Advances in Applied Mathematics 应用数学进展, 2026, 15(5), 96-102
Published Online May 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.155211

基于甲基化BS-seq数据的肿瘤异质性分解

杨　琴，张伟伟*

绍兴大学数理信息学院，浙江　绍兴

收稿日期：2026年4月13日；录用日期：2026年5月7日；发布日期：2026年5月14日

摘　要

肿瘤组织的细胞异质性是干扰表观基因组下游分析的关键因素。现有基于DNA甲基化重亚硫酸盐测序(BS-seq)数据的反卷积算法多局限于“正常－肿瘤”二元成分假设，难以准确解析复杂的肿瘤微环境。为此，本文提出一种基于极大似然估计与期望最大化(EM)算法的统计推断模型MethEML。该模型仅需肿瘤混合组织的甲基化谱数据，即可在无外部参考样本的条件下，联合推断多种细胞亚群的混合比例及各亚群特异性甲基化谱，并引入贝叶斯信息准则(BIC)实现亚群数量 K 的自适应确定。基于真实细胞系(HCC1954与HMEC)构建的模拟数据集实验表明，MethEML突破了现有主流算法MethylPurify在细胞类型数量($K=2$)上的应用限制。在不同测序覆盖度与混合比例的测试场景下，MethEML的预测精度显著优于对比算法，且展现出更低的均方误差与更强的鲁棒性，为精准解析肿瘤微环境异质性提供了高效的计算工具。

关键词

肿瘤异质性，DNA甲基化，重亚硫酸盐测序(BS-seq)，反卷积，期望最大化算法(EM)

Tumor Heterogeneity Decomposition Based on Methylation BS-seq Data

Qin Yang, Weiwei Zhang*

School of Mathematics Information, Shaoxing University, Shaoxing Zhejiang

Received: April 13, 2026; accepted: May 7, 2026; published: May 14, 2026

Abstract

Tumor cellular heterogeneity is a critical factor that confounds the downstream analysis of the epigenome. Current deconvolution algorithms based on DNA bisulfite sequencing (BS-seq) data frequently

*通讯作者。

文章引用: 杨琴, 张伟伟. 基于甲基化 BS-seq 数据的肿瘤异质性分解[J]. 应用数学进展, 2026, 15(5): 96-102.
DOI: 10.12677/aam.2026.155211

rely on a simplified normal-tumor binary composition hypothesis, which fails to accurately resolve the complexities of the tumor microenvironment. To address this challenge, we developed MethEML, a statistical inference model based on maximum likelihood estimation and the expectation-maximization (EM) algorithm. Without requiring external reference samples, MethEML can simultaneously infer the mixing proportions of multiple cell subpopulations and their corresponding cell-type-specific methylation profiles directly from the methylation data of bulk tumor samples. Furthermore, the model incorporates the Bayesian Information Criterion (BIC) to adaptively determine the optimal number of subpopulations (K). Experiments conducted on simulated datasets derived from real cell lines (HCC1954 and HMEC) demonstrate that MethEML circumvents the inherent limitation of the mainstream algorithm MethylPurify, which is restricted to a binary cell-type composition ($K = 2$). Under various scenarios of sequencing coverage and mixing proportions, MethEML significantly outperforms the baseline algorithm, exhibiting lower mean square error and superior robustness. This study provides an efficient computational tool for the precise characterization of tumor microenvironment heterogeneity.

