

Infinium^{MC} Global Diversity Array with Polygenic Risk Score Content-8 v1.0

Une solution évolutive et polyvalente destinée à la recherche sur les scores de risque polygénique (PRS, Polygenic Risk Score) dans une population multiethnique

- Contenu mis à jour comprenant 160 000 marqueurs de PRS qui couvrent un large éventail de troubles
- Couverture complète à l'ensemble du génome permettant une détermination pan-ethnique extrêmement précise des PRS
- Compatible avec le module logiciel Predict pour simplifier le calcul des PRS et la génération de rapports

illumina^{MD}

Introduction

L'évaluation précise du risque de maladie constitue un élément clé de la médecine préventive et personnalisée. Cependant, la prédiction du risque comporte encore des défis pour les maladies héréditaires polygéniques qui sont influencées par des milliers de variants génétiques interagissant avec des facteurs environnementaux. Les scores de risque polygénique (PRS, Polygenic Risk Score), également connus sous le nom de « scores polygéniques » ou « scores de risque génétique », sont des indicateurs numériques qui représentent la propension génétique d'un individu à développer une maladie ou un trait¹. Le PRS est un outil utile qui permet aux chercheurs d'étudier la stratification du risque et d'en savoir davantage sur les personnes qui pourraient bénéficier d'une surveillance supplémentaire ou d'interventions préventives précoces¹⁻³.

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 ([tableau 1](#)) est une solution de microréseaux de haute performance qui propose un flux de travail évolutif et flexible pour soutenir les études sur les PRS. La puce BeadChip à huit échantillons, conçue à partir de la chaîne principale éprouvée d'Infinium Global Diversity Array, présente un contenu de génotypage mis à jour comprenant 160 000 marqueurs pour une détermination précise des PRS ([tableau 2](#)). Les données de la puce BeadChip sont analysées de manière optimale avec le module logiciel Predict, offrant une solution complète d'évaluation du risque lié au génotype pour la recherche sur les maladies polygéniques.

Contenu optimisé pour la recherche sur les PRS

La chaîne principale pangénomique de l'Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 comporte environ 1,9 million de marqueurs pertinents d'un point de vue scientifique. Ce contenu complet de l'Infinium Global Diversity Array-8 v1.0, la puce à ADN de génotypage qui est utilisée dans le [programme de recherche All of Us](#), est complété par 160 000 marqueurs de PRS issus du vaste catalogue [Polygenic Score \(PGS\) Catalog](#)⁴ pour fournir une couverture étendue des régions génomiques de grande valeur ([tableau 2](#), [tableau 3](#)). Ces marqueurs, sélectionnés en collaboration avec [Allelica](#), fournissent une couverture exceptionnelle des six principales maladies à PRS, notamment le diabète de type 1 et 2, le cancer du sein, le cancer du côlon, le cancer de la prostate, la coronaropathie et la maladie d'Alzheimer.

Pour générer le contenu auxiliaire de PRS, les échantillons de la puce à ADN ont été mesurés par rapport aux scores de l'étalon de référence [1000 Genomes Project](#) et aux scores communs du PGS Catalog (p. ex. MAV313) pour tous les PRS Allelica. L'objectif de ce test de performance était d'identifier les variants présentant les erreurs d'imputation les plus courantes en comparant les résultats d'imputation avec 30× les données de séquençage du génome entier (WGS).

Pour développer la puce à ADN, on a sélectionné les variants avec l'ampleur d'effet la plus élevée sur le PRS et la qualité d'imputation la plus basse. Ces variants ont été inclus directement dans la puce à ADN pour être génotypés au lieu d'être imputés. Le contenu auxiliaire de PRS comprend des variants dont l'ampleur d'effet spécifique à chaque ascendance est la plus élevée dans les scores d'ascendances multiples Allelica pour optimiser la valeur de la puce à ADN pour les individus de toutes les ascendances et augmenter la concordance pour le PRS.

