Intégration des données protéomiques basées sur les puces à ADN et le SNG

Intégrer les résultats protéomiques basés sur puces à ADN de la plateforme SomaScan[MD] aux données de SNG d'Illumina Protein Prep permet une analyse unifiée

Maximisez les renseignements en harmonisant les données issues d'études multiplateformes

Limitez les biais spécifiques aux plateformes grâce à des données hautement corrélées et concordantes

Rationalisez l'analyse des données protéomiques avec Illumina Connected Multiomics

Introduction

L'intégration des mesures d'abondance des protéines dans différentes technologies de test est une exigence de plus en plus importante pour les études multiplateformes, les analyses longitudinales et le maintien de la continuité des données. À mesure que la taille des études augmente et que les technologies évoluent, il existe un besoin croissant d'interopérabilité entre les ensembles de données existants et les nouveaux ensembles générés sur les plateformes d'analyse protéomique. Le test Illumina Protein Prep et la plateforme SomaLogic[MD] SomaScan Platform fournissent tous deux des renseignements protéomiques de grande valeur, mais diffèrent considérablement en termes de chimie d'analyse, de structures de données et de conventions de normalisation. Combiner les données de ces deux plateformes est possible grâce à une stratégie d'intégration claire qui fournit des renseignements biologiques riches, tout en minimisant les biais spécifiques aux plateformes.

Les deux tests utilisent les réactifs SOMAmer[MD] (aptamère modifié à faible taux de dissociation) pour l'identification des protéines, mais utilisent des processus moléculaires fondamentalement différents pour générer des données protéomiques. Illumina Protein Prep utilise le séquençage de nouvelle génération (SNG) pour quantifier et détecter les protéines, tandis que SomaScan Platform utilise une lecture basée sur les puces à ADN. Les deux lectures étant optimisées pour réduire au minimum le bruit de fond, elles présentent une bonne corrélation lors de la comparaison d'un analyte donné dans plusieurs échantillons biologiques. Cependant, en raison des différences entre les deux tests, notamment leurs unités, leur plage dynamique et leurs références de normalisation (figure 1), les valeurs et les dénombrements relatifs entre les analytes diffèrent d'une lecture à l'autre.

L'objectif de l'intégration des données d'Illumina Protein Prep et de SomaScan Platform est d'harmoniser les mesures d'abondance afin que les résultats puissent être comparés plus facilement entre les plateformes.

Cette note technique décrit le cadre conceptuel pour intégrer les données d'Illumina Protein Prep aux ensembles de données existants de SomaScan Platform à l'aide du logiciel Illumina Connected Multiomics. Nous proposons une approche reproductible, transparente et scientifiquement justifiée qui permet à Illumina et à ses équipes collaboratives de tirer parti des ensembles de données existants de SomaScan Platform, tout en intégrant les nouvelles mesures de protéines d'Illumina dans des analyses unifiées.

Normalisation des données d'Illumina Protein Prep

Les données basées sur les puces à ADN peuvent être intégrées aux dénombrements par SNG en appliquant le pipeline de normalisation des données de SNG standard directement aux dénombrements bruts par puce à ADN.

