

## Supplementary Material

Main document: *A compositional model to assess expression changes from single-cell RNA-seq data*

Authors: Ma, Korthauer, Kendzierski, and Newton

Version: August 6, 2020.

This supplement is organized to match the sectioning of the main document. In summary,

1. Introduction
  - R package scDDboost
2. Modeling
  - 2.1 Data Structure, Sampling Model, and Parameters
    - Illustrating key issue
    - Proof of Theorem 2
  - 2.2 Method Structure and Clustering
    - EBSeq
    - modalClust
    - Randomized distances
    - Selecting  $K$
    - Robustness by randomized distances
  - 2.3 Double Dirichlet Mixture
    - Proof of Properties 1-6 and Theorem 3
3. Numerical Experiments
  - 3.1 Synthetic data, splatter
  - 3.2 Empirical study, conquer and Null case
  - 3.3 Bursting
  - 3.4 Time complexity
  - 3.5 Diagnostics
    - Negative Binomial assumption
    - Constant shape assumption
    - Clustering
4. Posterior consistency
  - Proof of Theorem 4

## 1. Introduction.

1.1. *R package.* R package `scDDboost` has tools for an end-to-end analysis of differential distribution via the proposed methodology. This includes clustering of unlabeled cells, EBSeq to assess probabilities of DE mean patterns among clusters, and final DD posterior probabilities. See <https://github.com/wiscstatman/scDDboost>

## 2. Modeling.

### 2.1. Data Structure, Sampling Model, and Parameters.

**Illustrating key issue.** Consider an example as in Figure 2 [main] wherein there are  $K = 7$  subtypes having proportions  $\phi = (\phi_1, \dots, \phi_7)$  in one condition and different proportions  $\psi = (\psi_1, \dots, \psi_7)$  in the other. Though  $\phi \neq \psi$ , the marginal distribution of data at one gene may not differ between the conditions; this depends on ties that may exist in the seven component distributions. For demonstration's sake, suppose that there are three distinct component distributions among the seven, say  $\alpha(x)$ ,  $\beta(x)$ , and  $\gamma(x)$ , and suppose there are ties:

$$\begin{aligned}\alpha &= f_{g,3} = f_{g,4} \\ \beta &= f_{g,2} = f_{g,5} = f_{g,6} \\ \gamma &= f_{g,1} = f_{g,7}.\end{aligned}$$

These ties represent the lack of differential expression between subtypes within each block, but there are differences between blocks since  $\alpha$ ,  $\beta$ , and  $\gamma$  are all different. For the negative binomial model assumed in main, equalities and inequalities of component distributions correspond to equalities and inequalities of component means, as the shape is assumed constant within each gene. Let's work out the marginal distribution of data in the first condition,  $f_g^1(x)$ , noting the proportion constraints in Figure 2:

$$\begin{aligned}f_g^1(x) &= \sum_{k=1}^7 \phi_k f_{g,k}(x) \\ &= (\phi_3 + \phi_4)\alpha(x) + (\phi_2 + \phi_5 + \phi_6)\beta(x) + (\phi_1 + \phi_7)\gamma(x) \\ &= \Phi_{3,4}\alpha(x) + \Phi_{2,5,6}\beta(x) + \Phi_{1,7}\gamma(x) \\ &= \Psi_{3,4}\alpha(x) + \Psi_{2,5,6}\beta(x) + \Psi_{1,7}\gamma(x) \\ &= f_g^2(x).\end{aligned}$$

To reiterate the key issue, a gene that does not distinguish several subtypes (it has ties in the component distributions) does not distinguish the conditions even if there are different overall proportions, as long as certain aggregate proportions are unaltered.

