



SeqOne platform

Instructions d'utilisation

Valable à partir de SeqOne platform v2.1

1. Introduction	7
2. Contact fabricant et mandataire suisse	8
3. Réactovigilance	8
4. Utilisation prévue de la plateforme	8
4.1. Utilisation prévue	8
4.2. Indication d'utilisation	9
4.3. Description de la plateforme	9
4.4. Population de patients	10
4.5. Utilisateurs visés	10
4.6. Caractéristiques	10
4.6.1. Performances analytiques	12
4.6.1.1. Modules bioinformatiques	13
4.6.1.2. Modules d'annotation	22
4.6.2. Performances cliniques	23
5. Précautions	24
5.1. Utilisation conforme	24
5.2. Avertissements	24
5.3. Limites	26
5.4. Dysfonctionnements	26
6. Prérequis	27
6.1. Matériel	27
6.2. Réseau	27
7. Authentification	28
7.1. Connexion	28
7.2. Déconnexion	29
7.3. Page d'accueil	29
7.4. Gestion des comptes utilisateurs	30
7.4.1. Gestion de "My Account"	31
7.4.2. Gestion des paramètres utilisateur - définitions	31
Entity	31
Users	31
Workset	31
Manifest	32
Panels - Transcript	32
Panels - Genes	32
Marker	32
Protocols	32
Reporting - Global settings	33
Reporting - Template list	33
Tags	33

API Keys	33
7.4.3. Gestion administrative - présentation de l'interface	34
7.4.4. Basculer entre les modes Main et Preview	37
8. Description des étapes opérationnelles	38
9. Création du projet	39
9.1. Créer un projet	39
9.2. Gestion de projets	40
9.3. Suppression d'un projet	40
9.4. Données ADN (DNA Sequencing, ctDNA Sequencing)	40
9.5. Données sur l'ARN (RNA Sequencing)	42
9.6. Paramètres du projet	43
9.6.1. Configurer les métriques QC par défaut du projet	43
9.6.2. Définir le profil de filtre par défaut du projet	43
10. Chargement d'échantillons	44
10.1. Importer des échantillons	44
Depuis un ordinateur	44
Via le SDS (SeqOne Data Sync)	44
Statut de l'importation des échantillons	44
Capacité d'importation	45
10.2. Format des fichiers d'entrée	45
10.2.1. Fichiers d'analyse primaire	45
10.2.2. Fichiers d'analyse secondaire	46
10.3. Métadonnées échantillons	48
10.4. Modifier un échantillon	49
10.5. Supprimer un échantillon	49
11. Analyse	50
11.1. Lancer une nouvelle analyse	50
11.1.1. Choix du workset d'analyse	51
Analyses disponibles à partir de fichiers FASTQ	51
GermlineVar	52
GermVar	52
GermlineVarLR	52
GermlineFamily	53
GermVar Family	53
CNVCapture	54
CNVExome	54
SomaVar	55
SomaCNVCapture	55
SomaRNA	56
SomaMSI	56

SomaHRD	57
SomaVar LF	57
SomaCGP	58
SomaHemato	58
SomaLBx	59
SomaMethyl	59
Analyses disponibles à partir de fichiers d'analyse secondaire	60
GermVar Tertiary	60
GermVar Family Tertiary	60
SomaVar Tertiary	61
11.1.2. Choix du mode de traitement des UMI	62
Présentation des modes de traitement des UMI	62
Processus de traitement des UMI	63
Compatibilité avec les worksets d'analyse	63
11.2. Statut de l'analyse	64
11.3. Qualité de l'échantillon	64
12. Contrôle de la qualité des données	66
12.1. Mesures et rapports de contrôle de qualité	66
12.1.1. Mesures QC et contrôles QC	66
12.1.2. Rapport QC	66
12.1.3. Rapport 'ID-Check'	70
12.1.4. Rapport technique	71
Score UPD	72
Graphique de fréquence de l'allèle B (BAF) à partir du rapport UPD	73
13. Interprétation des résultats de l'analyse	74
13.1. Résultats de l'analyse	74
13.1.1. Gestion du workflow de l'analyse et de l'interprétation	74
13.1.1.1. Attribution des analyses pour l'interprétation	74
13.1.1.2. Verrouillage d'analyse	75
13.1.1.3. Panneau latéral pour la gestion des analyses	76
Vue modale Workflow	77
Vue modale Patient	78
Vue modale HPO	79
Vue modale Panel	81
13.1.1.4. Panneau latéral pour la gestion des évaluations	82
13.1.2. Aperçu des résultats d'analyses dans l'onglet Dashboard	83
13.1.3. Couverture détaillée par exon	84
Méthodologie	84
Visualisation et interprétation	85
13.1.4. Petits variants	88

Méthodologie	88
Visualisation et interprétation	92
Trier les variants	105
Configuration des colonnes du tableau des variants	107
Editer les métadonnées échantillons depuis le volet d'analyse	108
Filtrer des variants	110
Filtres Genes & Variants	110
Section Samples & Quality	110
Section Frequencies	112
Section Genetics	113
Section Databases	115
Section Transmission & Phenotypes	116
Section Classification	118
Section Scores	119
Sélectionner une indication	121
Utilisation d'un profil de filtre prédéfini	122
Modification et enregistrement des profils de filtre	122
Profils de filtres favoris	122
Combinaison de profils de filtres	123
Profils de filtres SeqOne	123
Fonctionnalité "Mark as viewed"	124
Sélectionner des variants	124
Évaluation des variants	125
Page du variant	128
Générer un rapport clinique	137
13.1.5. CNV	138
13.1.5.1. CNV dans les panels de gènes	138
Méthodologie	138
Visualisation et interprétation	139
13.1.5.2. CNV dans l'analyse SomaCGP et SomaLBx	144
Méthodologie	144
Visualisation et interprétation	144
13.1.5.3. CNV dans les exomes et les génomes	145
Méthodologie pour l'exome	145
Méthodologie pour le génome	146
Visualisation et interprétation	146
Tableau des résultats des CNV	146
Page détaillée du CNV	150
13.1.6. Large Variant Viewer pour les CNVs et SVs (LVV)	155
Table des CNVs & SVs	155

Tri, filtrage, visualisation et interprétation	157
Panneau latéral CNV/SV (Details Drawer)	158
13.1.7. Signatures dans l'analyse SomaCGP	159
Méthodologie de calcul du score MSI	159
Méthodologie de calcul de la note du TMB	159
Visualisation et interprétation	160
13.1.8. Répétitions courtes en tandem (STRs)	162
Méthodologie	162
Visualisation et interprétation	162
13.1.9. Gènes de fusion	164
13.1.9.1. Fusions détectées à partir de l'ADN	164
Méthodologie	164
Visualisation et interprétation	164
13.1.9.2. Fusions détectées à partir de l'ARN	166
Méthodologie	166
Visualisation et interprétation	167
13.1.10. Clonalité	170
13.1.11. Variants d'épissage	171
Méthodologie	171
Visualisation et interprétation	171
13.1.12. Statut MSI/MSS	173
Méthodologie	173
Visualisation et interprétation	173
13.1.13. Statut HRD/HRP	175
Méthodologie	175
Visualisation et interprétation	177
13.1.14. Données de méthylation pour SomaMethyl	183
14. Rapport	186
14.1.1. Fichiers de sortie	186
15. Dépannage et messages d'erreur	187
16. Garantie	189
17. Avis juridique	189
17.1. Conditions d'utilisation	189
17.2. Limitation de la responsabilité	189
17.3. Droits d'auteur et marques déposées	189

1. Introduction

Il est conseillé de lire attentivement les instructions avant d'utiliser le logiciel. Une version papier de ces instructions peut être fournie sur demande à l'équipe Support de SeqOne (support@seqone.com). Pour obtenir les meilleurs résultats lors de l'utilisation de la plateforme,

lisez attentivement toutes les précautions mises en évidence par le symbole . Les divers symboles réglementaires et autres symboles utilisés dans ce guide et dans le logiciel sont les suivants :

Symbole	Signification
	Fabricant
	Indique le mandataire établi en Suisse.
	Date de fabrication (date de sortie du logiciel)
	Dispositif médical de diagnostic in vitro.
	Identifiant unique de l'appareil
	Référence du catalogue du fabricant
	Conforme aux exigences du règlement européen (UE) 2017/746 Marquage CE avec numéro d'identification de l'organisme notifié
	Consulter le mode d'emploi
	Avertissement Ce symbole de sécurité indique qu'une attention particulière est requise lors de l'utilisation de l'appareil.

L'étiquette (y compris les informations sur la version du logiciel SeqOne et les coordonnées de contact) et ce document décrivant les instructions d'utilisation (en plusieurs langues) sont disponibles dans la section "About".

2. Contact fabricant et mandataire suisse



SeqOne S.A.S

Web: <https://seqone.com/>

22 rue Durand
34000 Montpellier
FRANCE

Support client :

support@seqone.com



Arazy Group Swiss GmbH

Bruderholzallee 53

4059 Bâle, Suisse

swiss.ar@arazygroup.com

3. Réactovigilance

Si un incident grave se produit ou risque de se produire lors de l'utilisation du dispositif, les utilisateurs sont invités à le signaler sans délai au fabricant et à l'autorité compétente du pays où l'incident s'est produit ou risque de se produire.

La réglementation applicable et la liste des autorités compétentes dans l'UE sont disponibles à l'adresse suivante :

https://health.ec.europa.eu/medical-devices-sector/new-regulations/contacts_en



4. Utilisation prévue de la plateforme

4.1. Utilisation prévue

SeqOne Platform est un dispositif médical de diagnostic *in vitro* basé sur un logiciel, conçu pour aider les professionnels de la santé à analyser et à interpréter les données génomiques et de séquençage brutes des patients soumis à une analyse NGS ou CGH array. En fournissant des informations qualitatives et exploitables, la plateforme fonctionne comme un système d'aide à la décision, aidant les professionnels de la santé à prendre des décisions cliniques.

4.2. Indication d'utilisation

SeqOne Platform prend en charge l'analyse secondaire et tertiaire des données de séquençage brutes, ainsi que l'analyse tertiaire seule. Ses capacités analytiques indépendantes de la pathologie trouvent des applications dans divers domaines médicaux, notamment l'oncologie et les maladies rares.

4.3. Description de la plateforme

SeqOne Platform est un dispositif médical de diagnostic *in vitro* basé sur un logiciel qui intègre une suite d'applications fonctionnant de concert pour fournir aux utilisateurs experts une boîte à outils puissante pour l'interprétation de données génomiques et de séquençage complexes.

SeqOne Platform intègre des informations établies à partir d'un large éventail de **sources externes**, y compris des bases de données publiques, de la littérature scientifique et des bases de connaissances propriétaires. Ces sources de données externes fournissent un contexte et une signification aux variants génomiques identifiés.

Les fichiers d'entrée pour diverses applications au sein de la plateforme sont traités à l'aide d'une combinaison sur mesure de modules bioinformatiques et d'annotation, garantissant ainsi que chaque analyse est optimisée pour ses exigences spécifiques.

Les modules bioinformatiques effectuent une analyse secondaire à l'aide de méthodes informatiques telles que l'alignement et l'appel de variants pour identifier les variants génétiques dans les données de séquençage de l'ADN et de l'ARN en les comparant à un génome de référence (hg19 ou hg38). Les données, qui peuvent être utilisées pour des analyses germinales et somatiques, peuvent provenir du séquençage du génome entier, du séquençage de l'exome ou de panels de gènes ciblés.

Les modules d'annotation effectuent une analyse tertiaire pour annoter les variants avec des détails tels que le gène et sa fonction connue, l'emplacement du variant dans le gène et toute publication ou rapport clinique mentionnant le gène ou du variant. Cette annotation déterminera quelles règles de classification ACMG et AMP / CAP/ASCO s'appliquent à chaque variant et, à partir de là, un verdict de pathogénicité peut être rendu. L'annotation est également effectuée par DiagAI, une suite d'outils alimentée par l'IA qui aide à analyser les variants génétiques afin d'identifier rapidement les variants causals potentiels.

Cette architecture modulaire permet à SeqOne de s'adapter à différents types d'analyses et de questions de recherche. Les fichiers d'entrée des diverses applications de la plateforme sont traités à l'aide d'une combinaison personnalisée de ces modules bioinformatiques et d'annotation, ce qui garantit que chaque analyse est optimisée en fonction de ses exigences spécifiques. En intégrant ces diverses sources de données et outils d'analyse, SeqOne permet

aux utilisateurs experts d'acquérir une connaissance approfondie du paysage complexe des données génomiques et de ses implications pour la santé et les maladies humaines.

La plateforme fonctionne comme une solution SaaS basée sur le cloud, dont toutes les fonctionnalités sont accessibles à partir d'un navigateur web. Elle comprend SeqOne Flow, un outil optionnel qui peut être déployé soit dans le nuage AWS, soit sur les serveurs du client, afin de faciliter le transfert des données et l'intégration des flux de travail.

Pour la sécurité, toutes les données sont cryptées au repos à l'aide d'un cryptage AES-256, et le trafic entre les utilisateurs et la plateforme est sécurisé à l'aide de certificats SSL.

4.4. Population de patients

La solution convient à tout patient nécessitant une analyse génomique. Les patients ciblés peuvent être atteints de cancer, de maladies rares, de maladies héréditaires, d'une prédisposition génétique au cancer ou être en bonne santé.

4.5. Utilisateurs visés

SeqOne Platform est destiné aux professionnels de la santé chargés des analyses génomiques cliniques de routine. Il s'agit notamment d'analyser et d'interpréter les fichiers bruts générés par les séquenceurs à partir d'échantillons biologiques séquencés. Les utilisateurs cibles de la plateforme sont les généticiens, les bioinformaticiens, les biologistes, les ingénieurs et les techniciens de plateformes NGS.

4.6. Caractéristiques

Chaque application (workset) se compose de plusieurs **modules de bio-informatique et d'annotation**, chacun étant chargé de tâches analytiques spécifiques. Dans ce chapitre, les performances sont évaluées au niveau des modules afin de garantir une évaluation détaillée de la précision et de la fiabilité de chaque composant.

Les analytes détectés par la SeqOne Platform se répartissent en deux catégories d'altérations génomiques :

- Variants génétiques tels que les SNV, les indels, les MNV, les CNV, l'UPD, les fusions géniques, les variants structuraux, variants dépiçage, les insertions Alu et les FLT3-ITD (duplication interne en tandem)
- Signatures ou biomarqueurs complexes tels que le statut IGHV, l'instabilité des sites microsatellites (MSI), la charge mutationnelle tumorale (TMB), la méthylation ou les défauts de réparation des cassures double brin par recombinaison homologue (HRD, Homologous Recombination Deficiency).

Module de bio-informatique	Workset	Type d'analyte
Détection des CNV Exome Germline	CNVExome, GermVar, GermVar Family	Variant génétique
Détection des CNV Panel Germline	CNVCapture	Variant génétique
SNV/Indel Germline	GermlineVar, GermlineFamily	Variant génétique
SV Germline	GermlineVar, GermlineFamily	Variant génétique
UPD	GermVar, GermVar Family, GermVar Tertiary, GermVar Tertiary Family	Variant génétique
Insertions Alu	GermlineVar, GermlineFamily, SomaVar, SomaVar LF, SomaCGP, SomaHemato, GermVar, GermVar Family	Variant génétique
MNV	GermlineVar, GermlineFamily, GermVar, GermVar Family, GermVar Tertiary, GermVar Family Tertiary, SomaVar, SomaVar LF, SomaCGP, SomaHemato, SomaLBx, SomaVar Tertiary	Variant génétique
SNV/Indel somatique	SomaVar, SomaHemato, SomaRNA, SomaCGP	Variant génétique
SNV/Indel somatique LF	SomaVar LF	Variant génétique
CNV Panels somatiques larges	SomaCGP	Variant génétique
Index de méthylation	SomaMethyl	Signature
Fusion d'ADN somatique	SomaCGP, SomaVar, SomaVar LF, SomaHemato	Variant génétique
Fusion d'ADN somatique LBx	SomaLBx	Variant génétique
TMB	SomaCGP	Signature
MSI	SomaCGP, SomaMSI	Signature
Probabilité GI (GiP)	SomaHRD	Signature
Sentieon SNV/Indel Germline	GermVar, GermVar Family	Variant génétique
Variants mosaïques Germline	GermVar, GermVar Family	Variant génétique
CNV WGS	GermVar, GermVar Family	Variant génétique
Sentieon SNV/Indel somatique	SomaLBx	Variant génétique
ARN de fusion	SomaRNA	Variant génétique
Splicing d'ARN	SomaRNA	Variant génétique
CNV Petit Panel Somatique	SomaCNVCapture	Variant génétique
Statut IGHV	SomaHemato	Signature
Sous-ensembles IGHV (subsets)	SomaHemato	Signature
CFH	GermlineVarLR	Variant génétique
FLT3	SomaVar, SomaHemato	Variant génétique

4.6.1. Performances analytiques

Les performances indiquées dans cette notice d'utilisation correspondent à celles des dernières versions IVD des worksets, qui sont répertoriées dans le tableau ci-dessous :

Nom du workset	Version
CNVCapture	2.6
CNVExome	2.2
GermlineVar	2.6
GermlineFamily	2.6
GermlineVar LR	1.2
GermVar	3.12
GermVar Family	3.12
GermVar Tertiary	1.16
GermVar Tertiary Family	1.14
SomaCGP	1.6
SomaCNVCapture	2.4
SomaHemato	0.6
SomaHRD	1.5
SomaLBx	0.7
SomaMethyl	0.3
SomaMSI	1.1
SomaRNA	3.3
SomaVar	2.9
SomaVar LF	2.1
SomaVar Tertiary	2.4

4.6.1.1. Modules bioinformatiques

- **Détection des CNV Exome Germline et Panel Germline**

Type de variant	Version du génome	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
Détection des CNV Exome Germline (GermlineVar)	GRCh37	8	8	0	100%	100%
Détection des CNV Exome Germline (GermVar)	GRCh37	8	8	0	100%	100%
Détection des CNV Panel Germline	GRCh37	15	15	0	100%	100%
Détection des CNV Panel Germline	GRCh38	11	11	0	100%	100%

- **SNV / Indel Germline - Données du panel**

Version du génome	Type de variant	Critères d'acceptation	Sensibilité avant filtrage	Sensibilité après filtrage
GRCh37	SNV	100 %	98 %	100 %
GRCh38	SNV	100 %	98 %	100 %

Version du génome	Type de variant	Critères d'acceptation	Précision avant filtrage	Précision après filtrage
GRCh37	SNV	100 %	90 %	100 %
GRCh38	SNV	100 %	90 %	100 %

- **SNV / Indel Germline - Données d'exome**

Version du génome	Type de variant	Critères d'acceptation*	Sensibilité avant filtrage	Sensibilité après filtrage
GRCh37	SNV	> 98 %	98,7 %	98,5 %
	Indel	> 80 %	88,0 %	87,5 %
GRCh38	SNV	> 98 %	98,5 %	98,3 %
	Indel	> 80 %	84,9 %	84,3

* pour la détection des variants situés dans les régions à haute confiance de NA12878 et NA24385.

Version du génome	Type de variant	Critères d'acceptation*	Précision avant filtrage	Précision après filtrage
GRCh37	SNV	> 90 %	92,1 %	94,4 %
	Indel	> 80 %	84,7 %	89,0 %
GRCh38	SNV	> 90 %	94,2 %	95,4 %
	Indel	> 80 %	82,2 %	87,6 %

* pour la détection des variants situés dans les régions à haute confiance de NA12878 et NA24385.

- **SNV / Indel Germline - Repeatabilité and reproducibilité**

- **Répétabilité**

Type de variants	Sensibilité	Précision
SNV	98,59 % ± 0 %	95,92 % ± 0 %
indel	87,75 % ± 0 %	81,63 % ± 0 %
(all)	98,38 % ± 0 %	95,63 ± 0 %

- **Reproductibilité**

Type de variant	Sensibilité	Précision
SNV	98,57 % ± 0,03 %	95,93 % ± 0,08 %
indel	86,58 % ± 0,42 %	78,97 % ± 0,96 %
"All"	98,34 % ± 0,03 %	95,58 ± 0,09 %

- **Variants mosaïques Germline**

- **Sensibilité SNV**

Version du génome	Bins VAF	Nombre de variants réels	Critères d'acceptation	Sensibilité 100x
GRCh38	[0,05, 0,1[	3704	40 %	47,6 %
GRCh38	[0,1, 0,2[	747	70	73,1 %
GRCh38	[0,2,0,3[	685	90 %	90,7 %
GRCh38	[0,3,1,0[	76	90 %	93,4 %

- **Sensibilité aux indels**

Version du génome	Bins VAF	Nombre de variants réels	Critères d'acceptation	Sensibilité 100x
GRCh38	[0,05, 0,1[	87	40 %	43,7 %
GRCh38	[0,1, 0,2[	14	70	78,6 %
GRCh38	[0,2,0,3[	15	90 %	100 %
GRCh38	[0,3,1,0[	0	"N/A"	"N/A"

- **SV Germline**

Type de variant	Version du génome	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
SV de la lignée germinale	GRCh37	2	2	0	100 %	100 %

- **Insertions Alu**

Version du génome	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
GRCh37	1	1	0	100 %	100 %
GRCh38	1	1	0	100 %	100 %

- **MNV**

Version du génome	Type de variant	Critères d'acceptation	Précision de l'exome	Précision du génome
GRCh37	MNV	> 85 %	87,0 %	86,9 %
GRCh38	MNV	> 85 %	93,0 %	91,5 %

- **SNV / Indel somatique**

- **Sensibilité**

Version du génome	Type de variant	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation*	Sensibilité
GRCh37	SNV (≥ 5 %)	121	121	0	100 %	100 %
	Indel (≥ 5 %)	21	21	0	100 %	100 %
GRCh38	SNV (≥ 5 %)	121	121	0	100 %	100 %
	Indel (≥ 5 %)	21	21	0	100 %	100 %

*Le critère d'acceptation pour la détection des variants est défini comme une sensibilité de 100 % pour les SNV et les indels dont la VAF attendue est supérieure à 5 %.

○ **Exactitude**

			Exactitude			
			GRCh37		GRCh38	
'Gene'	Variant	Fréquence attendue	Fréquence *	Biais	Fréquence fréquence*	Biais
KIT	p.D816V	10 %	9,07 % ± 0,79 %	0,97 %	9,07 % ± 0,79 %	0,97 %
PIK3CA	p.E545K	8,80 %	7,88 % +/- 0,84 %	1,04 %	7,87 % ± 0,83 %	1,05 %
EGFR	p.E746_A750del	1,90 %	1,53 % ± 0,5 %	0,40 %	1,53 % ± 0,5 %	0,40 %
KRAS	p.G12D	6,30 %	6,27 % ± 0,76 %	0,57 %	6,27 % +/- 0,76 %	0,57 %
KRAS	p.G13D	15 %	14,59 % +/- 0,86 %	0,69 %	14,59 % ± 0,86 %	0,69 %
EGFR	p.G719S	24,50 %	24,5 % ± 1,39 %	1,15 %	24,5 % ± 1,39 %	1,15 %
PIK3CA	p.H1047R	17,50 %	17,47 % ± 1,04 %	0,77 %	17,47 % ± 1,04 %	0,77 %
EGFR	p.L858R	2,80 %	3,58 % ± 0,51 %	0,83 %	3,58 % ± 0,51 %	0,83 %
NRAS	p.Q61K	12,50 %	11,27 % +/- 1,19 %	1,27 %	11,27 % ± 1,19 %	1,27 %
EGFR	p.T790M	0,90 %	1,33 % ± 0,15 %	0,43 %	1,33 % ± 0,15 %	0,58 %
BRAF	p.V600E	10,70 %	11,68 % ± 0,62 %	0,98 %	11,68 % ± 0,62 %	0,98 %

*La moyenne des fréquences est obtenue à partir de 12 réplicats de l'échantillon HD200.

● **SNV/Indel Somatic LF**

Génome Version	Contrôles contrôles	Véritables Faux	Faux négatifs	Critères d'acceptation *	Sensibilité
GRCh37	42	42	0	100 %	100 %
GRCh38	42	42	0	100 %	100 %

*Le critère d'acceptation pour la détection des variants est défini comme une sensibilité de 100 % pour les SNV et les indels dont la fréquence allélique attendue (VAF) est supérieure à 0,1 %.

● **Grands panels de CNV somatiques**

Version du génome	Type de variant	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
GRCh37	Gain	27	26	1	90 %	96,3 %
	Loss	20	19	1	90 %	95,0 %
	LOH	11	11	0	90 %	100,0 %
GRCh38	Gain	27	26	1	90 %	96,3 %
	Loss	20	19	1	90 %	95,0 %
	LOH	11	10	1	90 %	90,9 %

- **Indice de méthylation**

Version du génome	Critères d'acceptation	Concordance pondérée
GRCh37	>99 %	99,88

- **Fusion d'ADN somatique**

Génome Version	Contrôles Contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	9	9	0	100 %	100 %
GRCh38	9	9	0	100 %	100 %

- **Fusion d'ADN somatique LBx**

Version du génome	Fraction tumorale	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
GRCh38	1 %	22	21	1	85 %	95,5 %
	0,5 %	22	19	3	85 %	86,4 %
	0,1 %	22	21	1	85 %	95,5 %

- **TMB**

Version du génome	Critères d'acceptation	Sensibilité	Précision
GRCh37	>80 %	83 %	85 %

- **MSI**

Version du génome	Témoins positifs	Contrôles négatifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité	Précision
GRCh37	22	24	22	0	100 %	100 %	100 %
GRCh38	6	17	6	0	100 %	100 %	100 %

- **GiP**

Génome version	Contrôles contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	171	165	6	90 %	96,5 %

Génome Version	Contrôles Témoins	Véritables Faux	Faux Faux	Acceptation Critères	Précision
GRCh37	171	165	7	90 %	95,9 %

- Sentieon SNV / Indel Germline

Version du génome	Type de variant	Critères d'acceptation *	Sensibilité Exome	Sensibilité du génome
GRCh37	SNV	> 90 %	98,9 %	98,5 %
	Indel	> 80 %	96,5 %	98,5 %
GRCh38	SNV	> 90 %	98,5 %	99,3 %
	Indel	> 80 %	95,5 %	99,1 %

* pour la détection des variants situés dans les régions à haute confiance de NA12878 et NA24385

Version du génome	Type de variant	Critères d'acceptation *	Précision Exome	Précision du génome
GRCh37	SNV	> 90 %	98,7 %	98,5 %
	Indel	> 80 %	96,5 %	98,6 %
GRCh38	SNV	> 90 %	99,6 %	99,6 %
	Indel	> 80 %	97,2 %	99,5 %

* pour la détection des variants situés dans les régions à haute confiance de NA12878 et NA24385.

- CNV WGS

Version du génome	Contrôles positifs	Vrais positifs	Faux positifs	Critères d'acceptation	Sensibilité
GRCh38	Duplication	1 kb - 5 kb	8	10 %	37,5 %
	Duplication	≥ 5 kb	3	10 %	100 %
	Délétion	1 ko - 5 ko	321	80 %	61,7 %
	Délétion	≥ 5 kb	132	80 %	92,4 %

Bien que la sensibilité aux petites délétions (1-5 kb) soit inférieure aux critères d'acceptation, cette réduction constitue un compromis acceptable afin de maximiser la détection d'événements de grande ampleur et de duplications ayant un impact plus important.

Version du génome	Contrôles positifs	Vrais Vrais	Faux positifs	Critères d'acceptation	Précision
GRCh38	Duplication	1 kb - 5 kb	8	"NA"	30 %
	Duplication	≥ 5 kb	3	"NA"	75 %
	Délétion	1 kb - 5 kb	321	"NA"	94,3 %
	Délétion	≥ 5 kb	132	"NA"	92,4 %

- **Sentieon SNV/Indel somatique**

Version du génome	Type de variant	Contrôles positifs	Vrais positifs	Faux positifs	Critères d'acceptation	Sensibilité
GRCh38	SNV	39	37	2	85 %	95 %
	Indel	21	16	5	75 %	76 %

- **ARN de fusion**

Génome Version	Contrôles contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	20	20	0	100 %	100 %
GRCh38	20	19	1	95 %	95 %

- **ARN de splicing**

Génome Version	Contrôles contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	3	3	0	100 %	100 %
GRCh38	3	3	0	100 %	100 %

- **Petit panel de CNV somatiques**

Génome Version	Contrôles Contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	4	4	0	100 %	100 %
GRCh38	4	4	0	100 %	100 %

- **Statut IGHV IMGT**

Génome version	Contrôles Contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	122	116	6	95 %	95,9 %

- **Statut IGHV igblast**

Version du génome	Contrôles positifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
GRCh37	122	117	5	95 %	95,1 %

- **Sous-ensembles IGHV**

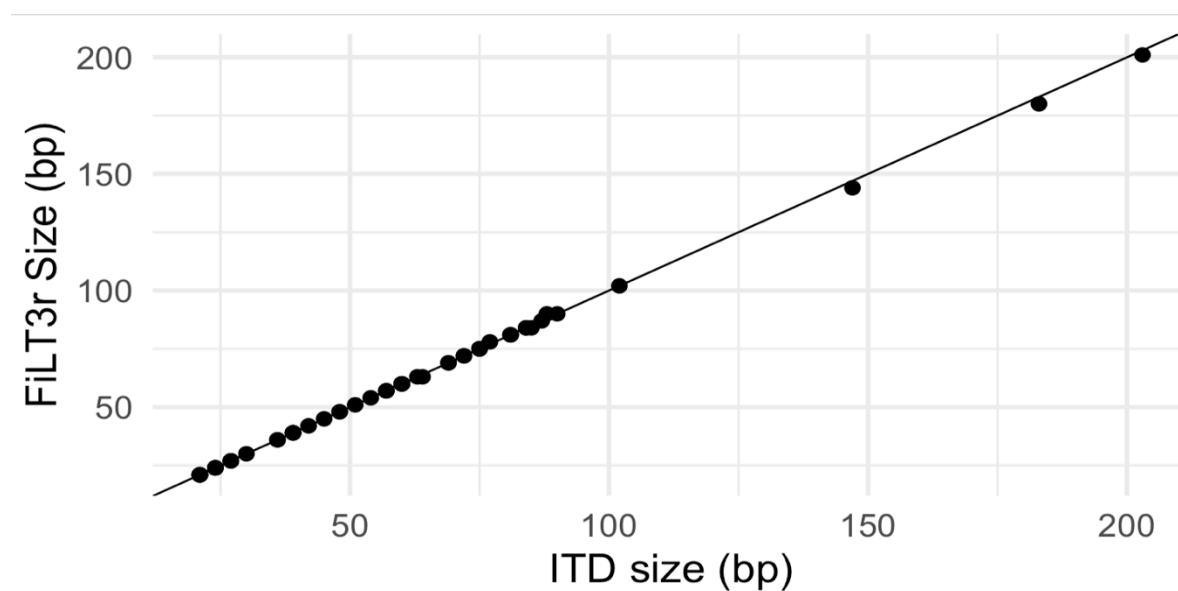
Génome Version	Contrôles véritables	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh37	61	10	0	95 %	100 %

Génome Version	Contrôles Témoins	Véritables Faux	Faux négatifs	Acceptation Critères	Précision
GRCh37	61	10	0	95 %	100 %

- CFH

Génome version	Contrôles Contrôles	Véritables Faux	Faux négatifs	Acceptation Critères	Sensibilité
GRCh38	4	4	0	100 %	100 %

- FLT3



Taille observée vs taille théorique des événements FLT3-ITD. Coefficient de Pearson R^2 **$> 0,99$**

Les résultats montrent une sensibilité de 100 % pour la détection des événements FLT3-ITD et une corrélation quasi parfaite entre les tailles d'événements attendues et observées issues de l'analyse.

- UPD

- Sensibilité

Type d'analyse	Version du génome	Contrôles positifs	Contrôles négatifs	Vrais positifs	Faux négatifs	Critères d'acceptation	Sensibilité
Singleton	GRCh37/38	7	23	7	0	80 %	100 %
Trio	GRCh37/38	7	2	6	1	80 %	86 %

- Précision

Type d'analyse	Version du génome	Témoins positifs	Contrôles négatifs	Vrais positifs	Faux positifs	Critères d'acceptation	Précision
Singleton	GRCh37/38	7	23	7	0	80 %	100 %
Trio	GRCh37/38	7	2	6	0	80 %	100 %

4.6.1.2. Modules d'annotation

- **Score DiagAI**

Ensemble de données	Analyse de type	Métrique évalué	Acceptation Critères	Métrique Valeur
Exomes positifs pour les maladies rares	"Germline"	Rang médian	= 1	1
Exomes positifs pour les maladies rares	"Germline"	top_10	> 90 %	95 %
Exomes CNV de maladies rares	"Germline"	Rang médian	= 1	1
Exomes CNV positifs pour les maladies rares	'Germline'	top_1	> 80 %	83,3 %

- **Score somatique DiagAI**

Ensemble de données	Analyse de type	Mesure évaluée	Critères d'acceptation	Valeur de la métrique
Cancer positif TSO500	'Somatic'	top_10	> 90 %	100 %

- **Listes de présélection DiagAI**

Ensemble de données	Analyse des types	Indicateur évaluée	Acceptation Critères	Valeur de la métrique
Exomes positifs pour les maladies rares	'Germline'	Sensibilité	> 90	95 %
Exomes positifs pour les maladies rares	"Germline"	Longueur moyenne	< 25	18,475
Exomes positifs pour les maladies rares	"Germline"	Longueur médiane	< 25	17

- **DiagAI SmartPicks**

Ensemble de données	Analyse des types	Mesure évaluée	Acceptation Critères	Valeur de la métrique
Exomes positifs pour les maladies rares	"Germline"	Sensibilité	> 40 %	45 %
Exomes positifs pour les maladies rares	"Germline"	Précision	> 80 %	78 %

- **ACMG CNV**

Ensemble de données	Analyse de type	Mesure évalué	Critères d'acceptation	Valeur de la métrique
Ensemble de données CNV annoté par des experts de l'ACMG	'Germline'	Sensibilité	> 70 %	76 %

- **SNV de l'ACMG**

Ensemble de données	Analyse de type	Mesure évalué	Critères d'acceptation	Valeur de la métrique
Ensemble de données annoté par des experts de l'ACMG	'Germline'	Précision	> 50 %	56 %

- **DiagAI UP2**

Ensemble de données	Analyse des types	Mesure évaluée	Critères d'acceptation	Métrique Valeur
Variants pathogènes	"Germline"	Sensibilité	90 %	92 %

- **DiagAI UP2 SV**

Ensemble de données	Analyse de type	Mesure évaluée	Critères d'acceptation	Métrique Valeur
Variants pathogènes	'Germline'	Sensibilité	90 %	98 %

- **DiagAI Phenogenius**

Ensemble de données	Analyse de type	Mesure évaluée	Critères d'acceptation	Métrique valeur
Associations gènes-symptômes	"Germline"	top_100	50 %	49,8 %
Associations gènes-symptômes	"Germline"	top_200	60 %	60,7 %
Associations gènes-symptômes	'Germline'	top_500	75 %	77,0 %
Associations gènes-symptômes	'Germline'	Médiane dans le top 100	<20	15
Associations gènes-symptômes	"Germline"	Moyenne dans le top 100	<30	25

- **Compermed**

Ensemble de données	Type d'analyse	Mesure évalué	Critères d'acceptation	Métrique valeur
Cancer positif TSO500	"Somatic"	Sensibilité	90 %	91 %

4.6.2. Performances cliniques

La SeqOne Platform n'est pas destinée à être utilisée pour l'évaluation d'une affection clinique ou d'un état physiologique particulier. Bien qu'elle fournisse des informations sur l'état physiologique d'un patient, telles que l'identification de certains variants, elle n'est pas destinée à évaluer des affections cliniques ou des états physiologiques spécifiques. Par conséquent, la SeqOne Platform elle-même ne nécessite pas de démonstration de performance clinique.

5. Précautions

5.1. Utilisation conforme

(Voir la section « **Utilisateurs visés** »)

Toute utilisation inappropriée de la plateforme est interdite.

Il incombe aux utilisateurs de vérifier que la plateforme est opérationnelle à tout moment et de garantir les conditions nécessaires à son bon fonctionnement.

5.2. Avertissements

 Les résultats fournis ne doivent pas constituer la seule base pour modifier la prise en charge du patient. Par conséquent, les utilisateurs doivent faire preuve de la prudence requise lors de l'utilisation du logiciel et de ses résultats. Les diagnostics doivent être confirmés par divers examens cliniques et médicaux.

L'analyse génomique de nouvelle génération (NGS) peut révéler des résultats secondaires ou fortuits, tels qu'une prédisposition au cancer, provenant de gènes sans rapport avec la pathologie principale (Kuhlen et Borkhardt 2015). Malgré les déclarations de principe d'organismes tels que l'AMP et l'ACMG (Green et al. 2013 ; Allyse et Michie 2013 ; Conseil d'administration de l'ACMG 2015 ; Hegde et al. 2015), les directives nationales relatives à la communication des résultats fortuits varient. Il est donc essentiel d'obtenir un document de « consentement éclairé » (Anderson et al. 2016). Le clinicien traitant est chargé de s'assurer que le patient ou son représentant légal comprend les résultats potentiels et les risques avant de prescrire tout test NGS.

Ce produit ne garantit pas la détection de toutes les variants génétiques présentes dans un échantillon. La probabilité de détecter une variants d'intérêt dépend intrinsèquement de la qualité et des caractéristiques de l'échantillon d'entrée (par exemple, l'intégrité de l'échantillon, la préparation de la bibliothèque, la profondeur et l'uniformité du séquençage, le filtrage bioinformatique, ainsi que le type ou le contexte de la variants). Un résultat négatif n'exclut pas définitivement la présence d'une variants pathogène si sa fréquence (VAF) est inférieure aux seuils de détection du module analytique utilisé. La détection fiable de réarrangements génomiques à grande échelle (variants structurales, y compris les CNVs et les délétions de chromosomes entiers) n'est pas garantie. Cela est dû à la limitation inhérente aux données de séquençage de nouvelle génération (NGS) à lectures courtes, qui rend ces événements difficiles à identifier. Les lectures NGS étant courtes (environ 150 à 200 pb), cette contrainte est une caractéristique de la technologie de séquençage elle-même et ne constitue pas une limitation de la SeqOne Platform. (Voir Tan R. et al., Hum Mutat. Juillet 2014 ; 35(7) : 899-907. doi : 10.1002/humu.22537 et les références qui y sont citées). Les utilisateurs doivent évaluer

l'adéquation de l'ensemble de travail sélectionné et la qualité du séquençage par rapport aux variants attendus afin de s'assurer que le contexte clinique approprié a été couvert.

Les décisions cliniques fondées sur les résultats du dépistage néonatal ou du dépistage des porteurs doivent être prises avec une extrême prudence et ne jamais reposer uniquement sur les résultats de ce logiciel. Pour le dépistage néonatal, un résultat positif nécessite une corrélation clinique immédiate avec les symptômes du nourrisson et des tests de confirmation urgents avant d'initier des traitements susceptibles de modifier le cours de la vie ou des restrictions alimentaires. Pour le dépistage des porteurs, un résultat négatif n'élimine pas le risque d'être porteur, mais le réduit à un niveau de risque résiduel, car certaines variants génomiques peuvent se situer en dessous de la limite de détection du logiciel. Étant donné que les résultats de ce dispositif influencent directement des choix reproductifs irréversibles et des soins néonataux critiques, tous les résultats donnant lieu à une action doivent être vérifiés à l'aide d'une autre méthode validée et interprétés par un professionnel qualifié dans le contexte de l'historique clinique complet du patient et des données biochimiques ou phénotypiques pertinentes.

Les variants récemment découverts qui ne sont pas encore documentés de manière exhaustive dans la littérature scientifique peuvent être mal classés.

Les résultats sont exacts au moment de l'analyse et reflètent les connaissances scientifiques actuelles. Cependant, une nouvelle analyse effectuée sur une version plus récente de la plateforme peut donner des résultats différents.

Les résultats de classification de la SeqOne Platform visent à s'aligner sur le consensus général. Sachez que tout résultat généré après modification des paramètres de classification ne relève pas du champ d'application de la directive CE-IVD et n'est donc pas validé.

Ce logiciel intègre des composants d'intelligence artificielle (IA). Les utilisateurs doivent être conscients des limites suivantes inhérentes à ces outils basés sur l'IA et en tenir compte lors de l'interprétation des résultats :

- Les données d'apprentissage proviennent principalement de populations d'ascendance européenne/caucasienne en raison de la composition des bases de données cliniques de référence actuelles ;
- Les performances peuvent varier pour les variants rares peu représentés dans les données d'apprentissage ;
- Les scores de confiance reflètent l'incertitude du modèle, et non la certitude clinique ; ils doivent être interprétés en tenant compte du contexte clinique ;
- Les performances peuvent se dégrader en raison de la dérive des données provenant de nouvelles plateformes de séquençage ou de protocoles de préparation de bibliothèques non représentés dans les données d'entraînement ;
- Les performances de DiagAI Phenogenius en matière de hiérarchisation des gènes sont évaluées sur la base des associations gène-symptôme et dépendent de la

qualité et de l'exhaustivité de l'annotation des termes HPO par le clinicien prescripteur.

5.3. Limites

Le dépistage des porteurs de maladies génétiques est soumis à une législation nationale et à des réglementations éthiques spécifiques. Il incombe au laboratoire et au prestataire de soins de santé de s'assurer que toutes les exigences légales – y compris le conseil génétique obligatoire, la prescription par un médecin et le consentement éclairé – sont respectées dans leur juridiction spécifique.

5.4. Dysfonctionnements

En cas de dysfonctionnement :

- Cessez immédiatement d'utiliser le logiciel
- Essayez d'identifier ou d'éliminer les causes en suivant la procédure de dépannage décrite dans ce guide (voir la section **Dépannage et messages d'erreur**)
- Si vous ne parvenez pas à identifier ou à éliminer la cause à l'aide de ce guide, veuillez envoyer un e-mail à l'équipe d'assistance à l'adresse support@segone.com .

6. Prérequis

6.1. Matériel

Les modules logiciels de la plateforme SeqOne sont installés dans le cadre d'un arrangement client-serveur. Le fournisseur de services doit être un serveur Linux pour garantir la compatibilité avec la plupart des outils couramment utilisés en bio-informatique. Chaque élément du système est encapsulé dans un environnement d'exécution tel que le conteneur Docker®, qui peut être déployé sur toute plateforme Docker® conforme, indépendamment de l'architecture matérielle sous-jacente. Les utilisateurs doivent disposer d'un ordinateur avec une connexion internet et un navigateur web (voir ci-dessous).

La configuration minimale requise pour un ordinateur client est la suivante :

- 4 GB RAM
- Processeur double cœur
- Accès à l'internet
- Navigateur Chrome (version 36.0.1985 et suivantes) ou Firefox (version 31 et suivantes) ou Microsoft Edge (toutes les versions basées sur Chromium).

6.2. Réseau

La vitesse minimale recommandée pour l'utilisation de la plateforme sur l'internet est de 100 Mbps. Le protocole HTTPS est utilisé pour se connecter à la plateforme. Pour accéder correctement à la plateforme, nous vous recommandons de demander à votre service informatique d'autoriser les adresses suivantes :

*.eu.seqone.com

*.common.seqone.com

7. Authentification

7.1. Connexion

Pour des raisons de sécurité, les utilisateurs ne peuvent accéder à la plateforme qu'après s'être identifiés sur l'écran d'accueil.

L'écran d'accueil est disponible via l'URL suivante :

- Pour l'Europe et l'Afrique : <https://auth.eu.seqone.com/>
- Pour la Suisse : <https://auth.ch.seqone.com/>
- Pour l'Asie et l'Océanie : <https://auth.au.seqone.com/>
- Pour les Amériques : <https://auth.us.seqone.com/>

Lorsque vous vous connectez sur l'écran d'accueil, vous devez saisir votre :

- Identifiant ou adresse email dans le champ "Username" or "email"
- Mot de passe dans le champ "Password"

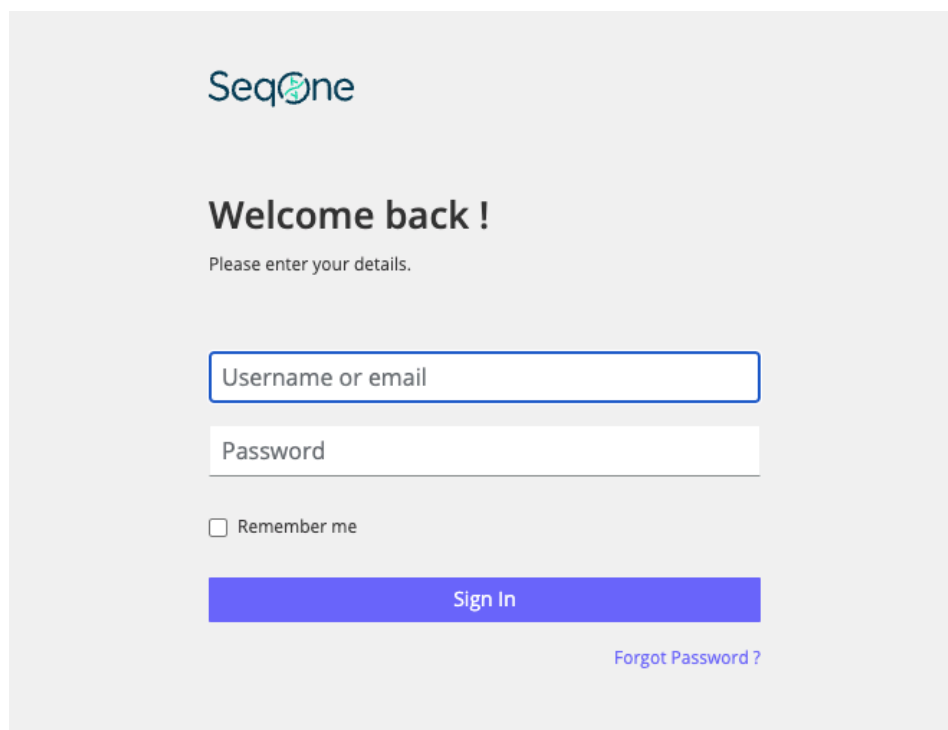


Figure 1 : Page de connexion

Si l'authentification à deux facteurs est activée pour votre compte, vous devrez saisir un code de vérification à 6 chiffres envoyé par courriel à l'adresse associée à votre compte SeqOne. Ce code est valable pendant 30 minutes.

Si vous avez accès à plusieurs espaces de travail ou entités, il vous sera demandé de sélectionner l'entité à laquelle vous souhaitez vous connecter via un menu déroulant sous "Choose a workspace".

Vous pouvez modifier votre mot de passe en cliquant sur "Forgot Password". Suivez les instructions en entrant l'adresse électronique associée à votre compte SeqOne, et vous recevrez un courriel vous permettant de réinitialiser votre mot de passe.

La politique en matière de mots de passe suit les règles suivantes :

- Longueur minimale : 12 caractères.
- Complexité : les mots de passe doivent comprendre au moins une lettre majuscule, une lettre minuscule, un chiffre et un caractère spécial. Plus le jeu de caractères est important, plus le mot de passe est fort.
- Historique des mots de passe : pas de réutilisation des mots de passe précédents.
- Pas de longueur maximale : pour permettre l'utilisation de phrases secrètes.

Il n'est pas possible pour les utilisateurs de modifier leur propre identifiant.

7.2. Déconnexion

Les utilisateurs peuvent se déconnecter en cliquant sur le logo du compte utilisateur dans le coin supérieur droit de la page, puis sur "Logout".

7.3. Page d'accueil

Après authentification, l'utilisateur accède à la page du tableau de bord, qui est divisée en deux parties :

- un bandeau bleu de recherche et d'action, accessible tout au long de la navigation, comprenant de gauche à droite :
 - l'icône SeqOne pour revenir à tout moment au tableau de bord,
 - une barre de recherche pour un accès direct à une analyse donnée,
- une icône de recherche VKB  pour la recherche de variants évalués dans la base de données VKB (voir la section **Évaluation des variants**),
- une icône  ouvrant la "To do list" de l'utilisateur, à partir de laquelle il peut accéder aux analyses qui lui ont été assignées (voir la section **Attribution des analyses pour l'interprétation**),
- une barre de menu, accessible via l'icône  , qui donne accès aux paramètres de l'utilisateur, au panneau de l'administrateur, au manuel de l'utilisateur, aux

conditions d'utilisation, à l'adresse électronique du support et qui permet à l'utilisateur de se déconnecter (voir la section **Gestion des comptes utilisateurs**).

- l'écran principal de gestion des projets et des analyses.