Keywords

Tumor Heterogeneity, DNA Methylation, Bisulfite Sequencing (BS-seq), Deconvolution, Expectation-Maximization Algorithm (EM)

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

临床获取的实体瘤组织通常表现出高度的细胞异质性。除恶性增殖的肿瘤细胞外，其肿瘤微环境(Tumor Microenvironment, TME)还广泛浸润了免疫细胞、基质细胞及血管内皮细胞等多种非肿瘤成分[1] [2]。这种多细胞混合模式不仅增加了样本的组内方差，更易掩盖真实的疾病生物学信号。在表观基因组数据分析中，若未对混合样本进行有效的异质性分解，直接开展差异甲基化或样本聚类等下游分析，极易导致推断出现严重偏差[3]。特别是在表观基因组关联研究(EWAS)中，细胞组分比例的变异往往与目标协变量(如年龄、病理分期)发生混杂(Confounding)，从而引发严重的统计偏倚及假阳性发现[4]。因此，精准解构肿瘤组织中的细胞亚群及其混合比例，对于消除微环境噪音、揭示真实的表观遗传异质性具有至关重要的意义[5]。

DNA 甲基化作为调控基因表达的核心表观遗传修饰，在维持染色质结构稳定性及细胞谱系发育等生理过程中发挥着关键作用[6]。研究表明，异常的 DNA 甲基化模式是几乎所有实体瘤的共有标志，典型特征为全基因组尺度的广泛低甲基化与特定抑癌基因启动子区域的局部高甲基化[7] [8]。当前，全基因组重亚硫酸盐测序(Bisulfite sequencing, BS-seq)凭借单碱基分辨率的优势，已成为刻画全基因组 DNA 甲基化图谱的“金标准”[9]。尤为重要的是，BS-seq 测序生成的每一条独立读段(Read)均严格溯源于单一细胞，这为解析单细胞层面的异质性提供了原生的分子条形码[10]。尽管该类数据蕴含极其丰富的异质性信息，但由于高通量数据的高维稀疏特性以及难以获取纯净的参考甲基化谱，针对 BS-seq 的计算反卷积(Deconvolution)工具仍面临巨大挑战[11]。早期经典算法如 MethylPurify 虽实现了基于纯 BS-seq 数据的反卷积，但其底层假设被严格限制在“正常–肿瘤”的二元组分空间内[12]。然而，真实的肿瘤组织是由多种动态演化的亚群构成，不同亚群间的甲基化特异性尚不完全明确，且组分比例在不同样本中具有高度的时空异质性。因此，亟需在统计推断框架上引入无参考(Reference-free)的新工具，以突破二元成分假设

的局限。

针对上述挑战，本文构建了一种基于极大似然估计(Maximum Likelihood Estimation, MLE)框架的无参考反卷积统计推断模型 MethEML。该模型通过引入期望最大化(Expectation-Maximization, EM)算法，能够在无外部先验参考谱的条件下，联合推断多种细胞亚群的混合比例及其特异性的 DNA 甲基化分布。基于真实细胞系(HCC1954 与 HMEC)模拟数据集的基准测试表明，MethEML 在应对不同测序覆盖度与多维复杂混合比例时，其参数估计的精度及模型鲁棒性均显著优于传统的二元解卷积模型 MethylPurify。本研究不仅为高维肿瘤甲基化数据的微环境异质性解析提供了创新的统计计算工具，亦为未来基于循环肿瘤 DNA (ctDNA)甲基化特征的无创液体活检与泛癌种早期筛查奠定了坚实的算法基础[13]-[15]。

2. 统计模型构建与算法推断

MethEML 算法的输入为“癌症–正常”混合样本的 BS-seq 测序数据，模型需要估计的参数记为 $\Theta=\{\pi_1,\pi_2,\cdots,\pi_{K-1};p_{11},\cdots,p_{1K},p_{21},\cdots,p_{2K},\cdots,p_{C1},\cdots,p_{CK}\}$。该参数空间包含了肿瘤混合组织中各细胞亚群的混合比例 $\{\pi_1,\pi_2,\cdots,\pi_{K-1}\}$，以及每一个 CpG 位点在特定细胞亚群中的固有甲基化水平 $\{p_{11},\cdots,p_{1K},p_{21},\cdots,p_{2K},\cdots,p_{C1},\cdots,p_{CK}\}$。

