Supplementary materials for "BLEND-M: cellular deconvolution with personalized DNA methylation references"

Contents

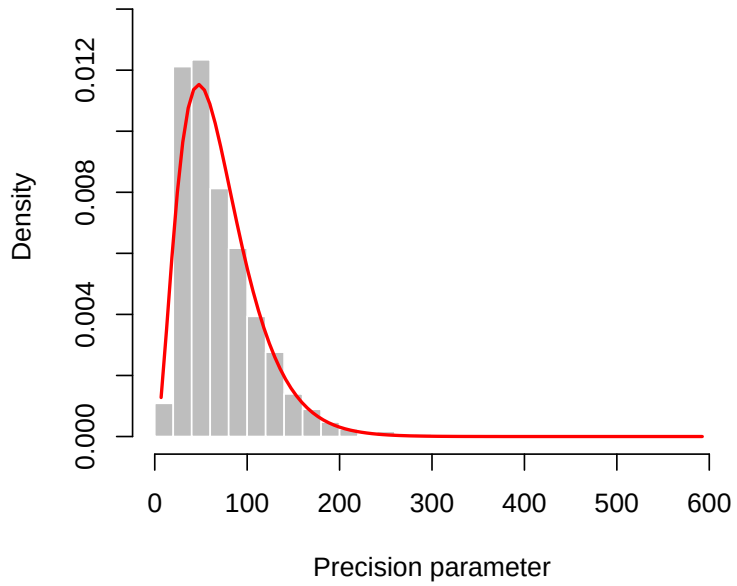
S1 Introduction

This supplement contains additional simulation and CpG marker selection details, as well as a proof of Theorem 1.

S2 Simulation settings

The means of the normal distributions used to sample cellular fractions were specified as B cell = 0.08, CD4 T cell = 0.20, CD8 T cell = 0.10, granulocyte = 0.42, monocyte = 0.05, natural killer = 0.10, nucleated red blood cell = 0.05 for cord blood samples, and B cell = 0.05, CD4 T cell = 0.20, CD8 T cell = 0.10, monocyte = 0.05, neutrophil = 0.55, natural killer = 0.05 for adult blood samples, with standard deviations of 0.02 for all cell types. Negative simulated fractions were set to zero and fractions for each sample were normalized to sum to one.

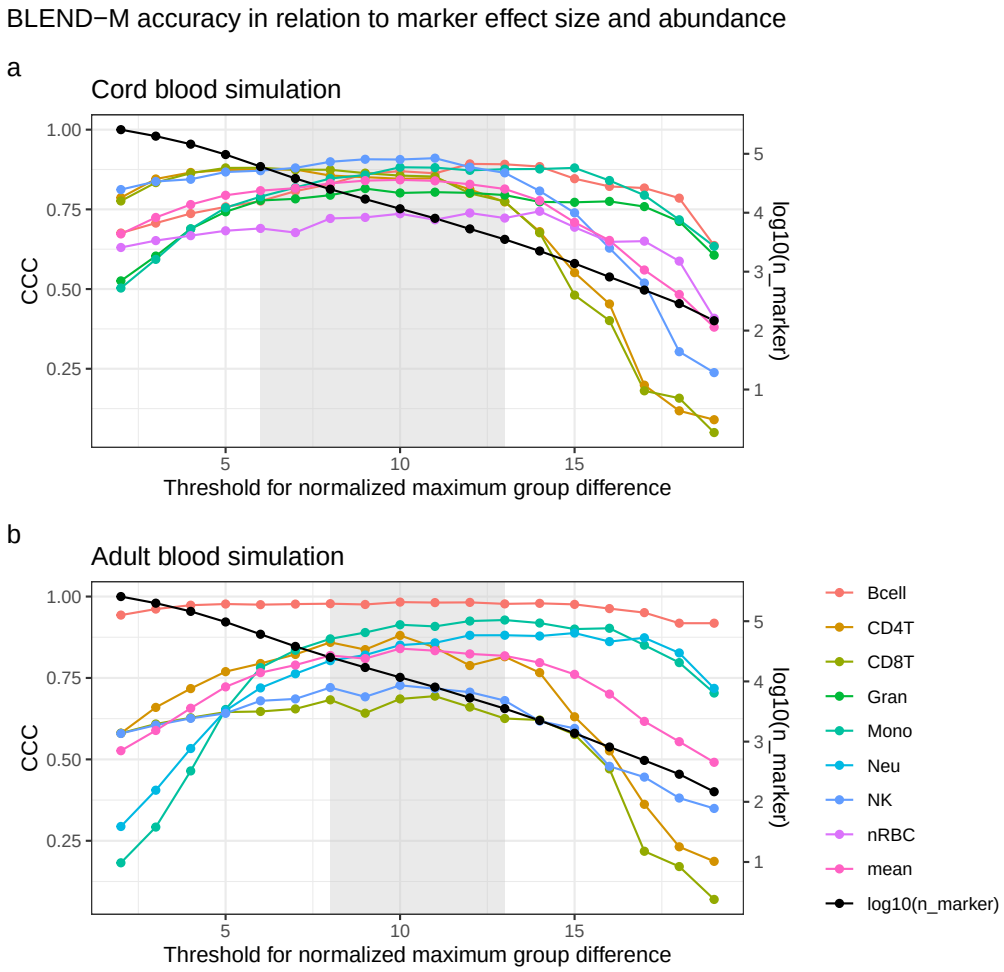
CpG-specific precision parameters ϕ_g were drawn from a gamma distribution with parameters learnt from the Holland 450k dataset. The histogram and fitted density curve are visualized in Supplemental Figure S1.



Supplemental Figure S1: Histogram and fitted distribution of the CpG-specific precision parameters ϕ_g.

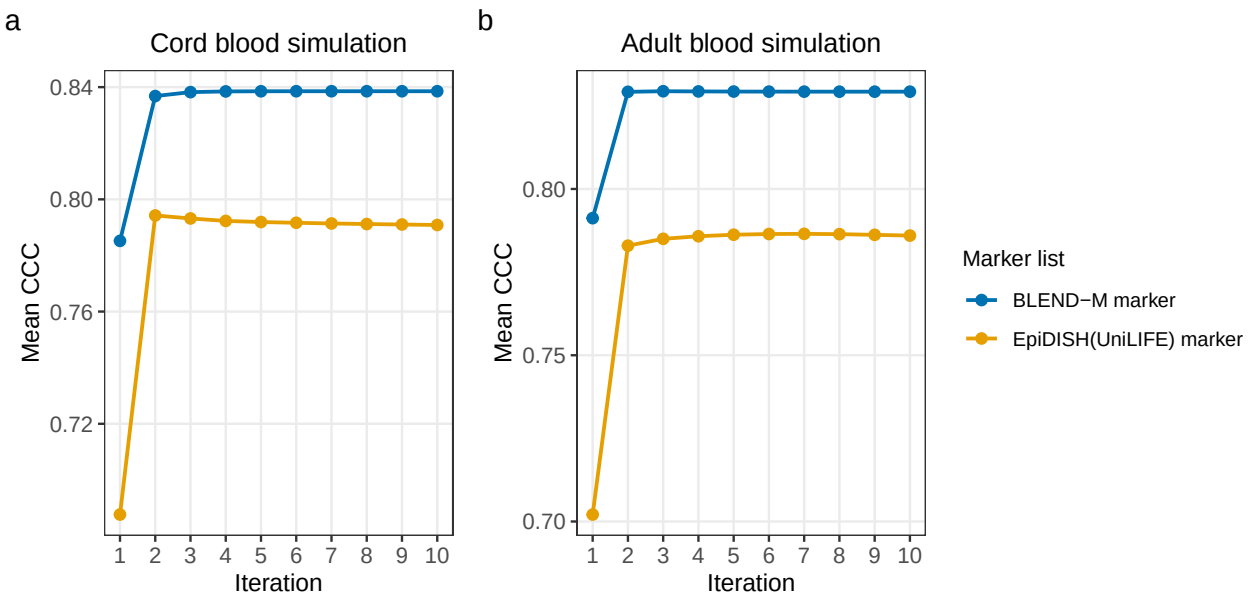
S3 Other marker CpG selection methods and thresholds

We also considered different selection criteria, including calculating p-values with ANOVA using beta values, using the unnormalized maximum group DNAm difference, and different



Supplemental Figure S2: **BLEND-M estimation accuracy in relation to marker effect size and abundance.** Lin's concordance correlation coefficient (CCC) is used to quantify deconvolution accuracy for each cell type, along with the mean CCC across cell types (left y-axis). The number of markers selected by the proposed procedure ($\log_{10}$ scale, right y-axis) is shown across varying thresholds for the normalized maximum group difference (x-axis)

filtering thresholds for p-values and differences. We simulated 30 leave-out bulk DNAm samples per group, in total 10 groups, and estimated fractions using BLEND-M with each marker set. We then calculated the mean CCC between ground truth and estimated cellular fractions. The marker selection procedure described in the main text achieved the highest CCC for both simulated cord blood and adult blood data. All cell type CCCs and mean CCCs for simulated cord and adult blood datasets are provided in Supplementary Tables 2 and 3, Additional File 2. Notably, Figure S2 shows that simulation results are stable across normalized maximum group difference thresholds defined in Step 3 of our marker selection pipeline (Section 2.5 in the main text).



Supplemental Figure S3: **Deconvolution accuracy stabilizes after iteration two.** Mean Lin's concordance correlation coefficient (CCC) across iterations 1–10 for (a) cord blood and (b) adult blood leave-out simulations, using marker CpGs selected by BLEND-M (blue) or the EpiDISH UniLIFE panel (orange).

S4 Estimates stabilize after two iterations

BLEND-M uses a two-step estimator to estimate cell type proportions, which is motivated by the results in Theorem 1 showing that BLEND-M's estimate is asymptotically equivalent to the oracle estimator when the variances of DNAm levels are know. We used simulations to further validate this approach. As shown in Figure S3, deconvolution accuracy improves substantially from iteration one to two under both marker lists, then stabilizes with negligible change in subsequent iterations. This is perfectly consistent with Theorem 1. We additionally note that our proposed marker list consistently outperforms the EpiDISH UniLIFE marker list across all iterations and both cord and adult blood, further supporting our marker selection procedure.

S5 Additional simulations with cell types drawn from a Dirichlet distribution

To evaluate BLEND-M's performance under alternative distributions for cell type proportions, we performed the same leave-out simulation described in main text but with cell fractions drawn from Dirichlet distribution. The Dirichlet distribution, which produces correlated proportions, is a natural distribution for compositional data such as cell type fractions [1]. We set the Dirichlet parameters to ensure the mean fractions are comparable with those in the main text simulation (cord blood, Bcell: 8, CD4T: 20, CD8T: 10, Gran: 42, Mono: 5, NK: 10, nRBC: 5; adult blood, Bcell: 5, CD4T: 20, CD8T: 10, Mono: 5, Neu: 55, NK: 5).

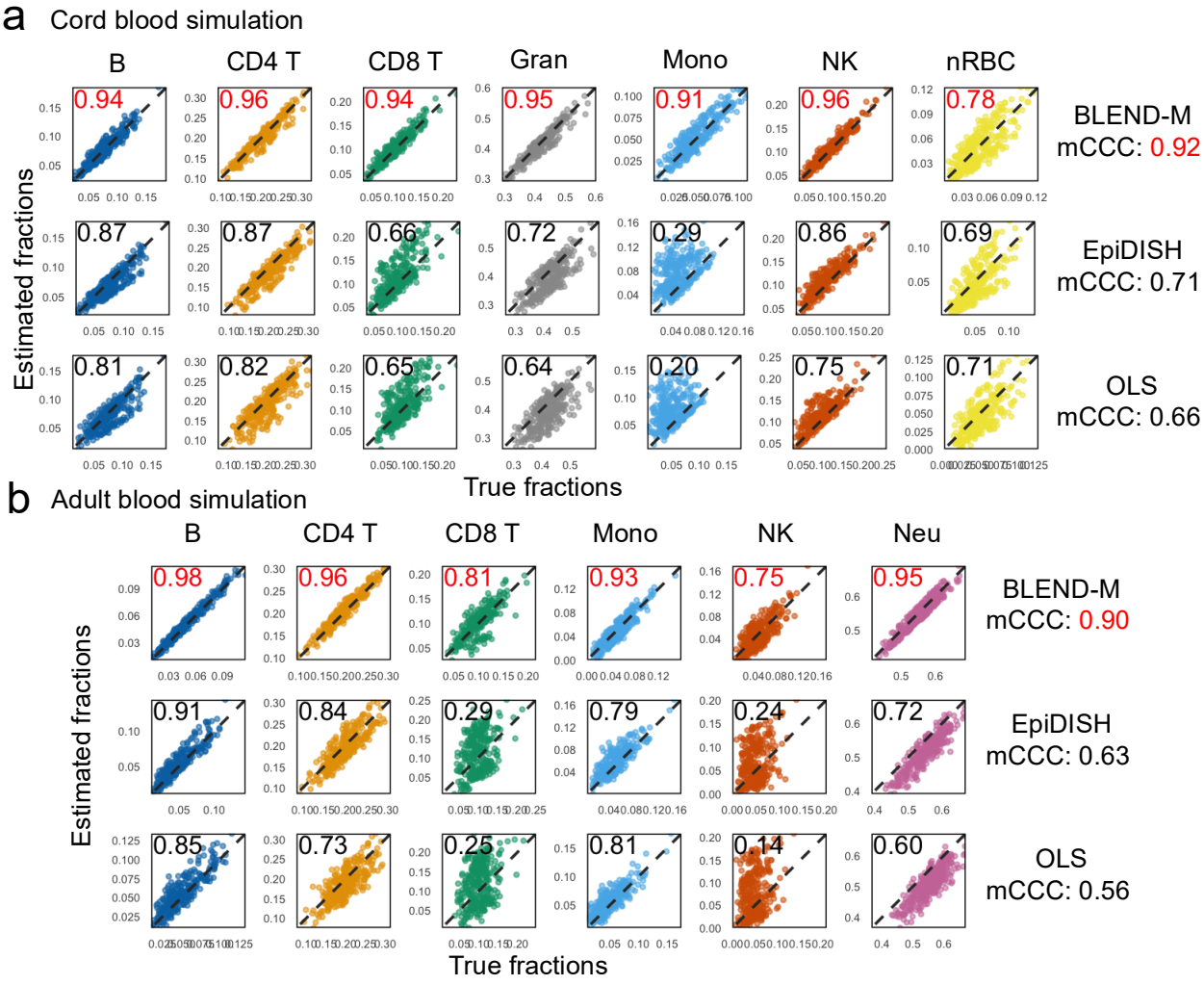
EMeth could not converge on these simulated bulk samples. We computed CCC and mean CCC between fraction estimates from methods and true fractions. We visualized estimates in Figure S4. BLEND-M provides the most accurate estimates with the highest CCC across all cell types in both cord blood and adult blood. BLEND-M's excellent performance in both the normal and Dirichlet settings is expected, as Theorem 1 in the main text, which helps establish BLEND-M's accuracy, is proven conditional on the true cell type proportions. Since proportions are non-random, issues regarding independence or correlation among them do not arise.

S6 Simulations to validate life-stage markers

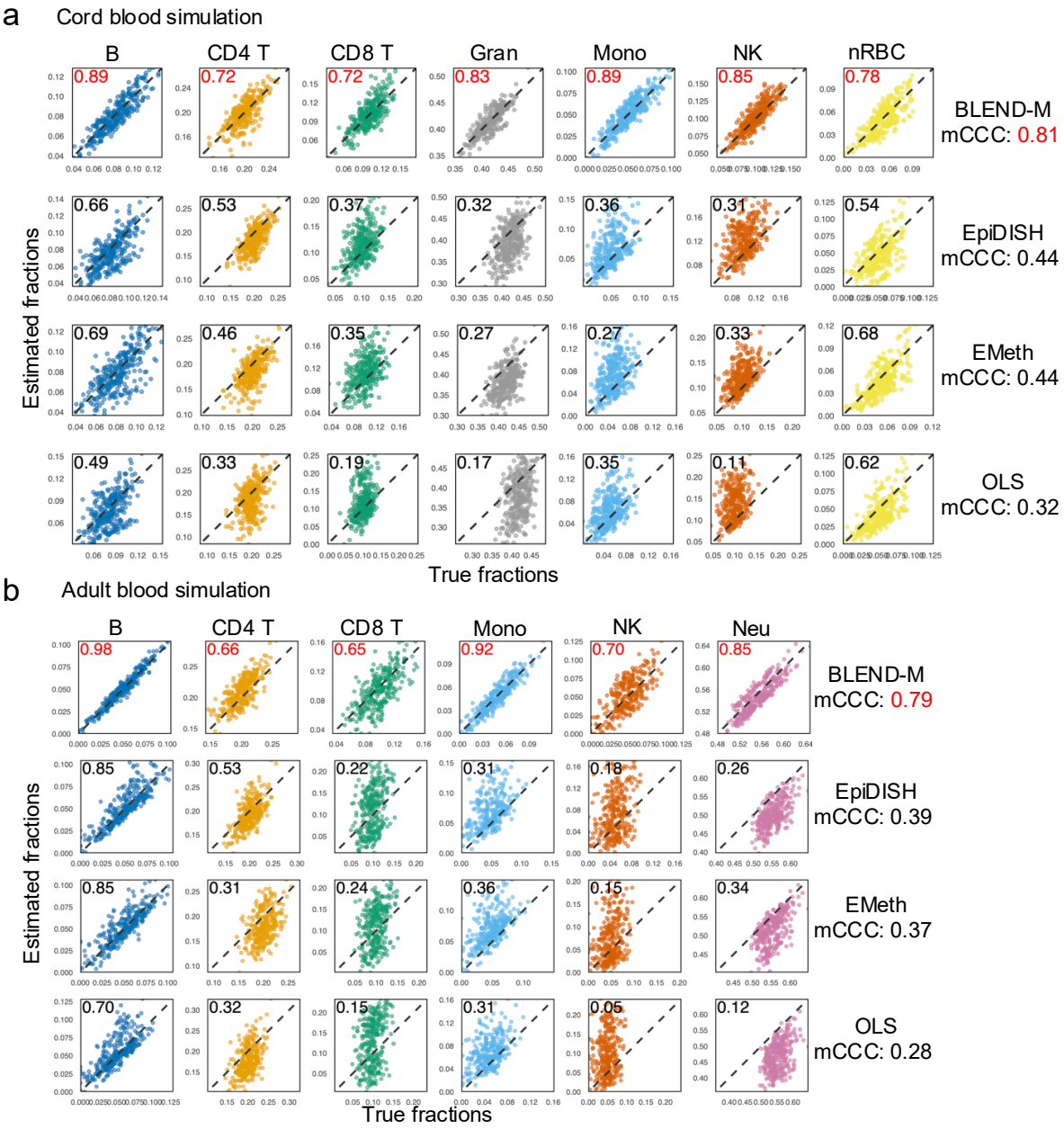
To justify our approach of requiring markers exhibit life-stage-related variation, we repeated the leave-out simulation described in our manuscript but with marker CpGs selected solely by cell type specificity. The thresholds in marker selection were kept the same. We selected 5,415 marker CpGs. We computed CCC between estimated fractions from deconvolution methods and true fractions, and the mean CCC across cell types. We visualized the estimated fractions in Figure S5. All four methods: BLEND-M, EpiDISH, EMeth, OLS, achieved lower mean CCC than when using cell type, life-stage informative markers (Table S1). However, BLEND-M is more robust to the new set of less discriminative markers as the mean CCC decrease is only 3.5% in cord blood and 4.8% in adult blood, while the mean CCC decrease ranges from 15.4% to 39.1% when using other methods. This is expected, as BLEND-M's reference blending strategy is designed to model inter-individual variability in CTS DNAm, including variation not explicitly identified during marker selection. Consistent with the large age-related variation observed in T cells, the major estimation performance decrease in BLEND-M is from CD4 T and CT8 T cells. In cord blood simulation, both CD4 and CD8 T had 14.3% decreases in CCC. In adult blood simulation, CD4 T had 23.3% decrease on CCC. Taken together, these results help validate our approach of selecting markers with both cell-type and life-stage specificity.

Table S1: **Performance decrease using only cell-type informative markers.** Percentage decrease in cross-cell type mean CCC when marker CpGs are selected with only cell-type specificity, relative to markers selected with cell-type and life-stage specificity, across deconvolution methods and blood tissue types.

	BLEND-M	EpiDISH	EMeth	OLS
Cord blood	3.6	15.4	15.4	28.9
Adult blood	4.8	25.0	26.0	39.1



Supplemental Figure S4: **BLEND-M outperforms competing methods under Dirichlet-distributed cell type fractions.** Scatter plots comparing true cell type fractions (x-axis) against estimated fractions (y-axis) for BLEND-M, EpiDISH, and OLS across (a) cord blood and (b) adult blood simulations. EMeth is excluded from the comparison as it did not converge on these simulated samples. Lin's concordance correlation coefficient (CCC) for each cell type is displayed in the upper-left corner of each scatter plot; the cross-cell type mean CCC (mCCC) is shown to the right of each method row, with the best performer highlighted in red. The dashed diagonal line ($y = x$) represents perfect agreement.



Supplemental Figure S5: **Simulations using marker selected with only cell-type specificity.** Scatter plots comparing true cell type fractions (x-axis) against estimated fractions (y-axis) from deconvolution methods in (a) cord blood and (b) adult blood simulations, with marker CpGs selected solely using cell type information. Lin's concordance correlation coefficient (CCC) between ground truth and estimates for each cell type is displayed in the upper-left corner of each scatter plot; the cross-cell type mean CCC (mCCC) is shown to the right of each method row, with the best performer highlighted in red.