Tableau 1 : Aperçu de l'Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0

Fonctionnalité	Description
Espèce	Être humain
Nombre total de marqueurs ^a	2 028 571
Nombre d'échantillons par puce BeadChip	8
Exigence d'entrée d'ADN	200 ng
Réplicats de SNP	15
Nombre de SNP nécessaires pour l'appel des variations du nombre de copies	10
Chimie de test	Infinium LCG
Système prenant en charge la puce	Système iScan
Débit d'échantillonnage maximal du système iScan ^b	~ 1 728 échantillons/semaine
Durée de balayage par échantillon ^b	3 à 5 minutes

a. Le contenu comprend environ 1,9 million de marqueurs de la chaîne principale pangénomique d'Infinium Global Diversity Array-8 v1.0 plus 160 000 marqueurs spécifiques aux PRS.
b. Valeurs approximatives. Les durées de balayage et les débits maximaux peuvent varier selon le laboratoire et la configuration du système.

Tableau 2 : Contenu à valeur élevée de l’Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0

Contenu	Nbre de marqueurs ^a	Application de recherche/remarques
Gènes RefSeq hg19	1 123 407	Tous les gènes connus
RefSeq hg19 ± 10 kb	1 266 608	Régions régulatrices ^b
Promoteurs RefSeq	48 249	2 kb en amont pour inclure les régions du promoteur
ADME exonique	16 528	Absorption, distribution, métabolisme et excrétion des médicaments
Gènes ADME hg19	32 117	
ADME hg19 ± 10 kb	37 468	Y compris les régions régulatrices
Marqueurs HLA	17 595	Défense contre la maladie, rejet de transplantation, affections auto-immunes
Gènes HLA hg19	1 297	
Marqueurs CMH	22 783	
Gènes COSMIC hg19	1 079 088	Mutations somatiques du cancer
Gènes GO hg19	324 039	Annotation de la Gene Ontology

- a. Le nombre de marqueurs de chaque catégorie peut varier.
- b. De tous les gènes connus.

Abréviations : ADME, absorption, distribution, métabolisme et excrétion; COSMIC, Catalog of Somatic Mutations in Cancer (catalogue des mutations somatiques intervenant dans un cancer); GO, base de données de la Gene Ontology; hg19, version du génome humain 19; HLA, antigène leucocytaire humain; CMH, complexe majeur d'histocompatibilité; RefSeq, base de données des séquences de référence du National Center for Biotechnology Information.

Tableau 3 : Renseignements relatifs aux marqueurs

Catégories de marqueurs	Nbre de marqueurs ^a
Marqueurs exoniques ^a	530 644
Marqueurs de non-sens ^b	28 287
Marqueurs de faux sens ^b	396 783
Marqueurs de synonymes ^b	33 442
SNP silencieux ^b	40 964
Marqueurs d’ADN mitochondrial ^b	1 346
Indels ^c	37 289
Marqueurs Chr X ^c	62 103
Marqueurs Chr Y ^c	6 449
Marqueurs PAR/homologue ^c	5 485

- a. RefSeq, base de données des séquences de référence de la NCBI, ncbi.nlm.nih.gov/refseq.
- b. Par rapport au système UCSC Genome browser, genome.ucsc.edu.
- c. NCBI Genome Reference Consortium, version GRCh37 (hg19), ncbi.nlm.nih.gov/assembly/GCF_000001405.13.

Abréviations : indel, insertion/suppression; NCBI, National Center for Biotechnology Information; PAR, région pseudo-autosomique; SNP, polymorphismes mononucléotidiques; UCSC, Université de Californie à Santa Cruz.

Vaste couverture de variants associés à des maladies

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 fournit une couverture des variants sélectionnés à partir du catalogue GWAS du National Human Genome Research Institute et European Bioinformatics Institute (NHGRI-EBI)⁵ représentant un large éventail de classifications de maladies et de phénotypes (figure 1). La sélection des variants comprend également une gamme de classifications des pathologies fondées sur les annotations ClinVar⁶ et du College of Medical Genetics (ACMG)⁷.

Les bases de données cliniques, telles que ClinVar, évoluent en permanence, à mesure que de nouveaux variants sont ajoutés et que d'autres changent de dénomination pour passer à « Pathogènes » ou « Probablement pathogènes ». La puce BeadChip fournit une couverture actualisée de bon nombre de ces variants de grande valeur qui sont contenus dans les bases de données annotées. Ce contenu complet offre des occasions exceptionnelles aux chercheurs souhaitant étudier diverses populations pour tester et valider des associations précédemment trouvées sur des populations européennes.