2.1. 概率模型与似然函数

假设肿瘤样本的 BS-seq 数据源自 K 种细胞亚群，π_k 表示第 k 种细胞亚群的真实混合比例($k=1,\cdots,K$)，且满足约束条件 $\sum_{k=1}^{K}\pi_k=1$。假设测序数据覆盖的 CpG 位点总数为 C，由 BS-seq 实验产生的总读段(Read)数为 F。对于 CpG 位点 c，令 p_{ck} 表示其在第 k 种细胞亚群中的甲基化水平。

对于任意读段片段 f，假设其覆盖了 n_f 个 CpG 位点，定义集合 $S_f=\{s_{fl},l=1,\cdots,n_f\}$ 表示片段 f 覆盖的 CpG 位点在所有 CpG 位点中位置索引构成的集合。令 $Y_f=\{y_{fl},l=1,\cdots,n_f\}$ 表示这些 CpG 位点观测到的甲基化状态，其中 $y_{fl}\in\{0,1\}$，取 1 代表甲基化，取 0 代表未甲基化。由于每一条独立读段必定且只能源于单一细胞，我们引入隐藏变量 Z 表示片段 f 的细胞亚群标签：当 $Z_f=k$ 时，表示片段 f 源自第 k 种细胞类型。在给定 $Z_f=k$ 的条件下，各 CpG 位点的甲基化状态服从独立的伯努利分布，即 $y_{fl}\mid Z_f=k\sim\text{Bernoulli}\left(p_{s_{fl},k}\right)$。

基于上述假设，片段 f 在 CpG 位点的甲基化状态的条件似然函数可表示为：

$$P\left(Y_f\mid Z_f=k\right)=\prod_{l=1}^{n_f}p_{s_{fl},k}^{y_{fl}}\left(1-p_{s_{fl},k}\right)^{1-y_{fl}}$$

将 Z_f 看作隐藏变量，引入指示函数 $\delta\left(Z_f=k\right)$，则片段 f 在 CpG 位点的联合似然函数为：

$$P\left(Y_f,Z_f\right)=\prod_{k}\left\{\pi_k\prod_{l=1}^{n_f}p_{s_{fl},k}^{y_{fl}}\left(1-p_{s_{fl},k}\right)^{1-y_{fl}}\right\}^{\delta\left(Z_f=k\right)}$$

令 $Y=\{Y_f;f=1,\cdots,F\},Z=\{Z_f;f=1,\cdots,F\}$，则整个样本集的完全数据似然函数为：

$$P\left(Y,Z\right)=\prod_{f}\prod_{k}\left\{\pi_k\prod_{l=1}^{n_f}p_{s_{fl},k}^{y_{fl}}\left(1-p_{s_{fl},k}\right)^{1-y_{fl}}\right\}^{\delta\left(Z_f=k\right)}$$

对上式取对数，得到参数的完全数据对数似然函数：

$$l\left(\Theta\right)=\sum_{f}\sum_{k}\delta\left(Z_f=k\right)\left\{\log\pi_k+\sum_{l=1}^{n_f}\left[y_{fl}\log\left(p_{s_{fl},k}\right)+\left(1-y_{fl}\right)\log\left(1-p_{s_{fl}}\right)\right]\right\}$$

2.2. 参数推断：EM 算法

由于片段的细胞来源 Z 是不可观测的隐变量，本文采用期望最大化(Expectation-Maximization, EM)算法对目标参数进行迭代求解。令 $\mu_{fk}=\delta\left(Z_f=k\right)$，设在第 t 次迭代时参数的估计值为 $\Theta^{(t)}$，μ_{fk} 的条件期望记为 $\mu_{fk}^{(t)}$。

E 步(Expectation Step)：计算隐变量在当前参数下的后验概率：

$$\mu_{fk}^{(t)}=E\left[\delta\left(Z_f=k\right)\mid Y,\Theta^{(t)}\right]=P\left(Z_f=k\mid Y,\Theta^{(t)}\right)=\frac{\pi_k^{(t)}P\left(Y_f\mid Z_f=k,\Theta^{(t)}\right)}{\sum_{k'=1}^{K}\pi_{k'}^{(t)}P\left(Y_f\mid Z_f=k',\Theta^{(t)}\right)}$$

