

# 靶向甲基化测序

## 简介

人类基因组测序开展的数十年以来，研究人员建立了遗传因素与许多疾病之间的联系。但是，编码基因组 DNA 的核苷酸只是遗传因素中影响细胞功能和整体健康的一部分，表观遗传学（指在某些情况下不修改 DNA 序列而发生的可遗传的改变）也具有重要功能。DNA 甲基化是一种已经过充分研究的表观遗传学标记，其可化学修饰胞嘧啶和腺嘌呤。胞嘧啶甲基化最常发生在基因组中的 CG 序列，称为 CpG 位点，它以细胞特异性的方式广泛地调控基因表达。<sup>1,2</sup> 全基因组关联研究 (GWAS) 将 DNA 甲基化的改变与复杂疾病例如癌症和肥胖以及复杂的生物学状态例如衰老和发育联系起来。<sup>1,3-5</sup> 近年来，甲基化测序已扩展到基因表达之外，开始在疾病诊断中获得关注。例如，研究表明差异甲基化是癌症检测中信息丰富且灵敏的标记，不论甲基化是否与基因表达相关。<sup>6</sup>

## 甲基化测序和捕获

甲基化测序涉及酶或化学方法，这些方法可使未甲基化的胞嘧啶通过一系列反应，最终通过脱氨基转化为尿嘧啶，而甲基化的胞嘧啶则保持完整（图 1）。在扩增过程中，尿嘧啶对应的互补链上会互补配对加上腺嘌呤，造成未甲基化的胞嘧啶的原始位置引入胸腺嘧啶。如图所示序列终产物是不对称的，转化后形成两条不同的双链 DNA 分子（图 1-上）；对于甲基化的 DNA，同样的过程则产生另外不同序列（图 1-下）。

转化可以在靶向富集捕获之前或捕获之后进行。捕获后转化靶向左边的原样本 DNA，而捕获前转化靶向右边四条转化后的序列（图 1）。虽然捕获后转化在探针设计中困难较少，但通常需要大量的起始 DNA 材料，因为 PCR 扩增不能保留甲基化模式，因此也不能在捕获前进行。所以，捕获前转化通常被用于低起始量、灵敏的应用，例如游离 DNA。

