Integration von per Array und NGS ermittelten Proteomikdaten

Die Verknüpfung der per Array ermittelten Proteomikausgabe der SomaScan® Platform mit Illumina Protein Prep NGS-Daten ermöglicht eine einheitliche Analyse

Maximaler Erkenntnisgewinn durch Harmonisierung von Daten aus Multiplattformstudien

Minimierung plattformspezifischer Verzerrung dank hochgradig korrelierter und konkordanter Daten

Optimierung der Analyse von Proteomikdaten mit Illumina Connected Multiomics

Einleitung

Die Zusammenführung von Proteinabundanzmessungen unterschiedlicher Assaytechnologien gewinnt bei Multiplattformstudien, Längsschnittanalysen und Aufrechterhaltung der Datenkontinuität zunehmend an Bedeutung. Vor dem Hintergrund zunehmender Studienumfänge und des technologischen Fortschritts besteht ein wachsender Bedarf in Bezug auf die gemeinsame Nutzbarkeit vorhandener und neu generierter Datensätze über unterschiedliche Proteomik-Analyseplattformen hinweg. Der Illumina Protein Prep-Assay und die SomaLogic® SomaScan Platform bieten hochwertige Proteomikinformationen, unterscheiden sich jedoch wesentlich in Hinblick auf Erfassungschemie, Datenstrukturen und Normalisierungskonventionen. Die Kombination von Daten aus diesen beiden Plattformen erfordert eine klare Verknüpfungsstrategie, die umfassende biologische Erkenntnisse liefert und zugleich plattformspezifische Verzerrungen minimiert.

Beide Assays verwenden SOMAmer®-Reagenzien (Slow Off-Rate Modified Aptamer) für den Proteinnachweis, nutzen jedoch grundverschiedene molekulare Prozesse zur Generierung der Proteomikdaten. Illumina Protein Prep verwendet NGS (Next-Generation Sequencing, Sequenzierung der nächsten Generation) für den Nachweis und die Quantifizierung von Proteinen, während die SomaScan Platform eine per Array generierte Ausgabe nutzt. Da beide Ausgaben für einen minimalen Hintergrund optimiert sind, zeigen sie beim Vergleich eines einzelnen Analyten mehrerer biologischer Proben eine hohe Korrelation. Aufgrund der Unterschiede zwischen den beiden Assays, einschließlich deren Einheiten, des Dynamikbereichs und der Normalisierungsreferenzen (Abbildung 1), unterscheiden sich die Werte und relativen Anzahlen der Analyten jedoch zwischen den Ausgaben.

Das Ziel der Verknüpfung von Illumina Protein Prep- und SomaScan Platform-Daten ist die Harmonisierung der Abundanzwerte, sodass sich die Ergebnisse einfacher plattformübergreifend vergleichen lassen.

Im vorliegenden technischen Hinweis wird der konzeptionelle Rahmen für die Verknüpfung von Illumina Protein Prep-Daten mit vorhandenen SomaScan Platform-Datensätzen mithilfe der Illumina Connected Multiomics-Software erläutert. Wir bieten ein reproduzierbares, transparentes und wissenschaftlich fundiertes Verfahren, mit dem Illumina- und Kooperationsteams vorhandene SomaScan Platform-Datensätze nutzen und neue Illumina-Proteinwerte in einheitlichen Analysen zusammenführen können.

Illumina Protein Prep-Datennormalisierung

Arraydaten lassen sich mithilfe von NGS-Zählungen verknüpfen, indem die standardmäßige NGS-Datennormalisierungspipeline direkt auf Array-Rohzählungen angewendet wird.

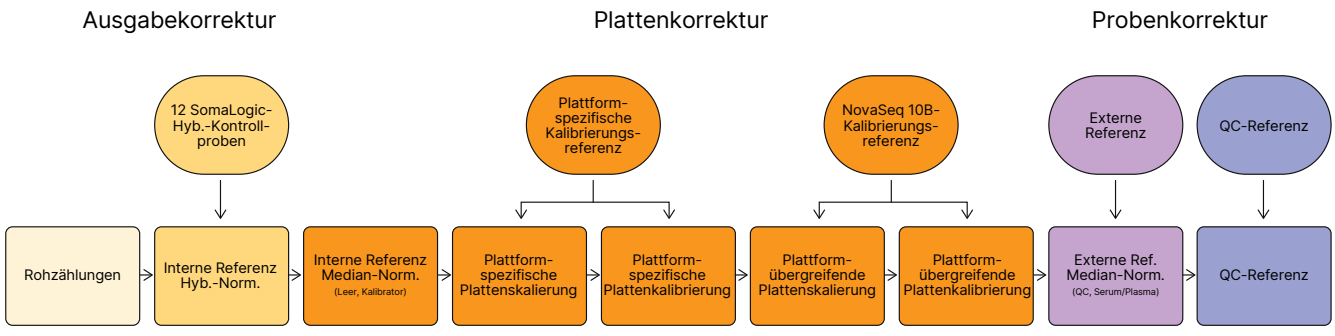


Abbildung 1: Sequenzielle Normalisierungsschritte in der Analysepipeline von Illumina Protein Prep

