Integración de datos proteómicos obtenidos mediante array y NGS

La conciliación de los resultados proteómicos basados en arrays en la plataforma SomaScan® con datos de NGS de Illumina Protein Prep permite un análisis unificado

Aproveche al máximo los conocimientos armonizando los datos de estudios multiplataforma.

Reduzca al mínimo los sesgos específicos de la plataforma con datos muy correlacionados y concordantes.

Optimice el análisis de datos proteómicos con Illumina Connected Multiomics.

Introducción

La integración de mediciones de abundancia de proteínas en diferentes tecnologías de ensayo es un requisito cada vez más importante para los estudios multiplataforma, los análisis longitudinales y el mantenimiento de la continuidad de los datos. Con el aumento del tamaño de los estudios y la evolución de las tecnologías, hay una creciente necesidad de interoperabilidad entre los conjuntos de datos existentes y los recién generados en las plataformas de análisis proteómico. El ensayo Illumina Protein Prep y la plataforma SomaLogic® SomaScan proporcionan información proteómica de alto valor, aunque difieren materialmente en la química de medición, las estructuras de datos y las convenciones de normalización. La combinación de datos de estas dos plataformas es posible con una estrategia de conciliación clara que ofrece una rica información biológica, al tiempo que reduce al mínimo los sesgos específicos de la plataforma.

Ambos ensayos utilizan reactivos SOMAmer® (aptámero modificado de velocidad lenta) para la identificación de proteínas, pero utilizan procesos moleculares fundamentalmente diferentes para generar datos proteómicos. Illumina Protein Prep utiliza la secuenciación de nueva generación (NGS, next-generation sequencing) para cuantificar y detectar proteínas, mientras que la plataforma SomaScan utiliza una lectura en array. Dado que ambas lecturas están optimizadas para un ruido de fondo mínimo, muestran una buena correlación al comparar un analito individual entre varias muestras biológicas. Sin embargo, debido a las diferencias entre ambos ensayos, incluidas sus unidades, el rango dinámico y las referencias de normalización (figura 1), los valores y los recuentos relativos entre analitos diferirán entre las lecturas.

El objetivo de la conciliación de los datos de Illumina Protein Prep y la plataforma SomaScan es armonizar las mediciones de abundancia para que los resultados se puedan comparar más fácilmente entre las plataformas.

En esta nota técnica se describe el marco conceptual para conciliar los datos de Illumina Protein Prep con los conjuntos de datos existentes de la plataforma SomaScan mediante el software Illumina Connected Multiomics. Proporcionamos un enfoque reproducible, transparente y con fundamentos científicos que permite a Illumina y a los equipos de colaboración aprovechar los conjuntos de datos de la plataforma SomaScan existentes, al tiempo que incorporan nuevas mediciones de proteínas de Illumina en análisis unificados.

Normalización de los datos de Illumina Protein Prep

Los datos obtenidos con array se pueden conciliar con los recuentos mediante NGS aplicando el proceso de normalización de datos de NGS estándar directamente a los recuentos de array sin procesar.

