Integração de dados proteômicos com base em matriz e no NGS

A amplificação em ponte de saídas proteômicas com base em matriz do SomaScan® Platform com dados NGS do Illumina Protein Prep permite uma análise unificada

Maximize os insights harmonizando dados de estudos multiplataforma

Minimize vieses específicos da plataforma com dados altamente correlacionados e concordantes

Simplifique a análise de dados de proteômica com o Illumina Connected Multiomics

Introdução

A integração de medições de abundância de proteínas em diferentes tecnologias de ensaio é um requisito cada vez mais importante para estudos multiplataforma, análises longitudinais e manutenção da continuidade dos dados. À medida que os tamanhos dos estudos aumentam e as tecnologias evoluem, há uma necessidade crescente de interoperabilidade entre os conjuntos de dados existentes e os recém-gerados nas plataformas de análise proteômica. O ensaio Illumina Protein Prep e o SomaLogic® SomaScan Platform fornecem informações proteômicas de alto valor, mas diferem materialmente na química da medição, nas estruturas de dados e nas convenções de normalização. A combinação de dados dessas duas plataformas é possível com uma estratégia clara de amplificação em pontes que forneça insights biológicos ricos, ao mesmo tempo minimizando os vieses específicos da plataforma.

Ambos os ensaios usam SOMAmer® Reagents (aptâmero modificado de baixa velocidade) para identificação de proteínas, mas usam processos moleculares fundamentalmente diferentes para gerar dados proteômicos. O Illumina Proteinom Prep usa sequenciamento de última geração (NGS) para quantificar e detectar proteínas, enquanto o SomaScan Platform usa uma leitura com base em matriz. Como ambas as leituras são otimizadas para um fundo mínimo, elas se correlacionam bem na comparação de um analito individual em várias amostras biológicas.

No entanto, devido a diferenças entre os dois ensaios, incluindo suas unidades, o intervalo dinâmico e referências de normalização (Figura 1), os valores e as contagens relativas entre os analitos serão diferentes entre as leituras.

O objetivo da amplificação em ponte entre os dados do Illumina Protein Prep e do SomaScan Platform é harmonizar as medições de abundância para que os resultados possam ser comparados mais facilmente entre as plataformas.

Esta nota técnica descreve a estrutura conceitual para a amplificação em ponte dos dados do Illumina Protein Prep com conjuntos de dados existentes do SomaScan Platform usando o software Illumina Connected Multiomics. Fornecemos uma abordagem reproduzível, transparente e cientificamente fundamentada que permite que a Illumina e as equipes colaborativas aproveitem conjuntos de dados existentes do SomaScan Platform, ao mesmo tempo incorporando novas medições de proteína da Illumina em análises unificadas.

Normalização de dados do Illumina Protein Prep

Dados com base em matriz podem ser combinados com as contagens do NGS aplicando o pipeline de normalização de dados do NGS padrão diretamente às contagens brutas de matrizes.

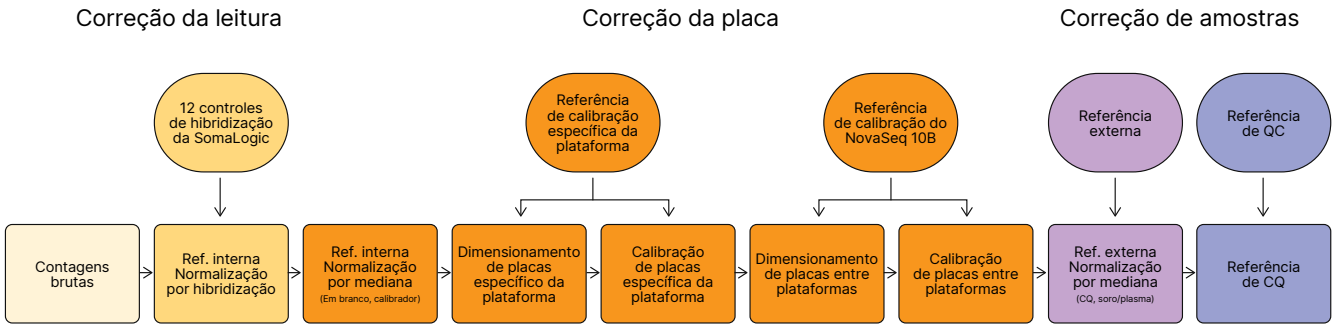


Figura 1: Etapas de normalização sequencial no pipeline de análise do Illumina Protein Prep

