domaines pour ce chapitre.

Sur la base d'une analyse des risques associée à chaque dossier de vérification/validation des méthodes, la structure doit établir les critères qualité à suivre pour s'assurer des performances de la méthode et les actions à mettre en place pour éliminer les risques identifiés à chacune des étapes de l'analyse à maîtriser (exemple pour l'analyse bioinformatique des étapes à maîtriser en Annexes 4 et 5).

Des échantillons de référence et des fichiers tests de référence (ou jeux de tests) doivent être utilisés pour la vérification/validation des méthodes avant mise en production pour s'assurer de la qualité des résultats en continu. La structure doit justifier de la représentativité des échantillons et fichiers de référence en fonction des cibles rendues. Les échantillons de référence doivent être documentés et représentatifs des échantillons patients. Les fichiers tests de référence doivent présenter des caractéristiques/spécifications compatibles avec la technologie et l'analyse à réaliser.

La structure doit s'assurer de la qualité et de la quantité d'ADN minimal pour la réalisation de l'analyse, ces deux paramètres étant cruciaux pour la vérification/validation des méthodes.

La structure doit déterminer les performances de la méthode adaptées à son dossier de validation de méthode (LOD, sensibilité, spécificité, fidélité intermédiaire, justesse, ..), les métriques qualité et les techniques adaptées (taux de couverture ou homogénéité, profondeur de lecteur ou valeur minimale à définir (ces valeurs varient selon les méthodes utilisées, le nombre d'échantillons, etc.)) pour démontrer que ces performances sont atteintes.

Pour chaque examen, la structure doit définir les cibles rendues au titre du diagnostic, faisant l'objet d'un compte-rendu et établir des critères de performances à atteindre (Exemple : 95% des exons de l'exome couvert à 20X ou bien bactéries, virus et champignons couverts sur 1% de leur génome au minimum par exemple pour la métagénomique).

Pour l'analyse bioinformatique, la **qualification** du pipeline bioinformatique peut être distinguée de la **validation de méthode qui est directement sous la responsabilité du directeur de la structure**. La validation de méthode doit intégrer l'ensemble des sous-processus de la phase analytique d'un examen. L'étape bioinformatique peut soit constituer un sous-processus dans le dossier de validation/vérification de méthode soit être intégrée dans le processus complet. L'étape de vérification ou validation de méthode ne peut être effectuée que lorsque le pipeline est qualifié.

**Dans le cas d'un recours à un service support**, la prestation bioinformatique doit être intégrée au dossier de validation/vérification de méthode en tant que sous-processus, mais la structure doit solliciter auprès du service support les éléments de qualification initiale qu'il jugera pertinents pour valider l'adéquation à son cahier des charges (informations sur les étapes de développement bioinformatique, par exemple, les preuves des fonctionnalités et de non-régression, la liste des modifications (sous forme de Release Notes,... ), gestion des données selon le RGPD et la réglementation en vigueur).

Conformément aux paragraphes 5.5.1.2 et 5.5.1.3, la supervision et l'approbation finale de la qualification des outils bioinformatiques doit relever d'un personnel possédant les compétences requises.



Toute modification d'une étape de l'analyse, notamment du pipeline bioinformatique ou des changements de versions des bases de données ou du génome de référence doit faire l'objet d'une nouvelle qualification avec des jeux de données-tests connus, des tests de non-régression associés dictés par une analyse de risques.

La structure devrait définir des conditions sur la base de critères pour identifier dans quels cas le résultat est à confirmer, par exemple, par une autre technique ou tester sur un deuxième prélèvement.

#### Cas particulier des tests CE-IVD ou analyse fournisseur clef en main (portée A):

La structure doit établir une vérification de méthode en conditions de production pour vérifier que les spécifications du fournisseur sont atteintes.

Pour la recherche non ciblée, et de la même manière que pour la recherche ciblée, la structure doit définir et vérifier/valider les performances de la méthode initialement et en continu en justifiant de la représentativité des échantillons et fichiers de référence utilisés en fonction des résultats rendus au titre du diagnostic.

Pour les applications du sous-domaine de la microbiologie et plus particulièrement lorsqu'une méthode de séquençage non ciblée est utilisée sur un échantillon biologique d'origine humaine, la structure devrait :

- documenter les proportions de séquences alignées à des génomes de référence d'origine microbiologique ou humaine
- définir les seuils minimaux de séquences d'origine microbiologique et humaine permettant l'identification et la quantification d'acides nucléiques d'origine microbiologique
- documenter les interférences liées à la présence d'ADN humain dans l'échantillon, la présence de séquences de microorganismes contaminants par exemple dans les réactifs utilisés, les détections croisées de microorganismes liées à la présence de microorganismes d'intérêt ou provenant de la flore commensale de l'échantillon.

Il est précisé que l'expression de la portée de la structure est mentionnée dans la liste détaillée des examens couverts par l'accréditation. La structure doit identifier l'examen ou le diagnostic précis réalisé en précisant le panel, les cibles, et/ou la pathologie. La structure indique la technologie utilisée « NGS » dans la colonne « principe de la méthode » ainsi que le nom de l'automate et la référence de l'automate ou des automates qui portent l'analyse.

