

# Obtenez une détection précise et fiable des protéines avec la solution Illumina Protein Prep

Guide des étapes de contrôle de la qualité (CQ) et de normalisation des données à l'aide du pipeline d'analyse secondaire DRAGEN<sup>MC</sup> Protein Quantification

## Résultats reproductibles

Obtenez une détection constante des protéines à partir d'échantillons et de plaques

## Analyse à lancement automatique

Obtenez un débit de haute qualité avec l'analyse secondaire DRAGEN

## Renseignements intégrés

Analysez facilement les fichiers de dénombrement normalisé des protéines avec Illumina Connected Multiomics

## Introduction

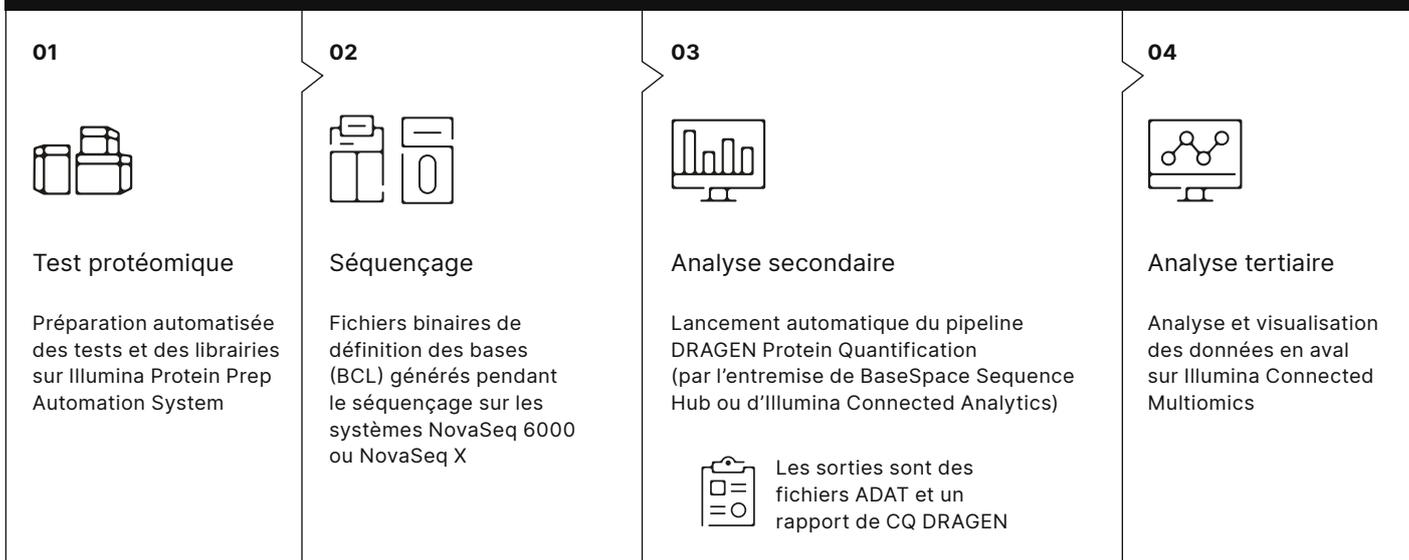
Les protéines jouent un rôle fonctionnel clé dans la biologie humaine, présentant un aperçu en temps réel de la santé et des états pathologiques. La quantification précise et reproductible des protéines est essentielle pour obtenir des renseignements biologiques significatifs. Illumina Protein Prep est un test protéomique innovant qui utilise les réactifs SOMAmer<sup>™</sup> (aptamère modifié à faible taux de dissociation) pour la capture des protéines, permettant une détection hautement sensible des protéines. Associer ce test avec lecture basée sur le séquençage de nouvelle génération (SNG) à la puissance bioinformatique de l'analyse secondaire DRAGEN offre une solution protéomique rationalisée et à débit élevé de « l'échantillon aux résultats ».