A normalização de dados inclui correção de leituras, correção de placas e etapas de correção de amostras para reduzir a variabilidade que pode ser introduzida durante a hibridização, a preparação de bibliotecas ou porções de manuseio de líquidos do ensaio. Essas etapas são realizadas sequencialmente e um arquivo ADAT individual é produzido para cada etapa de normalização.

A normalização de dados corrige a variabilidade intraensaio e interensaio, padroniza dados e permite comparações significativas entre amostras, placas, corridas de ensaios e plataformas. O processo de normalização de dados consiste em três etapas sequenciais que corrigem a variabilidade técnica que pode ser introduzida durante a hibridização, a preparação da biblioteca do NGS ou porções de manuseio de líquidos do ensaio (Figura 1).

O processo de normalização inclui as seguintes etapas:

- **Correção da leitura**: essa etapa usa controles do SOMAmer Reagent para reduzir a variabilidade técnica introduzida no processo de preparação da biblioteca do NGS
- **Correção da placa**: essa etapa usa controles do calibrador como controles positivos para corrigir vieses entre placas. Uma referência externa fornecida pela Illumina permite a comparação entre placas do Illumina Protein Prep com base em diferentes corridas de ensaio
- **Correção de amostras**: essa etapa usa a abundância de proteínas em cada amostra para reduzir a variabilidade técnica introduzida durante a quantificação de proteínas e corrigir as diferenças na concentração geral de proteínas

Estratégia de amplificação em ponte

A finalidade da amplificação em ponte dos conjuntos de dados é colocar as medições de abundância de proteínas originárias dos valores da unidade de fluorescência relativa (RFU, relative fluorescence unit) gerados pelo SomaScan Platform em uma escala que é interpretada com o ensaio Illumina Protein Prep baseado em NGS (Figura 2). A amplificação em ponte pode ser aplicada durante a integração retrospectiva de dados, a expansão contínua de coortes e estudos nos quais os ensaios Illumina Protein Prep e SomaScan sejam usados em paralelo ou sequencialmente.

Os ensaios Illumina Protein Prep 9.5K e SomaScan 11K usam controles de calibradores idênticos para correção de placas. Como resultado, quando a referência de calibração do NGS entre plataformas for aplicada a dados de matriz no Illumina Connected Multiomics, ela combinará estes dados com a escala do NGS, tornando os dados desses dois ensaios comparáveis. A avaliação do desempenho da matriz é preservada nessa normalização por meio da etapa de calibração específica da plataforma, na qual os calibradores são comparados a uma referência da matriz e uma métrica de controle de qualidade (CQ) é calculada.

