

数据的归一化与标准化

简介

借助 Olink 邻位延伸分析 (PEA) 技术 (在 Olink® Target 96/48 蛋白质生物标志物板块中也有所使用)，在读取步骤中可使用实时荧光定量 PCR (qPCR) 来测量蛋白质表达的相对变化。qPCR 检测的是连接到成对分析物特异性抗体的互补寡核苷酸标记在 DNA 聚合酶存在下进行杂交和延伸时形成的独特 DNA 序列。

Olink 借助一系列计算，将 qPCR 的 Ct 值转化为相对定量单位，即蛋白质归一化表达 (NPX)。这些操作旨在最大限度地减少技术差异，并增强结果的可解释性。

Olink 还开发了一个带有 NGS 读取 (Olink® Explore) 的 PEA 平台。虽然这两种读取方法的 NPX 计算方式不同，但数据的评估是以相同的基本方式进行的。然而，在此白皮书中，我们将只关注通过 qPCR 读取的标准化数据。

数据的分析可能会受到某些技术因素的影响。为了说明这一点，Olink 使用了一个质量控制 (QC) 系统，以便监测测定和样本的性能，然后进行适当归一化，从而减轻由处理样本或技术变化引起的系统噪音。

Olink 的内置质量控制 (QC) 体系

Olink 已开发了一个内置质量控制体系，并采用内部对照，可用于其多重生物标志物板块。该系统能对测定和样本的技术性能进行全面控制。

内部对照

质量控制系统由四个内部对照组成，它们被添加到每个样本中，旨在监测 Olink 协议的三个主要步骤：免疫反应、延伸和扩增/检测 (图 1)。

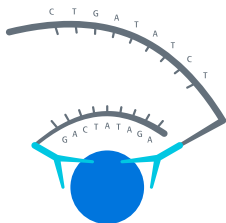
孵化对照：孵化对照 1 和 2 采用 PEA 测量的两种不同非人类抗原。这些对照可监测反应的全三个步骤中的潜在技术变化。

延伸对照：延伸对照含有与一对独特的 DNA 标记相耦合的单个抗体。这些 DNA 标记总是相当接近，因而此对照预计会给出一个独立于免疫反应的恒定信号。该对照可监测延伸和扩增/检测步骤中的变化，并可用于调整每个样本在延伸和扩增方面的信号。

检测对照：检测对照是一个完整的双链 DNA 扩增子，且不需要任何近似结合或延伸步骤来产生信号。该对照可监测扩增/检测步骤。

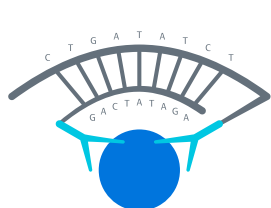
免疫反应

允许 92 个抗体探针与每份样本中的蛋白质相结合。



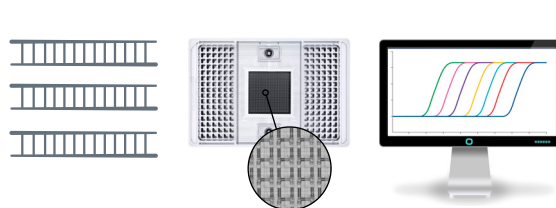
延伸和前置放大

延伸并前置放大 92 个独特的 DNA 报告序列，需借助近距离延伸。



扩增和检测

使用高通量实时 qPCR 对每个生物标志物的 DNA 报告进行量化。



免疫/孵化对照

延伸对照

检测对照

图 1. qPCR 读数及 PEA 测定中的主要步骤和控制。

样本对照

将六个必要的和两个推荐的外部对照添加到板块上不同的孔中。

● 板间对照

板间对照 (IPC) 在每个板上为一式三份, 且都是作为正常样本运行。IPC 是一个由 92 种抗体组成的池子, 每一种抗体都有一对独特的 DNA 标记, 其位置固定, 可看作是一个合成样本, 并预计在所有的测定中都会产生高信号。IPC 三倍体的中位值被用来将每项测定归一化, 以便补偿运行和板块之间的潜在变化。