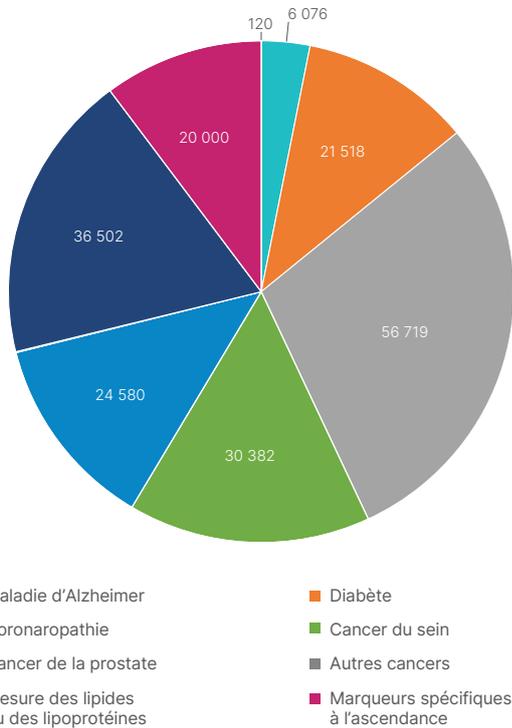


Figure 1 : Le contenu auxiliaire de PRS couvre plusieurs phénotypes – le contenu supplémentaire de PRS inclus sur l’Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 comporte des marqueurs de SNP sur un vaste ensemble de catégories de maladies. Un sous-ensemble de marqueurs assure une performance optimale pour des populations ancestrales spécifiques. Il est important de souligner que, comme de nombreux marqueurs sont partagés entre les phénotypes, la somme de ces valeurs n’est pas égale à 160 000 marqueurs.

Contenu exonique pan-ethnique exceptionnel

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 inclut le marquage renforcé des régions exoniques et une couverture enrichie pour faire correspondre avec précision les loci d’études d’association pangénomiques (GWAS, Genome-Wide Association Studies) avec des maladies précédemment identifiées ou de traits d’association. Plus de 400 000 marqueurs d’exome ont été réunis à partir de 36 000 individus d’origines ethniques diverses, y compris des Afro-Américains, Hispaniques, insulaires du Pacifique, populations d’Asie de l’Est, Européens et Métis. La puce à ADN se caractérise aussi par un contenu exonique diversifié de la base de données ExAC/gnomAD⁹, y compris les marqueurs pour l’ensemble des populations et propres à une population avec soit la fonctionnalité ou les fortes preuves aux fins d’association (tableau 4).

Tableau 4 : Couverture exonique pour l’ensemble des populations

Population(s) ^{a,b}	Nbre de marqueurs
NFE	346 340
EAS	146 281
AMR	272 178
AFR	257 690
SAS	224 431
NFE/EAS/AMR/AFR/SAS	69 432

a. internationalgenome.org/category/population.
 b. D’après gnomAD, gnomad.broadinstitute.org.

Abréviations : NFE, Européen non finnois; EAS, Asiatique de l’Est; AMR, Hispanique issu du métissage; AFR, Africain; SAS, Asiatique du Sud.

Marqueurs du contrôle qualité pour le suivi des échantillons

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 comprend environ 10 000 marqueurs du contrôle de la qualité (CQ). Ce contenu des marqueurs du CQ active d’importantes fonctions de suivi des échantillons, notamment la détermination de l’ascendance et la stratification pour faciliter les études à débit élevé (figure 2).