**PROOF OF THEOREM 2.** Recall the parameters  $\theta = (\phi, \psi, \mu, \sigma)$  and the dependence structure among variables indicated in the directed graph in Figure 3. We observe cell (sample) labels  $y_c$ , which accumulate into sample sizes  $n_1$  and  $n_2$  (numbers of cells within

each condition). Conditionally on these labels (and counts) subtype count vectors  $z^1, z^2$  are multinomial draws given frequencies  $\phi$  and  $\psi$ . Given subtype of a cell, measured expression  $X_{g,c}$  follows a negative binomial distribution with mean depending on the gene and on the cell subtype. Thus  $P(X, y, z|\theta) = P(y, z|\phi, \psi)P(X|z, \mu, \sigma)$ . Further,  $(\mu, \sigma)$  and  $(\phi, \psi)$  are taken independent a priori given pattern  $\pi$ . By Bayes's rule (and always conditioning on  $\pi$ ),

$$\begin{aligned} P(\theta|X, y, z) &\propto P(X, y, z|\theta)P(\theta) \\ P(X, y, z|\theta)P(\theta) &= P(y, z|\phi, \psi)P(X|z, \mu, \sigma)P(\mu, \sigma|z)P(\phi, \psi) \\ P(\phi, \psi|y, z) &\propto P(y, z|\phi, \psi)P(\phi, \psi) \\ P(\mu, \sigma|X, z) &\propto P(X|z, \mu, \sigma)P(\mu, \sigma|z) \\ \text{Thus } P(\theta|X, y, z) &\propto P(\phi, \psi|y, z)P(\mu, \sigma|X, z) \end{aligned}$$

It thus follows by integration over the parameter space that  $P(A_\pi \cap M_{g,\pi}|X, y, z) = P(A_\pi|y, z) P(M_{g,\pi}|X, z)$ .  $\square$

## 2.2. Method Structure and Clustering.

*EBSeq.* Here we recall some key elements from [Leng et al. \(2013\)](#) on the model behind EBSeq, which we adapt to get  $P(M_{g,\pi}|X, z)$ . Suppose we have  $K$  subtypes, let  $X_g^I = X_{g,1}^I, \dots, X_{g,S_1}^I$  denote transcripts at gene  $g$  from subtype  $I, I = 1, \dots, K$ . The EBSeq model assumes that normalized counts within subtype  $I$  are distributed as Negative Binomial:  $X_{g,s}^I|\sigma_g, q_g^I \sim NB(\sigma_g, q_g^I)$  with density  $P(X_{g,s}^I|\sigma_g, q_g^I) = \binom{X_{g,s}^I + \sigma_g - 1}{X_{g,s}^I} (1 - q_g^I)^{\sigma_g} (q_g^I)^{X_{g,s}^I}$ .

Following [Leng et al. \(2013\)](#), we assume a prior distribution on  $q_g^I : q_g^I|\alpha, \beta^g \sim \text{Beta}(\alpha, \beta^g)$ . The hyperparameter  $\alpha$  is shared by the whole genome and  $\beta^g$  is gene-specific. We force the size factor to be 1 for all cells and use the same procedure as EBSeq to estimate the shape parameter  $\sigma_g$ . Namely, we have

1. gene-level sample mean  $m_g = \frac{1}{n} \sum_{s=1}^n X_{g,s}$ , where  $n = n_1 + n_2$  is the total number of cells
2. average of sample variances over subtypes  $v_g = \frac{1}{K} \sum_{I=1}^K v_g^I$ .
3.  $v_g^I$  is the unadjusted sample variance for subtype  $I$ , i.e.  $v_g^I = \frac{1}{n^I} \sum_{s, z_s=I} (X_{g,s} - m_g^I)^2$  where  $m_g^I$  is the sample mean within subtype  $I$  and  $n^I$  is the number of cells within subtype  $I$ .