Illumina Protein Prep détecte simultanément des milliers de protéines dans un seul échantillon, fournissant ainsi une plateforme puissante pour la découverte de biomarqueurs. Cependant, le biais d'une plaque à l'autre et la variabilité technique des plateformes protéomiques à débit élevé peuvent masquer l'expression différentielle, en particulier

pour les biomarqueurs ayant des amplitudes d'effet réduites. Le pipeline DRAGEN Protein Quantification répond à ce défi lors de l'analyse secondaire, avec des méthodes de normalisation fiables pour fournir des données comparables de haute qualité dans l'ensemble des expériences. Sur le nuage, le pipeline se lance automatiquement une fois le séquençage terminé sur NovaSeq<sup>™</sup> 6000 System ou NovaSeq X System et génère un rapport de CQ DRAGEN et des fichiers ADAT avec des dénombrements de protéines normalisés qui peuvent être utilisés pour l'analyse tertiaire en aval avec Illumina Connected Multiomics (figure 1) pour obtenir des renseignements biologiques plus approfondis sur le protéome. Le pipeline DRAGEN Protein Quantification peut être installé sur un serveur DRAGEN sur site pour les utilisateurs ayant besoin d'une option d'analyse locale.

Cette note technique explique les étapes de contrôle de la qualité (CQ) et de normalisation incorporées dans le pipeline DRAGEN Protein Quantification pour éliminer la variabilité des échantillons et des techniques pour des données protéomiques de haute qualité.

Figure 1 : Un aperçu du flux de travail d'Illumina Protein Prep.



La solution protéomique à débit élevé utilise un test protéomique basé sur les réactifs SOMAmer pour la détection sensible de protéines à partir d'échantillons de sang ou de plasma, suivie de la préparation de bibliothèques Illumina et du séquençage sur les systèmes NovaSeq 6000 ou NovaSeq X. Le flux de travail du test protéomique et de la préparation des bibliothèques est automatisé sur Illumina Protein Prep Automation System, un poste de travail Tecan Fluent 780 personnalisé, pour délivrer des résultats cohérents et reproductibles. Le pipeline DRAGEN Protein Quantification se lance automatiquement après le séquençage dans BaseSpace Sequence Hub ou Illumina Connected Analytics, transformant les données protéomiques dans les fichiers BCL produits par les systèmes de séquençage d'Illumina en dénombrements protéomiques normalisés dans un fichier ADAT accompagné d'un rapport de CQ DRAGEN. Le fichier ADAT peut être intégré directement aux programmes d'analyse tertiaire, y compris Illumina Connected Multiomics, pour une interprétation plus approfondie des données.

## Pourquoi la normalisation des données est-elle nécessaire?

La variabilité technique est inhérente à tout test protéomique à débit élevé et peut provenir de plusieurs sources, notamment de la préparation des échantillons, de l'hétérogénéité intrinsèque des échantillons, de l'efficacité de l'hybridation, de l'efficacité de la PCR et du séquençage. En plus de la variabilité technique au sein d'un échantillon, des effets de lot peuvent survenir et affecter tous les échantillons d'une même plaque, en raison de différences au niveau des opérateurs, des instruments ou des effets environnementaux. Sans normalisation efficace ou correction des lots, ces variations systématiques pourraient masquer les véritables différences biologiques, générant ainsi des conclusions inexactes.

La normalisation des données est essentielle pour corriger cette variabilité, normaliser les données et permettre des comparaisons significatives entre les échantillons, les plaques et même les différentes analyses. La mise en œuvre de stratégies de normalisation complètes avec le pipeline DRAGEN Protein Quantification fournit des données de quantification des protéines fiables et reproductibles.

## Contrôles d'Illumina Protein Prep

Le test Illumina Protein Prep comprend des contrôles de plaques et des contrôles de réactifs SOMAmer dans chaque échantillon pour atténuer la variabilité du test. Les deux types de contrôles sont utilisés pour générer des dénombrements normalisés d'expression des protéines avec le pipeline DRAGEN Protein Quantification.

