

Analyse bio-informatique des données NGS

Campagne 2025

Instructions pour la soumission des fichiers bio-informatiques BAM, BED, et VCF

Les fichiers à soumettre : bed, bam, vcf

Pour l'analyse bio-informatique des données NGS, il est demandé de soumettre les fichiers suivants :

1. Le fichier de type 'bam'

Ce fichier contient l'alignement des reads produits par le séquenceur, alignés sur le génome de référence que vous avez choisi.

Les génomes acceptés sont : GRCh37/hg19 et GRCh38/hg38.

Le fichier 'bam' a un format standard (binaire) et peut être très grand (plusieurs Giga-octets, GB) selon la taille de la cible et la profondeur de lecture, donc aussi plus lent à télécharger.

2. Le fichier de type 'bed'

Il s'agit du fichier bed correspondant à votre test et décrivant votre cible génomique, fourni par le fabricant du kit. Le fichier décrit la cible du test en termes de coordonnées génomiques. Le fichier 'bed' a un format standard de trois colonnes obligatoires, qui sont (1) le chromosome de chaque zone d'intérêt ('feature' en anglais, une ligne dans le fichier), (2) le départ d'une zone (coordonnée sur le chromosome), et (3) la fin d'une zone (coordonnée sur le chromosome). Ce fichier est petit par rapport au fichier 'bam'.

Si vous n'avez pas la possibilité de télécharger votre fichier bed, veuillez télécharger le fichier « [genetiss_roi_hg19.bed](#) » ou le fichier « [genetiss_roi_hg38.bed](#) » en fonction du génome de référence utilisé. Ces fichiers ont été envoyés par courriel en même temps que ces instructions et peuvent aussi être téléchargés à partir du site web gen&tiss.

3. Le fichier de type 'vcf'

Ce fichier contient la liste des variants détectés par le pipeline secondaire sur la base des reads alignés sur le génome de référence que vous avez choisi. Ce fichier a un format bien précis et ne doit pas être modifié, par exemple en l'ouvrant dans un programme tel que Microsoft Excel. Le format accepté est la version 4.1 ou plus, correspondant à un format défini il y a plus de 10

ans et donc généré par tout outil bio-informatique. Ce fichier est petit par rapport au fichier 'bam'.

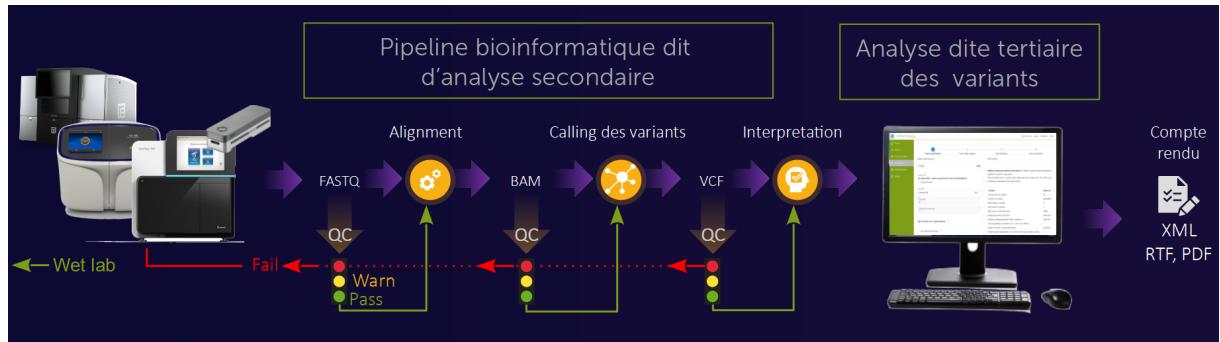


Figure 1: Pipeline typique NGS avec fichiers 'bam' et 'vcf'. Le fichier 'bed' est fourni avec le kit, par la compagnie qui le vend (pas sur l'image).

Le système de soumission

Pour cette campagne, les fichiers sont soumis sur un système pilote avec une interface en anglais. Ce système permet un certain niveau de vérification des fichiers soumis. Tous les fichiers seront reçus et sauvegardés, même si le format n'est pas correct.

Dans le cas d'une erreur de format il y aura un message d'erreur, qui peut être ignoré.

Pour plus de détails vous pouvez vous référer à la section '**Soumission des fichiers**'.

Veuillez noter que le système de soumission présente également d'autres fonctionnalités qui ne sont pas requises dans le cadre de la campagne d'EEQ gen&tiss.