S7 Comments on the optimal reference panel composition

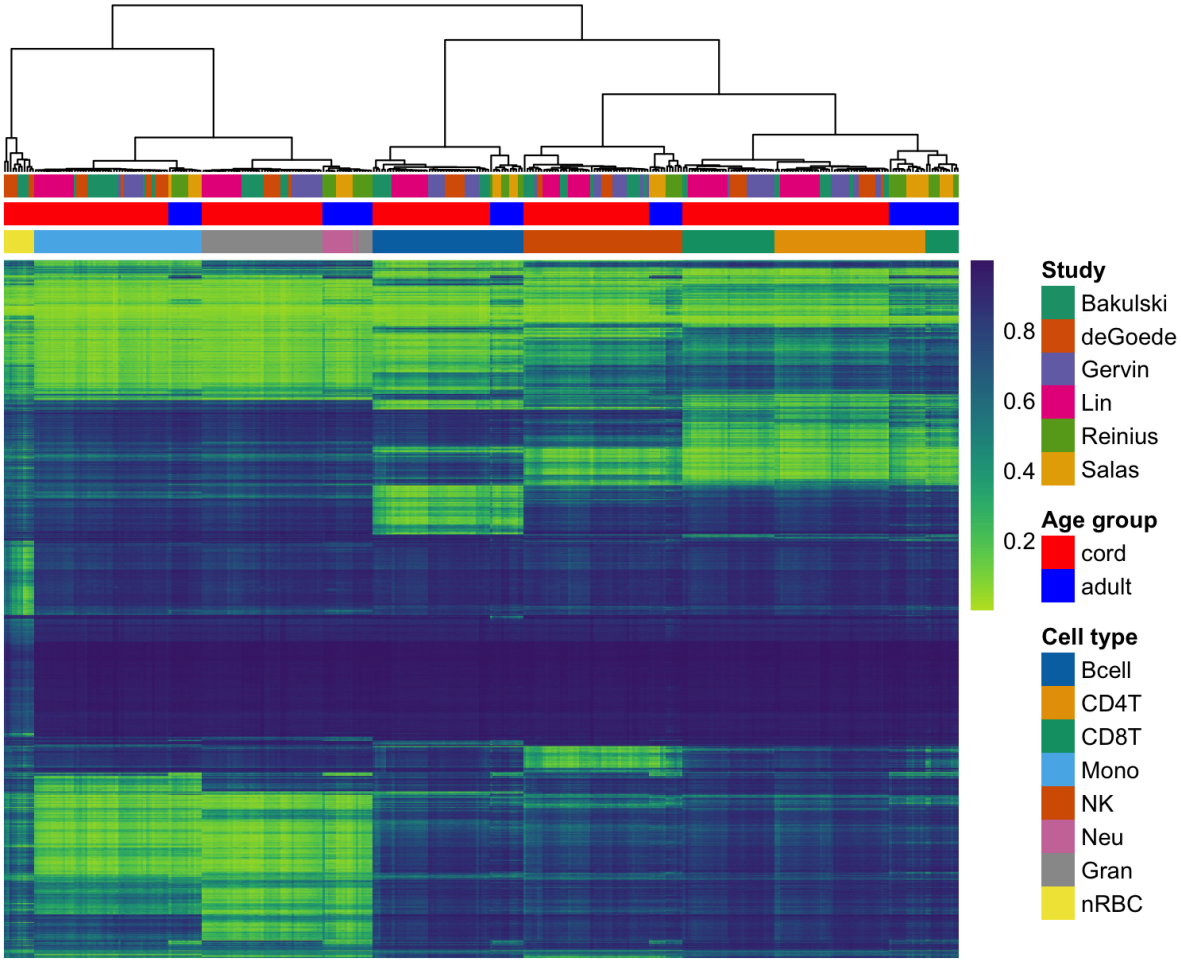
The "optimal" reference panel for deconvolution depends heavily on the population of bulk samples and on what is known about that population prior to analysis. For example, if the cell type-specific (CTS) DNAm levels of each bulk sample were known, the optimal signature would simply be the corresponding matched references. This small signature would likely perform better than a signature composed of a larger number of reference samples that did not capture the bulk samples' CTS DNAm. To illustrate, we used our leave-out simulation paradigm to simulate 100 cord blood and 100 adult blood bulk DNAm samples, ran BLEND-M with all remaining CTS profiles as the reference panel, and obtained the reference mixing proportions $\hat{\pi}_{nmt}$, which can be interpreted as bulk sample n's "preference" over the available references $m = 1, \ldots, M_t$ for cell type t. We labeled a reference "preferred" if its mean mixing proportion exceeded the uniform threshold $1/M_t$ and "not preferred" otherwise. The preferred panels contained only 66 and 24 profiles for cord and adult blood, respectively, versus 269 and 312 in the not-preferred panels. Running BLEND-M with each panel separately (Table S2), we found that expanding beyond the preferred set had little impact, while the larger not-preferred set performed substantially worse than the smaller preferred set.

Table S2: **Mean CCC for BLEND-M using full, preferred, and not preferred reference panels.** The corresponding reference panel sizes are 335, 66, and 269 for cord blood, and 336, 24, and 312 for adult blood.

	Full panel	Preferred panel	Not preferred panel
Cord blood	0.904	0.909	0.787
Adult blood	0.892	0.863	0.623

Although there is no one-to-one relationship between reference panel size or diversity and deconvolution performance, the near-identical CCCs for the full and preferred panels suggest that, given the choice, a more diverse reference panel is always preferable to a less diverse one. To further illustrate this, we simulated cord blood samples using the approach described in the main text—excluding nRBC, as this cell type is absent in adult blood—and applied BLEND-M with an adult-only reference panel to deconvolve them. The per-cell-type CCCs were Bcell 0.67, CD4T 0.43, CD8T 0.28, Gran 0.69, Mono 0.59, and NK 0.84, with a mean CCC of 0.58—substantially worse than in our main-text simulations, whose reference panel included both cord and adult blood references.

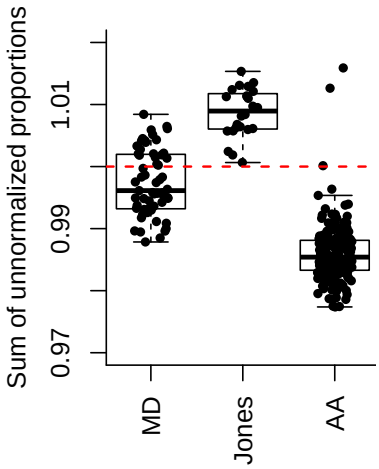
Taken together, these results suggest two things. First, there is no one-to-one relationship between the number or diversity of references and deconvolution performance: a small matched panel can outperform a large poorly matched one (Table S2). Second, expanding the reference panel is nonetheless always weakly preferable, since it can only leave performance unchanged (Table S2) or, as the adult-only simulation illustrates, improve it substantially.



Supplemental Figure S6: **Heatmap of DNAm with dataset source information.** Heatmap of DNAm beta values across 11,671 selected marker CpGs (rows) and 342 purified cell type reference samples (columns), with column annotation bars indicating cell type, age group, and dataset source.

S8 Heterogeneity in reference datasets

Here we examine heterogeneity in the datasets used to curate reference data. Figure S6 gives a heat map for BLEND-M's markers. While this shows a moderate tendency to cluster by dataset, there are also many cases where profiles from different datasets are intermixed. However, such dataset-specific variation is not a concern for BLEND-M. Dataset-level effects can arise from many non-technical sources—cohort composition, exposure histories, and so on—all of which are meaningful forms of heterogeneity that BLEND-M is designed to capture when building personalized references. To the extent that some dataset-level effects are technical in origin, they are unlikely to be shared by independent bulk data and therefore unlikely to influence deconvolution results. Moreover, technical variation is typically very small relative to the biological variability in DNAm levels across subjects [2].



Supplemental Figure S7: The sum of unnormalized cell proportions in the methyldeconv (MD), Jones, and atopic asthma (AA) datasets. Each point represents a bulk sample. The dashed red line is drawn at $y = 1$.

S9 Additional real data results

S9.1 P-values for median AUCs

In Section 3.6 of the main text, we computed p-values testing the null hypothesis that atopic asthma status was independent of each method's cell type proportion estimates. To obtain the null distribution, we generated 1,000 permuted atopic asthma outcomes and, for each one, computed the median AUC over 100 data partitions. P-values were computed as (#{permuted median AUCs that exceeded the method's observed median AUC}+1)/(1000+ 1). The addition of 1 helps account for the fact that a finite number of permutation datasets were used to compute the p-value.

S9.2 Missing cell types in real data

Here we evaluate the completeness of BLEND-M's signature—that is, whether it captures all cell types present in the bulk samples or omits some. Because BLEND-M does not impose a sum-to-one constraint in its non-negative least squares optimization, the sum of the unnormalized fractions should equal one if the signature is complete, whereas a sum below one would indicate the presence of cell types missing from the signature. We computed these sums for all three real datasets considered in the main text. As shown in Figure S7, the sums are all close to one, suggesting that BLEND-M's signature captures nearly all cell types present.

S10 The distribution of DNAm levels

S10.1 Checking the beta distribution assumption

Our model from Section 2.1 in the main text assumes DNAm levels y_{gn} follow a beta distribution:

$$y_{gn} \sim \text{Beta}\{\mu_{gn}\phi_g, (1-\mu_{gn})\phi_g\}.$$

Here, the expected value of y_{gn} is μ_{gn}—the weighted average of cell type-specific DNAm levels, where weights are cell type proportions (see Section 2.1 in the main text)—and the variance of y_{gn} is $\mu_{gn}(1-\mu_{gn})/(\phi_g+1)$.

Our two-step estimator for cell type fractions relies only on the mean and variance models being correctly specified. Since the mean model is fixed and given, we validate the beta assumption by verifying that we accurate estimate the variance model. We additional evaluate higher order moments below.

Briefly, we considered DNAm levels at the 11,671 marker CpGs in the atopic asthma dataset from our real data analysis in Section 3.6 [3]. We used BLEND-M's cellular proportion estimates to obtain $\hat{\mu}_{gn}$—estimates of μ_{gn}—for each marker CpG g and sample n. We then defined $\hat{\phi}_g$ as the maximum likelihood estimate of ϕ_g under the assumption that

$$y_{gn} \sim \text{Beta}\{\hat{\mu}_{gn}\phi_g, (1-\hat{\mu}_{gn})\phi_g\},$$

and defined the standardized residuals r_{gn} as

$$r_{gn} = (y_{gn} - \hat{\mu}_{gn})(\hat{\phi}_g + 1)^{1/2}.$$

If y_{gn} follows a beta distribution, then $\text{Var}(r_{gn}) \approx \hat{\mu}_{gn}(1-\hat{\mu}_{gn})$. To check this, we binned the standardized residuals by values of $\hat{\mu}_{gn}$ and, within each bin, computed the average squared residual r_{gn}^2, which serves as a model-free estimate of the data variance. Figure S8A shows that the variance model (red line) almost perfectly overlaps the observed variance, suggesting that the beta assumption is reasonable.

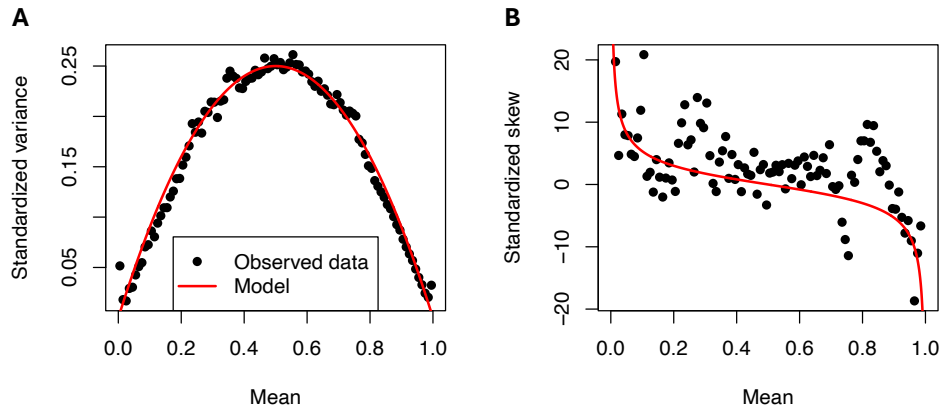
If the beta distribution were the true data-generating distribution, all moments of y_{gn}—not just the mean and the variance—would equal those of a beta distribution. We therefore also evaluated the skew of y_{gn}, which if y_{gn} follows a beta distribution is

$$\text{skew}(y_{gn}) = \frac{\mathbb{E}\{(y_{gn}-\mu_{ng})^3\}}{[\mathbb{E}\{(y_{gn}-\mu_{ng})^2\}]^{3/2}} = \frac{2(1-2\mu_{ng})\sqrt{\phi_g+1}}{(\phi_g+2)\sqrt{\mu_{ng}(1-\mu_{ng})}}.$$

Since our above analysis suggests our estimator for the variance is accurate, we considered the standardized skew of the residuals to be

$$\frac{(y_{gn}-\hat{\mu}_{gn})^3}{\{\widehat{\text{Var}}(y_{gn})\}^{3/2}} \frac{\hat{\phi}_g+2}{(\hat{\phi}_g+1)^{1/2}},$$

where $\widehat{\text{Var}}(y_{gn}) = \hat{\mu}_{ng}(1-\hat{\mu}_{ng})/(\hat{\phi}_g+1)$. If the beta distribution assumption was correct, these would have expectation approximately equal to $2(1-2\mu_{ng})/\sqrt{\mu_{ng}(1-\mu_{ng})}$. Figure S8B suggests this is approximately correct, providing further evidence that the beta assumption is reasonable.



Supplemental Figure S8: The accuracy of the beta assumption. (A): Variance model. Red line: assumed variance model Mean$*(1-$Mean). Black dots: average squared residual within each bin. (B) Skew. Red line: standardized skew under the beta distribution $2*(1-$ 2$*$Mean$)/\sqrt{\text{Mean}*(1-\text{Mean})}$. Black dots: average standardized skew of residuals.

S10.2 Relaxing the beta distribution assumption

The assumption that y_{gn} follows a beta distribution may be too restrictive for some datasets. To address this, we have included an option in BLEND-M that avoids specifying any explicit distribution for y_{gn}. Instead, it assumes only that the first two moments of y_{gn} satisfy

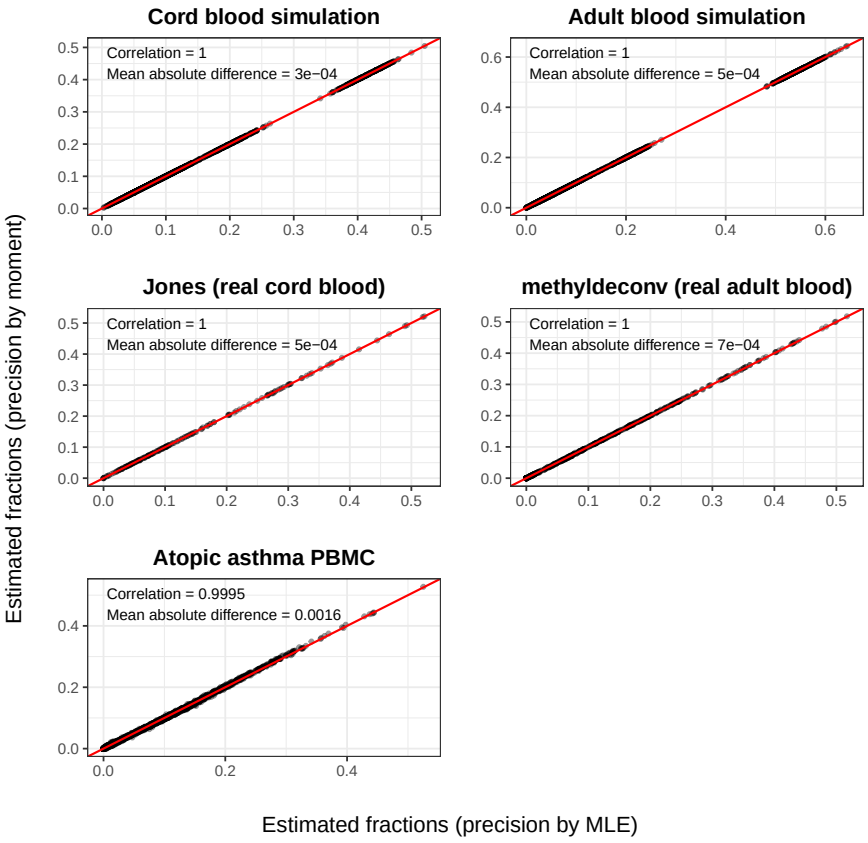
$$\mathbb{E}(y_{gn}) = \mu_{gn}, \quad \text{Var}(y_{gn}) = \frac{\mu_{gn}(1-\mu_{gn})}{\phi_g + 1},$$

which correspond to the mean and variance of the beta distribution. The mean model is well-established in deconvolution. The variance model is justified because it is the standard and most widely-used model for inter-subject variance of DNAm rates in sequencing studies [4, 5, 6, 7, 8], which is widely considered the gold-standard method for measuring DNAm levels [9, 10]. Moreover, previous studies have shown very high to near-perfect concordance between sequencing and microarray methylation rates [11, 12], further supporting the appropriateness of this variance model for microarray data.

Since BLEND-M only uses the beta distribution assumption to estimate ϕ_g, we need only modify this step to accommodate the relaxed assumption. To do so, we simply change our estimator for ϕ_g to be the following method of moments estimator:

$$\hat{\phi}_g = \max\left(\left[N^{-1}\sum_{n=1}^{N}\frac{\{y_{gn}-\hat{\mu}_{gn}^{(1)}\}^2}{\hat{\mu}_{gn}^{(1)}\{1-\hat{\mu}_{gn}^{(1)}\}}\right]^{-1} - 1, 0\right), \tag{S1}$$

where $\hat{\mu}_{gn}^{(1)}$ is BLEND-M's estimate for the mean based off of its first-step estimate for cellular proportions (see Section 2.2 in the main text). Notably, changing the estimator for ϕ_g from the maximum likelihood estimator to this method of moments estimator has no impact on any of our simulation or real data results (Figure S9). We prove in Section S13 that Theorem 1 still holds under this more general setting.



Supplemental Figure S9: BLEND-M's cellular proportion estimates when using the maximum likelihood estimate for ϕ_g (x-axis) and the method of moments estimator (y-axis) across all simulated and real datasets presented in the main text. The red line is the line $y = x$.

S11 Assumptions for Theorem 1

Here we provide the assumptions we use to prove Theorem 1 from the main text assuming DNAm levels y_{ng} follow a beta distribution. Section S13 provides the assumptions and proof when this does not hold. We make a slight change in notation by redefining "cell type proportions" $\boldsymbol{\beta}_n$ to be the regression coefficients $\boldsymbol{\Psi}_n$ defined in Section 2.2 of the main text. As such, we assume throughout that the true model for DNAm levels y_{ng} is

$$y_{ng} \sim \text{Beta}\{\phi_g \mu_{ng}, \phi_g(1-\mu_{ng})\}, \quad \mu_{ng} = \boldsymbol{x}_g^\top \boldsymbol{\beta}_n.$$

Recall our first- and second-step estimators first solve non-negative least squares problems and then are normalized to sum to one. We therefore define our first step estimator, $\bar{\boldsymbol{\beta}}_n^{(1)}$, to be

$$\hat{\boldsymbol{\beta}}_n^{(1)} = \underset{\boldsymbol{\beta} \geq 0}{\arg\min} \sum_{g=1}^{G} (y_{ng} - \boldsymbol{x}_g^\top \boldsymbol{\beta})^2, \quad \bar{\boldsymbol{\beta}}_n^{(1)} = \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(1)}\}^{-1} \hat{\boldsymbol{\beta}}_n^{(1)},$$

where $\mathbf{1}$ is the vector of all ones. Our estimators for CpG-specific precision parameters are

$$\hat{\phi}_g = \underset{\phi \in [c^{-1}, c]}{\arg\max} \sum_{n=1}^{N} \log \text{Beta}\{y_{ng}, ; \phi\hat{\mu}_{ng}, \phi(1-\hat{\mu}_{ng})\}, \quad \hat{\mu}_{ng} = \boldsymbol{x}_g^\top \bar{\boldsymbol{\beta}}_n^{(1)}$$

for some arbitrarily large constant $c > 1$, and our final estimator for cellular fractions, $\bar{\boldsymbol{\beta}}_n^{(2)}$, is

$$\hat{\boldsymbol{\beta}}_n^{(2)} = \underset{\boldsymbol{\beta} \geq 0}{\arg\min} \sum_{g=1}^{G} \frac{1}{\hat{\sigma}_{ng}^2}(y_{ng} - \boldsymbol{x}_g^\top \boldsymbol{\beta})^2, \quad \hat{\sigma}_{ng}^2 = \frac{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})}{\hat{\phi}_g + 1}, \quad \bar{\boldsymbol{\beta}}_n^{(2)} = \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(2)}\}^{-1} \hat{\boldsymbol{\beta}}_n^{(2)}.$$

We additionally assume that the following hold for some arbitrarily large constant $c > 1$.

Assumption 1 (Design conditions)**.** *The matrix* $\boldsymbol{X} = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_G)^\top$ *is non-random and satisfies:*

(i) $\boldsymbol{X}_{gk} \in [c^{-1}, 1 - c^{-1}]$ *for all* $g \in [G] = \{1, \ldots, G\}$ *and* $k \in [K]$.

(ii) The smallest eigenvalue of $G^{-1}\boldsymbol{X}^\top \boldsymbol{X}$ *is greater than* c^{-1}.

Assumption 2 (True coefficients)**.** *The sum of cell type proportions is equal to one, i.e.,* $\mathbf{1}^\top \boldsymbol{\beta}_n = 1$. *Additionally, let* $J_n \in \{2, \ldots, K\}$. *The first* J_n *coefficients of* $\boldsymbol{\beta}_n$ *are non-zero and the second set of* $K - J_n$ *coefficients are zero. The non-zero coefficients satisfy* $\beta_{nk} \geq c^{-1}$ *for all* $k \in [J_n]$.