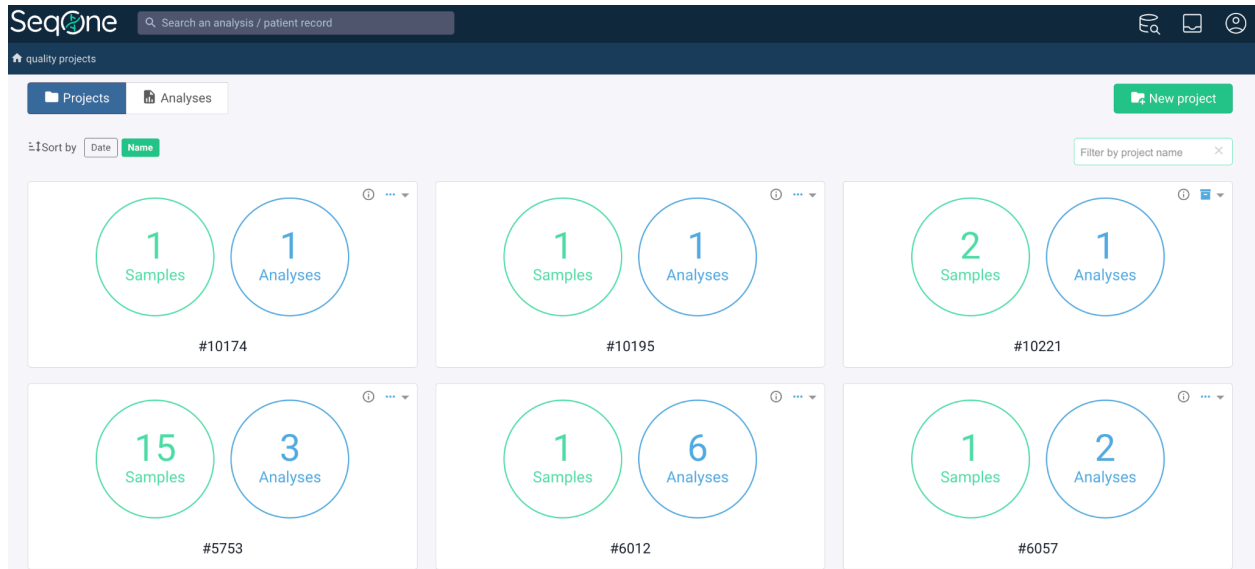


Figure 2 : Tableau de bord de l'utilisateur

Vous pouvez accéder au tableau de bord de l'utilisateur à tout moment en cliquant sur le logo  .

7.4. Gestion des comptes utilisateurs

Vous pouvez gérer votre compte utilisateur en cliquant sur le logo dans le coin supérieur droit de la barre des tâches.

Dans cette section, l'utilisateur a la possibilité de :

- accéder aux paramètres du compte en cliquant sur "My account"
- accéder aux paramètres utilisateur en cliquant sur "Entity Settings"
- changer l'entité à laquelle l'utilisateur est connecté en cliquant sur "Entity", et en sélectionnant l'entité de votre choix sous "Choose your workspace",
- accéder au panneau d'administration, en cliquant sur "Admin", permettant l'accès à la liste des utilisateurs ayant accès à l'entité, ainsi qu'à la création, la modification ou la suppression de comptes utilisateurs.
- cliquer sur "About" pour accéder au manuel d'utilisation en plusieurs langues, aux informations sur la version du logiciel SeqOne et aux coordonnées de contact,
- accéder aux conditions de votre licence SeqOne en cliquant sur "Terms of Use",

- cliquer sur "Support" pour accéder à l'adresse e-mail d'assistance pour toute question ou problème rencontré sur la plateforme,
- accéder aux raccourcis en cliquant sur "Keyboard Shortcuts"
- basculer entre la version standard et la version preview de l'application en cliquant sur "SeqOne Mode", et en sélectionnant soit "SeqOne" (standard mode) soit "SeqOne Preview" (preview mode). Le mode actuellement actif est indiqué par une coche.
- se déconnecter en cliquant sur "Logout".

7.4.1. Gestion de "My Account"

Dans la section "My account", l'utilisateur peut :

- télécharger sa signature pour l'inclure dans le rapport clinique,
- changer son mot de passe en cliquant sur "Change your password" (ouvre un nouvel onglet) dans la section "Manage your credentials".

7.4.2. Gestion des paramètres utilisateur - définitions

Dans les paramètres utilisateur, l'utilisateur peut gérer les paramètres, les fichiers et les configurations décrits ci-dessous.

Entity

L'entité désigne l'établissement ou l'équipe qui a souscrit une licence pour la plateforme SeqOne. La plateforme est conçue pour être utilisée par plusieurs membres d'un même groupe. Dans les paramètres de l'utilisateur, l'onglet *Account* permet de gérer les informations relatives à l'entité.

Users

Pour cet onglet, veuillez vous référer à l'administration - présentation de l'interface.

Workset

Cet onglet liste les différents worksets d'analyse disponibles sur l'entité à laquelle l'utilisateur est connecté. Plusieurs versions de certains pipelines pouvant coexister sur la plateforme SeqOne, l'utilisateur peut définir à partir de ce menu quelle version doit être pré-sélectionnée lors du lancement de l'analyse.

Les modifications apportées à cette liste sont tracées, de sorte que l'utilisateur qui a défini la version à lancer par défaut d'un workset est automatiquement affiché, ainsi que la date de la modification.

Manifest

Le manifeste affiche les régions génomiques d'intérêt (ROI) sous la forme d'un fichier BED, qui peut être accompagné des annotations correspondantes. Vous pouvez télécharger un nouveau manifeste via l'onglet *Manifest* dans les paramètres utilisateur en cliquant sur "Add manifest" et en sélectionnant le fichier de manifeste que vous souhaitez télécharger. Une fois téléchargé, ce manifeste sera disponible pour tous les utilisateurs lors de la création d'un nouveau projet et lors de l'application d'un panel in silico dans une analyse.

Panels - Transcript

Les utilisateurs peuvent télécharger une nouvelle liste de transcrits de référence (RefSeq) via l'onglet *Transcript*.

Panels - Genes

Cet onglet vous permet d'importer ou de générer des panels de gènes basés sur le NCBI GeneID en cliquant sur "Add gene list".

Le format de fichier attendu est un fichier tabulaire (par exemple un fichier .tsv) sans en-tête de colonne, qui est détaillé sur la plateforme. La première colonne du fichier attendu est consacrée aux identifiants des gènes (NCBI Entrez gene ID), la deuxième colonne (facultative) contient le transcrit d'intérêt (RefSeq ID) et la troisième colonne (facultative) contient le symbole HGNC du gène. Ce fichier peut également être généré en insérant la liste souhaitée de gènes, séparés par une virgule, dans le champ 'Gene symbols'.

Marker

Cet onglet permet d'importer une liste de marqueurs d'identité génétique en cliquant sur "Add marker config". Le format de fichier attendu est un fichier tabulaire (fichier .tsv ou .txt) avec un en-tête de colonne contenant 5 colonnes détaillant : le chromosome (CHR), l'identifiant du variant ou rsID (ID), la position génomique (POS), l'allèle de référence (REF) et l'allèle alternatif (ALT). Le fichier ne doit pas contenir plus de 100 lignes, y compris l'en-tête.

Protocols

Un protocole d'analyse est une combinaison prédéfinie de tâches conçues pour standardiser l'interprétation des analyses. L'onglet correspondant permet aux utilisateurs de visualiser la liste des protocoles disponibles sur l'entité, leurs versions, et de les créer.

Pour créer un protocole, cliquez sur "New protocol", puis ajoutez et ordonnez les tâches qui correspondent aux SOP de votre laboratoire. Pour modifier un protocole existant, cliquez sur l'icône crayon à côté de celui-ci et enregistrez la nouvelle version du protocole.

Sélectionnez un protocole lors du lancement de l'analyse, ou attachez-le plus tard depuis l'onglet Workflow dans le tiroir d'analyse. Marquez les tâches comme terminées directement depuis le tiroir d'analyse. Chaque tâche terminée enregistre automatiquement l'utilisateur et la date pour une traçabilité complète.

Reporting - Global settings

La section Global settings de l'onglet Reporting permet aux utilisateurs d'activer ou de désactiver le verrouillage automatique de l'analyse lorsque le statut passe à "Reporting Validating", et de personnaliser les sections de classification et de reporting. Pour ce faire, sous Germline / Somatic, cliquez sur "New section", saisissez un nom de section et sélectionnez un tag à associer à la classification souhaitée. Les variants qui correspondent à l'un des tags ou classifications sélectionnés sont automatiquement ajoutés à cette section. Vous devez d'abord créer des tags dans l'onglet correspondant pour utiliser cette fonctionnalité.

Reporting - Template list

L'onglet Reporting sous "Template list" permet aux utilisateurs de configurer les modèles utilisés lors de l'exportation des résultats d'analyse. À partir de cet onglet, vous pouvez personnaliser les informations incluses dans les rapports générés.

Tags

L'onglet "Tags" permet aux utilisateurs de gérer des étiquettes personnalisées pour organiser et filtrer les variants au sein de la VKB et du rapport. Les utilisateurs peuvent créer des tags en cliquant sur "Add tag", ou renommer et supprimer les tags qui sont disponibles pour tous les utilisateurs au sein de l'entité en utilisant les icônes de stylo et de corbeille correspondantes.

API Keys

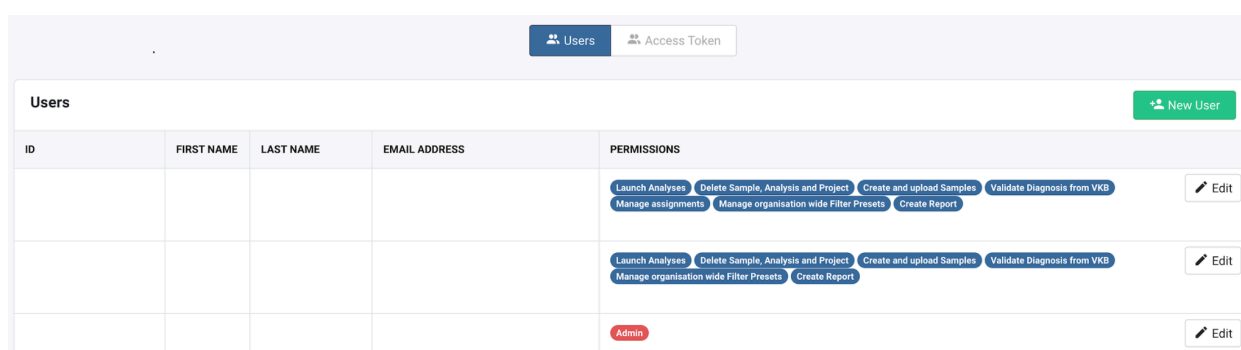
Une clé API est un jeton d'authentification unique permettant à des applications ou scripts externes d'interagir avec la plateforme SeqOne de manière programmatique. Depuis l'onglet API Keys, vous pouvez visualiser toutes les clés existantes ainsi que les informations qui leur sont associées.

Créez une nouvelle clé en cliquant sur "+ New API Key". Les clés expirées peuvent être affichées ou masquées à l'aide du bouton bascule "Show expired key". Chaque clé peut être supprimée individuellement. Les clés API doivent rester confidentielles, car elles donnent accès à la plateforme avec les permissions qui leur sont associées.

7.4.3. Gestion administrative - présentation de l'interface

Cette interface est dédiée aux administrateurs de comptes et leur permet de :

- accéder à la liste des utilisateurs de l'entité,
- créer de nouveaux utilisateurs,
- modifier les droits des utilisateurs : créer des projets et télécharger des échantillons, lancer des analyses, supprimer des échantillons, des analyses ou des projets, évaluer des variants dans la VKB, créer des rapports, filtrer la gestion des profils, gérer les affectations.
- activer la double authentification,
- désactiver un compte d'utilisateur.



ID	FIRST NAME	LAST NAME	EMAIL ADDRESS	PERMISSIONS
				Launch Analyses Delete Sample, Analysis and Project Create and upload Samples Validate Diagnosis from VKB Manage assignments Manage organisation wide Filter Presets Create Report Edit
				Launch Analyses Delete Sample, Analysis and Project Create and upload Samples Validate Diagnosis from VKB Manage organisation wide Filter Presets Create Report Edit
				Admin Edit

Figure 3 : Panneau de l'administrateur

L'ID correspond au nom d'utilisateur de l'utilisateur. Il doit être en minuscules et ne pas contenir d'accents ni d'espaces ; les caractères spéciaux tels que le point et le trait de soulignement sont acceptés. Il est conseillé de suivre la nomenclature prénom.nom lors de la création d'un nouvel utilisateur.

En ce qui concerne l'adresse électronique de l'utilisateur, nous déconseillons vivement l'utilisation d'une adresse électronique générique pour des raisons de sécurité.

Les permissions étant désactivées par défaut, il appartient à l'administrateur de les activer manuellement :

Permissions

Admin privileges

Grant full access to all system features, users management and settings



Create and upload Samples



Create Report



Delete Sample, Analysis and Project



Edit restricted panels



Launch Analyses



Manage assignments



Manage organisation wide Filter Presets



Manage VKB Knowledge Base



Protocols Management



Report template



Unlock Analyses



Update project settings (lock filters for now)



Validate Diagnosis from VKB



Advanced

Activate 2FA

Enable email two-factor authentication for enhanced security



Temporarily suspending the access. The data is retained, but the user cannot log in or use the system until re-enabled.

 Disable User

Figure 4 : Autorisations de l'utilisateur

Vous pouvez également renforcer la sécurité en activant la double authentification. Cette option envoie un code unique à l'utilisateur à chaque fois qu'il se connecte à la plateforme.

Le rôle "*Manage organization wide Filter Presets*" permet à l'utilisateur de modifier les profils de filtre publics. Si vous n'avez pas ce rôle, vous ne pouvez modifier que les profils de filtre privés.

Les permissions sont gérées par un administrateur d'entité depuis le panneau Admin. Lorsqu'un utilisateur est créé, toutes les permissions sont désactivées par défaut, et l'administrateur doit activer manuellement les droits requis pour cet utilisateur. Les droits attribués à un utilisateur s'appliquent à toutes les entités (espaces de travail) auxquelles l'utilisateur a accès sur la plateforme SeqOne (par exemple, si l'utilisateur peut accéder à plusieurs entités, le même ensemble de permissions s'applique à l'ensemble de celles-ci).

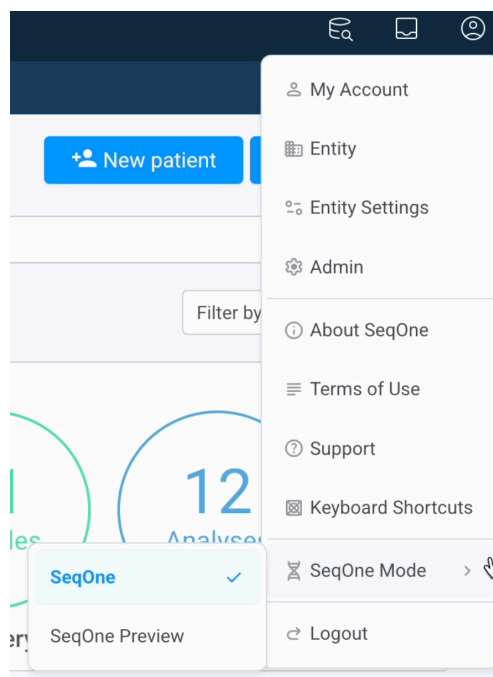
Si vous recevez un message d'erreur, veuillez contacter le service support à l'adresse support@seqone.com.

7.4.4. Basculer entre les modes Main et Preview

Se connecter à SeqOne via l'URL de connexion par défaut amène les utilisateurs au mode "Main" de leur entité. Ce mode par défaut héberge les jeux de travail marqués CE et d'autres certifications locales.

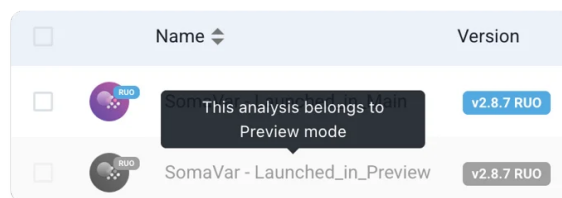
Pour un accès anticipé aux nouvelles fonctionnalités et aux versions de jeux de travail, passez au mode "Preview" (RUO: Research Use Only) en allant en bas du menu des paramètres et, sous SeqOne One, en sélectionnant "Preview".

Les modes "Main" et "Preview" partagent les mêmes données de base (projets, échantillons, analyses), autorisations, ressources (fichiers, modèles, manifestes, listes de gènes), connaissances (VKB) et fréquences (données de fréquence d'échantillons, occurrences).



Lors de la navigation en mode "Preview", le logo "SeqOne Preview RUO" apparaît pour rappeler que ce mode contient des fonctionnalités et des jeux de travail RUO.

En mode "Main", les analyses lancées en "Preview" apparaissent grisées. Si vous cliquez sur l'une de ces analyses grisées, la plateforme vous invite à passer en mode Preview pour la visualiser.



8. Description des étapes opérationnelles

Étape	Description
1	Création du projet Les analyses réalisées sur la plateforme SeqOne sont organisées au sein de projets, qui définissent les caractéristiques des échantillons analysés par le biais de métadonnées saisies par l'utilisateur (type de séquençage, méthode de préparation, gènes d'intérêt).
2	Exemple de téléchargement Les échantillons sont téléchargés sous forme de fichiers de séquençage bruts (FASTQ), qui doivent être placés dans le projet désigné. Une fois les fichiers téléchargés, une étape de vérification assure l'intégrité des fichiers soumis avant toute analyse.
3	Lancement de l'analyse Une fois les données téléchargées sur les serveurs de SeqOne, l'utilisateur peut choisir parmi une sélection d'applications ou "worksets", en fonction du type d'échantillon analysé et des types d'altérations génomiques recherchées (SNV, CNV, fusions de gènes, variants d'épissage, etc.) Ces applications effectuent des analyses bioinformatiques.
4	Contrôle de la qualité des données Les applications disponibles sur la plateforme SeqOne comprennent plusieurs modules dédiés à l'analyse de la qualité des données, en utilisant diverses mesures de séquençage, d'alignement et de profondeur de couverture.
5	Interprétation des résultats de l'analyse Les variations génomiques détectées dans l'échantillon sont affichées dans différents onglets ou "visionneuses", où l'utilisateur peut rechercher le variant diagnostique responsable de la pathologie du patient.
6	Obtenir des recommandations thérapeutiques et des essais cliniques associés à la pathologie du patient.
7	Délivrance d'un rapport Les résultats de l'analyse peuvent être exportés sous la forme d'un rapport PDF, qui peut ensuite être téléchargé.

9. Création du projet

Les analyses réalisées sur la plateforme SeqOne sont organisées au sein de projets, qui définissent les caractéristiques des échantillons analysés par le biais de métadonnées saisies par l'utilisateur (type de séquençage, méthode de préparation, gènes d'intérêt).

9.1. Créer un projet

Après avoir cliqué sur "New project", les utilisateurs sont invités à saisir les valeurs de plusieurs types d'informations, ou métadonnées, qui sont utilisées pour décrire le projet et les échantillons associés.

- **Description du projet**
- **Séquençage :**
 - Assay type : type d'analyse (lignée germinale ou somatique),
 - Les projets Germline (Assay type *Germline*) comportent une case à cocher UMI Trimming, qui permet de traiter les échantillons contenant des UMI. En sélectionnant cette option, les UMI seront éliminées des séquences FASTQ, comme dans le mode UMI désactivé pour les worksets somatiques (voir la section **Choix du mode de traitement des UMI**).
 - Sequencing platform: la plateforme de séquençage
 - Library type: type de séquençage de read effectué (*Paired-end* ou *Single-end*),
 - Data type: DNA, RNA, ctDNA ou Methylation Sequencing
 - Si l'ADN est sélectionné comme type de données, la méthode de séquençage doit être spécifiée à l'aide du menu déroulant "*DNA Sequencing method*", ainsi que la méthode d'amplification (Capture ou Amplicon) via "*Amplification method*".
 - File input type: *Fichiers Fastq*, *VCF*
 - Genome: la version du génome de référence (GRCh37 ou GRCh38),
 - Low-Quality Mapping Region: l'appel de variants dans les régions d'alignement de faible qualité en tenant compte des reads MAPQ0,
 - Sequencing method: Gene Panel, Whole exome sequencing, Comprehensive Genomic Profiling, Shallow WGS, Whole genome sequencing, CGH MicroArrays.
 - Secondary analysis provider: AmoyDX, Pillar, Thermofisher,
 - Manifest: le manifeste dans le cas d'un panel de gènes,
 - ID marker (Marqueurs d'identité (facultatif)) : le fichier de marqueurs d'identité.

- **Échantillons :**

- (Facultatif) De nouvelles métadonnées d'échantillon peuvent être ajoutées en cliquant sur "*You can also create your own metadata*". Lorsque vous y êtes invité, saisissez le nom et la valeur de la métadonnée.

Une fois les différentes métadonnées saisies pour le projet, les utilisateurs sont invités à créer leur projet en cliquant sur le bouton "*Create project*".

9.2. Gestion de projets

Via le tableau de bord, les utilisateurs peuvent gérer les différents projets qui ont été créés sur la plateforme. Ils peuvent notamment

- Accédez aux projets via l'onglet *Projets*, ainsi qu'aux échantillons et analyses de chaque projet en cliquant respectivement sur les logos "*Samples*" et "*Analysis*".
- Accédez directement aux analyses via l'onglet *Analyse*.
- Créez un nouveau projet en cliquant sur "*New project*" (voir la section **Créer un projet**).

9.3. Suppression d'un projet

Pour supprimer un projet, les utilisateurs doivent

- Sélectionner le projet
- Cliquez sur l'onglet "*Settings*" et ensuite sur "*Delete project*"
- Confirmez en cochant "Yes" dans le message affiché. Cette action est irréversible.

9.4. Données ADN (DNA Sequencing, ctDNA Sequencing)

Low-Quality Mapping Region. Lors du séquençage de panels de gènes, d'exomes ou de génomes, l'utilisateur peut inclure des régions d'alignement de faible qualité lors de l'appel de variants. Ces régions correspondent aux reads dits MAPQ0, dont la qualité d'alignement est égale à 0, telles que les pseudogènes et les régions homologues du génome.

L'option est applicable à des gènes ou régions cibles spécifiques. Pour utiliser cette option, l'utilisateur est invité à charger le manifeste ciblant les régions souhaitées, ou à le sélectionner via un menu déroulant.

Cette option peut réduire la qualité et la précision des résultats. Elle peut introduire un bruit de fond dans la fréquence des allèles et avoir un impact négatif sur la hiérarchisation des variants. Elle peut également allonger le temps d'analyse en raison d'une charge de calcul plus élevée.

Exome. Dans le cas du séquençage de l'exome, les utilisateurs n'ont pas besoin de spécifier le manifeste à utiliser ; ils sont invités à sélectionner le fabricant du kit utilisé en cliquant sur l'une des icônes de la catégorie Exome kit provider. Un menu déroulant permet alors à l'utilisateur de

sélectionner le manifeste souhaité. Si le manifeste souhaité n'est pas disponible sur la plateforme, l'utilisateur est invité à contacter le support utilisateur (support@seqone.com).

Gene panel. Dans le cas d'un panel de gènes, les utilisateurs sont invités à sélectionner la *méthode d'amplification* ainsi que le manifeste (au format BED) correspondant aux gènes d'intérêt dans un menu déroulant. Si le manifeste requis n'est pas disponible sur la plateforme, il peut être téléchargé en sélectionnant "*Upload new manifest*" et en suivant les instructions à l'écran.

Shallow WGS, CGH MicroArrays, Whole genome sequencing. Dans le cas du séquençage du génome, l'utilisateur n'a pas besoin de spécifier le manifeste à utiliser.

Capture. En choisissant la méthode d'amplification *capture*, une icône UMI (*Unique Molecular Identifiers*) est disponible. Les identificateurs moléculaires uniques sont des indices ajoutés avant l'amplification pour aider à éliminer les doublons de PCR et à réduire le bruit de fond après l'analyse.

Amplicon. Une attention particulière doit être accordée aux amorces d'amplification lors de l'analyse des amplicons. Lors de la création d'un projet pour des données d'amplicons, il est conseillé aux utilisateurs de sélectionner un manifeste dans lequel les régions correspondant aux cibles n'incluent pas les séquences d'amorces. Si un tel BED n'est pas disponible, une option peut être sélectionnée pour *clipper* une longueur fixe à la fin des séquences.

Identity monitoring. Pour les méthodes de séquençage correspondant au panel de gènes, à l'exome, au profilage génomique (CGP) ou au génome (shallow WGS) à partir de fichiers FASTQ, l'utilisateur est invité à sélectionner le fichier des marqueurs de contrôle d'identité dans un menu déroulant. Si le fichier souhaité n'est pas disponible sur la plateforme, il peut être importé en sélectionnant "*Upload new markers*" et en suivant les instructions à l'écran (voir également le paragraphe **Marker** dans la section **Gestion des paramètres utilisateurs**).

9.5. Données sur l'ARN (RNA Sequencing)

Si l'ARN est sélectionné comme type de données, les utilisateurs doivent spécifier si leur protocole conserve les informations sur le brin d'origine via le menu déroulant *Stranded protocol*. Si le protocole est spécifique à un brin, l'ordre de synthèse des brins (*avant-arrière* ou *arrière-avant*) doit être spécifié à l'aide de l'option *Read orientation (orientation de read)*. Enfin, la méthode de séquençage doit être spécifiée (gene panel).

Read orientation. Deux options sont disponibles pour spécifier l'orientation de la séquence :

- **Forward-Reverse** : le premier read (R1) de chaque paire correspond à la séquence de l'ARN messenger transcrit (brin sens).
- **Reverse-Forward** : le premier read (R1) de chaque paire correspond à la séquence complémentaire de l'ARN messenger transcrit (brin antisens).

En général, l'orientation des reads est la suivante pour les kits mentionnés ci-dessous :

- Agilent: Reverse-Forward
- Illumina: Unstranded
- Roche Kapa Hypercap: Reverse-Forward
- Twist: Reverse-Forward

En cas de doute, veuillez contacter le support client (support@seqone.com).

Une fois le type de données spécifié et le projet correctement configuré, l'utilisateur est invité à valider en cliquant sur le bouton "Next step".

9.6. Paramètres du projet

Dans l'onglet "Settings" du projet, vous pouvez :

- modifier le nom et la description du projet dans les champs "Project name" et "Project description",
- modifier le profil de contrôle de qualité (QC),
- modifier le profil de filtre par défaut dans le tableau des variants,
- supprimer définitivement le projet.

9.6.1. Configurer les métriques QC par défaut du projet

L'utilisateur peut choisir d'utiliser les mesures de contrôle de qualité par défaut ou de définir des valeurs seuils personnalisées à partir de l'onglet "Settings" du projet. L'utilisateur est invité à contacter le support utilisateur (support@seqone.com) pour définir les valeurs souhaitées.

Des mesures de contrôle de qualité personnalisées seront appliquées :

- QC checks dans la vue Analyses du projet,
- QC checks dans la section QC auto-validation de l'onglet QC de l'analyse,
- dans l'onglet GeneCov de l'analyse.

9.6.2. Définir le profil de filtre par défaut du projet

L'utilisateur peut choisir de définir le profil de filtre prédéfini appliqué à la table de variants par défaut à partir de l'onglet "Settings" du projet (voir également la section **Utiliser un profil de filtre prédéfini**).

Un profil de filtre peut être sélectionné pour chaque workset via un menu déroulant, à condition qu'il soit public. Le profil de filtre sera alors appliqué par défaut lors de l'affichage du tableau des variants, pour toutes les analyses de projet, pour tous les utilisateurs de l'organisation.

Si ce paramètre n'est pas modifié par l'utilisateur, lors de la création du projet, la valeur par défaut sera le profil défini par SeqOne.

Les profils de filtres par défaut peuvent être verrouillés en cliquant sur "Lock Variants Filters".

10. Chargement d'échantillons

10.1. Importer des échantillons

Les utilisateurs peuvent importer leurs échantillons via l'icône "Add samples" (ajouter des échantillons) dans chaque projet. Les échantillons peuvent être importés dans la plateforme SeqOne de deux manières :

Depuis un ordinateur

Les utilisateurs peuvent télécharger leurs fichiers FastQ directement depuis un dossier sur leur PC en cliquant sur l'onglet "From your computer" de l'icône "Add files". Les utilisateurs peuvent également glisser-déposer les fichiers de leur PC vers la zone dédiée de l'interface.

Via le SDS (SeqOne Data Sync)

SDS est un module premium optionnel qui permet de créer automatiquement des projets, de transférer des données, de lancer des analyses et de récupérer des données structurées (TSV) pour les intégrer dans le système de gestion de l'information du laboratoire (LIMS).

Ce module compatible avec Linux, MacOS et Windows est un exécutable qui permet aux utilisateurs de s'authentifier et d'interagir avec les API de SeqOne dans un environnement prédéfini et hautement sécurisé.

Exigences minimales :

- Linux 1.19 minimum
- MacOS 10.15 avec un processeur Intel minimum
- MacOS avec Apple Silicon CPU (puce M1) minimum
- Windows 10 64bits minimum
- 4 Go de RAM (8 Go pour Windows)
- Processeur quadricœur
- Accès à Internet et à la plateforme SeqOne (app.eu.seqone.com)
- Accès au système de stockage des données (directement ou via le réseau)

Si vous souhaitez utiliser cette solution ou obtenir de plus amples informations, contactez notre équipe Support (support@seqone.com).

Statut de l'importation des échantillons

La plateforme SeqOne utilise différents statuts pour permettre aux utilisateurs de suivre la progression de leurs téléchargements d'échantillons :

- *Upload pending* : importation en cours,
- *Verifying sample...* : conformité de l'échantillon en cours de vérification,
- *Mapping...* : étape d'alignement de l'échantillon,
- *Verified sample* : échantillon vérifié,
- *Upload failed* : échec de l'importation.

Capacité d'importation

La capacité de téléchargement des fichiers (en particulier pour les fichiers fastq.gz) dépend directement de la connexion internet de l'utilisateur. Par conséquent, la vitesse de la connexion internet doit être prise en compte lors du téléchargement des échantillons. Le tableau suivant donne une idée plus précise des temps de téléchargement estimés :

Estimation du temps d'attente				
	ADSL		Fibre optique (Mbits/s)	
	Ethernet : 10 Mbits/s	Wifi : 7,5 Mbits/s	Ethernet : 100 Mbits/s	Wifi : 90 Mbits/s
100Mo	1min 20s	1min 45s	8s	9s
Poids du fichier fastq				
1Go	13min 20s	17min 45s	1min 20s	1 min 30s
10Go	2h 13min 20s	2h 57mn	13min 20s	15 min
100Go	22h 22min	30h	2h 13min 20s	2h 30min

Figure 5 : Estimation du temps de téléchargement en fonction de la taille du fichier et de la vitesse de la connexion internet.

10.2. Format des fichiers d'entrée

10.2.1. Fichiers d'analyse primaire

Le format du fichier à télécharger pour chaque échantillon dépend de la plateforme de séquençage utilisée. Les noms de fichiers doivent être conformes à la convention de dénomination des fichiers de chaque séquenceur.

Les fichiers doivent être compressés à l'aide de l'extension .gz (gzip) avant de pouvoir être importés dans la plateforme. Dans le cas d'un séquençage en paires, deux fichiers sont attendus, correspondant aux reads des deux extrémités de chaque molécule.

Lorsqu'un échantillon a été correctement détecté par la plateforme au moment de l'importation, le nom complet du fichier est affiché. Il est alors possible de renommer le fichier en cliquant dessus. Les fichiers pair-end sont automatiquement combinés en un seul échantillon.

Les séquenceurs suivants sont pris en charge par la plateforme :

- [Illumina](#) : format fastq.gz, paired-end uniquement,
- [Element Biosciences](#) : format fastq.gz, paired-end uniquement,

- [MGI](#) : format fastq.gz, paired-end uniquement. Les noms de fichiers attendus sont : sample_1.fq.gz **and** sample_2.fq.gz **or** flowcellID_L01_sample_1.fq.gz **and** flowcellID_L01_sample_2.fq.gz

10.2.2. Fichiers d'analyse secondaire

Le format du fichier à télécharger pour chaque échantillon dépend de la plateforme de séquençage et de l'application utilisée pour l'analyse secondaire. Les fichiers doivent être compressés avec l'extension .gz (gzip) avant l'importation de la plateforme, sauf indication contraire.

Le format d'importation des fichiers d'analyse secondaire générés par les applications Illumina, Oxford Nanopore Technologies, Agilent, AmoyDx, Pillar, Thermofisher dépend du workset utilisé (voir la section **Analyses disponibles à partir des fichiers d'analyse secondaire**).

Lorsque la plateforme détecte correctement un échantillon lors de l'importation, elle affiche le nom complet du fichier. Vous pouvez renommer le fichier en cliquant dessus. Les fichiers qui partagent le même nom de base sont automatiquement combinés en un seul échantillon.

[Fichiers DRAGEN](#). Les fichiers et formats attendus pour chaque workset sont les suivants, lorsque les fichiers VCF - Secondary Analysis data et Illumina metadata sont sélectionnés lors de la création du projet :

GermVar Tertiary, GermVar Tertiary Family sample.hard-filtered.vcf.gz
sample.cnv.vcf.gz
sample.sv.vcf.gz
sample.repeats.vcf.gz

[Fichiers d'Oxford Nanopore Technologies](#) . Les fichiers et formats attendus sont les suivants lors de la sélection de VCF files - Secondary Analysis data, et des métadonnées d'Oxford Nanopore Technologies lors de la création du projet. Les utilisateurs peuvent importer un fichier de variants (SNP) seul ou le combiner avec des fichiers supplémentaires pour l'analyse CNV, STR ou SV :

GermVar Tertiary, GermVar Tertiary Family sample.wf_snp.vcf.gz
sample.wf_cnv.vcf.gz
sample.wf_str.vcf.gz
sample.wf_sv.vcf.gz

Thermofisher/Affymetrix files. Le format d'importation des fichiers d'analyse secondaire CGH est le format .txt généré par l'application Affymetrix de Thermofisher. Ce format n'est pris en charge que lorsque les métadonnées Thermofisher/Affymetrix et CGH MicroArrays sont sélectionnées lors de la création du projet. Les fichiers et formats attendus pour chaque workset sont les suivants :

GermVar Tertiary ARRAY_SampleID.txt

Fichiers CGH MicroArrays. Le format d'importation des fichiers d'analyse secondaire CGH est le format .xls généré par l'application Cytogenomics d'Agilent. Ce format n'est pris en charge que lorsque les métadonnées Agilent et CGH MicroArrays sont sélectionnées lors de la création du projet. Les fichiers et formats attendus pour chaque workset sont les suivants :

GermVar Tertiary IntervalBasedReport_sample.xls

Fichiers Alissa Reporter. Le format d'importation des fichiers d'analyse secondaire Alissa Reporter est le format .xls généré par l'application Alissa Reporter d'Agilent. Ce format n'est pris en charge que lorsque les métadonnées Agilent et CGH MicroArrays sont sélectionnées lors de la création du projet. Les fichiers et formats attendus pour chaque workset sont les suivants :

GermVar Tertiary IntervalBasedReport_sample.xls
ABERRATION_&_LOH_INTERVAL_sample.xls

PacBio-HiFi files. Lors de la création d'un projet, si les fichiers VCF (Secondary Analysis data) et les métadonnées PacBio-HiFi sont choisis, les fichiers et formats attendus pour chaque workset sont les suivants :

GermVar Tertiary, GermVar Tertiary Family sample.hard-filtered.vcf.gz
sample.cnv_sv.vcf.gz
sample.repeats.vcf.gz

Fichiers VCF pour panels de gènes somatiques. Le format d'importation des fichiers d'analyse secondaire SNV, CNV, Fusions est le format .vcf.gz généré par les applications Illumina, AmoyDx, Pillar, Thermofisher. Les noms de fichiers originaux doivent être convertis pour correspondre au format de dénomination requis décrit ci-dessous.

Les fichiers et formats attendus sont les suivants lors de la sélection des fichiers VCF files - Secondary Analysis data et métadonnées du panel de gènes lors de la création d'un projet somatique. Les utilisateurs peuvent importer un fichier de variants (SNV) seul ou le combiner avec des fichiers supplémentaires pour l'analyse des CNV ou des fusions :

SomaVar Tertiary sample.snv.vcf.gz
sample.cnv.vcf.gz

sample.fus.vcf.gz

Les fichiers et formats attendus sont les suivants lorsque vous sélectionnez des *VCF files* - *Secondary Analysis data* et *Comprehensive Genomic Profiling* dans les métadonnées lors de la création d'un projet somatique. Les utilisateurs peuvent importer un fichier VCF "multivariant" tel que le fichier VCF « All Variants » à partir de OncoPrint™ Comprehensive Assay V3/Plus.

SomaVar Tertiary OncoPrint_<LibPrepID>_<analysisID>.vcf

10.3. Métadonnées échantillons

Pour chaque échantillon téléchargé, les utilisateurs peuvent attribuer des valeurs aux métadonnées qu'ils ont sélectionnées lors de la création du projet ou définir de nouvelles valeurs en cliquant sur l'icône dans le coin supérieur droit de la liste des échantillons . 

Certaines métadonnées peuvent être téléchargées directement sous la forme d'un fichier de valeurs séparées par des tabulations en sélectionnant l'icône  .

Ces critères jouent un rôle important pour assurer la traçabilité des données et pour recouper les différentes analyses effectuées sur la plateforme. Certaines métadonnées, telles que le sexe, la maladie et les ID HPO, interviennent dans le processus d'identification des variants.

Lorsque le champ « Assay Type » est renseigné avec « Germline », les métadonnées associées à l'échantillon par défaut sont "HPO", "Sex" et "GnomAD group". Lorsque le champ « Assay Type » est renseigné avec « Somatic », les métadonnées associées à l'échantillon par défaut sont "Disease" et "Sex".

Pour ajouter ou modifier des HPO dans un échantillon, l'utilisateur peut cliquer sur "+ Add" dans la colonne HPO après avoir sélectionné un échantillon. Cela ouvre une fenêtre contextuelle permettant d'ajouter des HPO un par un, en recherchant un phénotype ou un code HPO dans la barre de recherche. Il est également possible d'importer un fichier texte contenant des codes

HPO en cliquant sur l'icône  . Un résumé des termes saisis est alors disponible, avec la possibilité de supprimer des termes si nécessaire.

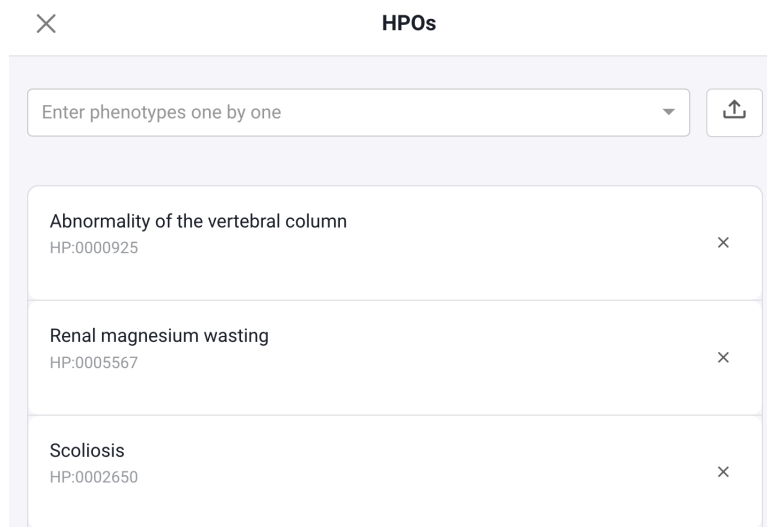


Figure 6 : Fenêtre pop-up d'importation HPO

10.4. Modifier un échantillon

Le nom, ainsi que les métadonnées associées à un échantillon peuvent être modifiés par l'utilisateur à tout moment, en sélectionnant l'échantillon dans la liste "Samples", puis en cliquant sur l'icône . Les modifications doivent ensuite être validées pour être effectives.

10.5. Supprimer un échantillon

Les utilisateurs peuvent supprimer les fichiers FastQ d'un échantillon de la plateforme en cliquant sur l'icône  après avoir sélectionné l'échantillon requis. Ces fichiers libéreront de l'espace de stockage et les résultats seront toujours disponibles sur SeqOne.

Pour supprimer un échantillon, cliquez sur l'icône  dans l'onglet "Samples" après avoir sélectionné l'échantillon souhaité. L'échantillon ne peut être supprimé définitivement qu'après avoir supprimé toutes les analyses qui lui sont associées.

11. Analyse

11.1. Lancer une nouvelle analyse

Les utilisateurs peuvent lancer une nouvelle analyse en cliquant sur "New analyse" dans l'écran d'*aperçu* de chaque projet.

Le lancement de l'analyse se fait en trois étapes distinctes, décrites ci-dessous.

L'étape "1. Workset" vous permet de choisir le workset d'analyse approprié (voir la section **Choix du workset d'analyse**).

L'étape "2. Select samples" est divisée en deux parties :

- 2.1. Set options :
 - décrit le workset et le génome de référence sélectionnés
 - permet d'ajouter du *remplissage* par le biais de la fonctionnalité "Extra calling padding", le cas échéant (voir la section **Choix du workset d'analyse**)
 - Permet de choisir le mode de traitement des UMI (voir la section **Choix du mode de traitement des UMI**), le cas échéant.
 - permet de choisir la baseline CNV, la baseline MSI ou d'ajuster le seuil VAF minimum, le cas échéant.
 - Permet de choisir le protocole.
- 2.2. Select sample(s) : permet de sélectionner les échantillons à analyser. Les utilisateurs peuvent sélectionner spécifiquement les nouveaux échantillons, qui sont marqués comme "New", en utilisant la fonction "select all new" (sélectionner tous les nouveaux).

L'étape "3. Confirm Analysis" résume les échantillons sélectionnés. Pour chaque analyse à lancer, il est possible d'appliquer un panel restreint en sélectionnant une ou plusieurs panels de gènes via l'option "Restricted panel" (voir aussi la section **Vue modale Panel**).



The screenshot shows a user interface for selecting a restricted panel. On the left, there is a box labeled 'GermlineVar - Patient'. To its right, the text 'Sample PATIENT' is displayed. Further right, the label 'Restricted panel' is followed by a dropdown menu and a trash icon. The dropdown menu is currently empty.

L'option "Restricted panel" applique un filtre à l'analyse sous la forme d'une ou plusieurs panels de gènes. Cette option n'affecte pas l'appel de variants, qui est déterminé par le manifeste du projet et le *remplissage* ajouté. Le choix des "Restricted panels" est réversible ; l'utilisateur peut modifier ces panels à partir de l'onglet *Panel* du panneau latéral de gestion de l'analyse après l'analyse (voir la section **Vue modale Panel**).

11.1.1. Choix du workset d'analyse

Plusieurs worksets d'analyse sont disponibles pour répondre à différentes questions biologiques. Les worksets proposés lors de la création d'une nouvelle analyse dépendent des métadonnées définies lors de la création du projet et peuvent donc varier en fonction des données.

Analyses disponibles à partir de fichiers FASTQ

Les worksets suivants sont disponibles lorsque les métadonnées *Fastq files - Primary Analysis data* sont sélectionnées lors de la création du projet.

À l'exception des projets où la méthode d'amplification est l'amplicon, certains worksets (c'est-à-dire GermlineVar, GermlineFamily, SomaVar, SomaVar LF (*versions* ≥ 2.0), SomaCGP (*version* ≥ 0.6), GermVar, GermVar Family, SomaLBx, SomaMethyl) vous permettent d'étendre la zone de recherche pour l'appel de variants et d'ajouter entre 0 et 100 pb de padding au manifeste initial.

Extra calling padding (between 0 and 100bp)  Apply extra padding on the manifest for calling, this won't affect the computation of QC metrics.

L'option est disponible à l'étape 2 (Select Sample) du lancement de l'analyse via la fonction "Extra calling padding". Par défaut, la valeur est fixée à 0 et doit être un entier naturel.

Certains Worksets (à savoir SomaVar (*versions* ≥ 2.6) et SomaHemato (*version* ≥ 0.2)) offrent la possibilité d'ajuster le seuil VAF minimum pour la détection des variants. Ce seuil peut être défini entre 0,01 % et 1 %, la valeur par défaut étant 1 %.

Minimal VAF threshold (between 0.01% and 1%)  Apply a custom minimal VAF threshold for short variant detection. Sensitivity and precision cannot be guaranteed when deviating from the default pipeline value (1% for SomaVar)

Pour plus d'informations sur cette option, contactez le Support utilisateur à l'adresse support@seqone.com.

GermlineVar



GermlineVar est utilisé pour détecter les mutations germinales sur la base du séquençage d'un panel de gènes ou de l'exome d'un patient.

Lors de l'exécution d'une analyse GermlineVar :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select",
- Sélectionnez les échantillons à analyser, puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

GermVar



GermVar permet la détection de mutations germinales, sur la base du séquençage d'un panel de gènes, de l'exome ou du génome d'un patient.

Lors de l'exécution d'une analyse GermVar :

- Sélectionnez la version du workset qu'ils souhaitent exécuter, puis cliquez sur "Select",
- Sélectionnez le modèle à utiliser,
- Sélectionnez le(s) échantillon(s) à analyser puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

GermlineVarLR



GermlineVarLR est utilisé pour détecter les mutations germinales dans les données de séquençage long-read provenant du séquençage de panels de gènes.

Lors de l'exécution d'une analyse GermlineVar LR :

- Sélectionnez la version du workset qu'ils souhaitent exécuter, puis cliquez sur "Select",
- Sélectionnez le modèle à utiliser et le(s) échantillon(s) à analyser, puis cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

GermlineFamily



GermlineFamily est utilisé pour détecter des mutations héréditaires sur la base du séquençage d'un panel de gènes, de l'exome ou du génome d'un patient et des membres de sa famille. Ce workset peut également être utilisé dans le cadre d'un test de préconception (*Carrier Screening*).

Lors de l'exécution d'une analyse GermlineFamily :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez au moins 2 échantillons
- Dans la section "Set conditions", indiquez si l'échantillon est celui du patient ou d'un membre de sa famille en définissant le lien familial sous "Family link". Le patient est automatiquement marqué comme "affecté". Dans le cadre d'un test de préconception (*Carrier Screening*), indiquez pour chaque échantillon le lien de parenté par "Mother" ou "Father" sans cocher la case "Is affected?"
- Vérifiez que les paramètres correspondent, puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

GermVar Family



GermVar Family permet la détection de mutations germinales, à partir du séquençage d'un panel de gènes, de l'exome ou du génome d'un patient.

Lors de l'exécution d'une analyse GermVar Family :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez le modèle à utiliser,
- Sélectionnez au moins 2 échantillons,
- Dans la section "Set conditions", indiquez si l'échantillon est celui du patient ou d'un membre de sa famille en définissant le lien familial sous "Family link". Le patient est automatiquement marqué comme "affecté". Dans le cas d'un dépistage de porteurs, indiquez pour chaque échantillon le lien de parenté par "Mother" ou "Father" sans cocher la case "Is affected?"
- Vérifiez que les paramètres correspondent, puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

CNVCapture



CNVCapture est utilisé pour détecter les variations du nombre de copies (CNV) germinales à partir des données de capture des panels de gènes.

Lors de l'exécution d'une analyse CNVCapture :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez un nombre minimum de 8 échantillons à analyser, puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

CNVExome

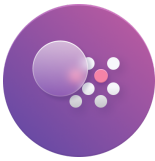


CNVExome est utilisé pour détecter les variations du nombre de copies (CNV) à partir des données de capture d'exome.

Lors de l'exécution d'une analyse CNVExome :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Sélectionnez le modèle à utiliser et le(s) échantillon(s) à analyser, puis cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaVar



SomaVar détecte les mutations somatiques à partir d'échantillons de tumeurs. Ce workset peut être utilisé avec des données de capture et d'amplicon.

Lors de l'exécution d'une analyse SomaVar :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaCNVCapture



SomaCNVCapture est utilisé pour détecter les variations du nombre de copies somatiques (CNV) à partir des données de capture des panels de gènes.

Lors de l'exécution d'une analyse SomaCNVCapture :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez un nombre minimum de huit échantillons à analyser, puis cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaRNA



SomaRNA peut être utilisé pour détecter les fusions de gènes à partir des données de capture RNA-seq.

Lors de l'exécution d'une analyse SomaRNA, :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser, puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

SomaMSI



SomaMSI détecte l'état des régions microsatellites (MSS ou MSI).