### Contrôles de plaque

Chaque plaque de 96 puits peut prendre en charge 85 échantillons et 11 contrôles de plaque (figure 2), y compris les témoins, les contrôles d'étalonneur et les échantillons de CQ. Ces contrôles sont inclus avec le test Illumina Protein Prep et sont analysés dans chaque plaque avec les échantillons de plasma ou de sérum pour la normalisation et le CQ. Ils servent de référence pour détecter et corriger les écarts techniques et fournissent une base pour l'alignement des échantillons analysés sur différentes plaques.

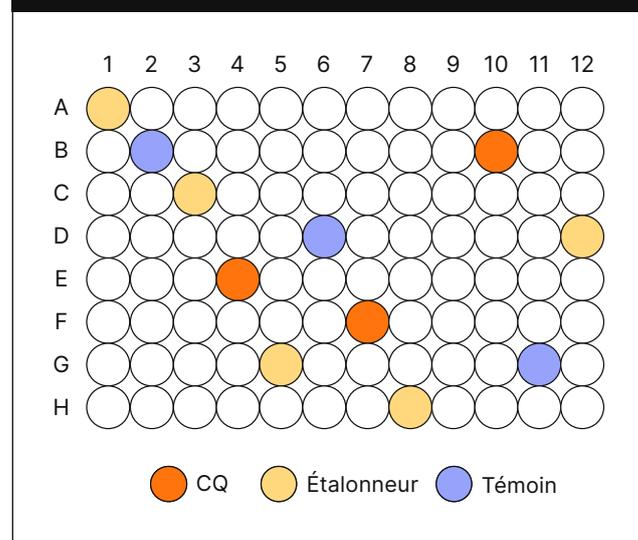
### Témoins

Les échantillons témoins sont constitués de tampons purs et fournissent une estimation du bruit de fond non spécifique dans le test Illumina Protein Prep. Ces contrôles devraient avoir des dénombrements significativement plus faibles que les échantillons de plasma ou de sérum. Des dénombrements élevés dans les puits témoins pourraient indiquer qu'une analyse présente un bruit de fond élevé ou qu'une contamination de l'échantillon a eu lieu.

### Contrôles d'étalonneur

Les contrôles d'étalonneur permettent de corriger la variabilité d'une plaque à l'autre. Ces contrôles sont soit des échantillons de plasma ou de sérum humain regroupés en fonction des types d'échantillons analysés. La normalisation de l'étalonnage de la plaque est effectuée en comparant la médiane des contrôles d'étalonneur par réactif SOMAmer à la référence d'étalonnage correspondante.

Figure 2 : Exemple de disposition de plaque avec contrôles.



## Échantillons de CQ

Les échantillons de CQ fournissent une vérification du contrôle de la qualité pour déterminer si la plaque a fonctionné comme prévu. Ces échantillons de contrôle sont soit des échantillons de plasma ou de sérum humain regroupés, en fonction des types d'échantillons analysés. La qualité de la plaque est évaluée en comparant la médiane de ces échantillons par réactif SOMAmer à la référence de CQ correspondante.

## Contrôles des réactifs SOMAmer

Les contrôles des réactifs SOMAmer sont incorporés dans chaque échantillon individuel à des concentrations connues avant l'hybridation et permettent de calculer les facteurs d'échelle spécifiques à l'échantillon pour normaliser les échantillons et évaluer la qualité.

## Étapes de normalisation des données

La normalisation des dénombrements protéomiques bruts implique plusieurs étapes pour éliminer les biais systématiques entre les échantillons et les plaques qui peuvent survenir pendant le flux de travail d'Illumina Protein Prep. Ces ajustements sont effectués en quatre étapes non consécutives : normalisation de l'hybridation, normalisation de la médiane de référence externe, étalonnage et ajustement de la plaque ([figure 3](#)).

## Normalisation de l'hybridation

Cette étape corrige la variabilité d'un échantillon à l'autre qui peut se produire pendant les étapes d'hybridation et de préparation de bibliothèques du test. Les facteurs d'échelle spécifiques à l'échantillon sont calculés en comparant le nombre de contrôles de normalisation de l'hybridation dans chaque échantillon à leurs valeurs médianes pour tous les échantillons de la plaque.