Assumption 3 (Precision parameters)**.** $\phi_g \in [c^{-1}, c]$ *for all* $g \in [G]$ *and are estimated on the parameter space* $[c^{-1}, c]$.

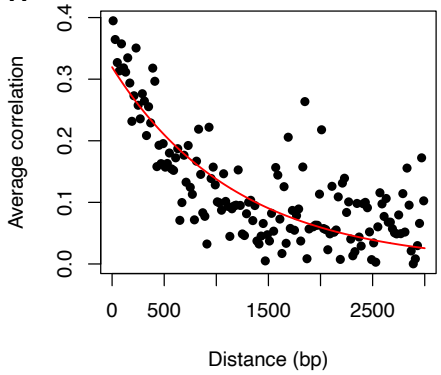
Assumption 4 (Independent observations)**.** *The random variables* $\{y_{ng}\}_{n \in [N]; g \in [G]}$ *are independent.*

Assumption 5 (Sample size)**.** $c^{-1}N \leq G \leq cN^{2-c^{-1}}$.

Assumption 1 assumes methylation probabilities in the signature are not exactly equal to 0 or 1 and cell type signatures are not linearly dependent. It also assumes signatures are fixed and non-random. As such, this ignores potential technical measurement error in the DNAm measurements used to create the signature. However, such measurement error has previously been observed to be very small in comparison to the biological variability in DNAm levels across subjects [2], indicating there is no need to account for it.

Assumption 2 simply states that cell type proportions add up to one, at least two entries of $\boldsymbol{\beta}_n$ are non-zero, and that some proportions are zero and others are non-zero. The assumption that at least two entries are non-zero will hold if each sample is not a single cell type, which is expected given that we are working with bulk tissue data. The fact that the first J_n are non-zero is without loss of generality. As is the case in traditional regression analysis, cell type proportions are treated as non-random model parameters. Assumption 3 is a common assumption used to ensure the consistency of maximum likelihood estimators [13].

A



B

Region size	Number of regions
1	7,336
2	814
3	227
4	78
5	39
6	13
7	13
8	7
9	2
10	1

Supplemental Figure S10: (A): Average correlation between markers in each 20 bp bin. The red line is the line-of-best-fit assuming exponential decay. (B): The number of markers in each region. No regions had more than 10 markers.

Assumption 4 is equivalent to requiring that the errors ϵ_{ng} in the regression model

$$y_{ng} = \boldsymbol{x}_g^\top \boldsymbol{\beta}_n + \epsilon_{ng}$$

be independent across samples $n \in [N]$ and CpGs $g \in [G]$. Independence across samples will hold if bulk samples correspond to unrelated subjects (which is common), while independence across CpGs amounts to assuming that DNAm levels are independent across CpGs conditional on cell type proportion. The latter is likely very close to true in practice, since cell type proportion accounts for most of the covariation between CpGs [14, 15], and it is a standard assumption of DNAm-based deconvolution methods [16, 17, 18].

We note that all of our results can easily be generalized when we relax Assumption 4 by replacing it with the requirement that CpGs cluster into regions satisfying (i) DNAm levels y_{ng} for CpGs g in different regions are independent and (ii) the number of CpGs in any region is small (i.e., bounded above by a constant that does not depend on the sample size N or the number of markers G). Under (i) and (ii), the rates in all of our proofs are unchanged. We prove Theorem 1 using Assumption 4 to make arguments easier to follow.

To assess whether the above relaxed conditions are reasonable, we estimated the spatial correlation between CpGs (conditional on cell type proportion) in the real data example from Section 3.6 [3]. Figure S10A shows the average pairwise correlation between DNAm levels at marker CpGs separated by base-pair distances from 2 to 3,000 bp (in 20 bp bins). Consistent with previous reports [19], the correlation between markers more than 2,200 bp apart is negligible (i.e., < 0.05, as determine by the red line of best fit in Figure S10A). We therefore grouped markers within 2,200 bp of one another into regions; because markers in distinct regions are more than 2,200 bp apart by construction, it is reasonable to treat their DNAm levels as independent, satisfying (i). Figure S10B reports the number of markers per region: most regions (86%) contain a single marker and nearly all (99.6%) contain five or fewer, so (ii) is also reasonable.

S12 Proof of Theorem 1

S12.1 Notation

In all proofs, we let $c > 1$ denote a generic constant, sufficiently large and not depending on N or G, whose value may change between occurrences. We use the standard "O_P" and "o_P" notation throughout [20], i.e., a sequence of random variables $X_n = O_P(a_n)$ if for all $\epsilon > 0$, there exist constants $C_\epsilon > 0$ and $n_\epsilon > 0$ so that $\Pr(|X_n| \geq C_\epsilon a_n) \leq \epsilon$ for all $n \geq n_\epsilon$; and $X_n = o_P(a_n)$ if for all $\epsilon > 0$ there exists a constant $n_\epsilon > 0$ so that $\Pr(|X_n| \geq \epsilon a_n) \leq \epsilon$ for all $n \geq n_\epsilon$. We define $\|\boldsymbol{V}\|_F$ to be the Frobenius norm of the matrix $\boldsymbol{V}$ and $\|\boldsymbol{v}\|_2$ to be the Euclidean norm if $\boldsymbol{v}$ is a vector and the matrix 2-norm if it is a matrix. A sequence of vectors or matrices $\boldsymbol{v}_n$ is $O_P(a_n)$ or $o_P(a_n)$ if $\|\boldsymbol{v}_n\|_2 = O_P(a_n)$ or $\|\boldsymbol{v}_n\|_2 = o_P(a_n)$. If $\boldsymbol{V} \in \mathbb{R}^{n\times m}$, we let $\boldsymbol{V}_{i*} \in \mathbb{R}^m$, $\boldsymbol{V}_{*j} \in \mathbb{R}^n$, and $\boldsymbol{V}_{ij} \in \mathbb{R}$ be its i-th row, j-th column, and (i,j)-th element. We define the sub-Gaussian and sub-exponential norms of a random variable v, $\|v\|_{\psi_2}$ and $\|v\|_{\psi_1}$, to be as they are defined in Definitions 5.7 and 5.13 of Eldar et al. [21]:

$$\|v\|_{\psi_2} = \sup_{p\geq 1} p^{-1/2}\{\mathbb{E}(|v|^p)\}^{1/p}, \quad \|v\|_{\psi_1} = \sup_{p\geq 1} p^{-1}\{\mathbb{E}(|v|^p)\}^{1/p}.$$

We also define $\|\boldsymbol{v}\|_{\psi_2}$ and $\|\boldsymbol{v}\|_{\psi_1}$ for a random vector $\boldsymbol{v}$ to be as defined in Definition 5.22 of Eldar et al. [21]:

$$\|\boldsymbol{v}\|_{\psi_2} = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \|\boldsymbol{x}^\top \boldsymbol{v}\|_{\psi_2}, \quad \|\boldsymbol{v}\|_{\psi_1} = \sup_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \|\boldsymbol{x}^\top \boldsymbol{v}\|_{\psi_1}.$$

Throughout the proof we define $\hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$ and $\bar{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$to be

$$\hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)} = \arg\min_{\boldsymbol{\beta}\geq 0} \sum_{h\in[G]\setminus\{g\}} (y_{nh} - \boldsymbol{x}_h^\top \boldsymbol{\beta})^2, \quad \bar{\boldsymbol{\beta}}^{(1)}_{n,(-g)} = \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}\}^{-1} \hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$$

which is the first-step estimator that excludes the g-th CpG. We first prove a series of results, which are utilized in the proof of Theorem 1 in Section S12.5.

S12.2 A road-map of the proof

The proof of Theorem 1 is broken up into three parts. The first (Section S12.3) studies the behavior of the first-step estimators $\hat{\boldsymbol{\beta}}^{(1)}_n$, $\bar{\boldsymbol{\beta}}^{(1)}_n$, $\hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$, and $\bar{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$. The results are summarized as follows:

- Lemma 1: derives an analytic expression for the the non-zero entries of $\hat{\boldsymbol{\beta}}^{(1)}_n$ in terms of its active set, i.e., the indices of $\hat{\boldsymbol{\beta}}^{(1)}_n$ corresponding to its non-zero entries.
- Lemma 2: proves the true set of indices of $\boldsymbol{\beta}_n$ with non-zero entries are contained within the active set for $\hat{\boldsymbol{\beta}}^{(1)}_n$ with high probability.
- Lemma 3: Proves $\hat{\boldsymbol{\beta}}^{(1)}_n - \boldsymbol{\beta}_n$ has sub-Gaussian-like properties.
- Corollary 1: uses the previous lemmas to derive the rate of convergence for $\hat{\boldsymbol{\beta}}^{(1)}_n$.

- Corollary 2: uses the previous results to derive approximate analytic expressions for $\bar{\boldsymbol{\beta}}_n^{(1)}$ and $\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ in terms of $\hat{\boldsymbol{\beta}}_n^{(1)}$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$.

- Lemma 4: proves that the active sets for $\hat{\boldsymbol{\beta}}_n^{(1)}$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ are the same with high probability.

The next part (Section S12.4) considers the behavior of $\hat{\phi}_g$:

- Lemma 5: Proves that the log-likelihood functions for the precision parameter when the set of means μ_{ng} are known and estimated using the first-step cellular fraction estimates are close.

- Corollary 3: Uses the previous lemma to show that the set of precision parameter estimates converge uniformly to their corresponding true values.

- Lemma 6: this proves useful properties about the first and second derivative of the log-likelihood functions.

- Lemma 7: derives an expression for the derivative of the log-likelihood function when μ_{ng} is estimated in terms of the derivative when μ_{ng} is known.

- Lemma 8: uses the previous lemmas to derive an approximate analytic expression for $\hat{\phi}_g$.

The final part (Section S12.5) uses all of the previous results—especially Lemmas 2, 6, and 8 and Corollaries 1 and 2—to prove Theorem 1. Section S12.6 contains additional technical results used to prove the above results.

S12.3 Behavior of the first-step estimates for cellular fractions

We first consider the behavior of the estimators $\hat{\boldsymbol{\beta}}_n^{(1)}$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$. To do so, we first state a useful result regarding non-negative least squares estimators.

Lemma 1. *Let $\mathcal{A}_n \subseteq [K]$ be the set of non-zero entries of $\hat{\boldsymbol{\beta}}_n^{(1)}$, define $\boldsymbol{X}_{\mathcal{A}}$ to be the sub-matrix of $\boldsymbol{X}$ containing only the columns in $\mathcal{A}$, and let $\boldsymbol{y}_n = (y_{n1}, \ldots, y_{nG})^\top$. Then the non-zero entries of $\hat{\boldsymbol{\beta}}_n^{(1)}$ can be expressed as*

$$(\boldsymbol{X}_{\mathcal{A}_n}^\top \boldsymbol{X}_{\mathcal{A}_n})^{-1} \boldsymbol{X}_{\mathcal{A}_n}^\top \boldsymbol{y}_n.$$

Proof. Let $\mathcal{A}_{\boldsymbol{\beta}} \subseteq [K]$ be the set of non-zero indices of $\boldsymbol{\beta}$, and let $\mathcal{I}_{\boldsymbol{\beta}} = [K] \setminus \mathcal{A}_{\boldsymbol{\beta}}$. Then the gradient of the least-squares objective function—up to multiplicative constants that do not depend on $\boldsymbol{\beta}$ and a re-ordering of the columns of $\boldsymbol{X}$—is

$$-\begin{pmatrix} \boldsymbol{X}_{\mathcal{A}_{\boldsymbol{\beta}}}^\top \\ \boldsymbol{X}_{\mathcal{I}_{\boldsymbol{\beta}}}^\top \end{pmatrix} (\boldsymbol{y}_n - \boldsymbol{X}_{\mathcal{A}_{\boldsymbol{\beta}}} \boldsymbol{\beta}).$$

Since $\hat{\boldsymbol{\beta}}_n^{(1)}$ is the minimizer, the KKT conditions imply

$$\boldsymbol{X}_{\mathcal{A}_n}^\top \{\boldsymbol{y}_n - \boldsymbol{X}_{\mathcal{A}_n} \hat{\boldsymbol{\beta}}_n^{(1)}\} = \boldsymbol{0}.$$

Solving for $\hat{\boldsymbol{\beta}}_n^{(1)}$ gives the desired result. □

Lemma 2. *Let $\mathcal{A}_n = \{k \in [K] : \hat{\beta}^{(1)}_{nk} > 0\}$ be the set of indices corresponding to the non-zero elements of $\hat{\boldsymbol{\beta}}^{(1)}_n$, and define $\mathcal{A}_{n,(-g)}$ to be the analogue for $\hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$. Then*

$$\Pr([J_n] \subseteq \mathcal{A}_n \text{ and } [J_n] \subseteq \mathcal{A}_{n,(-g)} \text{ for all } n \in [N] \text{ and } g \in [G]) \geq 1 - \exp\{c\log(G) - c^{-1}G\}$$

for all N, G sufficiently large.

Remark 1. The sets of non-zero elements of $\hat{\boldsymbol{\beta}}^{(1)}_n$ and $\bar{\boldsymbol{\beta}}^{(1)}_n$ are the same. The sets of non-zero elements of $\hat{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$ and $\bar{\boldsymbol{\beta}}^{(1)}_{n,(-g)}$ are also the same.

Proof. Let $E_1 = \{[J_n] \subseteq \mathcal{A}_n, \mathcal{A}_{n,(-g)} \text{ for all } n : J_n = K \text{ and } g \in [G]\}$ and $E_2 = \{[J_n] \subseteq \mathcal{A}_n, \mathcal{A}_{n,(-g)} \text{ for all } n : J_n < K \text{ and } g \in [G]\}$. To prove the result, note

$$\Pr([J_n] \subseteq \mathcal{A}_n, \mathcal{A}_{n,(-g)} \text{ for all } n \in [N], g \in [G]) \geq 1 - \Pr(E_1^{\mathrm{c}}) - \Pr(E_2^{\mathrm{c}}),$$

where E_1^{c} and E_2^{c} are the complements of E_1 and E_2. We therefore derive lower bounds for $\Pr(E_1^{\mathrm{c}})$ and $\Pr(E_2^{\mathrm{c}})$.

For the first, let $\hat{\boldsymbol{\beta}}^{(OLS)}_n = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}_n$ and $\hat{\boldsymbol{\beta}}^{(OLS)}_{n,(-g)}$ be the ordinary least squares estimators for $\boldsymbol{\beta}_n$, which satisfy

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(OLS)}_n &= \boldsymbol{\beta}_n + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n \\
\hat{\boldsymbol{\beta}}^{(OLS)}_{n,(-g)} &= \boldsymbol{\beta}_n + \{\boldsymbol{X}^\top_{(-g)}\boldsymbol{X}_{(-g)}\}^{-1}\boldsymbol{X}^\top_{(-g)}\boldsymbol{\epsilon}_{n,(-g)}.
\end{aligned}$$

Here, $\boldsymbol{\epsilon}_n = \boldsymbol{y}_n - \boldsymbol{X}\boldsymbol{\beta}_n$ and $\boldsymbol{\epsilon}_{n,(-g)} = \boldsymbol{y}_{n,(-g)} - \boldsymbol{X}_{(-g)}\boldsymbol{\beta}_n$. Note the entries of $\boldsymbol{\epsilon}_n$ are mean-zero, independent, and sub-Gaussian with uniformly sub-Gaussian norm [21]. As such, $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n$ satisfies

$$\begin{aligned}
&\mathbb{E}\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n\} = \boldsymbol{0} \\
&\mathbb{E}[\exp\{\boldsymbol{t}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n\}] \leq \exp(c\boldsymbol{t}^\top\boldsymbol{t}/G).
\end{aligned}$$

The same result holds for $\boldsymbol{\epsilon}_{n,(-g)}$. This implies

$$\begin{aligned}
\Pr(E_1^{\mathrm{c}}) \leq &\sum_{n:J_n=K} \Pr\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n \not\geq -\boldsymbol{\beta}_n\} \\
&+ \sum_{n:J_n=K}\sum_{g=1}^{G} \Pr[\{\boldsymbol{X}^\top_{(-g)}\boldsymbol{X}_{(-g)}\}^{-1}\boldsymbol{X}^\top_{(-g)}\boldsymbol{\epsilon}_{n,(-g)} \not\geq -\boldsymbol{\beta}_n] \leq \exp\{c\log(G) - c^{-1}G\},
\end{aligned}$$

where the first inequality follows by a union bound and the second by Lemma 5.5 of Eldar et al. [21].

To bound $\Pr(E_2^{\mathrm{c}})$, let $\boldsymbol{X}_{\mathcal{S}}$ be the sub-matrix of $\boldsymbol{X}$ restricted to columns in the set $\mathcal{S} \subseteq [K]$ and, for any matrix $\boldsymbol{Z}$, define $\boldsymbol{P}_{\boldsymbol{Z}}$ and $\boldsymbol{P}^{\perp}_{\boldsymbol{Z}}$ to be the orthogonal projection matrices that project vectors onto the image of $\boldsymbol{Z}$ and the kernel of $\boldsymbol{Z}^\top$, respectively. Let $a_n \in \{\text{all subsets of } [K] \setminus [J_n]\}$. Then Lemma 1 and the KKT conditions imply $\mathcal{A}_n = [J_n] \cup a_n$ if and only if the following hold:

$$\begin{aligned}
&\boldsymbol{e}_k^\top(G^{-1}\boldsymbol{X}^\top_{[J_n]\cup a_n}\boldsymbol{X}_{[J_n]\cup a_n})^{-1}G^{-1/2}\boldsymbol{X}^\top_{[J_n]\cup a_n}\boldsymbol{\epsilon}_n \geq -G^{1/2}\beta_{nk}, \quad k \in [J_n] \\
&\left(\tilde{\boldsymbol{X}}^\top_{a_n}\tilde{\boldsymbol{X}}_{a_n}\right)^{-1}\tilde{\boldsymbol{X}}^\top_{a_n}\boldsymbol{\epsilon}_n \geq \boldsymbol{0}, \quad \tilde{\boldsymbol{X}} = \boldsymbol{P}^{\perp}_{\boldsymbol{X}_{[J_n]}}\boldsymbol{X} \\
&\tilde{\boldsymbol{X}}^\top_{i_n}\boldsymbol{P}^{\perp}_{\tilde{\boldsymbol{X}}_{a_n}}\boldsymbol{\epsilon}_n < \boldsymbol{0}, \quad i_n = [K] \setminus \{[J_n] \cup a_n\},
\end{aligned}$$

where $\boldsymbol{e}_k$ is the k-th standard basis vector. Note $\tilde{\boldsymbol{X}}$ implicitly depends on n. As such, then the event $\{\mathcal{A}_n = [J_n] \cup a_n\}$ equals the event

$$\begin{aligned}
&\left\{\boldsymbol{e}_k^\top (G^{-1}\boldsymbol{X}_{[J_n]\cup a_n}^\top \boldsymbol{X}_{[J_n]\cup a_n})^{-1} G^{-1/2} \boldsymbol{X}_{[J_n]\cup a_n}^\top \boldsymbol{\epsilon}_n \geq -G^{1/2}\beta_{nk} \text{ for all } k \in [J_n]\right\} \\
\cap &\left\{a_n \text{ is the active set for the non-negative least squares problem} \right. \\
&\left. \arg\min_{\boldsymbol{\gamma}\geq \boldsymbol{0}} \|\boldsymbol{\epsilon}_n - \tilde{\boldsymbol{X}}_{[K]\setminus[J_n]}\boldsymbol{\gamma}\|_2 \right\} = E_{2,n}^{(1)}(a_n) \cap E_{2,n}^{(1)}(a_n).
\end{aligned}$$

Then

$$\begin{aligned}
\Pr([J_n] \subseteq \mathcal{A}_n) &= \sum_{a_n} \Pr\left\{E_{2,n}^{(1)}(a_n) \cap E_{2,n}^{(1)}(a_n)\right\} \geq \sum_{a_n} 1 - \Pr\left[\left\{E_{2,n}^{(1)}(a_n)\right\}^{\mathrm{c}}\right] - \Pr\left[\left\{E_{2,n}^{(2)}(a_n)\right\}^{\mathrm{c}}\right] \\
&\geq -\exp(-c^{-1}G) + \sum_{a_n} \Pr\left\{E_{2,n}^{(2)}(a_n)\right\} = 1 - \exp(-c^{-1}G),
\end{aligned}$$

where the second inequality follows by Lemma 5.5 of Eldar et al. [21] and the last equality because

$$\sum_{a_n} \Pr\left\{E_{2,n}^{(2)}(a_n)\right\} = \Pr\{\text{There is a solution to the problem } \arg\min_{\boldsymbol{\gamma}\geq \boldsymbol{0}} \|\boldsymbol{\epsilon}_n - \tilde{\boldsymbol{X}}_{[K]\setminus[J_n]}\boldsymbol{\gamma}\|_2\} = 1.$$

An identical analysis can be used to show that $\Pr\{[J_n] \subseteq \mathcal{A}_{n,(-g)}\} \geq 1 - \exp(-c^{-1}G)$. Then

$$\Pr(E_2^{\mathrm{c}}) \leq \sum_{n=1}^N \Pr([J_n] \nsubseteq \mathcal{A}_n) + \sum_{n=1}^N \sum_{g=1}^G \Pr\{[J_n] \nsubseteq \mathcal{A}_{n,(-g)}\} \leq \exp\{c\log(G) - c^{-1}G\},$$

which completes the proof. □

Lemma 3. *Let E be the event $\{[J_n] \subseteq \mathcal{A}_n, \mathcal{A}_{n,(-g)}$ for all $n \in [N], g \in [G]\}$. Then on the event E, the random vectors $\boldsymbol{Z}_n = G^{1/2}\{\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\}$ and $\boldsymbol{Z}_{n,(-g)} = G^{1/2}\{\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n\}$ are sub-Gaussian with uniformly bounded sub-Gaussian norm. That is,*

$$\begin{aligned}
\sup_{\substack{p\geq 1,\, \boldsymbol{v}\in\mathbb{R}^K \\ \|\boldsymbol{v}\|_2 = 1}} p^{-1/2} \left\{\mathbb{E}(|\boldsymbol{v}^\top \boldsymbol{Z}_n|^p \mid E)\right\}^{1/p} &\leq c \\
\sup_{\substack{p\geq 1,\, \boldsymbol{v}\in\mathbb{R}^K \\ \|\boldsymbol{v}\|_2 = 1}} p^{-1/2} \left[\mathbb{E}\{|\boldsymbol{v}^\top \boldsymbol{Z}_{n,(-g)}|^p \mid E\}\right]^{1/p} &\leq c
\end{aligned}$$

for all $n \in [N]$ and $g \in [G]$.