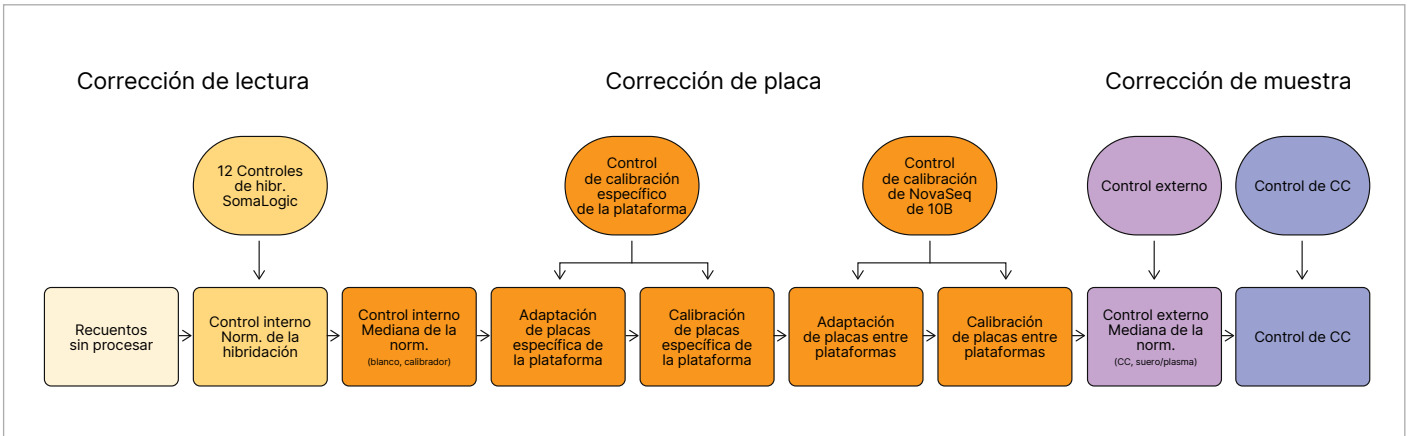


Figura 1: Pasos de normalización secuencial en el proceso de análisis de Illumina Protein Prep

La normalización de datos incluye pasos de corrección de lecturas, corrección de placas y corrección de muestras para reducir la variabilidad que se puede introducir durante la hibridación, la preparación de librerías o las partes de manipulación de líquidos del ensayo. Estos pasos se realizan secuencialmente y se produce un archivo ADAT individual para cada paso de normalización.

La normalización de datos corrige la variabilidad intra e interensayo, estandariza los datos y permite comparaciones significativas entre muestras, placas, tandas de ensayo y plataformas. El proceso de normalización de datos consta de tres pasos secuenciales que corrigen la variabilidad técnica que se puede introducir durante la hibridación, la preparación de librerías de NGS o en las partes de manipulación de líquidos del ensayo (figura 1).

El proceso de normalización incluye los siguientes pasos:

- **Corrección de lectura**: Este paso utiliza controles de reactivos SOMAmer para reducir la variabilidad técnica introducida en el proceso de preparación de librerías de NGS.
- **Corrección de placas**: Este paso utiliza controles de calibrador como controles positivos para corregir sesgos entre placas; una referencia externa proporcionada por Illumina permite la comparación entre placas de Illumina Protein Prep de diferentes tandas de ensayo.
- **Corrección de muestras**: Este paso utiliza la abundancia de proteínas en cada muestra para reducir la variabilidad técnica introducida durante la cuantificación de proteínas y corregir las diferencias en la concentración general de proteínas.

Estrategia de conciliación

El objetivo de la conciliación de los conjuntos de datos es conseguir que las mediciones de abundancia de proteínas a partir de los valores de unidades de fluorescencia relativas (RFU, relative fluorescence unit) generados por la plataforma SomaScan estén en la misma escala que los datos del ensayo Illumina Protein Prep mediante NGS para poder interpretarlos conjuntamente (figura 2). La conciliación se puede aplicar durante la integración de datos retrospectivos, la expansión continua de cohortes y los estudios en los que los ensayos Illumina Protein Prep y SomaScan se utilizan en paralelo o secuencialmente.

Los ensayos Illumina Protein Prep 9.5K y SomaScan 11K utilizan controles de calibrador idénticos para la corrección de placas. Como resultado, cuando se aplica la referencia de calibración de NGS multiplataforma a los datos del array en Illumina Connected Multiomics, conciliará estos datos a la escala de NGS, haciendo que los datos de ambos ensayos sean comparables. La evaluación del rendimiento del array se conserva en esta normalización a través del paso de calibración específico de la plataforma en el que se comparan los calibradores con una referencia de array y se calcula un criterio de medición de control de calidad (CC).