● 阴性对照

阴性对照在每个板块上为一式三份, 其由作为正常样本的缓冲液组成。这些用来监测当 DNA 标记接近时产生的背景噪音, 而事先并没有与相应的蛋白质结合。阴性对照为每种蛋白质测定设定了背景水平, 并用于计算检测极限。

● 样本对照

当运行超过 90 个样本时, 建议运行一式两份的集合血浆样本。这些用来评估运行和板块之间的潜在变化, 例如计算批间和批内的 CV。

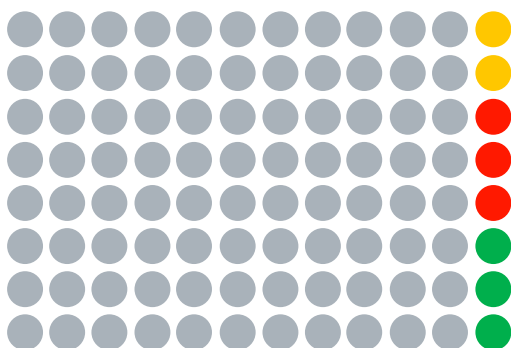


图 2. 建议 PEA 运行控制设置。

样本质量控制

每一个内部对照都以相同的浓度添加至样本中。由此, 预计这些信号在板块上是相同的。使用检测对照和孵化对照 2, 对最初计算的 NPX 值进行样本质量控制。在每个板块中, 对每个样本的对照水平进行监测, 并与所有样本的中位值进行比较。如果任何一个对照组的差异超过接受标准 ± 0.3 NPX, 则该样本被认为是偏差且不合格。内部对照的偏差值可能是由诸如移液错误或影响性能的样本分析前的变化等因素造成的, 例如基质效应。

术语解释

矩阵效应 的定义为分析读数的变化, 这种变化可能是由除待量化的特定分析物以外的所有其他样本成分引起的。

板块质量控制

内部对照也可用于板块质量控制。这评估了孵化对照 1 和 2 以及检测对照中每个板块的变化情况。如果其中一个对照的变化太大, 整个板块就被判定为不可靠。

使用板间对照进行归一化

NPX 的计算

Olink 使用一个任意的、相对的量化单位, 其称为归一化蛋白质表达量 (NPX)。在 qPCR 中, 反应曲线与阈值线相交点的 x 轴值被称为 Ct, 或“循环阈值”。这表明信号超过荧光信号阈值线所需的周期数。NPX 是由 qPCR 得到的 Ct 值用以下公式得出的:

延伸对照:

$$Ct_{\text{分析物}} - Ct_{\text{延伸对照}} = dCt_{\text{分析物}}$$

板间对照:

$$dCt_{\text{分析物}} - dCt_{\text{板间对照}} = ddCt_{\text{分析物}}$$

对照校正系数进行调整:

$$\text{校正系数} - ddCt_{\text{分析物}} = NPX_{\text{分析物}}$$

术语解释

校正系数 由 Olink 在验证板块的过程中进行计算。该值是预先确定的, 使用阴性控制, 并用于反转刻度, 从而使更高的值对应于更高的信号。其也被用来确保背景水平大约为零。

NPX 是一个相对定量单位, 其与蛋白质浓度呈对数关系。即使两个不同的蛋白质具有相同的 NPX 值, 它们的绝对浓度也可能不同。应在单次运行的样本之间分别对每项测定的 NPX 进行比较。在没有进行适当的板间归一化的情况下, 不应比较 NPX, 因为有可能将运行中的中位值变化错误地解释为生物学差异。然而, NPX 的相对差异可以更容易地进行比较, 且通常不需要额外的标准化步骤。