Proof. We prove the result for $\boldsymbol{Z}_n$. The proof of the corresponding result for $\boldsymbol{Z}_{n,(-g)}$ is identical. On E, the non-zero elements of $\boldsymbol{Z}_n$ can be expressed as

$$(G^{-1}\boldsymbol{X}_{\mathcal{A}_n}^\top \boldsymbol{X}_{\mathcal{A}_n})^{-1}(G^{-1/2}\boldsymbol{X}_{\mathcal{A}_n}^\top \boldsymbol{\epsilon}_n),$$

where $\boldsymbol{X}_{\mathcal{A}_n}$ is the sub-matrix of $\boldsymbol{X}$ restricted to the columns whose indices are in $\mathcal{A}_n$. Let $a_n \in \{\text{all subsets of } [K] \setminus [J_n]\}$ and define $\boldsymbol{X}_{a_n}$ to be the sub-matrix of $\boldsymbol{X}$ formed by keeping only columns in $[J_n] \cup a_n$. Then for any $p \geq 1$ and unit vector $\boldsymbol{v} \in \mathbb{R}^K$,

$$\begin{aligned}\{\mathbb{E}(|\boldsymbol{v}^\top \boldsymbol{Z}_n|^p \mid E)\}^{1/p} \leq & c\{\mathbb{E}(|\boldsymbol{v}^\top \boldsymbol{Z}_n|^p 1\{E\})\}^{1/p} \\ \leq & c \sum_{a_n} [\mathbb{E}\{1\{\mathcal{A}_n = [J_n] \cup a_n\} |\boldsymbol{v}^\top \boldsymbol{M}_{a_n} (G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} (G^{-1/2} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n)|^p\}]^{1/p} \\ \leq & c p^{1/2} \sum_{a_n} \|G^{-1/2} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\|_{\psi_2}\end{aligned}$$

where $\boldsymbol{M}_{a_n} \in \{0,1\}^{K \times (J_n + |a_n|)}$. The first inequality follows by Lemma 2, the second by a union bound and the triangle inequality applied to norms of random variables, and the third because the eigenvalues of $G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n}$ are bounded above zero by Assumption 1. Since $\|G^{-1/2} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\|_{\psi_2} \leq c$ by Lemma 5.9 of Eldar et al. [21], the result follows by Lemma 5.5 of Eldar et al. [21]. □

Corollary 1. *The estimates $\hat{\boldsymbol{\beta}}_n^{(1)}$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ satisfy*

$$\begin{aligned}&\|\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\|_2,\ \|\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n\|_2 = O_P(G^{-1/2}) \\ &\max_{n \in [N]} \|\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\|_2,\ \max_{\substack{n \in [N] \\ g \in [G]}} \|\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n\|_2 = O_P[\{\log(N)/G\}^{1/2}].\end{aligned}$$

Proof. We prove the result for $\hat{\boldsymbol{\beta}}_n^{(1)}$, as the proof for $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ is identical. The first is an immediate consequence of Lemma 3 above and Lemma 5.5 of Eldar et al. [21]. For the second, let $\boldsymbol{Z}_n$ be as defined in the statement of Lemma 3. Then for any $\lambda > 0$,

$$\begin{aligned}\Pr\left(\max_{n \in [N]} \|\boldsymbol{Z}_n\|_2 \geq t \mid E\right) \leq & c \Pr\left[\{\max_{n \in [N]} \|\boldsymbol{Z}_n\|_2 \geq t\} \cap E\right] \\ \leq & c \Pr\left[\sum_{n=1}^N \sum_{a_n} \exp\left\{\lambda \|(G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} G^{-1/2} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\|_2\right\} \geq \exp(\lambda t)\right].\end{aligned}$$

The result then follows by Lemma 5.5. of Eldar et al. [21] and an identical proof used to prove Lemma 10 in Section S12.6. □

Corollary 2. *Let $\hat{\boldsymbol{\delta}}_n^{(1)} = \hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n$ and $\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} = \hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n$. Then the estimates $\bar{\boldsymbol{\beta}}_n^{(1)}$ and $\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ satisfy*

$$\begin{aligned}&\|\bar{\boldsymbol{\beta}}_n^{(1)} - \{\hat{\boldsymbol{\beta}}_n^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)} \boldsymbol{\beta}_n\}\|_2,\ \|\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \{\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} \boldsymbol{\beta}_n\}\|_2 = O_P(G^{-1}) \\ &\max_{n \in [N]} \|\bar{\boldsymbol{\beta}}_n^{(1)} - \{\hat{\boldsymbol{\beta}}_n^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)} \boldsymbol{\beta}_n\}\|_2,\ \max_{\substack{n \in [N] \\ g \in [G]}} \|\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \{\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} \boldsymbol{\beta}_n\}\|_2 = O_P\{\log(N) G^{-1}\}.\end{aligned}$$

Remark 2. An immediate consequence of Corollaries 1 and 2 is that Corollary 1 holds with $\hat{\boldsymbol{\beta}}_n^{(1)}$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ replaced with $\bar{\boldsymbol{\beta}}_n^{(1)}$ and $\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$.

Proof. We prove the results for $\bar{\boldsymbol{\beta}}_n^{(1)}$. The proof for $\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ is identical. Recall $\bar{\boldsymbol{\beta}}_n^{(1)} = \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(1)}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(1)}$. By Taylor's Theorem,

$$\{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(1)}\}^{-1} = 1 - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)} + \frac{\{\mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)}\}^2}{(1+\tilde{\delta}_n)^3},$$

where $\tilde{\delta}_n$ lies between 0 and $\mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)}$. Therefore,

$$\|\bar{\boldsymbol{\beta}}_n^{(1)} - \{\hat{\boldsymbol{\beta}}_n^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)}\boldsymbol{\beta}_n\}\|_2 \leq K\|\hat{\boldsymbol{\delta}}_n^{(1)}\|_2^2 \frac{\|\hat{\boldsymbol{\beta}}_n^{(1)}\|_2}{(1+\tilde{\delta}_n)^3} + K^{1/2}\|\hat{\boldsymbol{\delta}}_n^{(1)}\|_2^2.$$

The result then follows by Corollary 1. □

Lemma 4. *Let $\mathcal{A}_n$ and $\mathcal{A}_{n,(-g)}$ be as defined in the statement of Lemma 2. Then*

$$\max_{n\in[N],g\in[G]} \Pr\{\mathcal{A}_{n,(-g)} \neq \mathcal{A}_n\} \leq c\{\log(G)/G\}^{1/2}.$$

Proof. We first consider samples n satisfying $J_n = K$. The ordinary least squares estimators defined in the proof of Lemma 2 satisfy

$$\hat{\boldsymbol{\beta}}_n^{(OLS)} = \boldsymbol{\beta}_n + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n$$
$$\hat{\boldsymbol{\beta}}_{n,(-g)}^{(OLS)} = \boldsymbol{\beta}_n + (\boldsymbol{X}_{(-g)}^\top\boldsymbol{X}_{(-g)})^{-1}\boldsymbol{X}_{(-g)}^\top\boldsymbol{\epsilon}_{n,(-g)}.$$

The error terms $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n$ and $(\boldsymbol{X}_{(-g)}^\top\boldsymbol{X}_{(-g)})^{-1}\boldsymbol{X}_{(-g)}^\top\boldsymbol{\epsilon}_{n,(-g)}$ are mean-zero sub-Gaussian random vectors with sub-Gaussian norms uniformly bounded above by $cG^{-1/2}$. We see that

$$\Pr(\mathcal{A}_{n,(-g)} = \mathcal{A}_n) \geq \Pr(\mathcal{A}_{n,(-g)}, \mathcal{A}_n = [K]),$$

which implies

$$\Pr(\mathcal{A}_{n,(-g)} \neq \mathcal{A}_n) \leq \Pr(\mathcal{A}_{n,(-g)} \neq [K]) + \Pr(\mathcal{A}_n \neq [K]),$$

where

$$\Pr(\mathcal{A}_n \neq [K]) \leq \sum_{k=1}^{K} \Pr\{\boldsymbol{e}_k^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n < -\beta_{nk}\} \leq c\exp(-cG).$$

The vector $\boldsymbol{e}_k \in \{0,1\}^K$ is the k-th standard basis vector. The last equality follows by the properties of the above error terms. An identical result holds for $\Pr(\mathcal{A}_{n,(-g)} \neq [K])$. We therefore only need to consider n satisfying $J_n < K$ to complete the proof.

When $J_n < K$,

$$\begin{aligned}
\Pr\{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\} &\geq \Pr[\{[J_n] \subseteq \mathcal{A}_{n,(-g)}, \mathcal{A}_n\} \cap \{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\}] \\
&= \Pr[\{[J_n] \subseteq \mathcal{A}_{n,(-g)}, \mathcal{A}_n\} \cap \{\mathcal{A}_{n,(-g)} \setminus [J_n] = \mathcal{A}_n \setminus [J_n]\}] \\
&= \sum_{a_n} \Pr[\{[J_n] \subseteq \mathcal{A}_{n,(-g)}, \mathcal{A}_n\} \cap \{\mathcal{A}_{n,(-g)} \setminus [J_n] = \mathcal{A}_n \setminus [J_n] = a_n\}],
\end{aligned}$$

where $a_n \in \{\text{all subsets of } [K]\setminus[J_n]\}$. For fixed a_n, the KKT conditions imply $\mathcal{A}_n = a_n \cup [J_n]$ will contain the non-zero coefficients if and only if the following hold:

$$\begin{aligned}
w_{nk}(a_n) &= \boldsymbol{e}_k^\top (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n \geq -\beta_{nk}, \quad k \in [J_n] \\
x_{nk}(a_n) &= G^{1/2} \boldsymbol{e}_k^\top (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n \geq 0, \quad k \in a_n \\
z_{nk}(a_n) &= G^{-1/2} \boldsymbol{X}_{*k}^\top \{I_G - \boldsymbol{X}_{a_n} (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top\} \boldsymbol{\epsilon}_n < 0, \quad k \in [K] \setminus \{[J_n] \cup a_n\}.
\end{aligned}$$

Here, we take $\boldsymbol{X}_{a_n}$ to be the sub-matrix of $\boldsymbol{X}$ formed by keeping only columns in $[J_n] \cup a_n$, i.e., the columns corresponding to the non-zero entries of $\hat{\boldsymbol{\beta}}_n^{(1)}$. The vector $\boldsymbol{e}_k \in \{0,1\}^{|[J_n]\cup a_n|}$ is the k-th standard basis vector. Equivalent conditions hold when we leave out the g-th CpG. This implies $\Pr\{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\}$ can be lower-bounded as

$$\begin{aligned}
\Pr\{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\} \geq & \sum_{a_n} \Pr[\{w_{nk}(a_n), w_{nk,(-g)}(a_n) \geq -\beta_{nk} \text{ for } k \in [J_n]\} \\
& \cap \{x_{nk}(a_n), x_{nk,(-g)}(a_n) \geq 0 \text{ for all } k \in a_n \\
& \cap \{z_{nk}(a_n), z_{nk,(-g)}(a_n) < 0 \text{ for all } k \in [K] \setminus \{[J_n] \cup a_n\}\}] \\
\geq & \sum_{a_n} \Pr[\{x_{nk}(a_n), x_{nk,(-g)}(a_n) \geq 0 \text{ for all } k \in a_n\} \\
& \cap \{z_{nk}(a_n), z_{nk,(-g)}(a_n) < 0 \text{ for all } k \in [K] \setminus \{[J_n] \cup a_n\}\}] \\
& - c \exp(-cG),
\end{aligned} \tag{S2}$$

where the last inequality follows because $w_{nk}(a_n)$ and $w_{nk,(-g)}(a_n)$ are sub-Gaussian with sub-Gaussian norms bounded above by $cG^{-1/2}$.

Define

$$\begin{aligned}
\Delta_{nk,g}^{(x)}(a_n) :=& x_{nk}(a_n) - x_{nk,(-g)}(a_n) = G^{-1/2}(1 - h_{g,a_n})^{-1} \boldsymbol{e}_k^\top (G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{g,a_n} \epsilon_{gn} \\
& - G^{-1}(1 - h_{g,a_n})^{-1} \boldsymbol{e}_k^\top (G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{g,a_n} \boldsymbol{X}_{g,a_n}^\top \{G^{-1/2} (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\} \\
\Delta_{nk,g}^{(z)}(a_n) :=& z_{nk}(a_n) - z_{nk,(-g)}(a_n) = G^{-1/2} \epsilon_{gn} \{X_{gk} \\
& - (1 - h_{g,a_n})^{-1} (G^{-1} \boldsymbol{X}_{*k}^\top \boldsymbol{X}_{a_n})(G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{g,a_n} \\
& + (1 - h_{g,a_n})^{-1} X_{gk} h_{g,a_n}^2\} \\
& + G^{-1} \boldsymbol{X}_{g,a_n}^\top \{G^{1/2} (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\} \{(G^{-1} \boldsymbol{X}_{*k}^\top \boldsymbol{X}_{a_n})(G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{g,a_n} \\
& - (1 - h_{g,a_n})^{-1} X_{gk} h_{g,a_n}\},
\end{aligned}$$

where h_{g,a_n} is the g-th leverage score for the design matrix $\boldsymbol{X}_{a_n}$ and $\boldsymbol{X}_{g,a_n}$ is the g-th row of $\boldsymbol{X}_{a_n}$ treated as a column vector. Both of the above quantities are sub-Gaussian with sub-Gaussian norm bounded above by $cG^{-1/2}$. Then ignoring the $c\exp(-cG)$ term (since it is $o(G^{-1/2})$) the lower bound in (S2) can be expressed as

$$\begin{aligned}
\Pr\{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\} \geq & \sum_{a_n} \Pr[\{x_{nk}(a_n) \geq |\Delta_{nk,g}^{(x)}(a_n)| \text{ for all } k \in a_n\} \\
& \cap \{z_{nk}(a_n) < -|\Delta_{nk,g}^{(z)}(a_n)| \text{ for all } k \in [K] \setminus \{[J_n] \cup a_n\}\}].
\end{aligned}$$

Further, the probabilities within the summand are bounded above by

$$\begin{aligned}
&\Pr[\{x_{nk}(a_n) \geq 0 \text{ for all } k \in a_n\} \cap \{z_{nk}(a_n) < 0 \text{ for all } k \in [K] \setminus \{[J_n] \cup a_n\}\}] \\
&- 2\Pr\{0 \leq x_{nk}(a_n) < |\Delta^{(x)}_{nk,g}(a_n)| \text{ for at least one } k \in a_n\} \\
&- \Pr[-|\Delta^{(z)}_{nk,g}(a_n)| \leq z_{nk}(a_n) < 0 \text{ for at least one } k \in [K] \setminus \{[J_n] \cup a_n\}].
\end{aligned} \tag{S3}$$

Here, $x_{nk}(a_n)$ and $z_{nk}(a_n)$ have sub-Gaussian norm uniformly bounded above by c. As such, an application of the Berry–Esseen theorem to $x_{nk}(a_n)$ and $z_{nk}(a_n)$ implies that both are asymptotically normally distributed with mean zero and covariance matrices whose eigenvalues are bounded within the compact interval $[c^{-1}, c]$. Moreover, the absolute difference between their cumulative distribution functions (CDFs) and the corresponding normal CDFs is bounded above by $cG^{-1/2}$. As such, for any $c > 1$,

$$\begin{aligned}
&\Pr\{0 \leq x_{nk}(a_n) < |\Delta^{(x)}_{nk,g}(a_n)| \text{ for at least one } k \in a_n\} \\
&\leq \sum_{k \in a_n} \underbrace{\Pr\{0 \leq x_{nk}(a_n) < c\log^{1/2}(G)G^{-1/2}\}}_{\leq c\log^{1/2}(G)G^{-1/2} \text{ by Berry–Esseen theorem}} \\
&+ \sum_{k \in a_n} \underbrace{\Pr\{|\Delta^{(x)}_{nk,g}(a_n)| \geq c\log^{1/2}(G)G^{-1/2}\}}_{\leq cG^{-c^2}},
\end{aligned}$$

where the last inequality follows because $\Delta^{(x)}_{nk,g}(a_n)$ is sub-Gaussian with sub-Gaussian norm bounded above by $cG^{-1/2}$. An identical result holds for the third quantity in (S3). Putting all this together give us

$$\begin{aligned}
\Pr\{\mathcal{A}_{n,(-g)} = \mathcal{A}_n\} \geq &\sum_{a_n} \Pr[\{x_{nk}(a_n) \geq 0 \text{ for all } k \in a_n\} \\
&\cap \{z_{nk}(a_n) < 0 \text{ for all } k \in [K] \setminus \{[J_n] \cup a_n\}\}] - cG^{-1/2}\log^{1/2}(G) \\
\geq &P\{[J_n] \subseteq \mathcal{A}_n\} - cG^{-1/2}\log^{1/2}(G) \\
\geq &1 - c\exp(-cG) - cG^{-1/2}\log^{1/2}(G),
\end{aligned}$$

which proves the claim. □

S12.4 Estimate for the precision parameter

After determining the first-step estimator, we set $\hat{\mu}_{ng} = \boldsymbol{x}_g^\top \bar{\boldsymbol{\beta}}_n^{(1)}$ and estimate the precision parameters ϕ_g to be

$$\hat{\phi}_g = \underset{\phi \in [c^{-1}, c]}{\arg\max}\, \hat{\ell}_g(\phi), \quad \hat{\ell}_g(\phi) = \sum_{n=1}^{N} \log \text{Beta}\{y_{ng}; \phi\hat{\mu}_{ng}, \phi(1 - \hat{\mu}_{ng})\} \tag{S4}$$

where $\text{Beta}(y; \alpha, \beta)$ is the likelihood of a beta distribution with parameters α, β evaluated at y. We first derive properties of $\hat{\ell}_g$.

Lemma 5. *Let $\ell_g(\phi)$ be the log-likelihood where $\hat{\mu}_{ng}$ in* (S4) *is replaced with the true μ_{ng}. Then*

$$\max_{g\in[G];\phi\in[c^{-1},c]}|N^{-1}\hat{\ell}_g(\phi)-N^{-1}\ell_g(\phi)|=O_P\{\log^c(G)G^{-1/2}\}.$$

Proof. The difference in log-likelihoods is

$$\begin{aligned}\hat{\ell}_g(\phi)-\ell_g(\phi)=&\sum_n\log\{\Gamma(\phi\mu_{gn})\}-\log\{\Gamma(\phi\hat{\mu}_{gn})\}\\&+\sum_n\log[\Gamma\{\phi(1-\mu_{gn})\}]-\log[\Gamma\{\phi(1-\hat{\mu}_{gn})\}]\\&+\sum_n\phi(\hat{\mu}_{ng}-\mu_{ng})\log(y_{gn})+\sum_n\phi(\mu_{ng}-\hat{\mu}_{ng})\log(1-y_{gn}).\end{aligned}\tag{S5}$$

We provide bounds for the first and third terms in this expression, where the bounds for the second and fourth are identical. Let $\Psi(x)=\frac{d}{dx}\log\{\Gamma(x)\}$. Then by Taylor's theorem, the first term can be expressed as

$$\sum_n\log\{\Gamma(\phi\mu_{gn})\}-\log\{\Gamma(\phi\hat{\mu}_{gn})\}=\phi\boldsymbol{X}_{g*}^\top\sum_n\Psi(\phi\tilde{\mu}_{gn})\{\boldsymbol{\beta}_n-\bar{\boldsymbol{\beta}}_n^{(1)}\},$$

where $\tilde{\mu}_{gn}$ lies between $\hat{\mu}_{gn}$ and μ_{gn}. Assumption 1 and Corollaries 1 and 2 imply

$$\max_{g\in[G];n\in[N]}|\tilde{\mu}_{gn}-\mu_{gn}|\leq\max_{g\in[G];n\in[N]}|\hat{\mu}_{gn}-\mu_{gn}|=O_P[\{\log(N)/G\}^{1/2}],$$

which implies

$$\max_{\substack{g\in[G];n\in[N]\\\phi\in[c^{-1},c]}}|\Psi(\phi\tilde{\mu}_{gn})|\leq c+O_P[\{\log(N)/G\}^{1/2}].$$

A second application of Corollaries 1 and 2 then implies

$$\begin{aligned}\sup_{\substack{g\in[G]\\\phi\in[c^{-1},c]}}\left|\sum_n\log\{\Gamma(\phi\mu_{gn})\}-\log\{\Gamma(\phi\hat{\mu}_{gn})\}\right|\leq&cN\left(\sup_{\substack{g\in[G];n\in[N]\\\phi\in[c^{-1},c]}}|\Psi(\phi\tilde{\mu}_{gn})|\right)\max_{n\in[N]}\|\boldsymbol{\beta}_n-\bar{\boldsymbol{\beta}}_n^{(1)}\|_2\\=&O_P\{NG^{-1/2}\log^c(G)\}.\end{aligned}$$

The third term in (S5) satisfies

$$\sum_n\phi(\hat{\mu}_{ng}-\mu_{ng})\log(y_{gn})=\phi\boldsymbol{X}_{g*}^\top\sum_n\log(y_{gn})\{\boldsymbol{\beta}_n-\bar{\boldsymbol{\beta}}_n^{(1)}\},$$

meaning

$$\begin{aligned}\sup_{\substack{g\in[G]\\\phi\in[c^{-1},c]}}\left|\sum_n\phi(\hat{\mu}_{ng}-\mu_{ng})\log(y_{gn})\right|\leq&cN\left\{\max_{n\in[N]}\|\boldsymbol{\beta}_n-\bar{\boldsymbol{\beta}}_n^{(1)}\|_2\right\}\left\{\max_{n\in[N];g\in[G]}|\log(y_{gn})|\right\}\\=&O_P\{NG^{-1/2}\log^c(G)\},\end{aligned}$$

where the final equality follows by Corollaries 1 and 2 and the fact that $\log(y_{gn})$ is sub-exponential with uniformly bounded sub-exponential norm. This completes the proof. □

Corollary 3. *The estimators $\hat{\phi}_g$ defined in* (S4) *converge uniformly, i.e.,*

$$\max_{g\in[G]}|\hat{\phi}_g - \phi_g| = o_P(1).$$

Proof. Let $\ell_g(\phi)$ be as defined in the statement of Lemma 5. The second derivative of ℓ_g satisfies

$$\begin{aligned}\ell_g''(\phi) =& N\Psi'(\phi) - \sum_n \mu_{ng}^2\Psi'(\phi\mu_{ng}) - \sum_n (1-\mu_{ng})^2\Psi'\{\phi(1-\mu_{ng})\} \\ =& \sum_n \Psi'(\phi) - \mu_{ng}^2\Psi'(\phi\mu_{ng}) - (1-\mu_{ng})^2\Psi'\{\phi(1-\mu_{ng})\} \\ =& -\sum_n \mathrm{Var}_\phi\left[\frac{d}{d\phi}\log \mathrm{Beta}\{y_{ng};\phi\mu_{ng},\phi(1-\mu_{ng})\}\right] \leq -c^{-1}N,\end{aligned}$$

where Var_ϕ is the variance assuming the true parameter is ϕ. The last inequality holds for all $\phi \in [c^{-1}, c]$ and $g \in [G]$ because the minus Fisher Information is bounded above by a constant for all $\phi \in [c^{-1}, c]$ and $g \in [G]$. This implies

$$\frac{d^2}{d\phi^2}\mathbb{E}\{N^{-1}\ell_g(\phi)\} = N^{-1}\mathbb{E}\left\{\frac{d^2}{d\phi^2}\ell_g(\phi)\right\} \leq -c^{-1},$$

i.e., $\mathbb{E}\{N^{-1}\ell_g(\phi)\}$ is concave for all $\phi \in [c^{-1}, c]$ and $g \in [G]$ with second derivatives bounded below by $-c^{-1}$.