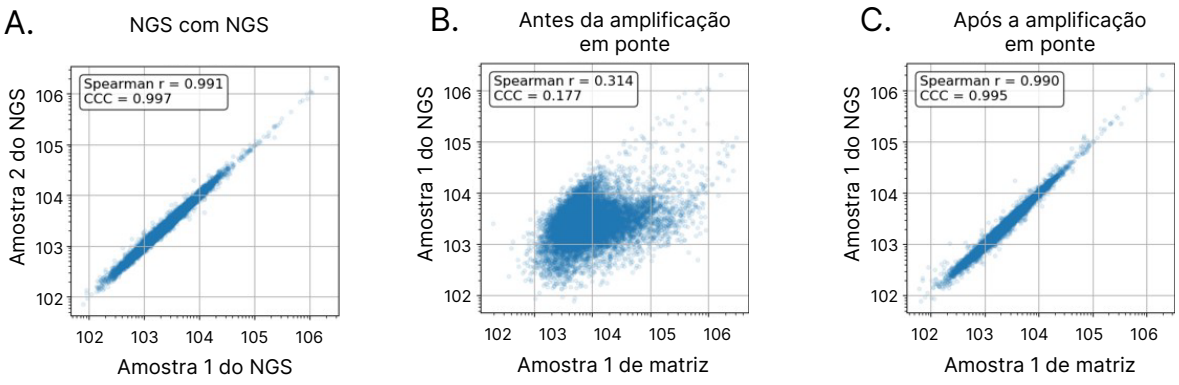


Figura 2: O coeficiente de correlação de Spearman entre dados com base em matriz e no NGS antes e depois da amplificação em ponte para uma única amostra demonstra harmonização de dados

Amostras de plasma em pool representativas de diferentes corridas de ensaios mostrando (A) excelente correlação entre dados de corridas do NGS individuais (B) baixa correlação entre o NGS e dados de matriz sem amplificação em pontes, principalmente devido à compressão do intervalo dinâmico específico do NGS e (C) correlação aprimorada entre dados do NGS e de matriz, aumentando para níveis de intraplataforma com amplificação em pontes.

A combinação fornece dados do NGS e de matriz altamente correlacionados

Para avaliar essa estratégia de amplificação em ponte, amostras de plasma em pool pareadas foram testadas no ensaio Illumina Protein Prep 9.5K Plasma e no ensaio SomaScan 11K. As amostras foram testadas uma segunda vez no Illumina Protein Prep para gerar uma linha de base para comparar os resultados da interplataforma (matriz com NGS) com resultados intraplatforma (NGS com NGS). Os conjuntos de dados do NGS foram normalizados usando a normalização do Illumina Protein Prep, enquanto o conjunto de dados de matriz foi normalizado usando a normalização de matriz antes da amplificação em ponte e a normalização do NGS após a amplificação (Figura 2).

A amplificação em ponte alinha as diferenças entre leituras com base em matriz e no NGS para produzir dados comparáveis (Figura 2). Um aspecto importante dessa correção é o dimensionamento relativo entre SOMAmer Reagents, que difere significativamente entre o NGS e a matriz devido a uma etapa de compressão de intervalo dinâmico (DRC, dynamic range compression) adicional na leitura do NGS que reduz a variabilidade do sequenciamento entre os SOMAmer Reagents.

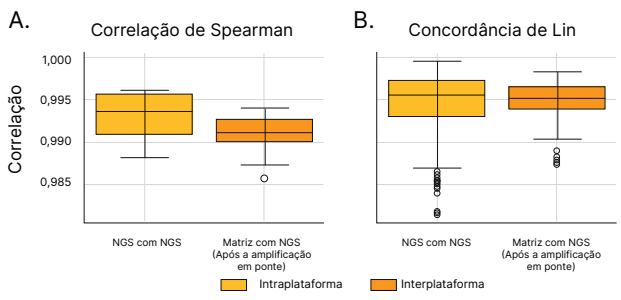


Figura 3: A distribuição da correlação de Spearman e dos coeficientes de correlação de concordância de Lin nas amostras mostra um efeito consistente da amplificação em ponte no alinhamento de dados

Distribuição de (A) correlação e (B) concordância em 18 comparações de plasma em pool entre plataformas com base em matriz e no NGS. As correlações e a concordância de leituras de matriz com NGS após a amplificação em ponte entre muitas amostras de plasma em pool são comparáveis às correlações de NGS com NGS.

A **DRC** é usada para uniformizar as concentrações do SOMAmer Reagent de um intervalo dinâmico de 5 log para 2 log. Sem DRC, os SOMAmer Reagents mais abundantes ocupariam a maioria da capacidade de leitura do sequenciamento. Cada amostra exigiria uma profundidade de sequenciamento extremamente alta para cobrir os SOMAmer Reagents não abundantes e quantificá-los com precisão. Com DRC, as sondas são agrupadas com base nas suas abundâncias observadas sem condição de DRC, e cada grupo é comprimido por uma razão de ajuste diferente.