将上述期望值代入完全数据对数似然函数中，构造辅助函数(Q 函数)：

$$Q\left(\Theta,\Theta^{(t)}\right)=E\left[l\left(\Theta\right)\mid Y,\Theta^{(t)}\right]=\sum_f\sum_k\mu_{fk}^{(t)}\left\{\log\pi_k+\sum_{l=1}^{n_f}y_{fl}\log\left(p_{s_{fl},k}^{(t)}\right)+\left(1-y_{fl}\right)\log\left(1-p_{s_{fl}}^{(t)}\right)\right\}$$

M 步(Maximization Step)：通过最大化 $Q\left(\Theta\mid\Theta^{(t)}\right)$ 更新模型参数。对 π_k 求偏导并令 $\frac{\partial Q}{\partial\pi_k}=0$，可得 π_k 的更新公式为：

$$\pi_k^{(t+1)}=\frac{\sum_f\mu_{fk}^{(t)}}{F}$$

同理，对 p_{ck} 求导以获得更新值。对于 CpG 位点 c，定义覆盖该位点的读段片段集合为 $A_c=\left\{f:c\in S_f\right\}$。对于任意 $f\in A_c$，存在局部索引 l_f 使得 $s_{f,l_f}=c$ (即位点 c 在片段 f 的相对位置)。定义集合对 $B_c=\left\{\left(f,l_f\right):f\in A_c,s_{f,l_f}=c\right\}$，则 p_{ck} 的解析更新公式为：

$$p_{ck}^{(t+1)}=\frac{\sum_{\left(f,l_f\right)\in B_c}\mu_{fk}^{(t)}y_{f,l_f}}{\sum_{f\in A}\mu_{fk}^{(t)}}$$

通过 EM 算法的交替迭代，模型最终收敛并输出各细胞亚群的混合比例与特异性甲基化谱。在实际应用中，隐类数量 K 需预先指定，可通过贝叶斯信息准则(BIC)或结合先验生物学背景进行自适应选择。值得注意的是，传统的全基因组 EM 算法对初始值极度敏感，极易陷入局部最优解。为保证参数估计的全局稳定性与可重复性，MethEML 首先基于测序读段的物理覆盖范围(相邻 CpG 位点的连通性)，将全基因组划分为多个独立的高连通性 CpG 区域；其次，在每个局部区域内，将混合比例均匀初始化为 1/K，组分甲基化水平在 0.1 至 0.9 之间等距设定，并独立运行局部 EM 推断；最后，为解决跨区域的组分失配问题，算法通过对各区域输出的局部混合比例大小进行全局排序与匹配，实现不同细胞亚群的“相位对齐”(对于包含 CpG 位点较少的短跨度区域，则采用均匀分布的随机赋值以增加鲁棒性)。这种经过严格对齐融合后的高置信度参数矩阵(结合了对齐后的局部甲基化谱与取全局均值后的混合比例)，才被作为最终全基因组 EM 迭代的初始输入(Initial Values)。该启发式策略大幅收窄了似然函数的搜索空间，在有效规避局部最优陷阱的同时，显著提升了高维表观数据反卷积的收敛效率与统计鲁棒性。

3. 模拟研究与结果分析

为系统评估 MethEML 算法在复杂肿瘤组织混合比例推断中的性能，本文设计了全面的模拟实验，并与现有的经典工具 MethylPurify 进行了基准对比。实验选取乳腺癌细胞系 HCC1954 (ER 阴性/HER2 阳性)与人类正常乳腺上皮细胞系 HMEC 的全基因组重亚硫酸盐测序(WGBS)数据作为源数据进行混合。两种