## 5.6 Garantie de qualité des résultats

La structure doit établir une stratégie de passage des contrôles de qualité interne et externe permettant d'assurer le suivi et le maintien des performances de la méthode en considérant l'ensemble du processus analytique.

La stratégie des CIQ doit être déterminée par la structure avec la définition de métriques suivies pour un run et/ou l'utilisation d'un CIQ fournisseur.

Dans le cas des analyses en génétique somatique, l'utilisation d'un CIQ représentatif des fréquences alléliques adapté aux limites de détection fixées est recommandée.

Dans le cadre de la microbiologie, le(s) CIQ doit/doivent être choisi(s) pour être le plus en adéquation possible avec les pathogènes recherchés. Par exemple, le CIQ peut comporter un mélange de bactéries, virus, champignons prédéterminés pour la Métagénomique pan-pathogène.



Les EEQ peuvent être réalisées par méthode comprenant comme critères la même technique d'enrichissement, et de séquençage. Les comparaisons interlaboratoires (CIL) sont un moyen pour s'assurer de la qualité des résultats. D'autres modalités alternatives de contrôles de qualité peuvent être mises en place en l'absence de CIL, notamment avec la réalisation d'analyses entre plusieurs structures sur un même matériau, de préférence de source externe (échange de matériaux/spécimens/échantillons biologiques). Afin d'assurer la confidentialité des résultats de génétique, et dans le cadre de l'utilisation d'échantillons biologiques, la structure doit s'assurer de respecter la réglementation en vigueur en termes de consentement des patients et de respect de la confidentialité.

La structure peut se référer aux recommandations des sociétés savantes du domaine pour ce chapitre (ESHG pour le NGS (CQ, PhiX, utilisation de patients traceurs à chaque test par exemple...) et les recommandations décrites pour la génétique constitutionnelle, ...). La structure doit s'assurer que les échantillons utilisés pour les contrôles qualité permettent de s'assurer de la conformité sur l'ensemble des processus pré, per et post-analytiques. Dans le cas où la structure a mis en place plusieurs pipelines bioinformatiques en interne, elle doit vérifier la comparabilité des résultats (cf § 5.6.4).

Dans le cadre de la microbiologie, en absence d'EEQ disponible pour le NGS, des EEQ visant à fournir les mêmes résultats que la technique NGS peuvent être utilisés.

- Par exemple, les EEQ « panels syndromiques » peuvent être utilisés pour la métagénomique, les EEQ de résistances prévus pour le Sanger pour la détection de variants ou des EEQ visant à comparer des souches par exemple lorsqu'ils existent.

Dans le cas de la recherche non ciblée, la structure veille à utiliser des EEQ les plus diversifiés possibles pour démontrer la largeur de détection de l'analyse.

- Par exemple en recherchant des pathogènes aussi diversifiés que possibles dans les approches de métagénomique.

## 5.7 Processus post-analytique

Les étapes de qualification clinique des variants et interprétation contextuelle doivent être strictement réalisées au sein de la structure, par le biologiste (modalités pour la signification des résultats connus et non connus et leur interprétation).

Pour l'étape d'interprétation et celle du compte-rendu, la structure doit connaître les paramètres sélectionnés aux étapes de traitement à maîtriser, aux logiciels d'alignement, d'appel de variant, d'annotation de variant, etc., ainsi qu'aux versions utilisées.

La structure doit garantir dans ses dispositions que les biologistes ont des pratiques harmonisées en matière de validation / interprétation des résultats de génétique. Elle peut choisir par exemple d'organiser des réunions d'harmonisation entre les différents biologistes à partir de dossiers réels de la structure, de mettre à disposition des données scientifiques (bases de données), de contacter des experts cliniques ou biologiques à consulter en cas de besoin (exemple: dans le cadre de Réunion de Concertation Pluridisciplinaire, ...). Néanmoins, la compétence pour valider certains comptes-rendus peut être réservée à certains personnels autorisés (exemple : agrément génétique constitutionnelle...) (cf. §5.1.2 du présent guide).



Pour la microbiologie, les seuils de positivité doivent être clairement établis par la structure pour le rendu positif ou négatif. Par exemple, dans le cas de recherche de mutations de résistance virale, les seuils d'interprétation (% de la quasi espèce) à partir desquels le biologiste rend le résultat doivent être clairement explicités. Cela peut être aussi une quantité relative de séquences reliées à un pathogène donné dans le cas de la métagénomique ou bien un seuil habituellement observé de différences entre deux souches dans le cadre d'études épidémiques.

Dans le cas de l'identification de pathogènes non décrits dans la pathologie identifiée, la structure doit prévoir des dispositions pour le rendu des résultats (cf 4.7).

En génétique, la structure doit décrire des dispositions et prévoir des conduites à tenir pour le rendu des résultats en cas de découverte de données incidentes et/ou secondaires (consentement des patients), de données sans critères d'interprétation validés, de résultats de variants de signification inconnue en accord avec la réglementation en vigueur (échanges avec les pairs, RCP, collection de données familiales, mises en place de tests fonctionnels).