Il est donc recommandé de bien suivre ces instructions.

La présence de ces autres fonctions n'a pas d'influence sur le processus de soumission.

Chaque participant a une zone de travail dédiée pour la soumission et a uniquement accès à cette zone. L'accès à ce système s'effectue via un code utilisateur associé à un mot de passe personnalisé (communiqué individuellement par courriel à chaque laboratoire).

Le système est accessible via l'adresse suivante : <https://www.genetiss.org/module-analyse-bio-informatique-des-donnees/ngs/> puis cliquez sur « Téléchargement des fichiers bio-informatiques »

Sur la page qui s'ouvre, on trouvera l'icône de connexion 'Login' en haut à droite.

Le reste du texte peut être ignoré, il n'est pas lié à l'EEQ gen&tiss.

The quality management system for your clinical NGS workflows

omnomicsQ provides the first line of defence against quality issues at all stages of your laboratory's NGS workflow. Make sure that your NGS data covers your region of interest, adheres to test requirements, and is at the highest quality, every time you run a clinical sample. Keep

Une fois connecté, sélectionner ‘Validation’ à partir du menu principal dans la marge, située à gauche de la page (Fig. 2).
Toutes les autres fonctions de la page peuvent être ignorées.

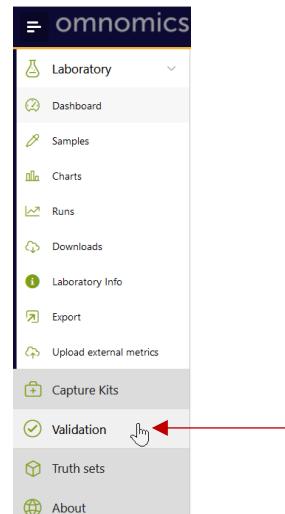


Figure 2: Menu principal après connexion

Soumission des fichiers – description des 4 étapes

La soumission des fichiers à proprement parler est un processus en 4 étapes :

1. Choix du type de processus (constitutionnel ou somatique)
2. Questionnaire, en anglais pour ce pilote
3. Téléchargement des fichiers puis
4. Validation des fichiers.

Le passage à l'étape suivante se fait à chaque fois en pressant sur ‘Next’, en haut à droite ou en bas de page.

Pour ce pilote, les trois premières étapes sont essentielles. Après soumission des fichiers, lorsque vous cliquez sur ‘Next’ pour progresser vers la quatrième étape, les fichiers auront été reçus par le système, même s'il y a un message d'erreur à cette étape 4.

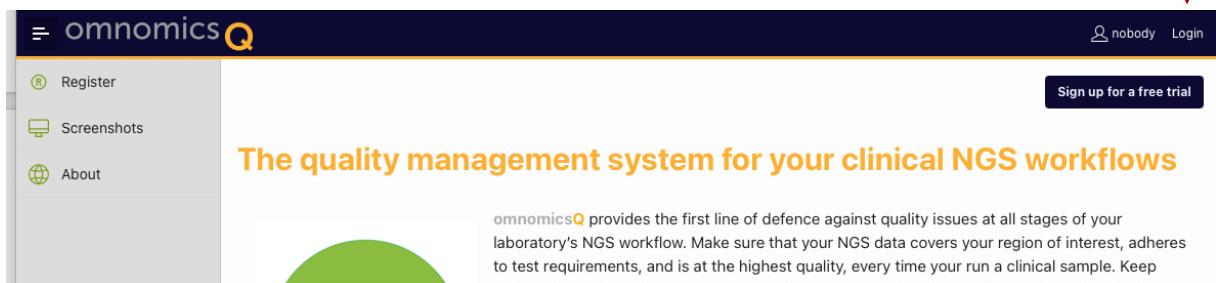
Après l'étape 4, sauf s'il y a erreur de format, le système progresse automatiquement (après environ 10-20 minutes) jusqu'au rapport, qui, sur ce système pilote, n'est qu'une preuve additionnelle que le format des fichiers est correct.