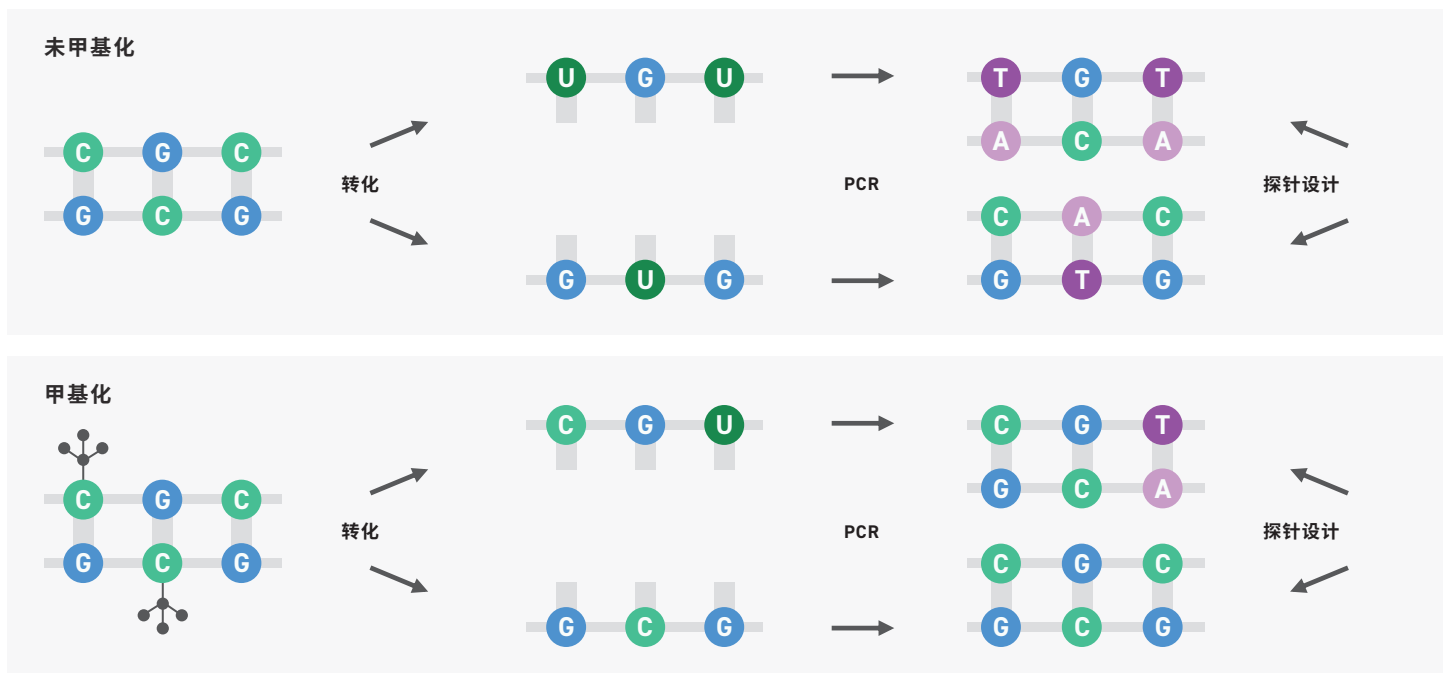


图 1. 甲基化转化。

## 组合设计

捕获前转化带来了更多复杂性，因为要设计靶向四种可能的转化后序列的探针（图 1）。实际上，目标基因组区域虽然增长数倍，但由于未甲基化胞嘧啶转化为胸腺嘧啶，使得总体序列复杂度降低。所以，该领域的靶向探针组合通常存在高脱靶率和低覆盖均一性的问题。

在此，我们尝试运用 Twist 的高质量寡核苷酸合成、捕获解决方案和数据驱动的方法解决靶向甲基化测序的困难，并获得了非常高均一性的靶向探针组合。采用 Twist 新开发的甲基化特异性基础设计算法和靶向甲基化测序方案，我们证明了不同大小组合的高性能指标。利用得到的数据进行机器学习，我们得到了更优的默认设计能力，可实现不同严谨度的设计。虽然芯片仍广泛用于甲基化检测，但对于一些甲基化靶标的检测，包括探寻动态和细胞特异性甲基化靶标，以及探索发生在不甚了解的非编码区、意义不明的单碱基对甲基化靶标，高质量靶向富集组合相对于固定的芯片设计提供了一种具有吸引力的替代检测方式。

## 材料与方法

将来自 NA12878 (Coriell Institute) 和 EpiScope® 低、高甲基化 gDNA 对照品（分别为 < 5% 和 > 95% 甲基化的 HCT116 DKO gDNA）的基因组 DNA 样本机械剪切为 ~300 bp 的大小（使用 Covaris® ME220）。通过混合已剪切的低、高甲基化对照品，制备各种模拟的不同甲基化水平的样本。文库制备分别采用化学转化和酶转化，其中，Swift Accel-NGS® Methyl-seq DNA 文库试剂盒的 DNA 起始量为 500 ng，结合重亚硫酸氢盐处理（Zymo EZ DNA Methylation Lightning 试剂盒、Omega Bio Tek Mag Bind RxnPure Plus SPRI 微珠和 KAPA HiFi Uracil + DNA 聚合酶）。NEBNext® Enzymatic Methyl-seq 文库试剂盒的 DNA 起始量为 200 ng。用 Agilent BioAnalyzer 7500 和 Invitrogen Qubit Broad Range 试剂盒验证已剪切的样本和文库。

Twist 的 4 小时快速杂交方案重新优化后用于 Twist 甲基化组合。本实验合成并使用 4 个覆盖不同靶标大小（0.05、1.0、1.5 和 3.0 Mb）的甲基化组合评估其性能。每个单一捕获使用 200 ng 文库，然后在 Illumina® NextSeq® 550 上使用 v2.5 高通量试剂盒进行 2x151 bp 测序。每个样本数据截取至原始覆盖度为 250x 后，使用 Bismark 19.1 和 Picard HsMetrics 进行比对和甲基化分析。