Die Datennormalisierung umfasst Ausgabe-, Platten- und Probenkorrekturschritte. Diese reduzieren die durch Hybridisierung, Bibliotheksvorbereitung oder Liquid-Handling-Prozesse des Assays bedingte Variabilität. Diese Schritte werden nacheinander durchgeführt. Für jeden Normalisierungsschritt wird eine eigene ADAT-Datei erstellt.

Die Datennormalisierung korrigiert die Intra- und Interassayvariabilität, standardisiert Daten und ermöglicht den aussagekräftigen Vergleich zwischen Proben, Platten, Assayläufen und Plattformen. Der Datennormalisierungsprozess umfasst drei aufeinanderfolgende Schritte zur Korrektur der technischen Variabilität, die sich während der Assayabschnitte Hybridisierung, Vorbereitung der NGS-Bibliothek und Liquid-Handling ergeben kann (Abbildung 1).

Der Normalisierungsprozess umfasst folgende Schritte:

- **Ausgabekorrektur**: In diesem Schritt wird die während des NGS-Bibliotheksvorbereitungsproze sses entstandene technische Variabilität anhand von SOMAmer-Reagenzkontrollproben reduziert.
- **Plattenkorrektur**: In diesem Schritt werden Kalibratorkontrollproben als Positivkontrollproben zur Korrektur von Verzerrungen zwischen Platten verwendet. Eine von Illumina bereitgestellte externe Referenz ermöglicht den Vergleich zwischen Illumina Protein Prep-Platten aus unterschiedlichen Assayläufen.
- **Probenkorrektur**: In diesem Schritt werden anhand der Proteinabundanz der einzelnen Proben die während der Proteinquantifizierung entstandene technische Variabilität reduziert und Unterschiede in der Gesamtproteinkonzentration korrigiert.

Verknüpfungsstrategie

Der Zweck der Verknüpfung von Datensätzen besteht darin, anhand von auf der SomaScan Platform generierten RFU-Werten (Relative Fluorescence Unit, relative Fluoreszenzeinheit) gemessene Proteinabundanzen auf einer Skala zu platzieren, die gemeinsam mit dem NGS-basierten Illumina Protein Prep-Assay interpretiert wird (Abbildung 2). Die Verknüpfung kann während der retrospektiven Datenintegration, während der laufenden Kohortenerweiterung und während Studien durchgeführt werden, in denen Illumina Protein Prep- und SomaScan-Assays parallel oder sequentiell zum Einsatz kommen.

Der Illumina Protein Prep 9.5K- und der SomaScan 11K-Assay verwenden identische Kalibratorkontrollproben für die Plattenkorrektur. Wenn die plattformübergreifende NGS-Kalibrierungsreferenz auf Arraydaten in Illumina Connected Multiomics angewendet wird, werden diese Daten über die NGS-Skala verknüpft, wodurch die Daten aus diesen beiden Assays vergleichbar sind. Die Beurteilung der Arrayperformance bleibt in dieser Normalisierung über den plattformspezifischen Kalibrierungsschritt erhalten, in dem die Kalibratoren mit einer Arrayreferenz verglichen und eine QC-Metrik (Quality Control, Qualitätssicherung) berechnet werden.

