

Analyse des VNC germinaux à l'aide du flux de travail Illumina DNA Prep with Exome 2.5 Enrichment

Détection efficace des variants
du nombre de copies à l'aide
d'un panel de référence



Introduction

Les variants du nombre de copies (VNC) sont une source majeure de diversité génétique chez l'humain¹. Cette classe de variation génomique se produit lorsque le nombre de copies d'un gène ou d'une région génomique varie d'un individu à l'autre et comprend de grandes duplications ou suppressions génomiques. L'analyse des VNC peut aider les chercheurs à identifier les variants associés aux caractéristiques complexes ou à la susceptibilité aux maladies, telles que le cancer, les maladies auto-immunes, les troubles génétiques héréditaires et bien plus encore¹⁻⁴.

Par rapport à l'utilisation de puces à ADN pour détecter des VNC, les méthodes de séquençage de nouvelle génération (SNG) révèlent plus de détails sur les variations structurelles. En outre, le SNG peut détecter certains variants < 50 kb que les puces à ADN n'arrivent pas à faire. Illumina DNA Prep with Exome 2.5 Enrichment fait partie d'une solution haute performance, rapide et fiable de séquençage de l'exome humain entier (WES, Whole-Exome Sequencing) qui génère des bibliothèques compatibles avec l'analyse des VNC par SNG. Combinée aux systèmes de séquençage et aux pipelines d'analyse d'Illumina, cette trousse de préparation de bibliothèques peut être utilisée pour détecter les variants mononucléotidiques (SNV, Single Nucleotide Variant), les petites insertions et suppressions (indels) et les VNC structurels de grande taille.

Cette note d'application démontre un flux de travail rationalisé de séquençage de l'exome pour analyser des VNC germinaux, de la préparation des bibliothèques aux renseignements (figure 1). La solution de WES intègre Illumina DNA Prep with Exome 2.5 Enrichment, le SNG éprouvé d'Illumina et l'analyse secondaire très précise de DRAGEN[™] pour l'identification des VNC.

Les utilisateurs peuvent également utiliser la plateforme de recherche et d'interprétation des variants Emedgene[™], activée par l'intelligence artificielle explicable (XAI, Explainable Artificial Intelligence). Le logiciel Emedgene permet l'analyse des variants et l'obtention de renseignements sur eux, la conservation et l'automatisation de la procédure normale d'exploitation (PNE) et comprend des options rationalisées pour la génération de rapports de recherche.

Méthodes

Échantillons

Pour permettre la détection des VNC germinaux dans l'exome, les utilisateurs doivent développer un fichier de référence à partir d'un panel d'échantillons supposément normaux à des fins de comparaison pendant l'analyse. Ce panel de référence d'échantillons normaux est utilisé pour normaliser le dénombrement de cibles et permet une détection précise des VNC pendant l'évaluation des échantillons pour analyse. Les échantillons du panel de référence doivent être préparés et séquencés dans les mêmes conditions que les échantillons pour analyse.

Pour créer le panel de référence, 54 échantillons (Coriell Institute for Medical Research) sans VNC pathogènes connus (tableau 1) ont été sélectionnés pour représenter la diversité de la variation naturelle entre les superpopulations humaines (basée sur le projet 1000 Genomes Project, tableau 2)⁵. Pour démontrer la performance du flux de travail du WES pour la détection des VNC, 17 échantillons contenant un VNC détecté précédemment dans un gène ou une région d'intérêt ont été utilisés dans une analyse avec le panel de référence.