Il y a trois choses à noter lors de la soumission :

1. Le temps requis pour le téléchargement dépend du débit du réseau du participant et de la taille des fichiers. Typiquement, les fichiers ‘vcf’ et ‘bed’ se chargent en quelques secondes. Cependant un grand fichier ‘bam’ peut prendre plusieurs minutes pour être téléchargé. L'interface indique le % de progression du téléchargement.
2. Après avoir cliqué sur ‘Next’ à la fin de l'étape 3, si l'interface passe à l'étape 4, les fichiers ont bien été reçus et on peut se déconnecter.
3. Si toutefois vous restez connecté sur le site à ce stade, il y aura une attente de plusieurs minutes supplémentaires : d'abord 5-15 minutes pour démarrer une phase de contrôle des fichiers soumis, puis le temps nécessaire pour effectuer certaines analyses additionnelles, ceci pouvant aller jusqu'à 20 minutes pour un fichier ‘bam’ de 2GB. Le rapport temporaire affiché sur l'interface après cette attente sert uniquement à prouver que les fichiers soumis ont pu être interprétés. Les résultats, indiqués après l'étape 4 n'ont aucune pertinence à ce stade pour l'évaluation gen&tiss. **Les résultats réels de l'analyse des données soumises seront rapportés ultérieurement par gen&tiss.**

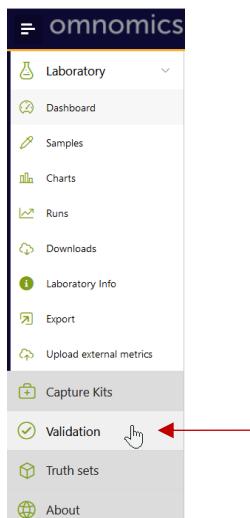
Soumission des fichiers pas à pas

1. Login



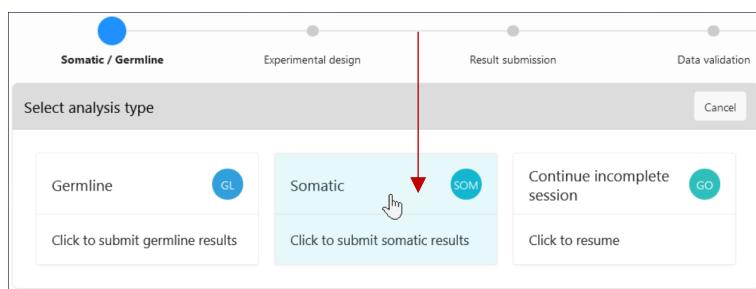
The screenshot shows the omnomicsQ login page. At the top right, there are links for 'nobody' and 'Login'. Below the header, a banner reads 'The quality management system for your clinical NGS workflows'. A green circular icon is visible at the bottom left. The main content area contains a paragraph about the service's purpose: 'omnomicsQ provides the first line of defence against quality issues at all stages of your laboratory's NGS workflow. Make sure that your NGS data covers your region of interest, adheres to test requirements, and is at the highest quality, every time you run a clinical sample. Keep your laboratory compliant with international quality standards.'

2. Sélectionner « Validation »



The screenshot shows the 'Validation' section of the omnomicsQ navigation menu. A red arrow points to the 'Validation' option, which is highlighted with a checkmark icon.

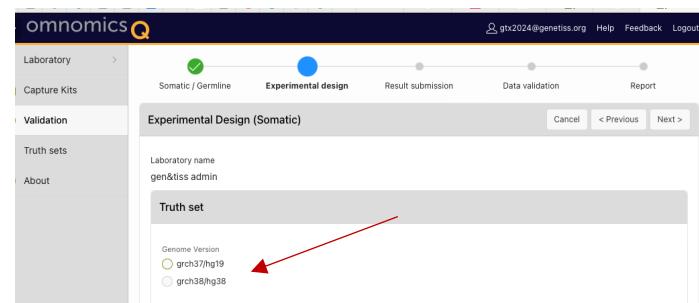
3. Sélectionner « Somatic »



The screenshot shows a dialog box titled 'Select analysis type'. It has four options: 'Germline' (GL), 'Somatic' (SOM), 'Continue incomplete session' (GO), and 'Click to resume'. The 'Somatic' option is highlighted with a blue background and a red arrow points to it. Below each option is a button labeled 'Click to submit [analysis type] results'.

