

# Rilevamento accurato e affidabile delle proteine con la soluzione Illumina Protein Prep



Guida alle fasi di QC e normalizzazione dei dati utilizzando la pipeline di analisi secondaria DRAGEN™ Protein Quantification

## Risultati riproducibili

Rilevamento coerente delle proteine su campioni e piastre

## Avvio automatico dell'analisi

Generazione di output di alta qualità con l'analisi secondaria DRAGEN

## Informazioni integrate

Analisi semplificata dei file dei conteggi normalizzati delle proteine con Illumina Connected Multiomics

## Introduzione

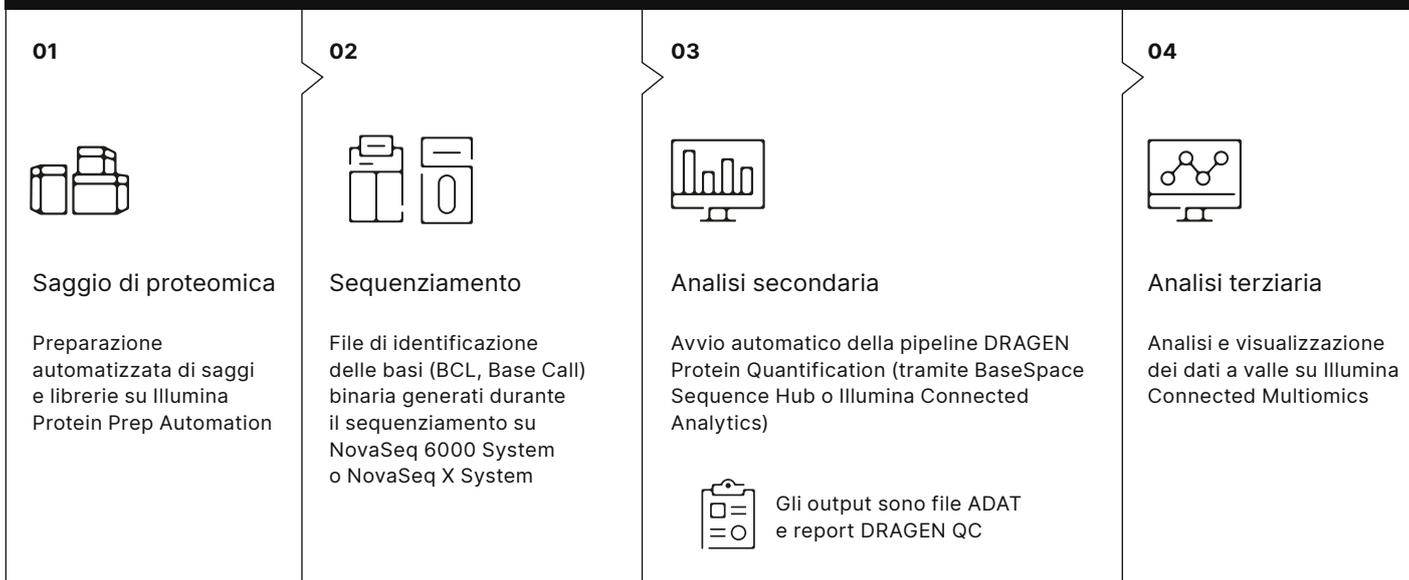
Le proteine svolgono un ruolo funzionale fondamentale nella biologia umana e riflettono in tempo reale lo stato di salute o le eventuali malattie. La quantificazione accurata e riproducibile delle proteine è fondamentale per ottenere informazioni biologiche significative. Illumina Protein Prep è un innovativo saggio di proteomica che utilizza i reagenti SOMAmer® (Slow Off-Rate Modified Aptamer), basati su aptameri modificati a bassa dissociazione con il ligando, per la cattura delle proteine, consentendone un rilevamento altamente sensibile. La combinazione di questo saggio con la lettura del sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) e la potenza bioinformatica dell'analisi secondaria DRAGEN offre una soluzione di proteomica ottimizzata e a elevata processività dal campione ai risultati.