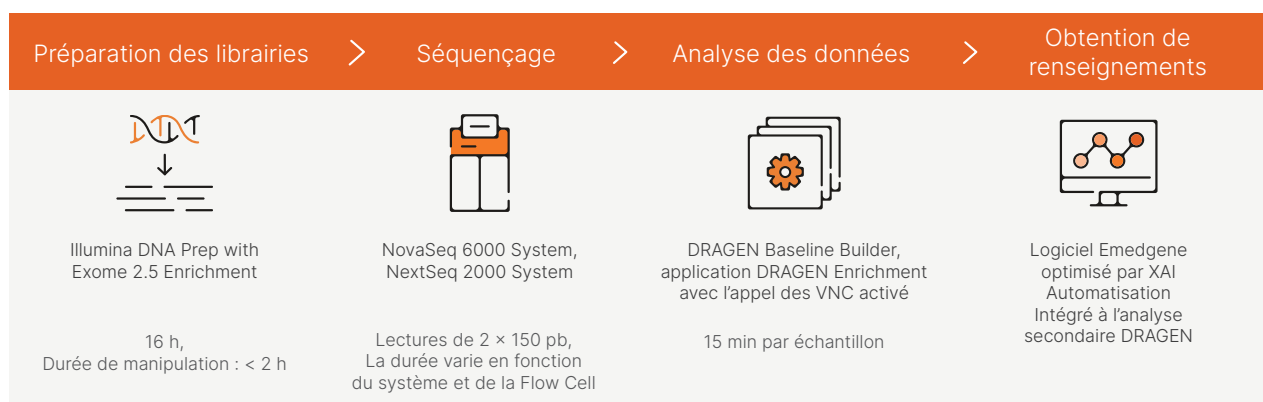


Figure 1 : Flux de travail du WES complet pour l'appel des VNC : combinez Illumina DNA Prep with Exome 2.5 Enrichment, les plateformes de SNG d'Illumina, l'analyse rapide des données avec l'analyse secondaire DRAGEN et l'interprétation des variants avec le logiciel Emedgene pour obtenir un flux de travail convivial et rationalisé pour un séquençage précis de l'exome afin d'interroger les VNC.

Tableau 1 : Identifiants d'échantillons du Coriell Institute inclus dans le panel de référence

HG00096	HG01393	HG01950	HG03006	NA10830	NA12877	NA18502	NA18970	NA21112
HG00190	HG01441	HG01985	HG03870	NA10831	NA12878	NA18508	NA19681	NA24143
HG00262	HG01551	HG01990	HG03882	NA10835	NA12889	NA18622	NA19720	NA24149
HG00626	HG01599	HG02013	HG03898	NA10838	NA12890	NA18637	NA20509	NA24631
HG00628	HG01896	HG02348	HG04090	NA10839	NA12891	NA18942	NA20875	NA24694
HG01392	HG01914	HG02521	HG04214	NA12249	NA12892	NA18957	NA21098	NA24695

Préparation des librairies

Les librairies de séquençage d'exomes ont été préparées en double conformément au protocole Illumina DNA Prep with Exome 2.5 Enrichment (document n° 1000000048041 v07) en utilisant 50 ng d'ADNg provenant d'échantillons du Coriell Institute, sauf indication contraire. Les ensembles A à D d'IDT for Illumina DNA/RNA UD Indexes, Sets A-D, Tagmentation (Illumina, références n° 20027213, 20027214, 20042666 et 20042667) ont été utilisés pour permettre un indexage unique de toutes les librairies préparées. Les librairies de pré-enrichissement ont été regroupées par masse (cible de 250 ng par échantillon) à 12 niveaux par groupement et hybridées pendant la nuit à l'aide du Twist Bioscience for Illumina Exome 2.5 Panel, disponible dans les trousseaux Illumina DNA Prep with Exome 2.5 Enrichment Kit (Illumina, références n° 20077595 et 20077596). Les librairies ont été préparées avec et sans panel de contrôle Twist Bioscience for Illumina Mitochondrial Panel (Illumina, référence n° 20093180). L'ajout du panel mitochondrial de contrôle a suivi la dilution au 1:100 décrite dans le guide de référence d'Illumina DNA Prep with Exome 2.5 Enrichment (document n° 1000000157112).