Lors de l'exécution d'une analyse SomaMSI :

- Sélectionnez un nombre minimum de huit échantillons à analyser, puis cliquez sur "Next step".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- sélectionner un nombre minimum de 8 échantillons à analyser puis cliquez sur "Next step",
- Lancez l'analyse en cliquant sur "Start analysis".

SomaHRD

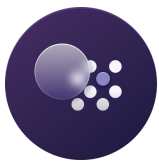


SomaHRD permet de détecter le statut HRD à partir de données WGS peu profondes, seules ou accompagnées d'un panel de gènes comprenant BRCA1 et BRCA2.

Lors de l'exécution d'une analyse SomaHRD :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Saisir les métadonnées de cellularité (% de cellules tumorales) pour chaque échantillon (voir la section **Métadonnées de l'échantillon**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step",
- Sélectionnez l'analyse SomaVar ou SomaCGP (GRCh37 ou GRCh38) appropriée pour chaque échantillon à l'aide de l'option "Select panel analysis (Optional)",
- Lancez l'analyse en cliquant sur "Start analysis".

SomaVar LF



SomaVar LF permet la détection de petits variants et de variants structuraux à faible fréquence, ainsi que de fusions de gènes, à partir d'échantillons de tumeurs (biopsies solides ou liquides, ADN tumoral circulant) enrichis par capture sur un panel de gènes.

Lors d'une analyse SomaVar LF :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaCGP



SomaCGP permet de détecter des mutations somatiques, des variants structuraux, des fusions de gènes, ainsi que les statuts MSI et TMB à partir d'échantillons tumoraux. Ce workset est dédié à l'analyse de l'ADN FFPE préparé avec le panel SureSelect Cancer CGP Assay d'Agilent, et n'est disponible que lorsque la métadonnée *Comprehensive Genomic Profiling* est sélectionnée lors de la création du projet. Ce workset peut être utilisé avec des données de capture.

Lors de l'exécution d'une analyse SomaCGP :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Si nécessaire, sélectionnez une « CNV baseline » et/ou une « MSI baseline » (si aucune base de référence n'est sélectionnée, l'analyse s'effectuera sans détection CNV/MSI).
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaHemato



SomaHemato permet la détection de mutations somatiques à partir d'échantillons présentant des spécificités hématologiques. Ce workset peut être utilisé avec des données de capture.

Lors de l'exécution d'une analyse SomaHemato :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaLBx



SomaLBx permet la détection de mutations somatiques de faible fréquence, de variants structuraux et de fusions de gènes, à partir de ctDNA. Ce workset est dédié à l'analyse de ctDNA et peut être utilisé avec des données de capture.

Lors de l'exécution d'une analyse SomaLBx :

- Sélectionnez la version du workset à exécuter et cliquez sur "Select".
- Sélectionnez une « CNV baseline » (si aucune base de référence n'est sélectionnée, l'analyse s'exécutera sans détection CNV).
- Dans les options, sélectionnez le mode de traitement des UMI applicable parmi les modes éligibles pour l'analyse (voir la section **Choix du mode de traitement des UMI**).
- Sélectionnez le(s) échantillon(s) à analyser et cliquez sur "Next step".
- Lancez l'analyse en cliquant sur "Start analysis".

SomaMethyl



SomaMethyl permet la détection d'îlots CpG méthylés à partir d'échantillons d'ADN tumoral circulant. Ce workset est dédié à l'analyse de la méthylation dans le ctDNA et peut être utilisé avec des données de capture au bisulfite.

Lors de l'exécution d'une analyse SomaMethyl, les utilisateurs sont invités à :

- Sélectionner la version du workset à exécuter et cliquer sur "Select",
- Sélectionner une "Methyl baseline" correspondant à la cohorte de contrôle (si aucune cohorte n'est sélectionnée, l'analyse s'exécutera sans détection de seuil ni méthylation différentielle),
- Sélectionner le ou les échantillons à analyser, puis cliquer sur "Next step",
- Lancer l'analyse en cliquant sur "Start analysis".

Analyses disponibles à partir de fichiers d'analyse secondaire

Les worksets suivants sont disponibles lorsque les *données VCF files - Secondary Analysis* sont sélectionnées lors de la création du projet ou lorsque les métadonnées Agilent et CGH MicroArrays sont sélectionnées (voir la section **Fichiers d'analyse secondaire**). Dans ce cas, la plateforme reçoit les données de sortie de l'application du fournisseur sélectionné, procède à l'annotation et à l'injection des variants dans leurs interfaces de visualisation respectives.

GermVar Tertiary



Pipeline d'analyse d'un seul échantillon, de variants détectés par l'application Illumina Germline ou l'application EPI2ME d'Oxford Nanopore Technologies, à partir de données WGS et WES ou de variants détectés par l'application Cytogenomics d'Agilent à partir de données CGH MicroArray.

Lors de l'exécution d'une analyse GermVar Tertiary :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez le(s) échantillon(s) à analyser, puis cliquez sur "Next Step".
- Lancez l'analyse en cliquant sur "Start analysis".

GermVar Family Tertiary



Pipeline d'analyse multi-échantillons de variants détectés par l'application Illumina Germline ou l'application EPI2ME d'Oxford Nanopore Technologies, à partir de données WGS et WES.

Lors de l'exécution d'une analyse GermVar Family Tertiary :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez au moins deux échantillons,
- Dans la section "Set conditions", indiquez si l'échantillon est celui du patient ou d'un membre de sa famille en définissant le lien familial sous "Family link". Le patient est automatiquement marqué comme "affecté".
- Vérifiez que les paramètres correspondent.
- Lancez l'analyse en cliquant sur "Start analysis".

SomaVar Tertiary



SomaVar Tertiary est un pipeline d'analyse tertiaire basé sur les fichiers VCF qui traite à la fois les fichiers VCF d'Illumina WES et les fichiers VCF de panels de gènes provenant de divers fournisseurs de kits (voir la section **Format des fichiers d'entrée**).

Lors de l'exécution d'une analyse SomaVar Tertiary :

- Sélectionnez la version du workset à exécuter, puis cliquez sur "Select",
- Sélectionnez le(s) échantillon(s) à analyser, puis cliquez sur "Next Step".
- Lancez l'analyse en cliquant sur "Start analysis".

11.1.2. Choix du mode de traitement des UMI

Présentation des modes de traitement des UMI

En pratique, la plateforme propose une ou plusieurs méthodes de traitement des UMI lors de l'exécution d'une analyse :

1. **Standard mode** (recommandé) : les consensus sont générés à partir des doublons de PCR présentant la même UMI lorsque leur nombre est supérieur ou égal à 2. Les UMI représentés par une seule séquence (singletons) sont également retenus. La fréquence allélique minimale des variants rapportés et le nombre minimal de reads supportant la mutation (AO, Alternate Observation) diffèrent selon le workset utilisé :
 - Dans SomaVar, SomaHemato, SomaRNA, SomaCNVCapture, SomaCGP et SomaMSI, la fréquence allélique minimale est de 1% et l'AO est un minimum de 3 reads.
 - Dans SomaVar LF et SomaLBx, la fréquence allélique minimale est de 0,1 % et l'AO minimal est de 2 reads. Ce workset permet de détecter des variants dont la fréquence allélique est inférieure à 1 %. Il est recommandé lorsque la profondeur de séquençage est supérieure à 5000x et pour des applications telles que le séquençage de l'ADNct.
2. **UMI Disabled** (disponible uniquement dans SomaVar, SomaRNA, SomaHemato, SomaLBx, SomaCNVCapture, SomaCGP et SomaMSI) : Les UMI ne sont pas utilisés pour la déduplication et sont coupés de la fin des séquences avant l'analyse. Par conséquent, les données générées par ce pipeline d'analyse sont brutes (non dédupliquées). La fréquence allélique minimale des variants signalés est de 1 % et le nombre minimal de reads soutenant la mutation (AO) est de 3.

Les différences de réglage entre ces modes sont résumées dans le tableau suivant :

	UMI disabled	UMI standard
Qualité minimale médiane/moyenne de la base mutée (score Phred)*	30	30
Nombre minimum de reads par consensus	N/A	2
Qualité minimale de chaque base dans le consensus (score Phred)	N/A	30
Reads sans consensus filtré	N/A	non

Tableau 1 : Paramètres des différentes configurations de traitement des UMI

* Ces filtres ne s'appliquent qu'aux variants qui ne sont pas sélectionnés par le modèle SVM après l'appel de variants (pipelines SomaVar et SomaRNA).

Processus de traitement des UMI

La structure des UMI dépend du kit de chaque fournisseur et les kits suivants sont actuellement disponibles sur la plateforme SeqOne :

- QIAGEN QIAseq
- Agilent XTBS/Low input
- Agilent XTBS V2
- Illumina TruSight Oncology 500
- Twist UMI Adapter System
- Roche KAPA HyperCap
- QIAseq targeted DNA Pro

 Warning: La configuration UMI "QIAseq Targeted DNA Pro" ne doit être utilisée que pour SomaVar, avec une longueur de lecture d'au moins 150 bp. Pour plus d'informations, contactez le support.

Le traitement des UMI est basé sur l'utilisation de la suite d'outils fgbio, qui peut traiter les UMI et les reads marqués par des UMI. Notre pipeline d'analyse UMI comprend les étapes suivantes :

1. Si nécessaire, élimination des UMI en fin de séquences et génération d'un BAM non aligné à partir de la paire de fichiers FASTQ,
2. Alignement des séquences sans UMI avec BWA-MEM2 (ADN/ADNct) ou STAR (ARN)
3. Fusion des fichiers BAM contenant les UMI avec les séquences alignées à l'aide de ZipperBam,
4. Regroupement des reads s'ils ont la même position et partagent le même UMI,
5. Génération de la séquence de consensus pour chaque groupe d'UMIs,
6. Filtre sur chaque base en fonction de sa qualité : la confiance dans la séquence consensuelle augmente avec le nombre de reads par groupe,
7. Filtre sur les séquences de consensus : si plus de 20 % des bases de la séquence de consensus sont filtrées, la séquence de consensus n'est pas conservée,
8. Conversion de ces séquences consensus filtrées en FASTQ et réalignement final sur le génome de référence avec BWA-MEM2 (ADN/ADNct) ou STAR (ARN).

Compatibilité avec les worksets d'analyse

Les worksets SomaVar, SomaHemato, SomaCGP, SomaLBx, SomaMethyl, SomaVar LF, SomaCNVCapture, SomaMSI et SomaRNA sont compatibles avec les données d'UMI. Les options de traitement disponibles lors du lancement des différents worksets sont décrites dans le tableau ci-dessous :

	UMI disabled	UMI standard
SomaVar	Oui	Oui
SomaCNVCapture	Oui	Oui
SomaMSI	Oui	Oui
SomaRNA	Oui	Oui

SomaVar LF	Non	Oui
SomaCGP	Oui	Oui
SomaLBx	Oui	Oui
SomaHemato	Oui	Oui
SomaMethyl	No	Yes

Tableau 2 : Options de traitement UMI disponibles lors du lancement des worksets compatibles.

11.2. Statut de l'analyse

Après avoir lancé une analyse, la plateforme SeqOne utilise différents statuts pour permettre aux utilisateurs de suivre la progression de leur analyse :

- **Bioinformatics Pending:** la plateforme attend la fin d'une autre analyse
- **Bioinformatics Running :** l'analyse est en cours ; ce statut indique également l'étape de l'analyse en cours.
- **Bioinformatics Finished :** l'analyse est terminée
- **Analysis failed :** une erreur s'est produite au cours de l'analyse ; dans ce cas, les utilisateurs sont invités à contacter l'équipe Support.
- **Interpretation - new :** l'analyse est terminée et prête à être interprétée, mais aucun utilisateur ne l'a encore ouverte.
- **Analysis failed - error :** L'analyse a rencontré une erreur. Dans ce cas, l'utilisateur est invité à contacter le support utilisateur.
- **Analysis failed - canceled :** l'analyse n'a pas démarré parce que l'échantillon qui lui est rattaché a rencontré une erreur lors du chargement ou de la vérification.

11.3. Qualité de l'échantillon

L'onglet QC du projet est divisé en deux sous-onglets : 'General statistics' et 'ID check summary'.

General statistics. Cet onglet affiche un tableau représentant la qualité globale de l'alignement et de la couverture pour chaque échantillon du projet. L'utilisateur peut exporter le tableau sous forme de fichier TSV en cliquant sur . 

Le calcul est basé sur l'outil SeQC, un programme développé par SeqOne pour calculer les statistiques d'alignement et de couverture à partir de tous les reads alignés sur le génome.

L'utilisateur peut choisir d'utiliser les mesures de contrôle de qualité par défaut ou de définir des valeurs seuils personnalisées à partir de l'onglet "Settings" du projet (voir la section **Configurer les métriques QC par défaut du projet**).

Les mesures proposées sont les suivantes :

- nombre total de reads
- longueur moyenne des reads
- taille médiane des fragments d'ADN dans la bibliothèque (inserts)
- pourcentage de reads de qualité >30
- pourcentage de reads alignés sur le génome de référence
- pourcentage de reads sur la cible
- pourcentage de reads dupliqués
- transition/taux de transition
- Couverture moyenne
- Fraction des cibles couvertes à différents seuils en fonction du contexte (amplicon, capture).

Les cibles du panel ou de l'exome sont déterminées à partir du fichier BED sélectionné lors de la création du projet.

ID check summary. Cet onglet détaille les génotypes de tous les marqueurs inclus dans ses analyses pour chaque échantillon (voir le paragraphe **Marker** dans la section **Gestion des paramètres utilisateur**). Une ligne correspond à un échantillon, et chaque colonne représente un marqueur avec son ID associé. L'utilisateur peut exporter ce tableau sous forme de fichier TSV en cliquant sur .

12. Contrôle de la qualité des données

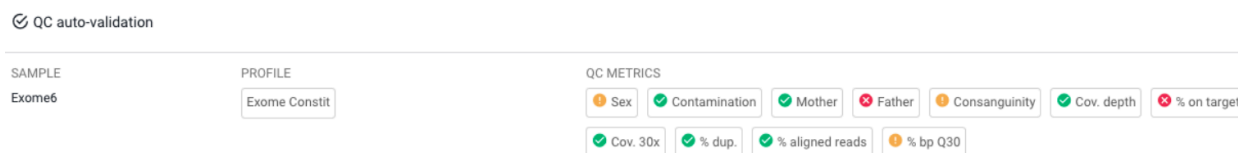
12.1. Mesures et rapports de contrôle de qualité

12.1.1. Mesures QC et contrôles QC

À partir de l'onglet QC, l'utilisateur accède aux contrôles de qualité appliqués à l'analyse, sous la forme de sous-onglets décrits dans les sections ci-dessous : *Rapport QC*, *Contrôle d'identité* (le cas échéant) et *Rapport technique*.

Dans le cas d'une analyse CGH MicroArray, l'onglet QC n'est pas disponible. Cependant, le rapport technique est disponible à partir de l'onglet *Files* de l'analyse au format *report.html*.

Dans la partie supérieure de l'onglet QC, une section "QC auto-validation" permet à l'utilisateur d'attribuer un code couleur aux mesures QC : *Passed*, *Warning* ou *Failed*. En cliquant sur l'icône associée, l'utilisateur accède aux seuils appliqués et aux détails des QC : sexe, contamination, parenté et consanguinité, profondeur de couverture, % on target, couverture 30X ou 100X, % duplicates, % reads alignés sur le génome de référence ou % bp Q30.



12.1.2. Rapport QC

Le "QC Report" global contient diverses mesures décrivant la qualité et l'alignement des séquences, la couverture des régions dans le manifeste et certaines statistiques relatives aux variants. Ce rapport est composé des sections suivantes :

[Alignment details \(Détails de l'alignement\)](#) (SomaMethyl). Nombre total de lectures traitées, lectures alignées de manière unique et non unique, et taux de mésappariement.

[Deduplication](#) (SomaMethyl). Métriques de duplication pour les lectures alignées, mettant en évidence les problèmes potentiels liés à la préparation de la bibliothèque.

[Cytosine coverage & methylation \(Couverture de la cytosine et méthylation\)](#) (SomaMethyl). Taux de conversion de la cytosine en uracile pour chaque contexte (CpG, CHG, CHH).

[Statistiques des groupes UMI](#) (SomaVar, SomaVar LF, SomaMethyl) et *SomaRNA UMI uniquement*). Voir le paragraphe **Choix du mode de traitement des UMI**.

Le rapport de contrôle de qualité (QC report) des analyses SomaVar, SomaVar LF, SomaMethyl et SomaRNA offre une vue plus détaillée de la composition de l'échantillon, en particulier la distribution des UMIs en fonction du nombre de séquences qui les hébergent.

Les statistiques décrivent en particulier

- La fraction de singletons (UMI portés par une séquence unique)
- La taille moyenne et médiane des groupes d'UMI, ainsi que l'écart-type
- La taille du groupe le plus important et la fraction des séquences totales qu'il comprend.

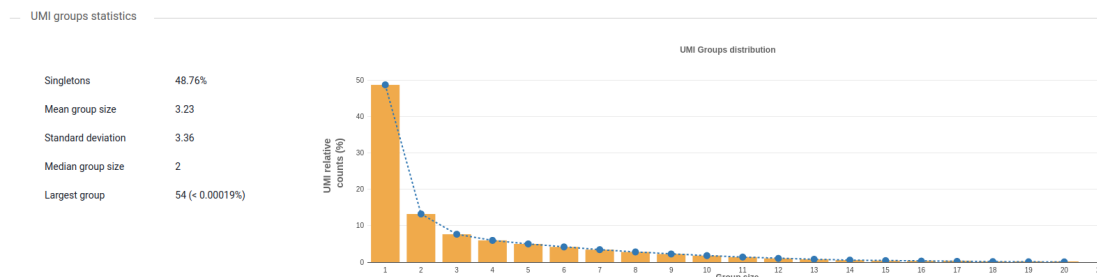


Figure 7 : Statistiques relatives à la distribution des UMI en termes de nombre de séquences dans un échantillon.

Les mesures globales de contrôle de qualité pour le projet sont également plus détaillées et divisées en trois sections :

- les paramètres de contrôle de qualité après le traitement UMI, qui indiquent la tendance de chaque paramètre (en pourcentage) sur la base des données brutes.
- Mesures brutes de contrôle de qualité calculées à partir de toutes les séquences (avant le traitement UMI).
- Distribution des UMI en termes de nombre de séquences dans chaque échantillon.

Cross-contamination estimate (SomaVar LF uniquement). La contamination croisée est estimée individuellement à l'aide de l'outil GATK CalculateContamination. Ce dernier utilise le signal des allèles de référence détectés aux loci homozygotes afin de calculer la fraction de séquences dans l'individu provenant de la contamination croisée.

Sample integrity (analyses de la lignée germinale uniquement). Cette section énumère les mesures relatives à l'intégrité et à l'identité de l'échantillon :

- **Lien familial** (Family link): métadonnées définies par l'utilisateur
- **Sexe d'origine** (Original sex): métadonnées relatives au sexe saisies par l'utilisateur (masculin/féminin/inconnu).
- **Sexe détecté** (Detected sex): sexe détecté (masculin/féminin/inconnu)

- **Coefficient de parenté** (*Relationship coefficient*): coefficient de parenté calculé à l'aide de l'outil Somalier pour chaque échantillon de l'analyse de famille, cette métrique affiche N/A pour l'échantillon lui-même.
- **Contamination** : estimation du pourcentage de contamination de l'échantillon par de l'ADN étranger, à l'aide de l'outil *verifyBamID2*.

Sont également indiqués le nombre de variants homozygotes et hétérozygotes, le rapport entre ces deux valeurs (rapport Hom/Het), le nombre de variants homozygotes et hétérozygotes sur le chromosome X, le rapport équivalent sur le chromosome X (rapport Het/Hom sur chr X), et le nombre de variants sur le chromosome Y.

Les métriques **Lien familial** (*Family link*) et **Coefficient de parenté** (*Relationship coefficient*) utilisées pour vérifier la relation entre les individus d'une famille, ne sont affichées que dans les analyses de famille en trio (GermlineVar Family ou GermVar Family Tertiary).

Les métriques de genre sont utilisées pour vérifier la correspondance entre le genre saisi et le genre détecté dans les données.

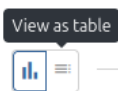
La métrique de contamination n'est pas calculée pour les worksets GermVar Tertiary et GermVar Tertiary Family.

Les contrôles de qualité (*QC checks*) sur la concordance du sexe, la contamination et la relation de parenté sont disponibles au niveau du projet ou de l'analyse afin d'identifier facilement toute concordance ou non-concordance.

Alignment metrics. Cette section énumère les principales mesures calculées à la fin de l'alignement, telles que :

- Le nombre total de séquences dans l'échantillon,
- La taille médiane des inserts, calculée à partir de la distance entre les deux reads de la même paire, dont la longueur respective est incluse,
- La longueur moyenne de la séquence,
- La fraction de séquences alignées avec succès sur le génome de référence,
- Le taux de duplication (capture uniquement),
- La fraction des bases dont la qualité est supérieure à 30,
- La couverture moyenne des régions ciblées, déterminée à partir du manifeste d'analyse.

Coverage. Diagramme représentant la fraction des régions cibles couvertes à différents seuils, indiqués sur l'axe des abscisses. Les valeurs de couvertures peuvent être affichées sous forme



de tableau en cliquant sur le bouton pour permettre à l'utilisateur de copier/coller ces valeurs.

QC graphs. Graphiques de l'analyse de qualité FastQC, illustrant notamment le contenu en GC des séquences ainsi que la qualité des bases en fonction de leur position dans les séquences.

Variant statistics. Une liste de statistiques sur les variants détectés dans l'analyse, qui comprend en particulier :

- Le nombre de transitions (Ts) et de transversions (Tv) parmi les substitutions détectées,
- Le rapport Ts / Tv,
- Le nombre total de mutations détectées,
- La fraction de SNV, MNV et indels dans ce total.

12.1.3. Rapport 'ID-Check'

L'onglet "ID check" est consacré aux résultats du contrôle d'identité dans l'onglet QC des worksets appelant des variants.

Les résultats sont affichés sous forme de tableau avec les informations suivantes :

- Description de chaque marqueur sur la base des coordonnées génomiques, par exemple : chr-start-ref-alt,
- une étiquette (rsID, étiquette personnalisée) lorsqu'elle est donnée par l'utilisateur,
- le génotype observé (GT) ou la fréquence de l'allèle (VAF),
- la profondeur de la couverture à l'emplacement du marqueur (COV),
- la qualité de la base de données (QUAL).

Lorsque plusieurs échantillons sont inclus dans l'analyse, les colonnes spécifiques à l'échantillon (GT/VAF, QUAL, COV) sont répétées pour chaque échantillon individuellement.

La sélection d'un variant dans le tableau ouvre un navigateur IGV dans un nouvel onglet, centré sur les coordonnées génomiques du variant.

12.1.4. Rapport technique

Le rapport technique (*Technical report*) liste les outils utilisés lors de l'analyse et leur version. Il est accessible depuis l'onglet QC d'une analyse, puis sous "Technical report".

Il décrit notamment :

- des informations générales sur la configuration et les métadonnées du projet et de l'analyse, telles que le type de séquençage, la version du génome et la version du workset utilisées ;

[QC report](#) [ID check](#) [Technical report](#)

Technical Report

This report describes the back-end process used for the secondary analysis.



General Infos

Project Name	VerificationTest - ID Marker Multiple samples
Experiment Type	dna-seq
Sequencing Type	paired-end
Sequencing Equipment	Illumina
Sequencing Method	panel
Amplification Method	capture
UMI Mode	standard
Genome Version	GRCh37
Analysis Name	SomaVar - toy_fusions48
Workset Name	SomaVar
Workset Description	Detection of somatic variants based only on the index sample
Workset Version	v2.5.11

- les détails du manifeste utilisé, tels que le nom, la taille, la longueur (pb) et le padding ajouté à l'analyse ;
- les détails des bases de données et des outils bioinformatiques utilisés.

12.1.5. Rapport UPD

Le Rapport UPD est disponible dans la section QC des analyses germinales compatibles (par ex., GermVar et les flux de travail associés). Il résume les résultats de détection UPD et fournit des éléments visuels pour faciliter l'interprétation.

Le Rapport UPD comprend généralement :

- Aperçu des résultats : indication de la détection ou non d'un événement de type UPD par l'analyse et du type d'événement (par ex., complet ou segmentaire ; dans les analyses en trio, le rapport peut également étayer l'origine parentale et les sous-types d'événements tels que UPD | Iso / UPD | Het).
- Liste des événements : une liste des événements détectés avec leur emplacement chromosomique et leur taille, ainsi que le ou les scores associés utilisés par l'algorithme.
- Graphiques de visualisation : graphiques basés sur les génotypes SNP (par ex., fréquence de l'allèle B (BAF) et signal d'homozygotie/ROH) qui aident à confirmer la présence de longues régions homozygotes et leurs limites.
- Informations de contrôle qualité : métriques liées à la fiabilité de la détection (par ex., densité SNP / nombre de SNPs) et avertissements lorsque les données ne sont pas suffisantes pour un appel confiant.

Les événements UPD détectés sont également affichés dans le Large Variant Viewer (LVV) avec des étiquettes spécifiques (par ex., UPD | Iso et UPD | Het).

Score UPD

Le score UPD est un indicateur numérique utilisé pour quantifier la force avec laquelle l'échantillon montre un signal de type UPD le long du génome. La détection UPD dans SeqOne repose sur les modèles de génotype SNP :

- Dans une analyse de singleton (solo), l'algorithme recherche de grandes régions d'homozygotie (ROH) et des changements brusques dans la distribution des SNPs le long des chromosomes.
- Dans une analyse en trio, le modèle SNP est en outre utilisé pour déterminer l'origine parentale de l'événement et pour distinguer l'isodisomie (UPD | Iso) de l'hétérodisomie (UPD | Het).

En pratique, le score UPD reflète la force et la cohérence du signal étayant un événement UPD (par ex., modèle ROH/BAF clair et limites bien délimitées). Un score plus élevé indique que l'événement a plus de chances d'être un véritable événement UPD, tandis que des valeurs plus faibles indiquent des signaux plus faibles ou plus bruyants.

Le score UPD doit toujours être interprété en conjonction avec les éléments fournis dans l'onglet Rapport UPD (section QC), notamment :

- les graphiques de visualisation UPD (par ex., fréquence de l'allèle B / signal d'homozygotie),
- la taille et l'emplacement chromosomique de l'événement,
- les métriques de qualité d'analyse (par ex., densité SNP / nombre de SNPs).

Graphique de fréquence de l'allèle B (BAF) à partir du rapport UPD

Le graphique de fréquence de l'allèle B (BAF) est une visualisation des génotypes SNP le long du génome et est utilisé pour évaluer le déséquilibre allélique et la perte d'hétérozygotie (LOH).

Pour chaque position SNP, la BAF correspond à la fraction de lectures de séquençage supportant l'allèle "B" et varie de 0 à 1 :

- Les valeurs proches de 0 correspondent à un génotype homozygote (AA).
- Les valeurs proches de 0,5 correspondent à un génotype hétérozygote (AB).
- Les valeurs proches de 1 correspondent à un génotype homozygote (BB).

Dans une région diploïde normale, le graphique BAF montre généralement trois bandes autour de 0, 0,5 et 1. Dans les longues régions d'homozygotie (ROH) ou l'isodisomie UPD, la bande hétérozygote (autour de 0,5) est réduite ou absente, ce qui corrobore la présence d'un signal de type UPD.

Limites importantes :

- Un signal de type UPD peut également être observé dans le contexte d'une consanguinité parentale (multiples régions ROH à travers le génome).
- L'appel UPD est principalement basé sur la distribution des SNPs et peut ne pas être suffisant à lui seul pour distinguer certains modèles UPD des événements liés au nombre de copies ; si nécessaire, vérifiez avec d'autres preuves disponibles (par ex., couverture / résultats CNV).

13. Interprétation des résultats de l'analyse

13.1. Résultats de l'analyse

Les utilisateurs peuvent accéder aux analyses d'un projet en cliquant sur l'icône "Analyses" dans la boîte associée ou en cliquant sur l'icône du même nom dans le coin supérieur gauche du tableau de bord de l'utilisateur.

Lorsqu'une analyse est terminée et qu'elle peut être consultée, elle est affichée avec le statut "*Interpretation - New*"

Les résultats de l'analyse sont ensuite disponibles sous la forme de plusieurs rapports, qui peuvent être ouverts directement en cliquant dessus.

Il existe plusieurs variants du statut "*Interprétation*", en fonction des progrès réalisés dans l'interprétation des résultats :

- **New** - le tableau des variants n'a pas encore été consulté
- **In progress** - le tableau des variants a été consulté, mais aucun variant n'a encore été validé.
- **Evaluating** - au moins un variant a été validé
- **Reporting validated** - un rapport a été généré pour l'analyse

13.1.1. Gestion du workflow de l'analyse et de l'interprétation

13.1.1.1. Attribution des analyses pour l'interprétation

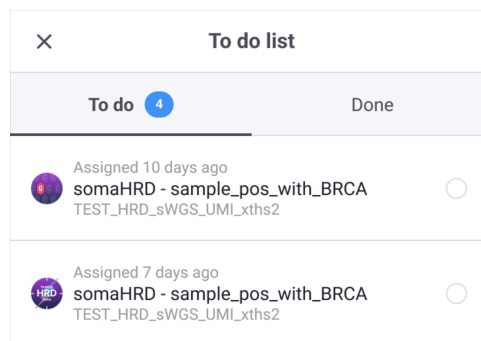
Les administrateurs d'entités peuvent choisir d'assigner une ou plusieurs analyses à interpréter par différents utilisateurs de l'entité à partir de l'onglet Analyses du projet. Pour ce faire, l'administrateur peut sélectionner une ou plusieurs analyses, puis cliquer sur le bouton

 Edit assignment

en bas de la page.

Une fenêtre "Assignment" permet d'affecter un ou plusieurs utilisateurs aux analyses sélectionnées, en recherchant l'identifiant de l'utilisateur (nom d'utilisateur) via la barre de recherche dédiée. Il est également possible de modifier l'affectation en ajoutant ou en supprimant des utilisateurs. Les utilisateurs auxquels une analyse est attribuée sont listés dans la colonne "Assigned to". Les utilisateurs qui ne se sont jamais connectés à SeqOne ne peuvent pas être assignés tant qu'ils n'ont pas créé leur identifiant.

Les utilisateurs peuvent accéder aux analyses qui leur ont été attribuées (onglet "To do") et aux analyses validées (onglet "Done") dans leur "To do list", accessible via l'icône  en haut à droite de la page. Une notification apparaît pour chaque nouvelle assignation.

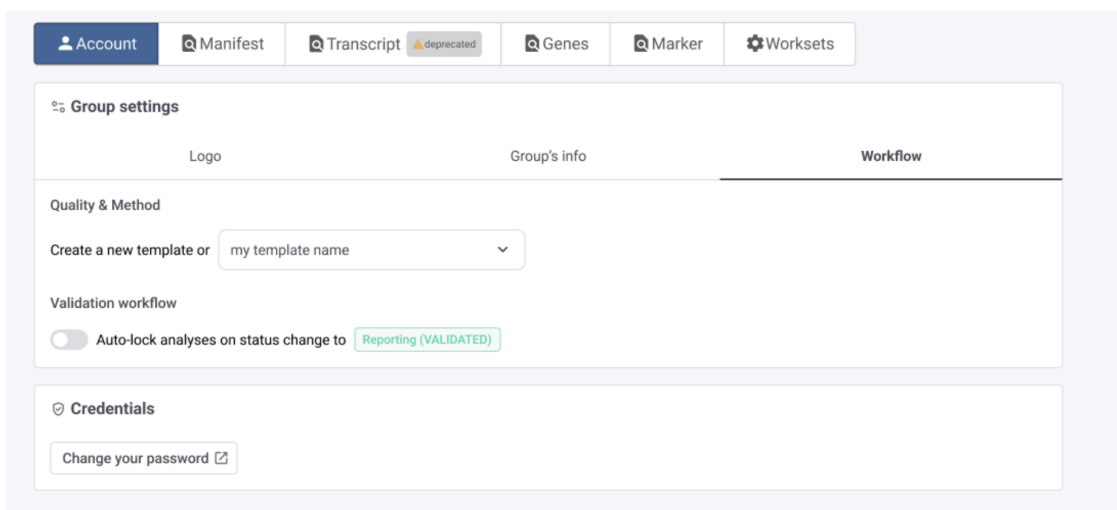


Les utilisateurs peuvent accéder directement aux analyses qui leur ont été attribuées en cliquant dessus. Ils peuvent valider leur tâche en cliquant sur la case à cocher correspondante dans l'onglet "To Do" pour la déplacer dans l'onglet "Done". Cette action est réversible.

Les analyses sont affichées dans l'ordre suivant en fonction de leur statut : terminé (*completed*), en cours (*running*) ou échoué (*failed*). Pour chaque analyse, la date d'affectation, le nom de l'analyse et le nom du projet sont détaillés.

13.1.1.2. Verrouillage d'analyse

L'administrateur d'une entité peut verrouiller ou déverrouiller une analyse. Dans les *Settings* de l'entité, sous l'onglet *Account* puis *Workflow*, il peut activer le verrouillage automatique des analyses validées en utilisant la fonctionnalité "Auto-lock analyses on status change to Reporting (VALIDATED)" :



À la fin de son travail d'interprétation, lorsqu'un utilisateur modifie le statut de l'analyse sous *Reporting - validated*, l'analyse est automatiquement verrouillée. Les analyses verrouillées affichent un statut "Reporting - Validated & Locked", visible avec l'icône  Reporting VALIDATED & LOCKED

Lorsqu'une analyse est verrouillée, les utilisateurs ne peuvent pas effectuer les actions suivantes à partir de l'interface :

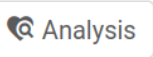
- ajouter ou modifier des évaluations de variants dans le VKB,
- ajouter ou modifier des commentaires associés aux variants,
- générer un nouveau rapport à partir de la fenêtre de *Selection*,
- générer un nouveau rapport de HRD,
- modifier manuellement le statut de l'analyse,
- effectuer l'une des actions suivantes dans le panneau d'analyse :
 - modifier l'affectation de l'essai (sous "Affectation")
 - modifier le sexe du patient
 - ajouter ou modifier des HPO
 - modifier les panels de l'analyse
 - supprimer le rapport clinique

Si un utilisateur tente l'une de ces actions, il sera invité à déverrouiller l'analyse, s'il en a l'autorisation, ou à contacter son administrateur s'il n'en a pas l'autorisation. Un utilisateur souhaitant déverrouiller une analyse doit disposer de l'autorisation "Unlock analyses".

13.1.1.3. Panneau latéral pour la gestion des analyses

Le panneau latéral d'analyse, également appelé "Analysis Drawer", est accessible à partir de n'importe quel onglet d'analyse.



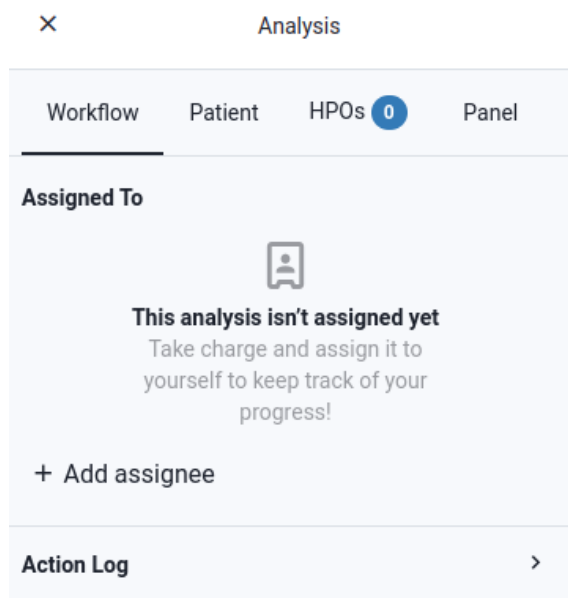
En cliquant sur le bouton  en haut à droite de la bannière des onglets d'analyse, l'utilisateur peut directement visualiser et gérer les données d'analyse et les HPO des patients sans quitter l'écran en cours.

Le volet d'analyse contient 4 vues :

- **Workflow** : vous permet d'organiser et de suivre votre Workflow à des fins d'analyse.
- **Patient** : permet de visualiser et de modifier les métadonnées du patient.
- **HPOs** : permet d'ajouter ou de modifier des termes HPO dans l'analyse, disponible uniquement pour les analyses dans un contexte constitutionnel.
- **Panel** : permet d'appliquer une liste de gènes à l'analyse et de filtrer les résultats de l'analyse sur la base de cette liste.

Vue modale Workflow

Cette vue vous permet d'organiser et de suivre les Workflows des utilisateurs pour l'analyse en cours d'interprétation.



Assigned To. Cette vue affiche les tâches assignées pour l'analyse en cours. Seuls les administrateurs ou les utilisateurs disposant de l'autorisation "Manage assignments" peuvent affecter des tâches (voir la section **Gestion administrative - Présentation de l'interface**).

Lorsque des tâches sont assignées à l'analyse en cours, une liste de tous les utilisateurs assignés s'affiche, avec la date et l'heure d'achèvement. Cette liste est mise à jour automatiquement dès qu'un utilisateur termine une tâche.

Les utilisateurs ne peuvent modifier que leurs propres tâches en cochant ou décochant la case correspondante. Lorsque l'utilisateur survole une tâche en lecture seule, une infobulle indique "You can only complete yours" (Vous pouvez seulement compléter les vôtres.).

Si aucune tâche n'a été affectée, la vue affiche "This analysis isn't assigned yet" (Cette analyse n'est pas encore assignée).

Si un utilisateur non autorisé tente d'assigner une tâche en cliquant sur le bouton "+ Add assignee", le message d'erreur indique "You must have administrator privileges to assign analysis. If needed, please contact your platform administrator to request the appropriate access" (Vous devez avoir des privilèges d'administrateur pour assigner une analyse. Si nécessaire, veuillez contacter l'administrateur de votre plateforme pour demander l'accès approprié).

Action log. La bannière "Analysis Action Log (Audit Trail)" affiche l'historique de toutes les actions effectuées sur le statut HRD dans l'analyse SomaHRD, les filtres, les profils de filtre et les panels de gènes restreints dans le tableau des variants, en détaillant le type d'action, l'utilisateur, l'horodatage et une description de l'action.

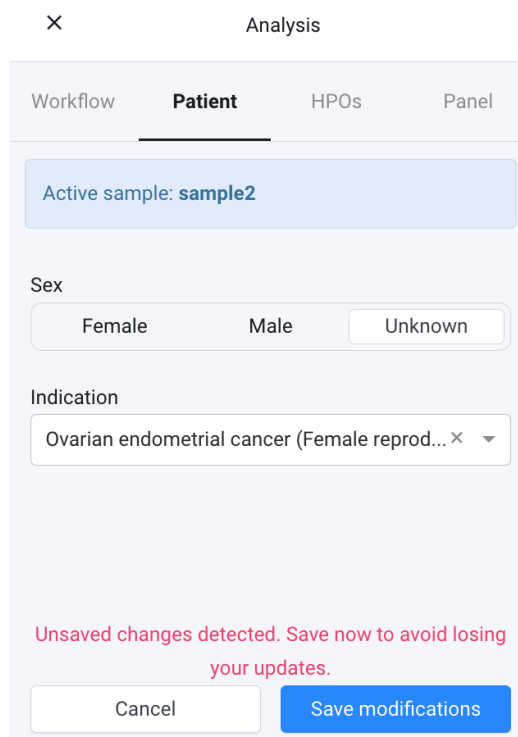
Les actions les plus récentes apparaissent en haut de la liste. La pagination permet aux utilisateurs de naviguer efficacement dans l'historique des actions. L'historique complet peut être téléchargé au format .csv via l'icône "Download Log".

Vue modale Patient

Cette vue vous permet de consulter et de modifier les métadonnées du patient telles que le sexe (Femme, Homme ou Inconnu), le groupe d'ascendance génétique GnomAD (Constitutionnel uniquement) ou l'Indication (Somatique uniquement).

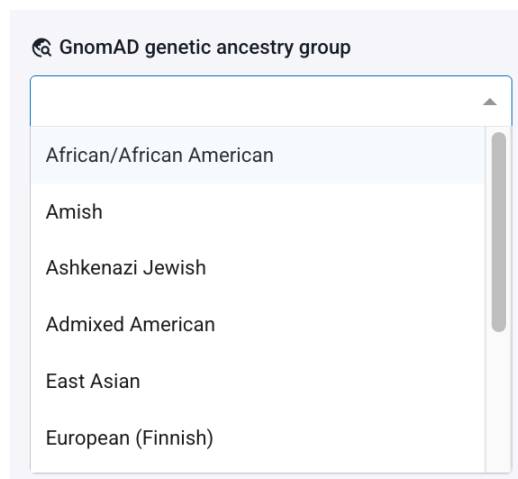
L'affichage et la modification de l'indication ne sont disponibles que pour les analyses somatiques :

- Si l'indication a été saisie avant le lancement de l'analyse (voir **Métadonnées de l'échantillon**), elle est définie par défaut pour l'échantillon, est alors marquée d'une étoile et apparaît en première position dans la liste.
- Si l'indication est modifiée au cours de l'interprétation, un message d'avertissement indique que l'indication sélectionnée est différente de l'indication initiale et la coloration de la pilule dans la sous-colonne "A" de la colonne "*Clinical Evidences*" est mise à jour (voir **Sélection d'une indication**).
- Si aucune indication n'est disponible pour l'analyse, le message "*No available indication*" (Aucune indication disponible) s'affiche.



L'affichage et la modification du « GnomAD genetic ancestry group » ne sont disponibles que pour les analyses **germline**. La valeur par défaut est vide et est traitée comme la population GnomAD combinée.

Les modifications effectuées dans cette vue doivent être sauvegardées ou annulées en cliquant sur les icônes "*Save modifications/Cancel*" correspondantes en bas du volet, afin de pouvoir changer d'onglet.



Vue modale HPO

Cette vue, disponible uniquement dans les projets Germline, vous permet de visualiser, d'éditer et d'attribuer des termes HPO relatifs à l'échantillon, soit manuellement, soit à partir d'une description clinique du patient.

L'utilisateur peut visualiser la liste des termes HPOs déjà associés à l'échantillon avant l'analyse. En l'absence de codes HPOs, le message "No HPOs yet" est affiché.

Pour ajouter des termes manuellement, l'utilisateur peut les saisir à partir de la barre de recherche "Enter phenotypes one by one", en recherchant un phénotype ou un code HPO. Il peut également importer un fichier texte contenant des codes HPO en cliquant sur l'icône  (voir également la section **Métadonnées d'échantillons**).

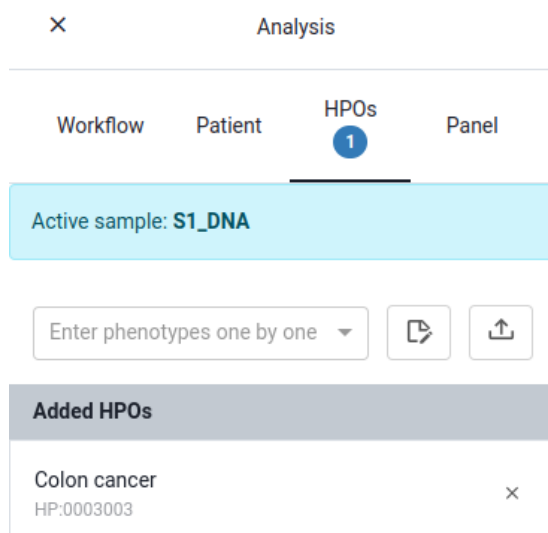


Figure 8 : Vue du panel d'analyse par les HPOs

L'utilisateur peut générer des termes HPO à partir de la description clinique du patient en utilisant DiagAI. En cliquant sur l'icône , une fenêtre de dialogue apparaît, invitant l'utilisateur à décrire le profil clinique du patient (cette description est limitée à 4000 caractères) et à cliquer ensuite sur "Extract HPOs with DiagAI" pour générer une liste des HPOs trouvés. Les termes HPO sont surlignés en couleur et l'utilisateur peut les associer au patient en passant la souris dessus et en cliquant sur "Add", ou ajouter tous les termes HPO trouvés en cliquant sur "Add all" (voir aussi la section **Éditer par les métadonnées de l'échantillon avec le volet d'analyse**).

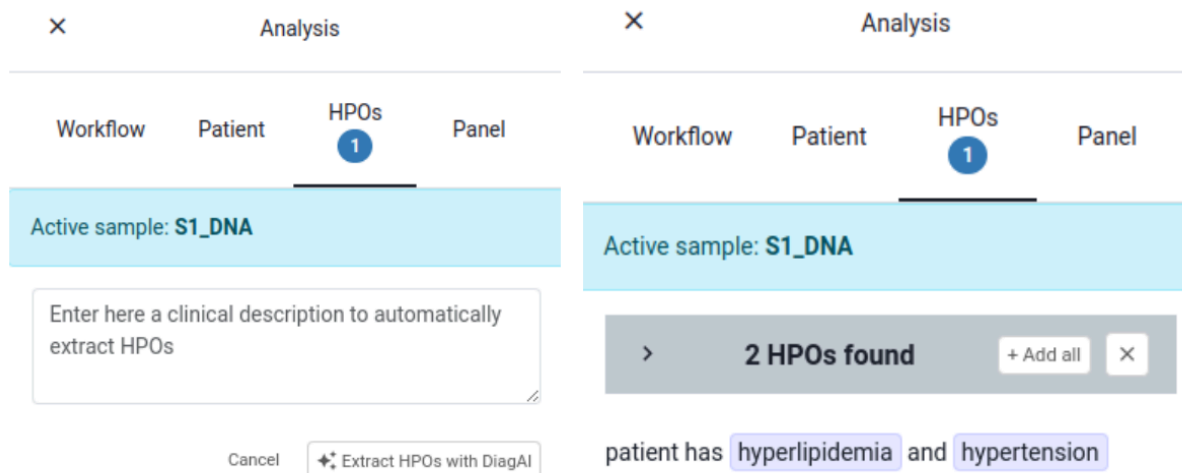
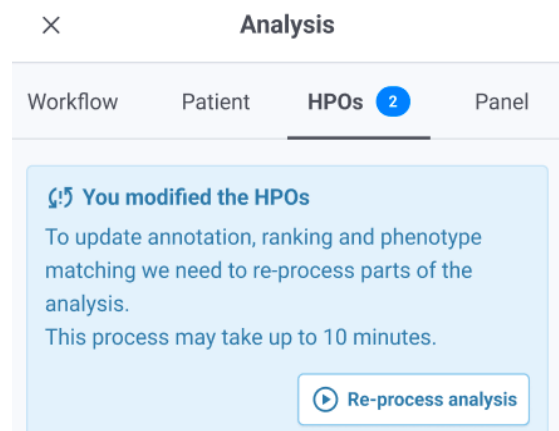


Figure 9 : Extraction des termes HPOs avec DiagAI à partir de la vue HPOs dans le volet d'analyse

Lorsque des HPO sont ajoutés ou modifiés à partir de cette vue, une bannière bleue apparaît, suggérant que l'analyse doit être retraitée avec les nouveaux HPO.

En sélectionnant "*Re-process analysis*", l'utilisateur peut mettre à jour le score DiagAI et la classification des variants (*Short list* ou *SmartPick*) en fonction des nouveaux HPO saisis.

Cette étape de retraitement fige le tableau des variants pendant plusieurs minutes afin de mettre à jour la classification des variants et la colonne DiagAI.

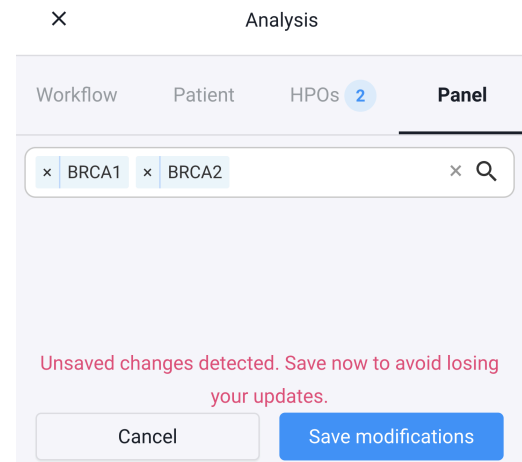


Vue modale Panel

Cette vue est disponible pour tous les worksets, à l'exception des worksets d'analyse CNV (CNVExome, CNVCapture et SomaCNVCapture).