曲线范例

原始数据

下方示例为一个带有校准器曲线的三次运行的原始数据。中间有偏离的样本，且一条曲线出现平行移动。

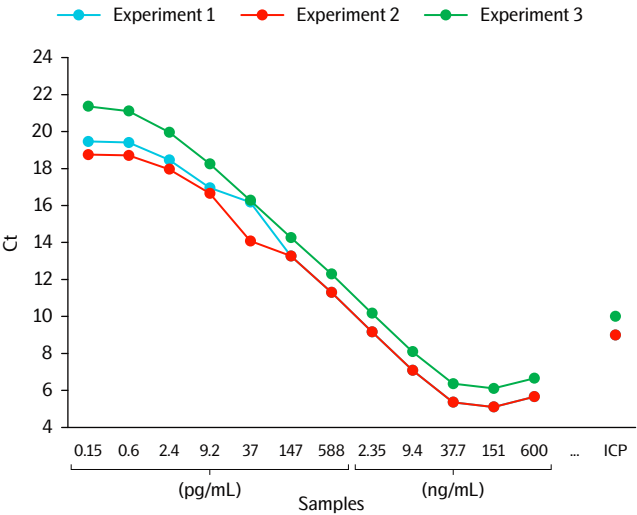


图 3. 三次运行的原始数据。

根据延伸对照进行调整

进而根据每个样本的延伸对照来调整原始数据，并通过减少延伸步骤中引入的技术变异来提高批内的可重复性。

$$Ct_{\text{分析物}} - Ct_{\text{延伸对照}} = dCt_{\text{分析物}}$$

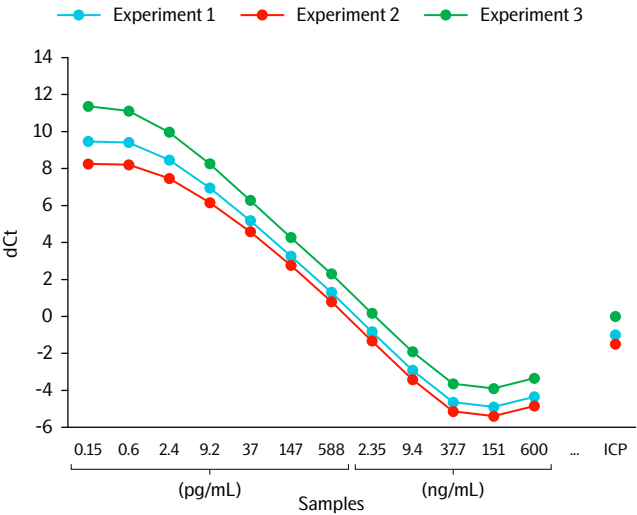


图 4. 根据延伸对照进行调整。

按照 IPC 进行归一化

为提高批间的可重复性，便根据 IPCs 对数据进行标准化处理。这是按照每个测定项目和板块进行的。根据研究的特点，可能还要进行强度或参考样本的归一化。

$$dCt_{\text{分析物}} - dCt_{\text{板间对照}} = ddCt$$

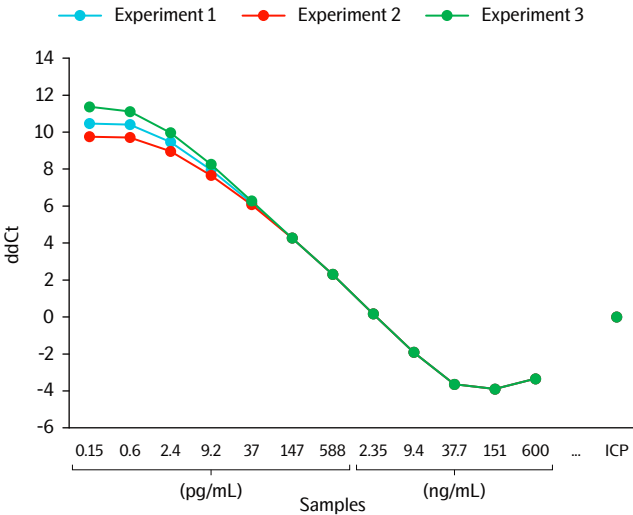


图 5. 按照 IPC 进行调整。

对照校正系数进行调整

然后将曲线倒置，从而高 NPX 值意味着高蛋白质浓度，解释也更为直观。1 NPX 的差异近似于蛋白质浓度的两倍，而无需考虑蛋白质。

$$\text{修正系数} - ddCt = NPX$$

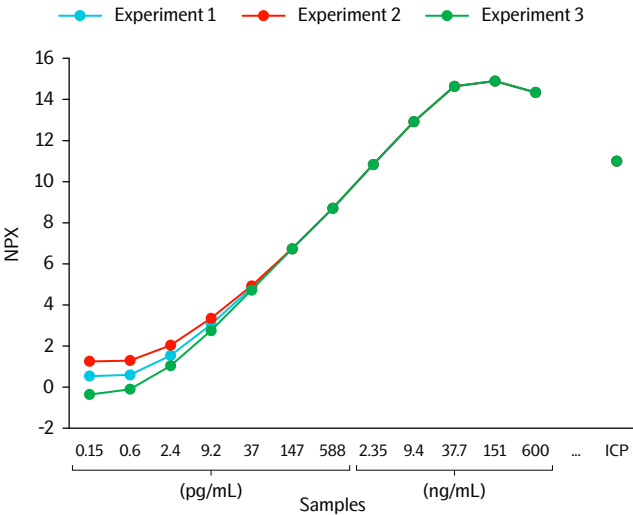


图 6. 对照校正系数进行调整。

板间归一化方法

Olink 推荐两种归一化方法中的一种，这取决于研究设计。对于随机研究而言，IPC 归一化应该被强度归一化所取代，这将提高统计能力并减少板块和项目之间的技术差异。只要样本分布被认为在数据集之间具有可比性，那么结合不同来源的研究时也可以使用强度归一化。

