illumina®

NOTA TECNICA

Integrazione dei dati proteomici basati su array e su NGS

Analisi unificata grazie al raccordo dei dati proteomici basati su array generati da SomaScan® Platform con i dati NGS di Illumina Protein Prep

Armonizzazione dei dati di studi multipiattaforma per informazioni ottimizzate

Distorsioni associate alla piattaforma ridotte al minimo con dati altamente correlati e concordanti

Semplificazione dell'analisi dei dati proteomici con Illumina Connected Multiomics

Introduzione

L'integrazione delle misurazioni dell'abbondanza proteica ottenute con saggi basati su tecnologie diverse è un requisito sempre più importante per gli studi multipiattaforma, le analisi longitudinali e il mantenimento della continuità dei dati. Con l'aumentare delle dimensioni degli studi e l'evolversi delle tecnologie, si evidenzia la crescente necessità di interoperabilità tra i set di dati esistenti e nuovi generati su diverse piattaforme di analisi proteomica. Tanto il saggio Illumina Protein Prep quanto SomaLogic® SomaScan Platform forniscono informazioni proteomiche di alto valore, ma differiscono materialmente per quanto riguarda chimica di misurazione, strutture dei dati e convenzioni di normalizzazione. La combinazione dei dati generati da queste due piattaforme è possibile con una chiara strategia di raccordo, che sia in grado di fornire informazioni biologiche approfondite riducendo al minimo le distorsioni associate alla piattaforma specifica.

Per l'identificazione delle proteine, entrambi i saggi impiegano reagenti SOMAmer® (Slow Off-Rate Modified Aptamer), basati su aptameri modificati a bassa dissociazione con il ligando, ma per generare dati proteomici utilizzano processi molecolari fondamentalmente diversi: Illumina Protein Prep sfrutta il sequenziamento di nuova generazione (NGS, Next-Generation Sequencing) per la quantificazione e il rilevamento delle proteine, mentre SomaScan Platform impiega la lettura basata su array.

Poiché sono ottimizzate per ridurre al minimo il rumore di fondo, le letture mostrano una buona correlazione nel confronto di singoli analiti in più campioni biologici. Tuttavia, a causa delle differenze tra i due saggi, ad esempio a livello di unità di misura, intervallo dinamico e riferimenti per la normalizzazione (Figura 1), i valori e i conteggi relativi tra analiti differiranno nelle due letture.

L'obiettivo del raccordo dei dati di Illumina Protein Prep e SomaScan Platform è quello di armonizzare le misurazioni dell'abbondanza in modo che i risultati generati dalle due piattaforme possano essere confrontati più facilmente.

Questa nota tecnica delinea il quadro concettuale alla base del raccordo dei dati di Illumina Protein Prep con i set di dati esistenti di SomaScan Platform per mezzo del software Illumina Connected Multiomics. Viene proposto un approccio riproducibile, trasparente e scientificamente fondato attraverso il quale Illumina e i team collaborativi possano sfruttare i set di dati esistenti di SomaScan Platform e incorporare al contempo in analisi unificate le nuove misurazioni proteiche ottenute con le piattaforme Illumina.

Normalizzazione dei dati di Illumina Protein Prep

Il raccordo dei dati basati su array con i conteggi NGS può essere effettuato applicando la pipeline di normalizzazione dei dati NGS standard direttamente ai conteggi non elaborati basati su array.