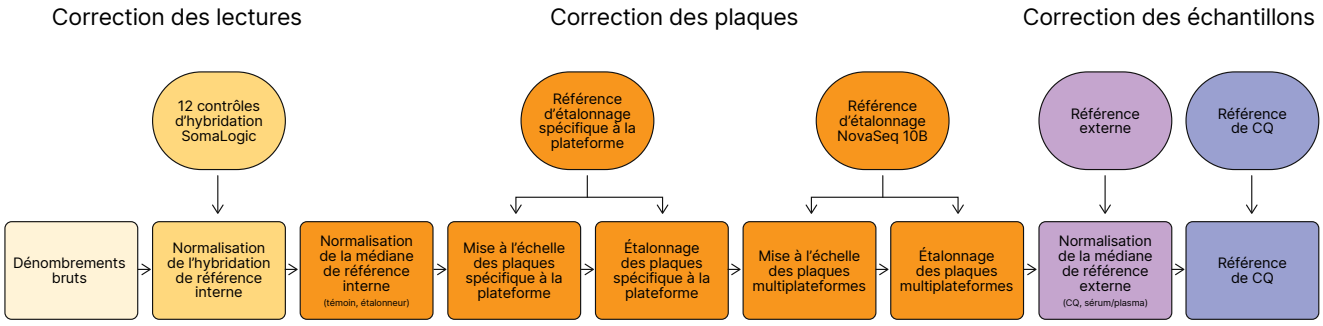


Figure 1 : Étapes séquentielles de normalisation dans le pipeline d'analyse d'Illumina Protein Prep

La normalisation des données comprend les étapes de correction des lectures, de correction des plaques et de correction des échantillons pour réduire la variabilité pouvant être introduite lors de l'hybridation, de la préparation des librairies ou des étapes de manipulation des liquides du test. Ces étapes sont effectuées de manière séquentielle et un fichier ADAT individuel est produit pour chaque étape de normalisation.

La normalisation des données corrige la variabilité intra et intertests, normalise les données et permet des comparaisons significatives entre les échantillons, les plaques, les analyses et les plateformes.
Le processus de normalisation des données comprend trois étapes séquentielles qui permettent de corriger la variabilité technique pouvant être introduite lors de l'hybridation, de la préparation des librairies de SNG ou des étapes de manipulation des liquides du test (figure 1).

Le processus de normalisation comprend les étapes suivantes :

- **Correction des lectures** : cette étape utilise les contrôles des réactifs SOMAmer pour réduire la variabilité technique introduite dans le processus de préparation des librairies de SNG.
- **Correction des plaques** : cette étape utilise les contrôles d'étalonneur comme contrôles positifs pour corriger les biais entre les plaques; une référence externe fournie par Illumina permet de comparer les plaques Illumina Protein Prep issues de différentes analyses.
- **Correction des échantillons** : cette étape utilise l'abondance des protéines dans chaque échantillon pour réduire la variabilité technique introduite lors de la quantification des protéines et pour corriger les différences de concentration protéique globale.

Stratégie d'intégration

L'objectif de l'intégration des ensembles de données est de transposer les mesures d'abondance des protéines, issues des valeurs d'unités relatives de fluorescence (RFU, Relative Fluorescence Unit) générées par SomaScan Platform, sur une échelle permettant leur interprétation de manière conjointe avec le test Illumina Protein Prep basé sur le SNG (figure 2). Cette approche peut être appliquée lors de l'intégration rétrospective des données, de l'expansion continue des cohortes, ainsi que dans le cadre d'études où les tests Illumina Protein Prep et SomaScan sont utilisés en parallèle ou de manière séquentielle.

Les tests Illumina Protein Prep 9.5K et SomaScan 11K utilisent tous deux des contrôles d'étalonneur identiques pour la correction des plaques.
Par conséquent, lorsque la référence d'étalonnage du SNG multiplateforme est appliquée aux données de la puce à ADN dans Illumina Connected Multiomics, elle intègre ces données à l'échelle du SNG, rendant ainsi comparables les données de ces deux tests.
L'évaluation des performances de la puce à ADN est préservée dans cette normalisation grâce à l'étape d'étalonnage spécifique à la plateforme, au cours de laquelle les étalonneurs sont comparés à une référence de puce à ADN et une mesure de contrôle qualité (CQ) est calculée.