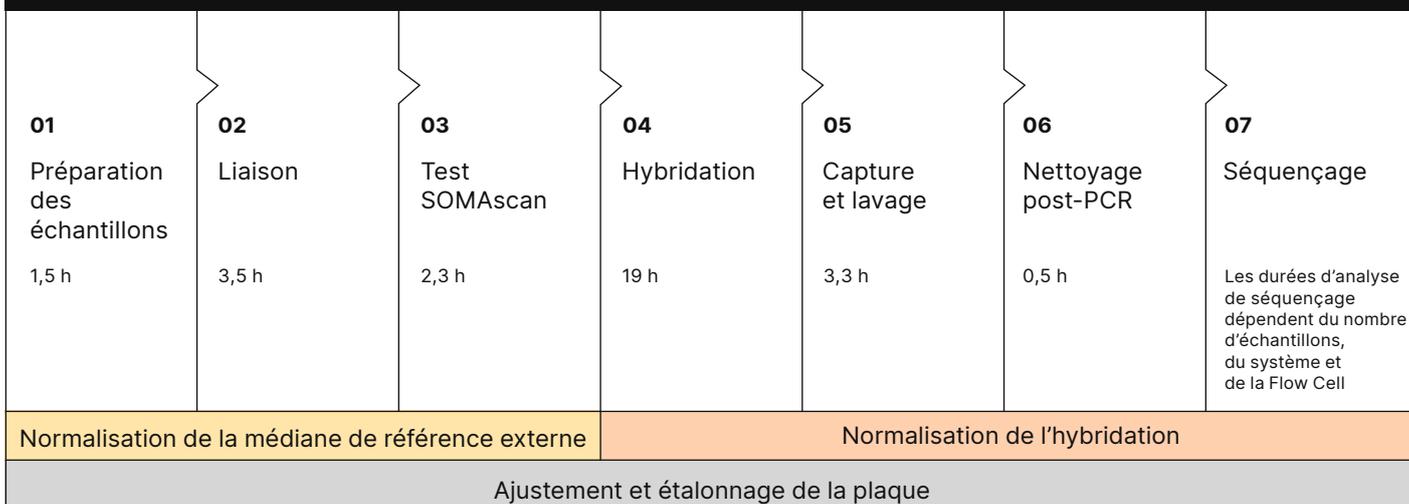
## Normalisation de la médiane de référence externe

Cette étape corrige les différences dans la mesure de l'abondance protéique totale des échantillons hors contrôles. Un facteur d'échelle est calculé en comparant les mesures de protéines observées à une référence des valeurs attendues pour chaque protéine.

## Étalonnage

L'étalonnage est effectué à l'aide des cinq échantillons d'étalonneur décrits dans la section des contrôles de plaque et corrige les effets de lot qui ont un impact sur les réactifs SOMAmer individuels. Cette étape permet des comparaisons précises entre les données produites à partir de différentes plaques et analyses. Pour chaque réactif SOMAmer, l'étalonnage corrige la variation entre les étalonneurs et les valeurs de référence attendues pour le contrôle d'étalonneur.

Figure 3 : Aperçu des étapes de normalisation effectuées par le pipeline DRAGEN Protein Quantification.



L'étalonnage est effectué en deux étapes :

- Étape 1 : compare les médianes de l'étalonneur à une référence issue du système de séquençage et ajuste chaque réactif SOMAmer en conséquence.
- Étape 2 : compare les médianes de l'étalonneur mises à jour à une référence issue des systèmes de séquençage, ce qui permet de comparer les données générées sur différents instruments. Le facteur d'échelle utilisé pour aligner la médiane des étalonneurs à la valeur de référence est appliqué à tous les échantillons de la plaque.

### Ajustement de la plaque

L'ajustement de la plaque traite les variations dans les mesures de protéines totales entre les plaques en utilisant la médiane de chaque mesure de réactif SOMAmer pour tous les étalonneurs et en les alignant sur une référence d'étalonnage externe.

