The main computational cost of our method relates to the MCMC algorithm. Once the MCMC chains have been produced, downstream analyses such as the detection of highly and lowly variable genes require simple (and fast) post-processing of the output provided by the MCMC algorithm. To assess the computational cost of our method, we performed the analysis of simulated datasets generated from the model described in the manuscript using different numbers of cells $n$ and genes $q$. For each combination of $n$ and $q$, we fixed the number of spike-in genes equal to 50. For each dataset, we ran the MCMC algorithm for $N = 10,000$ iterations. For each combination of $n$ and $q$, average results across 5 repetitions are displayed in Fig S4. It can be seen that the computational complexity of our MCMC algorithm is approximately linear with respect to the number of cells as well as to the number of genes.

Figure S4: Computational cost. Running times (in seconds) required to generate $N = 10,000$ MCMC samples for different numbers of cells and genes. Running times correspond to average results across 5 repetitions and dashed grey lines located at “$x=y$”. It can be seen that the computational complexity of our MCMC algorithm is approximately linear with respect to the number of cells as well as to the number of genes.