Il donne accès à la fonction *Restricted panels*, qui permet de restreindre et de verrouiller l'analyse à des gènes sélectionnés, à la manière d'un *panel in silico*.

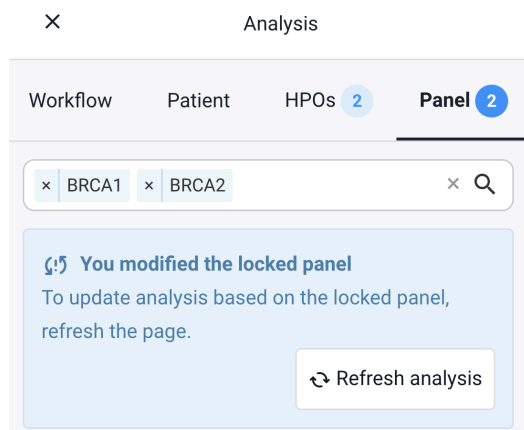
Pour ce faire, l'utilisateur peut sélectionner une ou plusieurs listes de gènes à appliquer, via un menu déroulant, parmi celles précédemment importées (voir paragraphe **Genes**, dans la section **Gestion des paramètres utilisateurs**). En sauvegardant les panels sélectionnés via 'Save modifications', un filtre est appliqué à l'ensemble de l'analyse en fonction du(des) panel(s) de gènes sélectionné(s).



Les panels restreints peuvent être sélectionnés dès le lancement de l'analyse (voir la section **Lancer une nouvelle analyse**) ou à partir de la vue Panel après l'analyse.

Lorsque des panels de gènes sont ajoutés ou modifiés, une bannière bleue avertit l'utilisateur du changement ("*You modified the locked panel*") et lui propose d'actualiser l'analyse avec les nouveaux panels via la fonction "*Refresh analysis*".

En cliquant sur "*Refresh analysis*", l'utilisateur choisit de mettre à jour la table des variants, l'onglet CNVs (si applicable) et l'onglet GeneCov en fonction du nouveau panel.

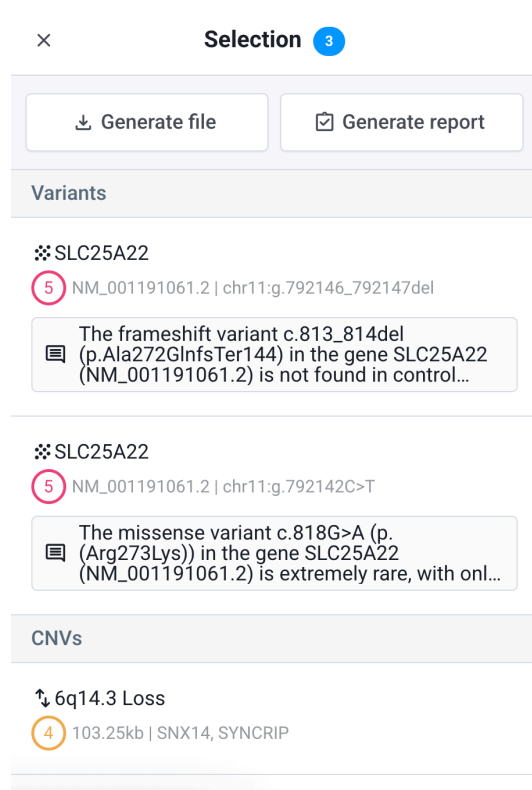


13.1.1.4. Panneau latéral pour la gestion des évaluations

Le panneau  **Selection** est accessible à partir de n'importe quel onglet d'une analyse. Il permet à l'utilisateur de visualiser les petits variants et les CNVs annotés dans VKB pour l'analyse en cours, de générer un fichier JSON ("Generate file") et d'éditer un rapport clinique ("Generate report").

Les CNVs validés dans VKB sont également disponibles dans cette section, uniquement dans le cas de CNVs provenant d'analyses d'exome, de génome et de CGH Microarray.

Les petits variants validés dans VKB sont différenciés par le logo  , tandis que les CNV sont affichés avec le logo  et les mots "gain" ou "loss".



13.1.2. Aperçu des résultats d'analyses dans l'onglet Dashboard

Pour certains worksets, un onglet "Dashboard" qui résume tous les résultats et les mesures de contrôle qualité est accessible aux utilisateurs et apparaît par défaut lors de l'ouverture d'une analyse. L'onglet "Dashboard" fournit une vue d'ensemble des résultats de l'analyse à travers plusieurs sections. Les sections disponibles varient en fonction du workset. Cet onglet est disponible pour SomaCGP, SomaVar Tertiary et SomaLBx.

Patient. Cette section résume les métadonnées de l'échantillon : nom de l'échantillon (*ID de l'échantillon*), sexe et indication de la pathologie (*Indication*).

Quality Controls. Cette section fournit des informations sur plusieurs mesures clés liées à la qualité des échantillons et du séquençage :

- Pureté de la tumeur,
- Ploïdie de l'échantillon,
- le pourcentage de bases dont la qualité est supérieure ou égale à 30 (% Q30 bases),
- la couverture moyenne des régions d'intérêt (ROI average coverage) avant et après le traitement UMI,
- pourcentage de régions d'intérêt avec une couverture supérieure à 200X (ROI > 200X) avant et après le traitement UMI.

Signatures. La section sur les signatures présente les chiffres clés des analyses TMB, MSI et HRD :

- Calcul du TMB (*Tumor mutational burden*) : nombre de mutations/Mb.
- Instabilité des *microsatellites* : pourcentage de sites instables.
- Statut HRD (*Homologous Recombination Deficiency*).

Les statuts des scores TMB et MSI peuvent être sélectionnés par l'utilisateur via les menus déroulants situés sous chaque score.

Selection. Cette section affiche les variants évalués dans la base de données VKB pour l'analyse en cours.

Suggested by SeqOne. Les altérations génomiques détectées dans l'analyse sont présentées sous la forme d'une liste réduite de variants cliniquement pertinents. Les variants sont sélectionnés si leur classe ComPerMed est 5, ou s'ils ont une assertion de niveau I correspondant à l'indication du patient, ou une assertion de niveau II pour des thérapies expérimentales liées à cette indication, ou une assertion de niveau I pour une autre indication. L'utilisateur peut ajouter des variants de cette catégorie à la section Sélection en cliquant sur « Add ».

Le fichier JSON peut être généré en cliquant sur « Generate JSON file » et contient toutes les informations de la page "Dashboard".

13.1.3. Couverture détaillée par exon

Dans l'onglet *GeneCov*, les utilisateurs peuvent visualiser le taux de couverture des gènes inclus dans le manifeste d'analyse pour chaque exon, qu'il s'agisse d'un panel de gènes ou d'un exome.

Methodologie

Ce module intersecte le manifeste de l'utilisateur avec les annotations RefSeq pour tous les gènes inclus dans le panel ou l'exome, en gardant 10bp de chaque côté afin de couvrir les jonctions d'épissage, puis procède au calcul de la profondeur de couverture pour chaque base incluse dans cette intersection à l'aide de l'outil *mosdepth*.

Différents seuils sont fixés afin d'attribuer à chaque exon un statut (*covered*, *warning*, *failed*) en fonction du pourcentage de bases en dessous de ces seuils de couverture.

Le statut est défini à partir de la couverture de base dans l'exon, y compris les 10 pb adjacents. Si des UMI sont utilisées, la couverture est calculée après déduplication.

Les seuils applicables aux différents types d'analyses sont détaillés dans le tableau ci-dessous* :

	GermlineVar /Germline Family/ GermVar / GermVar Family				SomaVar		SomaRNA		SomaVar LF		SomaCGP	
	Panel		Exome		Panel		Panel		Panel		Panel	
	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed
Capture	Cov. 100x < 100%	Cov. 30x < 100%	Cov. 30x < 100%	Cov. 10x < 100%	Cov. 500x <100%	Cov. 200x <100%	Cov. 200x <100%	Cov. 50x <100%	Cov. 5000x <100%	Cov. 2500x <100%	Cov. 100x < 100%	Cov. 50x < 100%
Amplicon	Cov. 100x < 100%	Cov. 30x < 100%			Cov. 1000x < 100%	Cov. 500x < 100%	Cov. 200x <100%	Cov. 50x <100%				

Tableau 3 : Seuils de couverture pour le module *Gene Coverage* en fonction du type d'analyse.

Si l'utilisateur a prédéfini des mesures de contrôle de qualité personnalisées à partir des paramètres du projet, celles-ci seront appliquées dans l'onglet *GeneCov* (voir la section **Configurer les métriques QC par défaut du projet).*

Visualisation et interprétation

Chaque gène peut être visualisé sur une représentation graphique, et les résultats de Gene Coverage peuvent être filtrés en fonction des gènes, des panels in silico et des listes de transcrits. Les exons ciblés sont représentés en vert (*Covered*), orange (*Warning*) ou rouge (*Failed*) en fonction des seuils décrits ci-dessus.

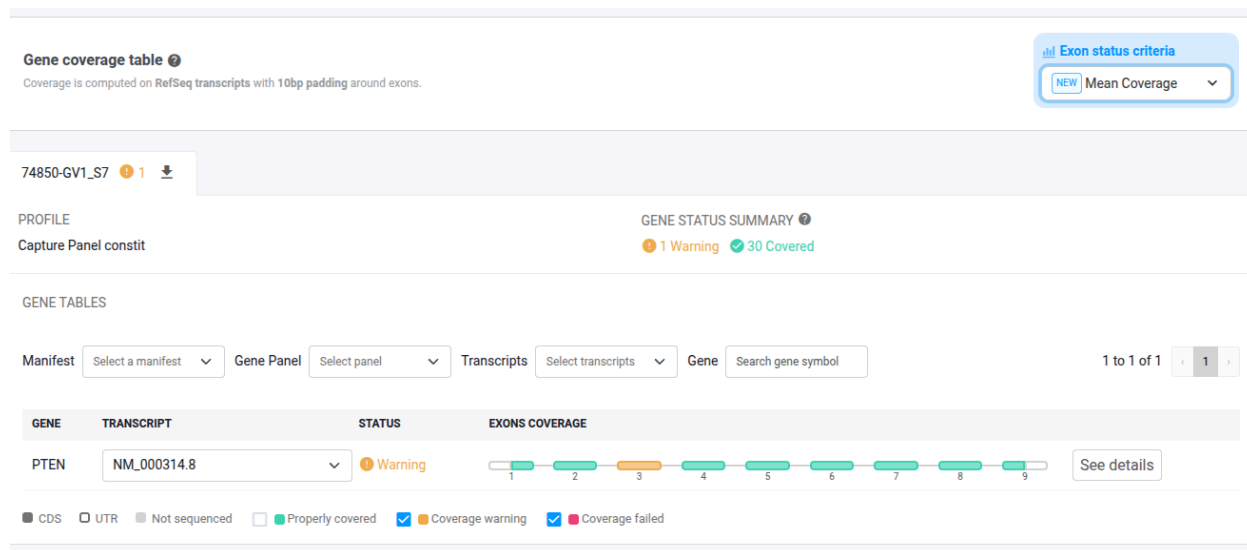


Figure 10 : Détails de l'onglet GeneCov

Options de filtrage

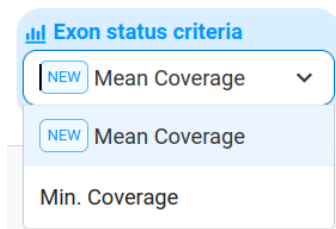
Un menu de filtres vous permet de :

- restreindre l'affichage à un sous-ensemble de régions à l'aide d'un fichier .bed préalablement importé dans l'interface, via le filtre "Manifest",
- appliquer une liste de gènes via le filtre "Gene Panel",
- appliquer une liste de transcrits via le filtre "Transcripts",
- rechercher un gène spécifique via le filtre "Gene".

Méthodes de calcul de la couverture

L'utilisateur peut sélectionner la méthode de calcul de la couverture à appliquer en utilisant le menu déroulant "Exon status criteria" en haut à droite de la page :

1. **Couverture moyenne (Mean Coverage)** (par défaut) : Utilise la profondeur de couverture moyenne sur l'ensemble de l'exon pour déterminer l'état de la couverture.
2. **Couverture minimale (Minimum Coverage)** : utilise la fraction minimale de bases qui répondent à un seuil de couverture spécifique. Cela permet de s'assurer qu'un certain pourcentage de bases a au moins la profondeur de couverture minimale requise.



Le menu déroulant "**Gene status criteria**" permet aux utilisateurs d'appliquer ces méthodes de calcul au niveau du gène :

- **Mean coverage (default):** Utilise la profondeur de couverture moyenne de tous les exons du gène pour déterminer le statut de couverture.
- **Worst exon status:** Détermine le statut de couverture du gène en fonction de l'exon le moins performant. Si un exon ne parvient pas à atteindre le seuil de couverture, le gène est considéré comme non couvert. Il s'agissait de la seule méthode prise en charge dans les versions précédentes.

Données affichées

Les exons qui ne sont pas ou partiellement couverts par le modèle (< 80 % de chevauchement entre l'intervalle manifeste et les limites RefSeq de l'exon) sont considérés comme non ciblés. Ces exons sont représentés en gris (*Not Sequenced*) sur le graphique.

Une vue détaillée de chaque gène peut être utilisée pour afficher les mesures de couverture des exons en cliquant sur "*See details*" : couverture moyenne, médiane, pourcentage de chaque région couverte à 10x, 30x, 50x, 100x, 200x, 300x et 500x.

Pour chaque exon, une fenêtre IGV permet aux utilisateurs de visualiser la couverture de la séquence et l'alignement de manière plus détaillée.

Exportation de données GeneCov

Les données complètes de l'analyse de couverture sont disponibles sous la forme d'un fichier `sample_genecov.v2.tsv.gz`, accessible à partir de l'onglet *Files* de l'analyse.

Dans l'onglet GeneCov, l'utilisateur peut exporter l'analyse de couverture selon la méthode de calcul de la couverture (Mean Coverage ou Min

Coverage), via l'icône  à côté du nom de l'échantillon dans le coin supérieur gauche de l'onglet de l'échantillon.

L'utilisateur peut choisir d'exporter les données en fonction de l'état de la couverture des gènes (*Failed*, *Warning*, *Covered*) et d'exporter le transcrit canonique uniquement (*Canonical*) ou toutes les transcrits disponibles (*All*) avant de cliquer sur "Download" pour obtenir le fichier souhaité.

Download genes from S1_DNA as .tsv file ×

Genes status

☒ Failed ☒ Warnings ☐ Covered

Transcripts

☒ Canonical ☐ All

Close Download

Les données sont exportées sous la forme d'un fichier .tsv contenant les colonnes suivantes :

- Gène (symbole du gène),
- Numéro du gène selon le HGNC (gene_id),
- Transcript canonique (is_canonical_transcript),
- Statut du transcrit (transcript_status)
- Exon (numéro d'exon)
- Type d'exon CDS ou UTR (exon_type)
- Statut de l'exon couvert ou non couvert (exon_status)
- Région couverte par le manifeste ou non (cible)
- Position génomique (interval_id)
- Numéro de chromosome (chr)
- Position de départ (start)
- Position finale (fin)
- Brin sens ou antisens (strand)
- % de couverture à 1, 10, 20, 30, 50, 100, 200, 300, 500 et 1000X (cov_)
- Couverture moyenne (cov_avg)
- Couverture médiane (cov_med)

13.1.4. Petits variants

Méthodologie

La détection des petits variants est assurée par une combinaison de plusieurs outils, qui peuvent varier d'un workset à l'autre.

Que ce soit pour l'analyse d'un exome ou d'un panel de gènes, l'appel de variants est déterminé par le manifeste (au format BED) associé au projet, incluant les intervalles des régions d'intérêt et le padding appliqué à ces régions (voir la section **Créer d'un projet**).

De plus, les variants sont appelés en tenant compte de la qualité d'alignement des reads sur le génome humain (version GRCh37 ou GRCh38). Les reads dits "MAPQ0" dont la qualité d'alignement (*mapping*) est égale à 0, parce qu'ils s'alignent sur plusieurs sites du génome, ne sont pas pris en compte lors de l'appel de variants pour les worksets d'analyse somatique. Dans les analyses "Germline", les reads MAPQ0 sont prises en compte pour l'appel de variants si un manifeste spécifiant des régions d'alignement de faible qualité est saisi lors de la création du projet sous les métadonnées "*Low-Quality Mapping Region*". Les variants détectés à partir des reads MAPQ0 sont identifiés dans le tableau des variants par une infobulle "_mapq0" visible dans la colonne CAL (Caller).

SNVs / indels. Pour la détection des SNV et des indels :

- *Freebayes*¹ est utilisé comme seul *caller* de variants par les worksets GermlineVar (exome), GermlineFamily (exome), SomaRNA et SomaVar LF. Les seuils de détection inférieurs sont les suivants :
 - VAF $\geq$ 10% dans GermlineVar pour l'exome,
 - VAF $\geq$ 5% dans GermlineVar pour les panels de gènes,
 - VAF $\geq$ 1% dans le SomaRNA,
 - VAF $\geq$ 0,1 % dans SomaVar LF.
- Il est accompagné de *GATK HaplotypeCaller*² dans le workset GermlineVar (panel).
- Son homologue somatique, *Mutect2*³, est utilisé avec Freebayes dans les worksets SomaVar, SomaHemato et SomaCGP. Le seuil de détection minimum est VAF $\geq$ 1% pour ces deux *callers* dans SomaVar.
- *Sentieon*⁴ est utilisé comme principal caller par les worksets GermVar, GermVar Family et SomaLBx.
- *Clair3*⁵ est utilisé comme seul appelant de variants par le workset GermlineVar LR.

¹ <https://github.com/freebayes/freebayes>

² <https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>

³ <https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2>

⁴ <https://github.com/Sentieon>

⁵ <https://github.com/HKU-BAL/Clair3>

Les analyses GermVar et GermVar Family s'appuient sur les composants bioinformatiques Sentieon pour les étapes clés de l'analyse secondaire. En particulier :

- L'alignement est effectué à l'aide de l'implémentation optimisée de BWA-MEM2 par Sentieon, et les lectures dupliquées sont marquées.
- L'appel germline (SNVs/Indels) est effectué à l'aide de Sentieon DNAScope⁶.
- Sentieon TNScope⁷ peut être utilisé pour détecter des variants en mosaïque à faible VAF (selon la configuration GermVar).

Note concernant les régions à faible mappabilité (MAPQ0) : dans les régions hautement homologues (par ex., SMN1/SMN2), les variants peuvent être filtrés par le filtrage d'apprentissage automatique de l'outil d'appel. Lorsque cela est nécessaire, SeqOne fournit un module MAPQ0 optionnel (module de régions de faible qualité) basé sur Freebayes avec des seuils adaptés pour augmenter la sensibilité dans des régions définies par l'utilisateur.

Les analyses SomaLBx s'appuient sur les composants bioinformatiques Sentieon pour les étapes clés de l'analyse secondaire de biopsie liquide. En particulier :

- L'alignement des lectures est effectué avec Sentieon (BWA-MEM), et le pipeline inclut le traitement/consensus UMI basé sur Sentieon pour prendre en charge les flux de travail ctDNA à haute sensibilité.
- L'appel somatique de variants courts (SNVs/Indels) est effectué à l'aide de Sentieon TNScope⁷ en mode tumeur seule (tumor-only), avec un support de fraction allélique très faible (VAF $\geq 0,1$ % selon la configuration).

Note : les génotypes des variants sont signalés comme hétérozygotes par défaut dans les résultats SNV/indel de SomaLBx car Sentieon TNScope n'effectue pas de prédiction de génotype.

SVs - Structural Variants. Les insertions et les délétions dont la longueur est proche ou supérieure à celle d'un read de séquençage sont détectées à l'aide du SV caller GRIDSS⁸ dans les worksets GermlineVar (panel/exome), SomaVar, SomaVar LF, SomaHemato et SomaCGP. Ces événements sont détectés à l'aide du caller SV Delly⁹ dans les worksets GermVar et GermVar Family. Seules les insertions et les délétions inférieures à 300 pb sont détectées dans les worksets somatiques, tandis qu'une limite de 10 kb est appliquée dans le workset GermlineVar (pour les versions ≥ 2.0) et les worksets GermVar et GermVar Family (versions ≥ 3.5). Les variants détectés par GRIDSS sont détectés quelle que soit la VAF.

FLT3-ITD. Un module spécifique de détection des duplications en tandem internes du gène FLT3 (FLT3-ITD) n'est disponible dans les worksets SomaVar et SomaHemato qu'à partir de la version v2.2. Ce module est basé sur l'outil *FILT3r*¹⁰.

⁶ <https://github.com/Sentieon/sentieon-dnascopy>

⁷ <https://github.com/Genomic-Medicine-Linkoping/TNScope>

⁸ <https://github.com/PapenfussLab/gridss>

⁹ <https://github.com/dellytools/delly>

¹⁰ <https://gitlab.univ-lille.fr/filt3r/filt3r>

Alu Insertions. Enfin, un module de détection des insertions d'éléments transposables (*AluMEI*), tels que les éléments Alu, a également été développé par SeqOne. Il est utilisé par les worksets GermlineVar (panel/exome), GermlineFamily (panel/exome), GermVar (panel/exome/génome), GermVar Family (panel/exome/génome), SomaHemato et SomaVar. Aucun seuil inférieur de VAF n'est appliqué pour les variants détectés par *AluMEI*.

Les variants combinés de *Freebayes*, *HaplotypeCaller* (ou *Mutect2*) et *FiLT3r* sont disponibles dans le fichier *final.vcf.gz* de l'analyse. Les variants spécifiques à *GRIDSS*, *FiLT3r* ou *AluMEI* sont disponibles dans leurs fichiers VCF respectifs (***gridss.vcf.gz***, ***filt3r.vcf.gz*** et ***variants.alumei.vcf.gz***), également disponibles en téléchargement à partir de l'onglet *Files* de l'analyse.

Dans le cas spécifique de *FiLT3r* :

- Les variants de *FiLT3r* communs à *Freebayes* et *Mutect2* sont disponibles dans le VCF final de l'analyse (***final.vcf.gz***),
- Les variants uniques à *FiLT3r* ainsi que les variants communs à *FiLT3r* et à *GRIDSS* sont répertoriés dans le fichier VCF de *FiLT3r* (***filt3r.vcf.gz***).

Deux modules supplémentaires développés par SeqOne permettent ensuite :

- La rétention des variants multinucléotidiques, ou MNV, qui sont codants et limités à un seul codon au cours du processus de décomposition des variants de type delins.
- LA recalibration de la fréquence allélique des indels et des delins détectées par *Freebayes* et *GATK* (outil *Kraf*).

Ces deux modules ne concernent que les worksets GermlineVar (panel/exome), SomaVar, SomaRNA et SomaVar LF.

Kraf utilise une stratégie de k-mers simples, sans alignement. La fréquence allélique recalibrée des indels et delins n'est retenue que si elle est supérieure à la fréquence allélique initiale du variant.

Les variants sont ensuite principalement annotés à l'aide de l'outil *Variant Effect Predictor*¹¹ (VEP), et des informations supplémentaires sont ajoutées à partir des bases de données suivantes :

- ClinVar
- OMIM (version commerciale) + PanelApp Australie + PanelApp Angleterre
- CKB
- dbSNP
- COSMIC (version commerciale)
- GnomAD (données exome + génome)

¹¹ <https://www.ensembl.org/info/docs/tools/vep/index.html>

- Dfam
- DGV
- LOVD

En outre, les scores *in silico* suivants sont disponibles :

- Scores d'épissage (*MaxEntScan, GeneSplicer, dbSCSNV*)
- Score de conservation (*GERP++, PhastCons100way, PhastCons20way, PhyloP100way Vertebrate, PhyloP20way Mammalian, SiPhy*)
- Prédiction fonctionnelles (*REVEL, DANN, FATHMM, FATHMM-MKL, LRT, MetaLR, MetaSVM, MutationAssessor, MutationTaster 2, PROVEAN, SIFT, PLI*)

Visualisation et interprétation

L'onglet Variants  permet de visualiser, de filtrer et d'annoter les variants obtenus lors d'une analyse. Ceux-ci sont affichés sous forme de tableau, chaque ligne correspondant à un variant différent. Le nombre de variants sélectionnés après filtrage (variants sélectionnés) et le nombre total de variants détectés (nombre total de variants) sont affichés dans le coin supérieur droit de

l'en-tête du tableau des variants à l'aide de l'icône suivante :  9/22 . Une pagination en bas de page permet de naviguer d'une page de résultats à l'autre. Si l'analyse utilise un panel restreint, l'en-tête du tableau comprendra également le nombre total de variants dans le panel restreint (*total number of variant in the restricted panel*).

Cette navigation est facilitée par des raccourcis clavier utilisant les touches ← et → du clavier.

Par défaut, plusieurs colonnes descriptives sont affichées dans le tableau. Les utilisateurs peuvent exporter le tableau sous forme de fichier TSV en cliquant sur .

Plusieurs transcrits peuvent être disponibles pour chaque variant, et leur nombre est affiché à gauche de chaque ligne. Les variants sont annotés sur les transcrits RefSeq, à l'exception des variants mitochondriaux qui sont annotés sur *Ensembl*.

Les transcrits MANE Select et MANE Clinical Plus sont identifiés par

une icône  ou , respectivement, accompagnée d'une infobulle.

TRANSCRIPT	HEVC
MANE Select transcript	
NM_005962.5 	c.536T>G p.V179G
NM_015252.5 	c.-296+2T>G
MANE Plus Clinical transcript	
NM_000132.4 	c.5375T>G p.V1792G

Pour un même variant, à l'exception des variants mitochondriaux, les transcrits sont classés dans l'ordre ci-dessous :

- l'impact ['high' > 'moderate' > 'low' > 'modifier']
- type de transcrit ['NM', 'XM', 'NR', 'XR'] (pour les transcrits RefSeq)
- Type de transcrit MANE [Mane Select > MANE Clinical Plus > non MANE]
- ordre alphanumérique

Pour les variants mitochondriaux, les transcrits sont triés dans l'ordre suivant :

- impact ['high' > 'moderate' > 'low' > 'modifier']
- si l'impact est 'modifier' :
 - l'effet 'non_coding_transcript_exon_variant'
 - le biotype [codage des protéines > Mt_rRNA > Mt_tRNA].
- ordre alphanumérique

DiagAI. Cette colonne indique le score de 0 à 100 basé sur la suite DiagAI utilisant le *machine learning*. Le score DiagAI est basé sur quatre composantes :

- L'algorithme UP², qui prédit la pathogénicité du variant
- La matrice PhenoGenius, qui évalue la correspondance entre le gène et les HPO du patient (si ceux-ci ont été saisis lors de l'analyse).
- La concordance entre le génotype observé et le mode de transmission de la maladie décrit dans les bases de données GenCC, OMIM et PanelApp (Australie et Angleterre).
- La qualité du variant

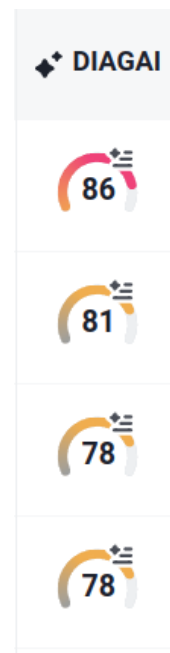
Plus le score est proche de 100, plus le variant est prédit comme étant pathogène et présentant un intérêt diagnostique. Plus il est proche de 0, plus le variant est considéré comme bénin.

Deux icônes attirent également l'attention sur certains variants d'intérêt :

- L'icône  signifie que le variant a une probabilité plus élevée d'être la cause de la maladie (*AI Smart Pick*).
- L'icône  indique que le variant est présent dans la liste restreinte des variants ayant une forte probabilité d'être à l'origine de la maladie du patient.

Les termes HPO peuvent être saisis avant le lancement de l'analyse en modifiant les métadonnées de l'échantillon (voir la section **Métadonnées de l'échantillon**). Ils peuvent également être modifiés pendant l'interprétation en utilisant la vue modale HPOs dans le panneau latéral de l'analyse (voir la section **Vue modale HPO**). Dans ce cas, l'utilisateur peut retraiter le score DiagAI et la classification des variants (Short list ou SmartPick) en fonction des nouveaux HPOs saisis, en utilisant la fonction "*Re-process with new HPOs*".

Si l'utilisateur clique sur l'icône du score DiagAI, la page du variant s'ouvre directement sur l'onglet DiagAI détaillant le score (voir la section *DiagAI* dans la page du variant ci-dessous).



ACMG. La classification de pathogénicité de chaque variant est calculée dans le cadre d'un processus semi-automatisé selon les recommandations émises par l'*American College of Medical Genetics and Genomics*. Les utilisateurs peuvent réviser la classification dans un onglet dédié sur la page du variant, qui est ensuite utilisée pour générer l'évaluation finale d'un variant.

ComPerMed. La classification de la pathogénicité de chaque variant est calculée dans le cadre d'un processus semi-automatisé selon les recommandations émises par le comité d'experts belge ComPerMed (Personalised Medicine Commission). Les utilisateurs peuvent accéder à des informations détaillées sur la classification dans un onglet dédié sur la page du variant.

AMP. (version d'analyse SomaVar v1.8 ou ultérieure, SomaVar LF ou SomaCGP uniquement). Cette colonne est affichée par défaut dans les analyses qui contiennent des annotations CKB et montre le niveau de AMP Tier de chaque variant lorsqu'il est disponible (niveau I ou II). Le niveau de AMP Tier affiché est basé sur les meilleures preuves cliniques pour tous les types, quelle que soit l'indication sélectionnée.

VKB. Les évaluations faites par l'utilisateur lors des analyses précédentes sont stockées dans la base de connaissances de l'utilisateur. Cette évaluation apparaît dans la colonne VKB du tableau pour chaque nouvelle occurrence du variant. Une infobulle affiche rapidement la classe et le commentaire associés au variant évalué ou précédemment évalué.

Sort by: DiagAI RUO	
	Likely Pathogenic Comment: this is my previous comment...
5	81 5 4 BRCA1 NM_007294.4

Gene. Nom des gènes selon le [HGNC](#) et drapeau OMIM associé :

- Drapeau rouge : "gène OMIM **avec** phénotypes **et** caractéristiques cliniques",
- Drapeau à contour rouge : "gène OMIM **avec** phénotypes mais **sans** caractéristiques cliniques",
- Point rouge : "gène OMIM **sans** phénotypes ni caractéristiques cliniques".

CB	GENE	TRANSC
	RPE65	NM_000...
	OMIM	
	Decipher	1_206
	Orphanet	
	MedGen	1_005
	GenCC	

Les gènes connus pour être des gènes d'empreinte selon PanelApp (Australie) sont mis en évidence par l'icône . L'infobulle associée indique si le gène est exprimé maternellement, paternellement, inconnu, ou si plusieurs empreintes sont référencées.

En cliquant sur le drapeau situé à côté du nom du gène, vous accédez directement à la page

OMIM. En cliquant sur le nom du gène, un menu déroulant apparaît, fournissant des liens vers les bases de données OMIM¹², Decipher¹³, Orphanet¹⁴, MedGen¹⁵ et GenCC¹⁶. Chaque lien mène directement à la page correspondant au gène sélectionné.

Les informations relatives à OMIM et aux bases de données ne sont disponibles que pour les analyses germinales.

HGVSC. Cette colonne fournit l'annotation HGVS du variant sur l'ADN complémentaire (ADNc) et sur la protéine.

HGVSC

c.10580G>A
p.R3527Q

Effect. Cette colonne décrit l'impact du variant sur la protéine et l'effet sur le transcrit.

L'impact est calculé selon 4 niveaux :

- **High** (impact élevé, en rouge) : le variant a un impact perturbateur élevé sur la protéine, entraînant probablement une troncature de la protéine, une perte de fonction ou le déclenchement de sa désintégration médiée par le non-sens.
- **Modéré** (impact modéré, en jaune) : variant non perturbateur susceptible de modifier l'efficacité de la protéine.
- **Faible** (impact faible, en gris) : variant supposé inoffensif ou peu susceptible de modifier le comportement de la protéine.
- **Modificateur** (impact modificateur, en bleu) : généralement des variants non codants ou des variants affectant des régions ou des gènes non codants, pour lesquels les prédictions sont difficiles ou pour lesquels il n'y a pas de preuve d'impact

Pour chaque variant, plusieurs effets sur le transcrit peuvent être disponibles dans la colonne *Effect*. Dans ce cas, seul l'effet ayant la conséquence la plus élevée selon VEP¹⁷ est affiché, suivi d'un nombre (par exemple +2) pour les autres effets disponibles. Les détails de ces derniers sont indiqués dans l'infobulle correspondante.

EFFECT	
Moderate	missense
Low	splice polypyrimidine tract + 1
Moderate	missense
Moderate	missense
Moderate	missense
Moderate	missense

¹² <https://www.omim.org/>

¹³ <https://www.deciphergenomics.org/genes>

¹⁴ <https://www.orpha.net/fr/disease/gene>

¹⁵ <https://www.ncbi.nlm.nih.gov/medgen/>

¹⁶ <https://search.thegencc.org/>

¹⁷ http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html

Transmission. Cette colonne indique le mode de transmission des maladies associées aux gènes signalés dans les bases de données GenCC et OMIM. Des badges de couleur sont visibles pour identifier rapidement les modes de transmission basés sur les prédictions de DiagAI :

- "De Novo"
- "Compound Het - Verified"
- "Recessive Hom"
- "Compound Het - Suspected"

Les badges *De Novo*, *Compound Het - Verified* et *Recessive Hom* ne sont affichés que dans les analyses de famille.

Un filtre peut être utilisé pour sélectionner les variants avec ces badges (voir la section **Filtrer les variants - Transmission et phénotypes**).

TRANSMISSION	
AD AR	CompHet Verified Nail disorder, nonsyndromic ... +1
AD	CompHet Suspected Coffin-Siris syndrome 2
AD	De Novo Advanced sleep phase syndrome..
AR	Recessive Hom Advanced sleep phase syndrome..
AD	Tarsal-carnal coalition syndrome +3

Transmission
AR, AD

Leber congenital amaurosis 9
OMIM:7638289

AD, AR | OMIM
AD | 3 GenCC
AR | 3 GenCC

Leber congenital amaurosis 2
OMIM:0019200

AD, AR | OMIM
AR | 3 GenCC

Leber congenital amaurosis 5
MONDO:739002

AR | 3 GenCC

Retinis pigmentosa
OMIM:10939229

AD, AR | OMIM

Definitive to Moderate
Supportive
Limited to Unknown

En cliquant sur les liens relatifs aux modes de transmission dans cette colonne, une fenêtre contextuelle s'ouvrira dans laquelle les utilisateurs pourront trouver les maladies répertoriées avec leur nom et les liens vers les bases de données (OMIM, MONDO - si OMIM n'est pas disponible - ou GenCC).

Les maladies sont organisées selon une hiérarchie spécifique : celles qui figurent à la fois dans OMIM et GenCC apparaissent en premier, suivies des entrées OMIM uniquement, puis des entrées GenCC uniquement.

Chaque entrée comporte des étiquettes codées par couleur : Les étiquettes OMIM ont des bordures vertes, tandis que les étiquettes GenCC utilisent un système de feux tricolores (vert/jaune/rouge) pour indiquer la force des preuves :

- Vert : *Definitive + Strong + Moderate*
- Jaune : *Supportive*
- Rouge : *Limited, Animal, Unknown*

Clinical Evidences. (analyse SomaVar v1.8 ou ultérieure, SomaVar LF ou SomaCGP uniquement). Cette colonne est affichée par défaut dans les analyses contenant des annotations CKB. Elle est divisée en trois sous-colonnes, chacune contenant le nombre total d'un type de preuve donné : A (Actionnabilité), D (Diagnostic) et P (Pronostic). La colonne A affiche également le nombre d'affirmations thérapeutiques sensibles (à gauche) et résistantes (à droite) sous forme de badge.

CLINICAL EVIDENCES		
A	D	P
81 81 8	0	0
54 53 2	0	0
54 53 2	0	0

Les valeurs de ces sous-colonnes sont codées par couleur :

- La résistance, lorsqu'elle est détectée (> 0 assertions), est toujours affichée en rouge, quelle que soit l'indication, et en gris dans le cas contraire.
- La sensibilité est mise en évidence en vert uniquement lorsque l'indication correspondante est sélectionnée, et en gris dans le cas contraire.

Si au moins une preuve de ce type correspond à l'indication sélectionnée dans le tableau des variants (panneau Indication), une coche apparaît dans l'infobulle devant le libellé de l'indication. Si aucune correspondance n'est trouvée, l'étiquette n'apparaît pas.

Phenotypes. Cette colonne, disponible pour les projets Germline, permet d'associer un gène à des termes HPO à l'aide de la matrice PhenoGenius¹⁸. Si des termes HPO sont associés à l'échantillon, le pourcentage le plus élevé de correspondance entre le gène et le phénotype est affiché, ainsi que le ratio des termes auxquels le gène est associé. Si aucun gène ne correspond aux HPO saisis, la colonne renvoie le message "pas de correspondance" ; si l'échantillon n'a pas de termes HPO, la colonne affiche le message "voir HPO". Dans les deux cas, les termes connus pour le gène sont néanmoins proposés dans la fenêtre pop-up associée.

Phenogenius

✓ **Patient phenotypes matching P3H1**

- 48% HP:0000773 Short ribs
- 47% HP:0003026 Short long bone
- 0% HP:0004443 Lambdoidal craniosynostosis
- 0% HP:0100832 Vitreous floaters
- 0% HP:0008479 Hypoplastic vertebral bodies

Other phenotypes described for P3H1

- Abnormal bone collagen fibril morphology
- Abnormal cartilage collagen
- Type 1 collagen overmodification
- Decreased calvarial ossification
- Femoral bowing

Phenotypes

- 48% | 2/5
- 48% | 2/5
- no match
- no match
- 24% | 1/5
- no match
- no match

Transcript	NA Mutation	AA Mutation	Impact	Effect	Transmission	Phenotypes
						see HPO
						see HPO
						see HPO

Phenogenius

Known phenotypes described for SLC25A22

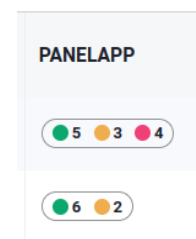
- Epileptic encephalopathy
- Encephalopathy
- EEG with burst suppression
- Focal myoclonic seizure
- EEG with spike-wave complexes

Figure 11 : Visualisation des informations sur Phenogenius

¹⁸ Kevin Yaury et al, " Learning phenotypic patterns in genetic diseases by symptom interaction modeling ", medRxiv, p. 2022.07.29.22278181, janv. 2022, doi : 10.1101/2022.07.29.22278181.

PanelApp. Cette colonne intègre les annotations de la base de données de panels PanelApp *Australia*¹⁹ et *PanelApp England*²⁰. Elle indique le nombre de panels associés au gène pour lequel un variant a été détectée, selon le niveau de classification décrit ci-dessous.

Si un gène n'est pas trouvé dans un panel, la colonne affichera "N/A".



La fenêtre contextuelle associée affiche les détails de l'association gène/phénotype, ainsi que le mode de transmission associé au phénotype, le mode de transmission associé au phénotype et la base de données à partir de laquelle le panel est dérivé, PanelApp Australia (**AUS**) ou PanelApp England (**ENG**) (**Figure 12**).

Chaque gène se verra attribuer un ou plusieurs panels selon un système de classification à trois niveaux basé sur les preuves de l'association entre le gène et le phénotype : Les preuves élevées sont indiquées en **vert**, les preuves modérées en **jaune** et les preuves faibles en **rouge**.

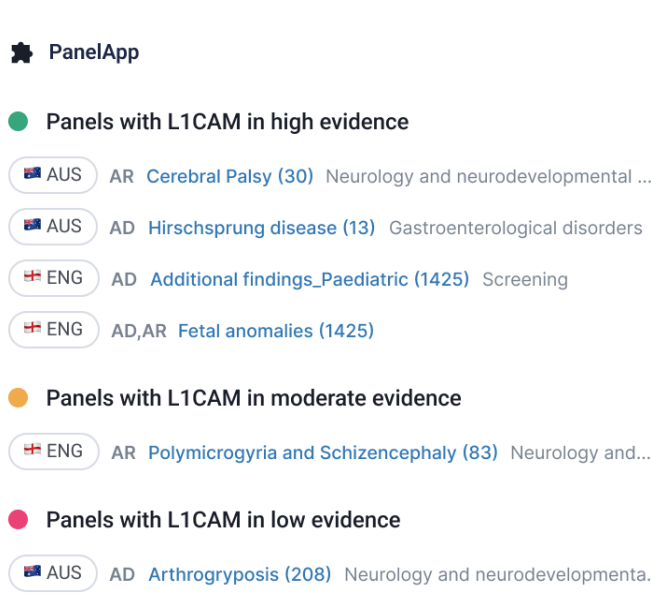
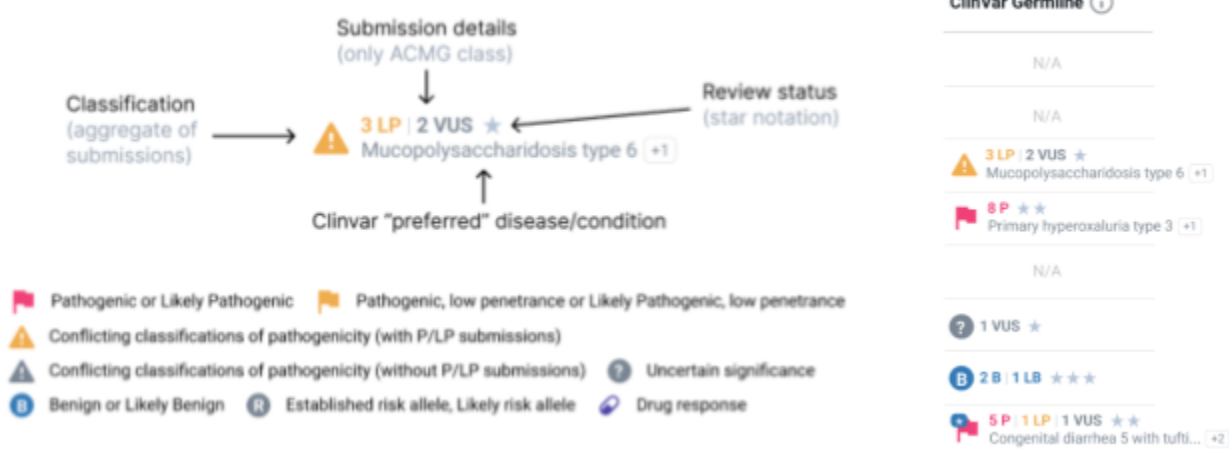


Figure 12 : Visualisation des informations de PanelApp.

¹⁹ <https://panelapp-aus.org>

²⁰ <https://panelapp.genomicsengland.co.uk/>

ClinVar Germline. Cette colonne affiche des informations extraites de ClinVar au cours de l'analyse. Elle fournit une série d'informations, notamment la classe de pathogénicité, les détails et le nombre de soumissions pour chaque classe ACMG, l'état de la révision et la maladie préférée de ClinVar. En cliquant sur la cellule du tableau des variants, l'utilisateur sera redirigé vers la page ClinVar correspondante.



Pour plus d'informations, voir https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/

GnomAD. Cette colonne affiche 6 éléments de la base de données GnomAD pour un variant spécifique dans la population : la fréquence allélique (AF), le nombre d'allèles (AC), le nombre d'homozygotes (HOM), le nombre d'hémizygotes (HEM), la fréquence de l'allèle majeur (MAF) et, pour les projets GRCh37, la version de GnomAD pour laquelle ce variant est trouvé (VER).

En cliquant sur la valeur affichée dans la colonne AF, l'utilisateur est redirigé vers la base de données GnomAD dans sa version v4 ou, à défaut, dans sa version v2. Si un variant n'est pas répertorié dans la base de données, la valeur affichée dans cette colonne sera N/A

GNOMAD				PROJ
	AC	HOM	VER	
e-1	47775	9503	v2 ▲	4 / 4

jeu(x) de données incriminé(s).

Une alerte est affichée dans la colonne VER si ce variant est présent dans GnomAD, mais ne passe pas tous les filtres de qualité pour au moins un des deux jeux de données (exome ou génome). Dans ce cas, les valeurs affichées s'adaptent aux valeurs disponibles pour ce variant, et les filtres qui ont déclenché l'alerte sont affichés, ainsi qu'un court message mentionnant le(s)

Score UP². Le score de prédiction fonctionnelle UP² (*Universal Pathogenicity Prediction*) allant de -1 à 1 classe la pathogénicité des variants à l'aide d'une approche de *Machine Learning* selon trois catégories : **P** (Pathogenic : score > 0,1), **VUS** (*Variant of Unknown Significance* : score ≥ -0,1 et ≤ 0,1) ou **B** (*Benign* : score < -0,1). Le score UP² est affiché par défaut dans le tableau des variants, mais l'utilisateur peut désactiver la colonne UP² à partir du panneau de gestion des

colonnes en cliquant sur l'icône .

CI-SpliceAI. Cette colonne comprend des scores de prédiction calculés à l'aide d'une approche de *machine learning* pour identifier les variants ayant un impact sur l'épissage alternatif à la manière de *CI-SpliceAI*.²¹

SpliceAI				
AG	DG	AL	DL	Link
0.01	0.02	0.62	0.84	↗
0.01	0.22	0.01	0.09	↗
0.01	0.02	0.12	0.01	↗
0.01	0.62	0.02	0.01	↗
N/A	N/A	N/A	N/A	↗
No splicing impact				

Position: -2bp

Open variant in SpliceAI Lookup

SpliceAI score was not significant (Delta score <0.2)

Les scores sont calculés pour un sous-ensemble de variants répondant aux critères suivants : variants d'épissage selon VEP (splice_*) avec une fréquence GnomAD inférieure à 10⁻³ et répondant aux critères de qualité suivants (QA > 20, AO > 10, DP > 20).

Si calculés, la colonne fournit 4 scores delta pour le variant concerné, pour le gain d'accepteur (AG), la perte d'accepteur (AL), le gain de donneur (DG) et la perte de donneur (DL).

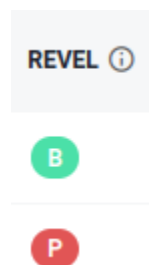
Le score delta d'un variant, compris entre 0 et 1, correspond à la probabilité que le variant modifie l'épissage. Un code couleur est appliqué aux seuils du score delta : **vert** (>0,2 ; rappel élevé), **jaune** (>0,5 ; recommandé) et **rouge** (>0,8 ; précision élevée). Si le score delta est inférieur à 0,2, nous considérons que le score est insignifiant. La colonne indique alors "No splicing impact".

Pour chaque variant, un lien est fourni vers l'interface de recherche *SpliceAI*²² correspondante.

²¹ https://github.com/YStrauch/CI-SpliceAI__Annotation

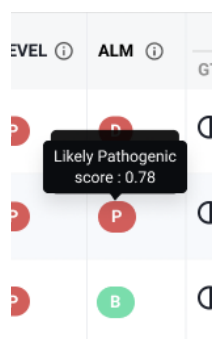
²² <https://spliceailookup.broadinstitute.org/>

REVEL. Le score de prédiction de REVEL est basé sur une méthode globale de prédiction de la pathogénicité des variants faux-sens utilisant une combinaison de scores provenant de 13 outils individuels : MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP et phastCons.



Le score peut varier de 0 à 1 pour un variant faux-sens, les scores les plus élevés reflétant une plus grande probabilité que le variant soit à l'origine de la maladie. Les seuils appliqués sont les suivants : score REVEL > 0,5 pour **P** (pathogène) et score REVEL < 0,5 pour **B** (bénin). L'utilisateur peut afficher le score dans le tableau des variants en cochant la case REVEL dans l'encart des colonnes, dans la section Prédiction fonctionnelle.

ALM. Le score de prédiction d'AlphaMissense²³ est disponible dans la colonne ALM. AlphaMissense est un modèle de prédiction qui évalue la pathogénicité des variants faux-sens à l'aide de l'apprentissage profond. Le modèle est construit sur les prédictions structurales d'AlphaFold pour améliorer sa précision.



AlphaMissense intègre la conservation évolutive et le contexte structural pour évaluer l'impact potentiel des variants faux-sens sur la fonction des protéines.

Le score est compris entre 0 et 1, les valeurs les plus élevées indiquant une plus grande probabilité qu'un variant soit délétère. Les seuils appliqués sont les suivants : score ALM > 0,564 pour **P** (Probablement pathogène) et score ALM < 0,34 pour **B** (Probablement bénin). Dans les autres cas (0,34 < score ALM < 0,564), il sera étiqueté **VUS** (Ambigu).