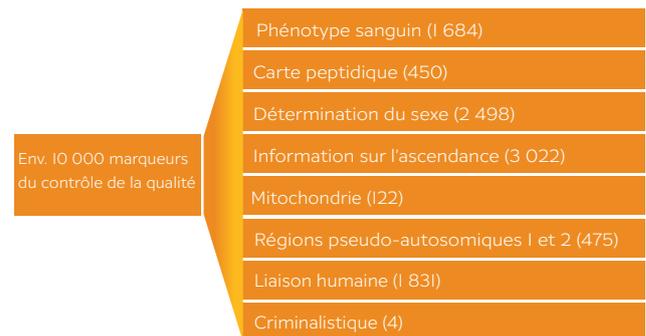


Figure 2 : Marqueurs du CQ par catégorie – les variants du CQ sur la puce à ADN permettent d’effectuer différentes fonctions de suivi des échantillons, notamment la détermination du sexe, l’ascendance continentale, l’identification humaine, et bien plus encore.

Flux de travail à débit élevé

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 utilise le format de puce à ADN à huit échantillons Infinium qui permet aux laboratoires de procéder efficacement à des mises à l'échelle. Afin d'obtenir un débit de traitement souple, le test Infinium LCG peut analyser jusqu'à 1 728 échantillons par semaine avec un seul système iScan^{MC}. Le test Infinium offre un flux de travail de trois jours qui permet aux chercheurs de rassembler des données et d'établir des rapports avec une durée de manipulation minimale (figure 3).

Test fiable de haute qualité

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 s'appuie sur une chimie de test Infinium dont la fiabilité a été établie pour produire des données reproductibles de la même haute qualité (tableau 5) que celles que les puces à ADN de génotypage d'Illumina fournissent depuis plus d'une décennie. De plus, le rapport signal/bruit élevé des débits individuels de génotypage du test Infinium donne accès à la définition de la variation du nombre de copies (VNC) sur l'ensemble du génome.

Tableau 5 : Performance des données et espace

Performance des données	Valeur ^{a,b}	Spécifications du produit ^c	
Taux de définition	99,78 %	> 99 % en moyenne	
Reproductibilité	99,99 %	> 99,90 % en moyenne	
Espace			
	Moyenne	Médiane	90e % ^d
Espace (kb)	1,439	0,619	3,680

a. Les valeurs sont tirées du génotypage de 2 228 échantillons de référence du projet HapMap.
 b. Valeurs attendues pour les projets types ayant recours aux protocoles standard d'Illumina. Les échantillons tumoraux et les échantillons préparés selon les protocoles non standard d'Illumina sont exclus.
 c. Les marqueurs du chromosome Y des échantillons féminins sont exclus.
 d. Basé sur les résultats de l'ensemble d'échantillons GenTrain.

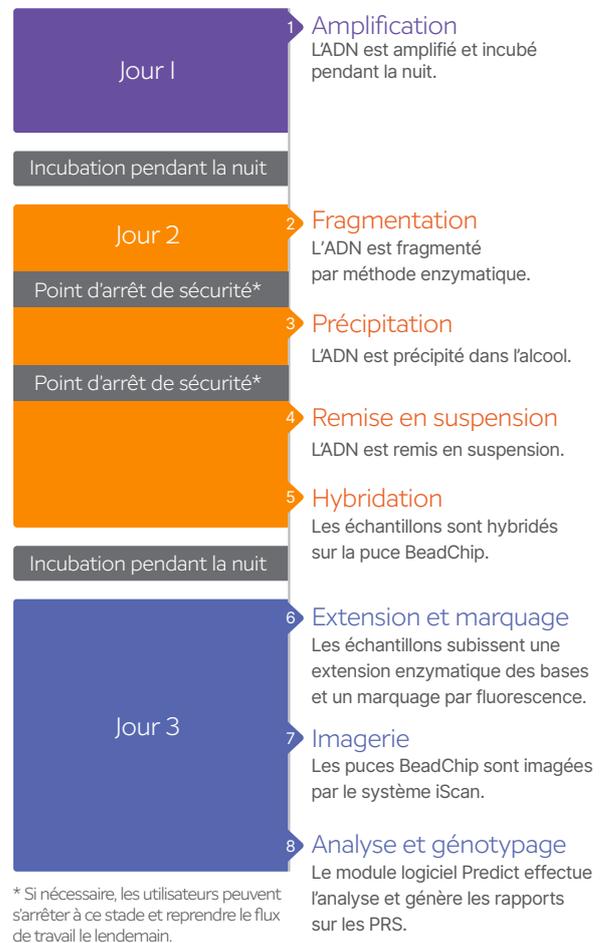


Figure 3 : Flux de travail Infinium LCG – le test Infinium utilise un flux de travail rapide de trois jours avec une durée de manipulation minimale.