## 5.8 Compte-rendu des résultats

En sus des exigences normatives, les recommandations des sociétés savantes précisent les informations à communiquer aux utilisateurs (prescripteurs, biologistes en cas de sous-traitance,) dans les comptes-rendus, par exemple :

- les résultats exprimés qualitativement, LOD (limite de détection) exprimée dans la méthode
- sur le compte-rendu, la fréquence allélique minimale, le cut-off d'interprétation des quasi-espèces, les règles décisionnelles utilisées (à définir par la structure),
- la version du pipeline NGS (ou des logiciels utilisés : pour l'alignement, l'appel de variants, etc.), la base de données, l'interface d'interprétation le cas échéant,
- Dans le cadre de la génétique, la nomenclature des variants doit suivre les standards internationaux. Dans le cadre de la microbiologie, et pour ce qui concerne les analyses de mutations, la structure doit s'assurer de rendre les résultats selon la nomenclature habituellement utilisée dans la littérature lorsqu'aucune règle internationale n'a été définie,
- les métriques qualité exploitables en prestation de conseil et en interprétation doivent être adaptées au type d'examen et au contexte clinique,
- les changements significatifs (i.e. susceptible d'impacter les résultats) dans le pipeline et/ou l'interface d'interprétation. Ces modifications doivent être enregistrées dans le système documentaire de manière à pouvoir être exploitées, entre autres choses, pour la prestation de conseil et l'interprétation, y compris de manière rétrospective.

Les exigences de la structure relatives aux logiciels ou aux services doivent respecter, *a minima*, celles du processus post-analytique.

En cas d'interprétation automatisée, la structure doit s'assurer des seuils minimaux fixés pour le rendu des résultats et de la compétence de son personnel en termes de règles de décision et d'interprétation des résultats.



## 5.10 Gestion des informations de la structure

Une analyse des risques liés à la transmission des fichiers NGS, à leur traitement et à leur conservation/stockage devrait être conduite au préalable de la mise en production du sous-processus bioinformatique. Les moyens de maîtrise correspondants devraient être établis au travers de dispositions ou d'actions spécifiques.

La structure peut formaliser des dispositions pour s'assurer de la maîtrise des risques potentiels suivants:

- rupture de confidentialité des données (ex : défaut de pseudonymisation, droits d'accès mal définis, insuffisamment limités, protocoles ou modalités de transfert mal sécurisés, etc).
- perte des données (ex : défaut de récupération des fichiers en cas de coupure du réseau, défaut de récupération des fichiers et ou du code source d'un outil bioinformatique en cas d'incident sur des serveurs, etc.)
- altération des résultats et d'erreurs d'interprétation: les causes potentielles sont nombreuses, s'agissant aussi bien des pipelines NGS, que des bases internes de variants et des interfaces d'interprétation. La structure devrait instruire ces analyses de risques en confrontant son expérience et ses besoins aux recommandations des sociétés savantes.
- stockage et archivage des données.  
La structure doit s'assurer de la conformité des services proposés avec la législation française en vigueur, notamment le RGPD et les recommandations de la CNIL.

LA VERSION ELECTRONIQUE EST PROHIBÉE



## 9 MODALITES D'EVALUATION DES LABORATOIRES DE BIOLOGIE MEDICALE ET STRUCTURES D'ACP POUR LES ANALYSES

Ce tableau présente les modalités d'instruction, d'expertise et d'évaluation en cas de demande d'extension pour le séquençage à haut-débit dans les cas suivants :

- N°1 : Réalisation de toute l'analyse par la structure en interne
- N°2 : Réalisation de toute l'analyse selon une méthode fournisseur
- N°3 : La structure a fait appel à un service support bioinformatique qui appartient à la même entité juridique pour la mise en place de l'analyse bioinformatique du NGS
- N°4 : L'analyse NGS est sous-traitée

Cas	Instructions du dossier	Evaluation sur site
N°1 En interne		DVM Résultats EEQ Cahier des charges Procédure validation et interprétation des résultats  <b>Réalisée par un ET NGS</b>
N°2 Analyse Fournisseur	Description de l'organisation Dossier de qualification du pipeline (1) Procédure en cas de modification du pipeline  <b>Expertise documentaire du dossier de qualification du pipeline Réalisée par un expert bioinformaticien</b>	DVM Résultats EEQ Cahier des charges Procédure validation et interprétation des résultats Accords avec le fournisseur  <b>Réalisée par un ET NGS</b>
N°3 Analyse bioinformatique mise en place par un service support		DVM Résultats EEQ Cahier des charges Procédure validation et interprétation des résultats Accords avec le service support Evaluation dédiée du service support  <b>Réalisée par un ET NGS</b>
N°4 Sous-traitance	Pas d'expertise	Contrat avec le sous-traitant Processus pré-post analytique  <b>Réalisée par un ET GENETIQUE</b>