Séquençage

Les librairies enrichies ont été séquencées sur les systèmes NovaSeq[™] 6000 et NextSeq[™] 2000 à l'aide de réactifs de séquençage par synthèse (SBS, Sequencing by Synthesis) standard et d'analyses à lecture appariée de 150 pb. Les sorties de données les plus élevées disponibles ont été utilisées pour obtenir une profondeur de couverture élevée par échantillon et permettre le sous-échantillonnage dans le cadre de l'analyse. Les échantillons du panel de référence et les échantillons pour analyse ont été séquencés à l'aide de la trousse NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, référence n° 20028312) et des réactifs NextSeq 2000 P3 Reagent (300 cycles) (Illumina, référence n° 20040561). Les trousseaux de réactifs pour NovaSeq 6000 System sont conçus pour être faciles à utiliser grâce à leurs trois cartouches prêtes à l'emploi préremplies avec tous les réactifs nécessaires.

Tableau 2 : Résumé des attributs du panel de référence

Sexe	Nbre d'échantillons
Masculin	27
Féminin	27
Superpopulation ⁵	
Européens (EUR)	18
Américains issus du métissage (AMR)	8
Africains (AFR)	7
Asiatiques du Sud (SAS)	9
Asiatiques de l'Est (EAS)	12

Tableau 3 : Débit de séquençage et couverture sur toutes les plateformes pour le panel de référence et les échantillons de VNC

Plateforme	Échantillons par analyse	Couverture moyenne par échantillon
NovaSeq 6000 System	72	~ 540×
NextSeq 2000 System	12	~ 280×

Le NextSeq 2000 System fournit une flexibilité au moyen de multiples configurations de Flow Cell et offre une analyse DRAGEN intégrée. Le nombre d'échantillons par analyse et la couverture moyenne obtenue par échantillon pour toutes les analyses sont indiqués dans le [tableau 3](#).

Analyse des données

L'application DRAGEN Baseline Builder App v4.2, accessible dans BaseSpace[®] Sequence Hub, a été utilisée pour créer les fichiers de dénombrement de cibles individuelles du panel de référence avec le génome de référence hg38 sans représentation graphique. Les données de séquençage pour les échantillons pour analyse de VNC ont été sous-échantillonnées à 80 millions de lectures (environ 120 à 140× la profondeur moyenne de la couverture exacte) et à 50 millions de lectures (environ 70 à 90× la profondeur moyenne de la couverture exacte) et analysées à l'aide de l'application DRAGEN Enrichment App v4.2, également accessible dans BaseSpace Sequence Hub, avec la détection des VNC activée. Pour chaque plateforme, le panel de référence correspondant a été utilisé pour la normalisation des échantillons. L'application DRAGEN Enrichment effectue une analyse secondaire précise et efficace pour l'appel complet des variants, notamment les SNV, les indels, les VNC et les variants structurels (SV, Structural Variant), entre autres applications. L'appel des variants DRAGEN pour Illumina DNA Prep with Exome 2.5 Enrichment peut également être effectué au sein du NextSeq 2000 System ou être entièrement intégré dans le flux de travail de recherche et d'interprétation des variants du logiciel Emedgene.

Résultats

Détection des VNC sur un éventail de variations

L'analyse avec le pipeline de VNC DRAGEN dans l'application DRAGEN Enrichment a révélé une détection fiable des VNC pour les variants évalués dans cette étude à seulement 50 millions de lectures, soit environ 70 à 90× la profondeur de couverture moyenne, sur toutes les plateformes. La détection des VNC est déterminée par la formule : (nombre de régions appelées)/(régions se chevauchant avec le panel d'exomes et la réalité). Les VNC étaient considérés comme détectés si cet indicateur était > 98 % (tableau 4).

Plusieurs des variants testés se trouvent dans le gène *DMD*, qui est une cible importante dans la recherche sur les maladies génétiques (tableau 4)⁶. À l'aide d'Integrated Genomics Viewer, les événements attendus de VNC du gène *DMD* détectés peuvent être visualisés pour les sept échantillons affectés testés (figure 2). Cela démontre le pouvoir de détecter les VNC sur ce gène lorsqu'on utilise le flux de travail Illumina Exome 2.5. La piste du fichier BED du Exome 2.5 Panel montre les régions couvertes par le panel (figure 2, en haut). Les résultats dépendront de divers facteurs. Il est recommandé de séquencer à une couverture moyenne > 150× pour le test initial d'un flux de travail de laboratoire afin de garantir une couverture suffisante pour la plupart des variants.