We estimate the pooled over-dispersion rate by  $o_g = \frac{v_g}{m_g}$  and obtain  $\sigma_g = m_g \frac{o_g}{1-o_g}$  from the first moment of NB. Our aim is to quantify the expression pattern among  $K$  groups:

$$M_{g,\pi} = \{\theta \in \Theta : \mu_{g,k} = \mu_{g,k'} \iff k, k' \in b, b \in \pi\}.$$

For example, if  $K = 3$ , there are 5 expression patterns, which may be written equivalently

in terms of parameters  $q$ :

$$\begin{aligned}
P1 : q_g^1 &= q_g^2 = q_g^3 \\
P2 : q_g^1 &= q_g^2 \neq q_g^3 \\
P3 : q_g^1 &\neq q_g^2 = q_g^3 \\
P4 : q_g^1 &= q_g^3 \neq q_g^2 \\
P5 : q_g^1 &\neq q_g^2 \neq q_g^3 \text{ and } q_g^1 \neq q_g^3
\end{aligned}$$

In a pattern where two groups  $I$  and  $J$  share the same  $q_g$  the counts from these groups are essentially pooled: i.e.  $X_g^{I,J} | \sigma_g, q_g \sim NB(\sigma_g, q_g)$ ,  $q_g | \alpha, \beta^g \sim \text{Beta}(\alpha, \beta^g)$ . The prior predictive function is

$$\begin{aligned}
f(X_g^{I,J}) &= \int_0^1 P(X_g^{I,J} | r_g, q_g) P(q_g | \alpha, \beta^g) dq_g \\
&= \left[ \prod_{s=1}^S \binom{X_{g,s} + \sigma_g - 1}{X_{g,s}} \right] \frac{\text{Beta}(\alpha + \sum_{s=1}^S \sigma_g, \beta^g + \sum_{s=1}^S X_{g,s})}{\text{Beta}(\alpha, \beta^g)}.
\end{aligned}$$

Consequently, the prior predictive function for  $P1, \dots, P5$  takes a convenient form if we further treat the distinct  $q$ 's as independently drawn from the common Beta mixing distribution:

$$\begin{aligned}
h_1^g(X_g) &= f(X_g^{1,2,3}) \\
h_2^g(X_g) &= f(X_g^{1,2})f(X_g^3) \\
h_3^g(X_g) &= f(X_g^1)f(X_g^{2,3}) \\
h_4^g(X_g) &= f(X_g^{1,3})f(X_g^2) \\
h_5^g(X_g) &= f(X_g^1)f(X_g^2)f(X_g^3)
\end{aligned}$$

where  $h_k^g(X_g) = P(X_g | M_{g,\pi_k}, z)$  for the associated pattern  $\pi_k$ . Then the marginal distribution of count vector  $X_g$  is  $\sum_{k=1}^5 p_k h_k^g(X_g)$ , where the mixing mass  $p_k = P(M_{g,\pi_k} | z)$  is shared by all genes. Then, the posterior probability of an expression pattern  $k$  is obtained by:

$$\frac{p_k h_k^g(X_g)}{\sum_{l=1}^5 p_l h_l^g(X_g)}.$$

In the optimization for determining the hyperparameters  $(\alpha, \beta^g, p)$ , we use EM for the mixing proportions and we use in each cycle a single gradient ascent step for  $\alpha$  and  $\beta^g$ , in contrast to a full root-finding step used by EBSeq.

*modalClust*. We review and extend Dahl’s modal clustering procedure (Dahl (2009)). This extension is part of the default cell clustering method of scDDboost. It operates on data from one gene at a time, and extends to Poisson-distributed observations the modal-clust procedure.

**Product Partition Model (PPM):** Let  $X = (X_1, X_2, \dots, X_n)$  be a vector of data (say at one gene). Given a partition  $\pi = \{S_1, \dots, S_q\}$ , where elements are disjoint subsets (i.e., blocks) of  $\{1, 2, \dots, n\}$  and  $\bigcup_{i=1}^q S_i = \{1, 2, \dots, n\}$ , a PPM for  $X$  entails

$$p(X|\pi) = \prod_{i=1}^q f(X_{S_i})$$

where  $X_{S_i}$  is the vector of observations corresponding to the items of component  $S_i$ . The component likelihood  $f(X_S)$  is defined for any non-empty component  $S$  and can take many forms. The partition  $\pi$ , which clusters cells, is the parameter we are interested in. Other parameters that may have been involved in the model are integrated out. (Note the partition here is not directly related to the partition of subtypes, as, e.g. in Figure 3.)