## 结果

### 各种不同大小组合的测试

尽管捕获前转化可用于高灵敏度的表观遗传学应用，但转化后基因组复杂度的降低带来了非常大的挑战。与非甲基化组合相比，这通常会显著的高脱靶率 (> 50%-60%)、较低的靶标覆盖度以及捕获均一性的显著降低 (fold 80 碱基罚分 > 2.5)。本测试使用 4 个不同大小，覆盖不同甲基化靶标的组合测试了 Twist 的靶向甲基化测序方案。利用 Twist 的基础组合设计能力，在以下实验中使用的组合显示了低于当前业界标准的脱靶水平，脱靶值低至 27%。0.05 Mb 组合的脱靶率显示高于其他三个组合，这可能是由于其极小的靶标区域所致。捕获均一性也明显优于业界标准的 > 2.5 fold 80，其值低至 1.75 和 1.5，显示出极佳的性能。重复率在 4 个测试的组合中都非常低，表明有效的捕获了靶向分子，能在整个工作流程中维持高样本复杂度。总之，在 250x 原始测序覆盖度下，我们实现了 30x 高于 70% 的覆盖度，即使是最小的组合也是如此。

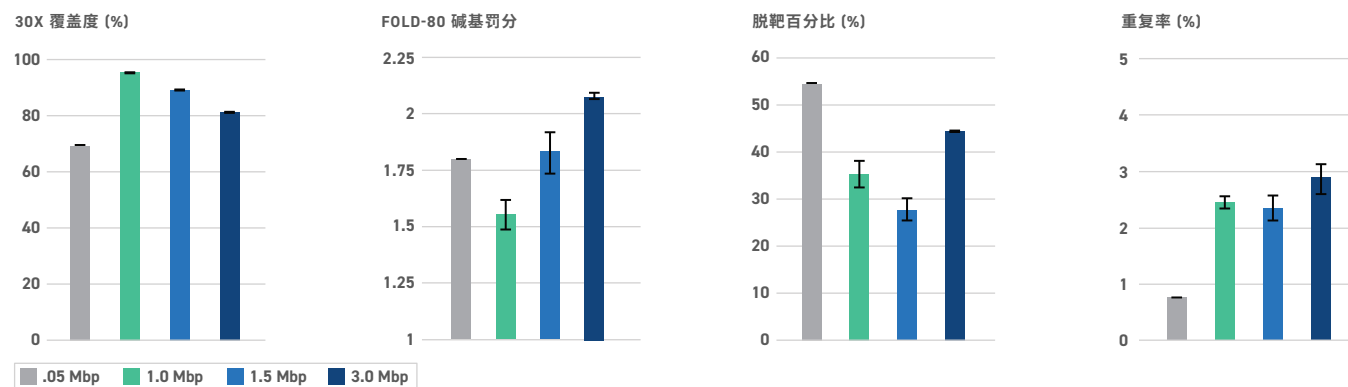


图 2. 组合大小。



### 高质量初始设计的严谨度控制

Twist 的适应性组合设计优化算法允许使用来自捕获实验的经验数据了解具体探针特征，以定量地调整性能。这一强大的方法对于以控制脱靶率为优先的甲基化组合尤其有用。因此，利用收集的超过 ~30000 个甲基化靶标的的数据，我们获得了信息丰富的序列特征，并使用这些特征开发了优化的默认组合设计算法，可实现 3 个不同严谨性下的设计。1 Mb 组合展示了低、中和高严谨度水平的默认组合设计可提供更好的脱靶率控制，同时对其它关键指标只有轻微改变（图 3）。