Sample Freq. Cette colonne indique pour chaque variant sa fréquence dans le projet (sous-colonne PROJECT) et dans tous les projets de l'entité avec des métadonnées similaires (sous-colonne TOUS).

SAMPLE FREQ.	
PROJECT	ALL
2 / 13	7.41%
2 / 13	7.41%

La fréquence d'échantillonnage du projet (PROJECT) correspond au nombre d'échantillons dans lesquels le variant est détecté par rapport au nombre total d'échantillons analysés dans le cadre du projet.

La fréquence d'échantillonnage ALL correspond au pourcentage d'échantillons analysés dans lesquels le variant est trouvé sur l'entité, dans des projets de métadonnées similaires. En cliquant sur la valeur affichée dans la colonne ALL, l'utilisateur peut accéder aux noms des projets et au nombre d'échantillons dans lesquels le variant est présent, ainsi qu'aux détails de tous les projets de métadonnées similaires (option

²³ <https://alphamissense.hegelab.org/>

"Show all projects").

Par défaut, la sous-colonne ALL est masquée dans le tableau des variants, mais elle peut être réaffichée à partir du panneau de gestion des colonnes en cliquant sur l'icône  en activant le bouton "Sample Frequency Entity (Entity)".

Exon. Cette colonne indique le nombre total d'exons dans le gène selon le transcrit concerné, ainsi que le nombre affecté par la mutation, sous forme de fraction. Une infobulle indique la distance entre le variant et le site d'épissage le plus proche. Si le variant est en amont du site d'épissage, la valeur sera positive et s'il est en aval, elle sera négative.

GT (genotype). Cette colonne indique le génotype prédit par les programmes de détection de variants. Les différents pictogrammes possibles sont décrits dans la section "**Filtrer les variants**" ci-dessous :

Dans une analyse en trio (GermlineFamily), si l'origine d'un variant hétérozygote peut être déterminée, elle sera indiquée par le code couleur suivant :

- **rose** → origine hétérozygote maternelle
- **bleu** → origine paternelle hétérozygote
- **rouge** → variant hétérozygote *de novo*
- **orange** → violation mendélienne

Aucune couleur n'est appliquée si la transmission est indéterminée (par exemple, si les deux parents sont hétérozygotes) ou pour les variants homozygotes.

Patient				Father				Mother			
GT	VAF	QUAL	COV	GT	VAF	QUAL	COV	GT	VAF	QUAL	COV
	50.8%	34	61		N/A	N/A	82		44.6%	33	83
	53%	34	83		N/A	N/A	49		N/A	N/A	100
	46.9%	34	64		48.8%	33	84		N/A	N/A	86



Figure 13 : Code couleur appliqué aux génotypes hétérozygotes sur la base de la transmission GermlineFamily

La mise en phase des variants est disponible pour les analyses GermVar Tertiary et GermVar Tertiary Family utilisant des fichiers VCF provenant de séquenceurs d'Oxford Nanopore Technologies. Lorsque les variants sont mis en phase, les icônes de génotype apparaissent

comme suit : . La coche indique que le variant est phasé. Les variants en phase avec la partie noire du même côté et appartenant au même "Phase ID" sont en phase, c'est-à-dire sur le même brin.

Vous pouvez cliquer sur le génotype pour obtenir une liste de variants cis ou trans, le cas échéant. Seuls les 5 variants les mieux classés par DiagAI sont présentés, comme indiqué ci-contre.

La liste complète des variants est disponible en cliquant sur "Browse all variants".

Variants in *trans*

Phase set ID 283928

4

SLC25A22 chr11:g.792142C>T

5

SLC25A22 chr11:g.792146_792147del

GT

VOF (Variant Allele Frequency). Cette colonne indique la fréquence de l'allèle variant, c'est-à-dire la proportion de reads soutenant l'allèle alternatif (AO) par rapport au nombre total de reads à cette position (COV), exprimée en pourcentage.

Les analyses utilisant les worksets SomaVar version ≥ 2.5 , SomaCGP version ≥ 0.7 et SomaVar LF version ≥ 2.1 affichent les informations supplémentaires suivantes :

- une infobulle affiche les valeurs de VAF, AO et COV calculées par les 2 *appelants de variants* Freebayes et Mutect2 détectant le variant.
- une icône d'avertissement  alerte l'utilisateur en cas de désaccord entre les deux appelants de variants pour les variants détectés avec une faible VAF. Ainsi, pour les variants affichant une VAF inférieure à 5% dans la colonne VAF, l'avertissement est affiché en cas de divergence de VAF $> 10\%$ entre les VAF calculées par Freebayes et Mutect2.

AO (Alternate Observation). Cette colonne indique le nombre de reads portant le variant.

CAL (Caller). Cette colonne indique le nom du *caller* (ou des *callers*) ayant détecté le variant.

Cov (Coverage). Cette colonne indique le nombre de reads totaux à la position du variant.

SOR. Le Strand Odds Ratio (SOR) est une métrique de biais de brin dérivée du test de l'Odds Ratio Symétrique qui mesure le déséquilibre des lectures supportant le variant entre les brins sens (forward) et anti-sens (reverse). Un SOR plus élevé suggère une probabilité plus élevée d'un artefact, tandis qu'un SOR plus faible suggère un support de brin plus équilibré et est plus cohérent avec un vrai variant.

Trier les variants

Il existe plusieurs options pour trier les variants dans l'onglet Variants. Celles-ci sont disponibles dans un menu déroulant situé dans le coin supérieur droit du tableau. Pour chaque option de tri, il est possible de trier par ordre croissant ou décroissant en cliquant sur l'icône de tri . Par défaut, toutes les options sont triées par ordre décroissant, à l'exception de l'option "Locus", qui est triée par ordre croissant.

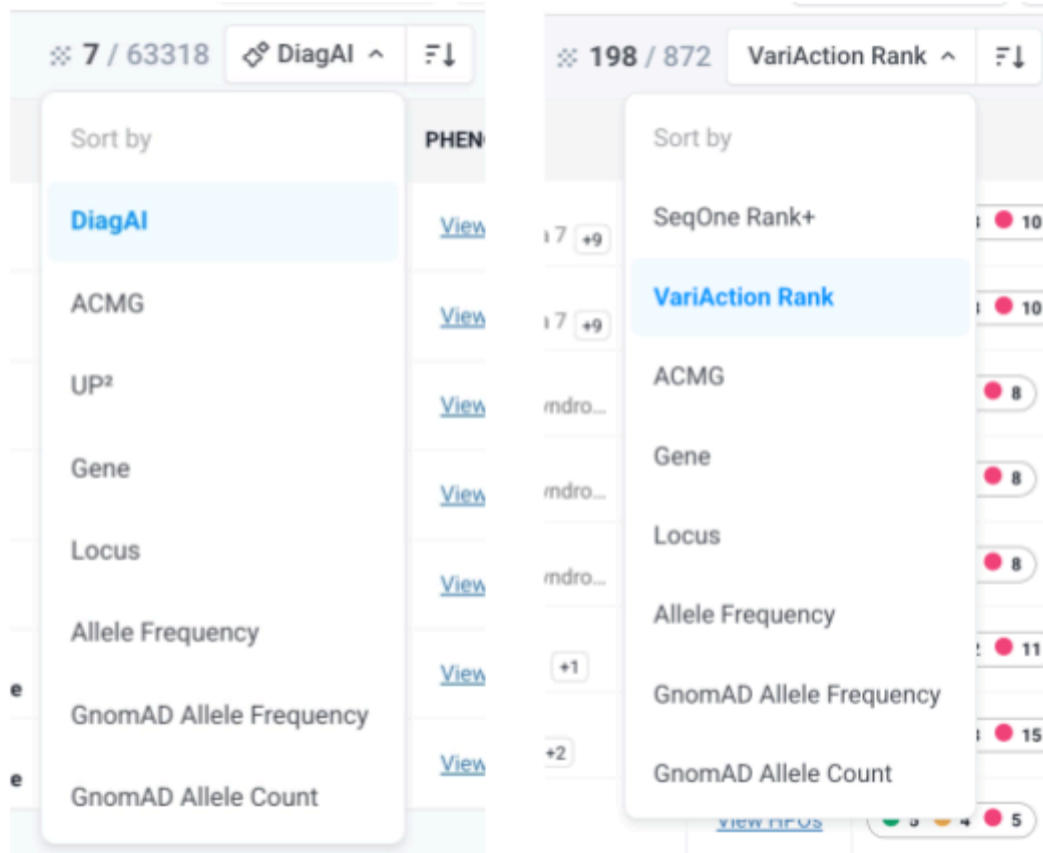


Figure 14 : Fenêtre de sélection de l'option de tri des variants

DiagAI Score. Cette option de tri est appliquée par défaut à toutes les analyses germinales et permet de trier les variants en fonction de leur DiagAI Score (remplace SeqOne Rank+ dans les analyses effectuées après le 2 octobre 2024).

SeqOne Rank+. L'algorithme de hiérarchisation SeqOne Rank+, basé sur un modèle d'intelligence artificielle plus ancien que DiagAI Score, reste disponible et est appliqué par défaut à toutes les analyses réalisées avant le 2 octobre 2024.

Il repose sur quatre éléments principaux :

- Une nouvelle implémentation de la classification ACMG, basée sur l'apprentissage automatique formé sur ClinVar (score UP²)
- Le phénotype du patient (si des termes HPO sont associés à l'échantillon)
- La concordance entre le génotype observé et le mode de transmission de la pathologie décrite dans les bases de données GenCC, OMIM et PanelApp (Australie)
- Qualité des variants

VariAction Rank. L'algorithme VariAction Rank pour la hiérarchisation des variants somatiques est proposé dans les analyses SomaVar à partir de la version v1.8. Ce système utilise l'annotation JaxCKB pour calculer les scores de priorisation des variants en fonction de leur

actionnabilité clinique. Ces scores de priorisation prennent en compte l'impact du variant, sa pathogénicité en utilisant le score ComPerMed, le Tier AMP, la spécificité du variant avec le profil moléculaire, ainsi que la correspondance avec l'indication, si la maladie a été saisie avant le lancement de l'analyse.

ACMG. Les variants peuvent être triés en fonction de leur classification ACMG en sélectionnant cette option. Par défaut, les variants sont triés par ordre décroissant, de la classe pathogène 5 à la classe bénigne 1.

Locus. Les variants peuvent être triés en fonction de leur position génomique en sélectionnant cette option.

Allele frequency. Les variants peuvent être triés en fonction de la fréquence des allèles en sélectionnant cette option. Par défaut, les variants sont triés par ordre décroissant. Il est possible de les trier par ordre croissant en cliquant sur l'icône de tri .

GnomAD Allele Frequency, GnomAD Allele Count. Les variants peuvent être triés par fréquence allélique ("Allele Frequency" (AF)) ou nombre d'allèles ("Allele Count" (AC)) GnomAD, par ordre croissant ou décroissant (si les données GnomAD sont disponibles).

UP². Les variants peuvent être triés par le score de pathogénicité UP², un composant du DiagAI Score.

Gene. Les variants peuvent être triés par ordre alphabétique croissant ou décroissant selon le symbole du gène.

Pour les variants affectant des gènes chevauchants, le symbole du gène utilisé pour le tri est celui associé au variant ayant l'impact le plus élevé, sur la base des règles de priorisation des transcrits décrites ci-dessus. Lors de l'extension de la liste des transcrits, l'utilisateur peut voir le deuxième gène chevauchant, qui peut apparaître non trié. Le tri peut également sembler ne pas être appliqué lorsque plusieurs filtres sont appliqués au tableau des variants.

Configuration des colonnes du tableau des variants

Les colonnes affichées dans le tableau des variants peuvent être configurées à partir du menu



Organise your columns, accessible via le bouton en haut à droite du tableau des variants.

L'affichage des colonnes peut être modifié par l'utilisateur en cliquant sur les icônes  pour activer la ou les colonnes souhaitées dans la section de gauche. La section de droite affiche les colonnes actives dans l'analyse et permet aux utilisateurs de créer et d'enregistrer des profils de

colonnes à partir d'un menu déroulant en cliquant sur l'icône « Select profile » (Sélectionner un profil).

Les colonnes actives peuvent être réorganisées en cliquant et en les faisant glisser à l'endroit souhaité. Par défaut, les colonnes sont organisées selon les profils « Germline » et « Somatic ».

Lorsque les utilisateurs survolent le nom d'un profil de colonnes dans le menu déroulant des pré-réglages, une infobulle indique le workset sur lequel a été créé ce profil. Cela permet d'avertir les utilisateurs que leur profil pourrait avoir des colonnes incompatibles avec le workset actuel.

Pour créer un nouveau profil de colonne, il suffit d'entrer un nom dans le champ « Create a new preset » (Créer un nouveau pré-réglage) et de cliquer sur « Save » (Enregistrer).

Editer les métadonnées échantillons depuis le volet d'analyse

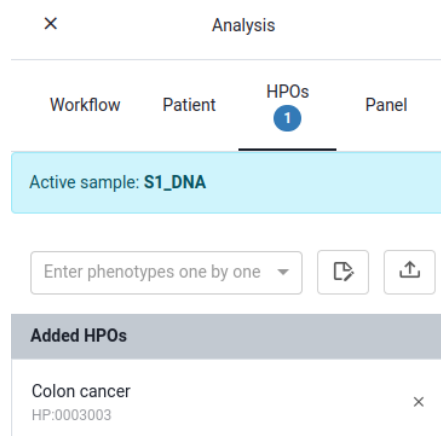
L'utilisateur peut modifier les métadonnées de l'échantillon directement à partir de l'analyse et saisir le sexe du patient ou les HPO (uniquement pour les analyses germinales) à partir du volet d'analyse, accessible en cliquant sur le bouton  Analysis en haut à droite de la page (voir la section **Panneau latéral pour la gestion de l'analyse**).

À partir de la vue Patient, l'utilisateur accède à l'édition du genre (voir la section **Vue modale Patient**).

Dans les analyses de projet de la lignée germinale, l'utilisateur a accès, à partir de la vue des HPO, à plusieurs fonctions d'édition des termes des HPO :

- Visualiser les termes déjà saisis avant de lancer l'analyse
- Ajout manuel de codes HPO ou de phénotypes via la barre de recherche "Saisir les phénotypes un par un".
- Importer un fichier texte contenant les codes HPO
- Génération de termes HPO à partir de la description clinique du patient, en appelant DiagAI

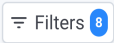
via l'icône . Cette fonction invite l'utilisateur à décrire le profil clinique du patient en anglais (description limitée à 4000 caractères). Ensuite, après avoir cliqué sur "Extract HPOs with DiagAI", une liste d'HPOs surlignés en couleur est générée. Enfin, l'utilisateur peut associer tous les termes HPO proposés au



patient en cliquant sur "Add all", ou les sélectionner individuellement en passant la souris sur les termes de son choix, puis en cliquant sur "Add".

L'ajout et la modification d'OPS à partir de cette vue donnent à l'utilisateur la possibilité de retraiter le score DiagAI et la classification des variants (*Short list* ou *SmartPick*) en fonction des nouveaux HPOs saisis via la fonctionnalité "*Re-process analysis*".

Filtrer des variants

Vous pouvez affiner les recherches dans l'onglet *Variants* en utilisant les catégories de filtres décrites ci-dessous. Le panneau des filtres est accessible via l'icône  située en haut à gauche du tableau des variants ou en appuyant sur la touche "F" du clavier. Les filtres sont disponibles sous l'onglet *Basic*. En haut du panneau des filtres, la section « Active Filters » indique le nombre de filtres actuellement appliqués.

Filtres Genes & Variants

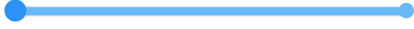
Nom et aperçu du filtre	Description
Genes ⓘ Search gene symbols	Restreint la liste des variants à un ou plusieurs gènes.
Variants ⓘ Search by HGVS	Restreint la liste des variants à un ou plusieurs variants selon la nomenclature HGVS ou HGVS.

Section Samples & Quality

Nom et aperçu du filtre	Description
Coverage ⓘ 1 836	Définit un seuil minimal pour la profondeur de couverture (COV). La valeur maximale affichée par ce filtre correspond à la valeur de couverture la plus élevée identifiée parmi les variants.
Allele Frequency ⓘ 0% 100%	Filtre les variants en fonction de la fréquence de l'allèle mutant dans l'échantillon analysé.
Quality ⓘ Low Medium High	Qualité moyenne (score phred) de chaque base portant l'allèle muté, indiquant la confiance dans la détection du variant : - Moyen ($\geq \text{QUAL}=20$) : 1 chance sur 100 d'appel erroné, - Élevé ($\geq \text{QUAL}=30$) : 1 chance sur 1000 d'appel erroné. Ce filtre s'applique à tous les échantillons.
Quality ⓘ All Low Medium High	Pour les analyses GermVar Tertiary et GermVar Tertiary Family sur les données d'Oxford Nanopore Technologies, la qualité basée sur le score Phred indique la confiance dans la détection des variants. $\geq \text{QUAL}=8$ (faible) : 1 chance sur 8 d'appel erroné, $\geq \text{QUAL}=20$ (moyen) : 1 chance sur 100 d'appel erroné, $\geq \text{QUAL}=30$ (élevé) : 1 chance sur 1000 d'appel erroné. Ce filtre s'applique à tous les échantillons.

▼ Genotype ⓘ ● Homozygous 387 ● Heterozygous 1163 ○ None 1942 ⦿ Biallelic Variant 222 ⦿ Biallelic Variant 84 ⦿ Unknown 76 ⦿ Unknown 0	Filtre les variants en fonction du génotype observé.
▼ Alternate Observations ⓘ 0 0458	Définit un nombre de séquences porteuses de la mutation (observation alternée ou AO). La valeur maximale affichée par ce filtre correspond à la valeur d'AO la plus élevée identifiée parmi les variants.
▼ Strand Odds Ratio ⓘ 0 01234≥5 min 0 — max 5	Filtre les variants selon le score strand odds ratio (SOR) calculé au site du variant.

Section Frequencies

Nom et aperçu du filtre	Description
▼ Sample frequency ⓘ all  all rare	Élimine les variants en fonction de leur récurrence dans le projet étudié, de la plus fréquente (all) à la moins fréquente (rare).
▼ Relative pop frequency ⓘ all 10-2 10-3 10-4 10-5 10-6 absent 	Sélectionne les variants en fonction de la fréquence de leur allèle mutant dans la population combinée des projets ExAc et gnomAD.
▼ Absolute pop frequency ⓘ 0 1609917 	Sélectionne les variants sur la base du nombre d'allèles dans la population combinée ExAc et GnomAD.
▼ Absolute homozygous pop frequency ⓘ 0 803576 	Sélectionne les variants sur la base de leur fréquence d'homozygotie dans la population combinée ExAc et GnomAD.

Section Genetics

Nom et aperçu du filtre	Description
▼ Impact ⓘ Low Moderate High ☒ Keep <i>Modifier</i>	L'impact de la mutation sur la protéine. <i>Modifier</i> : affecte les gènes non codants
▼ Effects ⓘ > Loss of Function 10 • Transcript ablation 10 • Splice acceptor 10 • Splice donor 10 • Stop gained 10 • Frameshift 10 • Stop lost 10 • Start lost 10 • Transcript amplification 10	Les termes SO (Sequence Ontology) qui décrivent les conséquences les plus perturbatrices au niveau du gène. Ce filtre vous permet de sélectionner l'impact de la mutation sur le gène, et les options sont organisées en catégories logiques pour améliorer la navigation : <i>Loss of Function, Coding, Splicing region, Silent variant, Intronic, UTR, ncRNA, Upstream/Downstream, Regulatory, Others.</i>
▼ Gene panel ⓘ Select panel	Sélectionner et appliquer un ou plusieurs panel(s) de gènes. Ce filtre affiche tous les panels de gènes disponibles pour l'entité. Ce filtre n'est pas disponible si un panel restreint a été appliqué à l'analyse à partir de l'onglet <i>Panel</i> du panneau de gestion de l'analyse (voir la section Vue modale Panel). Un message informe l'utilisateur que le filtre "Gene panel" ne peut pas être utilisé avec un panel restreint.
▼ Panel in-silico ⓘ Dyslipidemia_panel_1_Twist★	Sélectionne et applique un panel in-silico. Ce filtre affiche tous les fichiers BED disponibles pour le groupe. Le nom du fichier utilisé lors de l'analyse est mis en évidence par une étoile. Ce filtre ne retient que les variants dont la position de départ (start) est située dans l'un des intervalles du fichier de lit sélectionné.

PanelApp ⓘ Epi ENG Blepharophimosis ptosis and epicanthus ENG Early onset or syndromic epilepsy ENG Epidermodysplasia verruciformis	Permet de filtrer les variants selon une liste de panels in-silico dans les catalogues PanelApp Australie et PanelApp Angleterre. Un badge identifie la base de données d'où provient le panel in-silico : AUS (Australie) ou ENG (Angleterre). Dans les deux catalogues, les superpanels ne sont pas disponibles pour la sélection, de même que les panels qui n'incluent pas au moins un gène. Par exemple, un panel ne contenant que des STR ne pourra pas être sélectionné.
Gene occurrences ⓘ Usable only if they are fewer than 2000 selected variants.	Définit un seuil pour le nombre d'occurrences d'un gène donné dans le tableau des variants.
Chromosomes & MT ⓘ X 25 Y 0 M 0	Conserve les variants situées sur les chromosomes sexuels et/ou le chromosome mitochondrial (M)
RefSeq transcripts ⓘ BARD1_2transcripts.txt	Sélectionne une liste de transcrits RefSeq pour filtrer les variants.

Section Databases

Nom et aperçu du filtre	Description
Variant Databases ⓘ dbsnp 31 cosmic 31 civic 30 lovd 10 clinvar 3	Sélectionne les variants annotés par les bases de données suivantes : dbsnp, cosmic, lovd, clinvar et civic.
Clinvar ⓘ P or LP 3 P/LP low penetrance 0 ⚠ Conflicting (with P/LP subs) 1 ⚠ Conflicting (without P/LP subs) 0 ❓ Uncertain significance 4 B or LB 18 R Risk allele 0 💊 Drug response 2 ☐ Keep others	Permet de filtrer les variants en fonction de leur statut de pathogénicité sur ClinVar. La case 'Keep others' permet de conserver les variants non annotés sur ClinVar.
OMIM ⓘ omim morbid clinic 111 omim without phenotypes 23 omim morbid 6	Permet de filtrer sur OMIM morbid clinic, OMIM without phenotypes et OMIM morbid variants. Ce filtre dispose d'un bouton in/out qui peut être utilisé pour sélectionner ou supprimer un statut.
SeqOne Artifacts ⓘ ☐ Remove artifacts	Bouton permettant de supprimer les variants récurrents présents dans 5 % d'une cohorte de 500 exomes considérés comme des artefacts selon l'algorithme SeqOne.

SEQ-027_IFU_FR_V6

Publié le 2026-04-29

Transmission Mode ⓘ

▼ Dominant 10

AD Autosomal dominant inheritance 10

XLD X-linked dominant inheritance 10

DD XXX 10

PD XXX 10

> Recessive 10

> Other 10

Sources

OMIM 10

GenCC 10

Definitive 10

Supportive 10

Limited 10

Outside known sources 10

Permet aux utilisateurs de filtrer les variants à l'aide de liens de transmission héréditaires provenant de maladies liées à la génétique médicale humaine.

Les variants peuvent être filtrés selon les modes de transmission (dominants, récessifs, autres) et les sources de transmission (OMIM, GenCC ou autres bases de données).

Transmission pattern ⓘ

Compound Het - Suspected 10

Compound Het - Verified 10

Recessive Hom 10

De Novo 10

Permet de sélectionner les variants selon les modes de transmission hétérozygote composite (suspecté ou vérifié), homozygote récessif ou de novo, sur la base de DiagAI.

Veuillez noter que les filtres "Compound Het - Verified", "Recessive Hom" et "De novo" ne sont disponibles que dans les analyses familiales.

Section Classification

Nom et aperçu du filtre	Description
▼ VKB ⓘ New Unknown variant 5 Pathogenic 4 Likely Pathogenic 3 Uncertain Significance 2 Likely Benign 1 Benign Possible artifact ☐ Analysis only ⓘ	Permet de sélectionner des variants inconnus non évalués (New - Unknown variant) ou validés dans la VKB sur toutes les analyses de l'entité. Si la case "Analysis only" est cochée, les variants ne seront recherchés que sur l'analyse en cours d'annotation.
▼ ACMG ⓘ 5 Pathogenic 4 Likely Pathogenic 3 Unknown significance (VUS) 2 Likely Benign 1 Benign 5 1 56 49 6	Classification de la pathogénicité des variants germinales, calculée sur une échelle de 1 (bénigne) à 5 (pathogène).
▼ ComPerMed ⓘ 5 Pathogenic 4 Likely Pathogenic 3 VUS 2 Likely Benign 1 Benign 7 6 102 1 7	Sélectionne les variants somatiques en fonction de leur classe de pathogénicité ComPerMed sur une échelle de 1 (bénin) à 5 (pathogène).
▼ Evidence Type (CKB) ⓘ Actionable Diagnostic Prognostic 1346 0 0	Sélectionne les variants en fonction de 3 types de preuves (CKB) : Actionnable, Diagnostic, Pronostic.

AMP/ASCO/CAP Tier ⓘ I II 1346 0	Sélectionner les variants en fonction du niveau AMP qui leur a été attribué (niveaux I, II et N/A) dans l'analyse du contexte somatique uniquement.
--	---

Section Scores

Nom et aperçu du filtre	Description
CI-SpliceAI ⓘ all > 0.2 > 0.5 > 0.8 Keep unknown	Sélectionner les scores delta. La case "Keep unknown" permet de conserver les variants dont la valeur est "N/A", mais exclut les variants n'ayant aucun impact sur l'épissage (variants "No splicing impact").
Conservation ⓘ Low Conserved Highly Conserved	Degré de conservation du site, calculé à partir de multiples scores de conservation.
Metasvm ⓘ Tolerated Damaging Highly Damaging	Prend en compte le score de 10 applications de prédiction de variants et la fréquence maximale observée dans la population de 1000 génomes.
Universal Pathogenicity Predictor (UP²) ⓘ Benign VUS Pathogenic -1 0 1 Min -1 Max 1	Sélectionne les variants selon le prédicteur universel de pathogénicité (UP²).
LOEUF ⓘ Intolerant Tolerant 0 1 2 Min 0 Max 2 Keep unknown	Sélectionne les variants selon le score LOEUF.

▼ CADD ⓘ BenignPathogenic 1 99 Min 1 ^ ▼ Max 99 ^ ▼ ☒ Keep unknown	Sélectionner les variants selon le score CADD Phred.
▼ REVEL ⓘ BenignPathogenic 0 0.5 1 Min 0 ^ ▼ Max 1 ^ ▼ ☒ Keep unknown	Sélectionner les variants selon le score REVEL.

Sélectionner une indication

Dans SomaVar, SomaVar LF et SomaCGP, il est possible de sélectionner une indication pathologique prédéfinie dans la base de données JaxCKB à partir d'un menu déroulant disponible dans l'onglet *Patient* de la barre latérale de gestion des analyses (voir la section **Vue modale Patient**).

Le système d'ontologie est basé sur la *Disease Ontology*²⁴ et comprend des termes spécifiques à la JaxCKB. Les informations relatives à la sensibilité, dans la sous-colonne "A" de la colonne "Clinical Evidences" (voir la section **Page du Variant Page**), ne sont mises en évidence en vert que lorsque l'indication pour ce variant correspond à l'indication sélectionnée. Si des affirmations de résistance sont trouvées pour un variant, la cellule correspondante de la colonne 'A' sera surlignée en rouge, quelle que soit l'indication. Dans le cas contraire, elle sera grisée.

Si la maladie a été sélectionnée lors du chargement de l'échantillon et avant l'analyse (voir **Métadonnées de l'échantillon**), elle devient l'indication par défaut définie pour l'échantillon et dans l'onglet *Patient* du panneau latéral de gestion de l'analyse. Elle est marquée d'une icône en forme d'étoile et apparaît en premier dans la liste. modification de l'indication dans le panneau latéral de gestion des analyses ne met pas à jour la maladie présélectionnée pour l'échantillon (métadonnées de la maladie).

Analysis

Workflow Patient HPOs Panel

Active sample: CTDNA_toy

Sex

Female Male Unknown

Indication

★ Colorectal cancer (Colorectal cancer) × ▾

²⁴ <https://disease-ontology.org/>

Utilisation d'un profil de filtre prédéfini

Les utilisateurs peuvent sélectionner un profil de filtre prédéfini dans un menu déroulant en cliquant sur l'icône "Select profile". La plateforme SeqOne propose par défaut plusieurs profils de filtre prédéfinis : *GermVar*, *My Selection*, *Research*, *DiagAI*

Modification et enregistrement des profils de filtre

Les utilisateurs peuvent enregistrer la position actuelle des curseurs pour chaque filtre en tant que profil de filtres en donnant un nom au profil et en cliquant sur "Save as new". Le profil de filtre sera alors disponible via le même menu déroulant dans la catégorie *Private*.

Lors de l'enregistrement d'un profil de filtres, un formulaire de création s'ouvre et l'utilisateur peut sélectionner le bouton portant le nom de l'entité pour la rendre accessible aux autres utilisateurs de l'entité. Si l'utilisateur sélectionne le bouton "Private", le profil ne sera accessible qu'à l'utilisateur qui l'a créé.

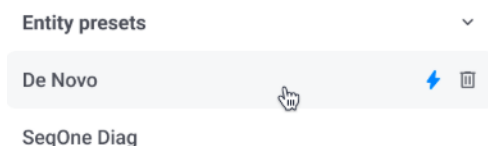
Seul l'utilisateur qui a initialement créé le profil de filtre pourra le modifier en cliquant sur "Update"; si le profil est modifié par un autre utilisateur de l'entité, celui-ci sera invité à le sauvegarder sous un nom différent. Si l'utilisateur n'a pas l'autorisation de modifier un profil de filtre public, il sera invité à contacter l'administrateur de son entité. Les profils de filtre SeqOne ne peuvent pas être modifiés et enregistrés sous un autre profil de filtres.

Un ou plusieurs filtres peuvent être réinitialisés. Il suffit de cliquer sur l'icône  à droite du filtre ou dans le coin supérieur droit de la boîte *Filters*.

Profils de filtres favoris

Les utilisateurs peuvent définir des profils de filtres comme favoris, en cliquant sur l'icône  à droite du nom du profil, de sorte qu'ils apparaissent comme un raccourci directement dans la barre d'accès rapide au-dessus du tableau des variants.

Pour supprimer un profil de filtre des favoris, il suffit de cliquer sur la même icône . Par défaut, le profil *My selection* est défini comme un favori.

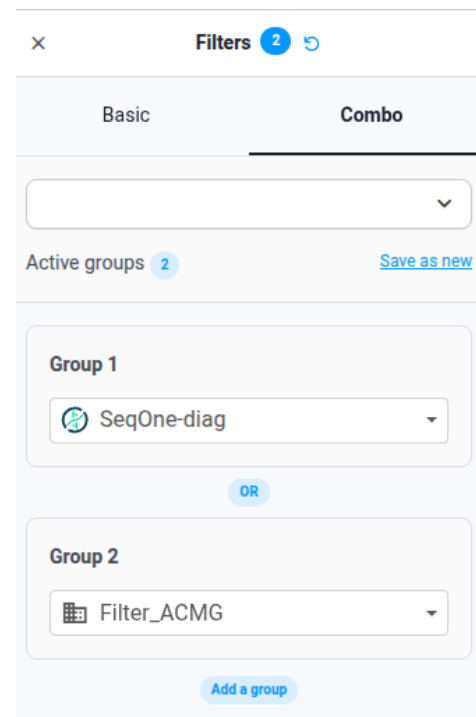


Combinaison de profils de filtres

Les profils de filtre peuvent être combinés sous la condition **OU** en cliquant sur l'onglet Combo dans le panneau des filtres, puis sur "Create a combo". Par défaut, le formulaire de création affiche 2 groupes de profils de filtres, mais les utilisateurs peuvent en ajouter en cliquant sur "Add a group".

Une fois les profils sélectionnés, les utilisateurs sont invités à sauvegarder le profil de leur combo en cliquant sur "Save as new". Comme pour les profils de filtres, l'utilisateur peut rendre sa combinaison accessible à tous les utilisateurs de l'entité dans le formulaire de création. Seul l'utilisateur qui a initialement créé le profil combo pourra le modifier en cliquant sur "Update".

Si le profil est modifié par un autre utilisateur de l'entité, celui-ci sera invité à le sauvegarder sous un autre nom. Si l'utilisateur n'est pas autorisé à modifier le profil d'un combo public, il lui sera demandé de contacter l'administrateur de son entité.



The screenshot shows a 'Filters' modal window with a 'Combo' tab selected. At the top, there's a 'Basic' tab and a 'Combo' tab. Below the tabs, there's a dropdown menu. Underneath, it says 'Active groups 2' with a link 'Save as new'. The main area contains two groups of filters. 'Group 1' has a dropdown menu with 'SeqOne-diag'. Below it is an 'OR' button. 'Group 2' has a dropdown menu with 'Filter_ACMG'. At the bottom, there's an 'Add a group' button.

Profils de filtres SeqOne

My Selection. Ce profil de filtres permet aux variants évalués comme 3 (VUS), 4 (Probablement pathogène) et 5 (Pathogène) d'être affichés directement dans VKB dans l'analyse en cours d'interprétation.

DiagAI. Ce profil de filtres, disponible dans les analyses de lignée germinale (GermlineVar, GermlineFamily, GermVar Tertiary, GermVar Tertiary Family), affiche une liste réduite de variants ayant une forte probabilité d'être la cause de la maladie du patient. Ce profil exploite l'algorithme de hiérarchisation de DiagAI et l'occurrence du variant dans le projet (filtre *Sample Frequency*). Il peut être appliqué lorsque des termes HPO sont ou ne sont pas associés à l'échantillon, mais sa sensibilité est accrue en présence de termes HPO. Dans le cas d'une analyse de famille, le profil du filtre s'adapte automatiquement à la taille de la famille analysée, car il tient compte du fait que l'occurrence des variants dans le projet peut être plus importante. Les utilisateurs peuvent également modifier à tout moment le filtre *Sample Frequency* dans le projet. Saisir les HPO permet d'afficher les variants *SmartPick* ayant la plus forte probabilité de provoquer une maladie.

Family profiles. Dans les analyses GermlineFamily et GermVar Tertiary Family, les profils de filtres *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* ne seront disponibles que

pour les analyses en trio comprenant un cas index affecté, le père non affecté et la mère non affectée. Dans le cas contraire, ces profils seront désactivés.

Il n'est pas recommandé de modifier les profils de filtre *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* lors de la création et de l'enregistrement d'un profil personnel. Il est recommandé de repartir de zéro lors de la création d'une variante de ces profils.

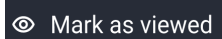
Carrier Screening. Ce profil de filtre n'est disponible dans Germlinefamily que pour les analyses en duo (parents sans échantillon index) et identifie les gènes récessifs avec un variant hétérozygote potentiellement pathogène chez les deux parents "non affectés" pour les tests préconceptionnels.

Somatic FFPE. Ce profil de filtre affiche une liste réduite de variants somatiques dans les analyses somatiques (SomaVar, SomaCGP) effectuées sur de l'ADN FFPE.

Somatic LF. Ce profil de filtre affiche une liste réduite de variants somatiques dans les analyses SomaVar LF.

Fonctionnalité "Mark as viewed"

La fonction "Mark as viewed" permet à l'utilisateur de masquer sélectivement certains variants. Pour ce faire, il suffit de sélectionner le ou les variants souhaités et de cliquer sur l'icône



affichée au bas du tableau des variants. La sélection des variants apparaît alors grisée. Pour supprimer la sélection, l'utilisateur peut soit cliquer sur l'icône



en haut du tableau des variants pour réinitialiser l'affichage, soit sélectionner le ou les variants et cliquer sur .

L'utilisateur peut masquer la fenêtre contextuelle en cliquant sur la croix. Tous les variants seront alors désélectionnés, mais resteront grisés.

Sélectionner des variants

Les variants peuvent être ajoutés au panneau « Selection » à côté des variants évalués (voir la section ci-dessous). Pour ce faire, sélectionnez les variants en cochant les cases situées à gauche du tableau des variants, puis cliquez sur l'icône « Add to selection » en bas du tableau.

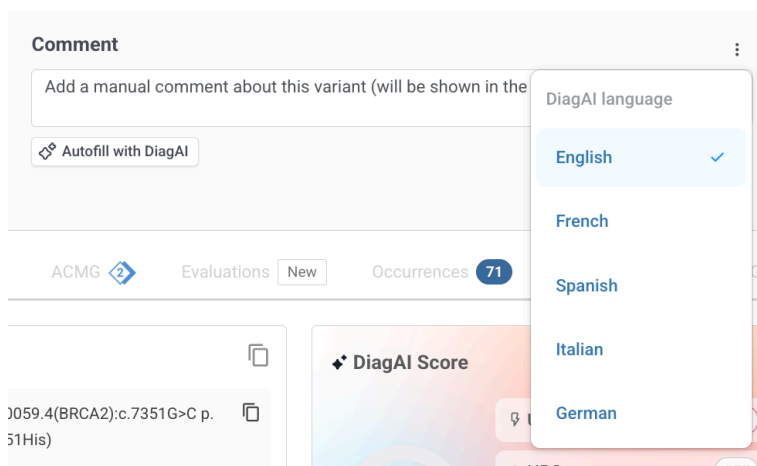
Évaluation des variants

Les variants peuvent être rapidement évalués à partir du tableau des variants en cliquant dans la colonne VKB et en spécifiant leur classe de pathogénicité et un commentaire facultatif. Ces évaluations et commentaires sont accessibles à tous les utilisateurs de la plateforme au sein d'une même entité via l'icône "Selection" dans l'analyse (voir la section **Panneau latéral pour la gestion des évaluations**). Un variant peut également être évalué à partir de sa page dédiée.

Les commentaires peuvent être saisis manuellement ou générés automatiquement. L'interface comprend un champ vide "Classification" pour les commentaires et un bouton "Autofill with DiagAI" pour la génération automatique. Un indicateur apparaît lors de la génération des commentaires.

Pour les commentaires existants, l'option "Improve with DiagAI" permet de les améliorer. Chaque commentaire généré peut être approuvé, modifié ou réinitialisé. Une fois validés, les commentaires sont sauvegardés avec leur classe de pathogénicité, leur date et leur contenu. Ils peuvent être modifiés ou supprimés après validation, tout en conservant leur historique.

Pour sélectionner la langue de DiagAI, cliquez sur les trois points dans le coin supérieur droit de la zone de commentaire et choisissez votre langue préférée dans le menu déroulant.

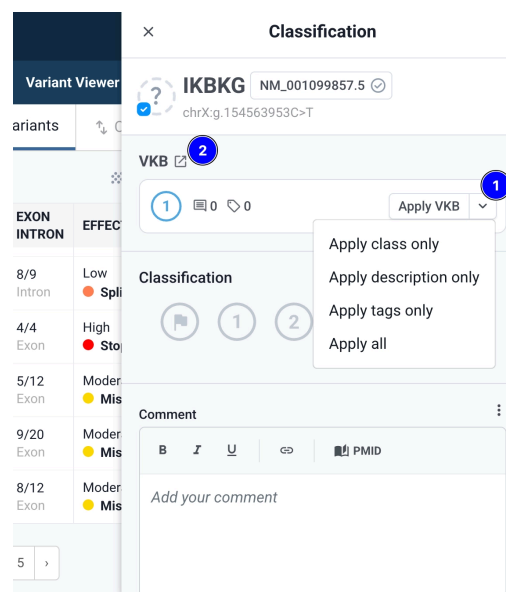


Les variants peuvent également être classés en cliquant dans la colonne de classification, dont l'en-tête comporte un drapeau. Pour classer un variant, cliquez sur l'icône +. Un panneau latéral de classification s'ouvre, permettant à l'utilisateur d'ajouter la classe de pathogénicité, un commentaire et un tag.

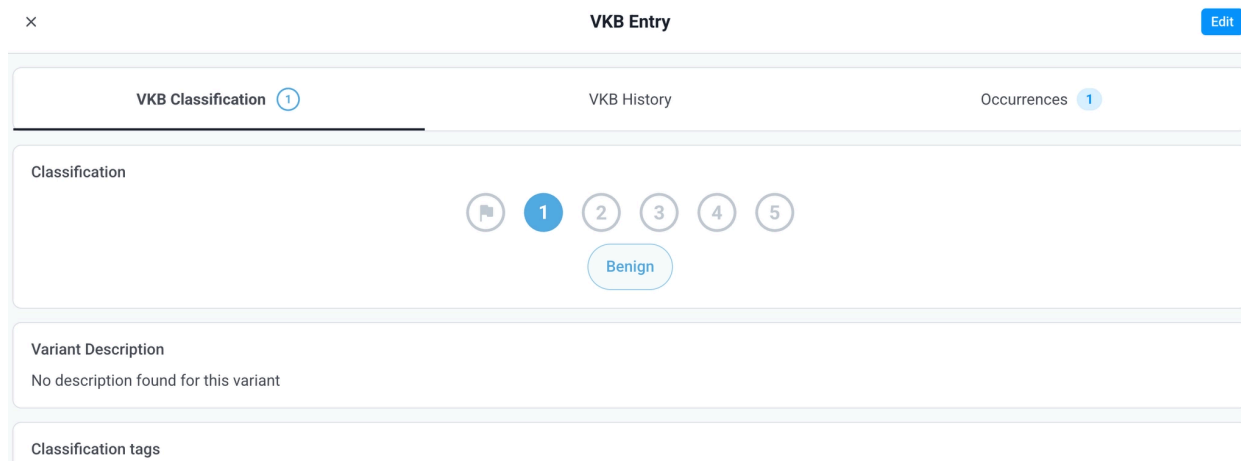
Pour ajouter cette classification à la base de données VKB, activez le commutateur "Add to VKB". Si une classification existe déjà dans la base de données VKB, cliquez sur "Apply VKB" pour l'appliquer à l'entrée actuelle. Des options supplémentaires sont disponibles en cliquant sur la flèche à côté de "Apply VKB" (voir l'étape 1 sur la figure ci-dessous).

Validez la classification en cliquant sur "Save" en haut à droite du panneau latéral.

Pour modifier une classification directement dans la base de données VKB, cliquez sur le lien de redirection VKB (voir l'étape 2 sur la figure ci-dessous) pour ouvrir la fenêtre "VKB Entry".



Dans la fenêtre "VKB Entry", cliquez sur "Edit" en haut à droite pour modifier l'évaluation, puis cliquez sur "Save" pour appliquer les modifications. L'historique VKB et les occurrences du variant sont accessibles en naviguant à travers les onglets dédiés.



VKB Search. Un panneau latéral est disponible en cliquant sur le bouton "VKB" dans la barre de recherche de l'en-tête de SeqOne. Il permet de rechercher des variants annotés dans la base de données VKB par nom de gène, transcription, annotation HGVS ou identifiant SeqOne (Genome_variant_chromosome_location_REF_ALT). La recherche peut être filtrée par génome de référence (GRCh37 ou GRCh38) et/ou par classe de pathogénicité.

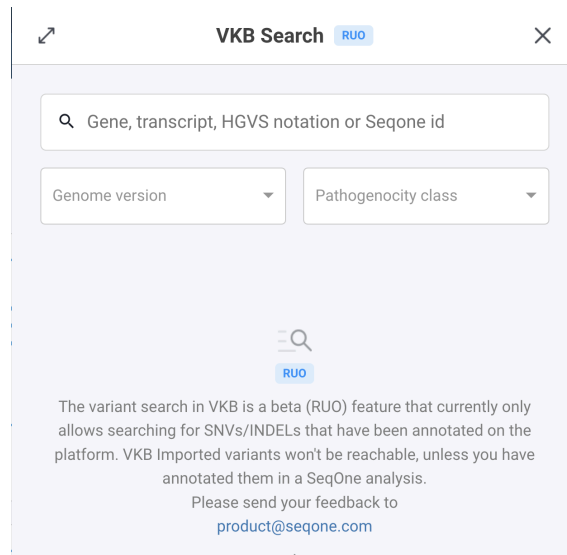


Figure 15 : Fenêtre de recherche VKB

Lorsque vous cliquez sur un variant évalué, une page plus détaillée s'ouvre avec des informations sur le variant, telles que sa position génomique, ses annotations HGVS sur l'ADN complémentaire (ADNc) et sur la protéine, les classes de pathogénicité, les analyses et leurs commentaires associés où il a été évalué, puis un lien vers les analyses.

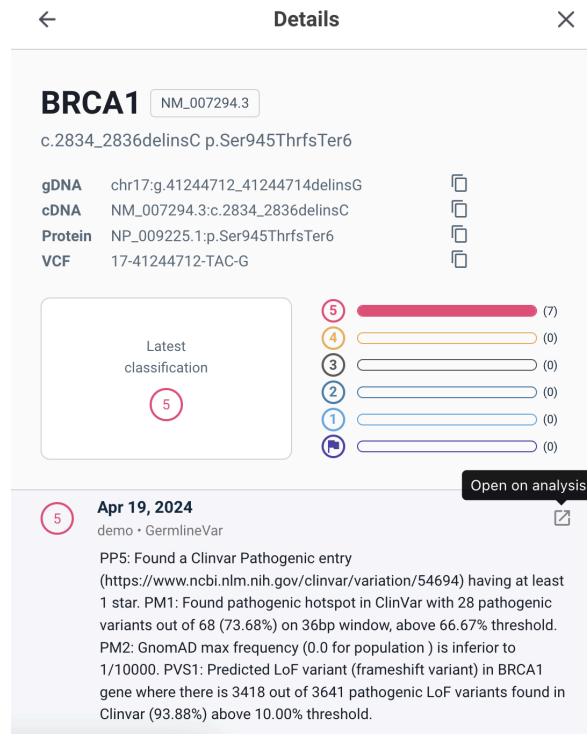


Figure 16 : Fenêtre des détails du variant dans le VKB

Page du variant

En cliquant sur un variant dans le tableau des variants, un panneau latéral contenant plus d'informations s'ouvre, avec plusieurs onglets. Les utilisateurs peuvent ouvrir la page du variant dans un autre onglet en cliquant sur "Open in new window" dans le coin supérieur droit. Le panneau latéral peut être fermé à l'aide de la touche "Echap".

Les utilisateurs peuvent évaluer et commenter un variant dans cette fenêtre. Comme les évaluations et les commentaires faits à partir du tableau principal des variants, ces évaluations et commentaires peuvent être vus par tous les membres du groupe qui utilisent la plateforme.

À partir de cette page, les utilisateurs peuvent également accéder à des informations sur les variants, notamment le génotype (GT), la fréquence allélique (VAF), la profondeur de la couverture (COV) et le score phred (QA).

Les onglets de la page des variants sont décrits ci-dessous.

Variant : l'onglet est divisé en plusieurs blocs, décrits ci-dessous.

- **Nomenclature** : les différentes nomenclatures (g., c., p., entrée VCF) sont répertoriées, avec la possibilité de copier les informations de tout le bloc ou de chaque ligne dans le presse-papiers à l'aide de l'icône .
- **Score DiagAI** : utilise le bloc de résumé du calcul du score DiagAI disponible dans l'onglet Score DiagAI de la page du variant (voir le paragraphe **Score DiagAI**).
- **ClinVar** : informations extraites de la colonne ClinVar Germline du tableau des variants, indiquant la classe de pathogénicité attribuée au variant, son état clinique et le nombre d'entrées dans ClinVar (voir **Visualisation et interprétation** dans la section **Petits variants**). L'icône  ouvre la page ClinVar dans un nouvel onglet.
- **Annotations** : impact sur la transcription, exon concerné, distance par rapport au site d'épissage le plus proche et type de mutation.
- **Transmission** : liens vers les bases de données OMIM et GenCC si une ou plusieurs maladies sont associées au gène.
- **Compound Heterozygous Candidates**, ce bloc est divisé en 2 parties :
 - La section *Short Variants* regroupe tous les SNV candidats situés sur le même gène (sur la base de l'ID du gène) et les classe en fonction de leur score DiagAI. Le variant sélectionné sur la page Variants est affiché en premier et est grisé dans la liste.
 - La section *CNVs* indique si des CNVs ont été trouvées, triées par classe ACMG. Cette section n'est disponible que si une analyse CNV a été effectuée, sinon un message d'avertissement s'affiche : "*CNVs not available*".

En cliquant sur l'un des résultats, un nouvel onglet s'ouvrira avec la page du variant (s'il s'agit d'un petit variant) ou la page du CNV (s'il s'agit d'un variant de copie). Si aucun CNV n'est trouvé, un message d'avertissement s'affiche avec la mention "no CNVs found".

