**Supplementary Note 1: Statistical Analyses.**

*1) Estimating sample-specific methylation rates*

We estimated for each cell $j$ at position $i$ the methylation rate $r_{i,j}$. To increase the coverage across cells, we employed a sliding window approach, which is conceptually similar to approaches that have been used for bulk BS-Seq [24,25]. With window size $w = 3000$bp and step size 600bp, we computed the sum of methylated ($c^+_{i,j}$) and unmethylated ($c^-_{i,j}$) read counts in each window:

$$s^+_{i,j} = \sum_{k=-w/2}^{+w/2} c^+_{i+k,j} \qquad s^-_{i,j} = \sum_{k=-w/2}^{+w/2} c^-_{i+k,j}$$

To estimate methylation rates, we modeled the sum $S^+_{i,j}$ of methylated counts as a Binomial random variable with methylation rate $r_{i,j}$:

$$S^+_{i,j} \sim \mathrm{Bin}(s^+_{i,j} + s^-_{i,j}, r_{i,j})$$

Assuming a Beta (1, 1) prior on $r_{i,j}$, leads to the maximum a posteriori estimator for methylation rates for each window and cell:

$$\hat{r}_{i,j} = \frac{s^+_{i,j} + 1}{s^+_{i,j} + s^-_{i,j} + 2}$$

We approximated the standard error of the rate estimator as follows:

$$\mathrm{SE}[\hat{r}_{i,j}]^2 = \frac{\hat{r}_{i,j}(1 - \hat{r}_{i,j})}{s^+_{i,j} + s^-_{i,j}}$$

*2) Estimating mean methylation rates*

We used the estimated sample-specific methylation rates $\hat{r}_{i,j}$ to estimate mean methylation rates and cell-to-cell variances. We modeled the mean methylation rate $r_i$ at position $i$ across all cells as a Gaussian random variable with mean $\bar{r}_i$ and variance $v_i$:

$$r_i \sim N(\bar{r}_i, v_i)$$

To account for differences in the standard errors $\mathrm{SE}[\hat{r}_{i,j}]$, we weighted sample $j$ and position $i$ by $w_{i,j} = \mathrm{SE}[\hat{r}_{i,j}]^{-2}$, and used the weighted maximum likelihood estimator

$$\hat{\bar{r}}_i = \frac{1}{\sum_j w_{i,j}} \sum_j w_{i,j} \hat{r}_{i,j}$$

to estimate $\bar{r}_i$. The corresponding standard error is given by

$$SE[\hat{\bar{r}}_i]^2 = \frac{1}{\sum_j w_{i,j}}.$$

The maximum likelihood estimator of the cell-to-cell methylation variance $v_i$ is

$$\hat{v}_i = \frac{\sum_j w_{i,j}}{(\sum_j w_{i,j})^2 - \sum_j {w_{i,j}}^2} \sum_j w_{i,j} (\hat{r}_{i,j} - \hat{\bar{r}}_i)^2,$$

which is the unbiased weighted sample variance. The chi-squared confidence interval of the variance estimator with confidence level $\alpha$ is

$$\left[\hat{v}_i^l, \hat{v}_i^u\right] = \left[\frac{n_i \hat{v}_i}{\chi^2_{1-\frac{\alpha}{2}, n_i}}, \frac{n_i \hat{v}_i}{\chi^2_{\frac{\alpha}{2}, n_i}}\right].$$

Here, $\chi^2_{p,n_i}$ is the $p$-quantile of the chi-squared distribution with $n_i$ degrees of freedom, where $n_i$ is the sum of sample weights:

$${n_i}^2 = \frac{\sum_j w_{i,j}}{(\sum_j w_{i,j})^2 - \sum_j {w_{i,j}}^2}$$

To determine highly variable methylated sites, we ranked these by the lower bound $\hat{v}_i^l$ of the chi-squared confidence interval and defined the top $k$ sites as the most variable

sites. This approach is selecting sites with large estimates of cell to cell variance while penalizing for uncertainty of these estimates, which typically stems from low read counts.

*3) Clustering*

To cluster cells and sites, we considered a complete linkage clustering, and employed the weighted Euclidean norm as distance measure for comparing sample $j$ with sample $j'$:

$$d(j,j') = \sqrt{\sum_{i=1}^{d} w_i^{j,j'} (\hat{r}_{i,j} - \hat{r}_{i,j'})^2}$$

We defined the weight $w_i^{j,j'}$ at position $i$ as

$$w_i^{j,j'} \propto \sqrt{w_{i,j} w_{i,j'}},$$

and normalized weights to sum up to the total number of positions $d$. This distance measure places most emphasis on sites that are well covered in both samples.