4. Renseigner les champs obligatoires marqués d'une étoile rouge *

a) Choisir la version du génome de référence



The screenshot shows the 'Experimental Design (Somatic)' step of the process. On the left, a sidebar lists 'Laboratory', 'Capture Kits', 'Validation', 'Truth sets', and 'About'. The main form has tabs for 'Somatic / Germline', 'Experimental design', 'Result submission', 'Data validation', and 'Report'. The 'Experimental design' tab is selected. It contains fields for 'Laboratory name' (set to 'gen&tiss admin') and 'Truth set'. Under 'Truth set', there is a dropdown menu for 'Genome Version' with two options: 'grch37/hg19' (selected) and 'grch38/hg38'. A red arrow points to the 'grch38/hg38' option.

b) Choisir « **gen&tiss somatique 2025** » comme matériel de référence

Laboratory >

Somatic / Germline Experimental design Result submission Data validation Report

Experimental Design (Somatic)

Laboratory name: genetiss participant 0007

Truth set

Genome Version: grch37/hg19 (selected)

Reference material used (truth set) *: gen&tiss somatique 2025 (grch37/hg19) (highlighted by a red arrow)

Truth Set Description: For collecting .bam and .vcf submission data, contains HD827

c) Choisir ‘Sample type’

Laboratory >

Somatic / Germline Experimental design Result submission Data validation

Experimental Design (Somatic)

Laboratory name: genetiss participant 0007

Truth set

Genome Version: grch37/hg19 (selected)

Reference material used (truth set) *: gen&tiss somatique 2025 (grch37/hg19)

Truth Set Description: For collecting .bam and .vcf submission data, contains HD827

Sequencing

Sample Type *: ✓ Fresh, Frozen, FFPE, Substandard quality, Fragmented, or kit, Cell free multiplex (highlighted by a red arrow)

d) Choisir le type de librairie (WES, capture, amplification, etc.)

Laboratory >

Capture Kits

Validation

Truth sets

About

Sequencing

Sample Type *: (dropdown menu)

Tissue / DNA source: (dropdown menu)

DNA extraction method or kit: (dropdown menu)

Method of DNA quantitation or quality check after extraction: (dropdown menu)

What type of library approach or protocol did you use? *: ✓ Whole exome capture, Commercial gene panel amplification, Commercial gene panel capture, Custom gene panel capture, Custom gene panel amplification (highlighted by a red arrow)

- e) Choisir le kit utilisé. On peut utiliser la fonction de recherche (loupe) pour scanner parmi les options listées. Si le kit n'est pas enregistré, utiliser 'Unknown', puis indiquer quel kit a été utilisé dans le champ « Notes » en bas de page.

The screenshot shows the 'Capture Kits' section of the omnomics software. The user has entered 'Exome Kit' into the search bar. A dropdown menu below the search bar lists several capture kit options: Exome Kit, Myeloid panel, RUNX1-SF3B1 Custom Panel, SureSelect_custom_01, Unknown Kit, 4bases BRACapanel, and 4bases HEMATOPro. The 'Exome Kit' option is highlighted.

- f) Choisir la plateforme de séquençage, puis le type.

The screenshot shows the continuation of the sequencing setup. The user has selected 'Illumina' from the 'What sequencing platform did you use?' dropdown. Below it, the 'What was your sequencing read type?' section is shown with two radio button options: 'Single End' and 'Paired End'. The 'Single End' option is selected.

- g) Dans la section pipeline bio-informatique, renseigner si possible les trois sections (même si elles ne sont pas obligatoires)

The screenshot shows the 'Bioinformatics pipeline' section. It includes three main sections: 'What analytical/informatics pipeline do you use to perform alignments?', 'What analytical/informatics tool(s) do you use to call variants?', and 'What variant types are you able to detect using your pipeline?'. Under the alignment tools, 'BWA-mem, Novoalign, Bowtie2, TMAP, StrandNGS, Dragen' are listed. Under the variant calling tools, 'GATK, GATK - Unified Genotyper, GATK - Haplotype Caller, VarScan2, Torrent Variant Caller, Mutect2, Dragen' are listed. Under the variant types, there are checkboxes for 'SNP', 'InDel (<45bp)', 'InDel large (>45)', 'Copy Number Variants (CNVs)', and 'Structural variants (SV)'.

- h) Dans la section « Quality targets » sélectionnez « Create new » puis donner un nom à votre SOP - **GTXXXX-MULTIPARAMETRIQUE ou GTXXXX-OVAIRE**

The screenshot shows the 'Quality targets' configuration page. The 'Create new' dropdown is set to 'Create new'. The 'SOP Name' input field contains 'GTXXXX-MULTIPARAMETRIQUE ou GTXXXX-OVAIRE'. The 'Read depth (average) - Warn low' input field contains '300'. The 'Read depth (average) - Fail low' input field contains '200'. The 'Minimum required single base Phred score for considering base call' input field contains '20'.