Illumina Protein Prep rileva migliaia di proteine contemporaneamente in un singolo campione, fornendo una potente piattaforma per l'individuazione dei biomarcatori. Tuttavia, la distorsione da piastra a piastra e la variabilità tecnica delle piattaforme

di proteomica a elevata processività possono nascondere l'espressione differenziale, in particolare per i biomarcatori con piccole dimensioni di effetto. La pipeline DRAGEN Protein Quantification affronta questa difficoltà durante l'analisi secondaria, con efficaci metodi di normalizzazione per fornire dati di alta qualità e comparabili tra gli esperimenti. Sul cloud, la pipeline viene avviata automaticamente dopo il completamento del sequenziamento su NovaSeq™ 6000 System o NovaSeq X System e genera un report DRAGEN QC e file ADAT con conteggi normalizzati delle proteine che possono essere utilizzati per l'analisi terziaria a valle con Illumina Connected Multiomics (Figura 1) per ottenere informazioni biologiche più approfondite sul proteoma. Per gli utenti che necessitano di un'opzione di analisi locale, la pipeline DRAGEN Protein Quantification può essere installata su un server DRAGEN in laboratorio.

Questa nota tecnica spiega le fasi di controllo qualità (QC) e normalizzazione integrate nella pipeline DRAGEN Protein Quantification per eliminare la variabilità tecnica e del campione per dati di proteomica di alta qualità.

Figura 1: panoramica del flusso di lavoro di Illumina Protein Prep



La soluzione proteomica a elevata processività utilizza il saggio di proteomica basato sul reagente SOMAmer per il rilevamento sensibile delle proteine in campioni di siero o plasma, seguito dalla preparazione delle librerie Illumina e dal sequenziamento su NovaSeq 6000 System o NovaSeq X System. Per risultati coerenti e riproducibili, il flusso di lavoro del saggio proteomico e della preparazione della libreria può essere automatizzato su Illumina Protein Prep Automation System, un sistema Tecan Fluent 780 personalizzato. La pipeline DRAGEN Protein Quantification viene avviata automaticamente dopo il sequenziamento in BaseSpace Sequence Hub o Illumina Connected Analytics, trasformando i dati di proteomica nei file BCL prodotti dai sistemi di sequenziamento Illumina in conteggi proteomici normalizzati in un file ADAT con un report DRAGEN QC di accompagnamento. Il file ADAT può essere integrato direttamente nei programmi di analisi terziaria, tra cui Illumina Connected Multiomics, per un'ulteriore interpretazione dei dati.

## Perché è necessaria la normalizzazione dei dati?

La variabilità tecnica è relativa a qualsiasi saggio di proteomica a elevata processività e può derivare da più fonti, tra cui la preparazione dei campioni, l'eterogeneità intrinseca del campione, l'efficienza di ibridazione, l'efficienza della reazione a catena della polimerasi (PCR, Polymerase Chain Reaction) e il sequenziamento. Oltre alla variabilità tecnica all'interno di un campione, possono verificarsi effetti del batch che influiscono su tutti i campioni presenti sulla stessa piastra, per l'intervento di operatori diversi, l'uso di strumenti differenti o effetti ambientali dissimili. Senza un'efficace normalizzazione o correzione del batch, queste variazioni sistematiche potrebbero nascondere le vere differenze biologiche, portando a conclusioni imprecise.

La normalizzazione dei dati è essenziale per correggere questa variabilità, standardizzare i dati e consentire confronti significativi tra campioni, piastre e persino diverse corse di saggi. L'implementazione di strategie di normalizzazione complete con la pipeline DRAGEN Protein Quantification fornisce dati affidabili e riproducibili sulla quantificazione delle proteine.

## Controlli di Illumina Protein Prep

Il saggio Illumina Protein Prep include i controlli delle piastre e i controlli del reagente SOMAmer in ciascun campione al fine di mitigare la variabilità del saggio. Entrambi i tipi di controlli vengono utilizzati per generare conteggi normalizzati dell'espressione proteica con la pipeline DRAGEN Protein Quantification.