Analyse supplémentaire possible avec le logiciel Emedgene

Le logiciel Emedgene rationalise et étend l'interprétation des variants, permettant ainsi un gain de temps de 50 à 75 % par cas de recherche⁷. De nombreuses fonctionnalités optimisent l'interprétation définie par l'utilisateur, notamment XAI pour un classement transparent, automatisé, sur la base de preuves, des variants potentiellement causaux pour les échantillons; un graphique de preuves et d'annotations toujours à jour; la visualisation des variants; la conservation des variants; l'automatisation définie par l'utilisateur; et plus encore pour promouvoir une interprétation éclairée des variants. Le logiciel Emedgene a été conçu pour offrir une expérience utilisateur efficace et intuitive.

Pour rationaliser l'examen des cas, la fonctionnalité de hiérarchisation ou de « présélection » des variants par l'IA compile les variants, y compris les VNC, les plus susceptibles de résoudre un cas. Cette fonctionnalité comprend des données probantes et permet à l'analyste de gagner beaucoup de temps. Dans une étude distincte⁸, pour 51 individus présentant un variant génétique unique (solo) précédemment résolu par un VNC, le variant résolvant a été identifié dans une liste restreinte de 10 variants dans 92 % des cas. Dans 6 % (n = 3), le variant résolvant était présent dans la liste des candidats. Lors de l'examen des cas, le logiciel Emedgene permet le démarquage des variants sélectionnés dans la liste restreinte ou le marquage manuel des variants non sélectionnés dans la liste restreinte. La classification automatisée de l'ACMG* couvre également les VNC.

Accès aux fichiers du panel de référence

Les fichiers du panel de référence, générés à l'aide d'échantillons du Coriell Institute, fournissent aux chercheurs un ensemble de données connues pour la comparaison avec les échantillons pour analyse sans avoir à consacrer du temps et des ressources à l'acquisition, au séquençage et à l'analyse de cet ensemble d'échantillons. Il est préférable d'utiliser ces fichiers par défaut pour le test initial du flux de travail Illumina DNA Prep with Exome 2.5 Enrichment pour l'analyse des VNC et les laboratoires sont encouragés à développer un panel de référence spécifique à leurs protocoles de laboratoire. En cas d'utilisation de ces fichiers pour le test initial, les bibliothèques doivent être préparées pour les échantillons à analyser en suivant le même protocole que celui utilisé pour préparer le panel de référence. Les écarts dans le protocole peuvent entraîner des différences dans les profils de couverture et réduire ainsi l'efficacité du panel de référence par défaut. Des fichiers et des renseignements supplémentaires sont disponibles sur le Site de l'assistance d'Illumina.

* ACMG, American College of Medical Genetics and Genomics.

Tableau 4 : Détection des VNC ordonnée par taille approximative de l'événement en fonction des coordonnées attendues de l'événement

Identifiant Coriell	Chromosome	Emplacement chromosomique ou génétique (exons affectés)	Événement attendu	Taille approximative de l'événement (kb) ^a	DéTECTÉ à 50 millions de lectures?
NA04315	X	DMD (44)	Perte	0,15	✓
NA05115	X	DMD (45)	Perte	0,18	✓
NA05117	X	DMD (45)	Perte	0,18	✓
NA23599	X	MECP2 (3–4)	Perte	2,25	✓
NA18949	17	BRCA1 (15–16)	Perte	3,59	✓
NA21939	15	FBN1 (42–43)	Perte	3,70	✓
HG03857	16	PALB2 (5–7)	Perte	4,23	✓
HG00343	22	CHEK2 (9–10)	Perte	4,25	✓
NA23127	X	DMD (27–28)	Gain	7,47	✓
NA04520	16	TSC2 (1–15)	Perte	16,28	✓
NA04327	X	DMD (5–7)	Gain	22,70	✓
NA10283	X	DMD (72–79)	Perte	54,38	✓
HG00500	2	SPAST (4–17)	Gain	58,84	✓
NA23675	X	MECP2 (tous)	Gain	76,14	✓
NA23087	X	DMD (2–30)	Gain	608,45	✓
NA06870	18	18p11.32-18p11.1	Gain	15 390,21	✓
NA20125	10	10q23.1-10q26.3	Gain	52 877,99	✓