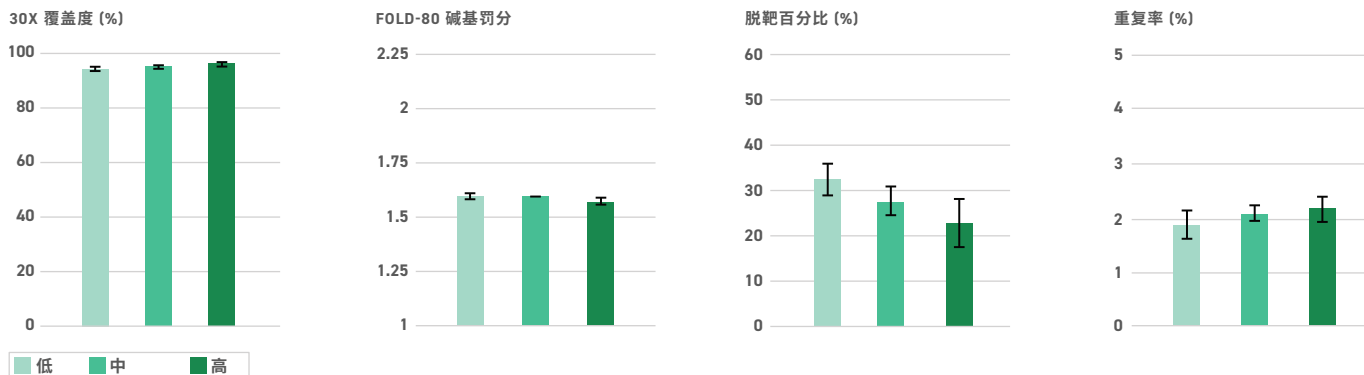


图 3. 设计严谨度。

### 在可能的靶标低、高甲基化范围内维持捕获性能

为评估在可能的甲基化水平范围内的兼容性，我们混合了低甲基化和高甲基化 gDNA 对照品以获得最终比例分别为 0%、25%、50%、75% 和 100% 甲基化的 gDNA 文库，接着用中等严谨度的 1 Mb 组合进行捕获。下图突出展示了关键捕获指标，柱状图显示了不同甲基化样本之间的捕获性能。结果显示，各指标在不同甲基化水平下均保持稳定，证明系统对各种甲基化状态，包括低、高甲基化 DNA 具有兼容性。

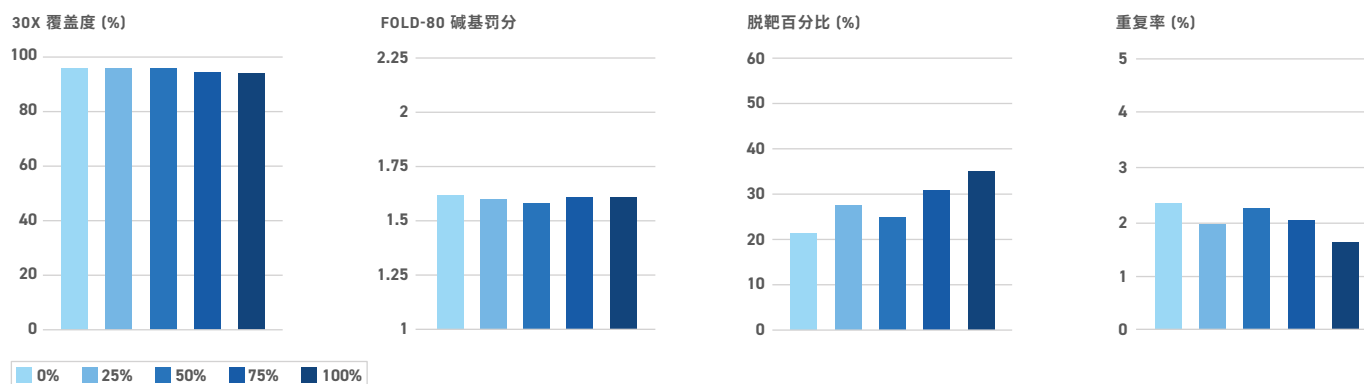


图 4. 不同甲基化程度 (%) 文库。

### 甲基化水平的检测

启动子和其他调控元件的甲基化水平改变正成为癌症早期检测最灵敏的标志物之一。<sup>7</sup> 我们证明了靶向甲基化测序能检测和定量不同水平的 DNA 甲基化。本检测混合低、高甲基化 DNA 至不同比例，并用 1 Mb 组合捕获。图 5 突出显示了临床相关的细胞周期素 D2 位点中的靶标和各个 CpG 位点不同 DNA 甲基化水平的检测结果，已知细胞周期素 D2 位点在某些癌症中可改变甲基化状态。<sup>8</sup>



图 5. 甲基化检测的灵敏度。

### 对于酶转化和重亚硫酸盐转化的兼容性

检测甲基化的胞嘧啶涉及将未甲基化的胞嘧啶转化为胸腺嘧啶，而甲基化的胞嘧啶不被转化。传统上，转化通过一种基于重亚硫酸氢盐的化学方法进行。但在这个领域内，采用其他方法包括酶法转化未甲基化的胞嘧啶（如本实验中所使用的）的比例正在增加。每种转化方法均有其优点和缺点，例如酶对转化反应条件的潜在敏感度更高，重亚硫酸氢盐对 DNA 有偏好的造成损伤。