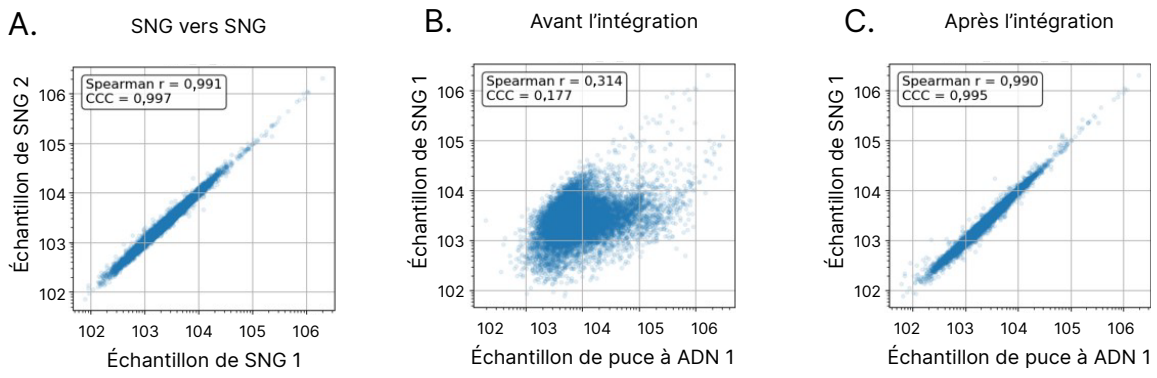


Figure 2 : Le coefficient de corrélation de Spearman entre les données issues du SNG et celles de la puce à ADN, avant et après l'intégration pour un seul échantillon, démontre l'harmonisation des données

Des échantillons de plasma regroupés représentatifs de différentes analyses montrant (A) une excellente corrélation entre les données des analyses de SNG individuelles, (B) une mauvaise corrélation entre les données du SNG et de la puce à ADN sans intégration, principalement en raison de la compression de la plage dynamique spécifique au SNG, et (C) une meilleure corrélation entre les données du SNG et de la puce à ADN, qui atteint des niveaux similaires à ceux des données intraplateforme grâce à l'intégration.

L'intégration fournit des données de SNG et de puce à ADN hautement corrélées

Pour évaluer cette stratégie d'intégration, des échantillons appariés de plasma regroupés ont été analysés avec le test Illumina Protein Prep 9.5K Plasma et le test SomaScan 11K. Les échantillons ont été analysés une deuxième fois avec Illumina Protein Prep pour générer une base de référence permettant de comparer les résultats interplateformes (puce à ADN vers SNG) aux résultats intraplateforme (SNG vers SNG). Les ensembles de données de SNG ont été normalisés à l'aide de la normalisation d'Illumina Protein Prep, tandis que l'ensemble de données de la puce à ADN a été normalisé à l'aide de la normalisation spécifique aux puces à ADN avant l'intégration, puis de la normalisation spécifique au SNG après l'intégration (figure 2).

L'intégration aligne les différences entre les lectures basées sur le SNG et celles basées sur la puce à ADN pour produire des données comparables (figure 2). Un aspect important de cette correction réside dans la mise à l'échelle relative entre les réactifs SOMAmer, qui diffère considérablement entre le SNG et la puce à ADN en raison d'une étape supplémentaire de compression de la plage dynamique (DRC, Dynamic Range Compression) lors de la lecture de SNG, visant à réduire la variabilité du séquençage entre les différents réactifs SOMAmer. Cette différence est flagrante même au sein d'un seul et même échantillon de plasma regroupé, où la variance intraplateforme (figure 2A) est nettement inférieure à la variance interplateformes (figure 2B) avant l'intégration. Cependant, l'intégration compense cette différence de lecture et aligne les données de la puce à ADN sur le dénombrement par SNG à une échelle comparable à la variance intraplateforme (figure 2C), mesurée par le coefficient de corrélation de Spearman et le coefficient de concordance de Lin. Cette similarité quantitative entre la variance intraplateforme et interplateformes après l'intégration est observée sur l'ensemble de la distribution des comparaisons d'échantillons de plasma regroupés (figure 3).

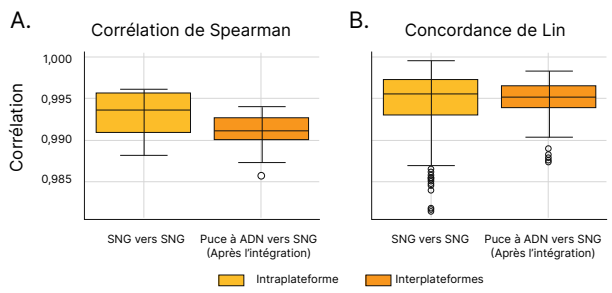


Figure 3 : La distribution des coefficients de corrélation de Spearman et de corrélation de concordance de Lin dans les échantillons montre l'effet constant de l'intégration dans l'alignement des données

Distribution de (A) la corrélation et de (B) la concordance dans 18 comparaisons d'échantillons de plasma regroupés entre les plateformes basées sur le SNG et sur les puces à ADN. Les corrélations et la concordance des lectures puce à ADN vers SNG après l'intégration entre de nombreux échantillons de plasma regroupés sont comparables aux corrélations SNG vers SNG.