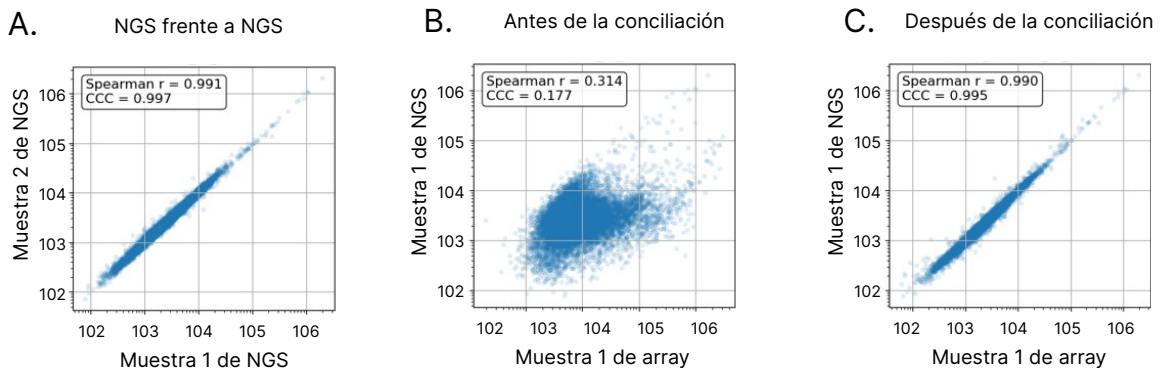


Figura 2: El coeficiente de correlación de Spearman entre los datos obtenidos mediante NGS y array antes y después de la conciliación para una sola muestra demuestra la armonización de los datos

Muestras de plasma agrupadas representativas de diferentes tandas de ensayo que muestran (A) una excelente correlación entre los datos de experimentos de NGS individuales, (B) una correlación deficiente entre los datos de la NGS y el array sin conciliación, principalmente debido a la compresión del rango dinámico específico de la NGS y (C) una correlación mejorada entre los datos de la NGS y el array, con aumento a niveles intraplataforma gracias a la conciliación.

La conciliación proporciona datos de NGS y array con una correlación muy alta

Para evaluar esta estrategia de conciliación, se analizaron muestras de plasma agrupadas emparejadas tanto del ensayo Illumina Protein Prep 9.5K Plasma como del ensayo SomaScan 11K. Las muestras se analizaron por segunda vez en Illumina Protein Prep para generar una referencia para comparar los resultados interplataforma (array frente a NGS) con los resultados intraplataforma (NGS frente a NGS). Los conjuntos de datos de NGS se normalizaron mediante la normalización de Illumina Protein Prep, mientras que el conjunto de datos del array se normalizó mediante la normalización de array antes de la conciliación y la normalización de NGS después de la conciliación (figura 2).

La conciliación alinea las diferencias entre las lecturas obtenidas mediante NGS y mediante array para producir datos comparables (figura 2). Un aspecto importante de esta corrección es el redimensionamiento relativo entre los reactivos SOMAmer, que difiere significativamente entre la NGS y el array debido a un paso de compresión de rango dinámico (DRC, dynamic range compression) adicional en la lectura de la NGS que reduce la variabilidad de secuenciación entre los reactivos SOMAmer.

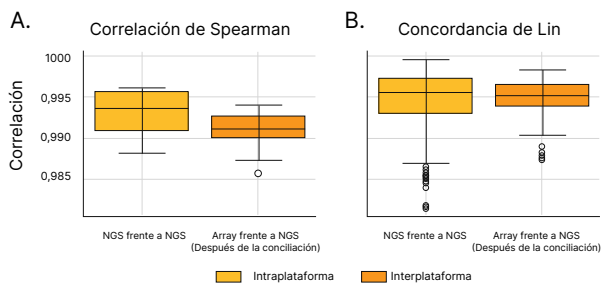


Figura 3: La distribución de la correlación de Spearman y los coeficientes de correlación de la concordancia de Lin entre las muestras revela un efecto uniforme de la conciliación en la armonización de los datos

Distribución de la correlación (A) y la concordancia (B) en 18 comparaciones de plasma agrupadas entre plataformas basadas en NGS y array. Las correlaciones y la concordancia de las lecturas en array frente a NGS después de la conciliación entre muchas muestras de plasma agrupadas son comparables a las correlaciones de NGS frente a NGS.