When the prior distribution for a partition  $\pi$  also takes a product form then so does the posterior. We aim to compute the MAP partition (maximizing the posterior  $p(\pi|X) \propto p(X|\pi)p(\pi)$ ) to be used as an initial estimated clustering. Dahl (2009) demonstrated that by some choice of  $f$  and prior of  $\pi$ , we can reduce the time complexity of finding the MAP partition to  $O(n^2)$ . The crucial condition for  $f$  is the **non-overlapping** condition: if  $X_{S_1}$  and  $X_{S_2}$  are overlapped in the sense that  $\min\{X_{S_2}\} < \max\{X_{S_1}\} < \max\{X_{S_2}\}$  or  $\min\{X_{S_1}\} < \max\{X_{S_2}\} < \max\{X_{S_1}\}$ , let  $X_{S_1^*}$  and  $X_{S_2^*}$  be the sets of swapping one pair of those overlapped terms and keep the other unchanged. Then  $f(X_{S_1})f(X_{S_2}) \leq f(X_{S_1^*})f(X_{S_2^*})$ . Here we confirm the non-overlapping condition for Poisson-Gamma observations.

Under the non-overlapping condition of density kernel  $f$ , the MAP partition  $\pi$  satisfies that for any two blocks  $b_1, b_2 \in \pi$ , either  $\max_{i \in b_1}(X_i) \leq \min_{j \in b_2}(X_j)$  or  $\min_{i \in b_1}(X_i) \geq \max_{j \in b_2}(X_j)$ . Thus we reduce the solution space and reduce the time complexity. In the Poisson-Gamma model we have:

$$\begin{aligned} X_i|\pi, \lambda &\sim \text{Poisson}(X_i|\lambda_1 \mathbf{I}\{i \in S_1\} + \dots + \lambda_q \mathbf{I}\{i \in S_q\}) \\ \pi &\sim p(\pi) \\ \lambda_j &\sim \text{Gamma}(\alpha_0, \beta_0) \end{aligned}$$

where  $p(\pi) \propto \prod_{i=1}^q \eta_0 \Gamma(|S_i|)$ . Integrate out  $\lambda$ ,  $f(X_S)$  is obtained as:

$$f(X_S) = \frac{\beta^\alpha}{(|S| + \beta)^{\sum_{i \in S} X_i + \alpha}} \frac{\Gamma\left(\sum_{i \in S} X_i + \alpha\right)}{\Gamma(\alpha)} \frac{1}{\prod_{i \in S} X_i}$$

To apply modal-clustering on Poisson-Gamma model, we need to show the kernel  $f(X_S)$  satisfies the non-overlapping condition.

PROOF. If  $X_{S_1}$  and  $X_{S_2}$  are overlapping, without loss of generality, we assume  $\min\{X_{S_2}\} < \max\{X_{S_1}\} < \max\{X_{S_2}\}$ . To resolve the overlapping state, we could either swap the max element of  $S_1$  with the min element of  $S_2$  or the max element of  $S_2$  while keeping the rest unchanged. Let  $S_1^*$  and  $S_2^*$  be the sets formed by swapping the max element of  $S_1$  with the min element of  $S_2$ . Let  $S_1^{**}, S_2^{**}$  be the sets formed by swapping the max element of  $S_1$  with the max element of  $S_2$ .