对于非随机研究，Olink 建议参考样本归一化。在所有板块上运行参考样本是一个很好的策略，这可以最大限度地减少技术差异。如果应用得当，强度归一化和参考样本归一化都可以借由减少技术变异来提高特定研究的能力，因为它们是基于真实的样本，而不是IPC样本。

术语解释

在这种情况下，**随机化** 适用于放置在板块上的样本。如果相关的实验变量可被认为是均匀地分布在各个板块上，那么样本集就是随机的。必须为每项研究决定好要考虑的变量。例如，它们可包括研究组、治疗、时间点或人口统计学。如果随机化在实验变量方面并不合适，那么正常化可能会消除真正的生物变异，反之则会被识别出来。

在研究中，如果样本在各板块之间为随机排列，则进行全局调整，使每个检测项目的数值围绕其中位值和所有板块的中位值集中。这就是强度归一化。如果随机化不可行或难以保证，那么 Olink 建议应涵盖对队列有代表性的参考样本。例如，汇集的血浆样本可以包括在所有板块上，以确保最大限度地控制任何系统性偏差。这种方法还可用于在合并研究时对技术变异的处理方式改进。

术语解释

偏差 可定义为由于实验、生物或技术方面的信号，导致系统产生误差。测量的蛋白质表达水平可能受到几种偏差来源的影响，包括分析前和技术上的差异，也可包括错误的采样。如果实施得当，归一化可以减少引入的偏差。Olink 白皮书《蛋白质生物标志物研究中的分析前差异 (Pre-analytical variation in protein biomarker research) 》提供了关于样本处理和加工的建议 (www.olink.com/downloads)。

强度归一化方法

强度归一化背后的假设是：平均而言，一个板块上检测的中位信号与另一个板块相比，并无预期差异。如果在不同的板块之间看到这种差异，它可以被解释为技术偏差。其可安全删除，并能产生更多的可比值。