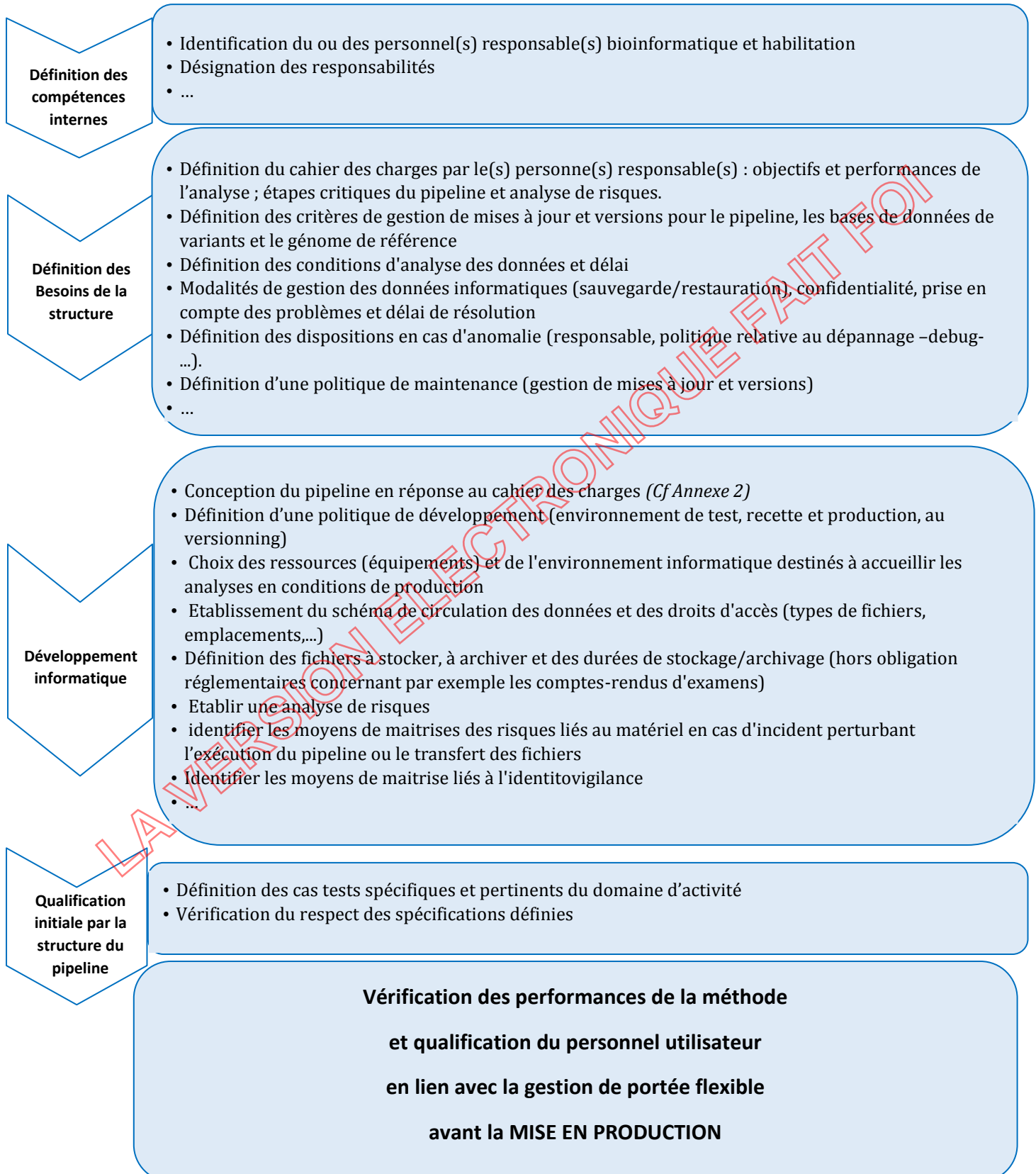
(1) Les documents relatifs à la qualification du pipeline doivent être expertisés en amont de l'évaluation notamment les dispositions de la structure et les enregistrements tels que le cahier des charges défini par la structure, la traçabilité des documents adaptée à toute évolution du pipeline, ou les enregistrements relatifs à la conception et/ou la mise en place du pipeline si faite en interne.