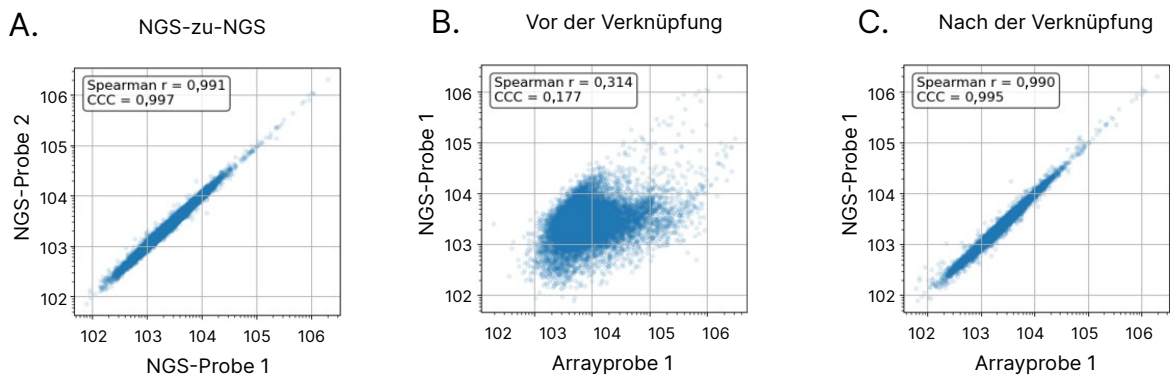


Abbildung 2: Spearman-Korrelationskoeffizient zwischen per NGS und per Array generierten Daten vor und nach der Verknüpfung für eine einzelne Probe zur Demonstration der Datenharmonisierung

Repräsentative gepoolte Plasmaproben aus unterschiedlichen Assayläufen, die eine (A) herausragende Korrelation zwischen Daten einzelner NGS-Läufe, (B) eine geringfügige Korrelation zwischen nicht verknüpften NGS- und Arraydaten (hauptsächlich aufgrund der NGS-spezifischen Dynamikbereichskompression) und (C) eine verbesserte Korrelation zwischen NGS- und Arraydaten zeigen, die dank der Verknüpfung auf Intraplattformwerte ansteigt.

NGS- und Arraydaten mit hoher Korrelation dank Verknüpfung

Zur Beurteilung dieser Verknüpfungsstrategie wurden gepaarte gepoolte Plasmaproben sowohl mit dem Illumina Protein Prep 9.5K Plasma-Assay als auch mit dem SomaScan 11K-Assay getestet. Die Proben wurden auf Illumina Protein Prep ein zweites Mal getestet, um Basiswerte für den Vergleich der Interplattformergebnisse (Array-zu-NGS) mit Intraplattformergebnissen (NGS-zu-NGS) zu generieren. NGS-Datensätze wurden mit der Illumina Protein Prep-Normalisierung normalisiert, während der Arraydatensatz mit der Arraynormalisierung vor der Verknüpfung und der NGS-Normalisierung nach der Verknüpfung normalisiert wurde (Abbildung 2).

Die Verknüpfung aligniert Unterschiede zwischen NGS- und Array-Ausgaben, woraus sich vergleichbare Daten ergeben (Abbildung 2). Ein wichtiger Aspekt dieser Korrektur ist die relative Skalierung zwischen SOMAmer-Reagenzien, die sich aufgrund eines zusätzlichen DRC-Schritts (Dynamic Range Compression, Dynamikbereichkompression) bei der NGS-Ausgabe, der die Sequenzierungsvariabilität zwischen SOMAmer-Reagenzien reduziert, bei NGS und Array signifikant unterscheidet.

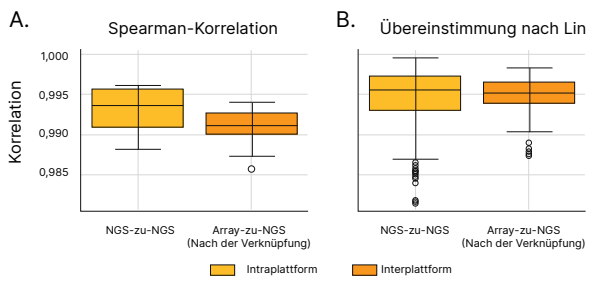


Abbildung 3: Die Verteilung der Spearman-Korrelation und der Konkordanz-Korrelationskoeffizienten nach Lin über die Proben hinweg zeigt eine konsistente Wirkung der Verknüpfung auf das Alignment der Daten

Verteilung der (A) Korrelation und (B) Konkordanz bei 18 gepoolten Plasmavergleichen zwischen NGS- und Array-Plattformen. Korrelationen und Konkordanz von Array-zu-NGS-Ausgaben nach der Verknüpfung über zahlreiche gepoolte Plasmaproben hinweg sind vergleichbar mit NGS-zu-NGS-Korrelationen.