### Controlli della piastra

Ogni piastra a 96 pozzetti può ospitare 85 campioni e 11 controlli della piastra (Figura 2), fra cui i bianchi, i controlli del calibratore e i campioni di controllo qualità (QC, Quality Control). Questi controlli sono inclusi nel saggio Illumina Protein Prep e vengono eseguiti in ciascuna piastra insieme ai campioni di plasma o siero per la normalizzazione e il QC, fungendo da riferimento per rilevare e correggere le deviazioni tecniche e fornendo una base per l'allineamento dei campioni eseguiti su diverse piastre.

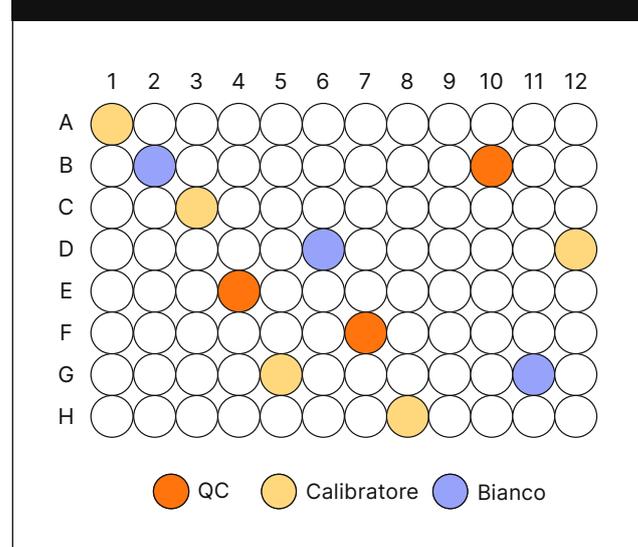
### Bianchi

I campioni bianchi sono costituiti da tampone puro e forniscono una stima del background non specifico nel saggio Illumina Protein Prep. Per questi controlli sono previsti conteggi significativamente inferiori rispetto ai campioni di plasma o siero. Conteggi elevati nei pozzetti bianchi potrebbero indicare che una corsa ha un background elevato o che si è verificata una contaminazione del campione.

### Controlli del calibratore

I controlli del calibratore consentono la correzione della variabilità da piastra a piastra. Questi controlli sono raggruppati in pool di plasma o siero umano, a seconda dei tipi di campione da eseguire. La normalizzazione della calibrazione della piastra viene eseguita confrontando la mediana dei controlli del calibratore per il reagente SOMAmer con il riferimento di calibrazione corrispondente.

Figura 2: esempio di layout della piastra con i controlli



## Campioni QC

I campioni QC forniscono un controllo qualità per determinare se la piastra ha funzionato come previsto. Questi campioni di controllo sono raggruppati in pool di plasma o siero umano, a seconda dei tipi di campione da eseguire. La qualità della piastra viene valutata confrontando la mediana di questi campioni per reagente SOMAmer con il riferimento QC corrispondente.

## Controlli del reagente SOMAmer

I controlli del reagente SOMAmer sono incorporati in ogni singolo campione a concentrazioni note prima dell'ibridazione e consentono di calcolare i fattori di scala specifici per il campione al fine di normalizzare i campioni e valutare la qualità.

## Fasi di normalizzazione dei dati

La normalizzazione dei conteggi di proteomica non elaborati prevede più fasi per rimuovere le distorsioni sistematiche tra i campioni e le piastre che possono verificarsi durante il flusso di lavoro di Illumina Protein Prep. Queste regolazioni vengono eseguite in quattro fasi non consecutive: normalizzazione dell'ibridazione, normalizzazione della mediana di riferimento esterno, calibrazione e adattamento delle piastre (Figura 3).

## Normalizzazione dell'ibridazione

Questa fase corregge la variabilità da campione a campione che può verificarsi durante le fasi di ibridazione e preparazione delle librerie del saggio. I fattori di scala specifici per il campione vengono calcolati confrontando i conteggi dei controlli di normalizzazione dell'ibridazione in ciascun campione con i loro valori mediani su tutti i campioni nella piastra.