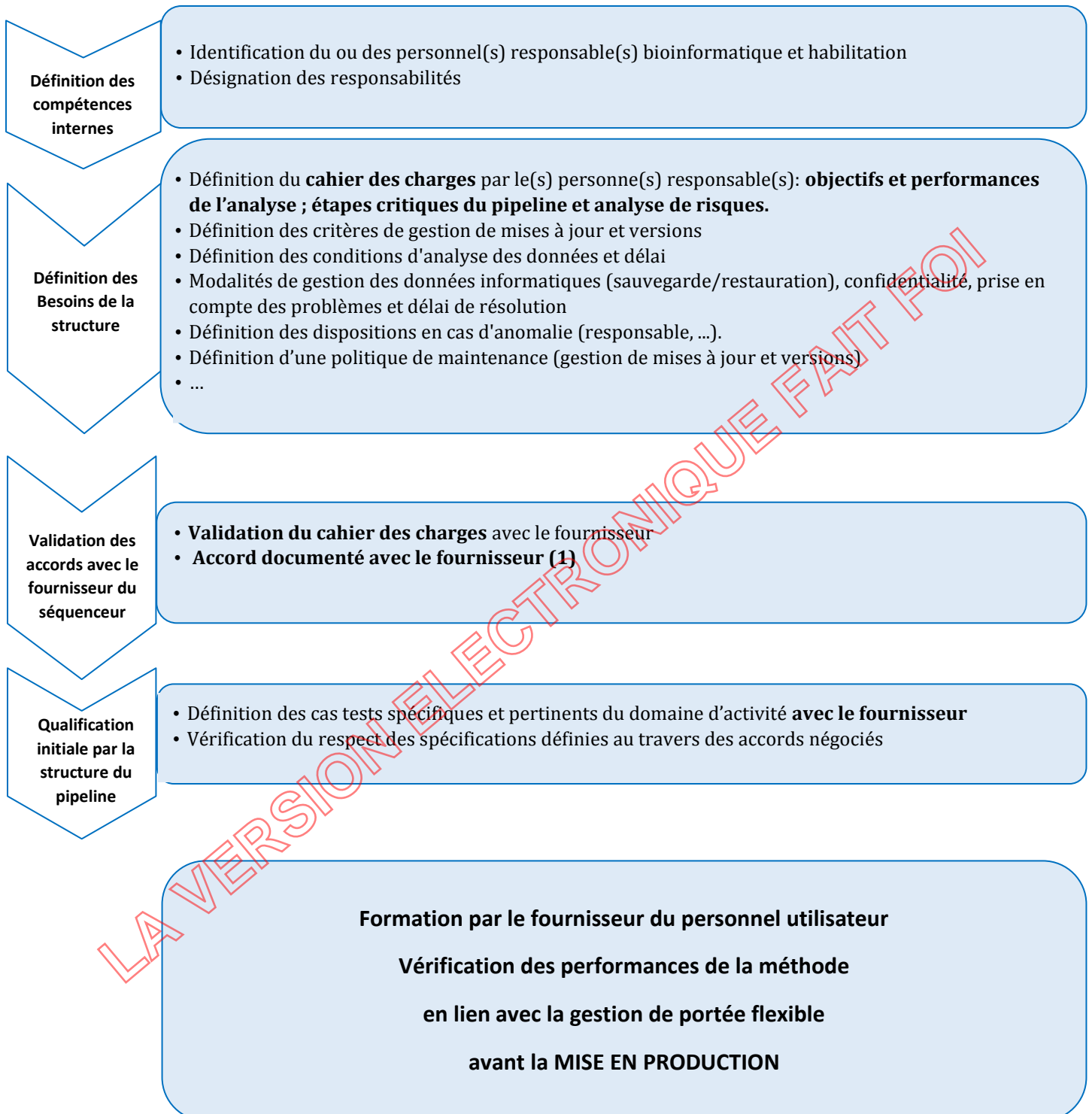
## Annexe 1 : Qualification du pipeline bioinformatique au travers de la maîtrise des risques