数据具有不同的测序特征：HCC1954 读段长度为 70 bp，测序深度为 27X；HMEC 读段长度为 100 bp，测序深度为 20X。在数据生成阶段，MethEML 首先按照设定的测序深度与混合比例，随机抽取序列生成混合数据集；随后筛选出覆盖 CpG 位点数 >10 且甲基化水平介于 0.1~0.9 之间的区域作为高信息量的差异甲基化区域输入模型。最终采用 1 减去 HCC1954 所占比例的相对误差作为模型的预测精度指标。由于该模拟数据集直接源于真实生物样本，故能高保真地还原临床测序数据的统计特性。

首先，本研究考察了读段覆盖度对 MethEML 推断精度的影响。在 HMEC 和 HCC1954 真实混合比例设定为 0.7:0.3 的条件下，分别在 5、10、20、30、40、50 倍的覆盖度下进行测试(每组独立模拟 100 次以检验稳定性)。如图 1(a)所示，除在极低覆盖度(5 倍)下精度出现下降外，在≥10 倍的测序深度下，MethEML 均保持了约 0.93 的高推断精度，证明了该算法对不同测序深度具有极强的鲁棒性。

随后，在覆盖度为 20 倍的前提下，研究评估不同混合比例下的模型表现。实验设置 HMEC 和 HCC1954 的混合比例为 1:0、0.9:0.1、0.8:0.2、0.7:0.3、0.6:0.4、0.5:0.5 (每组 100 次模拟)。图 1(b)结果表明，MethEML 在全梯度区间内均表现优异，全局平均精度达到 0.84。值得注意的是，在反卷积任务中，当各组分比例极其接近时通常难以准确推断，但即使在 0.5:0.5 的极端混合挑战下，MethEML 依然实现了约 0.89 的高精度解析。

为直观展现算法优势，我们在同等 20X 覆盖度下，通过 20 次独立重复实验，将 MethEML 与对比算法 MethylPurify 进行了精度与均方误差(MSE)的头对头比较。图 2(a)显示了两种算法在多组混合梯度下对 HCC1954 比例的预测分布。整体而言，MethEML 的预测中位数更贴近真实值；在 0.5:0.5 的困难场景中，MethEML 的精度明显优于 MethylPurify (0.41 vs. 0.37)。图 2(b)进一步证实，MethEML 在绝大多数梯度下的 MSE 均显著低于对比算法，特别是在 0.5:0.5 时，MethylPurify 的推断误差几乎是 MethEML 的 2 倍。

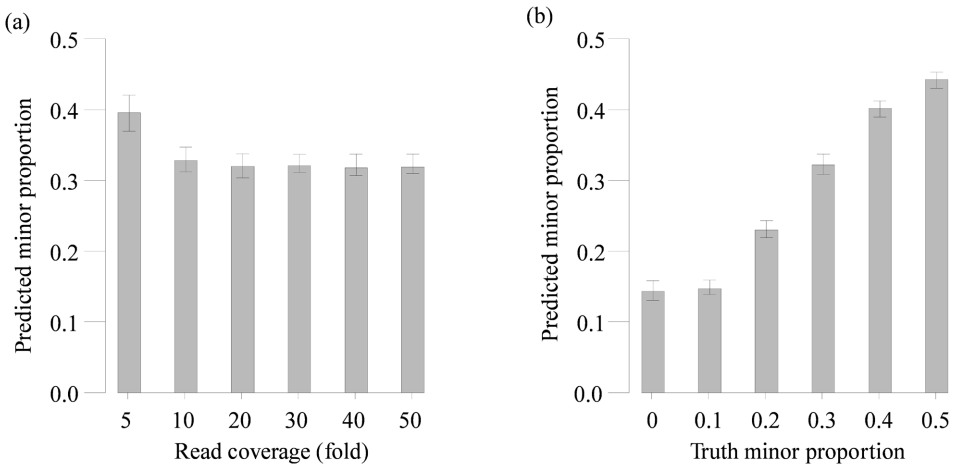

Figure 1. Effects of coverage and mixing ratio on the accuracy of MethEML
图 1. 测序覆盖度与细胞混合比例对 MethEML 推断精度的影响