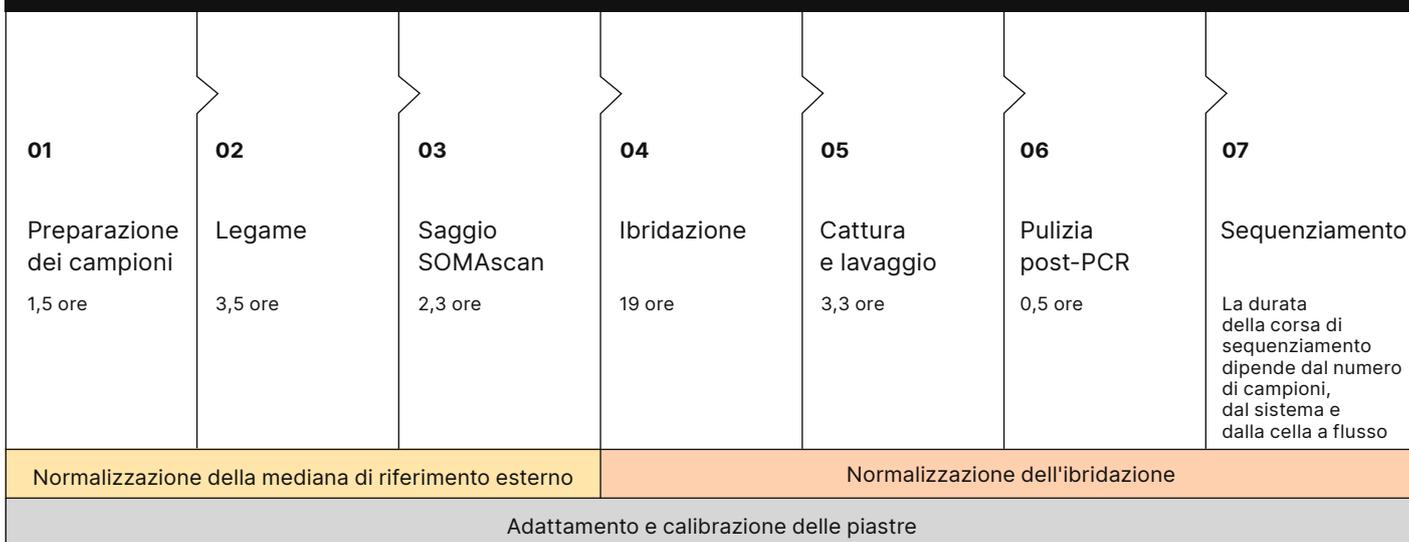
## Normalizzazione della mediana di riferimento esterno

Questo passaggio corregge le differenze nella misurazione dell'abbondanza proteica totale dei campioni non di controllo. Un fattore di scala viene calcolato confrontando le misurazioni delle proteine osservate con un riferimento dei valori previsti per ciascuna proteina.

## Calibrazione

La calibrazione viene eseguita utilizzando i cinque campioni calibratore descritti nella sezione dei controlli delle piastre e corregge gli effetti del batch che influiscono sui singoli reagenti SOMAmer. Questa fase consente confronti accurati tra i dati prodotti da diverse piastre e diverse corse. Per ciascun reagente SOMAmer, la calibrazione corregge la variazione tra i calibratori e i valori di riferimento previsti per il controllo del calibratore.

Figura 3: panoramica delle fasi di normalizzazione eseguite dalla pipeline DRAGEN Protein Quantification



La calibrazione viene eseguita in due fasi:

- Fase 1: confronta le mediane dei calibratori con un riferimento derivato dal sistema di sequenziamento e regola di conseguenza ciascun reagente SOMAmer.
- Fase 2: confronta le mediane dei calibratori aggiornate con un riferimento derivato dai sistemi di sequenziamento, consentendo il confronto tra i dati generati su strumenti diversi. Il fattore di scala utilizzato per allineare la mediana dei calibratori al valore di riferimento viene applicato a tutti i campioni sulla piastra.