Define

$$\begin{aligned}f_g(\phi) =& N^{-1}\ell_g(\phi) - \mathbb{E}\{N^{-1}\ell_g(\phi)\} = N^{-1}\phi\sum_n \mu_{ng}[\log(y_{gn}) - \mathbb{E}\{\log(y_{gn})\}] \\ &+ N^{-1}\phi\sum_n (1-\mu_{ng})[\log(1-y_{gn}) - \mathbb{E}\{\log(1-y_{gn})\}].\end{aligned}$$

Since $\{\log(y_{gn}) - \mathbb{E}\{\log(y_{gn})\}_{n\in[N]}$ are mean-zero, independent, and sub-exponential with uniformly sub-exponential norm, Lemma 11 in Section S12.6 implies

$$\sup_{\substack{g\in[G]\\ \phi\in[c^{-1},c]}} |f_g(\phi)| = O_P\{\log(N)N^{-1/2}\}.$$

This, together with the concavity of $\mathbb{E}\{N^{-1}\ell_g(\phi)\}$ and Lemma 5, proves the result. □

We next derive properties of the second derivative of $\hat{\ell}_g$.

Lemma 6. *Let $\ell_g(\phi)$ be the log-likelihood where $\hat{\mu}_{ng}$ in* (S4) *is replaced with the true μ_{ng}. Then*

$$\max_{g\in[G]}|N^{-1}\ell_g'(\phi) - \mathbb{E}\{N^{-1}\ell_g'(\phi)\}| = O_P\{\log^c(G)N^{-1/2}\}, \quad N^{-1}\ell_g''(\phi) \leq -c^{-1}$$

and

$$\max_{g\in[g]}|N^{-1}\hat{\ell}_g''(\phi) - N^{-1}\ell_g''(\phi)| = O_P\{\log^c(G)G^{-1/2}\}$$

for all $\phi \in [c^{-1}, c]$ and $g \in [G]$.

Proof. The second result follows directly from the proof of Corollary 3. For the third, let $\Psi(\phi) = \frac{d}{d\phi}\log\Gamma(\phi)$. Then differences in second derivatives is

$$\begin{aligned} N^{-1}\hat{\ell}''_g(\phi) - N^{-1}\ell''_g(\phi) =& N^{-1}\sum_n \{\mu_{ng}^2\Psi'(\phi\mu_{ng}) - \hat{\mu}_{ng}^2\Psi'(\phi\hat{\mu}_{ng})\} \\ &+ N^{-1}\sum_n [(1-\mu_{ng})^2\Psi'\{\phi(1-\mu_{ng})\} - (1-\hat{\mu}_{ng})^2\Psi'\{\phi(1-\hat{\mu}_{ng})\}]. \end{aligned}$$

We derive a bound the first term. The bound for the second term and its derivation are identical. We see that

$$\begin{aligned} \mu_{ng}^2\Psi'(\phi\mu_{ng}) - \hat{\mu}_{ng}^2\Psi'(\phi\hat{\mu}_{ng}) =& (\mu_{ng}^2 - \hat{\mu}_{ng}^2)\Psi'(\phi\mu_{ng}) + \{\Psi'(\phi\mu_{ng}) - \Psi'(\phi\hat{\mu}_{ng})\}\mu_{ng}^2 \\ &+ (\mu_{ng}^2 - \hat{\mu}_{ng}^2)\{\Psi'(\phi\mu_{ng}) - \Psi'(\phi\hat{\mu}_{ng})\}, \end{aligned}$$

where

$$|\hat{\mu}_{ng} - \mu_{ng}| \leq \|\boldsymbol{X}_{g*}\|_2\|\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_2\|_2 \leq c\|\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_2\|_2.$$

The result then follows by Corollaries 1 and 2 and an application of the mean-value theorem.

For the first result,

$$\begin{aligned} N^{-1}\ell'_g(\phi) - \mathbb{E}\{N^{-1}\ell'_g(\phi)\} =& N^{-1}\sum_{n=1}^N \mu_{gn}[\log(y_{gn}) - \mathbb{E}\{\log(y_{gn})\}] \\ &+ N^{-1}\sum_{n=1}^N (1-\mu_{gn})[\log(1-y_{gn}) - \mathbb{E}\{\log(1-y_{gn})\}]. \end{aligned}$$

Since $\log(y_{gn})$ and $\log(1-y_{gn})$ are sub-exponential with uniformly bounded sub-exponential norms, both of the terms after the equal sign are sub-exponential with sub-exponential norms bounded above by $cN^{-1/2}$. The result then follows by Lemma 5.15 of Eldar et al. [21] and Lemma 11 in Section S12.6. □

Lemma 7. *There exist non-random functions $\eta_g(\phi)$ satisfying $\max_{g\in[G]}|\eta_g(\phi)| \leq cG^{-1/2}$ for all $\phi \in [c^{-1}, c]$ and $g \in [G]$ such that the the first-derivatives of the log-likelihoods satisfy*

$$\max_{g\in[G]}|N^{-1}\hat{\ell}'_g(\phi) - \{N^{-1}\ell'_g(\phi) + \eta_g(\phi)\}| = O_P\{\log^c(G)N^{-1}\}$$

for all $\phi \in [c^{-1}, c]$ as $N, G \to \infty$.

Proof. Let $\Psi(\phi) = \frac{d}{d\phi}\log\Gamma(\phi)$. Then the difference in derivatives is

$$\begin{aligned} N^{-1}\hat{\ell}'_g(\phi) - N^{-1}\ell'_g(\phi) =& N^{-1}\sum_n \{\mu_{ng}\Psi(\phi\mu_{ng}) - \hat{\mu}_{ng}\Psi(\phi\hat{\mu}_{ng})\} \\ &+ N^{-1}\sum_n [(1-\mu_{ng})\Psi\{\phi(1-\mu_{ng})\} - (1-\hat{\mu}_{ng})\Psi\{\phi(1-\hat{\mu}_{ng})\}] \\ &+ N^{-1}\sum_n (\hat{\mu}_{ng} - \mu_{ng})\log(y_{ng}) + N^{-1}\sum_n (\mu_{ng} - \hat{\mu}_{ng})\log(1-y_{ng}). \end{aligned} \tag{S6}$$

We provide concentration bounds for the first and third quantities after the equal sign, as concentrations for second and fourth can be obtained analogously. For the first, we see that

$$\begin{aligned}\hat{\mu}_{ng}\Psi(\phi\hat{\mu}_{ng}) - \mu_{ng}\Psi(\phi\mu_{ng}) =&(\hat{\mu}_{ng} - \mu_{ng})\Psi(\phi\mu_{ng}) + \{\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng})\}\mu_{ng} \\ &+ (\hat{\mu}_{ng} - \mu_{ng})\{\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng})\},\end{aligned} \tag{S7}$$

where

$$N^{-1}\sum_n(\hat{\mu}_{ng} - \mu_{ng})\Psi(\phi\mu_{ng}) = \boldsymbol{X}_{g*}^\top N^{-1}\sum_n \Psi(\phi\mu_{ng})\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\}$$

By Lemma 13 in Section S12.6, there exists non-random functions $\eta_{g,1}(\phi)$ satisfying $|\eta_{g,1}(\phi)| \leq cG^{-1/2}$ for all $\phi \in [c^{-1}, c]$, $g \in [G]$ and

$$\begin{aligned}\max_{g\in[G]}\left|\boldsymbol{X}_{g*}^\top N^{-1}\sum_n \Psi(\phi\mu_{ng})\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\} - \eta_{g,1}(\phi)\right| =&O_P\{\log^c(G)G^{-1/2}N^{-1/2}\} \\ =&O_P\{\log^c(G)N^{-1}\}.\end{aligned}$$

For the second term in (S7), Taylor's theorem implies

$$\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng}) = \phi\Psi'(\phi\mu_{ng})(\hat{\mu}_{ng} - \mu_{ng}) + \frac{\phi^2}{2}\Psi''(\phi\tilde{\mu}_{ng})(\hat{\mu}_{ng} - \mu_{ng})^2, \tag{S8}$$

where $\tilde{\mu}_{ng}$ lies in between μ_{ng} and $\hat{\mu}_{ng}$. By Lemma 13 in Section S12.6, there exists non-random functions $\eta_{g,2}(\phi)$ satisfying $|\eta_{g,2}(\phi)| \leq cG^{-1/2}$ for all $\phi \in [c^{-1}, c]$, $g \in [G]$ and

$$\begin{aligned}&\max_{g\in[G]}\left|\boldsymbol{X}_{g*}^\top N^{-1}\sum_n \phi\Psi'(\phi\mu_{ng})(\hat{\mu}_{ng} - \mu_{ng})\mu_{ng} - \eta_{g,2}(\phi)\right| \\ =&\max_{g\in[G]}\left|N^{-1}\phi\boldsymbol{X}_{g*}^\top\sum_{n=1}^N \Psi'(\phi\mu_{ng})\mu_{ng}\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\} - \eta_{g,2}(\phi)\right| = O_P\{\log^c(G)G^{-1/2}N^{-1/2}\} \\ =&O_P\{\log^c(G)N^{-1}\}.\end{aligned}$$

For the second term in (S8), the Cauchy-Schwartz inequality implies

$$\begin{aligned}\max_{\substack{g\in[G]\\ \phi\in[c^{-1},c]}}\left|\frac{\phi^2}{2}\Psi''(\phi\tilde{\mu}_{ng})(\hat{\mu}_{ng} - \mu_{ng})^2\right| \leq&c\{1 + o_P(1)\}\max_{g\in[G]}\|\boldsymbol{X}_{g*}\|_2^2\|\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\|_2^2 \\ \leq&c\{1 + o_P(1)\}\|\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\|_2^2.\end{aligned}$$

Corollaries 1 and 2 then imply

$$\max_{g\in[G]}\left|N^{-1}\sum_n \mu_{ng}\frac{\phi^2}{2}\Psi''(\phi\tilde{\mu}_{ng})(\hat{\mu}_{ng} - \mu_{ng})^2\right| = O_P\{\log^c(G)G^{-1}\} = O_P\{\log^c(G)N^{-1}\}.$$

Putting these together implies the second term in (S7) satisfies

$$\max_{g\in[G]}\left|N^{-1}\sum_n\{\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng})\}\mu_{ng} - \eta_{g,2}(\phi)\right| = O_P\{\log^c(G)N^{-1}\}.$$

For the third term in (S7), we see that

$$(\hat{\mu}_{ng} - \mu_{ng})\{\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng})\} = \phi\Psi'(\phi\tilde{\mu}_{ng})(\hat{\mu}_{ng} - \mu_{ng})^2,$$

where $\tilde{\mu}_{ng}$ lies in between $\hat{\mu}_{ng}$ and μ_{ng} by the mean-value theorem. As such, identical techniques to those used above can be used to show that

$$\max_{g\in[G]}\left|N^{-1}\sum_n(\hat{\mu}_{ng} - \mu_{ng})\{\Psi(\phi\hat{\mu}_{ng}) - \Psi(\phi\mu_{ng})\}\right| = O_P\{\log^c(G)N^{-1}\}.$$

Putting this together gives us that the first term in (S6) satisfies

$$\max_{g\in[G]}\left|N^{-1}\sum_n\{\mu_{ng}\Psi(\phi\mu_{ng}) - \hat{\mu}_{ng}\Psi(\phi\hat{\mu}_{ng})\} - \{\eta_{g,1}(\phi) + \eta_{g,2}(\phi)\}\right| = O_p\{\log^c(G)N^{-1}\}$$

for all $\phi \in [c^{-1}, c]$.

For the third term in (S6), Corollary 2 implies

$$\begin{aligned}
N^{-1}\sum_n(\hat{\mu}_{ng} - \mu_{ng})\log(y_{ng}) =& \boldsymbol{X}_{g*}^\top N^{-1}\sum_n\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\}\log(y_{ng}) \\
=& \boldsymbol{X}_{g*}^\top N^{-1}\sum_n\{\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n\}\log(y_{ng}) \\
&+ N^{-1}\sum_n 1\{\mathcal{A}_n = \mathcal{A}_{n,(-g)}\}\frac{h_{g,n} - \tilde{h}_{g,n}\mu_{gn}}{1 - h_{g,n}}(y_{ng} - \hat{\mu}_{ng})\log(y_{ng}) \\
&+ \boldsymbol{X}_{g*}^\top N^{-1}\sum_n 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\}\{\bar{\boldsymbol{\beta}}_n^{(1)} - \bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}\}\log(y_{gn}) \\
&+ O_P\{\log^c(N)G^{-1}\}
\end{aligned}$$

where $h_{g,n}$ is the g-th leverage score of the design matrix $\boldsymbol{X}_{\mathcal{A}_n}$, $\tilde{h}_{g,n} = \mathbf{1}^\top(\boldsymbol{X}_{\mathcal{A}_n}^\top\boldsymbol{X}_{\mathcal{A}_n})^{-1}\boldsymbol{X}_{g,\mathcal{A}_n}^\top$, and the bound on the O_P error term holds uniformly across $g \in [G]$. Since $h_{g,n}, \tilde{h}_{g,n} \leq c/G$,

$$\begin{aligned}
\max_{g\in[G]}\left|N^{-1}\sum_n 1\{\mathcal{A}_n = \mathcal{A}_{n,(-g)}\}\frac{h_{g,n} - \tilde{h}_{g,n}\mu_{gn}}{1 - h_{g,n}}(y_{ng} - \hat{\mu}_{ng})\log(y_{ng})\right| &= O_P\{\log^c(G)G^{-1}\} \\
&= O_P\{\log^c(G)N^{-1}\}.
\end{aligned}$$

Next, Corollaries 1 and 2 and the fact that $\log(y_{gn})$ is sub-exponential imply

$$\begin{aligned}
&\max_{g\in[G]}\left|\boldsymbol{X}_{g*}^\top N^{-1}\sum_n 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\}\{\bar{\boldsymbol{\beta}}_n^{(1)} - \bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}\}\log(y_{gn})\right| \\
=& O_P\left[\log^c(N)G^{-1/2}N^{-1}\sum_n 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\}\right] = O_P\{\log^c(G)G^{-1}\},
\end{aligned}$$

where the final equality follows by Lemma 12 from Section S12.6. Lastly, Lemma 13 from Section S12.6 implies there exist non-random functions $\eta_{g,3}(\phi)$ satisfying $\max_{g\in[G]}|\eta_{g,3}(\phi)| \leq cG^{-1/2}$ and

$$\max_{g\in[G]}\left|\boldsymbol{X}_{g*}^{\top}N^{-1}\sum_{n}\{\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}-\boldsymbol{\beta}_n\}\log(y_{ng})\right| = O_P\{\log^c(G)G^{-1/2}N^{-1/2}\} = O_P\{\log^c(G)N^{-1}\},$$

which completes the proof. □

We next show that we can express the difference $\hat{\phi}_g - \phi_g$ in a tractable manner.

Lemma 8. *Let $\eta_g(\phi)$ be as defined in the statement of Lemma 7. Then the difference $\hat{\phi}_g - \phi_g$ satisfies*

$$\max_{g\in[G]}\left|(\hat{\phi}_g-\phi_g)+\frac{N^{-1}\ell_g'(\phi_g)+\eta_g(\phi_g)}{N^{-1}\ell_g''(\phi_g)}\right| = O_P\{N^{-1}\log^c(G)\}$$

as $N, G \to \infty$.

Proof. Taylor's Theorem implies

$$0 = N^{-1}\hat{\ell}_g'(\hat{\phi}_g) = N^{-1}\hat{\ell}_g'(\phi_g) + N^{-1}\hat{\ell}_g''(\phi_g)(\hat{\phi}_g-\phi_g) + \frac{1}{2}N^{-1}\hat{\ell}_g^{(3)}(\tilde{\phi}_g)(\hat{\phi}_g-\phi_g)^2$$

for some $\tilde{\phi}_g$ that lies between ϕ_g and $\hat{\phi}_g$, where Corollary 3 implies $\max_{g\in[G]}|\hat{\phi}_g-\phi_g| = o_P(1)$. Lemma 6 implies this can be re-written as

$$0 = \frac{N^{-1}\hat{\ell}_g'(\phi_g)}{N^{-1}\hat{\ell}_g''(\phi_g)} + (\hat{\phi}_g-\phi_g) + \frac{N^{-1}\hat{\ell}_g^{(3)}(\tilde{\phi}_g)}{2N^{-1}\hat{\ell}_g''(\phi_g)}(\hat{\phi}_g-\phi_g)^2$$

for all $g \in [G]$. Note third derivatives are uniformly bounded from above. After finding the roots of this quadratic function of $(\hat{\phi}_g - \phi_g)$, Lemmas 6 and 7 imply

$$\max_{g\in[G]}\left|(\hat{\phi}_g-\phi_g)+\frac{N^{-1}\hat{\ell}_g'(\phi_g)}{N^{-1}\hat{\ell}_g''(\phi_g)}\right| = O_P\left[\max_{g\in[G]}\left\{N^{-1}\hat{\ell}_g'(\phi_g)\right\}^2\right] = O_P\{\log^c(G)N^{-1}\}.$$

The result then follows by a another application of Lemmas 6 and 7. □

S12.5 Proof of Theorem 1

We next prove Theorem 1.

Proof of Theorem 1. Let $\hat{\boldsymbol{V}}_n = \text{diag}\{\hat{\mu}_{n1}(1-\hat{\mu}_{n1})/(\hat{\phi}_1+1), \ldots, \hat{\mu}_{nG}(1-\hat{\mu}_{nG})/(\hat{\phi}_G+1)\}$ and $\boldsymbol{V}_n = \text{diag}\{\mu_{n1}(1-\mu_{n1})/(\phi_1+1), \ldots, \mu_{nG}(1-\mu_{nG})/(\phi_G+1)\}$ and define

$$\hat{\boldsymbol{\beta}}_n^{(2)} = \underset{\boldsymbol{\beta}\geq 0}{\arg\min}\|\hat{\boldsymbol{V}}_n^{-1/2}(\boldsymbol{y}_n - \boldsymbol{X}\boldsymbol{\beta})\|_2, \quad \bar{\boldsymbol{\beta}}_n^{(2)} = \{\mathbf{1}^{\top}\hat{\boldsymbol{\beta}}_n^{(2)}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(2)}$$
$$\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} = \underset{\boldsymbol{\beta}\geq 0}{\arg\min}\|\boldsymbol{V}_n^{-1/2}(\boldsymbol{y}_n - \boldsymbol{X}\boldsymbol{\beta})\|_2, \quad \bar{\boldsymbol{\beta}}_n^{(\text{Obs})} = \{\mathbf{1}^{\top}\hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(\text{Obs})},$$

where $\bar{\boldsymbol{\beta}}_n^{(2)}$ is our estimator for cellular fractions and $\bar{\boldsymbol{\beta}}_n^{(\text{Obs})}$ is the estimator when the variance is observed. Our goal is to show that

$$\frac{\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2}{\|\bar{\boldsymbol{\beta}}_n^{(\text{Obs})} - \boldsymbol{\beta}_n\|_2} = \frac{G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2}{G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(\text{Obs})} - \boldsymbol{\beta}_n\|_2} = o_P(1),$$

which we prove by showing that $G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 = o_P(1)$ and $\{G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(\text{Obs})} - \boldsymbol{\beta}_n\|_2\}^{-1} = O_P(1)$.