当样本在各板块和项目随机出现时，Olink使用强度归一化，并以中位值作为归一化因子。强度归一化调整了数据，使每个板块上检测的中位值等于所有板块上的总体中位值。这种方法假定每个板块的实际中位值是相同的。确保这一点的办法是事先将样本进行随机化处理。如果可以假设完全随机化，那么这便是一种稳健的高性能规范化方法。强度归一化是以下列方式进行的：

步骤	描述
1	对于每项测定，计算所有样本和板块的总体中位值。
2	对于每个板块和每项测定，计算板块的特定中位值。
3	对于每项测定，从板块的每个数值中减去板块的特定中位值（等于集中到总体中位值 0）。
4	对于每项测定，加上总体中位值（等于集中到总体中位值）。

参考样本归一化方法

当样本不在各板块间随机排列时，每个板块上包含八个或更多的桥接参考样本，其可用于归一化。参考样本归一化是按照下列方式进行的：

步骤	描述
1	选择一个参考板块来进行归一化。
2	对于每项测定和每个板块，计算每个重叠样本与参考板块的成对差异。
3	通过计算步骤二中计算成对差异的中位值来估计板块和特定测定的归一化因子。
4	对于每项测定和板块，将步骤三中的板块和特定测定的归一化系数添加到每个值中，以使其与步骤一中选择的参考板块归一化。