为进一步衡量算法的方差控制能力，图 2(c)显示不同混合比例下 MethEML 20 次模拟的标准偏差。图中虚线代表混合组织中细胞系 HCC1954 真实所占的比例(分别为 0.1、0.2、0.3、0.4、0.5)，蓝色的误差条代表 20 次模拟的标准偏差。结果显示，模型在各比例下的估计方差均收敛于 10e−05 量级，表现出极高的统计稳定性(例如设定比例为 0.8:0.2 时，20 次预测均值为 0.19)。通过该图还可观察到两个底层统计学特性：第一，比例为 0.5:0.5 时估计方差略有增加，这是由于组分先验概率相近导致隐变量后验推断的熵增；第二，模型存在轻微低估肿瘤纯度的倾向。由于癌细胞系 HCC1954 较正常细胞 HMEC 具有更高

的固有表观异质性(即类内方差更大)，EM 算法在执行 M 步归属时，易将边缘状态的零散片段错分为低方差的 HMEC 组，从而造成轻微的估计偏倚。

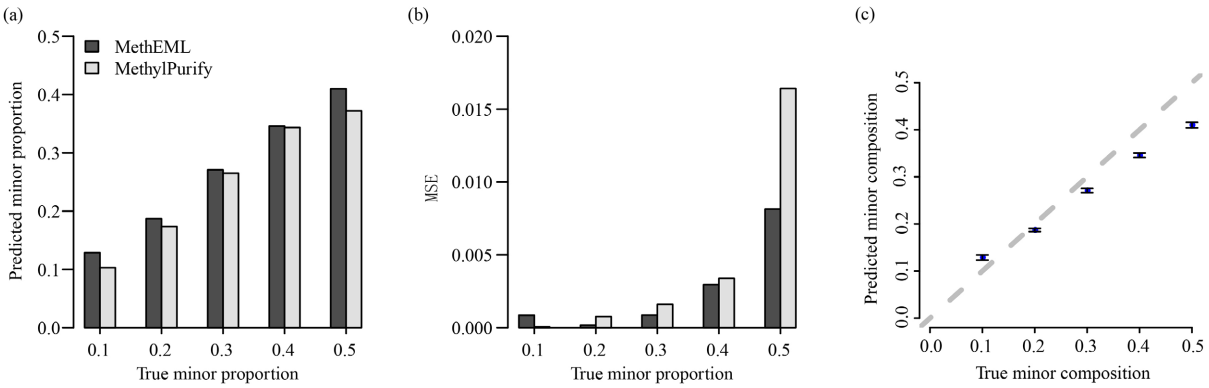

Figure 2. Comparison of accuracy and MSE between MethEML and MethylPurify under different mixing ratios
图 2. 不同混合比例下 MethEML 与 MethylPurify 的推断精度与均方误差对比

传统 MethylPurify 算法的底层逻辑受限于“正常–肿瘤”的二元假设(即强制 $K=2$)，而 MethEML 的 MLE 框架支持将 K 拓展至高维多态微环境。为验证此特性，本研究构建了混合比例为 0.2:0.3:0.5 的三组分模拟系统。在无先验输入的情况下，将备选的 K 值设定为 2 至 5，通过模型计算，图 3(a)显示当 $K=3$ 时 BIC 指标取得全局极小值(四组 BIC 分别为 32007、31994、32524、33041)，这验证了采用 BIC 准则作为内源性决定最佳亚群数量 K 的统计合理性。图 3(b)展示了在 $K=3$ 时的解卷积结果，100 次模拟的预测均值分别为 0.21、0.29 和 0.50，与真实比例高度吻合，证明了模型在解析多维肿瘤微环境方面的卓越能力。为了进一步探讨开发多组分模型的必要性，我们分析了传统二元模型 MethylPurify 在处理三组分数据时的失效表现。由于 MethylPurify 的概率模型严格限定 $K=2$，仅定义了两个细胞分量及对应的比例参数。当面对 0.2:0.3:0.5 的三元混合系统时，该算法的 EM 迭代过程被迫将所有测序读段强制划分至两个隐藏状态中。这种模型错设导致算法将甲基化特征相近的亚群进行了“错误合并”，例如将比例为 0.2 和 0.3 的两个少数类亚群识别为一个占比约 0.5 的单一分量。其结果不仅导致比例推断完全偏离真实值，更使得估算的亚群特异性甲基化谱成为了多个真实谱线的加权平均值，失去了生物学解析意义。