We need to show at least one of the following holds

- (1)  $f(X_{S_1^*})f(X_{S_2^*}) \geq f(X_{S_1})f(X_{S_2})$
- (2)  $f(X_{S_1^{**}})f(X_{S_2^{**}}) \geq f(X_{S_1})f(X_{S_2})$

Let  $a = \max\{X_{S_1}\}$ ,  $b = \min\{X_{S_2}\}$  and  $c = \max\{X_{S_2}\}$ , we have  $b < a < c$ . Let  $h_1 = \sum_{i \in S_1} X_i - a$  and  $h_2 = \sum_{i \in S_2} X_i - b$ ,  $n_1$  and  $n_2$  are the number of elements in  $S_1$  and  $S_2$ . Then

(1) can be equivalently expressed as

$$\begin{aligned}
 f(X_{S_1^*})f(X_{S_2^*}) &\geq f(X_{S_1})f(X_{S_2}) \\
 &\iff \\
 \frac{\Gamma(h_1 + a + \alpha)}{(n_1 + \beta)^{h_1 + a + \alpha}} \frac{\Gamma(h_2 + b + \alpha)}{(n_2 + \beta)^{h_2 + b + \alpha}} &\leq \frac{\Gamma(h_2 + a + \alpha)}{(n_2 + \beta)^{h_2 + a + \alpha}} \frac{\Gamma(h_1 + b + \alpha)}{(n_2 + \beta)^{h_1 + b + \alpha}} \\
 &\iff \\
 \frac{\Gamma(h_1 + a + \alpha)}{\Gamma(h_1 + b + \alpha)} \frac{\Gamma(h_2 + b + \alpha)}{\Gamma(h_2 + a + \alpha)} &\leq \left( \frac{n_1 + \beta}{n_2 + \beta} \right)^{a-b}
 \end{aligned}$$

The lefthand side of the above formula is  $\text{LHS}_1 = \frac{(h_1 + b + \alpha) \dots (h_1 + a - 1 + \alpha)}{(h_2 + b + \alpha) \dots (h_2 + a - 1 + \alpha)}$  by the property of Gamma function and that  $X_i$  are integers.

Similarly, (2) can be equivalently expressed as

$$\begin{aligned}
 f(X_{S_1^{**}})f(X_{S_2^{**}}) &\geq f(X_{S_1})f(X_{S_2}) \\
 &\iff \\
 \frac{\Gamma(h_2 + c + \alpha)}{\Gamma(h_2 + a + \alpha)} \frac{\Gamma(h_1 + a + \alpha)}{\Gamma(h_1 + c + \alpha)} &\leq \left( \frac{n_2 + \beta}{n_1 + \beta} \right)^{c-a}
 \end{aligned}$$

The left hand side of above formula is  $\text{LHS}_2 = \frac{(h_2 + a + \alpha) \dots (h_2 + c - 1 + \alpha)}{(h_1 + a + \alpha) \dots (h_1 + c - 1 + \alpha)}$ .

If  $h_1 \leq h_2$ , then  $\text{LHS}_1 \leq \left( \frac{h_1 + a - 1 + \alpha}{h_2 + a - 1 + \alpha} \right)^{a-b}$  and  $\text{LHS}_2 \leq \left( \frac{h_2 + c - 1 + \alpha}{h_1 + c - 1 + \alpha} \right)^{c-a}$ .

So if  $\frac{h_1 + a - 1 + \alpha}{h_2 + a - 1 + \alpha} \leq \frac{n_1 + \beta}{n_2 + \beta}$  then (1) holds, if  $\frac{h_2 + c - 1 + \alpha}{h_1 + c - 1 + \alpha} \leq \frac{n_2 + \beta}{n_1 + \beta}$  then (2) holds.

We multiply those two inequalities, and find that the left hand side  $\frac{h_1 + a - 1 + \alpha}{h_2 + a - 1 + \alpha} \times \frac{h_2 + c - 1 + \alpha}{h_1 + c - 1 + \alpha} = \frac{h_1 + a - 1 + \alpha}{h_1 + c - 1 + \alpha} \times \frac{h_2 + c - 1 + \alpha}{h_2 + a - 1 + \alpha} \leq 1$  as  $c > a$  and  $h_1 \leq h_2$ . But the right hand side  $\frac{n_1 + \beta}{n_2 + \beta} \times \frac{n_2 + \beta}{n_1 + \beta} = 1$ . Consequently at least one of (1) and (2) holds.