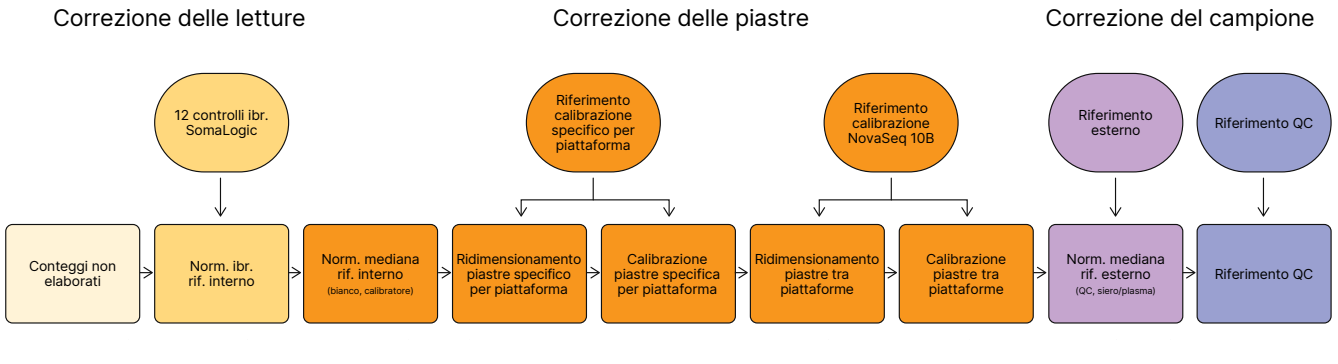


Figura 1: sequenza delle fasi di normalizzazione nella pipeline di analisi di Illumina Protein Prep

La normalizzazione dei dati include le fasi di correzione delle letture, correzione delle piastre e correzione del campione per ridurre la variabilità che può essere introdotta durante le attività di ibridazione, preparazione delle librerie o gestione dei liquidi del saggio. Le fasi di normalizzazione si svolgono in sequenza e per ognuna di esse viene generato un singolo file ADAT.

La normalizzazione dei dati corregge la variabilità nello stesso saggio e tra saggi, standardizza i dati e consente confronti significativi tra campioni, piastre, corse di analisi e piattaforme. Il processo di normalizzazione dei dati si compone di tre fasi sequenziali, responsabili di correggere la variabilità tecnica che può essere introdotta durante le attività di ibridazione, preparazione delle librerie NGS o gestione dei liquidi del saggio (Figura 1).

Il processo di normalizzazione include le seguenti fasi:

- **Correzione delle letture**. Questa fase utilizza i controlli dei reagenti SOMAmer per ridurre la variabilità tecnica introdotta nel processo di preparazione delle librerie NGS.
- **Correzione delle piastre**. Questa fase utilizza i controlli calibratori come controlli positivi per correggere le distorsioni tra le piastre; un riferimento esterno fornito da Illumina consente il confronto tra le piastre Illumina Protein Prep provenienti da diverse corse dei saggi.
- **Correzione del campione**. Questa fase utilizza l'abbondanza proteica in ciascun campione per ridurre la variabilità tecnica introdotta durante la quantificazione delle proteine e correggere le differenze a livello di concentrazione proteica complessiva.

Strategia di raccordo

Il raccordo dei set di dati ha l'obiettivo di adeguare le misurazioni dell'abbondanza proteica generate da SomaScan Platform ed espresse come valori di unità di fluorescenza relativa (RFU, Relative Fluorescence Unit) affinché siano interpretabili in parallelo ai risultati del saggio Illumina Protein Prep basato su NGS (Figura 2). Il raccordo può essere effettuato durante l'integrazione retrospettiva dei dati, l'espansione continua delle coorti e gli studi che prevedono l'uso dei saggi Illumina Protein Prep e SomaScan in parallelo o in sequenza.

I saggi Illumina Protein Prep 9.5K e SomaScan 11K utilizzano dei controlli calibratori identici per la correzione delle piastre. Di conseguenza, quando in Illumina Connected Multiomics il riferimento per la calibrazione NGS tra piattaforme viene applicato ai dati basati su array, tali dati vengono raccordati alla scala NGS, il che rende confrontabili i risultati dei due saggi. La valutazione delle prestazioni su array viene conservata in questa normalizzazione tramite la fase di calibrazione specifica per piattaforma, durante la quale i calibratori vengono confrontati con un riferimento array e viene calcolata una metrica di controllo qualità (QC, Quality Control).