### Cas n°1 : La structure réalise toute l'analyse en interne



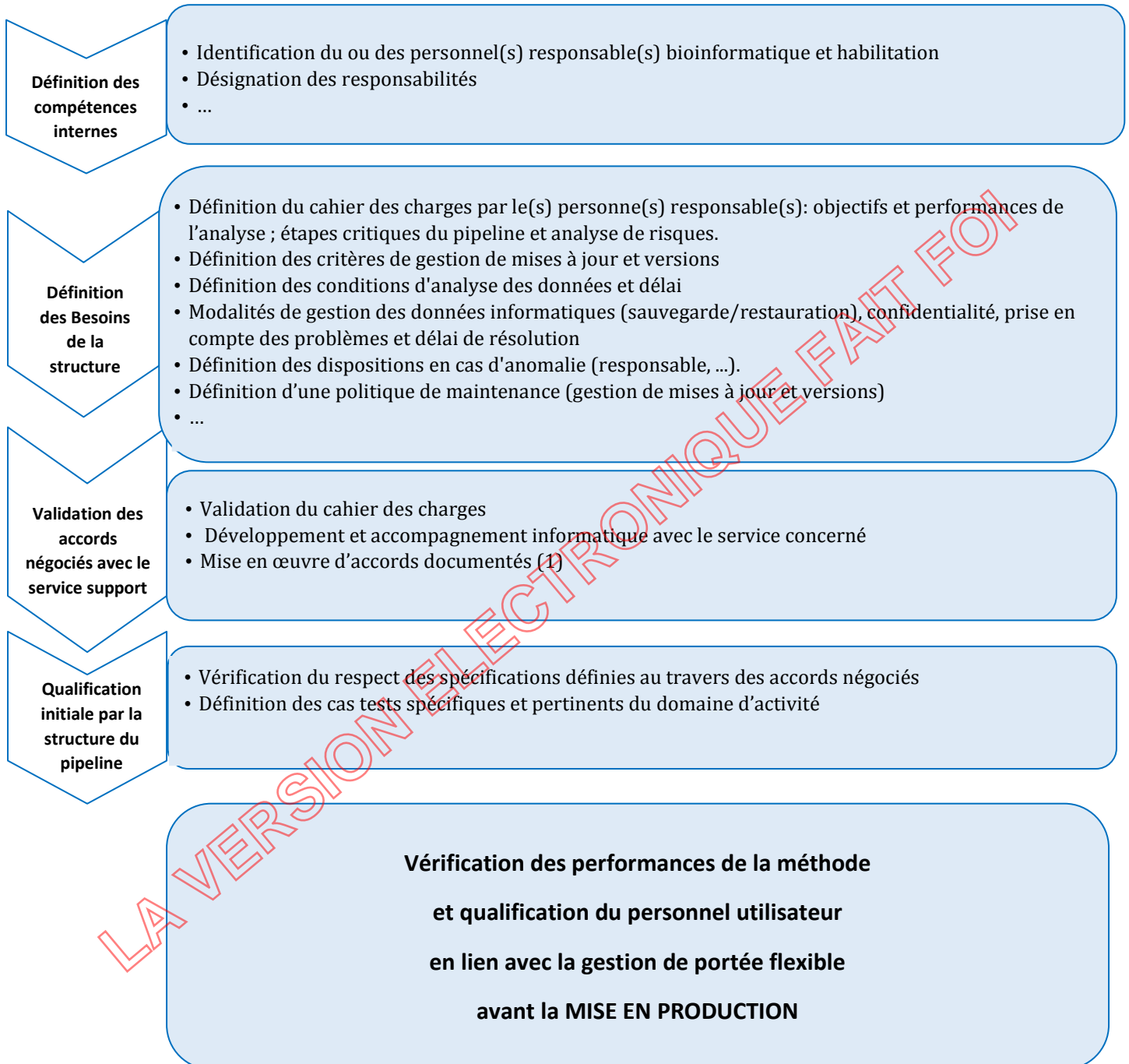


## Cas n°2 : Analyse fournisseur « Clef en main »





### Cas n°3 : La structure fait appel à un service support informatique appartenant à la même entité juridique



(1) La nature des accords établis (contrats, procédures, échanges formels, ...) sera adaptée par la structure.



## Annexe 2 : DESIGN, DEVELOPPEMENT ET MAITRISE DU PIPELINE

En lien avec le logigramme du cas N°1, pour le design et la conception du pipeline informatique

### Design, développement et maîtrise du pipeline

- **Mise en place d'un cahier des charge avec la description des besoins de la structure pour le pipeline à qualifier :** type de plateforme de séquençage (méthode de préparation de librairies et de séquençage), type d'échantillons primaires, type d'examen (constitutionnel, somatique, microbiologie, HLA), types de résultats à obtenir (SNV, indels  $\leq$  taille définie, CNV, SV, variants complexes, microorganismes), implication dans les délais de rendu des résultats
- **Conception du pipeline en réponse au cahier des charges:** choix des outils composant le pipeline, choix des paramétrages... La justification du choix des outils (logiciels, de normalisation des données, d'appel de variant, d'annotation...) peut être réalisée sur la base d'une étude de la littérature, d'une étude pilote de comparaison au sein de la structure ou mutualisée au sein de réseaux experts ou bien encore à partir de recommandations professionnelles.
- **Choix des ressources (équipements) et de l'environnement informatique** destinés à accueillir les analyses en conditions de production
- **Etablissement du schéma de circulation des données et des droits d'accès** (types de fichiers, emplacements...)
- **Définition des fichiers à stocker, archiver et des durées de stockage/archivage** (hors obligations réglementaires concernant par exemple les comptes-rendus d'examen)
- **Analyse de risque:**
  - Selon la méthodologie retenue par la structure (5M ou autre), directement dans le dossier ou en document lié.
  - Que le pipeline soit connecté ou non à des outils tels que des bases de données internes/interface d'interprétation par exemple, l'analyse peut être conduite de manière sectorisée, à titre d'exemple : *analyse primaire* des bcl au fastq ; *analyse secondaire* des fastq aux vcf ; *analyse tertiaire* des vcf jusqu'au rapport émis via une interface d'interprétation.
- Les moyens de maîtrises des risques liés au matériel doivent notamment prévoir des procédures dégradées en cas d'incident perturbant l'exécution du pipeline ou le transfert des fichiers ainsi que des procédures de récupération/sauvegarde des données en cas de perte. (Norme NF EN ISO 15189, § 5.10.3 ; Roy et al, 2018)
- **Les moyens de maîtrise des risques liés à l'identitovigilance :**
  - Il est recommandé de préserver la traçabilité d'un échantillon à chaque étape de l'analyse bioinformatique,
  - La structure doit s'assurer de la sécurisation du protocole de transfert des fichiers entre deux sites distants. Les droits d'accès doivent être maîtrisés,
  - En règle générale, la conformité réglementaire des moyens de maîtrise de ces risques doit être supervisée par le correspondant CNIL de la structure, de l'institution ou bien par la CNIL.



## Annexe 3 : Recommandations relatives au contenu du dossier de qualification

La qualification du pipeline bioinformatique peut comprendre une étape préliminaire d'ajustement suite au développement du pipeline.