Essa diferença é evidente mesmo dentro de uma única amostra de plasma em pool, em que a variância intraplataforma (Figura 2A) é acentuadamente menor do que a variância interplataforma (Figura 2B) antes da amplificação em ponte. No entanto, a amplificação em ponte supera essa diferença de leitura e alinha os dados da matriz com a contagem do NGS em uma escala comparável à variância intraplataforma (Figura 2C), quantificada pelo coeficiente de correlação de Spearman e pelo coeficiente de concordância de Lin. Essa similaridade quantitativa entre as variâncias intraplataforma e interplataforma após a amplificação em ponte é observada em toda a distribuição de comparações de amostras de plasma em pool (Figura 3).

A amplificação em ponte melhora a concordância da contagem entre o NGS e os dados da matriz em todo o painel SOMAmer

Como parte do mesmo estudo, uma coorte de 32 amostras de plasma pareadas foi analisada usando o ensaio Illumina Protein Prep 9.5K Plasma e o ensaio SomaScan 11K em duas configurações diferentes de manuseio de líquidos Tecan. As amostras foram testadas uma segunda vez no ensaio Illumina Protein Prep 9.5K para gerar uma linha de base para resultados intraplataforma (NGS com NGS). Os conjuntos de dados do NGS foram normalizados usando a normalização do Illumina Protein Prep, enquanto os conjuntos de dados com base em matriz foram normalizados usando a normalização de matriz (antes da amplificação em ponte) e a normalização do NGS (após a amplificação). As correlações intraplataforma com NGS usando as duas configurações e as correlações interplataforma entre NGS e saídas de matriz foram avaliadas (Figura 4, Figura 5).