We start by proving that $\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 = o_P(G^{-1/2})$, which will imply $G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 = o_P(1)$. To do so, we need only show that $\|\hat{\boldsymbol{\beta}}_n^{(2)} - \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 = o_P(G^{-1/2})$ and $\|\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} - \boldsymbol{\beta}_n\|_2 = O_P(G^{-1/2})$, as this would imply

$$\begin{aligned}
\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\text{Obs})} &= \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(2)}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(2)} - \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} \\
&\underbrace{=}_{\boldsymbol{\Delta}_n = \hat{\boldsymbol{\beta}}_n^{(2)} - \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}} \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(\text{Obs})} + \mathbf{1}^\top \boldsymbol{\Delta}_n\}^{-1}\{\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} + \boldsymbol{\Delta}_n\} - \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\}^{-1}\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} \\
&= O_P(\|\boldsymbol{\Delta}_n\|_2) = o_P(G^{-1/2}).
\end{aligned}$$

First, $\|\hat{\boldsymbol{\beta}}_n^{(\text{Obs})} - \boldsymbol{\beta}_n\|_2 = O_P(G^{-1/2})$ follows by replacing $\boldsymbol{X} \leftarrow \boldsymbol{V}_n^{-1/2}\boldsymbol{X}$ and $\boldsymbol{\epsilon}_n \leftarrow \boldsymbol{V}_n^{-1/2}\boldsymbol{\epsilon}_n$ and applying Lemma 2 and Corollary 1. To show $\|\hat{\boldsymbol{\beta}}_n^{(2)} - \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 = o_P(G^{-1/2})$, we start by by deriving properties of $\hat{\boldsymbol{\beta}}_n^{(\text{Obs})}$. Let $\mathcal{A}_n^{(\text{Obs})} \subseteq [K]$ be the indices of $\hat{\boldsymbol{\beta}}_n^{(\text{Obs})}$ with non-zero entries. By replacing $\boldsymbol{X} \leftarrow \boldsymbol{V}_n^{-1/2}\boldsymbol{X}$ and $\boldsymbol{\epsilon}_n \leftarrow \boldsymbol{V}_n^{-1/2}\boldsymbol{\epsilon}_n$, Lemma 2 implies

$$\Pr\{[J_n] \subseteq \mathcal{A}_n^{(\text{Obs})}\} = 1 - o(1)$$

as $N, G \to \infty$. It therefore suffices to only consider the event $[J_n] \subseteq \mathcal{A}_n^{(\text{Obs})}$. Suppose $\mathcal{A}_n^{(\text{Obs})} = a_n \cup [J_n]$, where $a_n \in \{\text{all subsets of } [K] \setminus [J_n]\}$. Just as we showed in the proof of Lemma 4, the KKT conditions imply $\hat{\boldsymbol{\beta}}_n^{(\text{Obs})}$ satisfies

$$\begin{aligned}
w_{nk}(a_n) &= G^{1/2}\boldsymbol{e}_k^\top(\boldsymbol{X}_{a_n}^\top \boldsymbol{V}_n^{-1}\boldsymbol{X}_{a_n})^{-1}\boldsymbol{X}_{a_n}^\top \boldsymbol{V}_n^{-1}\boldsymbol{\epsilon}_n \geq -G^{1/2}\beta_{nk}, \quad k \in [J_n] \\
x_{nk}(a_n) &= G^{1/2}\boldsymbol{e}_k^\top(\boldsymbol{X}_{a_n}^\top \boldsymbol{V}_n^{-1}\boldsymbol{X}_{a_n})^{-1}\boldsymbol{X}_{a_n}^\top \boldsymbol{V}_n^{-1}\boldsymbol{\epsilon}_n \geq 0, \quad k \in a_n \\
z_{nk}(a_n) &= G^{-1/2}\boldsymbol{X}_{*k}^\top\{I_G - \boldsymbol{V}_n^{-1}\boldsymbol{X}_{a_n}(\boldsymbol{X}_{a_n}^\top \boldsymbol{V}_n^{-1}\boldsymbol{X}_{a_n})^{-1}\boldsymbol{X}_{a_n}^\top\}\boldsymbol{V}_n^{-1}\boldsymbol{\epsilon}_n < 0, \quad k \in [K] \setminus \{[J_n] \cup a_n\},
\end{aligned}$$

where $\boldsymbol{X}_{a_n}$ is as defined in the proof of Lemma 4. When $J_n = [K]$, a_n is the null set and we only need to consider $w_{nk}(a_n)$. We also define $\hat{w}_{nk}(a_n)$, $\hat{x}_{nk}(a_n)$, and $\hat{z}_{nk}(a_n)$ to be the analogues of $w_{nk}(a_n)$, $x_{nk}(a_n)$, and $z_{nk}(a_n)$ with $\boldsymbol{V}_n$ replaced with $\hat{\boldsymbol{V}}_n$.

To prove the results, we first note that an application of the Lindeberg central limit theorem implies that for fixed a_n, $w_{nk}(a_n)$, $x_{nk}(a_n)$, and $z_{nk}(a_n)$ are mean-zero and asymptotically normally distributed with an asymptotic variance whose eigenvalues lie in the interval $[c^{-1}, c]$. Let $\mathcal{A}_n^{(2)} \subseteq [K]$ be the non-zero indices of $\hat{\boldsymbol{\beta}}_n^{(2)}$. If we can show

$$|\hat{w}_{nk}(a_n) - w_{nk}(a_n)|,\ |\hat{x}_{nk}(a_n) - x_{nk}(a_n)|,\ |\hat{z}_{nk}(a_n) - z_{nk}(a_n)| = o_P(1),$$

the normality of $w_{nk}(a_n)$, $x_{nk}(a_n)$, and $z_{nk}(a_n)$ would imply $\mathcal{A}_n^{(2)} = \mathcal{A}_n^{(\text{Obs})}$ with probability tending to 1. In this case, we would have

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}_n^{(2)} - \hat{\boldsymbol{\beta}}_n^{(\text{Obs})}\|_2 &\leq G^{-1/2} \sum_{a_n} \left[\sum_{k=1}^{J_n} \{\hat{w}_{nk}(a_n) - w_{nk}(a_n)\}^2 + \sum_{k \in [a_n]} \{\hat{x}_{nk}(a_n) - x_{nk}(a_n)\}^2 \right]^{1/2} \\ &= o_P(G^{-1/2}).\end{aligned}$$

To do so, we prove $|\hat{w}_{nk}(a_n) - w_{nk}(a_n)| = o_P(1)$, noting that the proofs for the remaining two quantities are identical. To simplify notation, we let $\boldsymbol{X}$ denote $\boldsymbol{X}_{a_n}$. We first see that

$$\|(G^{-1}\boldsymbol{X}^\top \hat{\boldsymbol{V}}_n^{-1}\boldsymbol{X})^{-1} - (G^{-1}\boldsymbol{X}^\top \boldsymbol{V}_n^{-1}\boldsymbol{X})^{-1}\|_2 = o_P(1) \tag{S9}$$

by Corollaries 1 and 3. As such, proving the result only requires showing that

$$\begin{aligned}G^{-1/2}\|\boldsymbol{X}^\top \hat{\boldsymbol{V}}_n^{-1}\boldsymbol{\epsilon}_n - \boldsymbol{X}^\top \boldsymbol{V}_n^{-1}\boldsymbol{\epsilon}_n\|_2 =& G^{-1/2} \left\| \sum_g \frac{\hat{\phi}_g + 1}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} - \sum_g \frac{\phi_g + 1}{\mu_{ng}(1-\mu_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} \right\|_2 \\ =& o_P(1).\end{aligned}$$

The difference inside the vector 2-norm function can be expressed as

$$\begin{aligned}&\sum_g \frac{\hat{\phi}_g - \phi_g}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} \\ &+ \sum_g (\phi_g + 1) \left\{ \frac{1}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} - \frac{1}{\mu_{ng}(1-\mu_{ng})} \right\} \boldsymbol{X}_{g*}\epsilon_{ng}.\end{aligned} \tag{S10}$$

To study the behavior of the first term, let $\eta_g(\phi)$ be as defined in the statement of Lemma 7. We see that

$$\begin{aligned}\sum_g \frac{\hat{\phi}_g - \phi_g}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} =& -\sum_g \frac{\ell_g'(\phi_g)/\ell_g''(\phi_g)}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} \\ &- \sum_g \frac{\eta_g(\phi_g)/\{N^{-1}\ell_g''(\phi_g)\}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} + O_P\{GN^{-1}\log^c(N)\} \\ =& -\sum_g \frac{\ell_g'(\phi_g)/\ell_g''(\phi_g)}{\mu_{ng}(1-\mu_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} - \sum_g \frac{\eta_g(\phi_g)/\{N^{-1}\ell_g''(\phi_g)\}}{\mu_{ng}(1-\mu_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} \\ &+ \underbrace{O_P\{GN^{-1}\log^c(N)\}}_{o_P(G^{1/2})} \\ =& -\sum_g \frac{\ell_g'(\phi_g)/\ell_g''(\phi_g)}{\mu_{ng}(1-\mu_{ng})} \boldsymbol{X}_{g*}\epsilon_{ng} + o_P(G^{1/2})\end{aligned} \tag{S11}$$

where the first equality follows by Lemma 8, the second equality by Corollary 1 and Lemma 6, the underbrace by Assumption 5, and the third equality because $\eta_g(\phi_g)$ and $\{N^{-1}\ell_g''(\phi_g)\}$

are non-random and satisfy $|\eta_g(\phi_g)| \leq cG^{-1/2}$ and $|\{N^{-1}\ell''_g(\phi_g)\}| \geq c^{-1}$ (see Lemmas 6 and 7). We next break up the sum on the last line of the above expression into two parts:

$$\begin{aligned}\sum_g \frac{\ell'_g(\phi_g)/\ell''_g(\phi_g)}{\mu_{ng}(1-\mu_{ng})}\boldsymbol{X}_{g*}\epsilon_{ng} =& \sum_g \frac{\ell'_{(-n)g}(\phi_g)/\ell''_g(\phi_g)}{\mu_{ng}(1-\mu_{ng})}\boldsymbol{X}_{g*}\epsilon_{ng} \\ &+ N^{-1}\sum_g \frac{\ell'_{ng}(\phi_g)/\{\ell''_g(\phi_g)/N\}}{\mu_{ng}(1-\mu_{ng})}\boldsymbol{X}_{g*}\epsilon_{ng},\end{aligned} \tag{S12}$$

where $\ell'_{ng}(\phi_g)$ is the n-th sample's contribution to the gradient of the log-likelihood, i.e.,

$$\begin{aligned}\ell'_{ng}(\phi_g) =& \Psi(\phi_g) - \mu_{ng}\Psi(\mu_{ng}\phi_g) - (1-\mu_{ng})\Psi\{(1-\mu_{ng})\phi_g\} + \mu_{ng}\log(y_{ng}) \\ &+ (1-\mu_{ng})\log(1-y_{ng}),\end{aligned}$$

and $\ell'_{(-n)g}(\phi_g) = \ell'_g(\phi_g) - \ell'_{ng}(\phi_g)$. For the second term in (S12),

$$\begin{aligned}\|N^{-1}\sum_g \frac{\ell'_{ng}(\phi_g)/\{\ell''_g(\phi_g)/N\}}{\mu_{ng}(1-\mu_{ng})}\boldsymbol{X}_{g*}\epsilon_{ng}\|_2 &\leq cN^{-1}G\left\{\max_{g\in[G]}|\ell'_{ng}(\phi_g)|\right\}\left(\max_{g\in[G]}|\epsilon|_{ng}\right) \\ &= O_P\{\log^c(G)N^{-1}G\} = o_P(G^{1/2}),\end{aligned}$$

where the first equality follows by Lemmas 10 and 11 and the second by Assumption 5. For the first term in (S12), let $\boldsymbol{Z}_n$ be the first term in (S12) and $\mathcal{F} = \{\ell'_{(-n)1}(\phi_1), \ldots, \ell'_{(-n)G}(\phi_G)\}$. Since $\ell'_{(-n)g}(\phi_g)$ is independent of ϵ_{ng} and $\ell''_g(\phi_g)$ is non-random,

$$\mathbb{E}(\boldsymbol{Z}_n \mid \mathcal{F}) = \boldsymbol{0}$$

and

$$\mathrm{Var}(\boldsymbol{Z}_n \mid \mathcal{F}) = \sum_g \frac{\{\ell'_{(-n)g}(\phi_g)/\ell''_g(\phi_g)\}^2}{(\phi_g+1)\mu_{ng}(1-\mu_{ng})}\boldsymbol{X}_{g*}\boldsymbol{X}_{g*}^\top = O_P\{GN^{-1}\} = o_P(G),$$

where the second equality follows because $\mathbb{E}[\{\ell'_{(-n)g}(\phi_g)/\ell''_g(\phi_g)\}^2] \leq cN^{-1}$ for all $g \in [G]$ by Lemma 6. This implies the first term in (S10) is $o_P(G^{1/2})$.

For the second term in (S10), let $f(x) = \{x(1-x)\}^{-1}$. Taylor's theorem implies the expression in brackets within the summand is equal to

$$\begin{aligned}&f'(\mu_{ng})(\hat{\mu}_{ng} - \mu_{ng}) + f''(\tilde{\mu}_{ng})(\hat{\mu}_{ng} - \mu_{ng})^2/2 \\ =& f'(\mu_{ng})\boldsymbol{X}_{g*}\{\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_n\} + \frac{f''(\tilde{\mu}_{ng})}{2}[\boldsymbol{X}_{g*}^\top\{\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_n\}]^2,\end{aligned}$$

where $\tilde{\mu}_{ng}$ lies between μ_{ng} and $\hat{\mu}_n g$. Since $\max_{g\in[G]}|f''(\tilde{\mu}_{ng})|[\boldsymbol{X}_{g*}^\top\{\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_n\}]^2 = O_P(G^{-1})$ by Corollary 1, the second term in (S10) can be expressed as

$$G^{-1}\sum_g (\phi_g+1)f'(\mu_{ng})\boldsymbol{X}_{g*}\boldsymbol{X}_{g*}^\top(G^{-1}\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\epsilon}_n\epsilon_{ng} + o_P(G^{1/2}).$$

Let $\boldsymbol{Z}_n = (Z_{n1}, \ldots, Z_{nK})^\top$ be the first term in the above expression. (Please note we abuse notation, as this $\boldsymbol{Z}_n$ is different from the $\boldsymbol{Z}_n$ defined above.) Then

$$Z_{nk} = G^{-1} \sum_g \boldsymbol{v}_{gk}^\top \boldsymbol{\epsilon}_n \epsilon_{ng}, \quad \boldsymbol{v}_{gk} = (\phi_g + 1) f'(\mu_{ng}) \boldsymbol{X}_{gk} \boldsymbol{X} (G^{-1} \boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_{g*} \in \mathbb{R}^G.$$

Since the elements of $\boldsymbol{v}_{gk}$ are uniformly bounded across $g \in [G]$ and $k \in [K]$,

$$\mathbb{E}(Z_{nk}) = G^{-1} \sum_g \boldsymbol{v}_{gk,g} \mathbb{E}(\epsilon_{ng}^2) = O(1),$$

where $\boldsymbol{v}_{gk,g}$ is the g-th elements of $\boldsymbol{v}_{gk}$. To compute its variance, we observed that

$$Z_{nk} = \boldsymbol{\epsilon}_n^\top \boldsymbol{H} \boldsymbol{D}_{nk} \boldsymbol{\epsilon}_n, \quad \boldsymbol{D}_{nk} = \mathrm{diag}\{(\phi_1 + 1) f'(\mu_{n1}) \boldsymbol{X}_{1k}, \ldots, (\phi_G + 1) f'(\mu_{nG}) \boldsymbol{X}_{Gk}\} \preceq c I_G,$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$ is the orthogonal projection matrix that projects vectors onto the image $\boldsymbol{X}$ and has trace equal to K. Therefore,

$$\mathrm{Var}(Z_{nk}) \leq c \|\boldsymbol{H} \boldsymbol{D}_{nk}\|_F^2 = c \mathrm{Tr}(\boldsymbol{D}_{nk}^2 \boldsymbol{H}) \leq c \mathrm{Tr}(\boldsymbol{H}) = cK.$$

This implies the second term in (S10) is $o_P(G^{1/2})$. Putting this all together implies $\|\hat{\boldsymbol{\beta}}_n^{(2)} - \hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})}\|_2 = o_P(G^{-1/2})$ which, by the discussion above, proves $\|\bar{\boldsymbol{\beta}}_n^{(2)} - \bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})}\|_2 = o_P(G^{-1/2})$.

To complete the proof of Theorem 1, we must show that $\{G^{1/2} \|\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2\}^{-1} = O_P(1)$, i.e., that

$$\Pr(G^{1/2} \|\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2 \geq c^{-1})$$

can be made arbitrarily close to 1 by choosing c large enough. As we have already shown that $\|\hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2 = O_P(G^{-1/2})$ above,

$$\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n = \{\mathbf{1}^\top \hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})}\}^{-1} \hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n = \{I_K - \boldsymbol{\beta}_n \mathbf{1}^\top\} \hat{\boldsymbol{\delta}}_n^{(\mathrm{Obs})} + \underbrace{O_P(G^{-1})}_{= o_P(G^{-1/2})}$$

by Taylor's Theorem, where $\hat{\boldsymbol{\delta}}_n^{(\mathrm{Obs})} = \hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n$. Note $I_K - \boldsymbol{\beta}_n \mathbf{1}^\top$ has rank equal to $K - 1$, since $\boldsymbol{\beta}_n$ lies in its kernel. As such,

$$\|\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2 \leq c \left\| \boldsymbol{Q}_n^\top \hat{\boldsymbol{\delta}}_n^{(\mathrm{Obs})} \right\|_2 + o_P(G^{-1/2}),$$

where the columns of $\boldsymbol{Q}_n \in \mathbb{R}^{K \times (K-1)}$ are orthonormal and orthogonal to $\boldsymbol{\beta}_n$. Next, since $[J_n] \subseteq \mathcal{A}_n^{(\mathrm{Obs})}$ with probability tending to 1 by the discussion above (i.e., by extension of Lemma 2 to $\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})}$), it suffices to prove the result assuming $[J_n] \subseteq \mathcal{A}_n^{(\mathrm{Obs})}$. Let a_n be as defined above, let $i_n = [K] \setminus \{[J_n] \cup a_n\}$, and re-define $\boldsymbol{X} \leftarrow \boldsymbol{V}_n^{-1/2} \boldsymbol{X}$ and $\boldsymbol{\epsilon}_n \leftarrow \boldsymbol{V}_n^{-1/2} \boldsymbol{\epsilon}_n$. The KKT conditions imply $\mathcal{A}_n^{(\mathrm{Obs})} = [J_n] \cup a_n$ if

$$\begin{aligned}
\boldsymbol{w}_n(a_n) &= (G^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} G^{-1/2} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n \geq \begin{pmatrix} -G^{1/2} \boldsymbol{\beta}_{n,1} \\ \mathbf{0} \end{pmatrix} \\
\boldsymbol{z}_n(a_n) &= G^{-1/2} \boldsymbol{X}_{i_n}^\top P_{\boldsymbol{X}_{a_n}}^\perp \boldsymbol{\epsilon}_n < \mathbf{0},
\end{aligned}$$

where $\boldsymbol{\beta}_{n,1}$ are the first J_n elements of $\boldsymbol{\beta}_n$ and $P^{\perp}_{\boldsymbol{X}_{a_n}}$ is the orthogonal projection matrix that projects vectors onto the kernel of $\boldsymbol{X}_{a_n}^{\top}$. The Lindeberg central limit theorem implies

$$\begin{pmatrix} \boldsymbol{w}_n(a_n) \\ \boldsymbol{z}_n(a_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{d}_{n,1}(a_n) \\ \boldsymbol{d}_{n,2}(a_n) \end{pmatrix} + \boldsymbol{\zeta}_n, \quad \begin{pmatrix} \boldsymbol{d}_{n,1}(a_n) \\ \boldsymbol{d}_{n,2}(a_n) \end{pmatrix} \sim N_K \left(\mathbf{0}, \begin{pmatrix} (G^{-1}\boldsymbol{X}_{a_n}^{\top}\boldsymbol{X}_{a_n})^{-1} & \mathbf{0} \\ \mathbf{0} & G^{-1}\boldsymbol{X}_{i_n}^{\top} P^{\perp}_{\boldsymbol{X}_{a_n}} \boldsymbol{X}_{i_n} \end{pmatrix} \right),$$

$\|\boldsymbol{\zeta}_n\|_2 = o_P(1)$.