评估

为了评估某个归一化方法的有效性，Olink 调查了 NPX 分布，并比较了不同归一化方法的平均 %CV。

强度归一化评估

下图描述了一种蛋白质在三个不同板块上的 NPX，其中的样本随机。不同的颜色表示不同的板块。

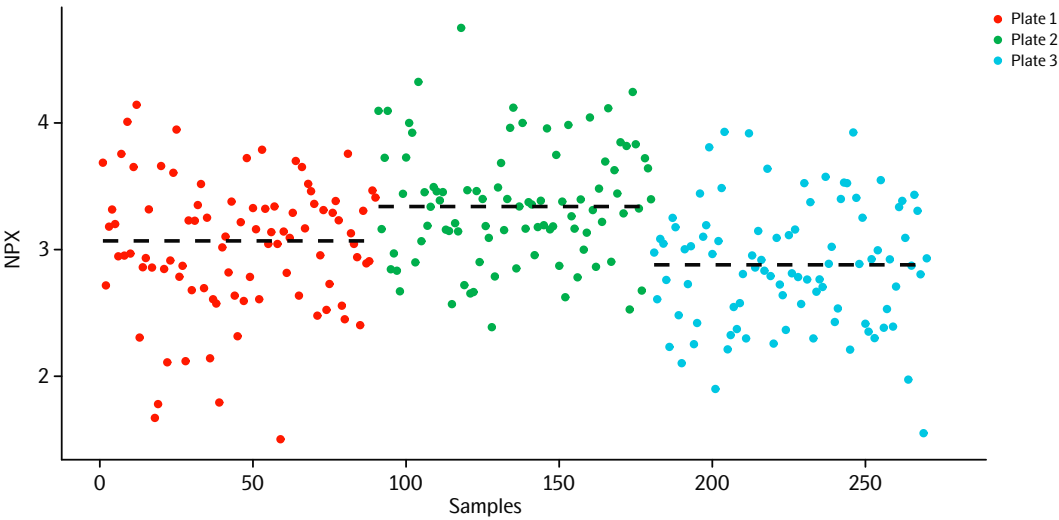


图 7. 每种蛋白质的 NPX 覆盖各板块。

在强度归一化之后，中位值 NPX 被拉平

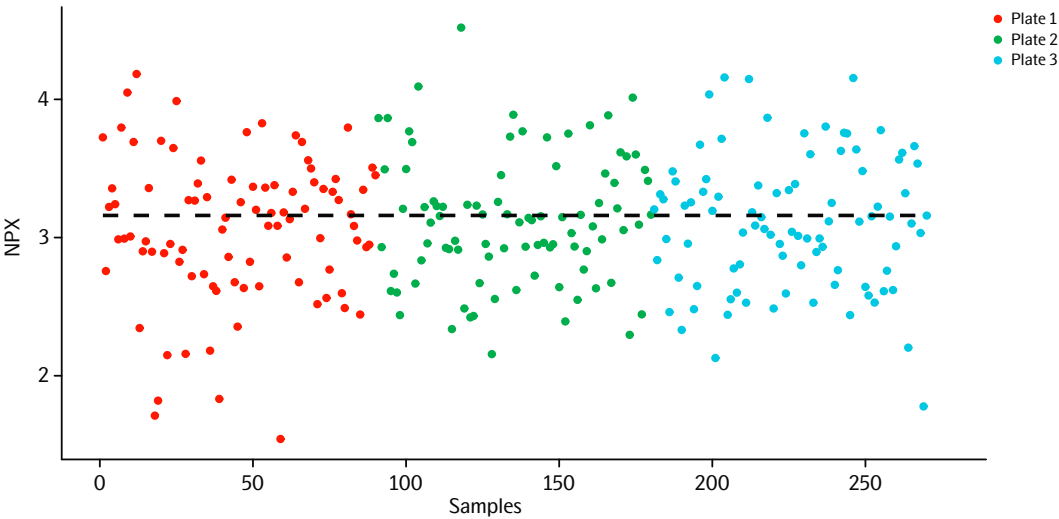


图 8. 强度归一化后，每种蛋白质的 NPX 覆盖各板块。

在适当的随机研究中，强度归一化通常能够将平均 %CV 提高几个百分点。

参考样本归一化评估

在下例中，样本的随机性很差，中间的板块主要包含一个组的样本。这意味着板块之间的技术差异与可通过比较样本组来研究的生物差异相混合。在这种情况下，在校正板块之间的技术差异时，板块的强度归一化将消除一些相关的生物变化。图中显示了随机性差的样本中蛋白质 NPX 值和参考样本的 NPX 值。参考样本显示为紫色。