a. Basée sur les coordonnées d'événement attendues de l'exon affecté ou de l'emplacement chromosomique.

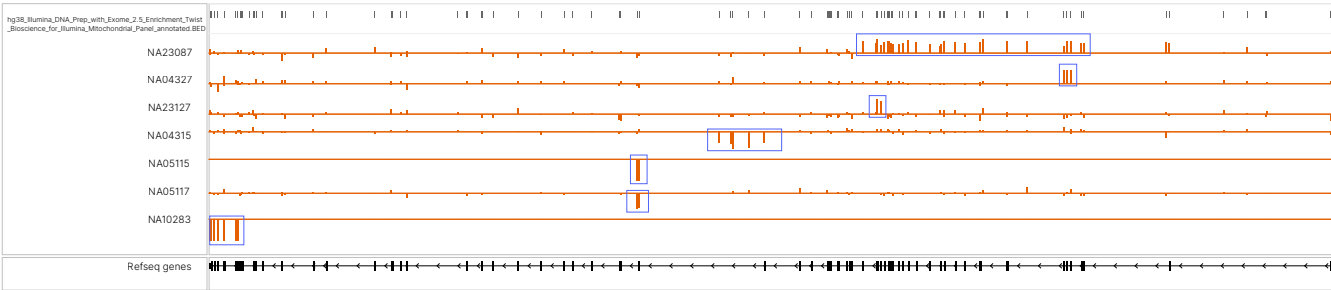


Figure 2 : Visualisation des VNC détectés sur sept échantillons contenant les événements attendus au sein du gène *DMD* : les fichiers fournissant une représentation BigWig du signal moyen normalisé (*.tn.bw) sont générés dans le cadre de l'analyse de l'application DRAGEN Enrichment et montrent les gains et les pertes en fonction des régions ciblées dans le panel Twist Bioscience for Illumina Exome 2.5 Enrichment Panel. Les pistes sont affichées avec une mise à l'échelle automatique sous forme de graphique à barres.

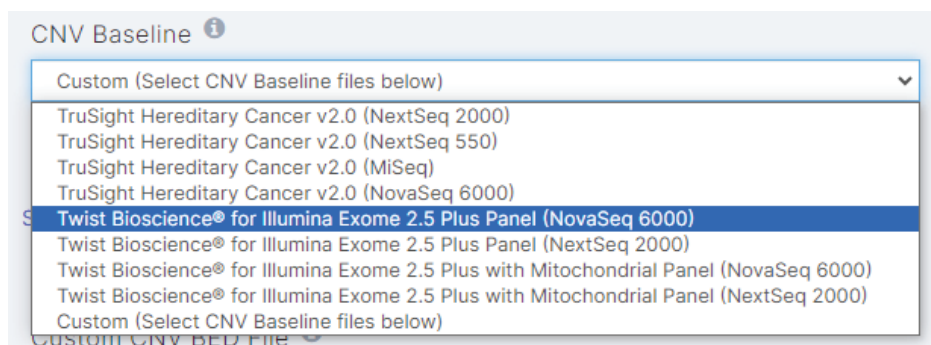


Figure 3 : Menu déroulant du panel de référence dans l'application DRAGEN Enrichment App v4.3 dans BaseSpace Sequence Hub.