La compresión del rango dinámico (DRC) se utiliza para igualar las concentraciones de reactivo SOMAmer de un rango dinámico de 5 log a uno de 2 log. Sin DRC, los reactivos SOMAmer más abundantes ocuparían la mayor parte de la capacidad de lectura de secuenciación. Cada muestra requeriría una profundidad de secuenciación extremadamente alta para cubrir los reactivos SOMAmer no abundantes y cuantificarlos con precisión. Con DRC, las sondas se agrupan en función de sus abundancias observadas sin condiciones de DRC y cada grupo se comprime con una relación de ajuste diferente.

Esta diferencia es evidente incluso en una única muestra de plasma agrupada, donde la variación intraplataforma (figura 2A) es notablemente menor que la variación interplataforma (figura 2B) antes de la conciliación. Sin embargo, la conciliación supera esta diferencia de lectura y alinea los datos del array con el recuento de NGS a una escala comparable a la variación intraplataforma (figura 2C), cuantificada por el coeficiente de correlación de Spearman y el coeficiente de concordancia de Lin. Esta similitud cuantitativa entre la variación intraplataforma e interplataforma después de la conciliación se observa en toda la distribución de las comparaciones de muestras de plasma agrupadas (figura 3).

La conciliación mejora la concordancia de recuentos entre la NGS y los datos del array en todo el panel SOMAmer

Como parte del mismo estudio, se analizó una cohorte de 32 muestras de plasma emparejadas con los ensayos Illumina Protein Prep 9.5K, Plasma y SomaScan 11K en dos configuraciones diferentes del robot de manipulación de líquidos Tecan. Las muestras se analizaron por segunda vez con el ensayo Illumina Protein Prep 9.5K para generar una referencia para comparar los resultados intraplataforma (NGS frente a NGS). Los conjuntos de datos de NGS se normalizaron mediante la normalización de Illumina Protein Prep, mientras que los conjuntos de datos del array se normalizaron mediante la normalización de array (antes de la conciliación) y la normalización de NGS (después de la conciliación).