- **Tools** : liens vers d'autres bases de données (LitVar, LOVD, VarSome, dbSNP, Cosmic, GnomAD, CKB, UCSC, Decipher, Ensembl).
- **GnomAD** : si le variant est connu de cette base de données, un tableau résumant la fréquence du variant dans différentes populations est affiché, et un accès direct à GnomAD est fourni via l'icône .

Nomenclature

gDNA chr4(GRCh38)g.154611883G>C

cDNA NM_021870.3(FGG):c.323C>G

Protein NP_068656.2 p.Ala108Gly

VCF 4-154611883-G-C

Clinvar

Likely Pathogenic

Criteria Provided, Multiple Submitters, No Conflicts ★★☆☆

Review status **Reclassified by SeqOne**

Submissions **2 P | 9 LP | 2 VUS**

Disease Congenital Fibrinogen Deficiency, Thrombus

Annotations

Impact/Effect **Moderate** ● missense

Exon/Intron 4/9

Biotype protein coding

Transmission

AR

DISEASE	OMIM	MEDGEN
Hypofibrinogenemia, congenital	AR	N/A

Compound Heterozygous candidates

Short variants

79

4

New

c.323C>G

missen...

7

1

New

c.*16C>T

3 prime ...

1 to 2 of 2 1

CNVs

No CNVs found

Tools

Variant database

Gene database

Genome database

GnomAD

MAF 3.84e-3 ALL 3.03e-3

QC Exome | PASS QC Genome | PASS

Population	AF	AC	HOM
African/American population	6.13e-4	46	0
American population	6.67e-4	40	0
Amish	0.00e+0	0	0
Ashkenazi Jewish	1.52e-3	45	0
East Asian population	0.00e+0	0	0
Finnish population	1.06e-3	67	0
Middle Eastern population	6.61e-4	4	0
Non-Finnish European population	3.84e-3	4526	11
Other populations	1.63e-3	102	1
South Asian population	6.05e-4	55	0
Non-UKB combined population	2.38e-3	1854	5
Combined population	3.03e-3	4885	12

Figure 17 : Onglet Variant sur la page du variant

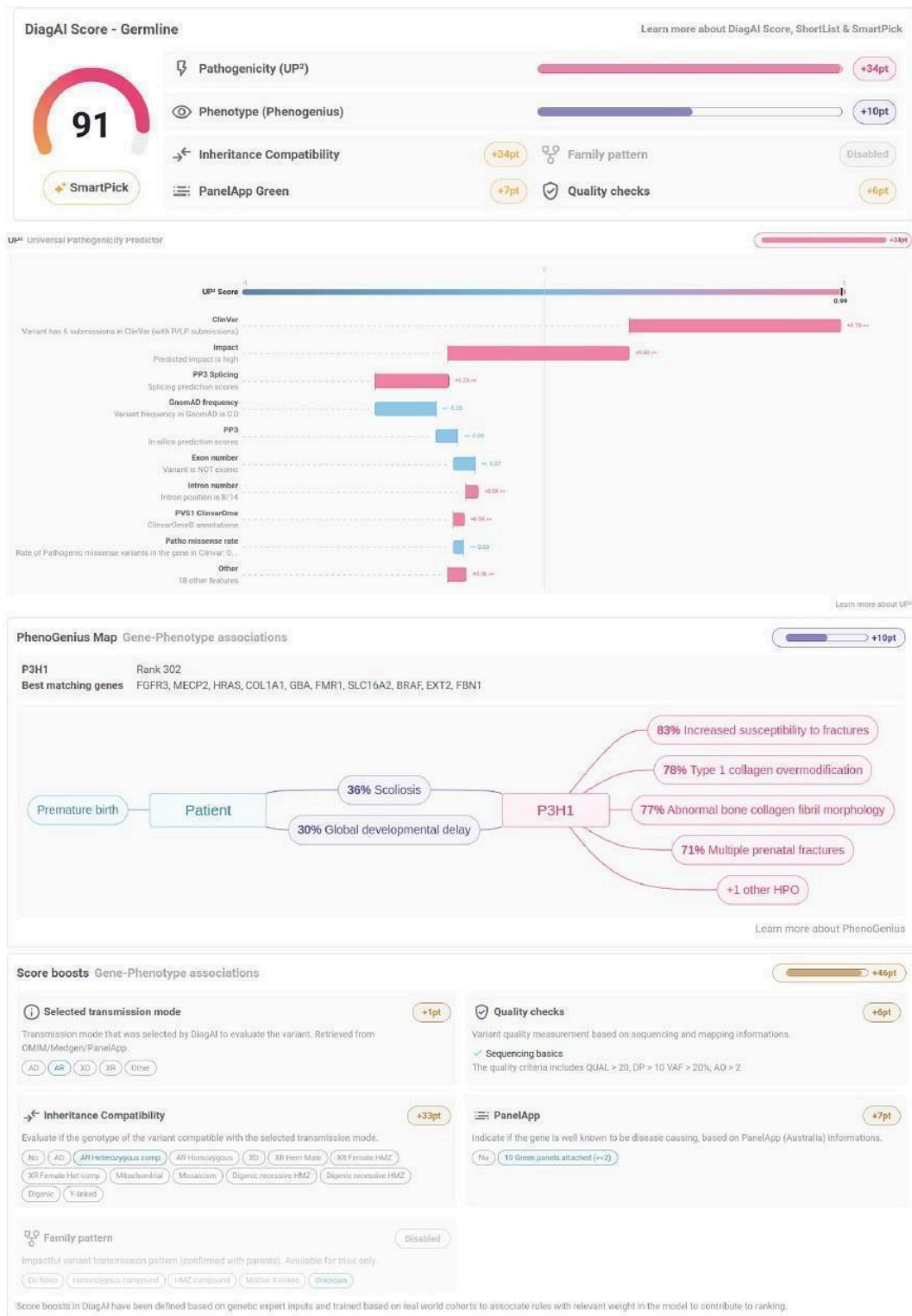


Figure 18 : Visualisation de l'onglet DiagAI sur la page variant

DiagAI. L'onglet contient des explications sur le score DiagAI attribué au variant. Cet onglet n'est disponible que pour les analyses GermlineVar, GermlineFamily, GermVar Tertiary et GermVar Tertiary Family (à partir de la version 1.8).

L'onglet DiagAI est divisé en 4 sections principales :

- Un bloc de résumé du **score DiagAI - Germline** indiquant le score, le statut SmartPick et énumérant les principales caractéristiques contribuant au score global et le nombre de points associé.
- Un graphique en cascade expliquant le score **UP² (Universal Pathogenicity Predictor)**. Un message d'avertissement  s'affiche lorsque l'utilisateur consulte la page d'un variant dont le transcrit diffère de celle utilisée pour calculer l'UP².
- Un bloc d'explicabilité de la carte **PhenoGenius**, coloré si des termes HPOs sont associés à l'échantillon. Le bloc détaille le rang du gène, une liste des 10 gènes les mieux associés, et la carte PhenoGenius, avec un code couleur en fonction de la correspondance entre les HPO du patient et le gène : bleu pour les HPO du patient uniquement, violet pour les HPO du patient et correspondant au gène, rouge pour les HPO du gène uniquement. La section est grisée si aucun terme HPO n'est saisi pour le patient.
- La section **"Score boosts"** intègre 5 sous-caractéristiques qui peuvent ou non contribuer (Disabled) au score : Mode de transmission, Compatibilité avec l'héritage, Modèle de famille, PanelApp et Contrôle de qualité.

ACMG. L'onglet permet aux utilisateurs de visualiser et de revoir les critères utilisés lors de la classification automatisée ACMG. Ils peuvent modifier la sélection en cochant les différents critères requis, puis utiliser la classification pour évaluer le variant en cliquant sur *Use for Evaluation*. Il est également possible de modifier le niveau de preuve (*Supporting*, *Moderate*, *Strong* et *Very Strong*) afin d'attribuer un nombre de points différent à chaque critère, qui sera pris en compte lors du calcul du score ACMG final. Chaque niveau de preuve attribuera : 1 point pour *"Supporting"*, 2 points pour *"Moderate"*, 4 points pour *"Strong"* et 8 points pour *"Very Strong"*. Les points attribués à chaque critère seront ensuite additionnés pour attribuer un statut ACMG à le variant :

- *Pathogenic* : 10 points
- *Likely Pathogenic* : 9 à 6 points
- *Uncertain Significance* : 5 à 0 points
- *Likely Benign* : -6 à -1 points
- *Benign* : à partir de -7 points

Pour le critère BA1, l'utilisateur ne peut pas sélectionner le niveau de preuve, car il s'agit d'une classe autonome.

The screenshot shows the ClinGen variant page for TP53. At the top, the variant is identified as NM_000546.6 c.1083del p.S362AfsTer8, which is a frameshift mutation. The sample information shows a GT/VAF of 0/1.25% and a COV/QA of 319/34. The ACMG tab is active, displaying a predicted pathogenicity score of 'Likely Pathogenic'. Below this, automatic tags are listed: PVS1 (Strong), PM2 (Moderate), and PP5 (Moderate). A detailed description for the PVS1 tag is provided, stating it is a predicted null variant in a gene where LOF is a known mechanism of disease, and it is a non-sense or frameshift variant that is not predicted to undergo NMD, but the truncated region is critical to protein function.

Figure 19 : Visualisation de l'onglet ACMG sur la page variant

TP53. L'onglet est disponible uniquement si le variant est annoté sur le gène TP53. Cet onglet fournit des informations sur les mutations du gène TP53 et leur signification clinique selon la base de données du CIRC²⁵. L'onglet TP53 est divisé en plusieurs blocs, décrits ci-dessous.

Un bloc d'annotations (*Annotations*) affiche les informations suivantes associées au variant :

- Mutation Rate : taux de substitution associé à la position du variant
- Hotspot (yes/no) : présence du variant dans un site de mutation récurrent (hotspot)
- Structural Motif : motif structurel affecté par la mutation
- Domain : domaine fonctionnel de la protéine affectée par la mutation

Le bloc de données fonctionnelles (*Functional Data*) affiche les classifications et les informations fonctionnelles associées à le variant :

- Classe DNE : effet dominant négatif (DNE) sur la transactivation de la protéine p53 de type sauvage
- Classe DNELof : classification fonctionnelle pour la perte de la suppression de la croissance et des activités dominantes négatives.
- Classe de transactivation : classification fonctionnelle basée sur l'activité transcriptionnelle globale (TA).
- Structure Fonction Classe : prédictions fonctionnelles.

²⁵ <https://tp53.cancer.gov/>

Les blocs “Graphics” et “Sample details” fournissent des détails sur les échantillons référencés dans la base de données du CIRC pour le variant. L'utilisateur peut choisir de visualiser les données associées aux échantillons somatiques ou germinaux, en sélectionnant “Somatic” ou “Germline”.

Le bloc “Graphics” affiche deux graphiques, avec des menus déroulants, montrant la répartition des échantillons en fonction de leurs caractéristiques.

Le bloc “Sample details” affiche des informations détaillées sur l'échantillon sous forme de tableau.

ComPerMed (analyse somatique uniquement). La classe de pathogénicité de chaque variant est calculée selon les recommandations du comité d'experts belge ComPerMed (Commission pour la médecine personnalisée). L'utilisateur a accès à des informations détaillées sur cette classification à partir d'un onglet dédié sur la page du variant.

Evidences. L'onglet fournit une liste de thérapies associées au variant sélectionné dans la base de données JaxCKB. L'utilisateur peut faire défiler chaque ligne pour trouver des détails sur les thérapies ou les indications.

BRCA2 protein coding
 NM_000059.4 c.1310_1313del p.K437IfsTer22 frameshift

sample
 GT / VAF 43.7%
 COV / QA 126 29

1 2 3 4 5

Add a manual comment about this variant (will be shown in the output report) ...

Add comment Cancel

Variant ACMG 5 Compermed 6 **Evidences** Evaluations New Occurrences 6 Scores IGV ⋮ Gene

ⓘ The following information is provided for informational purposes only

Actionable 25

Prognostic & Diagnostic 0

Risk factor & emerging

Therapy	Tier	Response	Indication	
Olaparib	IA	Sensitive	Breast cancer	Her2-receptor negative breast cancer ✓FDA +6
Evidence type Actionable Molecular profile BRCA2 inact mut Drugs Olaparib (PARP Inhibitor (Pan)) Description Lynparza (olaparib) is included in guidelines as systemic treatment for patients with recurrent or metastatic breast cancer harboring BRCA1/2 germline mutations (NCCN.org).				
Rucaparib	IA	Sensitive	Ovarian cancer ✓FDA	Prostate cancer ✓FDA +3

Figure 20 : Visualisation de l'onglet *Evidences* dans la page du variant

Evaluations. Cet onglet permet d'accéder à l'historique des évaluations et des interprétations du variant au fil du temps :

- Dans l'analyse actuelle (section gauche de l'onglet)
- Dans les autres analyses effectuées au sein du groupe (partie droite de l'onglet)

Occurrences. L'onglet identifie les autres échantillons contenant le variant, indépendamment de toute évaluation. Pour chaque occurrence, le nom de l'échantillon, la fréquence allélique, le lien vers l'analyse correspondante et le nom du projet sont affichés.

Publications. L'onglet permet d'accéder à la littérature scientifique associée au variant et au gène. Cliquez sur le sous-onglet "Variant" ou "Gene" pour afficher les publications PubMed faisant référence au variant ou au gène. Un lien PMID redirige vers l'enregistrement PubMed, et le lien du texte intégral ouvre l'article complet lorsqu'il est disponible.

Scores. L'onglet est disponible pour les variants présentant un effet *missense*, *start lost* ou *stop gained*. A l'exception de ces cas, le score sera grisé. Cet onglet permet de visualiser les résultats des scores de prédiction de plusieurs outils tels que *REVEL*, *DANN*, *FATHMM*, *FATHMM-MLK*, *LRT*, *MetaLR*, *MetaSVM*, *MutationTaster 2*, *PROVEAN*, *SIFT* et *PLI*. Il donne également accès à des scores de conservation tels que *GERP++*, *PhastCons100way Vertebrate*, *PhastCons20way Mammalian*, *PhyloP100way Vertebrate*, *PhyloP20way Mammalian* et *SiPHhy*.

IGV. L'onglet permet à l'utilisateur de visualiser les reads alignés sur la position du variant grâce à la version web du Genome Browser d'IGV. Dans les zones à forte couverture de read, les reads sont sous-échantillonnés par défaut, c'est-à-dire que tous les reads ne sont pas affichés. Les paramètres de visualisation par défaut sont appliqués, c'est-à-dire un maximum de 100 reads par fenêtre de 50 nucléotides. Ces régions peuvent être identifiées par la ligne noire située sous le profil de couverture (downsampling). En cliquant dessus, on obtient le nombre de reads

retirés de l'affichage (# downsampled). Un menu est disponible à partir de l'icône  , offrant trois options :

- Ouvrir dans un nouvel onglet (*Open in new tab*) : ouvre la vue du variant IGV dans un nouvel onglet.
- Réglage de la piste (*Tracks settings*) : permet d'ajouter à la vue d'autres échantillons provenant du même projet ou de différents projets.
- Ouvrir dans IGV local (*Open local IGV*) : redirige l'application IGV Desktop vers les coordonnées du variant en cours d'interprétation. Pour visualiser les SNV ou d'autres variants, l'échantillon BAM doit d'abord être chargé dans l'application IGV Desktop. Cette option n'est compatible qu'avec les versions ≥ 17 d'IGV Desktop.

IGV prend en charge les vues fractionnées, par exemple en copiant deux ensembles de coordonnées dans la barre de recherche de locus, séparés par un espace. Différents paramètres sont disponibles en cliquant sur les roues dentées à droite du Genome Browser. Par exemple,

pour visualiser directement les reads non alignés (soft-clips), sélectionnez l'option "Show soft clips".

[Gene](#). L'onglet contient des informations supplémentaires sur le gène et la bibliographie associée.

Générer un rapport clinique

L'utilisateur peut éditer un rapport clinique exportable au format PDF et .docx en cliquant sur le bouton "Generate report" (Générer le rapport) depuis l'onglet "Selection" lors de la navigation dans l'analyse.

En haut de la page d'édition, l'utilisateur peut sélectionner un modèle, définir le statut du rapport sur nouveau ou brouillon, sélectionner un échantillon (plusieurs échantillons peuvent être sélectionnés pour les analyses familiales) et sélectionner la langue d'édition. L'utilisateur peut également basculer entre la vue "Edition" et la vue "Preview" (Aperçu) avant d'enregistrer un brouillon ou de passer à l'étape suivante, à l'aide des boutons "Save Draft" et "Next Step".

Les modèles de rapport peuvent être créés dans la section "Reporting" sous "Template list" dans les paramètres de l'entité (Entity settings).

Dans la vue d'édition, l'utilisateur peut modifier les différentes sections du rapport et sélectionner plusieurs options. Dans la vue d'aperçu, l'utilisateur peut revoir le rapport avant de continuer l'édition, d'enregistrer le brouillon ou de valider le rapport.

L'utilisateur peut enregistrer un brouillon et y accéder à nouveau depuis l'onglet "Selection" puis "Generate report", comme lors de la création d'un nouveau rapport. Lorsqu'un rapport PDF est généré, il s'affiche automatiquement dans l'onglet "Reporting" (Rapports) de l'analyse. Ce n'est qu'après validation que le rapport peut être exporté au format .PDF ou .DOCX. L'utilisateur peut supprimer le rapport ou créer des rapports supplémentaires pour la même analyse.

The screenshot shows the 'Reporting' tab of the 'GermVar Family v3.12.33' application. The top navigation bar includes 'Selection' (active) and 'Analysis'. Below the navigation bar, there are several tabs: 'Template' (set to 'None'), 'Report' (set to 'new'), 'Sample', and 'Language' (set to 'English'). On the right, there are buttons for 'Save draft' and 'Next step >'. The main content area is titled 'Edition' and 'Preview'. It contains a 'Header' section with the following fields: 'Report title', 'Patient indication', 'Patient HPO(s)', a 'Refresh HPO from sample' link, and a 'Select metadata to include from the list below (Optional)' dropdown menu. At the bottom, there is a checkbox for 'Include bioinformatic information in the report' with a note: 'When enabled, reports will include sample ID, analysis name, date, user, and test information'.

Figure 21 : Affichage de l'onglet du rapport clinique.

13.1.5. CNV

13.1.5.1. CNV dans les panels de gènes

Methodologie

La détection des CNV dans les données de panels de gènes, dans GermlineVar, GermVar/GermVar Family, SomaVar et SomaHemato, est basée sur la sélection d'échantillons contrôles dans la cohorte co-analysée, en fonction de leur corrélation.

La corrélation est calculée à partir d'une table de couverture, elle-même calculée à partir de toutes les régions du manifeste, à l'exclusion des chromosomes sexuels.

Les critères d'éligibilité d'un échantillon pour l'analyse CNV sont les suivants :

- Pour qu'une région soit retenue dans l'analyse CNV, la couverture minimale est de 50 reads.
- La valeur seuil de corrélation pour la sélection des échantillons de contrôle est de 0,90 pour le contexte germlinal et de 0,75 pour le contexte somatique.
- Le nombre minimum de contrôles requis pour chaque échantillon est de six.

Le sexe des échantillons (métadonnées de l'échantillon) doit être défini avant de lancer CNVCapture ou SomaCNVCapture afin d'effectuer l'appel sur les gènes localisés sur les chromosomes sexuels. Enfin, au moins huit échantillons sont nécessaires pour effectuer une analyse CNVCapture ou SomaCNVCapture sur SeqOne.

Les critères appliqués au manifeste de l'utilisateur pour qu'un gène du panel soit éligible à l'analyse CNV sont les suivants :

- Les régions de manifeste sont annotées à partir du transcrit canonique déterminé selon la ressource MANE Select du NCBI. Si aucun transcrit MANE Select n'est disponible, l'annotation est basée sur la ressource RefSeq du NCBI.
- Seuls les exons pour lesquels il existe un chevauchement ≥ 80 % entre l'intervalle manifeste et l'exon sont analysés. Si aucun exon de gène ne chevauche ≥ 80 % avec l'intervalle manifeste, le gène ne sera pas inclus dans l'analyse CNV.
- Seuls les introns pour lesquels il existe un chevauchement ≥ 60 % entre l'intervalle manifeste et l'intron sont analysés.

L'analyse CNVCapture (version ≥ 2.4) inclut les régions situées 20kb en amont et en aval des gènes couverts par le manifeste, si ces régions sont incluses dans la conception du manifeste. Plus précisément, chacun des 20 segments de 1kb constituant la région de 20kb est analysé si au moins 20% du segment chevauche le manifeste initial.

Visualisation et interprétation

L'onglet CNVs affiche un graphique représentant la liste des gènes dans lesquels des variations de copie ont été identifiées. Cet onglet est disponible dans les worksets décrits dans la section Méthodologie ci-dessus.

La terminologie utilisée pour décrire le statut des gènes appelés est la suivante :

- Gain : le gène est affecté par une amplification partielle ou totale
- Loss : le gène est porteur d'une délétion partielle ou totale
- Normal : le gène ne porte pas de CNV
- Partially not called : le gène est partiellement appelé
- Not called : le gène est exclu de l'analyse CNV

Le statut partiellement non appelé est attribué à un gène lorsque le statut de sa copie est normal mais qu'au moins un exon n'a pas été appelé (avec le statut "failed").

Malgré son éligibilité à l'analyse des CNV, un gène est exclu (Not called) de l'analyse si :

- le gène se trouve sur un chromosome sexuel et les métadonnées relatives au sexe n'ont pas été remplies,
- tous les exons échouent selon les critères spécifiés ci-dessous : nombre de contrôles, pourcentage de variation. Le statut "not called" d'un gène est défini sur la CDS uniquement. Les régions introniques et les régions en amont/en aval sont exclues si les exons CDS sont appelés.

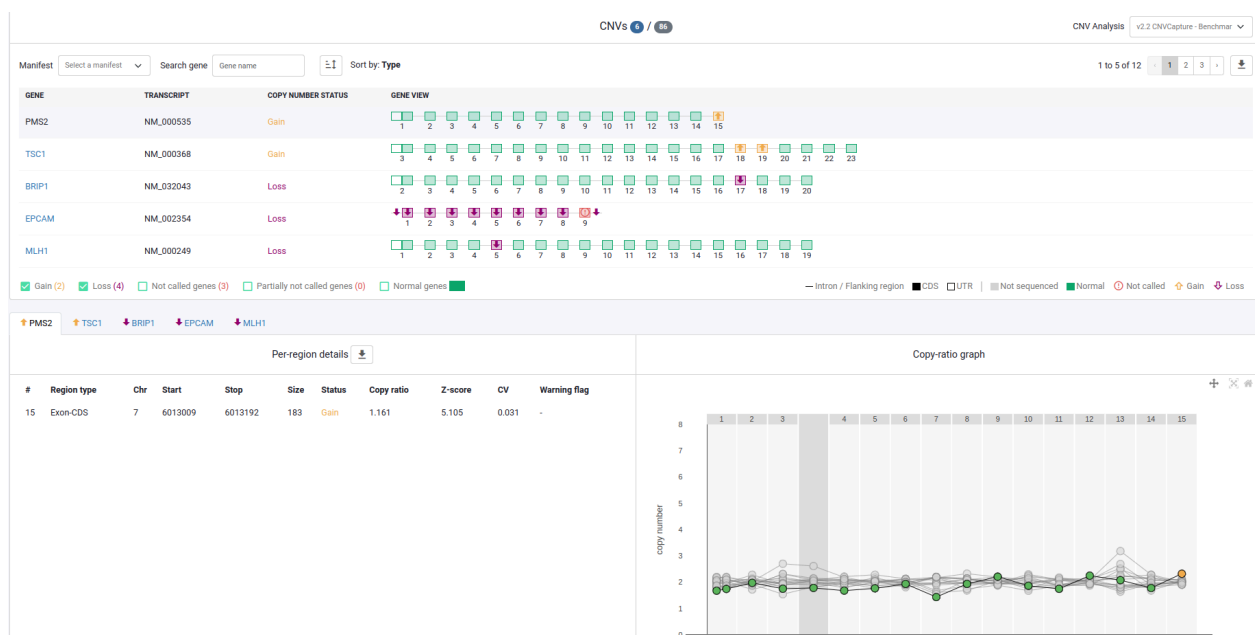


Figure 22 : Visualisation de l'onglet CNVs dans une analyse GermlineVar.

Filtres et options de tri

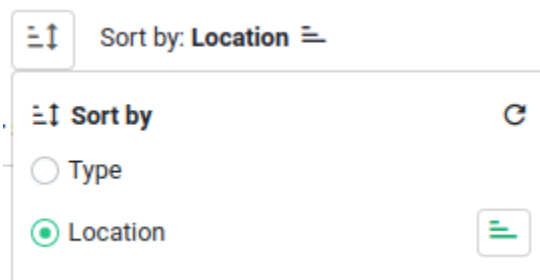
Le menu de filtres peut être utilisé pour :

- Restreindre l'affichage à un sous-ensemble de régions/gènes à l'aide d'un fichier BED qui a été préalablement téléchargé dans l'interface avec la fonction *Manifest*.
- Recherche d'un gène spécifique grâce au filtre *Gene*.

Deux options de tri des CNV sont disponibles :

- **Type** (par défaut) : logique de tri du courant (Gain puis Loss)
- **Location** : permet de trier les CNV en fonction des chromosomes et de leur position de départ, par ordre croissant ou décroissant, en

cliquant sur l'icône de tri. 



L'utilisateur peut également sélectionner les gènes en fonction de leur statut : Gain, Perte, Gènes normaux, Gènes partiellement non appelés ou Gènes non appelés.

☒ Gain (0) ☒ Loss (0) ☐ Not called genes (69) ☐ Partially not called genes (0) ☐ Normal genes 

Détails par région

Le tableau des détails par région présente les détails de chaque région pour chaque gène :

- **#** : numéro de la région (exons et introns)
- **Région type** : type de région (amont, utr-exonique, codant-exonique, intronique, aval)
- **Chr** : chromosome
- **Start/Stop** : position de début et de fin de la CNV
- **Size** : taille de la CNV (pb)
- **Status** : statut de la région (gain = duplication, perte = délétion, normal = pas de CNV, non appelé = la région n'a pas été incluse dans l'analyse CNV).
- **Copy-ratio** : rapport du nombre de copies par rapport à la normale (n=2).
- **Z-score** : nombre d'écarts types séparant un résultat de la moyenne. Des valeurs positives ou négatives élevées indiquent une grande confiance dans l'appel CNV.
- **CV** : coefficient de variation mesurant la dispersion relative au sein des échantillons de la cohorte.
- **Warning flag** : Un message d'erreur peut parfois s'afficher en fonction des critères suivants :
 - **Low zscore** : Valeur absolue du Z-score < 4
 - **High variation** : coefficient de variation (CV) > 0,1
 - **Low coverage** : couverture de la région pour cet échantillon < 100

- **Number of controls** : Le nombre de contrôles est insuffisant pour pouvoir réaliser le calling sur cette région.
- **Percentage of variation** : La variation est trop importante entre les échantillons pour cette région pour pouvoir réaliser le calling
- **Read depth** : couverture insuffisante de la région (< 50) et ratio de copies > 0,1
- **Missing gender** : la région est située sur un chromosome sexuel et les métadonnées relatives au sexe n'ont pas été renseignées.

Attention : une région riche en soft-clips, telle qu'une grande indel ou un insert Alu, peut être interprétée comme une baisse de couverture par les worksets CNVCapture et SomaCNVCapture et détectée comme une perte de copie, si la taille de l'insertion est du même ordre que celle de la région appelée.

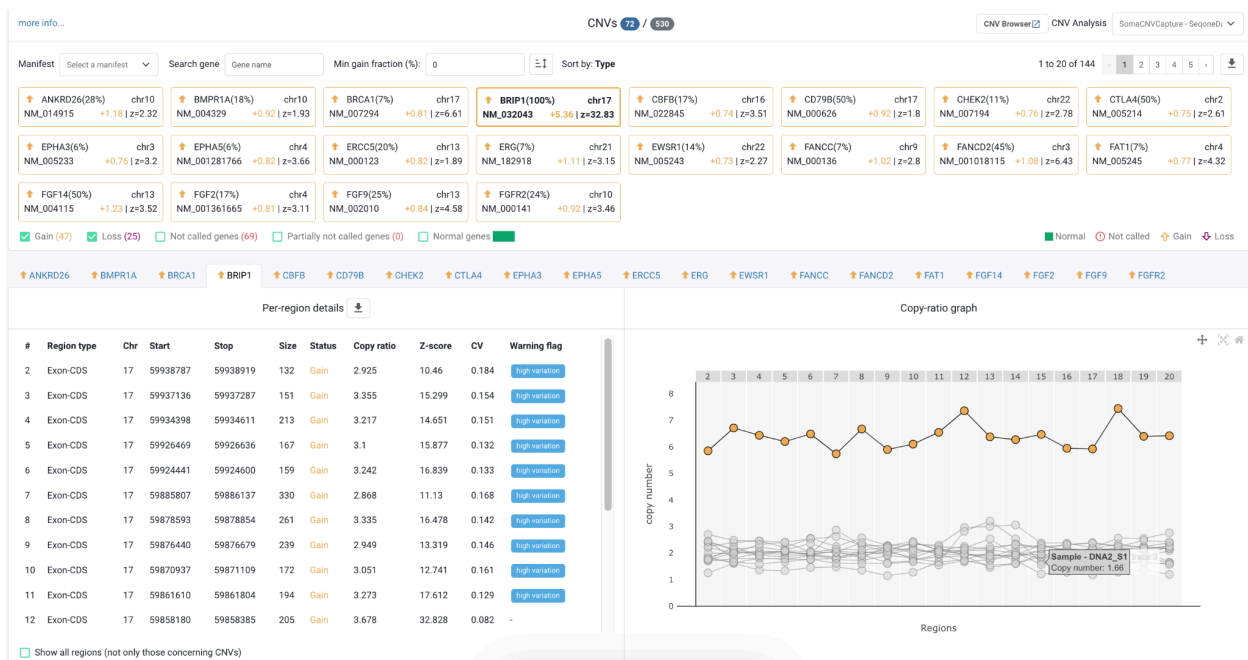
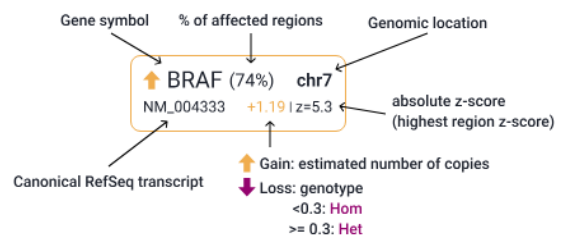


Figure 23 : Visualisation de l'onglet CNVs dans une analyse SomaVar

Dans le cas des analyses **SomaCNVCapture**, bien que la méthode de détection soit similaire à celle de CNVCapture, l'échelle utilisée pour la représentation graphique des variants est le gène, plutôt que l'exon. La représentation se présente sous la forme d'une série de boîtes. Elle fournit le symbole du gène, la position génomique, le transcrit canonique Refseq, le score z absolu et, en cas de gain, le nombre de copies ou, en cas de perte, le génotype.



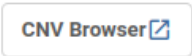
Une fenêtre contextuelle expliquant ces cases est disponible au-dessus des filtres dans la section "**Plus d'infos...**".

Le pourcentage affiché dans chaque case reflète la fraction du gène (*min gain fraction*) affectée par la CNV détectée. Le chunk ratio peut être utilisé pour filtrer les résultats. Par défaut, le filtre est fixé à 0 % et ne s'applique qu'aux gains de copies.

Les délétions détectées par ce pipeline ne sont pas soumises à ce filtre et doivent être interprétées avec prudence. En effet, la capacité de cet outil à détecter les variations d'une copie de gène dépend de la nature et de la qualité de l'échantillon, en particulier de la fraction tumorale.

Attention : Si une amplification et une délétion sont détectées sur le même gène, le gène sera considéré comme amplifié par défaut, et donc soumis au filtre "min gain fraction" décrit ci-dessus.

CNV Browser

Une icône  est disponible dans l'onglet d'affichage CNVs des analyses GermlineVar, GermlineFamily, SomaVar ou SomaHemato pour ouvrir un CNV Browser dans un nouvel onglet. Le graphique CNV Browser affiche la fréquence des allèles B (BAF) des SNP (Single Nucleotide Polymorphism) présents dans l'échantillon analysé et présents dans un ensemble de données de SNP arrays d'Illumina et de SNP d'autres backbones tels que OneSeq d'Agilent, afin de faciliter la visualisation de certains types d'événements tels que la LoH (Loss of Heterozygosity).

Le BAF est la fréquence de l'allèle de non-référence (allèle B) observée pour un SNP hétérozygote germlinal.

Par défaut, le graphique représente le BAF des variants (axe des y) sur l'ensemble du génome (axe des x), ventilé par chromosome (1-22, X et Y). Un filtre permet à l'utilisateur d'afficher les variants sur un chromosome donné dans le coin supérieur gauche.

Dans le contexte des analyses CNV, ces informations BAF peuvent être utilisées pour mettre en évidence une altération du nombre de copies spécifique à un allèle, une perte d'hétérozygotie ou un déséquilibre allélique.

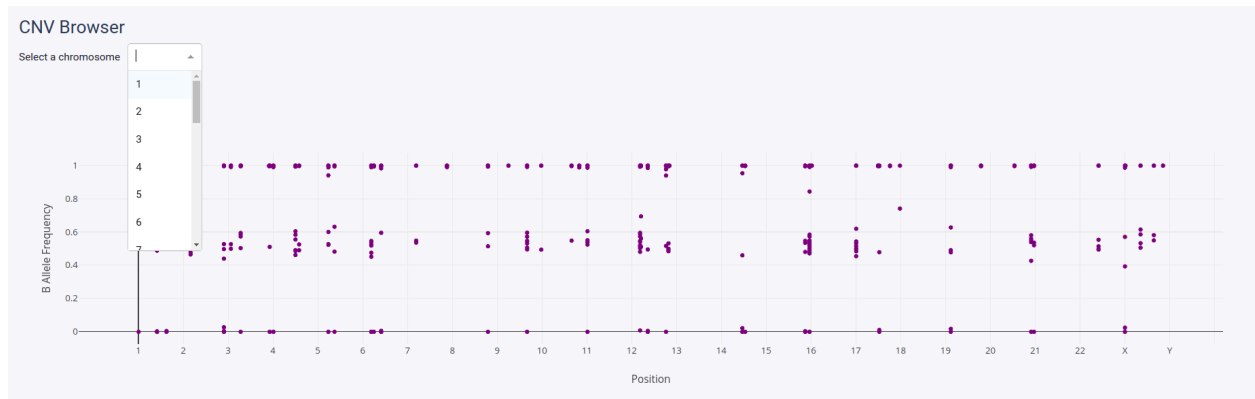


Figure 24 : Visualisation de l'onglet CNV Browser dans une analyse GermlineVar.

Exportation des données CNV

L'icône  en haut à droite permet d'exporter les données de l'analyse CNV sous la forme d'un fichier .tsv contenant les colonnes suivantes :

- gène (nom du gène),
- transcrit RefSeq (refseq_id),
- le statut du gène (statut),
- chromosome (chr),
- coordonnées génomiques de la CNV : positions de départ et d'arrivée,
- fraction du gène affectée par la CNV (gene_fraction_cnv),
- rapport du nombre de copies (copy_ratio) par rapport à la normale (n=2),
- nombre d'écarts types séparant un résultat de la moyenne (zscore)
- nombre d'exons chevauchant la CNV (exons), dans le cas d'une analyse CNVCapture.

Pour chaque gène, l'icône  à côté de "Per-region details" peut être utilisée pour exporter spécifiquement les données d'analyse CNV pour le gène sélectionné sous la forme d'un fichier .tsv contenant les champs suivants :

- gène (nom du gène),
- chromosome (chr),
- coordonnées génomiques : positions de départ et d'arrivée,
- le numéro de l'exon ou de l'intron (exon_intron_number),
- statut,
- le type (exon ou intron, type),
- les messages d'erreur associés (warning_flags, fail_flags),
- valeur de probabilité critique ajustée (padj),
- la valeur critique de probabilité (valeur p),
- rapport entre le nombre de copies et la normale (ratio),
- mesure de la dispersion relative au sein des échantillons (variabilité),
- écart-type du résultat par rapport à la moyenne (zscore)

13.1.5.2. CNV dans l'analyse SomaCGP et SomaLBx

Methodologie

SomaCGP et SomaLBx utilisent PureCN²⁶ pour détecter les CNV et les LOH (perte d'hétérozygotie). PureCN identifie les variants somatiques à partir des données de détection des variants afin d'évaluer la ploïdie et la pureté tumorale de l'échantillon. Ces informations sont utilisées pour déterminer le nombre de copies de chaque événement.

Visualisation et interprétation

L'onglet de visualisation des CNV dans SomaVar Tertiary analysis fonctionne de manière identique à celui de SomaCGP et SomaLBx, comme décrit ci-dessous.

Les CNV sont présentées dans un tableau filtrable. Chaque CNV affiche les informations suivantes :

- **Gene** : nom du gène
- **Location** : cytobande
- **Event coordinates / Event size** : coordonnées et taille de l'événement
- **Transcript**
- **Exons** : Exons affectés
- **Type** : type d'événement (LoH, Loss ou Gain)
- **Copy number** : nombre absolu de copies
- **Function** : fonction du gène le cas échéant, oncogène ou gène suppresseur de tumeur (TSG)

L'utilisateur peut rechercher les CNV affectant un gène donné à l'aide d'un champ de recherche et choisir d'afficher ou non les gènes présents dans les régions LOH en cochant la case "*Show genes in LOH regions*". Le tableau des résultats peut être exporté en utilisant l'icône de téléchargement dans le coin supérieur droit.

 CNV Browser

L'icône ouvre un *CNV Browser* dans un nouvel onglet. Cet onglet affiche un graphique montrant la fréquence des allèles B (BAF) des SNP dans l'échantillon analysé et des SNP LoH provenant de publications sélectionnées et du backbone OneSeq Agilent. Un second graphique représente les données logRatio générées par PureCN. Plus de détails sont disponibles dans la section **CNV : Gene Panels** sous **CNV browser**.

²⁶ <https://github.com/lima1/PureCN>

13.1.5.3. CNV dans les exomes et les génomes

L'onglet de visualisation CNV pour les données d'exome est disponible dans les analyses GermlineVar, GermlineFamily, GermVar et GermVarFamily lors de l'utilisation de fichiers FASTQ comme entrée principale et de l'exécution d'une analyse CNVExome. Cet onglet affiche également les résultats CNV pour les données génomiques dans les analyses GermVar et GermVar Family utilisant des fichiers FASTQ, ainsi que dans les analyses GermVar Tertiary et GermVar Tertiary Family utilisant des fichiers VCF ou des données de microréseau CGH.

Méthodologie pour l'exome

Cette méthodologie s'applique spécifiquement à l'outil CNVExome. Tout comme les pipelines d'analyse CNV sur les données de panels de gènes, le processus d'identification des variants dépend dans ce cas d'une analyse de la profondeur de couverture. Cependant, l'analyse CNVExome à l'aide de l'outil [GATK GermlineCNVCaller](#) est individuelle, puisque chaque échantillon est comparé à un modèle HHMM (*Hierarchical Hidden Markov Model*) calculé par SeqOne.

Pour tenir compte des variations techniques liées à la préparation des échantillons, ce modèle est basé sur une large cohorte allant d'une à plusieurs centaines d'échantillons séquencés par le laboratoire qui les utilise :

- Le même protocole de préparation des bibliothèques
- Le même kit de capture
- La même plateforme de séquençage et la même longueur de read

Ces exigences et le nombre variable d'échantillons nécessaires à la mise en place du modèle s'expliquent par la nécessité de disposer d'une cohorte offrant un niveau de couverture homogène. Plusieurs facteurs doivent être pris en compte, tels que la profondeur de séquençage, le degré de variabilité au sein de la cohorte et la reproductibilité des protocoles de préparation des échantillons.

Les métadonnées relatives au sexe (voir **Métadonnées de l'échantillon**) doivent être remplies avant de lancer l'analyse CNVExome afin de permettre la vérification de la ploïdie des chromosomes sexuels. Si elles ne sont pas remplies, un message d'avertissement "Ploidy" s'affiche sur l'interface d'affichage CNV.

Limitation: la documentation de GATK GermlineCNVCaller recommande d'exclure les régions PAR (PseudoAutosomal Regions), c'est-à-dire les régions PAR1 et PAR2, si leur contribution à la couverture totale du chromosome X est disproportionnée. En fait, la ploïdie est difficile à calculer pour les extrémités du chromosome X correspondant aux régions PAR pour de nombreux modèles calculés par SeqOne. Par conséquent, la ploïdie n'est pas calculée

correctement pour ces régions, qui sont détectées de manière artéfactuelle à 1 copie, alors que l'ensemble du chromosome X est détecté à 2 copies.

Les utilisateurs sont invités à contacter l'équipe d'assistance aux utilisateurs à l'adresse support@seqone.com pour en savoir plus sur la procédure de mise en place de la CNVExome dans leur groupe.

Méthodologie pour le génome

Pour les versions jusqu'à la v3.6, les analyses GermVar et GermVar Family sur les données génomiques intègrent l'outil Sentieon CNVScope²⁷ pour la détection des CNV. À partir de la version v3.7, les worksets GermVar et GermVar Family utilisent l'outil Cue²⁸ pour l'appel des CNV sur les données génomiques.

Les données de duplication segmentaire sont détectées à l'aide de Sentieon's SegDup Caller, qui est spécifiquement conçu pour identifier les événements de nombre de copies dans des régions génomiques hautement homologues sur les données WGS uniquement en utilisant GermVar. Cet outil d'appel cible un ensemble soigneusement sélectionné de gènes cliniquement pertinents, connus pour être complexes pour les approches standard de détection de variants. Les événements résultants sont automatiquement intégrés dans le pipeline d'analyse et visualisés dans le visualiseur de CNVs & SVs en tant que CNV (à la fois gain et perte), permettant l'identification de variants dans des régions génomiques qui étaient auparavant inaccessibles à l'analyse standard.

Visualisation et interprétation

Une interface d'interprétation est disponible dans l'onglet **CNV** pour les analyses CNVExome, GermVar et GermVarFamily, GermVar Tertiary ou GermVar Tertiary Family. L'interface est divisée en deux sections : un tableau de résultats CNV sur la gauche et une page CNV détaillée avec plusieurs onglets sur la droite.

Tableau des résultats des CNV

Un tableau de résultats affiche les CNV identifiées, ainsi qu'un menu de filtres basés sur leur présence dans les gènes OMIM morbides [<https://omim.org/>], leur taille, la fréquence d'apparition des CNV dans la cohorte, ou le LogRatio dans le cas de l'analyse MicroArray CGH.

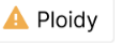
²⁷ <https://support.sentieon.com/appnotes/cnvscope/>

²⁸ <https://github.com/PopicLab/cue>

Le nombre de CNV filtrées (CNV sélectionnées) et le nombre total de CNV détectées (éléments au total) sont affichés en haut du tableau CNV.

Une icône de cadenas indique que l'analyse utilise un panneau restreint, et les détails des CNVs apparaissent dans les infobulles associées.

0 in locked panel
392 elements in total
CNVs 0 / 0

Un message d'avertissement  s'affiche lorsqu'une valeur de ploïdie anormale est détectée et/ou lorsque les métadonnées relatives au sexe (voir **Métadonnées de l'échantillon**) n'ont pas été remplies avant le lancement de l'analyse.

En cliquant sur l'icône "Ploïdie", l'utilisateur peut visualiser la ploïdie de chaque chromosome et le sexe du patient. Les valeurs anormales sont surlignées en jaune.

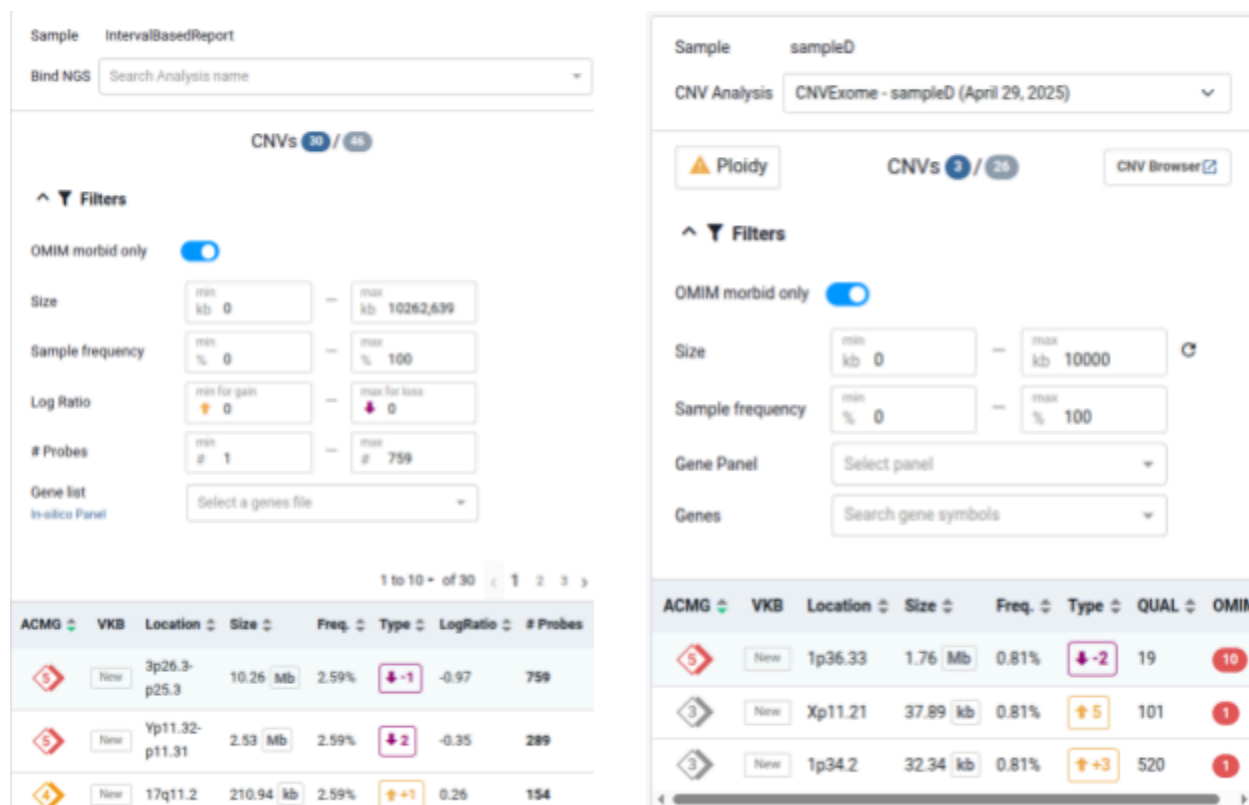


Figure 25 : Vue d'ensemble des CNV détectées et menu de filtres d'une analyse CGH (à gauche) et d'une analyse standard (à droite).

Dans le cas d'une analyse CGH MicroArray combinée à une matrice SNP (CGH+SNP), les événements de perte d'hétérozygotie (LOH) détectés sont rapportés dans le tableau des résultats CNV. Ces événements correspondent à une perte d'hétérozygotie sans effet sur la

copie (CN-LOH). En d'autres termes, la région conserve deux copies, mais identiques, caractérisées par la présence de SNP homozygotes.

Dans le cas d'une analyse CGH MicroArray, il est possible de sélectionner une analyse SNV non-CGH (GermlineVar, GermlineFamily, GermVar Tertiary et GermVar Tertiary Family) dans le menu déroulant "Bind NGS" afin d'afficher les variants qui se chevauchent avec la CNV détectée.

OMIM morbid only. Ce filtre est activé par défaut pour identifier facilement les CNVs affectant des gènes cliniques, annotés OMIM Morbid sur la base de leur identifiant NCBI (NCBI gene_id) ou, à défaut, sur la base du nom du gène et de ses synonymes (gene_symbol et gene_synonyms). Il suffit de lever ce filtre pour afficher toutes les CNVs détectées dans l'échantillon.

Size. L'utilisateur peut filtrer les CNV en fonction de leur taille (en kb).

Sample frequency. L'utilisateur peut filtrer les CNV en fonction de leur taille et de leur fréquence (Sample frequency). Cette fréquence correspond à l'occurrence de la CNV parmi tous les échantillons analysés jusqu'à la date d'analyse et comparés au même modèle, y compris les échantillons utilisés pour construire le modèle. La fréquence de l'échantillon est basée sur un chevauchement de 95% et considère le même type de CNV (gain/perte/délétion). En passant la souris sur sa valeur affichée en pourcentage, l'utilisateur verra apparaître la fraction correspondante. La mise à jour de la fréquence peut prendre plusieurs minutes après l'ouverture de l'analyse.