Further,

$$G^{1/2}\{\hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\} = \boldsymbol{M}_{a_n}\boldsymbol{w}_n(a_n) = \boldsymbol{M}_{a_n}\boldsymbol{d}_{n,1}(a_n) + o_P(1),$$

where $\boldsymbol{M}_{a_n} \in \{0,1\}^{K \times (J_n + a_n)}$. Putting all this together gives us

$$\begin{aligned} G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2 =& \sum_{a_n} 1\{\mathcal{A}_n^{(\mathrm{Obs})} = [J_n] \cup a_n\} G^{1/2}\|\boldsymbol{Q}_n^{\top}\{\hat{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\}\|_2 + o_P(1) \\ =& \sum_{a_n} 1\{\mathcal{A}_n^{(\mathrm{Obs})} = [J_n] \cup a_n\} \|\boldsymbol{Q}_n^{\top}\boldsymbol{M}_{a_n}\boldsymbol{d}_{n,1}(a_n)\|_2 \\ & + o_P(1) \\ =& \sum_{a_n} 1\{\mathcal{A}_n^{(\mathrm{Obs})} = [J_n] \cup a_n\} \{\|(I_{J_n} - \boldsymbol{\beta}_{n,1}\mathbf{1}^{\top})\boldsymbol{d}_{n,1,1}(a_n)\|_2^2 \\ & + \|\boldsymbol{d}_{n,1,2}(a_n)\|_2^2\}^{1/2} + o_P(1). \end{aligned}$$

where $\boldsymbol{d}_{n,1,1}(a_n)$ are the first J_n elements of $\boldsymbol{d}_{n,1}(a_n)$ and $\boldsymbol{d}_{n,1,2}(a_n)$ are the remaining elements. This decomposition then implies that any $\gamma > 0$, there exists a $\delta = \delta_\gamma > 0$ such that

$$\begin{aligned} \Pr\{G^{1/2}\|\bar{\boldsymbol{\beta}}_n^{(\mathrm{Obs})} - \boldsymbol{\beta}_n\|_2 \geq \gamma\} \geq& 1 - \sum_{a_n} \Pr\{\|(I_{J_n} - \boldsymbol{\beta}_{n,1}\mathbf{1}^{\top})\boldsymbol{d}_{n,1,1}(a_n)\|_2 < \delta_\gamma \text{ or } \|\boldsymbol{d}_{n,1,2}(a_n)\|_2 < \delta_\gamma \} \\ & + o(1). \end{aligned}$$

Since $(I_{J_n} - \boldsymbol{\beta}_{n,1}\mathbf{1}^{\top})\boldsymbol{d}_{n,1,1}(a_n)$ and $\boldsymbol{d}_{n,1,2}(a_n)$ are normally distributed with mean-zero and covariance whose largest eigenvalue exceeds c^{-1}, the probabilities inside the summand can be made arbitrarily small by choosing γ, and therefore δ_γ, to be sufficiently small. The result then follows. □

S12.6 Technical lemmas

Here we state and prove the technical lemmas referenced in the above proofs.

Lemma 9. *Suppose $y_{gn} \sim Beta\{\phi_g\mu_{gn}, \phi_g(1-\mu_{gn})\}$, where $\phi_g \in [c^{-1}, c]$ and $\mu_{gn} \in [c^{-1}, 1-c^{-1}]$ for all $g \in [G]$ and $n \in [N]$. Then $\log(y_{gn})$ and $\log(1-y_{gn})$ are sub-exponential with sub-exponential norms bounded above by a constant c.*

Proof. We prove the result for $\log(y_{gn})$, as the proof for $\log(1-y_{gn})$ is identical. Lemma 5.5 of Eldar et al. [21] implies we need only find a constant C that does not depend on n or g so that

$$\mathbb{E}[\exp\{|\log(y_{gn})|/C\}] \leq e.$$

Let $\alpha_{gn} = \phi_g \mu_{gn}$ and $\gamma_{gn} = \phi_g(1 - \mu_{gn})$. We see that for $\mathcal{B}(\cdot, \cdot)$ the beta function,

$$\mathbb{E}[\exp\{|\log(y_{gn})|/C\}] = \frac{\mathcal{B}(\alpha_{gn} - C^{-1}, \gamma_{gn})}{\mathcal{B}(\alpha_{gn}, \gamma_{gn})}.$$

Since α_{gn} and γ_{gn} lie on a compact set that does not include zero, we can make the right-hand side arbitrarily close to one for all $g \in [G]$ and $n \in [N]$ by choosing C large enough. □

Lemma 10. *Suppose $X_1, \ldots, X_N$ are mean-zero sub-Gaussian random variables with sub-Gaussian norm bounded above by C (which may depend on N). Then*

$$\max_{n \in [N]} |X_n| = O_P\{\log^{1/2}(N) C\}.$$

Proof. We see that

$$\Pr\left(\max_{n \in [N]} |X_n| \geq t\right) \leq \Pr\left(\max_{n \in [N]} X_n \geq t\right) + \Pr\left(\max_{n \in [N]} -X_n \geq t\right)$$

for any $t > 0$. We bound the first probability on the right hand side, noting that the bound for the second probability is identical. Lemma 5.5 of Eldar et al. [21] implies $\mathbb{E}\{\exp(\lambda X_n)\} \leq \exp(c\lambda^2 C^2)$ for some universal constant $c > 0$. As such,

$$\Pr\left(\max_{n \in [N]} X_n \geq t\right) \leq \Pr\left\{\sum_n \exp(\lambda X_n) \geq \exp(\lambda t)\right\} \leq \exp(c\lambda^2 C^2 - \lambda t + \log N),$$

for all $\lambda > 0$, where the second inequality follows by Markov's inequality. The result follows by minimizing λ. □

Lemma 11. *Suppose $X_1, \ldots, X_N$ are mean-zero sub-exponential random variables with sub-exponential norms bounded above by a constant $c > 0$. Then*

$$\max_{n \in [N]} |X_n| = O_P\{\log(N)\}.$$

Proof. The proof follows directly from Lemma 5.15 of Eldar et al. [21] and the proof of Lemma 10 above. □

Lemma 12. *Let $\mathcal{A}_n$ and $\mathcal{A}_{n,(-g)}$ be as defined in the statement of Lemma 4, and let $Z_g = \sum_{n=1}^N 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\}$ be the number of samples n where $\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}$. Then*

$$\mathbb{E}(Z_g) \leq c \log^{1/2}(G) N G^{-1/2}$$

and

$$\max_{g \in [G]} |Z_g - \mathbb{E}(Z_g)| = O_P\{\log(G)(NG^{-1/2})^{1/2}\} = o_P\{\log^{1/2}(G) N G^{-1/2}\}.$$

Proof. The bound on $\mathbb{E}(Z_g)$ follows directly from Lemma 4. To prove the bound on the maximum deviation, define $W_g = Z_g - \mathbb{E}(Z_g)$ and $p_{gn} = \Pr\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\} \leq c\{\log G\}^{1/2} G^{-1/2}$ by Lemma 4. Then the moment generating function for W_g satisfies

$$\log[\mathbb{E}\{\exp(\lambda W_g)\}] = \sum_{n=1}^{N} \log\{1 + p_{gn}(e^{\lambda} - 1)\} - \lambda \mathbb{E}(Z_g) \leq (e^{\lambda} - 1)\mathbb{E}(Z_g) - \lambda \mathbb{E}(Z_g)$$

for all $\lambda \in \mathbb{R}$, where the inequality follows by Taylor's Theorem. Let $t > 0$. Then

$$\Pr\left(\max_{g\in[G]} |W_g| \geq t\right) \leq \Pr\left(\max_{g\in[G]} W_g \geq t\right) + \Pr\left(\max_{g\in[G]} -W_g \geq t\right).$$

We provide an upper bound for $\Pr(\max_{g\in[G]} W_g \geq t)$. The bound for the second term is identical. For any $\lambda > 0$,

$$\begin{aligned}\Pr\left(\max_{g\in[G]} W_g \geq t\right) &\leq \Pr\left\{\sum_{g=1}^{G} \exp(\lambda W_g) \geq \exp(\lambda t)\right\} \leq \frac{\sum_{g=1}^{G} \mathbb{E}\{\exp(\lambda W_g)\}}{\exp(\lambda t)} \\ &\leq \exp\{(e^{\lambda} - 1)\mu - \lambda(\mu + t) + \log(G)\},\end{aligned}$$

where the second inequality follows by Markov's inequality and $\mu = c \log^{1/2}(G) N G^{-1/2}$ for some sufficiently large c. Note that $\mu = O(N^{c^{-1}})$ for some $c > 1$ by Assumption 5. Since $\lambda > 0$ was arbitrary, we can choose so as to minimize the right hand side of the inequality, which results in $\lambda = \log\{(\mu + t)/\mu\}$, implying

$$\begin{aligned}\Pr\left(\max_{g\in[G]} W_g \geq t\right) &\leq \exp[t - (\mu + t)\log\{(\mu + t)/\mu\} + \log(G)] \\ &\leq \exp\left\{-\frac{t^2}{2\mu}\left(1 - \frac{t}{\mu}\right) + \log(G)\right\}.\end{aligned}$$

The second inequality follows by Taylor's Theorem. Setting $t = c\{\log(G)\mu\}^{1/2}$ implies the right hand side can be made arbitrarily small by choosing the constant c large enough, which proves the result. □

Lemma 13. *The following hold for some non-random functions $\eta_{g,j}(\phi)$, $j = 1, 2, 3$, satisfying $\max_{g\in[G]} |\eta_{g,j}(\phi)| \leq cG^{-1/2}$:*

$$\begin{aligned}&\max_{g\in[G]} \left| N^{-1} \boldsymbol{X}_{g*}^{\top} \sum_{n=1}^{N} \Psi(\phi \mu_{ng})\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\} - \eta_{g,1}(\phi) \right| = O_P\{\log^c(G) G^{-1/2} N^{-1/2}\} \\ &\max_{g\in[G]} \left| N^{-1} \phi \boldsymbol{X}_{g*}^{\top} \sum_{n=1}^{N} \Psi'(\phi \mu_{ng}) \mu_{ng}\{\bar{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\} - \eta_{g,2}(\phi) \right| = O_P\{\log^c(G) G^{-1/2} N^{-1/2}\} \\ &\max_{g\in[G]} \left| N^{-1} \boldsymbol{X}_{g*}^{\top} \sum_{n=1}^{N} \{\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n\} \log(y_{gn}) - \eta_{g,3}(\phi) \right| = O_P\{\log^c(G) G^{-1/2} N^{-1/2}\}.\end{aligned}$$

Remark 3. The function $\eta_{g,3}(\phi)$ is a constant with respect to its argument ϕ. We call it a function of ϕ for consistency.

Proof. Let $\hat{\boldsymbol{\delta}}_n^{(1)} = \hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n$. Corollary 2 implies we can prove the above expressions after replacing $\bar{\boldsymbol{\beta}}_n^{(1)}$ and $\bar{\boldsymbol{\beta}}_{n,(-g)}^{(1)}$ with $\hat{\boldsymbol{\beta}}_n^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)} \boldsymbol{\beta}_n$ and $\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} \boldsymbol{\beta}_n$, respectively.

We prove the first result, as the proofs of the other two are identical. The first term can be expressed as

$$\begin{aligned} N^{-1} \boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi \mu_{ng}) \{\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n - \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)} \boldsymbol{\beta}_n\} =& N^{-1} \boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi \mu_{ng}) \hat{\boldsymbol{\delta}}_n^{(1)} \\ & - N^{-1} \sum_{n=1}^N \mu_{ng} \Psi(\phi \mu_{ng}) \mathbf{1}^\top \hat{\boldsymbol{\delta}}_n^{(1)}. \end{aligned} \tag{S13}$$

Define x_g to be the first term to the right of the equal sign and let the event E to be as defined in the statement of Lemma 3. Then for any $t > 0$, Lemma 2 implies

$$\begin{aligned} \Pr(\max_{g \in [G]} |x_g - \mathbb{E}(x_g)| > t) \leq & \Pr(\max_{g \in [G]} \{|x_g - \mathbb{E}(x_g 1\{E\})| > t - \max_{g \in [G]} |\mathbb{E}(x_g 1\{E^{\mathrm{c}}\})|\} \cap E) \\ & + c \exp\{c \log(G) - c^{-1} G\}. \end{aligned}$$

We first prove that $\max_{g \in [G]} G |\mathbb{E}(x_g 1\{E^{\mathrm{c}}\})|$ converges to 0 and that $|\mathbb{E}(x_g)| \leq c G^{-1/2}$. We then prove that

$$\Pr\left(\{\max_{g \in [G]} |x_g - \mathbb{E}(x_g 1\{E\})| > \lambda \log^c(G) G^{-1/2} N^{-1/2}\} \cap E \right)$$

can be made arbitrarily small by choosing a large enough λ.

Let a_n and $\boldsymbol{X}_{a_n}$ be as defined in the proof of Lemma 4. Then on the event E, the first term to the right of the equals sign in (S13) can be written as

$$N^{-1} \boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi \mu_{ng}) \hat{\boldsymbol{\delta}}_n^{(1)} = \boldsymbol{X}_{g*}^\top N^{-1} \sum_{n=1}^N \sum_{a_n} \boldsymbol{M}_{a_n} 1\{\mathcal{A}_n = a_n \cup [J_n]\} \Psi(\phi \mu_{ng}) (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n,$$

where $\boldsymbol{M}_{a_n} \in \{0,1\}^{K \times (J_n + |a_n|)}$ is a matrix whose columns are all 0 except for one entry that is equal to 1. Note $\boldsymbol{X}_{g*}^\top \boldsymbol{M}_{a_n}$ equals the g-th row of $\boldsymbol{X}_{a_n}$. On the event E,

$$\begin{aligned} & \left\| \mathbb{E}[\boldsymbol{M}_{a_n} 1\{\mathcal{A}_n = a_n \cup [J_n]\} \Psi(\phi \mu_{ng}) (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n 1\{E\}] \right\|_2 \\ & \leq \mathbb{E}\left[\left\| \boldsymbol{M}_{a_n} 1\{\mathcal{A}_n = a_n \cup [J_n]\} \Psi(\phi \mu_{ng}) (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n \right\|_2 1\{E\} \right] \\ & \leq c \left[\mathbb{E}\left\{ \left\| (\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1} \boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n \right\|_2^2 \right\} \right]^{1/2} \leq c G^{-1/2}, \end{aligned}$$

which also implies

$$\left| \mathbb{E}\left[N^{-1} \boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi \mu_{ng}) \{\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n\} 1\{E\} \right] \right| \leq c G^{-1/2}.$$

On the event E^{c}, the first term in (S13) satisfies

$$\left| \mathbb{E}\left[N^{-1} \boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi \mu_{ng}) \hat{\boldsymbol{\delta}}_n^{(1)} 1\{E^{\mathrm{c}}\} \right] \right| \leq c \max_{n \in [N]} \mathbb{E}\{\|\hat{\boldsymbol{\delta}}_n^{(1)}\|_2\} \leq c + c \max_{n \in [N]} \mathbb{E}\{\|\hat{\boldsymbol{\beta}}_n^{(1)}\|_2\}.$$

where the first inequality follows because $\|\boldsymbol{X}_{g*}\|_2, |\Psi(\phi\mu_{ng})| \leq c$ and the second because $\|\boldsymbol{\beta}_n\|_2 \leq c$. Since the non-zero elements of $\hat{\boldsymbol{\beta}}_n^{(1)}$ equal the ordinary least squares estimator for the regression of $\boldsymbol{y}_n$ onto some subset of the columns of $\boldsymbol{X}$, $\mathbb{E}\{\|\hat{\boldsymbol{\beta}}_n^{(1)}\|_2 \leq c$. Together, these results and Lemma 2 imply

$$\left|\mathbb{E}\left[N^{-1}\boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi\mu_{ng})\hat{\boldsymbol{\delta}}_n^{(1)}\right]\right| \leq cG^{-1/2}$$
$$\left|\mathbb{E}\left[N^{-1}\boldsymbol{X}_{g*}^\top \sum_{n=1}^N \Psi(\phi\mu_{ng})\hat{\boldsymbol{\delta}}_n^{(1)} 1\{E^{\mathrm{c}}\}\right]\right| \leq c\exp\{c\log(G) - c^{-1}G\},$$

which proves $\max_{g\in[G]} G|\mathbb{E}(x_g 1\{E^{\mathrm{c}}\})| \to 0$ and $|\mathbb{E}(x_g)| \leq cG^{-1/2}$.

To complete the proof for the first term in (S13), define $\boldsymbol{\alpha}_{a_n,g}$ to be

$$\boldsymbol{\alpha}_{a_n,g} = \boldsymbol{M}_{a_n} 1\{\mathcal{A}_n = a_n \cup [J_n]\}\Psi(\phi\mu_{ng})(\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1}\boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n,$$

where

$$x_g = \boldsymbol{X}_{g*}^\top N^{-1}\sum_{n=1}^N \sum_{a_n} \boldsymbol{\alpha}_{a_n,g}, \quad \mathbb{E}(x_g 1\{E\}) = \boldsymbol{X}_{g*}^\top N^{-1}\sum_{n=1}^N \sum_{a_n} \mathbb{E}(\boldsymbol{\alpha}_{a_n,g})$$

on E. For any unit vector $\boldsymbol{v} \in \mathbb{R}^K$ and $p \geq 1$,

$$\begin{aligned}\{\mathbb{E}(G^{1/2}|\boldsymbol{v}^\top \boldsymbol{\alpha}_{a_n,g}|^p)\}^{1/p} \leq& c\{\mathbb{E}(|\boldsymbol{v}^\top \boldsymbol{M}_{a_n}(G^{-1}\boldsymbol{X}_{a_n}^\top \boldsymbol{X}_{a_n})^{-1}G^{-1/2}\boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n|^p)\}^{1/p} \\ \leq& cp^{1/2}\|G^{-1/2}\boldsymbol{X}_{a_n}^\top \boldsymbol{\epsilon}_n\|_{\psi_2} \leq cp^{1/2},\end{aligned}$$

meaning $G^{1/2}\boldsymbol{\alpha}_{a_n,g}$ is sub-Gaussian with uniformly bounded sub-Gaussian norm. Since the $\boldsymbol{\alpha}_{a_n,g}$'s are independent across n,

$$\boldsymbol{X}_{g*}^\top N^{-1}\sum_{n=1}^N \sum_{a_n} \{\boldsymbol{\alpha}_{a_n,g} - \mathbb{E}(\boldsymbol{\alpha}_{a_n,g})\}$$

is mean-zero and sub-Gaussian with sub-Gaussian norm bounded above by $cG^{-1/2}N^{-1/2}$ by Lemma 5.9 of Eldar et al. [21]. Then

$$\Pr\left[\max_{g\in[G]}\{|x_g - \mathbb{E}(x_g 1\{E\})| > t\} \cap E\right] \leq \Pr\left[\max_{g\in[G]}\left|\boldsymbol{X}_{g*}^\top N^{-1}\sum_{n=1}^N \sum_{a_n}\{\boldsymbol{\alpha}_{a_n,g} - \mathbb{E}(\boldsymbol{\alpha}_{a_n,g})\}\right| > t\right].$$

The desired bound then follows by Lemma 10. Putting all of this together implies $|\mathbb{E}(x_g)| \leq cG^{-1/2}$ and that

$$\Pr\{\max_{g\in[G]}|x_g - \mathbb{E}(x_g)| > \lambda\log^c(G)G^{-1/2}N^{-1/2}\}$$

can be made arbitrarily small by choosing λ sufficiently large.

An identical proof can be used to show the same holds for $z_g = N^{-1}\sum_{n=1}^{N}\mu_{ng}\Psi(\phi\mu_{ng})\mathbf{1}^\top\hat{\boldsymbol{\delta}}_n^{(1)}$. This implies the first result in the statement of the lemma holds with

$$\eta_{g,1}(\phi) = \mathbb{E}\left[N^{-1}\boldsymbol{X}_{g*}^\top\sum_{n=1}^{N}\Psi(\phi\mu_{ng})\{\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n - \mathbf{1}^\top\hat{\boldsymbol{\delta}}_n^{(1)}\boldsymbol{\beta}_n\}\right].$$

The remaining two statements hold with

$$\eta_{g,2}(\phi) = \mathbb{E}\left[N^{-1}\phi\boldsymbol{X}_{g*}^\top\sum_{n=1}^{N}\Psi'(\phi\mu_{ng})\mu_{ng}\{\hat{\boldsymbol{\beta}}_n^{(1)} - \boldsymbol{\beta}_n - \mathbf{1}^\top\hat{\boldsymbol{\delta}}_n^{(1)}\boldsymbol{\beta}_n\}\right]$$
$$\eta_{g,3}(\phi) = \mathbb{E}\left[N^{-1}\boldsymbol{X}_{g*}^\top\sum_{n=1}^{N}\{\hat{\boldsymbol{\beta}}_{n,(-g)}^{(1)} - \boldsymbol{\beta}_n - \mathbf{1}^\top\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\boldsymbol{\beta}_n\}\log(y_{gn})\right].$$

□

S13 Proof of Theorem 1 when data do not follow a beta distribution

Here we prove Theorem 1 when DNAm levels y_{gn} do not follow a beta distribution.

S13.1 Assumptions

The below assumptions are assumed to hold for some large constant $c > 1$ in all proofs below.

Assumption 6 (Design conditions)**.** *The matrix $\boldsymbol{X} = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_G)^\top$ is non-random and satisfies:*

(i) $\boldsymbol{X}_{gk} \in [c^{-1}, 1-c^{-1}]$ for all $g \in [G]$ and $k \in [K]$.

(ii) The smallest eigenvalue of $G^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ is greater than c^{-1}.

Assumption 7 (True coefficients)**.** *The sum of cell type proportions is equal to one, i.e., $\mathbf{1}^\top\boldsymbol{\beta}_n = 1$. Additionally, let $J_n \in \{2, \ldots, K\}$. The first J_n coefficients of $\boldsymbol{\beta}_n$ are non-zero and the second set of $K - J_n$ coefficients are zero. The non-zero coefficients satisfy $\beta_{nk} \geq c^{-1}$ for all $k \in [J_n]$.*

Assumption 8 (Precision parameters)**.** $\phi_g \in [c^{-1}, c]$ *for all* $g \in [G]$.

Assumption 9 (Observed DNAm levels)**.** *The random variables $\{y_{ng}\}_{n\in[N];g\in[G]}$ are independent, bounded between zero and one, and satisfy*

$$\mathbb{E}(y_{ng}) = \mu_{ng} = \boldsymbol{X}_{g*}^\top\boldsymbol{\beta}_n, \quad \mathrm{Var}(y_{ng}) = \mu_{ng}(1-\mu_{ng})/(\phi_g+1), \quad n \in [N]; g \in [G].$$

Assumption 10 (Sample size)**.** $c^{-1}N \leq G \leq cN^{2-c^{-1}}$.

Assumptions 6, 7, and 10 are identical to Assumptions 1, 2, and 5 from Section S11. Assumption 8 is slightly more general than Assumption 3 from Section S11, as it does not explicitly require ϕ_g be estimated in the compact set.