**Etape 1 : La phase préliminaire à la qualification correspond au temps nécessaire pour la mise en œuvre et la maîtrise du pipeline en dehors de l'environnement de production.**

**La mise en œuvre est évaluée dans le cadre d'une étude pilote :**

- Jeu de données test : données issues de matériel biologique et préalablement caractérisées par une autre méthode, données *in silico* vérifiées. Dans le premier cas, et pour la recherche de variants, ceux qui sont caractérisés doivent être en adéquation avec la finalité du pipeline (type de variant, d'analyse : constitutionnelle/somatique/microbiologie). Il peut s'agir de données internes à la structure ou externes (fichiers de CIL, EEQ, issus de publication, références de type NA12478...). Dans le cas de données externes, il n'est pas nécessairement obligatoire, au stade de l'étude pilote, que les variants sur lesquels portent l'évaluation des performances soient inclus dans des gènes correspondant aux examens envisagés par la structure.
- Evaluation des performances du pipeline : elle peut être effectuée sur la base de métriques intermédiaires et terminales superposables à celles utilisées en validation, et le délai de mise à disposition du résultat. Ces métriques servent autant à la qualification du pipeline qu'à la vérification des performances du séquenceur ; par exemple : la profondeur, le score d'alignement des lectures, ...
- En cas d'identification d'erreurs ou de points critiques non anticipés, la structure doit mettre en place une organisation de suivi des incidents (cf. §4.9).

**La structure devrait adapter la phase préliminaire en fonction des résultats obtenus, et, en cas de modifications, s'assurer de l'enregistrement des versions.**

**A l'issue de cette phase, les configurations retenues (versions de logiciel, paramètres...) devraient être verrouillées** en vue d'entreprendre la qualification et la vérification/validation de méthode.

### **Etape 2 : la qualification**

**La phase de qualification correspond au déploiement dans l'environnement de production.**

**Elle doit s'effectuer obligatoirement :**

- dans l'environnement informatique de production de la structure
- avec la configuration du pipeline telle qu'arrêtée au terme de la phase d'optimisation
- au moyen d'un jeu de données de référence adapté à l'utilisation diagnostique du test. Des recommandations sont disponibles concernant l'utilisation de données *in silico* et l'inclusion de variants représentatifs dans le jeu de référence, ainsi que la



revalidation en cas de modification du pipeline (cf. références : recommandations 14, 15 et 17, Roy *et al*, 2018). Les modalités de calcul du nombre d'échantillons et de variants à inclure dans le jeu sont explicitées dans la référence suivante : Jennings et al, 2017 (PMID : 28341590).

Les critères utilisés dans la phase préliminaire à la phase de qualification doivent être suivis dans la phase de qualification puis dans la phase de validation. Ils doivent correspondre à la fois à des métriques intermédiaires et des critères de performance terminaux.

LA VERSION ELECTRONIQUE FAIT FOI



## Annexe 4 : Exemple d'étapes de l'analyse bioinformatique à maîtriser pour la détection de variants

Le pipeline bioinformatique utilisé dans l'analyse NGS est composé de différentes étapes, pré-processing et variant calling, qui présentent des phases à contrôler.

A titre d'exemple, sont présentées les étapes suivantes. Chaque structure devrait adapter si besoin son analyse.

### 1- L'étape du pré-processing :

Cette étape est composée des éléments suivants : la réception des données, le QC initial, la phase d'alignement, la normalisation des alignements et l'obtention de données alignées.

#### ➤ Pour la réception des données :

Il s'agit, à ce niveau, de vérifier la qualité des données initiales.

##### Points à contrôler :

- La qualité des bases : cette étape est importante pour l'étape du variant calling. La technologie n'est pas parfaite, des erreurs de séquençage peuvent apparaître (<0.1% d'erreur chez Illumina), qui sont différentes des variations réelles par rapport à la séquence de référence
- La longueur des séquences
- Le contenu en GC
- La quantité de lecture.

#### ➤ Pour la phase d'alignement :

##### Points à contrôler :

- Choix du génome de référence : étape cruciale ; quelques subtilités sont à vérifier pour ce choix au niveau de la nomenclature des chromosomes (exemples : Chr1/c1/1 ou ChrM/ChrMT/M/MT). De plus, il est important, dans le choix du génome, de porter attention aux séquences alternatives haplotypaires ou inconnues qui induiraient des alignements non spécifiques et la non détection potentielle de variations dans la région concernée (exemple de SYNGAP1 contenu dans une région haplotypaire\* alternative de hg19).
- Les paramètres de l'alignement.

#### ➤ Pour la phase de normalisation des alignements.

##### Les points à contrôler :

- Le marquage des duplicats PCR en fonction de la technologie utilisée (un duplicat de PCR correspond aux exactes mêmes coordonnées pour les 2 reads de la paire. Il est peu probable que 2 molécules soient fragmentées de la même manière, sur le même point de cassure pour produire 2 lectures identiques. Ceci est en fait dû à un artefact de la PCR qu'il faut identifier car si ce duplicat est pris en compte cela aurait un impact sur la détection de la mutation.)
- La recalibration des scores de qualité des bases : la qualité des bases dépend du contexte.
- Le réalignement autour des sites d'insertions et de délétions : les méthodes d'alignement peuvent être imparfaites, ils sont à vérifier sur ces sites.
- La profondeur et la couverture :



- Définition de la profondeur : nombre de lectures qui couvre 1 position (exprimée en X)
- Définition de la couverture : nombre de bases séquencées par au moins une lecture (exprimée en %)

On peut aussi exprimer un taux de couverture à une profondeur déterminée. Lorsque les régions d'intérêts ne sont pas couvertes par au moins une lecture, on ne peut détecter de mutations. Il est recommandé d'indiquer la limite de l'analyse dans le compte-rendu.