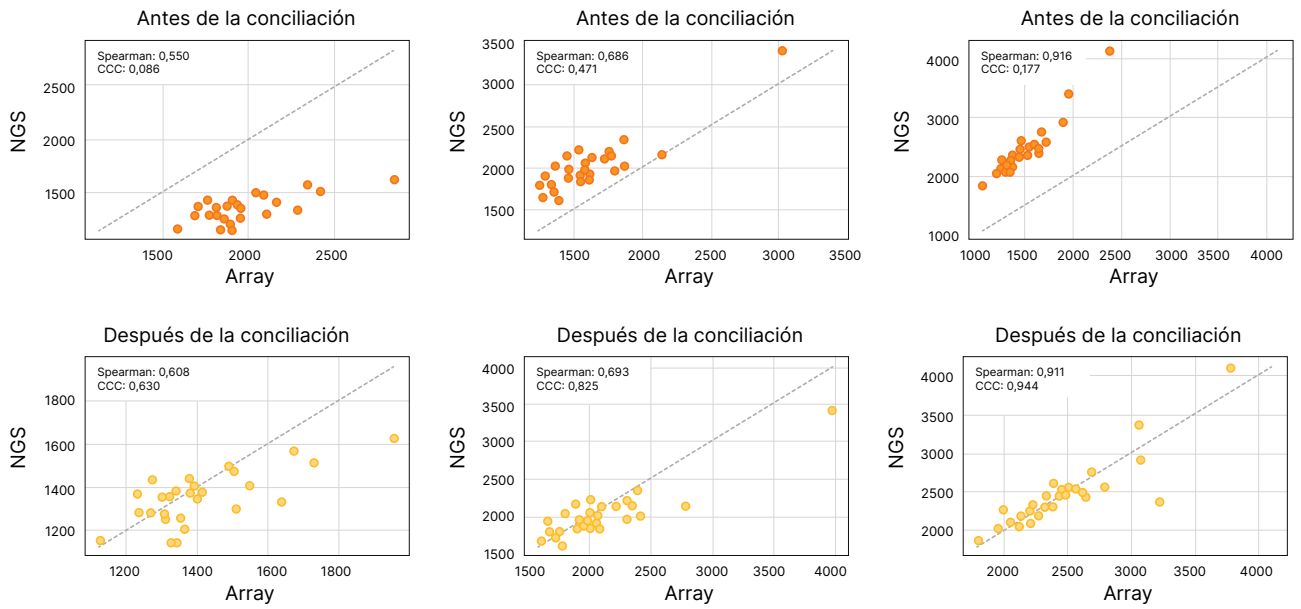


Figura 4: Ejemplos de correlaciones de array frente a NGS para tres reactivos SOMAmer antes y después de la conciliación
La conciliación mantiene la correlación y aumenta la concordancia entre los recuentos de la NGS y el array para cada reactivo SOMAmer en una cohorte biológica.

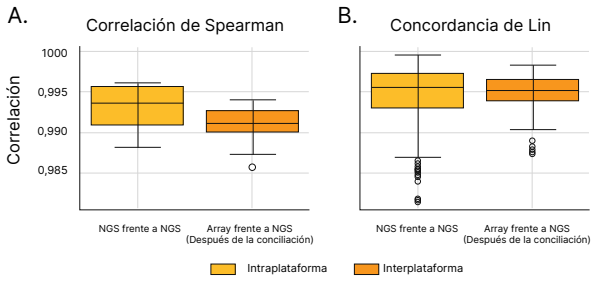


Figura 5: La distribución de la correlación de Spearman y los coeficientes de correlación de la concordancia de Lin en todos los reactivos SOMAmer humanos demuestra una concordancia uniforme con la conciliación

(A) La distribución de las correlaciones NGS-NGS intraplataforma entre todos los reactivos SOMAmer es comparable a las correlaciones NGS-array. (B) Con la conciliación, la distribución de la concordancia interplataforma se aproxima a la de la concordancia intraplataforma.

Se evaluaron las correlaciones intraplataforma con NGS utilizando ambas configuraciones y las correlaciones interplataforma entre los resultados de la NGS y del array (figura 4, figura 5).

Dado que ambas lecturas miden la concentración relativa del reactivo SOMAmer, los recuentos individuales de reactivos SOMAmer en la cohorte biológica suelen mostrar correlación. Es decir, los datos se mueven en la misma dirección entre las lecturas en array y mediante NGS, incluso sin conciliación, y se cuantifican utilizando el coeficiente de correlación de Spearman (figura 4, figura 5A). Sin embargo, la concordancia del recuento de reactivo SOMAmer, que es una medida de lo cerca que están los recuentos reales, es significativamente menor en las comparaciones interplataforma (array frente a NGS) que en las comparaciones intraplataforma (NGS frente a NGS). Esto se debe a las diferencias entre las metodologías de lectura, incluidas las unidades, el rango dinámico y las referencias de normalización. La conciliación armoniza los datos del array con la escala de lectura de la NGS, lo que permite la comparación entre recuentos de reactivo SOMAmer por muestra. Esto se cuantifica mediante el coeficiente de correlación de la concordancia de Lin, que aumenta a niveles intraplataforma con la conciliación (figura 5B).

Resumen

En esta nota técnica se demuestra que la conciliación del ensayo proporciona datos con una alta correlación y concordancia entre las plataformas basadas en NGS y array. Una estrategia de conciliación eficaz permite a los laboratorios ampliar los datos generados previamente en la plataforma SomaScan con nuevas mediciones de Illumina Protein Prep, lo que permite el descubrimiento de biomarcadores multiplataforma, respalda la planificación de la transición de ensayos y garantiza la continuidad de los análisis en los cambiantes procesos proteómicos.

Más información →

Illumina Protein Prep

Consiga una detección de proteínas precisa y fiable con Illumina Protein Prep

1 800 809 4566 (llamada gratuita, EE. UU.) | tel.: +1 858 202 4566
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. Todos los derechos reservados. Todas las marcas comerciales pertenecen a Illumina, Inc. o a sus respectivos propietarios. Si desea consultar información específica sobre las marcas comerciales, consulte www.illumina.com/company/legal.html.
M-GL-04147 ESP v1.0