### Adattamento delle piastre

L'adattamento delle piastre affronta le variazioni nelle misurazioni delle proteine totali tra le piastre utilizzando la mediana di ciascuna misurazione del reagente SOMAmer tra i calibratori e allineandola a un riferimento di calibrazione esterno.

Il processo di adattamento avviene in due fasi:

- Fase 1: regola le misurazioni in base a un riferimento derivato dallo strumento di sequenziamento, ovvero NovaSeq 6000 System o NovaSeq X System.
- Fase 2: perfeziona la regolazione utilizzando un riferimento che rimane tra i sistemi di sequenziamento, consentendo il confronto tra i dati generati su strumenti diversi. Nel report DRAGEN Protein Quantification sono mostrate le metriche QC relative solo alla prima fase di adattamento.

### Percentuale controllo QC in tailing

Al termine della normalizzazione viene eseguita una fase QC utilizzando i controlli QC su ciascuna piastra. Questa fase confronta la mediana di ogni misurazione con reagente SOMAmer tra i tre replicati del campione QC con un riferimento QC esterno per assegnare un fattore di scala specifico per il reagente SOMAmer e una metrica QC (QCCheckTailPercent).

Le varie procedure di normalizzazione implementate nella pipeline DRAGEN Protein Quantification riducono al minimo la variabilità estranea originata da fasi distinte del flusso di lavoro di Illumina Protein Prep (Figura 4).

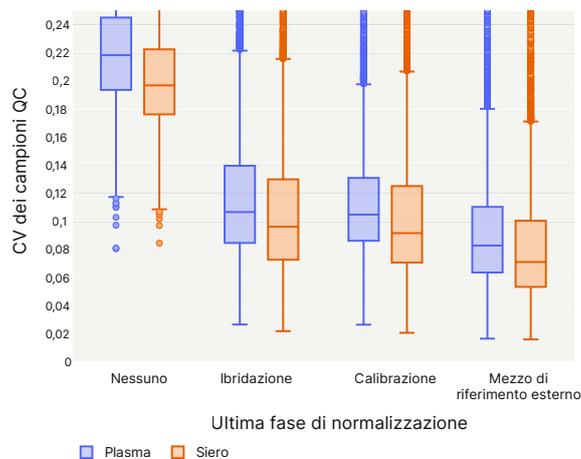
## Output di DRAGEN Protein Quantification

La pipeline DRAGEN Protein Quantification genera due tipi di file fondamentali per l'analisi dei dati di Illumina Protein Prep. Il primo è una serie di file ADAT derivati da ogni fase di normalizzazione e il secondo è un report DRAGEN QC che riassume il QC della corsa.

### File ADAT

La pipeline DRAGEN Protein Quantification processa i conteggi non elaborati e genera un file di testo ASCII delimitato da tabulazioni con un'estensione ADAT come output per ogni fase di normalizzazione eseguita. Il file ADAT contiene le misurazioni per una serie di analiti (colonne) su una serie di campioni (righe) e include la descrizione degli analiti e la descrizione dei campioni.

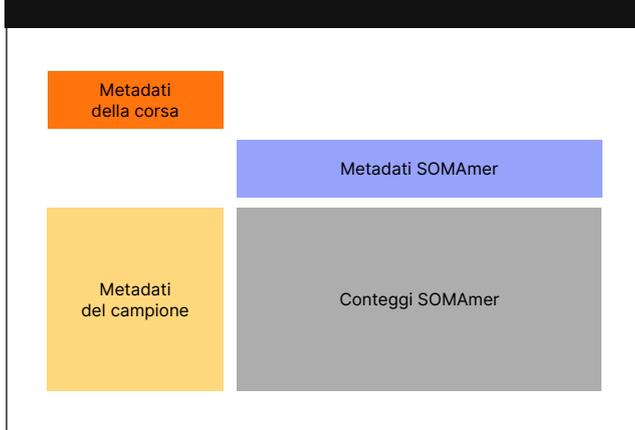
Figura 4: impatto della normalizzazione sulla variabilità dei dati