## 2- L'étape du Variant Calling

Une fois les données alignées, l'étape de recherche de Variant Calling débute. Cette étape comprend différentes phases : la recherche de variations, l'annotation associée aux nomenclatures et aux bases de données, les filtres.

### ➤ Pour la phase de recherche de variations

Il s'agit, avant tout, d'adapter les outils utilisés car en fonction des cas, ils ne peuvent pas tous toujours détecter les mutations (exemple : mutation mosaïque à 7%, indétectable via GATK unified Genotyper).

#### Points à contrôler :

- Biais de batch
- Méthode de VC adaptée (somatique, germline, ...), spécificité, sensibilité.

### ➤ Pour la phase d'annotation

#### Points à contrôler :

- Nomenclature : en fonction de la méthode utilisée, il est possible de détecter des types de mutations différentes. Il est donc difficile dans certains cas de détecter des éléments complexes, cela dépend des pathologies (exemple Nomenclature HGVS : rapporter les mutations en 3', difficulté de détection « propre » des insertions/délétions).
- Bases d'annotations : la structure justifie son choix des bases d'annotation utilisées.

### ➤ Pour la phase de filtration

#### Points à contrôler :

- Les seuils utilisés pour les filtres
- L'adaptation des échantillons contrôle
- La gestion de la découverte secondaire

Il s'agit ici de porter attention à la sélection des bases et de leurs filtres. Lorsque des filtres sont utilisés pour les variants, ils devraient être adaptés à la pathologie utilisée.

Exemple du facteur V de Leiden (Thrombophilie) : la fréquence dans GnomAD est de 98.05% mais la base de référence n'est pas la bonne, rétablissement de la bonne base dans GRCH38 avec une fréquence de variation de 1.95%. Pour le facteur V, le filtre à 1% ne permet pas de retenir la variation.





## Annexe 5 : Exemple d'étapes de l'analyse bioinformatique à maîtriser pour la détection de microorganismes

### 1- L'étape du pré-processing :

Cette étape est composée des éléments suivants :

- l'analyse de la qualité des données brutes et démultiplexées,
- le QC initial,
- la phase d'alignement,
- la normalisation des alignements et l'obtention de données alignées.

➤ Pour l'analyse de la qualité des données brutes :

Il s'agit, à ce niveau, de vérifier la qualité des données initiales.

Points à contrôler :

- La qualité globale du run, la qualité de chaque base n'est pas impactante pour une identification de séquence dans sa globalité, mais la qualité du run devrait être suivie pour vérifier que les analyses se font dans les règles définies par le fabricant et /ou les utilisateurs.
- La densité des amplifications clonales (ou clustering) peut servir également d'indicateur pour le suivi

➤ Pour l'analyse de la qualité des données démultiplexées :

- Le nombre de séquences obtenu pour chaque patient conditionne la sensibilité de l'analyse, elle devrait être rigoureusement suivie et les seuils d'acceptabilité précisés par la structure.

### 2- L'étape d'identification

Une fois les données alignées, l'étape de recherche d'appel des variants débute. Cette étape comprend l'annotation associée aux nomenclatures et aux bases de données, les filtres.

➤ **Pour la phase d'identification :**

Points à contrôler :

- Choix de la base de données : étape cruciale qui permet de déterminer les limites de diversité détectable dans l'analyse.
- Les paramètres d'identification lorsque cela est possible (e-value, % d'identité...)

➤ **Pour la phase de filtration**

Points à contrôler :

- Le contrôle environnemental : Il s'agit d'un témoin qui suit le process expérimental et analytique de manière identique à un échantillon mais qui ne contient pas d'échantillon. Ce témoin est destiné à capter toutes les contaminations environnementales (eau, air, kits...) qui seraient susceptibles de rendre l'échantillon faussement positif. Ce témoin peut servir pour élaborer un seuil de positivité technique et doit être utilisé régulièrement.
- Le contrôle positif : Dans la mesure où il est possible dans un run que l'ensemble des échantillons soient négatifs, un témoin positif couvrant a minima les grands groupes de pathogènes (bactéries, champignons, virus) doit être utilisé régulièrement pour



vérifier la capacité de la technique à détecter les micro-organismes. Cela permet d'assurer les résultats négatifs.

### 3- L'étape d'interprétation

Points à contrôler :

- Si une interprétation automatisée est prévue, les seuils minimaux de nombre de séquence, quantité relative... devraient être fixés par la structure pour le rendu des résultats.

LA VERSION ELECTRONIQUE FAIT FOI