DRC (Dynamic Range Compression, Dynamikbereichkompression) dient dem Ausgleich der SOMAmer-Reagenzkonzentrationen von einem Dynamikbereich von 5-log auf 2-log. Ohne DRC würden die SOMAmer-Reagenzien mit der höchsten Abundanz den Großteil der Sequenzierungsausgabekapazität ausmachen. Dadurch wäre für die einzelnen Proben eine extrem hohe Sequenzierungstiefe erforderlich, um die SOMAmer-Reagenzien von geringer Abundanz zu erfassen und genau zu quantifizieren. Bei der DRC werden Sonden anhand der ohne DRC-Bedingung erfassten Abundanz gruppiert. Die einzelnen Gruppen werden in einem anderen festgelegten Verhältnis komprimiert.

Dieser Unterschied zeigt sich auch bei einer einzelnen gepoolten Plasmaprobe, bei der die Intraplattformvarianz (Abbildung 2A) deutlich geringer ist als die Interplattformvarianz (Abbildung 2B) vor der Verknüpfung. Die Verknüpfung gleicht diesen Ausgabeunterschied jedoch aus und aligniert die Arraydaten sowie die NGS-Zählung auf ein mit der Intraplattformvarianz vergleichbares Niveau (Abbildung 2C). Die Quantifizierung erfolgt anhand des Korrelationskoeffizienten nach Spearman und des Konkordanzkoeffizienten nach Lin. Diese quantitative Ähnlichkeit zwischen Intra- und Interplattformvarianz nach der Verknüpfung wird über die gesamte Verteilung gepoolter Plasmaprobenvergleiche beobachtet (Abbildung 3).

Verknüpfung verbessert die Zählungskonkordanz zwischen NGS- und Arraydaten im SOMAmer-Panel

Im Rahmen derselben Studie wurde eine Kohorte von 32 gepaarten Plasmaproben mit dem Illumina Protein Prep 9.5K, dem Plasma-Assay und dem SomaScan 11K-Assay auf zwei verschiedenen Tecan-Liquid-Handler-Konfigurationen getestet. Die Proben wurden mit dem Illumina Protein Prep 9.5K-Assay ein zweites Mal getestet, um Basiswerte für Intraplattformergebnisse (NGS-zu-NGS) zu generieren. NGS-Datensätze wurden mit der Illumina Protein Prep-Normalisierung normalisiert, während Arraydatensätze sowohl mit der Arraynormalisierung (vor der Verknüpfung) und mit der NGS-Normalisierung (nach der Verknüpfung) normalisiert wurden.