Twist 的靶向甲基化测序与基于酶和重亚硫酸氢盐的方法均兼容（图 6）。以非 CpG 位点胞嘧啶的转化比例为依据的转化率都 > 99.5%（数据未展示）。对于两种文库制备方法，总体捕获指标类似，处于同一数量级，但对于重亚硫酸氢盐方法，某些指标例如均一性和中靶率有所偏低。有证据表明，均一性的降低至少部分是由于基于重亚硫酸氢盐的文库制备方法引入的固有 GC 偏好（数据未展示）。

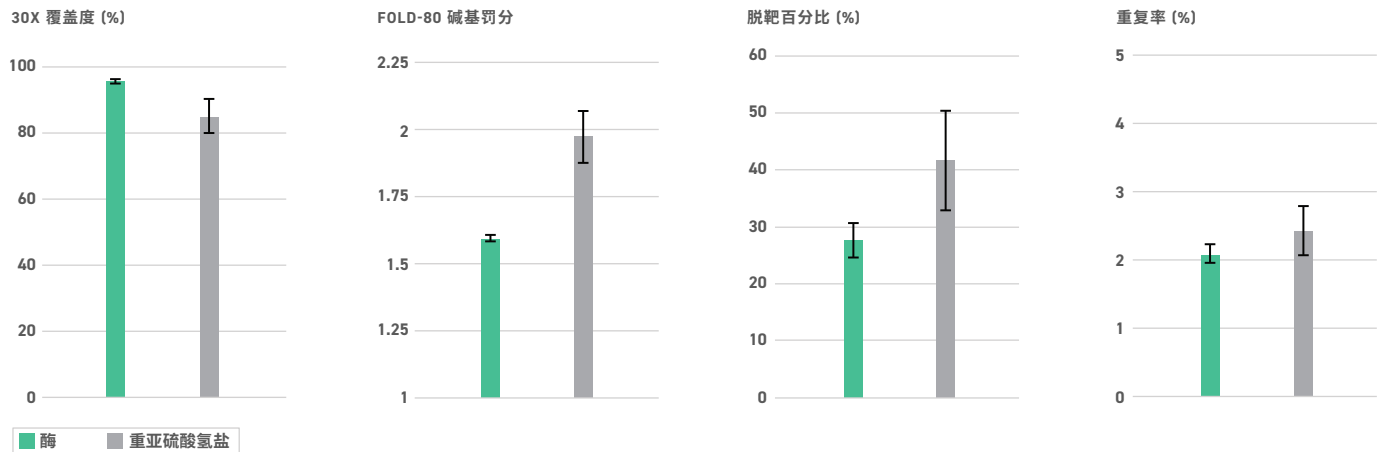


图 6. 文库制备中的甲基化转化方法。

### 结论

测试结果证明了 Twist 的技术在甲基化测序靶向富集组合中的独特能力。我们展示了产品在不同甲基化靶标和水平下的卓越性能，高度优化的设计软件以及其设计严谨度的灵活性。Twist 的高均一性甲基化捕获组合与酶和重亚硫酸氢盐转化兼容，可作为新工具推动高灵敏度表观遗传学应用的开发。

### 参考文献

1. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484–492 (2012).
2. Yang, X., Shao, X., Gao, L. & Zhang, S. Comparative DNA methylation analysis to decipher common and cell type-specific patterns among multiple cell types. *Brief Funct Genomics* 15, 399–407 (2016).
3. Loh, M., Zhou, L., Ng, H. K. & Chambers, J. C. Epigenetic disturbances in obesity and diabetes: Epidemiological and functional insights. *Mol Metab* 27S, S33–S41 (2019).
4. Ferrucci, L. et al. Measuring biological aging in humans: A quest. *Aging Cell* 19, e13080 (2020).
5. Pérez, R. F. et al. Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. *J Transl Med* 17, 15 (2019).
6. Locke, W. J. et al. DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Front Genet* 10, (2019).
7. Hao, X. et al. DNA methylation markers for diagnosis and prognosis of common cancers. *PNAS* 114, 7414–7419 (2017).
8. Evron, E. et al. Loss of cyclin D2 expression in the majority of breast cancers is associated with promoter hypermethylation. *Cancer Res.* 61, 2782–2787 (2001).