The main change from Section S11 is Assumption 9. In additional to the random variables being independent (which was also required in Assumption 4 from Section S11), Assumption 9 assumes we have the correct form for the mean and variance model. As the data do not need to follow a specific distribution, this is much more general than requiring the data follow a beta distribution.

S13.2 BLEND-M's estimators

Here we detail BLEND-M's estimates for cellular proportions when only the mean and variance model are known. Since this only impacts our estimates for ϕ_g, we simply change its estimate from the maximum likelihood estimator to the method of moments estimator given in (S1). All other estimates, i.e., its first- and second-step estimates for cellular proportions, remain the same.

S13.3 Proof of Theorem 1

We first note that the lemmas and corollaries presented in Section S12.3 do not depend on the way the precision parameter is estimated and only depend on the beta assumption through the fact that when this holds, y_{gn} is bounded between zero and one. As such, all of these results hold under the new set of assumptions. The same is true for Lemmas 10-12 from Section S12.6.

Since the only part of BLEND-M that changes is the estimate for the precision parameter, we need only modify the results regarding $\hat{\phi}_g$ from Section S12.4. In particular, we prove an analogue of Lemma 8, which states that the method of moments estimator can be expressed in a tractable manner. We then prove Theorem 1.

Lemma 14. *Let $\hat{\phi}_g$ be as defined in* (S1) *and assume Assumptions 6–10 hold. Then there exist non-random quantities η_g satisfying $\max_{g\in[G]}|\eta_g| \leq cG^{-1/2}$ for all $g \in [G]$ such that*

$$\max_{g\in[G]}\left|(\hat{\phi}_g - \phi_g) + (\phi_g + 1)^2\left\{N^{-1}\sum_{n=1}^{N}\frac{\epsilon_{ng}^2 - \mathbb{E}(\epsilon_{ng}^2)}{\mu_{ng}(1-\mu_{ng})} + \eta_g\right\}\right| = O_P\{\log^c(G)N^{-1}\}.$$

Remark 4. We abuse notation in the statement of the above lemma, as the quantities η_g are not the same as the non-random functions defined in the statement of Lemma 8. We use the same notation because the two sets of quantities share the same properties.

Remark 5. Since ϵ_{ng} is bounded between zero and one under Assumption 9, $N^{-1}\sum_{n=1}^{N}\frac{\epsilon_{ng}^2-\mathbb{E}(\epsilon_{ng}^2)}{\mu_{ng}(1-\mu_{ng})}$ is sub-Gaussian with sub-Gaussian norm bounded above by $cN^{-1/2}$ for all $g \in [G]$. As such, Lemma 10 implies a direct consequence of Lemma 14 is that $\max_{g\in[G]}|\hat{\phi}_g - \phi_g| = O_P\{\log^{1/2}(G)N^{-1/2}\}$.

Proof. Define

$$\hat{\alpha}_g = N^{-1} \sum_{n=1}^{N} \frac{(y_{ng} - \hat{\mu}_{ng})^2}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})},$$

where $\hat{\phi}_g = \max(\hat{\alpha}_g^{-1} - 1, 0)$. To prove the desired result, we need only show that

$$\max_{g \in [G]} \left| \{\hat{\alpha}_g - (\phi_g + 1)^{-1}\} - N^{-1} \sum_{n=1}^{N} \frac{\epsilon_{ng}^2 - \mathbb{E}(\epsilon_{ng}^2)}{\mu_{ng}(1 - \mu_{ng})} - \eta_g \right| = O_P\{\log^c(G) N^{-1}\}.$$

This is because $\zeta_g = N^{-1} \sum_{n=1}^{N} \frac{\epsilon_{ng}^2 - \mathbb{E}(\epsilon_{ng}^2)}{\mu_{ng}(1-\mu_{ng})}$ is sub-Gaussian with sub-Gaussian norm bounded above by $cN^{-1/2}$ (by Lemma 5.5 in Eldar et al. [21]). As such, Taylor's theorem would then imply

$$\begin{aligned} \max_{g \in [G]} \left| \hat{\alpha}_g^{-1} - (\phi_g + 1) + (\phi_g + 1)^2 \{\zeta_g + \eta_g\} \right| \leq & O_P \left\{ \max_{g \in [G]} |\{\zeta_g + \eta_g\}^2| + \log^c(G) N^{-1} \right\} \\ = & O_P\{\log^c(G) N^{-1}\}. \end{aligned}$$

Let $\bar{\boldsymbol{\delta}}_n^{(1)}$, $\bar{\boldsymbol{\delta}}_{n,(-g)}^{(1)}$, $\hat{\boldsymbol{\delta}}_n^{(1)}$, and $\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}$ be as defined in Corollary 2, and define $\tilde{\boldsymbol{X}} = \boldsymbol{X}(I_K - \boldsymbol{\beta}_n \mathbf{1}^\top)$ and $\hat{\boldsymbol{\Delta}}_{n,g} = \hat{\boldsymbol{\delta}}_n^{(1)} - \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}$. To simplify notation, we note that the O_P remainders in all of the below expressions are valid over maxima taken over $g \in [G]$. That is, if $z_g = O_P(a_n)$ for some sequence a_n, it is taken to mean $\max_{g \in [G]} |z_g| = O_P(a_n)$.

First, $\hat{\alpha}_g$ can be expressed as

$$\begin{aligned} \hat{\alpha}_g = & N^{-1} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \boldsymbol{X}_{g*}^\top \bar{\boldsymbol{\delta}}_n^{(1)}\}^2}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} \underbrace{=}_{\text{Corollary 1,2}} N^{-1} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_n^{(1)}\}^2}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} + O_P\{\log^c(G) N^{-1}\} \\ = & N^{-1} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}\}^2}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} + O_P\{\log^c(G) N^{-1}\} \\ \underbrace{=}_{\text{Corollary 1}} & N^{-1} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\}^2}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} - \frac{2}{N} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\} \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} \\ & + O_P\{\log^c(G) N^{-1}\}. \end{aligned}$$

To bound the second term in the last line, note that

$$\begin{aligned} & \frac{2}{N} \sum_{n=1}^{N} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\} \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} \\ = & \frac{2}{N} \sum_{n=1}^{N} 1\{\mathcal{A}_n = \mathcal{A}_{n,(-g)}\} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\} \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})} \\ & + \frac{2}{N} \sum_{n=1}^{N} 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\} \frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\} \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1 - \hat{\mu}_{ng})}. \end{aligned} \tag{S14}$$

By Lemma 2, it suffices to assume $[J_n] \subseteq \mathcal{A}_n, \mathcal{A}_{n,(-g)}$. On this event and when $\mathcal{A}_n = \mathcal{A}_{n,(-g)}$,

$$\hat{\boldsymbol{\Delta}}_{n,g} = \frac{\epsilon_{ng} - (\hat{\mu}_{gn} - \mu_{gn})}{1 - h_{gn}} \left(\boldsymbol{X}_{\mathcal{A}_n}^\top \boldsymbol{X}_{\mathcal{A}_n}\right)^{-1} \boldsymbol{X}_{g,\mathcal{A}_n},$$

where h_{gn} is the g-th leverage score of the design matrix $\boldsymbol{X}_{\mathcal{A}_n}$ and $\boldsymbol{X}_{g,\mathcal{A}_n}$ is the g-th row of $\boldsymbol{X}_{\mathcal{A}_n}$, expressed as a column vector. Since $\max_{g\in[G],n\in[N]}|h_{gn}| \leq cG^{-1}$,

$$\begin{aligned}
&\max_{g\in[G]}\left|\frac{2}{N}\sum_{n=1}^{N} 1\{\mathcal{A}_n = \mathcal{A}_{n,(-g)}\}\frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\}\tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})}\right| \\
&\leq cG^{-1} \max_{g\in[G],n\in[N]}\left|\frac{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})}\right| \max_{g\in[G],n\in[N]}|\epsilon_{ng} - (\hat{\mu}_{gn} - \mu_{gn})| = O_P\{\log^c(G)G^{-1}\} \\
&= O_P\{\log^c(G)N^{-1}\}.
\end{aligned}$$

Next, Corollary 1 and Lemma 12 imply

$$\begin{aligned}
&\max_{g\in[G]}\left|\frac{2}{N}\sum_{n=1}^{N} 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\}\frac{\{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\}\tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\Delta}}_{n,g}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})}\right| \\
&\leq c \max_{g\in[G],n\in[N]}\left|\frac{\epsilon_{ng} - \tilde{\boldsymbol{X}}_{g*}^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})}\right| \max_{g\in[G],n\in[N]}\|\hat{\boldsymbol{\Delta}}_{n,g}\|_2 \max_{g\in[G]} N^{-1}\sum_{n=1}^{N} 1\{\mathcal{A}_n \neq \mathcal{A}_{n,(-g)}\} \\
&= O_P\{\log^c(G)N^{-1}\}.
\end{aligned}$$

We now return to the expression for $\hat{\alpha}_g$. The above bounds and Corollary 1 imply it can be expressed as

$$\hat{\alpha}_g = N^{-1}\sum_{n=1}^{N}\frac{\epsilon_{ng}^2}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} - \frac{2}{N}\sum_{n=1}^{N}\frac{\epsilon_{ng}\tilde{\boldsymbol{X}}_g^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} + O_P\{\log^c(G)N^{-1}\}. \tag{S15}$$

We start by considering the second term after the equality. Corollaries 1 and 2 imply $\max_{n\in[N],g\in[G]}|\hat{\mu}_{ng} - \mu_{ng}| = O_P\{\log^c(N)G^{-1/2}\} = O_P\{\log^c(G)N^{-1/2}\}$. As such, a second application of Corollary 1 and Taylor's theorem imply the second term can be expressed as

$$\frac{2}{N}\sum_{n=1}^{N}\frac{\epsilon_{ng}\tilde{\boldsymbol{X}}_g^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}}{\mu_{ng}(1-\mu_{ng})} + O_P\{\log^c(G)N^{-1}\}.$$

Since ϵ_{ng} and $\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}$ are independent, the first term is sub-Gaussian with sub-Gaussian norm bounded above by $cN^{-1}\left[\sum_{n=1}^{N}\{\tilde{\boldsymbol{X}}_g^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\}^2\right]^{1/2}$ conditional on $\hat{\boldsymbol{\delta}}_{1,(-g)}^{(1)}, \ldots, \hat{\boldsymbol{\delta}}_{N,(-g)}^{(1)}$ by Lemma 5.5 of Eldar et al. [21]. Corollary 1 and Lemma 10 then imply

$$\max_{g\in[G]}\left|\frac{2}{N}\sum_{n=1}^{N}\frac{\epsilon_{ng}\tilde{\boldsymbol{X}}_g^\top \hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}}{\mu_{ng}(1-\mu_{ng})}\right| = O_P\{\log^c(G)N^{-1}\}.$$

To complete the proof, it remains to consider the first term on the right hand side of (S15). To do so, let $f(x) = \{x(1-x)\}^{-1}$. Then Taylor's theorem and Corollary 2 imply

$$\begin{aligned} N^{-1}\sum_{n=1}^{N}\frac{\epsilon_{ng}^2}{\hat{\mu}_{ng}(1-\hat{\mu}_{ng})} =& N^{-1}\sum_{n=1}^{N}\frac{\epsilon_{ng}^2}{\mu_{ng}(1-\mu_{ng})} + N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_n^{(1)} \\ &+ N^{-1}\sum_{n=1}^{N}\frac{f''(\tilde{\mu}_{gn})}{2}(\hat{\mu}_{gn}-\mu_{gn})^2 + O_P\{\log^c(G)N^{-1}\}, \end{aligned}$$

where $\tilde{\mu}_{gn}$ lies between μ_{gn} and $\hat{\mu}_{gn}$. Corollary 1 and 2 imply the term term on the right hand side of the equality is $O_P\{\log^c(G)N^{-1}\}$. Therefore, to complete the proof, we need only show that

$$N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_n^{(1)} = \eta_g + O_P\{\log^c(G)N^{-1}\},$$

where η_g satisfies the properties in the statement of the lemma.

To prove this, we see that

$$N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_n^{(1)} = N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} + N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\Delta}}_{n,g}^{(1)}.$$

Identical techniques used to bound (S14) can be used to show that the second term on the right hand side of the equality is $O_P\{\log^c(G)N^{-1}\}$. Finally, an identical proof used to prove Lemma 13 can be used to show that

$$\eta_g = \mathbb{E}\left\{N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)}\right\}$$

satisfies the properties in the statement of the lemma and $N^{-1}\tilde{\boldsymbol{X}}_{g*}^{\top}\sum_{n=1}^{N} f'(\mu_{gn})\epsilon_{ng}^2\hat{\boldsymbol{\delta}}_{n,(-g)}^{(1)} = \eta_g + O_P\{\log^c(G)N^{-1}\}$. □

We next prove Theorem 1 assuming Assumptions 6–10 hold. Since the only change we have made from the proof in Section S12.5 is the estimator for ϕ_g, proving the theorem only requires updating that proof to account for the change in estimator. We therefore need only show (S9) still holds (see the proof of Theorem 1 in Section S12.5), and that the term on the left hand side of the equality in (S11) is $o_P(G^{1/2})$. We prove these two below, which together with the proof from Section S12.5 in, will complete the proof of Theorem 1 under the new set of assumptions.

Proof that (S9) *holds.* To prove (S9) holds, we need only show that

$$\max_{g\in[G]}\left|\frac{\hat{\phi}_g+1}{\hat{\mu}_{gn}(1-\hat{\mu}_{gn})} - \frac{\phi_g+1}{\mu_{gn}(1-\mu_{gn})}\right| = o_P(1). \tag{S16}$$

This follows directly from Corollaries 1 and 2, Lemma 14, and Remark 5. □

Proof that the term on the left hand side of the equality in (S11) *is* $o_P(G^{1/2})$. To prove this, we re-define $\ell'(\phi_g)$, $\ell''(\phi_g)$, and $\eta_g(\phi_g)$ from the proof of Theorem 1 in Section S12.5 to be

$$\ell'(\phi_g) = (\phi_g + 1)^2 \sum_{n=1}^{N} \frac{\epsilon_{ng}^2 - \mathbb{E}(\epsilon_{ng}^2)}{\mu_{ng}(1-\mu_{ng})}, \quad \ell''(\phi_g) = N, \quad \eta_g(\phi_g) = (\phi_g + 1)^2 \eta_g,$$

where η_g is as defined in the statement of Lemma 14. The result then follows by the same proof we used to show that (S11) is $o_P(G^{1/2})$ in Section S12.5. □

References

[1] T. Sun, J. Yuan, Y. Zhu, J. Li, S. Yang, J. Zhou, et al. "Systematic evaluation of methylation-based cell type deconvolution methods for plasma cell-free DNA". In: *Genome Biology* 25.1 (Dec. 2024). ISSN: 1474-760X. DOI: 10.1186/s13059-024-03456-8. URL: http://dx.doi.org/10.1186/s13059-024-03456-8.

[2] M. Forest, K. J. O'Donnell, G. Voisin, H. Gaudreau, J. L. MacIsaac, L. M. McEwen, et al. "Agreement in DNA methylation levels from the Illumina 450K array across batches, tissues, and time". In: *Epigenetics* 13.1 (Jan. 2018), pp. 19–32. ISSN: 1559-2308. DOI: 10.1080/15592294.2017.1411443. URL: http://dx.doi.org/10.1080/15592294.2017.1411443.

[3] I. V. Yang, B. S. Pedersen, A. Liu, G. T. O'Connor, S. J. Teach, M. Kattan, et al. "DNA methylation and childhood asthma in the inner city". In: *Journal of Allergy and Clinical Immunology* 136.1 (2015), pp. 69–80. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2015.01.025. URL: http://dx.doi.org/10.1016/j.jaci.2015.01.025.

[4] A. Akalin, M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, et al. "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles". In: *Genome Biology* 13.10 (Oct. 2012). ISSN: 1474-760X. DOI: 10.1186/gb-2012-13-10-r87. URL: http://dx.doi.org/10.1186/gb-2012-13-10-r87.

[5] H. Feng, K. N. Conneely, and H. Wu. "A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data". In: *Nucleic Acids Research* 42.8 (Feb. 2014), e69–e69. ISSN: 0305-1048. DOI: 10.1093/nar/gku154. URL: http://dx.doi.org/10.1093/nar/gku154.

[6] D. Sun, Y. Xi, B. Rodriguez, H. J. Park, P. Tong, M. Meong, et al. "MOABS: model based analysis of bisulfite sequencing data". In: *Genome Biology* 15.2 (Feb. 2014). ISSN: 1474-760X. DOI: 10.1186/gb-2014-15-2-r38. URL: http://dx.doi.org/10.1186/gb-2014-15-2-r38.

[7] M. D. Robinson, A. Kahraman, C. W. Law, H. Lindsay, M. Nowicka, L. M. Weber, et al. "Statistical methods for detecting differentially methylated loci and regions". In: *Frontiers in Genetics* 5 (2014). ISSN: 1664-8021. DOI: 10.3389/fgene.2014.00324. URL: http://dx.doi.org/10.3389/fgene.2014.00324.

[8] Y. Park and H. Wu. "Differential methylation analysis for BS-seq data under general experimental design". In: *Bioinformatics* 32.10 (Jan. 2016), pp. 1446–1453. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btw026. URL: http://dx.doi.org/10.1093/bioinformatics/btw026.

[9] M. Kernaleguen, C. Daviaud, Y. Shen, E. Bonnet, V. Renault, J.-F. Deleuze, et al. "Whole-Genome Bisulfite Sequencing for the Analysis of Genome-Wide DNA Methylation and Hydroxymethylation Patterns at Single-Nucleotide Resolution". In: *Epigenome Editing*. Springer New York, 2018, pp. 311–349. ISBN: 9781493977741. DOI: 10.1007/978-1-4939-7774-1_18. URL: http://dx.doi.org/10.1007/978-1-4939-7774-1_18.

[10] N. Chapin, J. Fernandez, J. Poole, and B. Delatte. "Anchor-based bisulfite sequencing determines genome-wide DNA methylation". In: *Communications Biology* 5.1 (2022). ISSN: 2399-3642. DOI: 10.1038/s42003-022-03543-1. URL: http://dx.doi.org/10.1038/s42003-022-03543-1.

[11] A. Morin, E. E. Thompson, B. A. Helling, L. E. Shorey-Kendrick, P. Faber, T. Gebretsadik, et al. "A functional genomics pipeline to identify high-value asthma and allergy CpGs in the human methylome". In: *Journal of Allergy and Clinical Immunology* 151.6 (2023), pp. 1609–1621. ISSN: 0091-6749. DOI: 10.1016/j.jaci.2022.12.828. URL: http://dx.doi.org/10.1016/j.jaci.2022.12.828.

[12] O. S. Meyer, M. M. Andersen, C. Børsting, N. Morling, H. C. Wulf, P. A. Philipsen, et al. "Comparison of global DNA methylation analysis by whole genome bisulfite sequencing and the Infinium Mouse Methylation BeadChip using fresh and fresh-frozen mouse epidermis". In: *Epigenetics* 18.1 (Nov. 2022). ISSN: 1559-2308. DOI: 10.1080/15592294.2022.2144574. URL: http://dx.doi.org/10.1080/15592294.2022.2144574.

[13] J. Wang, Q. Zhao, T. Hastie, and A. B. Owen. "Confounder adjustment in multiple hypothesis testing". In: *The Annals of Statistics* 45.5 (2017), pp. 1863–1894.

[14] C. McKennan and D. Nicolae. "Accounting for unobserved covariates with varying degrees of estimability in high-dimensional biological data". In: *Biometrika* 106.4 (Sept. 2019), pp. 823–840. ISSN: 0006-3444. DOI: 10.1093/biomet/asz037.

[15] C. McKennan and D. Nicolae. "Estimating and Accounting for Unobserved Covariates in High-Dimensional Correlated Data". In: *Journal of the American Statistical Association* 117.537 (June 2020), pp. 225–236. DOI: 10.1080/01621459.2020.1769635. URL: https://doi.org/10.1080/01621459.2020.1769635.

[16] E. Rahmani, N. Zaitlen, Y. Baran, C. Eng, D. Hu, J. Galanter, et al. "Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies". In: *Nature Methods* 13.5 (Mar. 2016), pp. 443–445. ISSN: 1548-7105. DOI: 10.1038/nmeth.3809. URL: http://dx.doi.org/10.1038/nmeth.3809.

[17] E. Rahmani, R. Schweiger, L. Shenhav, T. Wingert, I. Hofer, E. Gabel, et al. "BayesCCE: a Bayesian framework for estimating cell-type composition from DNA methylation without the need for methylation reference". In: *Genome Biology* 19.1 (Sept. 2018). ISSN: 1474-760X. DOI: 10.1186/s13059-018-1513-2. URL: http://dx.doi.org/10.1186/s13059-018-1513-2.

[18] I. Unterman, D. Avrahami, E. Katsman, T. J. Triche, B. Glaser, and B. P. Berman. "CelFiE-ISH: a probabilistic model for multi-cell type deconvolution from single-molecule DNA methylation haplotypes". In: *Genome Biology* 25.1 (June 2024). ISSN: 1474-760X. DOI: 10.1186/s13059-024-03275-x. URL: http://dx.doi.org/10.1186/s13059-024-03275-x.

[19] T. J. Peters, M. J. Buckley, Y. Chen, G. K. Smyth, C. C. Goodnow, and S. J. Clark. "Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate". In: *Nucleic Acids Research* 49.19 (2021), e109–e109. ISSN: 1362-4962. DOI: 10.1093/nar/gkab637. URL: http://dx.doi.org/10.1093/nar/gkab637.

[20] S. Janson. "Probability asymptotics: notes on notation". In: *arXiv preprint arXiv:1108.3924* (2011). arXiv: 1108.3924 [math.PR]. URL: https://arxiv.org/abs/1108.3924.

[21] Y. Eldar and G. Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.