La normalizzazione dei dati non elaborati del saggio determina una riduzione del coefficiente di variazione (CV, coefficient of variation) intrapietra. I dati sono stati generati utilizzando il saggio Illumina Protein Prep 6K con campioni QC di plasma e siero. È disponibile il saggio Illumina Protein Prep 9.5K con CV mediano di circa 5,5%<sup>a</sup>.

a. CV mediano previsto calcolato utilizzando campioni di donatori sani.

Figura 5: struttura dei file ADAT



I file ADAT sono file delimitati da tabulazioni con la seguente struttura (Figura 5):

- I **metadati della corsa** contengono metriche rilevanti per l'intera corsa o le singole piastre (ad es. RunID e metriche associate alla percentuale di controllo QC in tailing).
- I **metadati SOMAmer** contengono metriche rilevanti per ciascun reagente SOMAmer o target proteico (ad es. ID SOMAmer e ID gene e target specifici).
- I **metadati del campione** contengono metriche rilevanti per ciascun campione (ad es. ID campione, ID piastra e metriche QC).
- I **conteggi SOMAmer** sono una matrice di conteggi NGS che riflettono l'abbondanza relativa di ciascun SOMAmer in ciascun campione.

Il formato è progettato per fornire flessibilità per il numero di campioni, nonché il numero e i tipi di analiti e descrittori dei campioni e può essere utilizzato per l'analisi terziaria a valle per ottenere informazioni biologiche.

#### Report di controllo qualità DRAGEN

Il file ADAT comprende un report DRAGEN Protein Quantification QC e riepiloga il QC di ogni singolo campione e dell'intera piastra. Il QC dei singoli campioni deriva dalla normalizzazione dell'ibridazione e dalle fasi di normalizzazione mediana di riferimento esterno. Il QC della piastra deriva dalla normalizzazione della calibrazione e dalle fasi QC.

## Utilizzo dei file ADAT per l'analisi terziaria

Il file ADAT normalizzato può essere utilizzato immediatamente per un'ulteriore analisi dei dati con Illumina Connected Multiomics tramite Illumina Connected Analytics. Illumina Connected Multiomics è dotato di un'interfaccia grafica che consente all'utente di gestire e selezionare i dati e può essere utilizzato per l'interpretazione avanzata dei dati, tra cui l'espressione proteica differenziale, l'analisi dei percorsi e l'analisi di arricchimento del set genico. Il contenuto del file ADAT può anche essere analizzato utilizzando pacchetti open-source forniti in R (SomadataIO) e Python (Canopy).

## Riepilogo

È essenziale normalizzare i dati proteomici non elaborati per ridurre al minimo la distorsione del saggio e del campione che potrebbe verificarsi durante il flusso di lavoro di Illumina Protein Prep. La pipeline DRAGEN Protein Quantification utilizza più fasi di normalizzazione e calibrazione per ridurre la variazione sistematica e minimizzare il rischio di eliminazione della variazione biologica reale, fornendo risultati più accurati. Questa pipeline è disponibile sul cloud o sui server in laboratorio. Sul cloud, la pipeline DRAGEN Protein Quantification viene avviata automaticamente dopo il sequenziamento su NovaSeq 6000 System o su NovaSeq X System, generando un report di controllo qualità DRAGEN e file ADAT con conteggi normalizzati delle proteine. L'output del file ADAT può quindi essere facilmente utilizzato per l'analisi terziaria a valle con Illumina Connected Multiomics per offrire informazioni biologiche più approfondite sul proteoma.

**Maggiori informazioni** →

[Illumina Protein Prep](#)

[Analisi secondaria DRAGEN](#)

[Illumina Connected Analytics](#)



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566  
techsupport@illumina.com | www.illumina.com

© 2025 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari. Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-03205 ITA v1.0