Options adaptables de contenu

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 peut être personnalisé de manière à inclure les types de billes qui répondent aux besoins spécifiques en matière de recherche sur les PRS. Le contenu supplémentaire peut être sélectionné et exporté depuis le catalogue PGS pour être incorporé dans la puce BeadChip au moyen de l'outil DesignStudio^{MC} Microarray Assay Designer. Ceci permet d'accroître la couverture des variants.

Puissant pipeline d'analyse pour la recherche sur les PRS

Les données produites par l'Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 sont compatibles avec l'analyse à l'aide de [BaseSpaceMC Sequence Hub](#), développé à partir de l'infrastructure infonuagique fiable [Illumina Connected Analytics \(ICA\)](#). Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 est analysée de manière optimale avec le module logiciel Predict pour répondre aux besoins des laboratoires en quête d'un flux de travail complet pour la prédiction du risque lié au génotype. Ce logiciel d'analyse facile à utiliser est conçu pour produire des valeurs de PRS individuelles par rapport à celles qui sont calculées dans une population de référence avec des phénotypes connus afin de générer des rapports de PRS automatisés avec une prédiction du risque. Le module logiciel Predict peut calculer jusqu'à 24 PRS pour 1 152 échantillons en 15 minutes, pour une détermination rapide et précise des PRS fondés sur l'ascendance.

Résumé

Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 fait partie intégrante d'une trousse à outils complète offrant une solution simplifiée de prédiction du risque lié au génotype pour la recherche sur les PRS. La trousse à outils PRS d'Illumina simplifie le processus de développement des PRS et permet ainsi aux bio-informaticiens dédiés de gagner des mois d'analyse de données. Cette solution flexible est compatible avec les applications à débit faible et élevé, et peut être personnalisée pour créer des rapports exhaustifs de PRS fondés sur l'ascendance.

En savoir plus

Scores de risque polygénique, [illumina.com/areas-of-interest/complex-disease-genomics/polygenic-risk-scores](https://www.illumina.com/areas-of-interest/complex-disease-genomics/polygenic-risk-scores)

Infinium Global Diversity Array, [illumina.com/products/by-type/microarray-kits/infinium-global-diversity](https://www.illumina.com/products/by-type/microarray-kits/infinium-global-diversity)

Module logiciel Predict, [illumina.com/products/by-type/informatics-products/polygenic-risk-score-software](https://www.illumina.com/products/by-type/informatics-products/polygenic-risk-score-software)

Renseignements relatifs à la commande

Produit	N° de référence
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (16 échantillons)	2090683
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (48 échantillons)	2090684
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (96 échantillons)	2090685
Infinium Global Diversity Array with Polygenic Risk Score Content-8 v1.0 (384 échantillons)	2090686

Références

1. Khera AV, Chaffin M, Aragam KG, et al. [Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations](#). Nat Genet. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z
2. Lambert SA, Abraham G, Inouye M. [Towards clinical utility of polygenic risk scores](#). Hum Mol Genet. 2019;28(R2):R133-R142. doi:10.1093/hmg/ddz187
3. Lewis CM, Vassos E. [Polygenic risk scores: from research tools to clinical instruments](#). Genome Med. 2020;12(1):44. doi:10.1186/s13073-020-00742-5
4. The Polygenic Score (PGS) Catalog website. pgscatalog.org. Accessed August 18, 2022.
5. Buniello A, MacArthur JAL, Cerezo M, et al. [The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019](#). Nucleic Acids Res. 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
6. NCBI. ClinVar Database website. ncbi.nlm.nih.gov/clinvar. Accessed August 18, 2022.
7. Green RC, Berg JS, Grody WW, et al. [ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing](#). Genet Med. 2013;15(7):565-574. doi:10.1038/gim.2013.73
8. Broad Institute. Genome Aggregation Database (gnomAD) Browser website. gnomad.broadinstitute.org. Accessed August 18, 2022.

illumina^{MD}

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

© 2023 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-01187 FRA V1.0