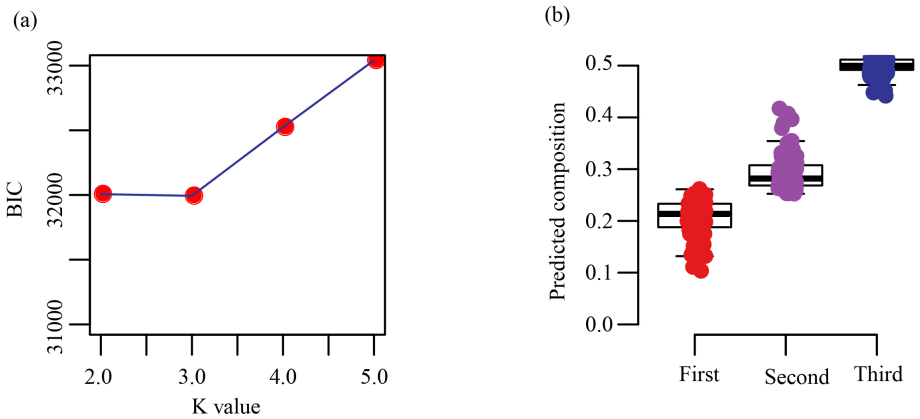

Figure 3. Determination of optimal K via BIC and deconvolution accuracy of MethEML for three cell types
图 3. 基于 BIC 准则的最优 K 值选择及 MethEML 对三元细胞混合物的解卷积精度

4. 结论

本文创新性地提出并开发了 MethEML，一种基于极大似然估计与期望最大化框架的无参考 BS-seq 数据反卷积算法。理论推导与模拟实验均充分证明，该算法能够高鲁棒性地分离并量化混合样本中的多维组分，有效解决了复杂肿瘤微环境中异质性细胞混合比例推断的计算难题。MethEML 不仅突破了传统二元分解模型在应用场景与推断维度上的双重瓶颈，也为下游更精准的表观遗传差异分析与临床分子分型提供了坚实的量化基础。

基金项目

研究得到了江西省自然科学基金(20212BAB202001)的支持。

参考文献

[1] Xu, Y.L., Ma, S.Y., Xu, M.Y., Zhu, H., Wang, Y., Dong, W., *et al*. (2025) DNA Methylation Heterogeneity in Complex Tumor Microenvironment: Quantitative Methods, Influencing Factors, and Clinical Implications. *Genes & Diseases*, **13**, Article ID: 101832. https://doi.org/10.1016/j.gendis.2025.101832

[2] Zhou, Y., Liu, J., Shi, B., Ma, T., Yu, P., Li, J., *et al*. (2025) Evaluation of Pan-Cancer Immune Heterogeneity Based on DNA Methylation. *Genes*, **16**, Article No. 160. https://doi.org/10.3390/genes16020160

[3] Ferro dos Santos, M.R., Giuili, E., De Koker, A., Everaert, C. and De Preter, K. (2024) Computational Deconvolution of DNA Methylation Data from Mixed DNA Samples. *Briefings in Bioinformatics*, **25**, bbae234. https://doi.org/10.1093/bib/bbae234