- i) *Read depth (average) - Warn low* : Renseigner le seuil au-dessus duquel la profondeur moyenne est considérée correcte (vert)- **indiquer 300 comme référence**.
- j) *Read depth (average) - Fail low* : Renseigner le seuil au-dessous duquel la profondeur de moyenne n'est plus considérée acceptable (rouge). Il peut donc y avoir une zone entre (i) et (j) considérée juste suffisante (jaune) - **indiquer 200 comme référence**
- k) Minimum required single base Phred score for considering base call: **indiquer 20 (Q20)**
- l) Tout en bas de la page, dans la section ‘Autre information’ (Other information) indiquez le programme Multiparamétrique ou programme Ovaire. En plus, indiquez le kit s'il n'était pas dans la liste en (e), ou si le kit est fait sur mesure (custom)

The screenshot shows the 'Other information' section. A large text area labeled 'Other information or comments' is present, intended for entering the program name and kit information.

- m) Sauvegarder les données du questionnaire en pressant sur 'Next' en bas à droite (ou en haut à droite)
5. Télécharger les trois fichiers, vcf, bed, bam (**Étape 3**) :

L'ordre n'a pas d'importance, mais les deux premiers sont plus rapidement chargés que le troisième (bam). **Rappel : si vous ne pouvez télécharger votre fichier bed, veuillez utiliser le fichier bed spécifique au génome de référence utilisé, fourni avec ces**

instructions. Une fois chargés, confirmer la soumission en pressant sur ‘Next’. Sauf erreur sur l’interface, les fichiers sont à ce stade sauvegardés dans le système, et l’interface passe à l’étape 4. À ce stade, il est possible de faire une deuxième soumission (autre programme) à partir de (2) ci-dessus, ou de se déconnecter (logout.).

Sample key Somatic-2024-08-25

* Vcf Done Uploaded VCF Sample_8.1.hg38-filtered.vcf.gz

* Target BED file Done Uploaded bed file Twist_merged_target_bases_ONCOSOM-HRR-Total_TE-92360841_GI_hg38.bed

Fasta.gz files No files selected. Uploaded fastq files

Sorted alignment BAM/CRAM Sample_8.1.hg38.bam Done Uploaded bam/cram file Sample_8.1.hg38.bam

Etape 3 : Upload - Les trois fichiers sont chargés sur le système et la procédure se termine par 'Next' qui clos la page et ouvre la suivante sur l’étape 4. A ce stade on peut terminer la procédure par la déconnexion (logout).

File Name	Truth set name	Bed Filename	Status	Added	Started	Ended	Progress	Log output
Sample_8.1.hg38-filtered.vcf.gz	A_2024_b38_Gustave_Roussy	Twist_merged_target_bases_ONCOSOM-HRR-Total_TE-92360841_GI_hg38.bed	WAITING	25-AUG-2024 21:54:31	-	-	[Progress Bar]	-
Sample_8.1.hg38.bam	A_2024_b38_Gustave_Roussy	Twist_merged_target_bases_ONCOSOM-HRR-Total_TE-92360841_GI_hg38.bed	WAITING	25-AUG-2024 21:54:31	-	-	[Progress Bar]	-

This page will proceed to the report page automatically in 30 seconds after all files have been validated successfully.

Etape4 :

Confirmation - Après une attente variant de 10 à 30 minutes, cette page va progresser vers une dernière pagre 'Report', qui à ce stade n'a pas de pertinence, mais en soi confirme que tous les fichiers soumis ont pu être lus.

Notes

En cas de besoin, p.ex. la soumission de faux fichiers, la procédure peut être reprise depuis le début et une nouvelle soumission sera enregistrée. **S'il y a reprise, la section (h) ci-dessus aura déjà enregistré le nom de la SOP et il peut donc être choisi directement au lieu de 'create new'. Attention à ne pas oublier de renseigner le programme Multiparamétrique ou Ovaire dans 'Autre information', voir point 4.I) ci-dessus.**

Le système ne permet pas d'ouvrir des onglets parallèles.

Le système effectue un auto-logout au bout d'environ 1h s'il n'y a pas d'activité sur le site. L'interface peut alors générer une erreur de type 'page pas trouvée'. Il est donc recommandé de télécharger d'un trait, ou de faire un nouveau login.

Le chargement du fichier bam peut durer des minutes. Le système indique combien a été chargé, ce qui permet d'estimer le temps restant jusqu'au transfert complet.

Après le chargement complet à l'étape 3, terminer en pressant sur ‘Next’. Ensuite il n'est plus nécessaire d'attendre, on peut se déconnecter. Si l'on attend la suite, cela peut prendre jusqu'à environ 30 minutes selon la taille des fichiers soumis.