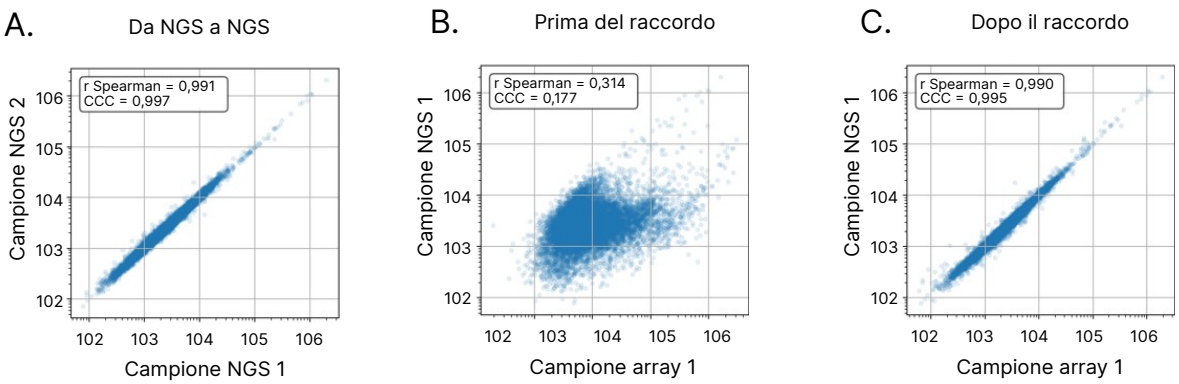


Figura 2: il confronto del coefficiente di correlazione di Spearman relativo ai dati basati su NGS e su array prima e dopo il raccordo per un singolo campione dimostra l'armonizzazione dei risultati

I campioni rappresentativi di plasma raggruppati da diverse corse di analisi mostrano (A) eccellente correlazione dei dati delle singole corse NGS, (B) scarsa correlazione dei dati basati su NGS e su array senza raccordo, principalmente a causa della compressione dell'intervallo dinamico specifica per l'NGS, e (C) migliore correlazione dei dati basati su NGS e su array, che aumenta fino a raggiungere i livelli intrapiattaforma con il raccordo.

Il raccordo produce dati basati su NGS e array altamente correlati

Per valutare questa strategia di raccordo, i campioni di plasma raggruppati e accoppiati sono stati analizzati sia con il saggio Illumina Protein Prep 9.5K Plasma sia con il saggio SomaScan 11K. I campioni sono stati analizzati una seconda volta su Illumina Protein Prep al fine di generare un livello di riferimento (baseline) per il confronto tra risultati interpiattaforma (da array a NGS) e intrapiattaforma (da NGS a NGS). I set di dati NGS sono stati normalizzati con la normalizzazione Illumina Protein Prep, mentre il set di dati basato su array è stato normalizzato con la normalizzazione array prima del raccordo e la normalizzazione NGS dopo il raccordo (Figura 2).

Il raccordo allinea le differenze tra le letture basate su NGS e su array per ottenere dati comparabili (Figura 2). Un aspetto importante di questa correzione è la scalabilità relativa tra reagenti SOMAmer, che mostra differenze significative tra le due modalità a causa di una fase aggiuntiva di compressione dell'intervallo dinamico (DRC, Dynamic Range Compression) nella lettura NGS, la quale riduce la variabilità del sequenziamento con reagenti SOMAmer diversi. Tale differenza è evidente anche in singoli campioni di plasma raggruppati, in cui la variazione intrapiattaforma (Figura 2A) è notevolmente inferiore a quella interpiattaforma (Figura 2B) prima del raccordo. Il raccordo consente tuttavia di superare questa differenza tra letture: i dati basati su array vengono infatti allineati al conteggio NGS con uno scostamento paragonabile alla variazione intrapiattaforma (Figura 2C), quantificato dal coefficiente di correlazione di Spearman e dal coefficiente di concordanza di Lin. Questa somiglianza quantitativa tra variazione intrapiattaforma e interpiattaforma dopo il raccordo dei dati è osservabile sull'intera distribuzione dei campioni di plasma raggruppati (Figura 3).

Per uniformare le concentrazioni di reagenti SOMAmer, si **utilizza la compressione dell'intervallo dinamico (DRC, Dynamic Range Compression)**, che riduce l'intervallo dinamico da 5 a 2 log. Senza DRC, i reagenti SOMAmer più abbondanti occuperebbero la maggior parte della capacità di lettura del sequenziamento. Ogni campione richiederebbe una profondità di sequenziamento estremamente elevata per coprire i reagenti SOMAmer non abbondanti e quantificarli accuratamente. Con la DRC, le sonde vengono raggruppate in base alle abbondanze osservate in assenza di DRC e ciascun gruppo viene compresso con un diverso rapporto fisso.