Le processus d'ajustement se déroule en deux étapes :

- Étape 1 : ajuste les mesures en fonction d'une référence issue de l'instrument de séquençage, à savoir les systèmes NovaSeq 6000 ou NovaSeq X.
- Étape 2 : affine l'ajustement à l'aide d'une référence qui reste disponible pour tous les systèmes de séquençage, permettant ainsi des comparaisons entre les données générées sur différents instruments. Les indicateurs de CQ de la première étape d'ajustement seulement sont affichés dans le rapport DRAGEN Protein Quantification.

### Pourcentage de vérification du CQ dans les extensions homopolymériques

Une étape de CQ est effectuée à la fin de la normalisation à l'aide des contrôles de CQ sur chaque plaque. Cette étape compare la médiane de chaque mesure de réactif SOMAmer sur les trois répliquats d'échantillons de CQ à une référence de CQ externe pour attribuer un facteur d'échelle spécifique au réactif SOMAmer et un indicateur de CQ (QCCheckTailPercent).

Les diverses procédures de normalisation mises en œuvre dans le pipeline DRAGEN Protein Quantification minimisent la variabilité externe provenant de différentes étapes du flux de travail d'Illumina Protein Prep (figure 4).

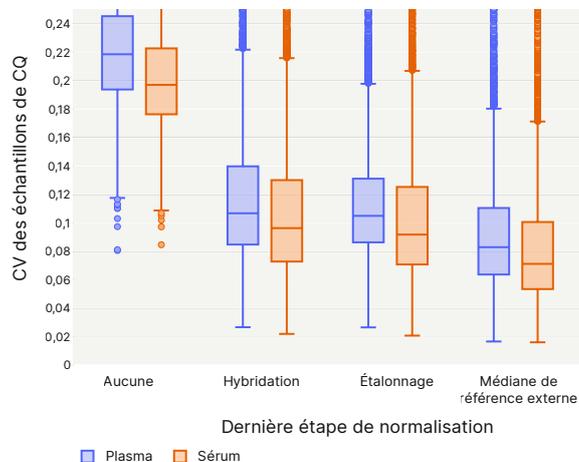
## Fichiers de sortie de DRAGEN Protein Quantification

Le pipeline DRAGEN Protein Quantification produit deux types de fichiers essentiels pour l'analyse des données d'Illumina Protein Prep. Le premier est une série de fichiers ADAT à chaque étape de normalisation et le deuxième est un rapport de CQ DRAGEN résumant le CQ de l'analyse.

### Fichiers ADAT

Le pipeline DRAGEN Protein Quantification traite les dénombrements bruts et génère un fichier texte ASCII délimité par des tabulations avec une extension ADAT comme fichier de sortie pour chaque étape de normalisation effectuée. Le fichier ADAT contient les mesures d'une série d'analytes (colonnes) pour une série d'échantillons (lignes) et comprend la description de l'analyte et la description de l'échantillon.

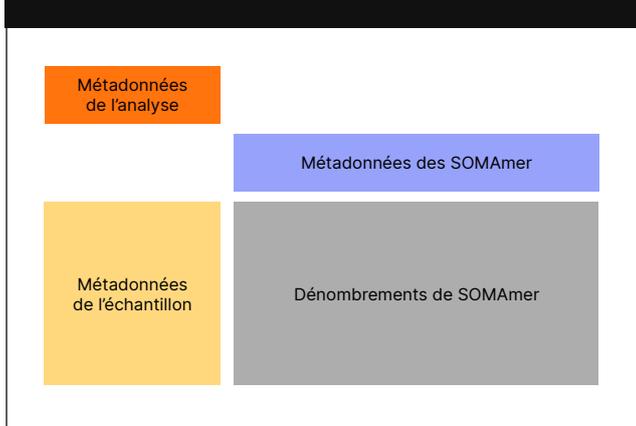
Figure 4 : Impact de la normalisation sur la variabilité des données.