The proof for case  $h_1 > h_2$  follows similarly. □

*Randomized distances.* scDDboost reports average posterior probabilities, where the averaging is taken over clusterings of cells induced from randomized-distance matrices. A Bayesian argument for this technique is presented in the Appendix. Random distance  $d_{i,j}^* = d_{i,j}/w_{i,j}$ , where  $w_{i,j} = e_i + e_j$  and the  $\{e_i\}$  are random  $\text{Gamma}(a/2, a)$  deviates. Parameter  $a$  is estimated by an empirical Bayes argument, using the deduction

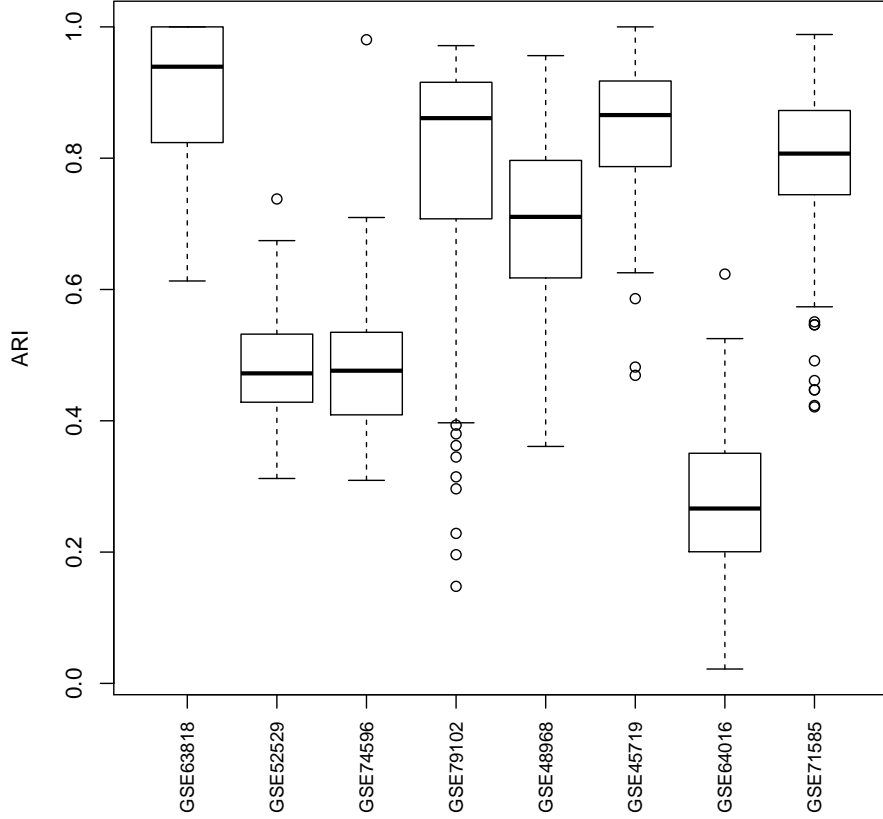
$$(3) \quad P(d_{i,j}|a_0, a_1, d_0) = \frac{\Gamma(a_0 + a_1)}{\Gamma(a_0)\Gamma(a_1)} \frac{d_0^{a_0} d_{i,j}^{a_1-1} a_1^{a_1}}{(d_0 + a_1 d_{i,j})^{a_0+a_1}}$$

for parameters  $a_0, a_1$ , and  $d_0$ , and recognizing  $a = a_0 + a_1$ . We estimate these parameters from the collection of observed distances  $\{d_{i,j}\}$ . First  $d_0$  is estimated by the method of moments, noting  $d_0 = E(1/d_{i,j})/\text{var}[1/d_{i,j}]$ . Next we estimate  $a_0$  and  $a_1$  by optimization of a marginal log-likelihood using (3) and program `nlminb` in R. We limit tolerance to avoid overflow issues. Supplementary Figure S1 shows the level of variation in clusterings produced by this randomized distance approach in 8 example data sets.