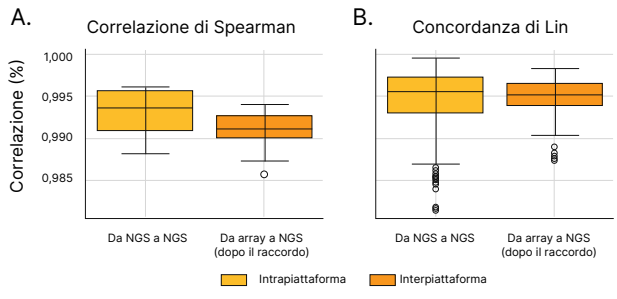


Figura 3: la distribuzione dei coefficienti di correlazione di Spearman e di concordanza di Lin nei campioni mostra l'effetto coerente del raccordo sull'allineamento dei dati

Distribuzione della (A) correlazione e della (B) concordanza dei dati delle piattaforme basate su NGS e su array in 18 confronti di campioni di plasma raggruppati. Le correlazioni e la concordanza delle letture da array a NGS dopo il raccordo per molti campioni di plasma raggruppati sono paragonabili alle correlazioni da NGS a NGS.

Il raccordo migliora la concordanza dei conteggi tra dati basati su NGS e su array sul pannello SOMAmer

Nell'ambito dello stesso studio, Illumina Protein Prep 9.5K, il saggio Plasma e il saggio SomaScan 11K sono stati utilizzati per analizzare una coorte di 32 campioni di plasma accoppiati su due diverse configurazioni Tecan di gestione dei liquidi. I campioni sono stati analizzati una seconda volta con il saggio Illumina Protein Prep 9.5K al fine di generare un livello di riferimento (baseline) per il confronto dei risultati intrapiattaforma (da NGS a NGS). I set di dati NGS sono stati normalizzati con la normalizzazione Illumina Protein Prep, mentre i set di dati basati su array sono stati normalizzati sia con la normalizzazione array (prima del raccordo) sia con la normalizzazione NGS (dopo il raccordo). Sono state valutate sia le correlazioni intrapiattaforma con l'NGS utilizzando le due configurazioni sia le correlazioni interpiattaforma dei risultati basati su NGS e array (Figura 4, Figura 5).