[4] Ma, S., Pan, X., Gan, J., Guo, X., He, J., Hu, H., *et al*. (2024) DNA Methylation Heterogeneity Attributable to a Complex Tumor Immune Microenvironment Prompts Prognostic Risk in Glioma. *Epigenetics*, **19**, Article ID: 2318506. https://doi.org/10.1080/15592294.2024.2318506

[5] Dietrich, A., Willruth, L.L., Pürckhauer, K., *et al*. (2025) Unifying DNA Methylation-Based *in Silico* Cell-Type Deconvolution with deconvMe. *Bioinformatics Advances*, **5**, vbaf201.

[6] Li, L.Y. and Sun, Y.L. (2024) Circulating Tumor DNA Methylation Detection as Biomarker and Its Application in Tumor Liquid Biopsy: Advances and Challenges. *MedComm*, **5**, e766. https://doi.org/10.1002/mco2.766

[7] Zhang, Y., Naderi Yeganeh, P., Zhang, H., Wang, S.Y., Li, Z., Gu, B., *et al*. (2024) Tumor Editing Suppresses Innate and Adaptive Antitumor Immunity and Is Reversed by Inhibiting DNA Methylation. *Nature Immunology*, **25**, 1858-1870. https://doi.org/10.1038/s41590-024-01932-8

[8] Rendek, T., Pos, O., Duranova, T., Saade, R., Budis, J., Repiska, V., *et al*. (2024) Current Challenges of Methylation-Based Liquid Biopsies in Cancer Diagnostics. *Cancers*, **16**, Article No. 2001. https://doi.org/10.3390/cancers16112001

[9] Cai, M., Zhou, J., McKennan, C. and Wang, J. (2024) scMD Facilitates Cell Type Deconvolution Using Single-Cell DNA Methylation References. *Communications Biology*, **7**, Article No. 1. https://doi.org/10.1038/s42003-023-05690-5

[10] Qi, T., Lakshmanan, L.N., Yang, Y., Zhou, Y., Pan, M., Skanderup, A.J., *et al*. (2025) Read-Level DNA Methylation Deconvolution Enhances Circulating Tumor DNA Detection. *Briefings in Bioinformatics*, **26**, bbaf551. https://doi.org/10.1093/bib/bbaf551

[11] Wang, Y.X., Li, J.Y., Li, J.Q., *et al*. (2025) cfDecon: Accurate and Interpretable Methylation-Based Cell Type Deconvolution for Cell-Free DNA.

[12] Zheng, X., Zhao, Q., Wu, H., Li, W., Wang, H., Meyer, C.A., *et al*. (2014) MethylPurify: Tumor Purity Deconvolution and Differential Methylation Detection from Single Tumor DNA Methylomes. *Genome Biology*, **15**, Article No. 419. https://doi.org/10.1186/s13059-014-0419-x

[13] Zhao, P.P. and Hu, R. (2025) Application of Circulating Tumor DNA Methylation Characteristics in Early Diagnosis and Prognosis Monitoring of Lung Cancer. *American Journal of Translational Research*, **17**, 8939-8952. https://doi.org/10.62347/eljo6418

[14] Zhou, S., Yin, H., Yan, L., Xie, N. and Fu, C. (2025) ctDNA Methylation Profiling Reveals NBL1 as a Promising Biomarker for Early Ovarian Cancer Screening. *World Journal of Surgical Oncology*, **23**, Article No. 305. https://doi.org/10.1186/s12957-025-03957-1

[15] Liang, S.I., Quandt, Z., Wienke, S., Wang, J., Gordon, S., Barnett, R.M., *et al*. (2025) Methylation-Based ctDNA Tumor Fraction Changes Predict Long-Term Clinical Benefit from Immune Checkpoint Inhibitors in RADIOHEAD, a Real-World Pan-Cancer Study. *Cancer Research Communications*, **5**, 1384-1395. https://doi.org/10.1158/2767-9764.crc-25-0151