La compression de la plage dynamique (DRC) est utilisée pour uniformiser les concentrations des réactifs SOMAmer d'une plage dynamique de 5-log à 2-log. Sans DRC, les réactifs SOMAmer les plus abondants représenteraient la majorité de la capacité de lecture du séquençage. Chaque échantillon nécessiterait une profondeur de séquençage extrêmement élevée pour couvrir les réactifs SOMAmer peu abondants et les quantifier avec précision. Avec la DRC, les sondes sont regroupées en fonction de leurs abondances observées en l'absence de DRC et chaque groupe est compressé selon un rapport défini différent.

L'intégration améliore la concordance des dénombrements entre les données de SNG et de puces à ADN dans le panel SOMAmer

Dans le cadre de la même étude, une cohorte de 32 échantillons de plasma appariés a été analysée à l'aide du test Illumina Protein Prep 9.5K Plasma et du test SomaScan 11K sur deux systèmes de manipulation de liquides Tecan différents. Les échantillons ont été analysés une deuxième fois avec le test Illumina Protein Prep 9.5K pour générer une base de référence pour les résultats intraplateforme (SNG vers SNG).

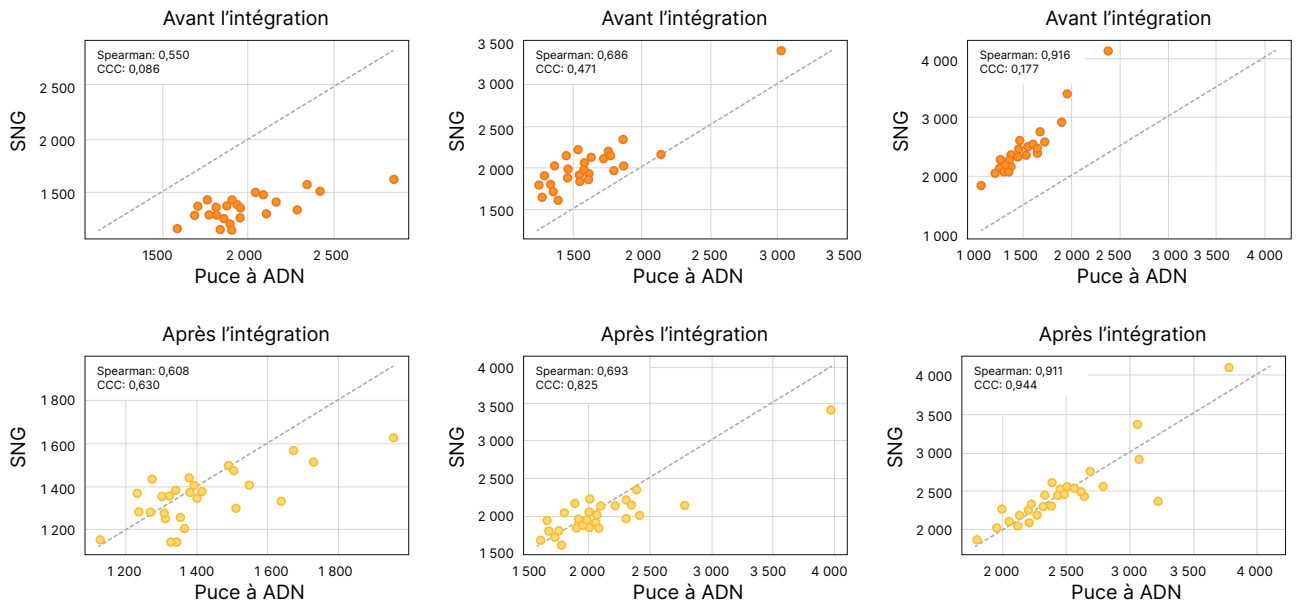


Figure 4 : Exemples de corrélations puce à ADN vers SNG pour trois réactifs SOMAmer avant et après l'intégration

L'intégration maintient la corrélation et augmente la concordance entre les dénombrements par SNG par puce à ADN pour chaque réactif SOMAmer dans une cohorte biologique.