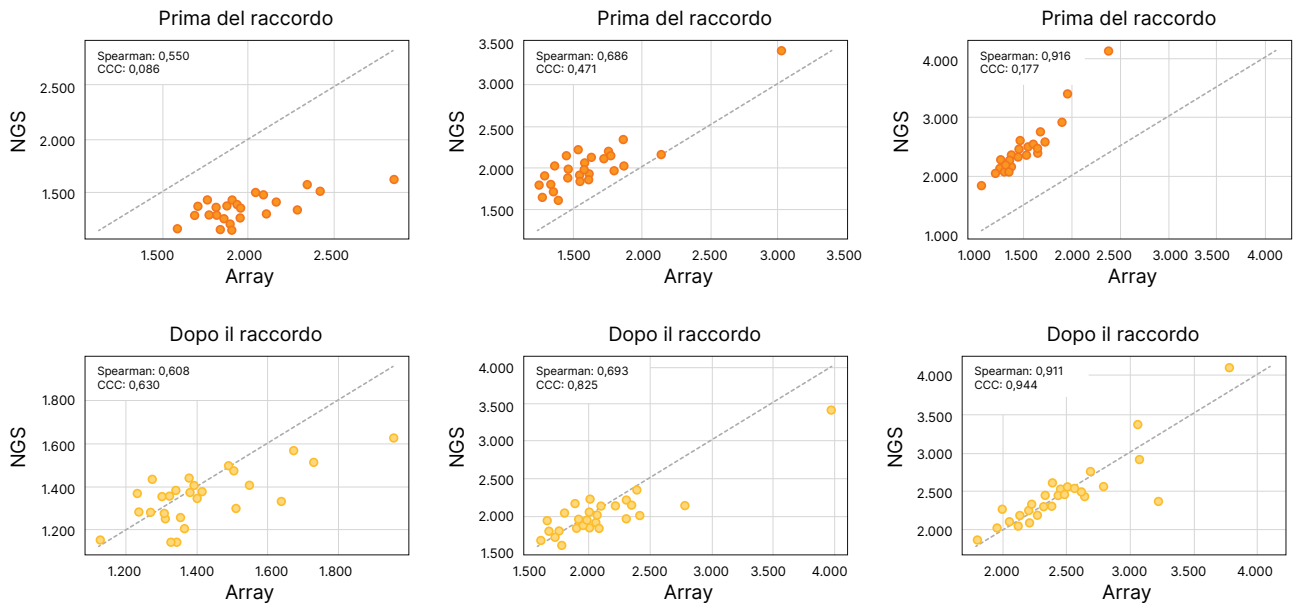


Figura 4: esempi di correlazioni da array a NGS per tre reagenti SOMAmer prima e dopo il raccordo

Il raccordo mantiene la correlazione e aumenta la concordanza dei conteggi basati su NGS e su array per ciascun reagente SOMAmer in una coorte biologica.

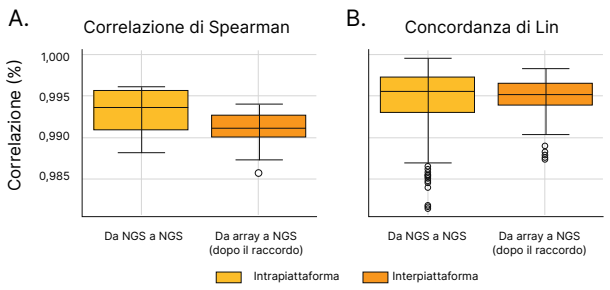


Figura 5: la distribuzione della correlazione di Spearman e dei coefficienti di correlazione di concordanza di Lin in tutti i reagenti SOMAmer umani dimostra una concordanza coerente con il raccordo

(A) La distribuzione delle correlazioni intrapiattaforma da NGS a NGS per tutti i reagenti SOMAmer è paragonabile alle correlazioni da NGS a array. (B) Con il raccordo, la distribuzione della concordanza interpiattaforma si avvicina a quella della concordanza intrapiattaforma.

Poiché entrambe le letture misurano la concentrazione relativa di reagenti SOMAmer, i singoli conteggi di questi ultimi nella coorte biologica sono generalmente correlati. Ciò vale a dire che i dati delle letture basate su array e delle letture basate su NGS, quantificati mediante il coefficiente di correlazione di Spearman, vanno nella stessa direzione anche senza raccordo (Figura 4, Figura 5A). Tuttavia, la concordanza dei conteggi dei reagenti SOMAmer, che misura la prossimità ai conteggi effettivi, risulta notevolmente inferiore nei confronti interpiattaforma (da array a NGS) rispetto a quelli intrapiattaforma (da NGS a NGS). Ciò è dovuto alle differenze tra le metodologie di lettura, ad esempio a livello di unità di misura, intervallo dinamico e riferimenti per la normalizzazione. Il raccordo allinea i dati basati su array al sistema di lettura NGS, consentendo il confronto tra i conteggi per reagente SOMAmer e per campione. La quantificazione avviene tramite il coefficiente di concordanza di Lin, che con il raccordo raggiunge i livelli intrapiattaforma (Figura 5B).

Riepilogo

Questa nota tecnica dimostra che il raccordo dei saggi consente di ottenere dati altamente correlati e concordanti tra piattaforme basate su NGS e su array. Un'efficace strategia di raccordo permette ai laboratori di integrare i dati già generati da SomaScan Platform con le nuove misurazioni di Illumina Protein Prep; ciò favorisce la scoperta di biomarcatori multipiattaforma, la pianificazione della transizione tra saggi e la continuità delle analisi con diverse pipeline proteomiche man mano che si evolvono.

Ulteriori informazioni →

Illumina Protein Prep

Rilevamento accurato e affidabile delle proteine con Illumina Protein Prep

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

© 2026 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.
M-GL-04147 ITA v1.0