Gene Panel. Comme un panel in silico, ce filtre permet à l'utilisateur de sélectionner et d'appliquer une liste de gènes pour filtrer les CNV détectées dans l'analyse à l'aide du menu déroulant (voir **Gestion des paramètres utilisateur**). Ce filtre n'est pas disponible si un panel restreint a été appliqué à l'analyse à partir de l'onglet Panel du panneau latéral de gestion de l'analyse (voir la section **Vue modale Panel**).

Genes. Permet de filtrer les CNV détectées contenant les gènes sélectionnés via ce filtre.

Log Ratio. Pour les analyses basées sur les données CGH MicroArray, le Log Ratio (entre le signal brut et le signal de contrôle) présent dans les fichiers d'entrée Agilent est affiché dans une colonne supplémentaire. Le filtre associé permet à l'utilisateur de sélectionner des valeurs minimales pour les gains de copies (min pour le gain) et des valeurs maximales pour les suppressions de copies (max pour la perte). Les événements LoH (Loss of Heterozygosity) sont conservés même après l'application de ce filtre.

#Probes. Pour les analyses basées sur les données CGH MicroArray, la colonne #Probes indique le nombre de sondes utilisées dans chaque région (exonique, intergénique ou intronique). Ces

valeurs proviennent du fichier d'entrée IntervalBasedReport.xls, et le filtre associé permet à l'utilisateur de sélectionner le nombre minimum et maximum de sondes utilisées. Dans une analyse combinée CGH+SNP, le nombre de sondes indique le nombre de SNP homozygotes dans la région LOH.

Tri des CNV

La liste des CNV peut également être triée par classe de pathogénicité ACMG (ACMG), position génomique (*Location*), taille (*Size*), fréquence de l'échantillon (*Freq.*), nombre de copies (*Type*), score de qualité (QUAL) dans GermVar Tertiary et par LogRatio dans CGH MicroArray (*LogRatio*).

ACMG. La classe de pathogénicité de chaque variant est calculée de manière semi-automatique selon les recommandations de l'*American College of Medical Genetics and Genomics*. La classe de pathogénicité est le mode de tri par défaut dans le tableau CNV.

Les événements LoH (Loss of Heterozygosity) dans les CGH MicroArray sont toujours assignés à la classe 3 (VUS).

Type (Copy number). La variation du nombre de copies est affichée (par exemple : +2, -1, +3) pour les CNV détectées sur les chromosomes autosomiques (1 à 22) car le nombre de copies est fixé à 2 par défaut.

Cependant, le nombre absolu de copies est affiché (ex : 0, 1, 2) pour les CNV détectés sur les gonosomes (chromosomes sexuels X et Y) et non la variation du nombre de copies.

Contrairement à l'analyse des panels de gènes, les métadonnées relatives au sexe de l'échantillon ne sont pas nécessaires pour inclure les chromosomes sexuels dans l'identification des CNV de l'exome.

Pour les analyses combinées CGH+SNP, les événements CN-LOH, qui conservent deux copies, sont représentés dans la colonne Type par l'insigne  .

QUAL. Le score de qualité Phred est calculé dans CNVExome par le caller GATK. Lorsqu'une CNV est constituée de plusieurs intervalles génomiques, le score QUAL le plus élevé parmi tous les intervalles est alors attribué à la CNV. Pour les analyses tertiaires, le score QUAL affiché est celui obtenu à partir du fichier VCF d'entrée (DRAGEN ou ONT).

CNV Browser

L'icône ouvre un navigateur CNV dans un nouvel onglet affichant la fréquence des allèles B (BAF) des SNP (Single Nucleotide Polymorphism) présents dans l'échantillon analysé. Le graphique est décrit dans la section précédente (CNV : Gene Panels, sous CNV browser).

Page détaillée du CNV

CNV summary. Cet onglet initial présente une multitude d'informations et d'annotations sur la CNV sélectionnée, notamment :

- Détails de la classification automatique ACMG
- Les régions cliniques et les gènes inclus dans l'intervalle affecté par le variant. L'annotation des régions cliniques est basée sur les bases de données [ClinGen](#) et [OMIM](#). Seuls les gènes cliniques annotés OMIM Morbid sont affichés par défaut. Pour afficher tous les gènes cliniques, l'utilisateur peut activer la fonction "Afficher tous les gènes" à droite de la page.
- Chevauchements entre les CNV détectés et la base de données [DGV](#).
- Les petits variants (SNV et indels) détectés dans les gènes affectés par la CNV. Par défaut, le filtre SeqOne-diag est activé. Pour désactiver ce filtre, cliquez sur SeqOne-diag et désélectionnez SeqOne-diag dans les profils de filtre. Pour afficher tous les variants, l'utilisateur peut activer la fonction "Show all variants" à droite de la page.
- Une section dédiée à la base de connaissances des variants (VKB) pour identifier les variants identiques à plus de 95 % qui ont déjà été validés dans le groupe et leurs évaluations respectives.

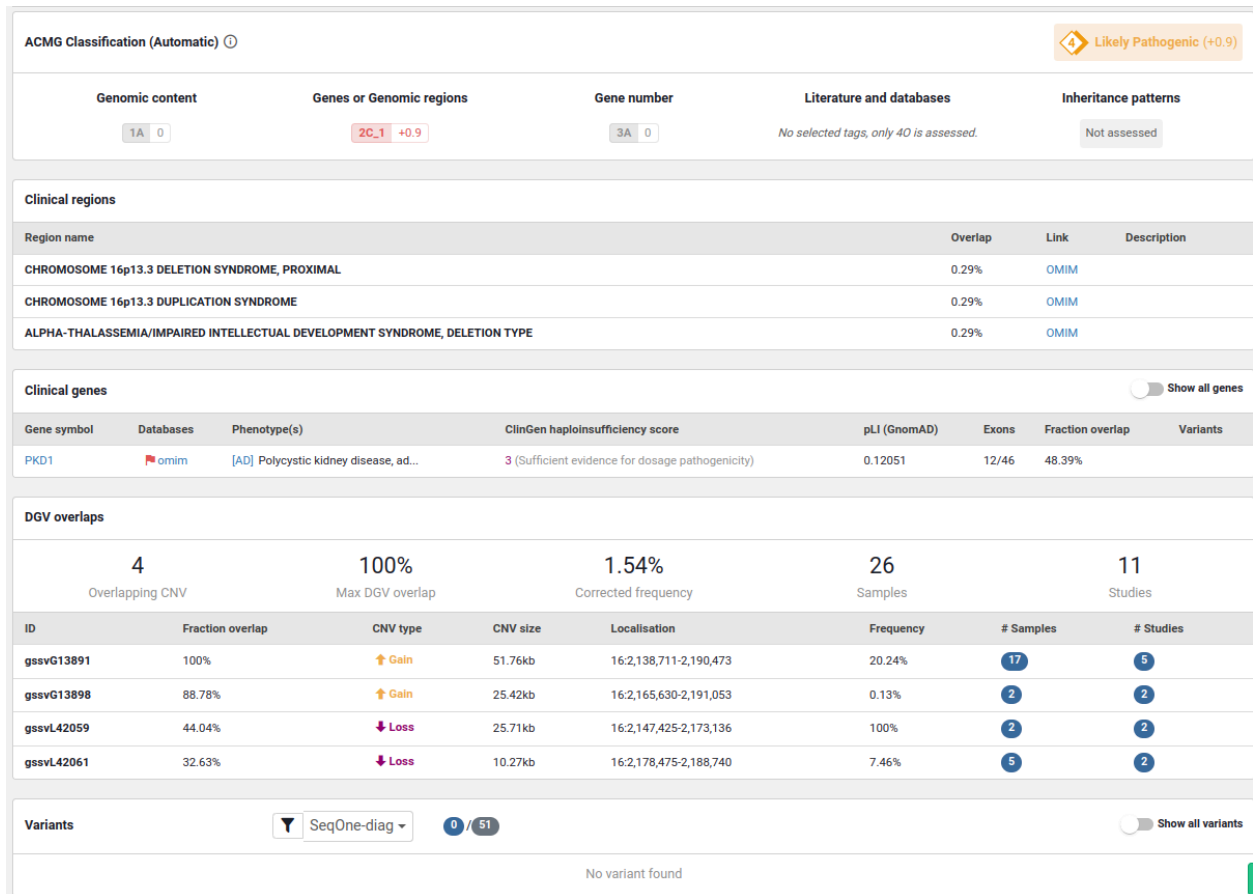


Figure 26 : Vue d'ensemble de l'onglet résumé CNV

- L'IGV du Genome browser où les CNV peuvent être visualisés (*non disponible dans les analyses GermVar Tertiary et GermVar Tertiary Family*). Pour vérifier la stabilité du signal, il est possible d'afficher jusqu'à cinq autres patients en plus de l'individu analysé à partir du menu de configuration  du graphique représentant le nombre de copies (DCN).

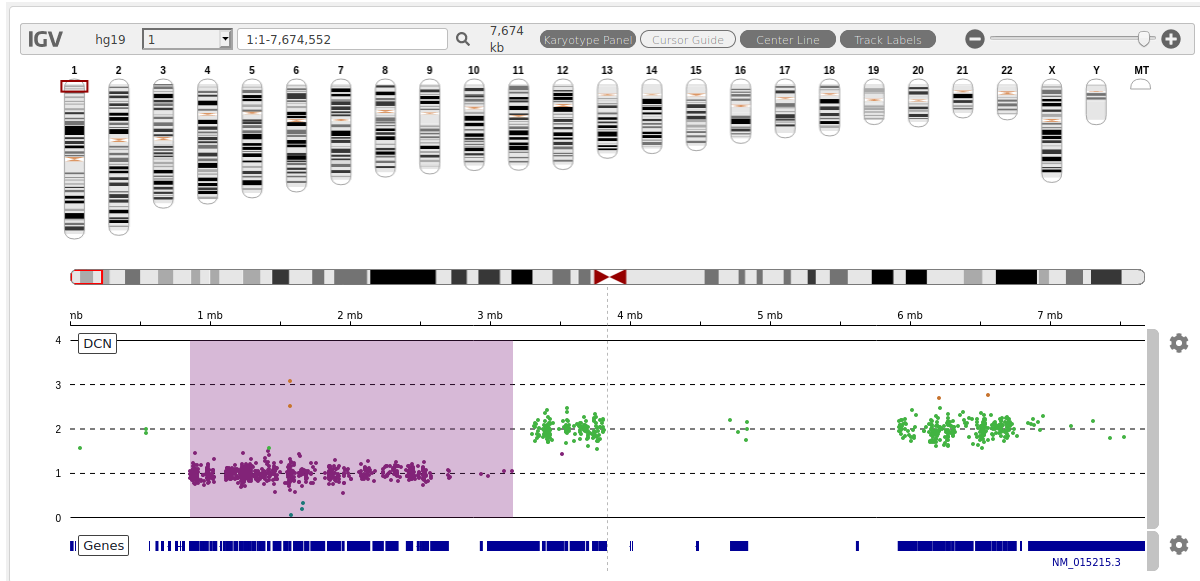


Figure 27 : Vue des CNV à partir d'un navigateur IGV intégré

Le code couleur est le suivant pour le patient sélectionné : bleu (0 à 0,5 copies), violet (0,5 à 1,5 copies), vert (1,5 à 2,5 copies), orange (2,5 à 3,5 copies) et rouge (>3,5 copies). Les autres échantillons sont affichés en gris.

CNV Occurrences. Cet onglet est utilisé pour identifier d'autres échantillons contenant la CNV, indépendamment de toute évaluation. Pour chaque occurrence, les informations suivantes seront saisies :

- le nom de l'échantillon (Sample),
- le nom de l'analyse et un lien vers celle-ci (Analyse),
- le nom du projet (Projet),
- la localisation génomique de la CNV (localisation),
- la taille de la CNV (Size),
- le type de CNV,
- le chevauchement des fractions.

Les résultats sont triés par ordre décroissant du pourcentage de chevauchement. Il est possible de filtrer les CNVs affichées en fonction de la fraction minimale de chevauchement en remplissant le champ "min fraction overlap (%)" en haut de la liste. Par défaut, ce filtre est fixé à 80 %.

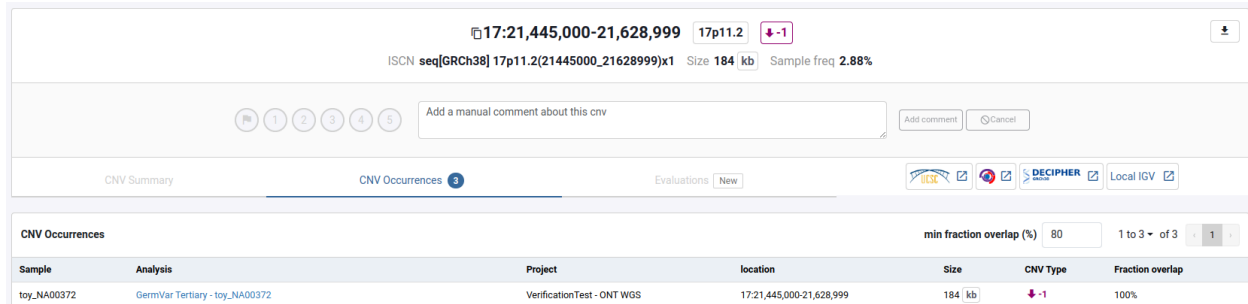


Figure 28 : Vue d'ensemble de l'onglet des occurrences de CNV dans l'analyse CNVExome

Evaluations. La classification de pathogénicité de chaque CNV peut être définie par l'utilisateur, accompagnée d'interprétations sous forme de commentaires. Chaque nouvelle évaluation permet également d'alimenter la base de connaissances personnelle du laboratoire.

UCSC. Un lien externe vers le Genome Browser de [l'UCSC](#) permet aux utilisateurs de visualiser la région génomique de la CNV. Cette page s'ouvre dans un onglet supplémentaire du navigateur et inclut les paramètres personnels de l'utilisateur pour ce navigateur, offrant ainsi une vue personnalisée.

CNV-Hub. Un lien externe vers le classificateur [CNV-Hub](#) vous permet de visualiser la classification des CNV en tenant compte de plusieurs outils de classification. Le lien ouvre un onglet supplémentaire du navigateur et intègre les paramètres personnels du navigateur de l'utilisateur, offrant ainsi une visualisation personnalisée.

DECIPHER Genome Browser. Un lien externe vers le navigateur de génome [DECIPHER](#) vous permet de visualiser les CNV sur le génome. Le lien ouvre un onglet supplémentaire du navigateur et intègre les paramètres personnels du navigateur de l'utilisateur, offrant ainsi une visualisation personnalisée.

IGV local. Ce lien redirige l'application IGV Desktop²⁹ vers les coordonnées de la CNV en cours d'interprétation. Pour visualiser la CNV, il est nécessaire d'avoir préalablement chargé l'échantillon BAM sur l'application IGV Desktop. Cette option n'est compatible qu'avec les versions ≥ 17 d'IGV Desktop.

En plus des fichiers d'alignement, divers fichiers peuvent être téléchargés à partir de l'onglet *Files* dans l'analyse CNVExome :

- Les résultats de l'analyse CNV sous forme de fichier tabulaire (cnv.tsv)
- La ploïdie de chaque chromosome (contig_ploidy.tsv)
- Le nombre prédit de copies pour chaque exon dans le manifeste (denoise.tsv)

²⁹ <https://igv.org/doc/desktop/>

Exportation des données CNV

L'icône  en haut à droite vous permet d'exporter un fichier TSV contenant les données suivantes pour la CNV concernée :

- sa notation VKB et son commentaire,
- Note de l'ISCN,
- Classification ACMG,
- chromosome, positions de début et de fin et cytobande,
- Taille et type de CNV,
- nombre de copies
- nombre de copies
- fréquence de l'échantillon
- score de qualité (échelle Phred)
- Log2Ratio
- nombre de sondes impliquées (sondes)
- nombre de gènes affectés par les CNV.

Les informations sur le(s) gène(s) associé(s) sont décrites dans les colonnes suivantes :

- le nom du gène (Gene_Symbol),
- Phénotype OMIM (OMIM_Phenotypes),
- probabilité d'intolérance génétique à la LoF (Pli_Score)
- Scores ClinGen d'haploinsuffisance et de triplosensibilité (HI_Score, TS_Score)
- le nombre d'exons chevauchés par la CNV (Nb_overlap_exon),
- le nombre total d'exons dans le gène (Nb_total_exon).

Si certains champs descriptifs ne concernent que certaines CNV (dans le cas de MicroArray CGH, par exemple), ces champs ont la valeur "NA".

13.1.6. Large Variant Viewer pour les CNVs et SVs (LVV)

Le visualiseur de grands variants (Large Variant Viewer, LVV) permet à l'utilisateur de visualiser et d'analyser les CNVs et les variants structuraux (SVs) dans un seul onglet. L'interface se compose d'un tableau central des variants, d'un panneau de filtres à gauche et d'un panneau latéral qui s'ouvre à droite lorsqu'un variant est sélectionné, similaire à l'onglet des variants (section Petits variants).

Dans les analyses compatibles, le visualiseur est situé dans un onglet intitulé "CNVs & SVs" (CNVs et SVs) ou "SVs", selon le type de données :

- L'onglet CNVs & SVs affiche les CNVs (Perte/Gain), les SVs (Délétion/Duplication) et les événements LOH/UPD.
- Onglet "CNV" existant : les CNVs sont également affichés dans l'ancien visualiseur pour maintenir la rétrocompatibilité avec les flux de travail existants.
- Les SVs ne sont accessibles que dans l'onglet "CNVs & SVs" ou "SVs".

La disponibilité des données dans le visualiseur de grands variants dépend du type d'analyse :

- WGS : Entièrement compatible (CNVs, SVs, UPDs).
- WES : Entièrement compatible (CNVs, SVs, UPDs). Les CNVs n'apparaissent que si le module d'appel de CNV est exécuté dans le cadre de l'analyse.
- Panels : Partiellement compatible (SVs uniquement).
- Analyse tertiaire (VCF/aCGH) :
- aCGH : Tous les événements de nombre de copies sont affichés.
- VCF : Les CNVs et les SVs (allèles ALT symboliques) sont affichés. Les SVs avec des séquences ALT spécifiques sont affichés dans le visualiseur de petits variants.

Table des CNVs & SVs

Similar to the small variant table, the main table displays one row per event, with the following columns providing variant-specific metrics:

DiagAI. Score de priorisation de 1 à 100, adapté aux grands variants, classant les variants selon leur probabilité d'être responsables de la maladie et cliniquement pertinents dans le contexte du cas (en particulier lorsque des termes HPO sont fournis). Le score DiagAI combine plusieurs composants, notamment : la prédiction de pathogénicité via le score UP², la correspondance de phénotype via PhenoGenius (basée sur HPO) et des règles d'expert/d'hérédité et de qualité.

ACMG. L'indicateur de classification ACMG basé sur les grands variants est affiché sous forme de score de classification de 1 à 5 dans la table LVV. Dans le tiroir de détails, un onglet ACMG affiche les critères sous-jacents et un résumé de classification en direct allant de Pathogène à Bénin (principalement pour les CNV Gain/Perte).

Classification (icône de drapeau) qui permet à l'utilisateur de classer le variant dans la VKB.

Type de CNV/SV. Type de CNV/SV (gain ou perte), taille, emplacement et cytobande,

Régions (exonique, intronique, UTR) et gènes chevauchant le CNV (noms de gènes HGNC), présence OMIM, fraction de gènes codants et détails sur les exons chevauchants. Les variants sont annotés avec les gènes chevauchants et les gènes situés à moins de 5 kb des points de cassure (marge de 5 kb).

Gènes associés à un variant sont triés par pertinence clinique (par ex., statut morbide OMIM, pourcentage de chevauchement, scores de contrainte). Les trois gènes les plus pertinents sont affichés dans le tableau principal. La liste complète est disponible dans l'onglet "Genes & Regions" du panneau latéral ("Details Drawer").

RMSK %. Pourcentage de l'événement chevauchant les annotations RepeatMasker (par ex., régions répétitives).

ClinVar & Decipher (Bases de données cliniques). Les deux colonnes affichent le nombre de correspondances de la base de données clinique correspondante. La meilleure correspondance est définie comme un événement où le chevauchement de la base de données est > 50 %. Les classifications ClinVar et Decipher sont mappées sur Bénin (B), VUS ou Pathogène (P).

DGV Gold & GnomAD (Bases de données de population). Le tableau affiche la fréquence de population maximale trouvée pour le variant. Les événements sont considérés comme correspondants si le chevauchement de variant est > 80 % et si le type d'événement est équivalent. L'onglet Fréquences dans le panneau latéral répertorie tous les événements chevauchants avec un chevauchement > 1 %.

Phenotypes. En cliquant sur les nombres affichés dans les colonnes correspondant aux résultats de l'algorithme Phenogenius, l'utilisateur peut accéder :

- au pourcentage de correspondance, qui est le score de correspondance le plus élevé entre un seul gène affecté et les termes HPO du patient.
- à la fraction de correspondance, qui est le nombre de termes HPO du patient qui correspondent à au moins un gène affecté.
- Pour l'analyse familiale, la correspondance est effectuée à l'aide des termes HPO du cas index (proband).

SampleFreq (Fréquence d'échantillon interne). La fréquence du variant au sein du projet et de l'entité actuels est affichée dans les colonnes correspondantes.

- Les correspondances sont basées sur > 80 % de chevauchement de variant et un type d'événement identique.
- Fréquence de mise à jour : les données sont basées sur un instantané des analyses effectuées avant le lancement de l'analyse en cours. Un bouton manuel "Refresh" (Actualiser) est disponible dans le Details Drawer pour mettre à jour ces données.

Colonnes de scores. L'utilisateur peut afficher plusieurs scores :

- UP² : score de prédiction de pathogénicité qui est un composant de DiagAI.
- pLI, LOEUF : intolérance à la perte de fonction.
- pTriplo, pHAPLO : Triplosensibilité et Haploinsuffisance.
- HI ClinGen, TS ClinGen.

Colonnes d'échantillon. Ces colonnes affichent des métriques d'événement spécifiques à la source de données :

- Pour les analyses familiales (à l'exclusion de aCGH), les métriques d'échantillon sont affichées pour tous les membres de la famille.
- Genotype : pour les CNVs où un génotype n'est pas fourni par l'outil d'appel, le système déduit le génotype en fonction du nombre de copies absolu par rapport à la ploïdie :
 - Homozygote (hmz) avec un point entièrement sombre
 - Hétérozygote (htz) avec un point à moitié sombre
- VAF / #Reads : fréquence allélique du variant et lectures de support (disponible pour les SVs).
- #Copy : nombre absolu de copies (disponible pour les CNVs).
- LogRatio / Probes : disponible pour les données aCGH.
- QUAL : score de qualité de l'outil d'appel de variant ou du fichier d'entrée.

Tri, filtrage, visualisation et interprétation

Filtrage. Le panneau de filtre permet à l'utilisateur de filtrer les variants en fonction des données de colonne, de manière similaire au système de filtrage de l'onglet Petits variants (voir la section correspondante).

- Panel de gènes : filtre les événements chevauchant des gènes dans une liste de gènes spécifique.
- Qualité : filtre en fonction de la colonne QUAL. Par défaut, les événements sans score de qualité (N/A) sont masqués sauf si "Keep unknown" (Garder les inconnus) est sélectionné.
- Préréglages : les configurations de filtre peuvent être enregistrées en tant que préréglages pour une utilisation future.
- Pour les analyses familiales, le filtre Qualité utilise la logique "OR" (OU). Le variant est affiché si au moins un membre de la famille atteint le seuil de qualité.

Tri. Par défaut, le tableau est trié par DiagAI Score (décroissant). L'utilisateur peut trier par UP2, Size, Location ou Sample Freq. Entity, par ordre croissant ou décroissant.

Comme dans l'ancien onglet CNV, l'utilisateur peut ouvrir les fenêtres de ploïdie et le navigateur CNV en haut du tableau (voir la section CNV : Visualisation et interprétation). Une icône de gestion des colonnes est également disponible tout à droite de l'en-tête.

Panneau latéral CNV/SV (Details Drawer)

Semblable à la page des variants (section Petits variants), la sélection d'un variant dans le Large Variant Viewer ouvre une page détaillée sur le côté droit de l'écran. Cette page contient plusieurs onglets :

Variants. Cet onglet récapitulatif comprend la nomenclature du variant, le score DiagAI, la liste des hétérozygotes composites suspectés (Suspected Het Compound), les informations sur le gène et la région, et les liens vers les bases de données.

ACMG. Classification détaillée.

Gènes et régions. la liste complète des gènes et des régions associés au variant, avec des informations cliniques détaillées telles que le statut morbide OMIM, le pourcentage de chevauchement, les scores de contrainte et la correspondance de phénotype.

Classification. Classez le variant actuel dans votre VKB en haut du panneau latéral, et enregistrez-le. La classification actuelle et les occurrences de ce variant dans d'autres analyses se trouvent sur cette page.

Fréquences. L'utilisateur peut accéder aux détails sur les fréquences des variants dans les bases de données DGV Gold et GnomAD, ainsi que pour l'entité et le projet actuel.

Bases de données cliniques. ClinVar & Decipher.

Viewer. L'utilisateur peut choisir de visualiser le variant dans le SeqOne viewer ou IGV. Dans le sous-onglet IGV, l'utilisateur peut afficher le "Start" (Début) ou le "End" (Fin) de l'événement, ou "Start to End" (Du début à la fin). Dans le sous-onglet SeqOne, l'utilisateur peut accéder au Genome Browser et visualiser l'événement dans l'une des multiples couches disponibles : CNV, SV, UPD, Gene, CN (copy number), BAF, DGV.

13.1.7. Signatures dans l'analyse SomaCGP

Méthodologie de calcul du score MSI

L'analyse MSI est une analyse individuelle réalisée par rapport à une référence standardisée. Cette référence consiste en un ensemble d'échantillons négatifs préalablement sélectionnés et regroupés. L'échantillon tumoral est comparé à cette référence à l'aide de l'outil *MSIsensor*. Un score est alors calculé, représenté par le pourcentage de régions microsatellites instables couvertes par le panel.

Le panel de normaux actuellement pris en charge par SomaCGP correspond au panel Agilent SureSelect Cancer CGP. L'utilisation d'un autre kit ou d'une conception personnalisée dérivée du CGP nécessitera la création d'un panel de normaux dédié.

Méthodologie de calcul de la note du TMB

Le TMB est calculé en trois étapes :

1. Calcul de la taille des régions codantes incluses dans le panel
2. Identification des variants somatiques
3. Calcul du nombre de mutations somatiques par mégabase (mut/Mb)

L'identification des variants somatiques est basée sur un ensemble de filtres décrits ci-dessous :

- Type de variant et filtres d'effet
 - SNV uniquement (les MNV sont exclus)
 - Effets retenus : gain de codon d'arrêt, décalage du cadre de read, variant d'épissage (accepteur et donneur), ablation de transcription, délétion/insertion en phase, variant de faux-sens, variant synonyme.

Les variants synonymes sont exclus du calcul du score TMB non synonyme.

- Filtres de qualité variable
 - variants détectées par Mutect2 qui passent les filtres FilterMutectCalls
 - Profondeur de champ (DP) $\geq 50X$
 - Score de qualité > 30
 - Rapport de soutien allélique (ASR) $> 0,1$
- Filtres de variants somatiques
 - Fréquence des allèles variants (VAF) entre 5% et 40%.
 - Fréquence dans la population générale (GnomAD) $\leq 0,0001$

Risques liés à l'utilisation d'échantillons de contrôle synthétiques de TMB

- Distribution de mutations non naturelles :
 - Les mutations créées peuvent ne pas refléter le contexte génomique des tumeurs réelles (par exemple, regroupement des mutations, biais de brin, contexte de séquence).
- Manque d'hétérogénéité tumorale :
 - Les échantillons synthétiques ne reproduisent pas la diversité clonale ou la pureté tumorale, ce qui affecte la distribution de la fréquence des allèles variants (VAF).
 - Pas de bruit de fond ni d'artefacts FFPE : les échantillons réels présentent des artefacts de séquençage, des dommages FFPE et des variations du nombre de copies, ce qui n'est pas le cas du matériel synthétique.
- Taille des fragments et comportement chimique différents :
 - Les constructions synthétiques se comportent souvent différemment lors de la préparation des librairies, de la capture des hybrides et du séquençage.

Visualisation et interprétation

L'onglet *Signatures* affiche les résultats des analyses MSI et TMB en 2 sections :

Microsatellite Instability. Cette vue montre des informations telles que :

- le nombre de régions microsatellites dont la couverture est supérieure ou égale à 50X (sites dont la couverture est supérieure à 50X),
- le nombre de sites somatiques,
- Score MSI en pourcentage et intervalle de confiance à 95 % (IC),
- Statut MSI

Tumor Mutational Burden. Cette vue montre les informations suivantes :

- la taille de la région codante exprimée en Mb (taille de la région codante),
- le nombre de variants somatiques et non synonymes passant les filtres (variants somatiques passant les filtres et variants somatiques non synonymes passant les filtres),
- le score TMB de tous les variants somatiques passant les filtres en mut/Mb (score TMB),
- le score TMB des variants non synonymes passant les filtres mut/Mb.

Chaque résultat peut être téléchargé en cliquant sur l'icône  en haut à droite de chaque section en tant que fichier tsv.

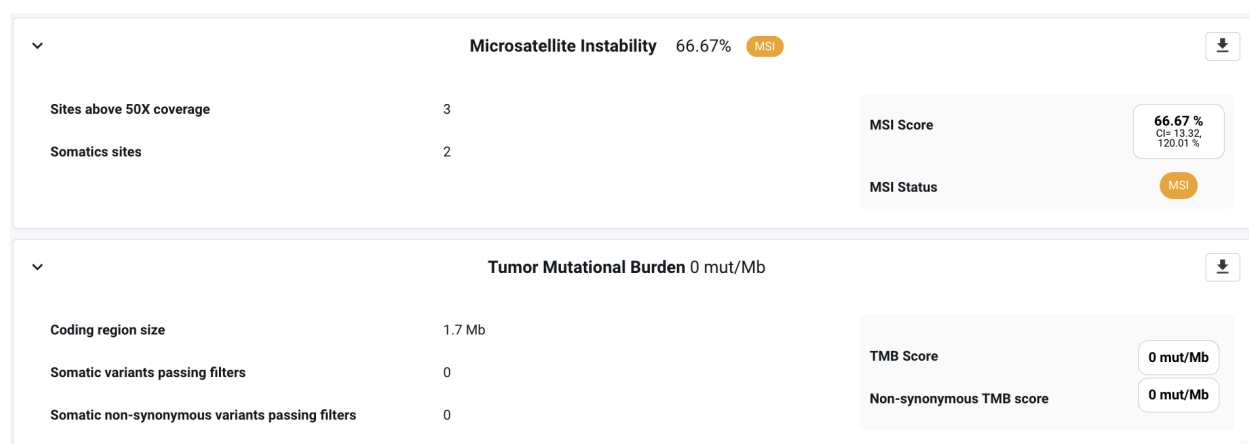


Figure 29 : Visualisation de l'onglet Signatures d'une analyse SomaCGP.

13.1.8. Répétitions courtes en tandem (STRs)

La visualisation des répétitions courtes en tandem (STR) est disponible pour les worksets d'analyse GermVar Tertiary et GermVar Tertiary Family dédiés à l'analyse des données de séquençage du génome entier (WGS), lorsqu'elles sont générées à partir du séquençage d'Oxford Nanopore Technologies.

La détection des répétitions en tandem courtes (STR) est également disponible dans les ensembles de travail d'analyse GermVar et GermVar Family pour le WGS à partir de données FASTQ générées avec Illumina, Element Biosciences, MGI et Oxford Nanopore Technologies.

Méthodologie

Répétitions en tandem courtes à partir de FASTQ

Les données de répétition en tandem courte (STR) sont détectées à l'aide de ExpansionHunter, qui traite le fichier BAM avec un génome de référence et un catalogue de variants définissant les locus cibles. Le VCF résultant est ensuite annoté par Stranger, qui classe chaque locus comme Normal, Pré-mutation ou Pathogène en fonction du nombre de répétitions observé et des seuils de maladie connus.

Les deux outils sont intégrés directement dans les pipelines GermVar et GermVar Family.

ExpansionHunter: <https://github.com/Illumina/ExpansionHunter>

Stranger: <https://github.com/moonso/stranger>

Répétitions en tandem courtes à partir de VCF

Les données relatives aux répétitions en tandem sont extraites du fichier VCF (sample.wf_str.vcf.gz) généré par le workflow wf-human-variation v2.3.0 de l'application EPI2ME (Oxford Nanopore Technologies) et fourni par l'utilisateur. Comme ces données ont été annotées pour chaque locus porteur de STR, à partir d'un catalogue de variations maintenu par EPI2ME, notre pipeline ne modifie pas les annotations. Le statut prémuté ou pathogène de chaque allèle est déterminé en fonction du nombre de répétitions séquencées et des seuils de prémutation et de pathogénicité connus.

Visualisation et interprétation

L'onglet STRs vous permet de visualiser les *Short Tandem Repeats* sous la forme d'un tableau affichant les détails suivants pour chaque locus génotypé :

Informations et annotations sur le locus

- [Gene](#) : nom du gène (HGNC) et lien correspondant.
- [Site](#) : les coordonnées génomiques de la répétition.

- **Repeat Unit** : motif nucléotidique répété.
- **Disease** : la maladie associée selon la base de données OMIM et le lien correspondant.
- **Transmission** : le mode de transmission connu.
- **Premutation range** : plage de prémutation correspondant au nombre minimal et maximal de répétitions déterminant le statut de prémutation.
- **Pathogenic range** : seuil inférieur de répétitions déterminant le statut pathogène.

Détails spécifiques à chaque échantillon

- **GT** : le génotype homozygote, biallélique ou hétérozygote du locus.
- **Allele repeats** : nombre de répétitions par allèle. Les loci homozygotes portent une seule valeur. Pour les loci bialléliques et hétérozygotes, le nombre de répétitions pour les deux allèles est indiqué. Le statut de chaque allèle est déterminé par le nombre de répétitions : type sauvage en vert, prémuté en orange et pathogène en rouge.
- **Supp. reads** : le nombre de reads de soutien pour chaque allèle. Les loci homozygotes portent une seule valeur, tandis que les valeurs pour 2 allèles sont indiquées pour les loci bialléliques et hétérozygotes.

STRs 34										
Gene	Site	Repeat unit	Disease	Transmission	Premutation range	Pathogenic range	GT	Allele repeats	Supp. reads	
PABPN1	chr14:23321472-23321490	GCG	Oculopharyngeal Muscular Dystrophy	AD	10 - 11	> 11	●	16 - 6	0 - 0	
ATXN3	chr14:92071009-92071042	CTG	Machado-Joseph disease	AD	44 - 60	> 60	●	15	0	
ATXN10	chr22:45795354-45795424	ATTCT	Spinocerebellar Ataxia types 10	AD	32 - 800	> 800	●	14 - 12	0 - 0	
NIPA1	chr15:22786677-22786701	CGCGGG	Amyotrophic lateral sclerosis-parkinsonism/dementia complex, susceptibility to	AD	8 - 10000	> 10000	●	6 - 4	0 - 0	
NOP56	chr20:2652733-2652757	GCCTGG	Spinocerebellar Ataxia types 36	AD	14 - 650	> 650	●	8 - 5	0 - 0	
NOP56	chr20:2652757-2652775	GCCTGG	Spinocerebellar Ataxia types 36	AD	14 - 650	> 650	●	7 - 4	0 - 0	
ATN1	chr12:6936716-6936773	CAG	Dentatorubropallidoluysian Atrophy	AD	35 - 48	> 48	●	25 - 19	0 - 0	

Figure 30 : Affichage de l'onglet STRs d'une analyse GermVar Tertiary.

Le tableau peut être téléchargé en tant que fichier repeats.tsv via l'icône  en haut à droite du tableau. Ce fichier est également disponible dans l'onglet Fichiers de l'analyse.

13.1.9. Gènes de fusion

13.1.9.1. Fusions détectées à partir de l'ADN

Méthodologie

Le module [GRIDSS](#) (Genome Rearrangement IDentification Software Suite), spécialisé dans la détection des réarrangements génomiques et implémenté dans le workset SomaVar et SomaVar LF, est utilisé pour détecter les fusions de gènes à partir du séquençage de l'ADN, enrichi par la capture ciblant un panel de gènes.

Visualisation et interprétation

L'onglet Fusions d'une analyse somatique (SomaVar, SomaVar LF, SomaCGP, SomaVar Tertiary) affiche la liste des fusions passant les différents filtres GRIDSS sous la forme d'un tableau détaillant les informations relatives à la fusion et à chaque partenaire impliqué :

Informations sur la fusion

- avec les symboles génétiques de chaque partenaire,
- le type d'événement (translocation, duplication, inversion ou délétion),
- l'occurrence de la fusion dans la base de données COSMIC,
- la fréquence de l'allèle (VAF),
- le nombre de reads soutenant la fusion,
- couverture totale,
- les filtres attribués par le GRIDSS.

Informations relatives aux partenaires

- les coordonnées génomiques du point de fusion (breakend),
- le transcrit RefSeq concerné,
- le site correspondant au point de fusion (breakpoint) : 5'UTR, 3'UTR, Intron, Exon (point de rupture non codant), CDS (point de rupture codant) ou Site d'épissage,
- la numérotation de l'exon ou de l'intron impliqué dans la fusion, par rapport au nombre total d'exons ou d'introns.

Un menu de filtres vous permet de

- rechercher une fusion spécifique ou un gène impliqué dans une fusion à l'aide d'une barre de recherche,
- afficher les fusions qui n'ont pas passé les filtres GRIDSS (LOW_QUAL) en cochant la case "Show filtered fusions". Cette case est décochée par défaut.

FUSION	BREAKEND 1	BREAKEND 2	TYPE	DATABASES	TRANSCRIPT 1	SITE 1	EXON/INTRON 1	TRANSCRIPT 2	SITE 2	EXON/INTRON 2	VAF	SUPP. READS	COVERAGE	FILTER
ALK::EML4	2:29446685 (-)	2:42523824 (+)	inversion	cosmic	NM_004304	intron	19/28	NM_019063	intron	12/22	0.329 %	14	4251	PASS
NRAS::LINC00486	1:115256634 (-)	2:33141297 (+)	translocation	N/A	NM_002524	intron	2/6	NR_027098	intron	3/5	5.629 %	128	2274	PASS
LINC00486::RET	2:33141301 (+)	10:43612238 (+)	translocation	N/A	NR_027098	intron	3/5	NM_020975	intron	11/19	3.617 %	118	3262	PASS
LINC00486::FGFR3	2:33141349 (+)	4:1803528 (+)	translocation	N/A	NR_027098	intron	3/5	NM_000142	intron	5/17	0.352 %	28	7963	PASS
PRIM2::HRAS	6:57331998 (+)	11:533875 (-)	translocation	N/A	NM_000947	intron	6/13	NM_005343	CDS	3/6	0.126 %	6	4753	LOW_QUAL

Figure 31 : Visualisation de l'onglet Fusions pour une analyse SomaVar ou SomaVar LF.

Les données du tableau peuvent être triées en fonction des coordonnées génomiques (breakend), de la présence de la fusion dans COSMIC (Databases), de la fréquence allélique (VAF) et du nombre de reads soutenant la fusion (Supporting reads).

Le lien associé aux coordonnées génomiques de chaque point de rupture ouvre une vue fractionnée du point de fusion sur IGV. Les infobulles détaillent le nombre de reads scindés et de paires non appariées dans la colonne Supporting reads, et la couverture à chaque point de rupture (Cov. 1 et Cov. 2) dans la colonne Coverage.

La colonne Filtre affiche 3 filtres de qualité utilisés par GRIDSS et donnant une indication de la confiance dans l'appel de fusion :

- PASS : la fusion a passé les filtres de qualité. Le score de qualité est supérieur ou égal à 500 et est considéré comme un appel de confiance moyen ou élevé (s'il est > 1000).
- LOW_QUAL : la fusion a un score de qualité inférieur à 500 pour au moins l'une des deux ruptures impliquées dans la fusion, et est considérée comme un appel de faible confiance.
- ASSEMBLY_ONLY : la fusion est appelée sur la base d'un assemblage et non sur la base du read qui la soutient. L'assemblage fait référence à la reconstruction des séquences autour des points de fusion potentiels, qui sont ensuite réalignés pour identifier la fusion.

Si plusieurs filtres sont disponibles pour la même fusion, un signe "+" affiche tous les filtres dans la colonne Filtre.

En plus des résultats présentés dans l'interface, un fichier fusions.tsv détaillant les fusions

détectées est disponible en téléchargement depuis l'onglet Fichiers ou depuis l'icône  en haut à droite de la page. Il contient les informations suivantes :

- nom de la fusion avec les symboles des gènes de chaque partenaire (fusion)

- coordonnées génomiques de chaque point de fusion (breakend_1 et breakend_2)
- transcrit canonique pour chaque partenaire (transcript_canonical_1 et 2)
- site correspondant au point de fusion pour chaque partenaire (site_1 et site_2)
- numérotation des exon ou des introns selon le transcrit RefSeq associé (exon_1 et 2)
- 60 pb de séquences flanquantes entourant le point de fusion (flanking_region_1 et 2)
- filtres de qualité attribués par GRIDSS (filtre)
- type d'événement (type)
- nombre de split reads (split_reads)
- nombre de paires de reads discordantes (discordant_pairs)
- couverture pour chaque partenaire (couverture_1 et couverture_2)
- fraction des reads soutenant la fusion (read_fraction)
- nombre de reads soutenant la fusion (supporting_reads)
- couverture totale (couverture)
- Identifiant de fusion COSMIC (cosmic_id)

Ce fichier contient :

- Fusions affichées dans l'interface, ayant passé les filtres de qualité (filtre = PASS),
- Fusions détectées mais avec un score de qualité inférieur à 500 pour au moins l'une des deux ruptures impliquées dans la fusion (filtre = LOW_QUAL).

Le score de qualité pour chaque breakend se trouve dans le fichier gridss.raw.vcf.gz (colonne QUAL), téléchargeable à partir de l'onglet Fichiers.

13.1.9.2. Fusions détectées à partir de l'ARN

L'outil SomaRNA permet d'analyser les données RNA-seq provenant d'approches ciblées telles que la capture et l'amplicon, ce qui signifie que les utilisateurs peuvent identifier et visualiser les événements de fusion de gènes au niveau de la transcription.

Méthodologie

Une combinaison de deux outils est utilisée pour détecter et visualiser les fusions par SomaRNA :

- [STAR](#) : un aligneur de données RNA-seq conçu pour aligner des séquences non contiguës directement sur le génome de référence.
- [Arriba](#) : détecte les fusions de gènes à partir d'alignements chimériques

Sur la base des alignements chimériques de STAR, Arriba applique un ensemble de filtres pour éliminer les artefacts connus et les transcrits observés dans des contextes non pathologiques. Il attribue à chaque événement de fusion potentiel une note de confiance qui dépend de plusieurs critères :

- Le nombre de séquences qui soutiennent la fusion
- L'équilibre entre les split reads et les paires discordantes
- La distance entre les points de rupture, leur proximité (intragénique ou autre)
- Le type d'événement

Les fusions détectées avec un degré de confiance moyen ou élevé par Arriba sont affichées sur SeqOne. Celles dont la confiance est faible sont répertoriées dans les fichiers "arriba_discarded" (voir ci-dessous). La confiance d'Arriba reflète la probabilité d'un transcrit aberrant non observé dans les tissus sains, la possibilité d'un réarrangement génomique sous-jacent et l'exclusion des artefacts.

Plus de détails sur le niveau de confiance dans la documentation d'Arriba : <https://github.com/suhrig/arriba/blob/master/documentation/08-Interpretation-of-results.md>

Pour comprendre pourquoi une fusion est "Low confidence", consultez la colonne "filters" (AA) dans le fichier "discarded". Elle indique le filtre qui a écarté l'événement. Détails de chaque filtre : <https://github.com/suhrig/arriba/blob/master/documentation/11-Internal-algorithm.md>

Visualisation et interprétation

A partir de l'onglet Fusions d'une analyse SomaRNA, l'utilisateur obtient une liste de prédictions de fusion, sous la forme de différents onglets.

La partie supérieure de la page donne des détails sur la fusion et sur chacun des partenaires impliqués :

Informations générales sur la fusion

- le nombre de reads de la jonction, égal à la somme des reads des divisions respectives de chacun des gènes 5' et 3' impliqués dans la fusion,
- le nombre de paires discordants,
- le degré de confiance attribué par Arriba (élevé, moyen ou faible),
- le type d'événement à l'origine de la fusion (translocation, duplication, inversion ou délétion),
- l'impact de la fusion sur le cadre de lecture,
- le cas échéant, l'occurrence de la fusion dans la base de données COSMIC.

Informations sur chaque partenaire de fusion

- le symbole de chaque gène,
- le transcrit Refseq ou, à défaut, ENSEMBL pour chaque gène,
- la numérotation des exons impliqués dans la fusion,
- les coordonnées génomiques de chaque point d'arrêt,

- la couverture totale au point de fusion (Couverture),
- la fraction de read soutenant la fusion.

Les split reads sont des reads alignés majoritairement sur le gène 1 ou le gène 2 et chevauchant la fusion. Dans le cas de paires de reads discordantes, l'un des reads est sur le gène 1 et l'autre sur le gène 2, sans nécessairement chevaucher la fusion.

La fraction de read est la somme des reads fractionnés et des paires de reads discordantes associées à un gène, divisée par la couverture totale. La valeur de la fraction de read peut être supérieure à 100 % lorsque les split reads associées à un gène font également partie de paires de reads discordantes. Ces reads sont alors comptés deux fois au lieu d'une.

DHRX - TCEANC

E3 - E4

POLR2J2 - EFCAB13

E3 - E18

POLR2J2 - EFCAB13

E4 - E18

MEGF8 - CNFN

E3 - E4

General informations

of spanning junction reads

568

of discordant mates

246

Confidence ⓘ

high

Fusion type

translocation

Reading frame

in-frame

Database

cosmic

Partners informations

5prim

3prim

Gene

NPM1

ALK

Transcript

ENST00000517671 ⓘ

NM_004304.4

Exons

5 / 12

20 / 29

Breakpoint

5 : 170818803 (+)

2 : 29446394 (-)

Coverage

14004

8455

Read fraction

3.9%

6.08%

Figure 32 : Visualisation des informations de fusion provenant d'une analyse SomaRNA.