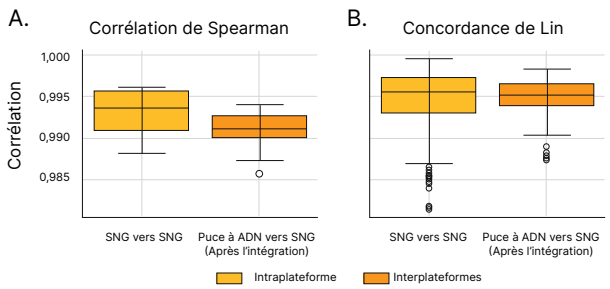


Figure 5 : La distribution du coefficient de corrélation de Spearman et du coefficient de corrélation de concordance de Lin à travers l'ensemble des réactifs SOMAmer humains démontre une concordance constante grâce à l'intégration

(A) La distribution des corrélations SNG vers SNG intraplateforme dans l'ensemble des réactifs SOMAmer est comparable aux corrélations SNG vers puce à ADN. (B) Grâce à l'intégration, la distribution de la concordance interplateformes se rapproche de celle de la concordance intraplateforme.

Les ensembles de données de SNG ont été normalisés à l'aide de la normalisation d'Illumina Protein Prep, tandis que les ensembles de données basés sur la puce à ADN ont été normalisés à la fois à l'aide de la normalisation spécifique aux puces à ADN (avant l'intégration) et de la normalisation spécifique au SNG (après l'intégration). Les corrélations intraplateforme avec le SNG à l'aide des deux systèmes, ainsi que les corrélations interplateformes entre les résultats du SNG et de la puce à ADN, ont été évaluées (figure 4, figure 5).

Étant donné que les deux lectures mesurent la concentration relative des réactifs SOMAmer, les dénombrements individuels de réactifs SOMAmer dans la cohorte biologique sont généralement corrélés. Autrement dit, les données se déplacent dans la même direction entre les lectures de puce à ADN et de SNG, même sans intégration et sont quantifiées à l'aide du coefficient de corrélation de Spearman (figure 4, figure 5A). Cependant, la concordance des dénombrements des réactifs SOMAmer, qui mesure à quel point les dénombrements réels sont proches, est nettement plus faible dans les comparaisons interplateformes (puce à ADN vers SNG) que dans les comparaisons intraplateforme (SNG vers SNG).

Cela est dû aux différences entre les méthodologies de lecture, notamment les unités, la plage dynamique et les références de normalisation. L'intégration aligne les données de la puce à ADN sur l'échelle de lecture du SNG, ce qui permet de comparer les dénombrements par réactif SOMAmer par échantillon. Cela est quantifié par le coefficient de corrélation de concordance de Lin, qui atteint des niveaux similaires à ceux des données intraplateforme grâce à l'intégration (figure 5B).

Résumé

Cette note technique démontre que l'intégration des tests fournit des données hautement corrélées et concordantes à partir de plateformes basées sur le SNG et sur les puces à ADN. Une stratégie d'intégration efficace permet aux laboratoires de compléter les données générées précédemment sur SomaScan Platform avec les nouvelles mesures d'Illumina Protein Prep, ce qui permet la découverte de biomarqueurs multiplateformes, soutient la planification de la transition entre les tests et assure la continuité des analyses au sein des pipelines protéomiques en constante évolution.

En savoir plus →

Illumina Protein Prep

Détection précise et fiable des protéines avec Illumina Protein Prep

illumina[MD]

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

© 2026 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-04147 FRA v1.0