La normalisation des données brutes du test entraîne une réduction du coefficient de variation (CV) intraplaque. Les données ont été générées à l'aide du test Illumina Protein Prep 6K avec des échantillons de CQ de plasma et de sérum. Le test Illumina Protein Prep 9.5K avec un CV médian d'environ 5,5 %<sup>a</sup> est désormais disponible.

a. CV médian attendu calculé à l'aide d'échantillons provenant de donneurs sains.

Figure 5 : Structure du fichier ADAT.



Les fichiers ADAT sont des fichiers délimités par des tabulations dont la structure est la suivante (figure 5) :

- Les **métadonnées de l'analyse** contiennent des indicateurs qui sont pertinents pour l'ensemble de l'analyse ou des plaques individuelles (p. ex. identifiant de l'analyse et indicateurs associés au pourcentage de vérification du CQ dans les extensions homopolymériques).
- Les **métadonnées SOMAmer** contiennent des indicateurs pertinents pour chaque réactif SOMAmer ou cible protéique (p. ex. identifiant SOMAmer et identifiants de gène et de cible spécifiques).
- Les **métadonnées de l'échantillon** contiennent des indicateurs pertinents pour chaque échantillon (p. ex. identifiant de l'échantillon, identifiant de la plaque et indicateurs de CQ).
- Les **dénombrements de SOMAmer** sont une matrice de dénombrements par SNG qui reflètent l'abondance relative de chaque SOMAmer dans chaque échantillon.

Le format est conçu pour offrir une flexibilité pour le nombre d'échantillons ainsi que le nombre et les types d'analytes et de descripteurs d'échantillons et peut être utilisé pour l'analyse tertiaire en aval afin d'obtenir des renseignements biologiques.

### Rapport de CQ DRAGEN

Un rapport de CQ DRAGEN Protein Quantification accompagne le fichier ADAT et résume le CQ de chaque échantillon individuel et de l'ensemble de la plaque. Le CQ d'échantillons individuels est dérivé des étapes de normalisation de l'hybridation et de normalisation de la médiane de référence externe. Le CQ de la plaque est dérivé des étapes de normalisation de l'étalonnage et du CQ.

## Utilisation des fichiers ADAT pour l'analyse tertiaire

Le fichier ADAT normalisé peut être immédiatement utilisé pour une analyse plus approfondie des données avec Illumina Connected Multiomics par l'entremise d'Illumina Connected Analytics. Illumina Connected Multiomics dispose d'une interface utilisateur graphique pour la gestion et la sélection des données et peut être utilisé pour l'interprétation avancée des données, y compris l'expression différentielle des protéines, l'analyse des voies et l'analyse d'enrichissement des ensembles de génétiques. Le contenu du fichier ADAT peut également être analysé à l'aide d'ensembles libres fournis dans R (SomadataIO) et Python (Canopy).

## Résumé

La normalisation des données protéomiques brutes est essentielle pour minimiser le biais du test et des échantillons qui peut survenir pendant le flux de travail d'Illumina Protein Prep. Le pipeline DRAGEN Protein Quantification utilise plusieurs étapes de normalisation et d'étalonnage pour réduire les variations systématiques tout en minimisant le risque de supprimer les véritables variations biologiques, offrant ainsi des résultats plus précis. Ce pipeline est disponible sur des serveurs infonuagiques ou sur site. Sur le nuage, le pipeline DRAGEN Protein Quantification se lance automatiquement après le séquençage sur les systèmes NovaSeq 6000 ou NovaSeq X, générant un rapport de CQ DRAGEN et des fichiers ADAT avec des dénombrements de protéines normalisés. La sortie sous forme de fichiers ADAT peut ensuite être facilement utilisée pour l'analyse tertiaire en aval avec Illumina Connected Multiomics pour permettre l'obtention de renseignements biologiques plus approfondis sur le protéome.

En savoir plus →

[Illumina Protein Prep](#)

[Analyse secondaire DRAGEN](#)

[Illumina Connected Analytics](#)



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

© 2025 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](http://www.illumina.com/company/legal.html).  
M-GL-03205 FRA v1.0