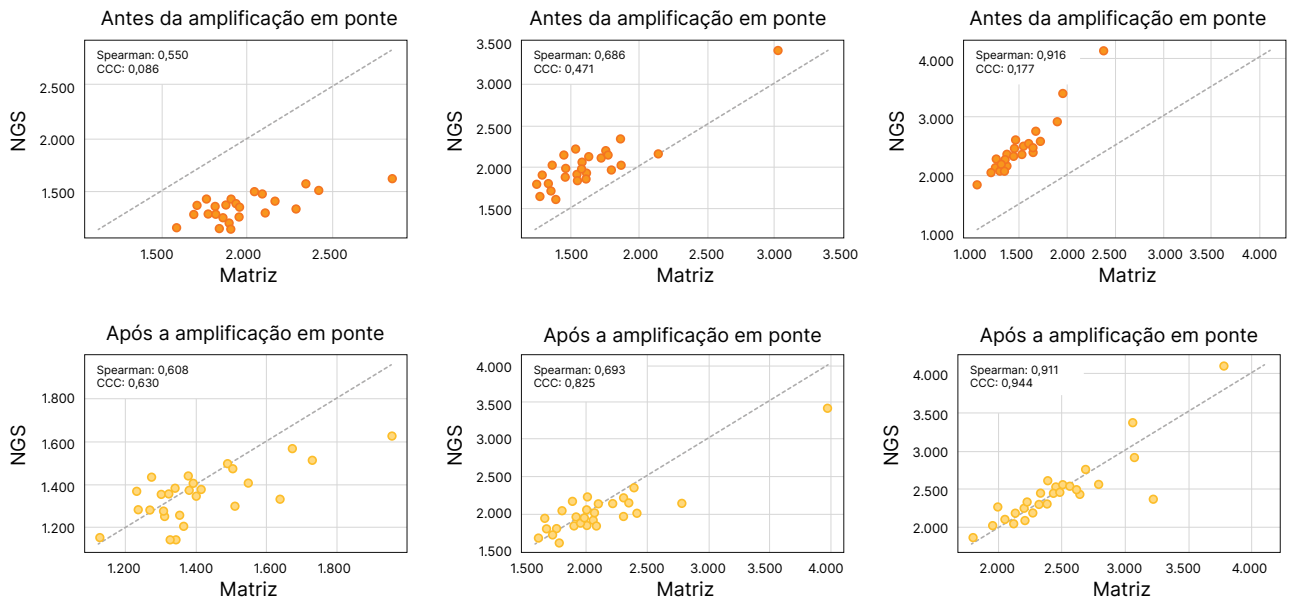


Figura 4: Exemplos de correlações de matriz com NGS para três reagentes SOMAmer antes e depois da amplificação em ponte

A amplificação em ponte mantém a correlação e aumenta a concordância entre o NGS e as contagens de matrizes para cada SOMAmer Reagent em uma coorte biológica.

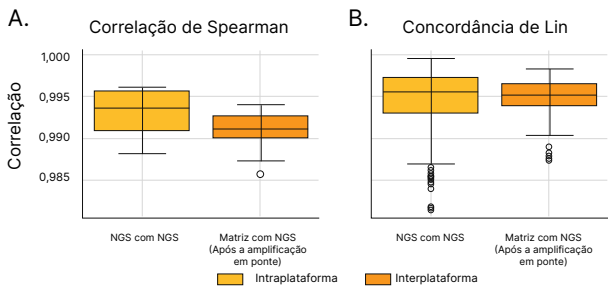


Figura 5: A distribuição da correlação de Spearman e dos coeficientes de correlação de concordância de Lin em todos os SOMAmer Reagents humanos demonstra concordância consistente com a amplificação em ponte

(A) A distribuição de correlações de NGS com NGS intraplataforma em todos os SOMAmer Reagents é comparável com correlações de NGS com matriz. (B) Com a amplificação em ponte, a distribuição da concordância interplataforma se aproxima da concordância intraplataforma.

Como ambas as leituras medem a concentração relativa do SOMAmer Reagent, as contagens individuais do SOMAmer Reagent em toda a coorte biológica geralmente se correlacionam. Ou seja, os dados se movem na mesma direção entre leituras de matriz e do NGS, mesmo sem amplificação em ponte e são quantificados usando o coeficiente de correlação de Spearman (Figura 4, Figura 5A). No entanto, a concordância com a contagem do SOMAmer Reagent, que é uma medida da proximidade das contagens reais, é significativamente menor nas comparações interplataforma (matriz com NGS) do que nas comparações intraplataforma (NGS com NGS). Isso se deve a diferenças entre as metodologias de leitura, incluindo unidades, intervalo dinâmico e referências de normalização. A amplificação em ponte alinha os dados da matriz com a escala de leitura do NGS, permitindo a comparação entre contagens por SOMAmer Reagent por amostra. Isso é quantificado pelo coeficiente de correlação de concordância de Lin, que aumenta para níveis intraplataforma com a amplificação em ponte (Figura 5B).

Resumo

Esta nota técnica demonstra que a amplificação em ponte do ensaio fornece dados altamente correlacionados e concordantes de plataformas com base em matriz e no NGS. Uma estratégia eficaz de amplificação em ponte permite que os laboratórios ampliem os dados gerados anteriormente do SomaScan Platform com as novas medições do Illumina Protein Prep, permitindo a descoberta de biomarcadores multiplataforma, o suporte ao planejamento da transição do ensaio e a garantia da continuidade das análises em pipelines proteômicos em evolução.

Saiba mais →

Illumina Protein Prep

Obter uma detecção precisa e confiável de proteínas com o Illumina Protein Prep

illumina®

+1 (800) 809-4566, ligação gratuita (EUA) | tel. +1 (858) 202-4566
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. Todos os direitos reservados. Todas as marcas comerciais pertencem à Illumina, Inc. ou aos respectivos proprietários. Para obter informações específicas sobre marcas comerciais, consulte www.illumina.com/company/legal.html.
M-GL-04147 PTB v1.0