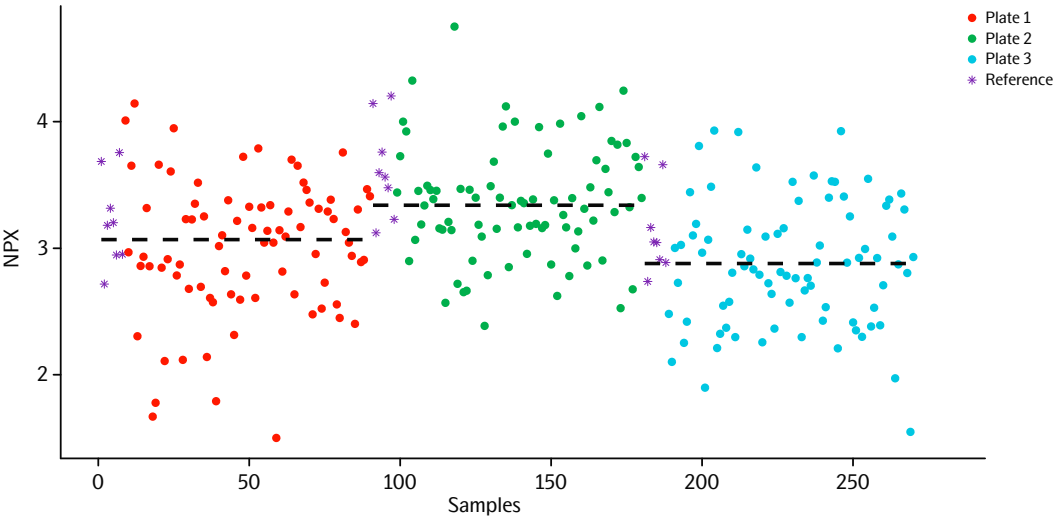


图 9. 每种蛋白质的 NPX 覆盖各板块，包括参考样本。

然后用参考样本进行归一化处理。在参考样本归一化后，各板块之间的 NPX 中位值仍可能不同，此反映了样本组的不同组成部分。

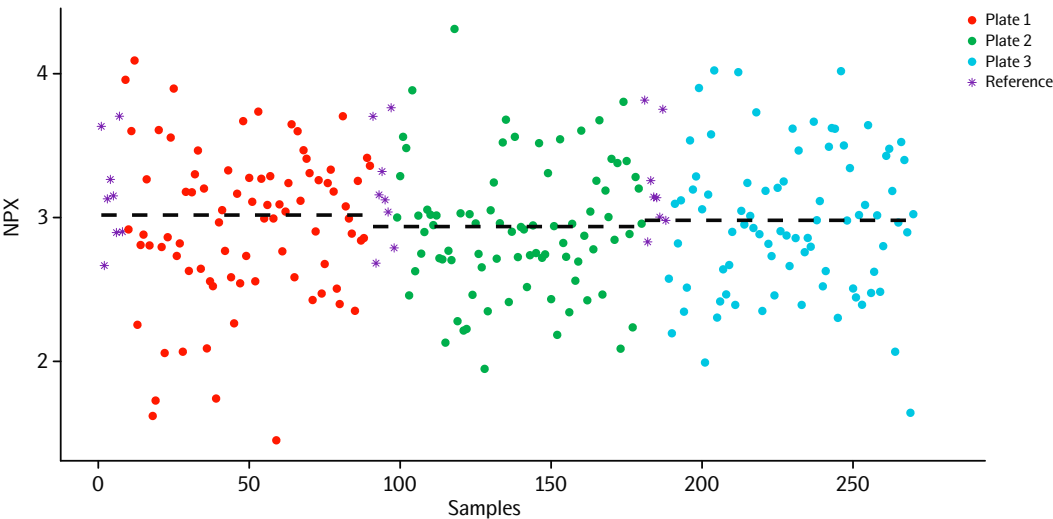


图 10. 参考样本归一化后，每种蛋白质的 NPX 覆盖各板块。

误读的数据

为了进一步说明归一化不当的问题，下图给出了一个例子，说明当归一化因素由生物因素引入的方差驱使，而并非简单地由不适当的样本制备或基于仪器的变异性驱使时，数据集中引入的偏差。此插图显示了一组随机性较差的样本，两组之间的差异具有统计学意义 ($P<0.05$)。在这种情况下，强度归一化将消除这种差异。