La partie inférieure de la page présente des représentations graphiques de la fusion, tant au niveau du gène et de l'exon qu'au niveau de la protéine. Cette dernière section est divisée en deux onglets :

- Le premier onglet présente la structure de la fusion et les partenaires impliqués, où la couverture des exons couverts par le manifeste d'analyse est représentée par un diagramme sur fond jaune :

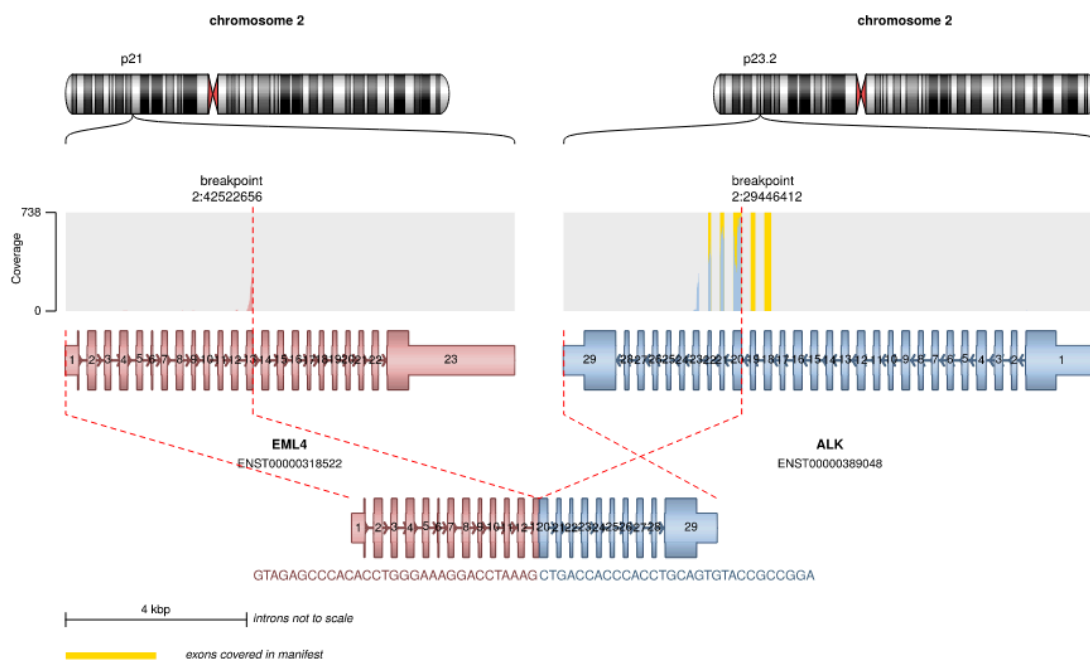


Figure 33 : Vue d'un événement de fusion entre deux gènes

- Les domaines protéiques conservés après l'événement de fusion sont mis en évidence dans une deuxième vue :

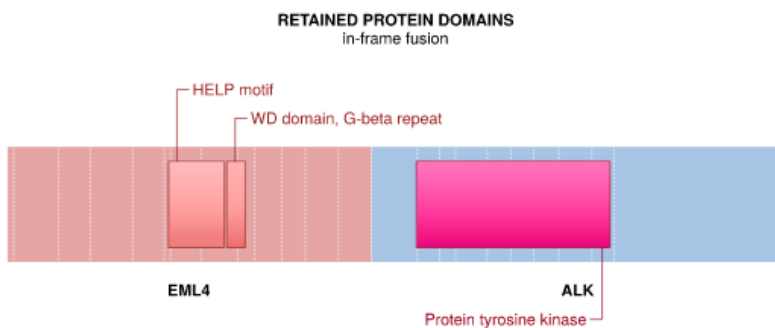


Figure 34 : Vue du polypeptide résultant de la fusion de deux gènes

En plus des résultats présentés dans l'interface, différents fichiers sont disponibles au téléchargement :

- Les fusions retenues par Arriba au format TSV (fichier arriba.tsv),
- Les fusions éliminées par les filtres Arriba au format TSV (fichier arriba_discarded_standard.tsv),
- les fusions éliminées par les filtres Arriba et répertoriées dans COSMIC au format TSV (fichier arriba_discarded_cosmic.tsv),
- les différentes figures affichées dans l'interface au format SVG

13.1.10. Clonalité

L'onglet 'IGHV mutational status report' dans l'onglet "Clonality" des analyses SomaHemato présente les résultats du statut mutationnel.

La section 'IGHV Results' affiche :

- Le statut mutationnel de l'IGHV et le pourcentage d'homologie :
 - Mutated: si l'homologie est < 98% d'identité avec la lignée germinale
 - Unmutated: si l'homologie est $\geq$ 98% d'identité avec la lignée germinale
 - Borderline: pour les clones dont l'identité se situe entre 97 et 98 %.
 - Polyclonal: pas de clone majeur au-dessus de 20 %.
 - Non-Conclusive: si l'un des contrôles de qualité essentiels échoue
- Le sous-ensemble déterminé (avec impact clinique) et confiance : *2 (risque élevé), *8 ou None
- Une notification d'avertissement si des avertissements (warnings) apparaissent dans le rapport.

La section 'Clonality' énumère les clones identifiés avec leur fréquence et leur statut de productivité (Productive, Un-productive). Chaque clone dispose d'une section dédiée indiquant son statut mutationnel, sa fréquence, le lien vers la page IMGT_quest et des détails supplémentaires (homologie, exhaustivité) ainsi que les contrôles de qualité.

13.1.11. Variants d'épissage

Methodologie

Le workset SomaRNA utilise la suite d'outils *Regtools* pour détecter les sauts d'exon et identifier les variants d'épissage alternatives à partir du séquençage de l'ARN, enrichi par la capture ciblant un panel de gènes. RegTools extrait les jonctions exon-exon du fichier BAM généré par l'aligneur de données RNA-seq STAR, et les annote selon une structure de transcription connue basée sur un fichier GTF et des annotations Ensembl.

Visualisation et interprétation

L'onglet Variants d'épissage d'une analyse SomaRNA est axé sur la visualisation des variants d'épissage des gènes cliniquement pertinents suivants et de leur variant principal pouvant faire l'objet d'une action :

- MET (MET Proto-Oncogène, Récepteur Tyrosine Kinase) et le variant actionnable MET Δ 14
- EGFR (récepteur du facteur de croissance épidermique) et le variant d'épissage actionnable EGFRvIII
- AR (Androgen Receptor) et le variant d'épissage actionnable AR-V7

Une liste complète des variants d'épissage détectés dans l'analyse SomaRNA, en plus de ceux énumérés ci-dessus, peut être téléchargée directement sous forme de fichier TSV via l'icône Download dans l'interface ou à partir de l'onglet Files dans l'analyse.

Pour chaque gène mentionné ci-dessus, l'état de détection (détecté ou non détecté) du principale variant pouvant faire l'objet d'une action est affiché, ainsi que le nombre d'autres événements alternatifs détectés pour ce gène.

Les utilisateurs peuvent consulter les détails des principales variants d'épissage détectées en ouvrant un encart dédié, divisé en trois parties :

- Sur la gauche, des informations sur le principal variant d'épissage exploitable sont affichées : coordonnées génomiques, nom du gène, transcrit RefSeq, type de variant détecté détaillant les exons manquants, phase de lecture, fraction de read et nombre de reads soutenant le variant. Enfin, la moyenne des reads de support canonique est spécifiée. Cela correspond à la moyenne des reads 5' et 3' supportant l'inclusion de l'exon.
- À droite, un diagramme sashimi (Junction Plot) illustre les jonctions détectées entre les exons, ainsi que le nombre de reads soutenant chaque jonction.

- En bas, des informations sur les événements alternatifs détectés sont énumérées, c'est-à-dire les isoformes autres que le variant principal d'intérêt décrit ci-dessus : type de variant détecté (numéros d'exon manquants entre parenthèses), phase de lecture, fraction et nombre de reads soutenant le variant, nombre moyen de reads soutenant le transcrit canonique (Avg. canonical support), statut PASS ou FAILED en fonction des filtres décrits ci-dessous.

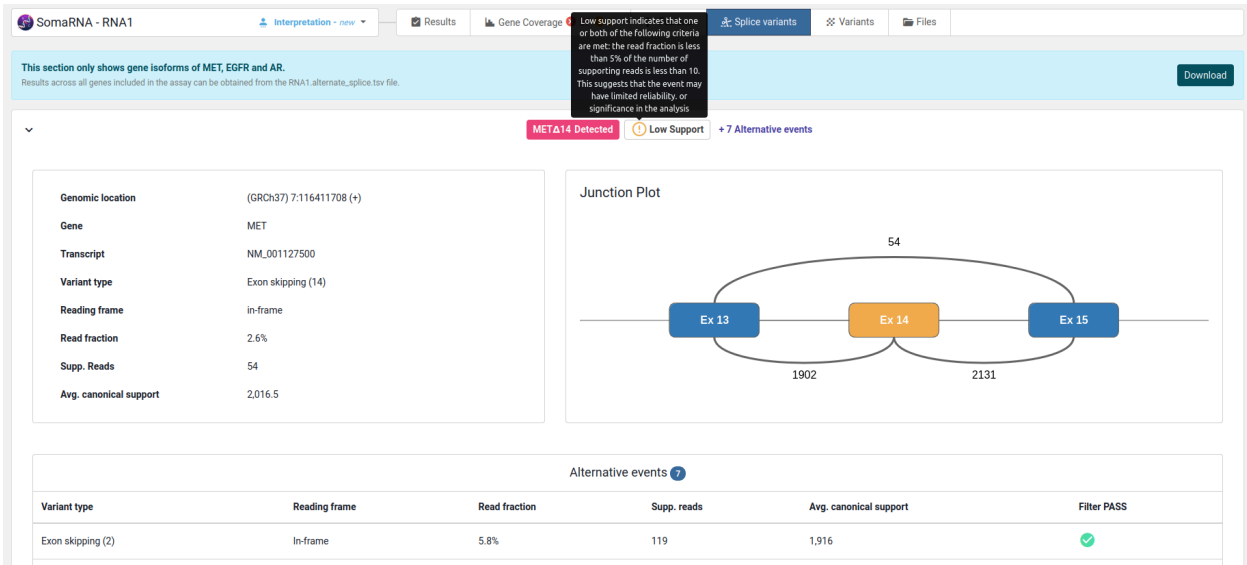


Figure 35 : Visualisation de l'onglet Splice variants d'une analyse SomaRNA.

La détection des variants d'épissage dans les gènes affichés dans l'onglet variants d'épissage n'est pas soumise à l'application de filtres de détection.

Néanmoins, une icône "Low support" associée à la détection de METΔ14 et EGFRvIII indique à l'utilisateur que les métriques relatives aux supporting reads sont inférieures aux seuils suivants : fraction de reads < 5 % ou nombre de supporting reads < 10.

Le statut PASS ou FAILED d'autres événements alternatifs utilise également ces seuils :

- *PASS* : fraction de reads >= 5% et nombre de reads >= 10.
- *FAILED* : fraction de reads < 5% ou nombre de reads < 10.

Le fichier TSV téléchargeable contient les informations suivantes pour tous les événements d'épissage détectés : transcrit RefSeq, gène, direction du brin (+ ou -), fraction de reads et nombre de reads soutenant le variant, nombre d'exons manquants et numéros d'exons correspondants, phase de lecture, coordonnées génomiques des jonctions. Le fichier précise également sous FILTER_PASS (oui/non) si le variant passe les seuils appliqués aux reads de support (fraction de reads >= 5% et nombre de reads >= 10) et sous Actionable (oui/non), l'actionnabilité du variant.

13.1.12. Statut MSI/MSS

Le workset d'analyse SomaMSI est disponible lors de l'exécution d'une nouvelle analyse sur les données d'un panel de gènes et est utilisé pour déterminer l'état MSI ou MSS d'un échantillon de tumeur.

Methodologie

Cette analyse de cohorte non individuelle est similaire à une analyse CNVCapture ou SomaCNVCapture. Au moins huit échantillons sont nécessaires pour effectuer l'analyse.

L'outil génère un échantillon de référence à partir des échantillons de la cohorte dont on estime qu'ils sont négatifs. Cet état des échantillons n'étant pas connu à l'avance, cette estimation est basée sur le calcul du taux de délétions observées dans les régions microsatellites : la moitié des échantillons de la cohorte présentant le taux le plus faible est donc sélectionnée pour générer un échantillon de référence, chaque échantillon apportant ainsi une fraction de ses séquences.

Chaque échantillon de tumeur est ensuite comparé à l'échantillon de référence à l'aide de l'outil MSIsensor. Un score est calculé, qui représente le pourcentage de régions microsatellites étudiées qui présentent une instabilité. Les régions étudiées sont les régions microsatellites couvertes par le panel et qui sont détectées par le MSIsensor.

Sur la base de ce score, l'état MSI ou MSS des différents échantillons est dicté par un seuil calculé comme suit : $\max(10, \text{median} + \max(3 * \text{median-abs-dev}, 0.3 * \text{median}))$

Le principe de ce score est le suivant :

- Dans une analyse présentant des échantillons MSI et MSS, le seuil sera généralement la médiane + 3*médiane-abs-dev.
- Toutefois, si tous les échantillons ont l'état MSS, si les scores sont très faibles et si leur distribution est extrêmement proche, le seuil sera fixé à 10 pour éviter de générer des faux positifs.

Visualisation et interprétation

Une fois l'analyse terminée, les résultats de tous les échantillons peuvent être visualisés à partir de l'analyse SomaMSI. L'onglet *Results* est divisé en deux sections :

MSI graph : représentation graphique des résultats où chaque échantillon est représenté en fonction de son score. Le seuil déterminé par la cohorte d'échantillons crée deux zones dans le graphique représentant les échantillons considérés comme MSI et MSS respectivement.

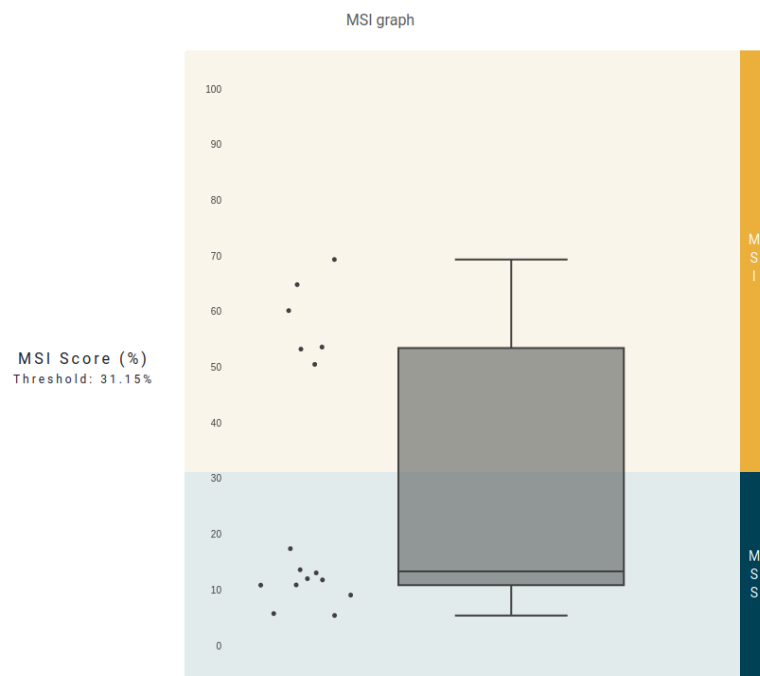


Figure 36 : Diagramme en boîte représentant la distribution des scores MSI pour la cohorte de l'échantillon.

Per sample details : la répartition des scores et l'état attribué à chaque échantillon sont disponibles dans le panneau de droite.

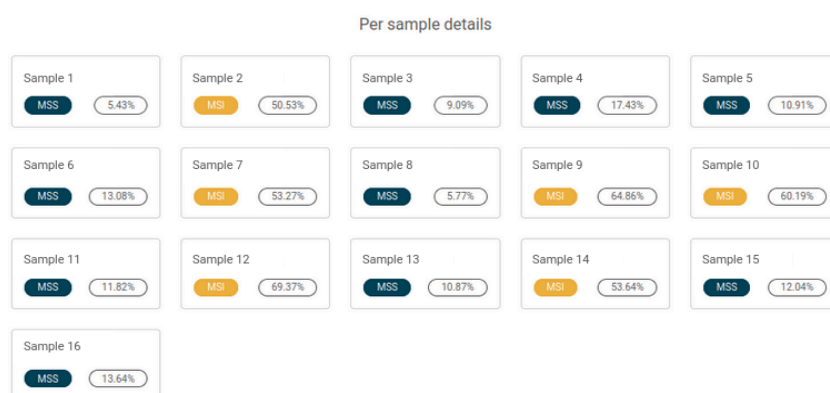


Figure 37 : Répartition des scores et de l'état de tous les échantillons inclus dans l'analyse SomaMSI.

13.1.13. Statut HRD/HRP

Methodologie

La détection du statut de HRD proposée par SeqOne est basée sur l'analyse des données de séquençage du génome à faible couverture (sWGS), auxquelles peuvent être ajoutés les résultats d'une analyse de panel de gènes, si ce dernier inclut les gènes BRCA1 et BRCA2.

Le principe est le suivant :

- Quantifier l'instabilité génomique à partir des données sWGS et identifier les grands réarrangements (LGA) ainsi que les pertes de copies parentales (LPC), qui reflètent une perte d'hétérozygotie.
- Identifier les variations du nombre de copies (CNV) dans les gènes clés décrits comme pouvant affecter le statut de HRD.
- Détecter éventuellement les variants pathogènes dans les gènes BRCA1, BRCA2 et TP53, si les résultats du panel sont disponibles sur l'interface SeqOne pour le patient correspondant. Dans ce cas, les types de variants détectés par le workset SomaVar qui sont pris en compte sont les SNV, les indels et les SV (détectés par Freebayes, Mutect2 et GRIDSS).

HRD status. Le statut HRD positif est défini par au moins un des éléments suivants (**versions SomaHRD v1.2 et v1.3 uniquement**) : (1) au moins une mutation pathogène BRCA1/2 et/ou (2) une probabilité de HRD supérieure ou égale à 0,5.

Depuis la version v1.4 du workset SomaHRD, le statut de l'analyse est défini uniquement par le score d'instabilité génomique (score GI). Les utilisateurs peuvent sélectionner un statut HRD (*Positive, Negative, ou Non-conclusive*) à côté du statut GI pour l'inclure dans le rapport à la place du statut GI.

HRD probability (versions SomaHRD v1.2 et v1.3 uniquement). La probabilité HRD de SeqOne correspond à la probabilité que la tumeur soit HRD positive sur la base des caractéristiques d'instabilité génomique LGA et LPC déterminées à partir des données sWGS. Les grandes altérations génomiques (LGA) sont définies comme le nombre de copies des points de rupture, un point de rupture correspondant à un changement du nombre de copies entre deux segments génomiques d'au moins 10 Mb et séparés de 3 Mb au maximum. Les pertes de copies parentales (LPC) sont définies comme le nombre de segments haploïdes d'une longueur d'au moins 10 Mb. La probabilité de HRD est une combinaison linéaire des deux caractéristiques d'instabilité génomique, convertie à l'aide d'une régression logistique pour placer les valeurs sur une échelle de probabilité entre 0 et 1.

GI probability (GiP) (à partir de SomaHRD v1.4). La probabilité d'instabilité génomique (IG) représente la probabilité que la tumeur présente une instabilité génomique, sur la base du séquençage peu profond du génome entier (sWGS). Le calcul de cette probabilité dépend des deux types de caractéristiques d'instabilité génomique LGA et LPC. La LGA est définie comme le nombre de points de rupture du nombre de copies. Un point de rupture correspond à un changement du nombre de copies entre deux segments génomiques suffisamment grands et proches l'un de l'autre. Le score LGA n'est pas une mesure unique, mais plutôt un ensemble de plusieurs variations. Ces variations sont calculées en utilisant différents paramètres pour la taille minimale des fragments et la distance entre les fragments. Le score LPC, quant à lui, se réfère au nombre de segments haploïdes d'au moins 10 Mb de long. La probabilité d'instabilité génomique de SeqOne n'est pas une fonction directe des scores LGA et LPC moyens. Elle est calculée à l'aide d'un modèle de renforcement qui intègre plusieurs caractéristiques LGA et LPC, améliorant ainsi la précision de la prédiction. La probabilité finale est la moyenne de 100 probabilités (allant de 0 à 1) calculées en appliquant différents paramètres du modèle.

GI status (à partir de SomaHRD v1.4). Un statut d'instabilité génomique positif (statut IG) est défini par une probabilité d'instabilité génomique supérieure à 50 %.

BRCA Status. Pour la détection des variants pathogènes, la détermination du statut BRCA indépendamment de la probabilité HRD ou de la probabilité GI, les critères suivants sont appliqués aux variants BRCA1 et BRCA2 à partir de l'analyse SomaVar du patient :

- Fréquence des allèles $\geq 15\%$ (**versions SomaHRD v1.2 et v1.3**),
- Fréquence des allèles $\geq 10\%$ (**d'après SomaHRD v1.4**)
- Couverture ≥ 35
- Qualité moyenne de chaque base portant le variant (score Phred) ≥ 20
- Observation alternative (nombre de séquences portant la mutation) ≥ 5
- Classification 4 (*probablement pathogène*) ou 5 (*pathogène*) selon notre outil de classification de la pathogénicité des variants basé sur les critères de l'ACMG.

TP53 variants. Les critères suivants sont appliqués aux variants pathogènes TP53 issus de l'analyse SomaVar effectuée sur le patient :

- Couverture ≥ 100
- Qualité moyenne de chaque base portant le variant (score Phred) ≥ 20
- Observation alternative (nombre de séquences portant la mutation) ≥ 5
- Classification 4 (*probablement pathogène*) ou 5 (*pathogène*) selon notre outil de classification de la pathogénicité des variants basé sur les critères de l'ACMG.

Ces variants sont uniquement affichés sans seuil minimum de fréquence d'allèle dans le rapport d'analyse SomaHRD, mais ne sont pas déterminants pour le statut HRD. Seul le variant pathogène TP53 ayant la fréquence allélique la plus élevée sera signalé dans le rapport HRD.

Cellularity. Les métadonnées de cellularité (% de cellules tumorales dans l'échantillon) sont nécessaires pour lancer l'analyse SomaHRD. Elles peuvent être saisies lors du chargement de l'échantillon ou avant le lancement de l'analyse, en modifiant les métadonnées de l'échantillon correspondant (voir **Métadonnées de l'échantillon**). Si la valeur est inférieure à 20 %, le résultat ne sera pas concluant. Cependant, le statut HRD peut être positif si l'analyse est accompagnée d'un panel comprenant BRCA1 et BRCA2, qui donne un statut BRCA positif.

Tumor Fraction Warning Procedure : L'analyse SomaHRD comprend un mécanisme d'alerte pour indiquer quand un échantillon peut avoir un faible contenu tumoral. Ce mécanisme est basé sur une mesure statistique appelée "statistique de chevauchement", qui évalue la similarité entre des segments génomiques consécutifs en termes de rapport du nombre de copies. Les échantillons présentant un chevauchement important ont souvent des fractions tumorales plus faibles, ce qui rend plus difficile la distinction de l'instabilité génomique.

Afin d'établir un seuil fiable pour cette alerte, nous avons corrélié la statistique de chevauchement avec la fréquence des allèles variables (VAF) des mutations TP53 à l'aide de données cliniques. Une alerte est déclenchée si la statistique de chevauchement dépasse un seuil prédéfini. Cette alerte met en évidence les divergences potentielles avec le pourcentage de contenu tumoral fourni manuellement, qui peut indiquer que la fraction tumorale est inférieure à 20 %. Les échantillons dont la fraction tumorale est inférieure à ce niveau présentent un risque élevé de produire des faux négatifs. Toutefois, ces échantillons ne sont pas rejetés et l'alerte sert de guide pour une interprétation prudente des résultats de la HRD.

Minimum coverage. La couverture minimale requise pour exécuter le test est de 0,1X, mais nous recommandons une couverture minimale de 0,5X, idéalement de 1X. Lorsque la couverture est réduite à 0,1X, le taux d'erreur de classification augmente de 5 %.

Bien que les données sWGS puissent avoir été générées à partir d'un kit contenant des UMI, celles-ci ne sont pas prises en compte lors de l'analyse SomaHRD. Une étape d'élagage pour les adaptateurs et les UMI est incluse dans le workset.

Visualisation et interprétation

Une fois l'analyse terminée, les résultats sont disponibles dans l'analyse SomaHRD. En plus des métriques QC et du rapport technique, un rapport HRD est directement accessible depuis l'onglet Rapport HRD (**versions SomaHRD v1.2 et v1.3**).

Depuis la **version v1.4 du workset SomaHRD**, les résultats de l'analyse sont affichés dans un onglet HRD dédié. Chaque section de l'onglet reflète une section du rapport HRD décrit ci-dessous (Résumé HRD, Qualité de l'échantillon, Caractéristiques d'instabilité génomique, Mutations BRCA1/BRCA2 & TP53, Variations du nombre de copies).

A partir de cet onglet, l'utilisateur peut exporter ces résultats sous la forme d'un fichier CSV ou d'un rapport HRD au format PDF. Dans le cas d'une exportation au format PDF, l'utilisateur peut sélectionner la langue du rapport HRD dans une liste de langues disponibles : Français, Espagnol, Anglais, Allemand, Italien. La sélection par défaut est l'anglais.

Le rapport sur le développement des ressources humaines est divisé en plusieurs sections.

Un **en-tête** énumère les métadonnées du patient et de l'analyse, le type de test effectué et le statut HRD (**SomaHRD v1.2 & v1.3**) ou le statut GI (**SomaHRD ≥ v1.4**).

Sample quality. Cette section résume les principales mesures de qualité de l'analyse ainsi que les seuils minimaux requis.

Les contrôles QC sont affichés sous forme de tableau. Pour chaque paramètre, les informations suivantes sont disponibles :

- **Name** : le nom du paramètre
- **Value** : la valeur du paramètre, y compris :
 - Une icône indique si le filtre est *réussi*, *averti* ou *échoué*.
 - La valeur du paramètre pour l'échantillon
- **Description** : brève description du paramètre

SAMPLE QUALITY

NAME	VALUE	DESCRIPTION
% of tumoral cells	✓ 90%	Low value of tumoral cells is likely to generate false negative results
Coverage	⚠ 0.43X	Average sequencing coverage of the genome
% correct mapping	✓ 96.5%	Proportion of reads correctly mapped
Confidence in genomic instability	✓ 1.00	Large values indicates high confidence in genomic instability computation
Low tumor fraction	✓ NORMAL	Low tumor fraction detected from the shallow WGS

SAMPLE QUALITY

NAME	VALUE	DESCRIPTION
% of tumoral cells	⚠ 30%	A percentage of tumoral cells below 20% is likely to generate false negative results.
Coverage	✓ 0.7X	The average sequencing coverage of the genome should be above 0.1x.
GI Dispersion	✓ 0.93	A large dispersion reflects high variability in the genomic instability computation, indicating low precision in the estimated score.
Low tumor profile	⚠ WARNING	Low cellularity detected from the genomic profile.

Figure 38 : Résumé des mesures de contrôle qualité relatives à l'échantillon sWGS dans SomaHRD v1.2 (en haut) et SomaHRD v1.5 (en bas).

Les contrôles de qualité sont appliqués aux paramètres suivants, en fonction des seuils présentés à la fin du rapport HRD (section **SAMPLE QUALITY CHECKS**) :

- le pourcentage de cellules tumorales, sur la seule base de la valeur fournie par l'utilisateur.
- la couverture moyenne du séquençage du génome,
- proportion de reads correctement alignés, calculée après l'étape de découpage des adaptateurs (**versions SomaHRD v1.2 et v1.3 uniquement**),
- la métrique de confiance dans l'instabilité génomique (LGA et LPC) basée sur les données de couverture (**version SomaHRD v1.2 uniquement**). Ce contrôle de qualité est utilisé pour mesurer la confiance dans le statut HRD renvoyé. Certains résultats positifs et négatifs sont marqués d'une étiquette "low confidence" pour identifier les échantillons dont le statut est le moins fiable.
- un intervalle de confiance de la probabilité HRD (*intervalle de confiance de la probabilité HRD*) (**version SomaHRD v1.3 uniquement**). Cette métrique aura un statut d'avertissement si l'erreur standard (écart-type) est supérieure à 0,07 ; elle aura un statut d'échec si l'intervalle de confiance de la probabilité HRD inclut la valeur 0,5.
- une dispersion GI (**version SomaHRD v1.5 uniquement**) qui indique la dispersion des 100 calculs de probabilité GI. Cet indicateur affichera un statut « Warning » avec un tag « Moderate » à côté de la valeur lorsque la dispersion est > 2,6 et un statut « Failed » avec un tag « High » si elle est > 2,7. Sinon, un tag « Low » apparaît à côté du score.
- en cas de suspicion de faible contenu tumoral (*Low tumoral fraction*) sur la base du swGS, le message **WARNING** s'affiche. Dans le cas contraire, le contrôle affiche le message **NORMAL**.

Si l'une des mesures QC suivantes est inférieure aux seuils de défaillance prévus, le résultat de l'analyse sera non concluant (NON CONCLUSIVE) :

- % de cellules tumorales
- couverture
- % de reads correctement alignées (**uniquement pour les versions SomaHRD v1.2 et v1.3**).
- Dispersion GI > 2,7 **si également** $0,37 < GiP < 0,59$ (**uniquement pour SomaHRD v1.5**)

Dans ce cas, les mesures de confiance dans l'instabilité génomique et la faible fraction tumorale ne sont pas calculées et affichent N/A dans leurs sections respectives (pour **SomaHRD version < v1.5**).

Pour SomaHRD v1.3, si la confiance dans la mesure de l'instabilité génomique est égale à Échec, c'est-à-dire inférieure au seuil d'échec prévu, le résultat de l'analyse sera :

- POSITIVE LOW CONFIDENCE si la probabilité de HRD est supérieure ou égale à 50 %.
- NEGATIVE LOW CONFIDENCE si la probabilité de HRD est inférieure à 50 %.

Dans ce cas, le résultat de l'analyse est accompagné d'un avertissement.

Néanmoins, si l'analyse est accompagnée d'un panel comprenant BRCA1 et BRCA2, et que le panel renvoie un statut BRCA positif, le statut HRD sera positif, même si une métrique QC relative à l'échantillon sWGS est inférieure au seuil d'échec (**uniquement pour les versions SomaHRD v1.2 et v1.3**).

Copy number variations. Cette section est consacrée aux variations du nombre de copies des gènes CCNE1 et RAD51 en termes de nombre de copies détectées. Ces variations ne sont pas prises en compte dans le calcul de la probabilité du HRD et de la probabilité de l'IG.

Genomic instability features. Cette section fournit des détails sur les caractéristiques qui ont été calculées pour déterminer l'instabilité génomique de la tumeur. Ces caractéristiques sont utilisées pour déterminer la probabilité HRD de SeqOne (**SomaHRD v1.2 et v1.3**) et la probabilité GI (**SomaHRD version $\geq$ v1.4**). Cette section a pour but de faire comprendre comment la probabilité est calculée. Pour chacune des caractéristiques, les informations suivantes sont fournies :

- **Name** : Nom de la caractéristique (par ex. LGA Status ou LGA Score)
- **Value** : La valeur qui a été calculée pour le patient
- **Description** : Brève description des caractéristiques

En cas de résultats *NON CONCLUSIVE*, liés à des mesures de contrôle qualité inférieures aux seuils attendus, les sections *AMPLIFICATIONS* et *GENOMIC INSTABILITY FEATURES* du rapport apparaissent en grisé.

GENOMIC INSTABILITY FEATURES

The following variables are used to determine the HRD Probability.

NAME	VALUE	DESCRIPTION
LGA Status	23	Large Genomic Alteration Status
LPC Status	15	Loss of Parental Copy Status

Figure 39 : Caractéristiques de la probabilité de la maladie d'Alzheimer liées à l'instabilité génomique.

BRCA1/BRCA2 & TP53 mutations. Dans le cas d'un test HRD complet comprenant une analyse SomaVar, nous fournissons une section supplémentaire consacrée aux variants pathogènes détectés dans les gènes BRCA1, BRCA2 et TP53. Tous les variants pathogènes BRCA1 et BRCA2 correspondant aux critères décrits dans la section **Methodology** sont répertoriés. Toutefois, seul le variant TP53 pathogène ou probablement pathogène ayant la fréquence d'allèle la plus élevée sera signalé. Les autres variants pathogènes de TP53 peuvent être

consultés dans l'analyse SomaVar associée. Cette section est masquée pour les rapports sWGS uniquement.

Les mutations détectées sont reportées dans un tableau comportant les colonnes suivantes :

- **Gene & Transcript** : Nom du gène HGNC et ID de le transcrit RefSeq
- **Variant** : nomenclatures HGVSC et HGVSP du variant
- **Evaluation** : Évaluation du variant selon les critères de l'ACMG (*pathogène* ou *probablement pathogène*)
- **VAF** : fraction allélique du variant

Si aucune mutation pathogène du gène BRCA n'a été identifiée dans l'analyse du panel lié, cette section du rapport affiche le message suivant : "Aucune mutation pathogène n'a été identifiée sur les gènes BRCA1 et BRCA2".

Si aucune mutation pathogène du gène TP53 n'a été identifiée dans l'analyse du panel lié, cette section du rapport affiche le message suivant : "Aucune mutation pathogène n'a été identifiée sur le gène TP53".

BRCA1/BRCA2 & TP53 MUTATIONS

GENE & TRANSCRIPT	VARIANT	EVALUATION	VAF
BRCA1 NM_007294.4	c.3116_3117del p.Ala1039GlufsTer6	Pathogenic	60.68%
BRCA1 NM_007294.4	c.3113_3117delinsG p.Glu1038GlyfsTer9	Pathogenic	58.18%
BRCA1 NM_007294.4	c.3113_3114del p.Glu1038GlyfsTer7	Pathogenic	60.08%
TP53 NM_000546.6	c.524G>A p.Arg175His	Pathogenic	39.29%

Figure 40 : Variants pathogènes BRCA1/BRCA2 et TP53.

HRD summary. Cette dernière section du rapport est consacrée aux résultats. Cette section vise à fournir une vue détaillée des résultats, en affichant les informations suivantes :

- **HRD Probability (SomaHRD v1.2 et v1.3)** : La valeur de la probabilité de HRD, comprise entre 0 et 1. Le statut est positif lorsque la valeur est supérieure ou égale à 0,5.
- **GI probability (SomaHRD version ≥ v1.4)** : La valeur de la probabilité d'IG, comprise entre 0 et 1. Le statut est positif lorsque cette valeur est supérieure ou égale à 0,5.
- **BRCA Status** : statut déterminé par la détection d'au moins une mutation BRCA pathogène : MUTÉE, TYPE SAUVAGE, ou S/O en l'absence d'analyse du panel.
- **HRD Status or GI Status** : POSITIVE, NEGATIVE, NEGATIVE LOW CONFIDENCE, POSITIVE LOW CONFIDENCE ou NON CONCLUSIVE tel qu'affiché dans l'en-tête du rapport.
- **Method** : La méthode (ou mode) qui a été utilisée pour produire le score HRD :
 - HRD complet ⇒ "WGS SHALLOW + BRCA1/2 GENE PANEL" (*panel de gènes*)
 - sWGS uniquement ⇒ "SHALLOW WGS ONLY"

En cas de résultats *NEGATIVE LOW CONFIDENCE*, *POSITIVE LOW CONFIDENCE* ou *CONCLUSIVE*, ceux-ci sont affichés en gris clair.

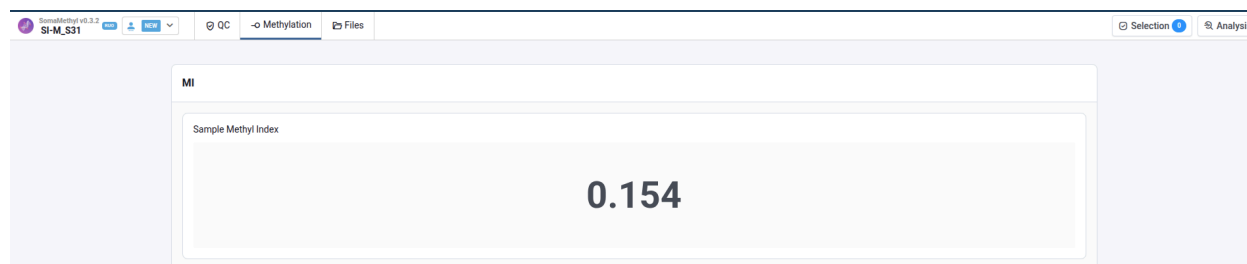
Les détails concernant les différentes valeurs affichées dans le rapport, ainsi que la méthodologie générale du test, sont présentés dans les sections **Footnotes** et **Methodology**, situées à la suite du résultat.

En plus du rapport HRD, les graphiques d'instabilité génomique associés à l'analyse HRD peuvent être téléchargés à partir de l'onglet *Files* de l'analyse SomaHRD. Ces graphiques représentent la couverture normalisée sur tous les chromosomes à partir des données sWGS.

En cas de résultats *NON CONCLUSIVE*, liés à des paramètres de contrôle qualité inférieurs aux seuils prévus, les graphiques d'instabilité génomique ne sont pas générés.

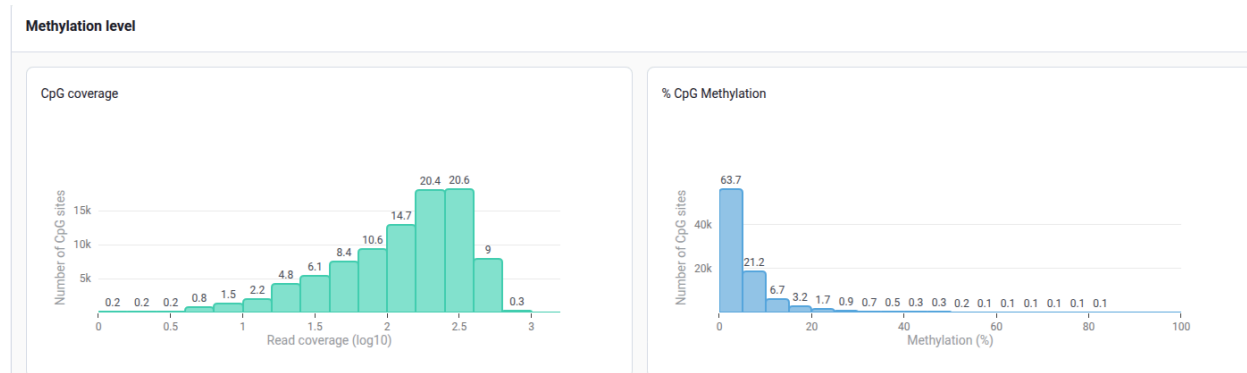
13.1.14. Données de méthylation pour SomaMethyl

Methylation Index (Indice de méthylation)



L'indice de méthylation (MI) est un score numérique unique qui quantifie l'état de méthylation global d'un échantillon biologique et résume le signal de méthylation associé à la tumeur. Il peut détecter de manière sensible même de faibles niveaux d'ADN dérivé de la tumeur dans un bruit de fond d'ADN normal. Dans SomaMethyl, le MI est calculé en mode Standard à l'aide de l'algorithme exclusif d'Agilent (MethylAlgo). Lorsqu'une cohorte de référence est disponible, le MI de l'échantillon peut être comparé aux valeurs MI de la cohorte et interprété à l'aide d'un seuil défini par l'utilisateur. Pour évaluer le MI, Agilent a fourni des échantillons de titrage avec des niveaux de méthylation attendus. Les valeurs MI calculées par le module SomaMethyl en mode Standard sont comparées aux valeurs MI attendues fournies par Agilent, et les performances sont résumées à l'aide d'une métrique de concordance pondérée définie comme $1 - \text{wMAPE}$ (erreur absolue moyenne en pourcentage pondérée).

Methylation level (Niveau de méthylation)



CpG coverage (histogram) (Couverture CpG (histogramme))

Une distribution de la profondeur de lecture sur les sites CpG dans les régions ciblées (manifeste).

Axe X (couverture de lecture log10) : couverture par site CpG (nombre de lectures soutenant l'appel de méthylation à ce CpG, c'est-à-dire les comptes méthylés + non méthylés).

Axe Y (nombre de sites CpG) : nombre de sites CpG dans le manifeste ayant cette couverture.
Vous pouvez vérifier si les CpG ciblés sont suffisamment couverts et si la couverture est assez uniforme pour interpréter les pourcentages de méthylation avec confiance.

% CpG methylation (histogram) (% de méthylation CpG (histogramme))

Une distribution des niveaux de méthylation sur les sites CpG dans les régions ciblées (manifeste).

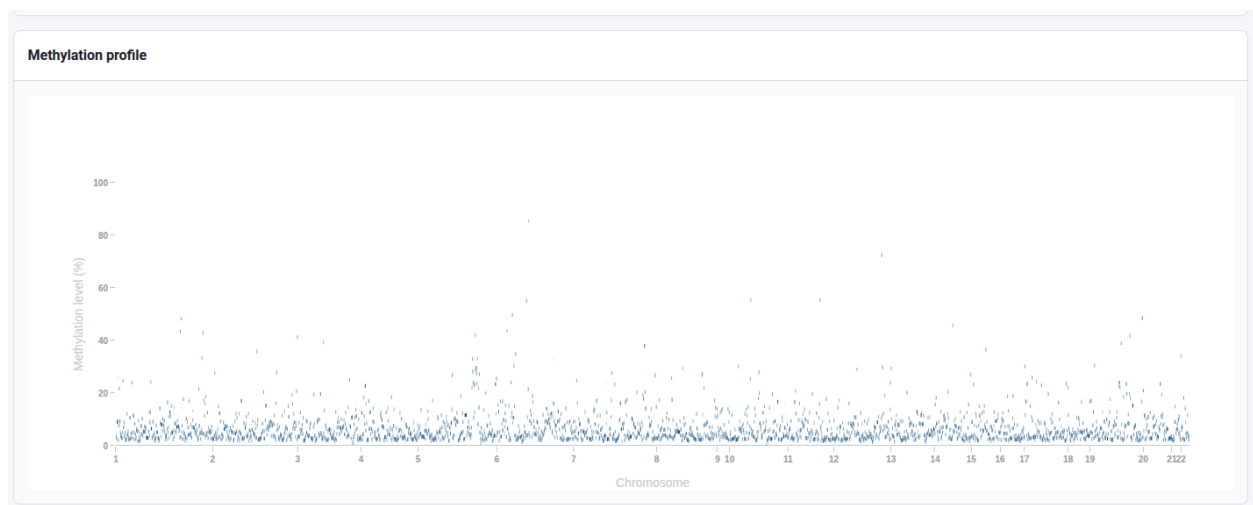
Axe X (% de méthylation) : pour chaque site CpG, la fraction de lectures méthylées parmi toutes les lectures couvrant ce CpG.

Axe Y (nombre de sites CpG) : nombre de sites CpG dans le manifeste montrant ce niveau de méthylation.

Il fournit une vue au niveau de l'échantillon des états de méthylation CpG à travers les régions ciblées (montrant souvent des pics près d'une méthylation faible et élevée, selon les régions capturées et la biologie de l'échantillon).

Les sites CpG avec une couverture très faible produisent des pourcentages de méthylation moins fiables car un petit nombre de lectures peut faire basculer le % vers le haut ou vers le bas. Pour cette raison, l'histogramme de % de méthylation CpG doit toujours être interprété avec l'histogramme de couverture CpG (et les autres métriques de contrôle qualité).

Methylation profile (Profil de méthylation)



Le graphique du profil de méthylation montre le signal de méthylation CpG de l'échantillon à travers le génome, affiché par chromosome. Chaque point correspond au pourcentage moyen de méthylation CpG calculé sur des fenêtres génomiques successives (comme dans la sortie .mcpG.track.bedGraph). Cette vue aide à visualiser les grands modèles de méthylation et à identifier les régions avec une méthylation plus ou moins élevée par rapport au reste du génome.

Methylation Table (Tableau de méthylation)

Table					
			Search genes		Locus
REGION	METH. LEVEL	CpG NBR.	CpG REGIONS	GENOMIC FEATURE	REFSEQ CDS
C1orf94 chr1:34176744-34177348	11	53	Island +1	silencer	
chr1:35577191-35577675	3	42			
chr1:37033668-37034149	5	51			
chr1:37034542-37035023	3	61			
EPHA10 chr1:37735070-37735507	24	26		enhancer	NM_001099439.2
chr1:38559361-38559798	5	22			
chr1:39080823-39081222	12	28	Island +1		
chr1:39672120-39672554	3	28			
HPCAL4 chr1:39683853-39684273	11	47	Island		
chr1:40884160-40884512	5	43			
31 to 40 of 2217					

Le tableau de méthylation liste les régions génomiques (chromosome:début-fin) sur lesquelles SomaMethyl résume la méthylation CpG (fenêtres fixes ou segments fusionnés, selon l'étape). Pour chaque région, METH. LEVEL rapporte le niveau de méthylation agrégé calculé à partir des appels de méthylation CpG dans cet intervalle, et CpG NBR. indique combien de sites CpG ont contribué au calcul. Chaque ligne est ensuite annotée avec des CpG REGIONS (îlot, rivage, plateau le cas échéant) et des annotations fonctionnelles, y compris des éléments régulateurs de GENOMIC FEATURE chevauchants (promoteur, amplificateur, silencieux) et des informations REFSEQ CDS (séquence codante).

14. Rapport

14.1.1. Fichiers de sortie

L'onglet *Files* répertorie les fichiers générés aux différentes étapes du workset. Chaque fichier peut être téléchargé en cliquant sur l'icône  .

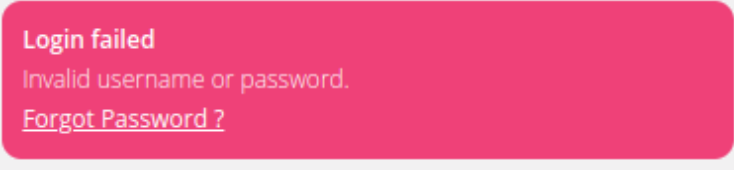
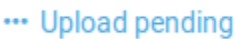
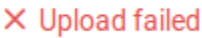
Certains fichiers ne comportent pas le nom de l'échantillon comme préfixe. Pour les télécharger avec ce préfixe, vous devez activer l'option "*Prefix with sample name*" en haut à droite de la page.

Les fichiers peuvent être téléchargés à partir des différentes sections de l'interface :

- Les fichiers FastQ sont disponibles dans l'onglet *Samples* du projet en sélectionnant l'échantillon concerné, puis l'icône  .
- Les manifestes (fichiers BED) sont disponibles dans les *paramètres de l'utilisateur* via l'icône située dans le coin supérieur droit du tableau de bord, puis dans l'onglet *Manifest*.
- Le(s) rapport(s) d'analyse au format PDF est/sont disponible(s) dans l'onglet *Reporting* après avoir sélectionné une analyse.
- Les fichiers générés lors d'une analyse (BAM, VCF, etc.) et les différents rapports HTML (qualité, technique, etc.) sont disponibles dans l'onglet *Fichiers* pour chaque analyse.

15. Dépannage et messages d'erreur

Plusieurs messages sont utilisés pour informer les utilisateurs d'une erreur

Message d'erreur	Description
 Login failed Invalid username or password. Forgot Password ?	Le mot de passe et/ou le login est/sont incorrect(s). Les utilisateurs sont invités à saisir à nouveau leur login et leur mot de passe, ou à cliquer sur "Forgot Password" pour réinitialiser leur mot de passe.
 ... Upload pending	Le téléchargement de l'échantillon est en cours. Il est conseillé aux utilisateurs d'attendre, de ne pas fermer la fenêtre du navigateur et de ne pas couper leur connexion internet.
 × Upload failed	L'échantillon n'a pas pu être téléchargé. <i>Cette erreur peut se produire si la fenêtre du navigateur a été fermée ou si la connexion internet a été interrompue pendant le téléchargement.</i> Les utilisateurs sont invités à réessayer de télécharger le(s) fichier(s).
 62%	Indique la progression du téléchargement en cours Il est conseillé aux utilisateurs d'attendre, de ne pas fermer la fenêtre du navigateur et de ne pas couper leur connexion internet.

	Chargement de la page. Les utilisateurs sont invités à patienter.
... Bioinformatics - <i>pending</i>	L'analyse est en cours. Les utilisateurs sont invités à patienter.
× Failed - <i>error</i>	L'analyse a échoué. Les utilisateurs sont invités à vérifier les paramètres et les données utilisés pour exécuter l'analyse, puis à redémarrer l'analyse.
⚠ Something went wrong Something went wrong, you should refresh this page before continue. Feel free to contact the support if the problem persist.	Une erreur s'est produite. L'utilisateur est invité à cliquer sur la touche "Esc". Si le problème persiste, contactez le service d'assistance à l'adresse support@seqone.com
× Missing fastq files	Les fichiers FastQ associés à cet échantillon sont manquants, ce qui signifie qu'aucune nouvelle analyse ne peut être lancée pour cet échantillon.
unexpected EOF × Verification failed	Le message "Unexpected EOF" (End Of File) correspond à un fichier tronqué. L'utilisateur est invité à vérifier les données et à recharger les fichiers concernés.

Les utilisateurs peuvent contacter l'équipe Support en cliquant sur le logo  en haut à droite du tableau de bord de l'utilisateur, puis sur l'icône  **Support**

16. Garantie

Veuillez vous référer aux conditions de garantie de votre contrat de maintenance.

17. Avis juridique

17.1. Conditions d'utilisation

Les informations relatives à la licence d'utilisation sont accessibles en cliquant sur le logo du compte utilisateur suivi des *Terms of Use*. Les utilisateurs peuvent visualiser et télécharger les informations au format PDF à partir de la fenêtre contextuelle en cliquant sur "*OPEN PDF*".

17.2. Limitation de la responsabilité

SeqOne décline toute responsabilité en cas de mauvaise utilisation ou d'utilisation non professionnelle de la plateforme SeqOne.

SeqOne est un outil d'aide à la décision clinique et n'est pas responsable des diagnostics dérivés des résultats obtenus.

Par conséquent, les utilisateurs doivent faire preuve de prudence dans l'utilisation du logiciel et de ses résultats. Les diagnostics doivent être confirmés par divers examens cliniques et médicaux.

17.3. Droits d'auteur et marques déposées

Tous les noms de sociétés et de produits mentionnés dans ce guide peuvent être des marques déposées par leurs propriétaires respectifs.