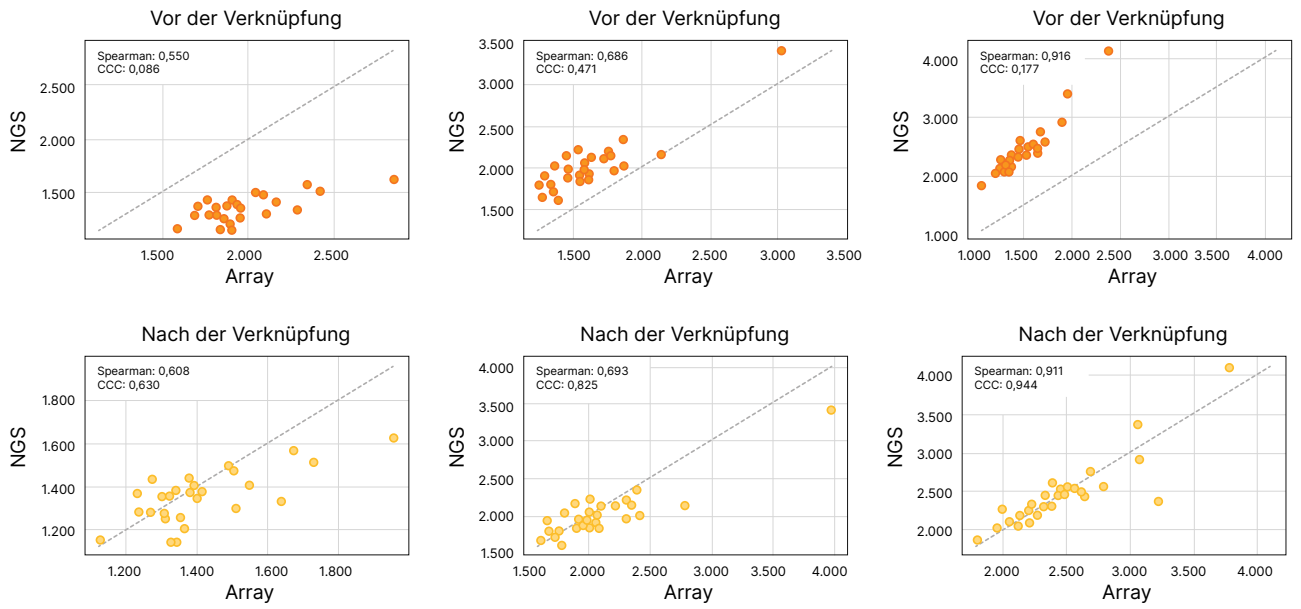


Abbildung 4: Beispiele für Array-zu-NGS-Korrelationen für drei SOMAmer-Reagenzien vor und nach der Verknüpfung
Die Verknüpfung hält die Korrelation aufrecht und erhöht die Konkordanz zwischen NGS- und Array-Zählungen für jedes SOMAmer-Reagenz in einer biologischen Kohorte.

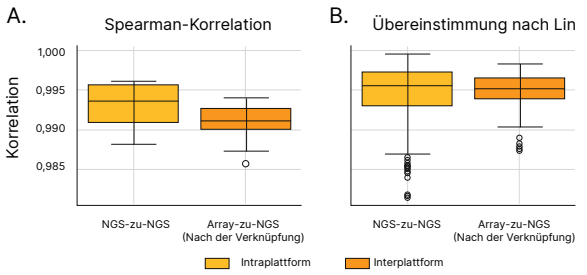


Abbildung 5: Die Verteilung der Spearman-Korrelation und der Konkordanz-Korrelationskoeffizienten nach Lin für alle SOMAmer-Humanreagenzien zeigt mit der Verknüpfung eine konsistente Konkordanz

(A) Die Verteilung der NGS-zu-NGS-Intraplattform-korrelationen für alle SOMAmer-Reagenzien ist mit den NGS-zu-Array-Korrelationen vergleichbar. (B) Bei der Verknüpfung nähert sich die Verteilung der Interplattformkonkordanz der der Intraplattformkonkordanz an.

Intraplattformkorrelationen mit NGS wurden anhand der beiden Konfigurationen und Interplattformkorrelationen zwischen NGS- und Array-Ausgaben bewertet (Abbildung 4, Abbildung 5).

Da beide Ausgaben die relative SOMAmer-Reagenzkonzentration messen, korrelieren die einzelnen SOMAmer-Reagenzzählungen in der biologischen Kohorte in der Regel. Das heißt, die Daten bewegen sich bei Array- und NGS-Ausgaben auch ohne Verknüpfung in die gleiche Richtung und werden mit dem Korrelationskoeffizienten nach Spearman quantifiziert (Abbildung 4, Abbildung 5A). Die Konkordanz der SOMAmer-Reagenzzählung, ein Maß für die Nähe der tatsächlichen Anzahlwerte, ist jedoch bei Interplattformvergleichen (Array-zu-NGS) signifikant geringer als bei Intraplattformvergleichen (NGS-zu-NGS). Die Ursache hierfür sind Unterschiede bei den Ausgabemethoden, darunter Einheiten, Dynamikbereich und Normalisierungsreferenzen. Die Verknüpfung aligniert die Arraydaten auf die NGS-Ausgabeskala und ermöglicht den Vergleich zwischen den Zählungen je SOMAmer-Reagenz pro Probe. Dies wird anhand des Konkordanz-Korrelationskoeffizienten nach Lin quantifiziert, der mit der Verknüpfung auf Intraplattformwerte ansteigt (Abbildung 5B).

Zusammenfassung

Der vorliegende technische Hinweis zeigt, dass die Assayverknüpfung hochgradig korrelierte und konkordante Daten aus NGS- und Arrayplattformen liefert. Eine wirksame Verknüpfungsstrategie ermöglicht Laboren, vorhandene SomaScan-Plattformdaten durch neue Illumina Protein Prep-Messwerte zu erweitern, was die Bestimmung von Multiplattform-Biomarkern ermöglicht, die Planung von Assaywechseln vereinfacht und die Kontinuität von Analysen während der Weiterentwicklung von Proteomikpipelines gewährleistet.

Weitere Informationen →

Illumina Protein Prep

Genauer und zuverlässiger Proteinnachweis mit Illumina Protein Prep

1 800 8094566 (USA, gebührenfrei) | +1 858 2024566 (Tel. außerhalb der USA)
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. Alle Rechte vorbehalten. Alle Marken sind Eigentum von Illumina, Inc. bzw. der jeweiligen Inhaber. Spezifische Informationen zu Marken finden Sie unter www.illumina.com/company/legal.html.
M-GL-04147 DEU v1.0