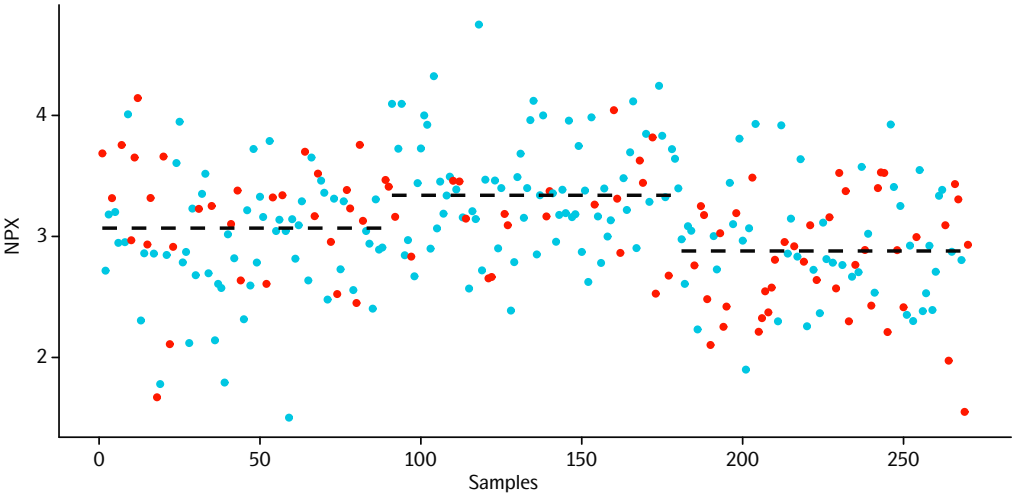


图 11. 每种蛋白质的 NPX 覆盖各板块，包括随机性较差的样本。

第二个板块包含的样本主要来自同一个组。如下图所示，在这里，技术变量（板块）与生物变量（样本组）混合在一起，而强度归一化无法将两者分开。

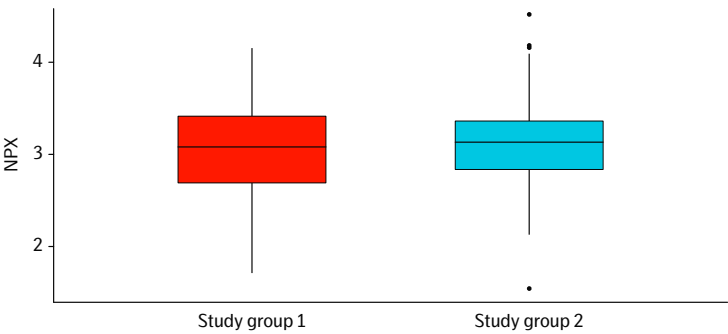


图 12. 强度归一化消除了此差异 ($p = 0.52$)。

将参照物归一化，便有可能看到明显的差异。

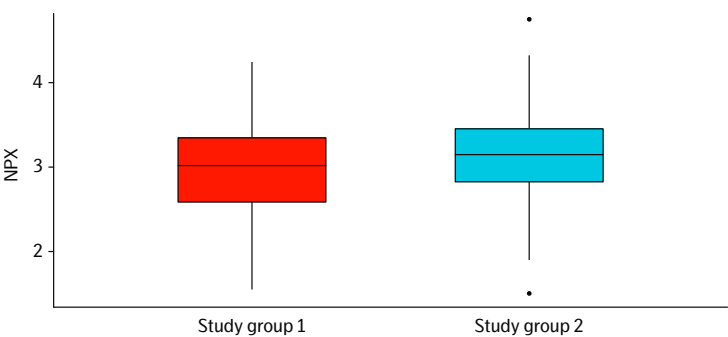


图 13. 与参考样本归一化有明显差异 ($P<0.05$)。

Olink 可提供帮助

归一化对消除系统性变异很重要，但需要谨慎应用，以便尽量减少不慎消除真正生物变异的风险。作为一项服务，Olink 可以在分析开始前与您讨论规范化方法和研究设计。

如果您将样本送到 Olink 分析服务团队，他们便会处理数据分析的步骤。在您的研究完成后，他们会向您提供一份全面的 NPX 数据和质量控制报告。

如果您需要更广泛的数据分析服务，Olink 数据科学团队的生物统计学家可以帮助您定制统计分析服务。

Olink 数据科学服务目前包括

- **Olink®洞察统计分析 (Insights Stat Analysis)**

此为免费的基于网络的应用程序，可用于基本数据的可视化和统计分析 (Shiny App)。

www.olink.com/biostat-apps

- **OlinkAnalyze (GitHub Rpackage)**

此为处理 Olink 数据的多功能工具箱，包括 QC 绘图功能、归一化、各种统计测试和建模。

www.olink.com/biostat-apps

- **Olink 统计服务**

交由经验丰富的处理 Olink 数据的专家执行。

www.olink.com/biostat-services

欧邻科生物科技（上海）有限公司

中国技术中心

地址：上海市浦东新区沪南路2157弄2号1011室

电话：021-50778771

邮箱：china@olink.com

网址：www.olink.com

版本号：1096,CN1.1, 2022-11-01



关注官方微信公众
号可获取更多详细信
息