Lors de l'utilisation de BaseSpace Sequence Hub, l'application DRAGEN Enrichment App v4.3 fournit des options déroulantes pour effectuer l'analyse initiale des VNC avec Twist Bioscience for Illumina Exome 2.5 Enrichment Panel pour de nombreuses références (hg19, hg38 et hs37d5 avec et sans représentation graphique) sur plusieurs plateformes (NextSeq 2000 System et NovaSeq 6000 System) à l'aide de plusieurs librairies (Twist Bioscience for Illumina Exome 2.5 Enrichment Panel avec et sans panel de contrôle Twist Bioscience for Illumina Mitochondrial Panel) (figure 3). Les échantillons pour analyse utilisés avec ce flux de travail d'analyse doivent être préparés comme il est décrit dans la section Méthodes de cette note d'application.

Résumé

Le flux de travail de SNG de l'exome pour analyser des VNC utilise Illumina DNA Prep with Exome 2.5 Enrichment pour la préparation des librairies, le séquençage sur NovaSeq 6000 System ou NextSeq 2000 System et les applications DRAGEN pour l'analyse des données. Les résultats démontrent l'application d'un panel de référence pour l'analyse des VNC provenant d'échantillons à analyser à l'aide du flux de travail du WES décrit. Le rappel élevé des VNC par le logiciel d'analyse secondaire DRAGEN présente une bonne corrélation avec d'autres méthodes. Ce flux de travail, y compris l'utilisation du panel de référence, pour l'analyse des VNC peut également être adapté à d'autres plateformes de séquençage d'Illumina. Le logiciel Emedgene aide les laboratoires à effectuer des analyses supplémentaires telles que l'interprétation des variants et la génération de rapports de recherche.

En savoir plus

[Données de référence du panel de référence](#)

[Analyse des variants du nombre de copies](#)

[Illumina DNA Prep with Exome 2.5 Enrichment](#)

[Plateformes de séquençage d'Illumina](#)

[Analyse secondaire DRAGEN](#)

[Application DRAGEN Germline](#)

[Logiciel Emedgene](#)

Références

1. Zhang F, Gu W, Hurles ME, Lupski JR. [Copy number variation in human health, disease, and evolution](#). *Annu Rev Genomics Hum Genet.* 2009;10:451-481. doi:10.1146/annurev.genom.9.081307.164217
2. Mehawej C, Maalouf JE, Abdelkhalik M, Mahfouz P, Chouery E, Megarbane A. [CNV Analysis through Exome Sequencing Reveals a Large Duplication Involved in Sex Reversal, Neurodevelopmental Delay, Epilepsy and Optic Atrophy](#). *Genes (Basel).* 2024;15(7):901. doi:10.3390/genes15070901
3. Fang X, Ma M, Rong W, et al. [Exome sequencing confirms the clinical diagnosis of both joubert syndrome and klinefelter syndrome with keratoconus in a han Chinese family](#). *Front Genet.* 2024;15:1417584. doi:10.3389/fgene.2024.1417584
4. Zeng Y, Ding H, Wang X, et al. [High positive predictive value of CNVs detected by clinical exome sequencing in suspected genetic diseases](#). *J Transl Med.* 2024;22(1):644. doi:10.1186/s12967-024-05468-1
5. Harrison PW, Amode MR, Austine-Orimoloye O, et al. [Ensembl 2024](#). *Nucleic Acids Res.* 2024;52(D1):D891-D899. doi:10.1093/nar/gkad1049
6. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. [Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model](#). *BMC Med Genomics.* 2018;11(1):91. doi:10.1186/s12920-018-0404-4
7. Greenwood Genetic Center. GGCC reduces turn around time on genomic analysis by 75% with Emedgene's AI platform. ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform. Publié le 12 septembre 2019. Consulté le 14 août 2024.
8. Données internes. Illumina, Inc. 2024.



Numéro sans frais aux États-Unis : + (1) 800 809-4566 | Téléphone : + (1) 858 202-4566
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d'Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez la page www.illumina.com/company/legal.html.
M-GL-01453 FRA v1.0