We further check validity of randomized distances as a Bayesian approximation by comparing it to Dirichlet-process-based clustering (Jara et al. (2011)) on a simulated dataset. We simulate one-dimensional data  $X$  from a mixture of 5 normal distributions with different means and same variance ( $\mu = (-6, -2, 0, 2, 10), \sigma = 1$ ). We compare clustering results between randomized distances and Bayesian clustering using the Dirichlet process prior (using `DPpackage`) in terms of posterior probabilities that two elements belong to the same class given the whole data. We also compare accuracy of the two procedures by looking at the ARI comparing to true class label (Supplementary Figure S2). We find that randomized distance scheme closely matches the distributional features of the Dirichlet-process computation, and, in this case, tends to put more mass close to the data-generating partition.

*Selecting  $K$ .* In this section, we give the criterion to select the number of subtypes  $K$ . We implement a procedure inspired by validity, as defined in Ray and Turi (2000).

We consider a modified  $\text{validity} = \frac{\text{intra}}{\text{inter}}$ , where  $\text{intra} = \frac{1}{N} \sum_{i=1}^K \sum_{x \in C_i} \|x - z_i\|^2$ ,  $\text{inter} = \text{mean}(\|z_i - z_j\|^2), i, j = 1, 2, \dots, K$ , and  $z_i$  is the center (medoid) of cluster  $i$ . **intra** is the average of distance of a point to the center of its corresponding cluster, which measures the compactness of clusters. **inter** is the average distance between centers, which measures the separation between clusters. In the original paper **inter** was defined as minimum distance between medoids (Ray and Turi, 2000). Here, we instead use the average for smoothness. By minimizing validity (contrary to what the name suggests) we aim for small intra-cluster distance and large inter-cluster distance. We find empirically that validity is monotonic decreasing with  $K$  and this trend stabilizes when  $K$  is sufficiently big. We select the first  $K$  satisfying  $\text{validity}_K < \epsilon$ . We set the default value of  $\epsilon$  to be 1, as we found this yields good performance in simulation.



**Supplementary Figure S1:** Adjusted RAND index of clusterings generated by randomizing distances. The plot shows variation of clustering induced by randomizing distance matrices in 8 data sets and for 100 randomized matrices per data set. In each data set, ARI is computed between the clusterings from all randomized distances and the clustering induced by the non-randomized distance matrix.

*Robustness.* In this section, we demonstrate change of the posterior probability of DD under different  $K$ , and also the robustness provided by randomized distances. We also give an example where very large  $K$  inflates FDR.

The number of subtypes  $K$  is an important parameter. Taking  $K$  too small may end up underfitting such that cells within same subtype can still be very different, the mean expression change among subtypes is incapable to capture the marginal distribution change. This would lead to reduced power. Too large  $K$  may end up overfitting such that two subtypes can be very similar. Given that we have a fixed number of cells, allowing more clusters will not only increase the burden of computation but decrease the certainty of our inference on DE pattern. Empirically we find that taking  $K \leq 10$  is often sufficient (Supplementary Table S1). In any case, we note here that  $K$  affects the posterior probability of DD (PDD).

To demonstrate the change of PDD over different  $K$ , we present an example using dataset GSE75748. When we increase  $K$ , the variance of the differential term  $PDD_{K+1} -$





**Supplementary Figure S2:** Comparison between randomized-distance scheme and Dirichlet-process procedure. Top: heatmap of probabilities that two elements belong to the same class given the whole data. Bottom: scatterplot of these posterior probabilities (left), and adjusted RAND index comparing to the underlying true class label (right).

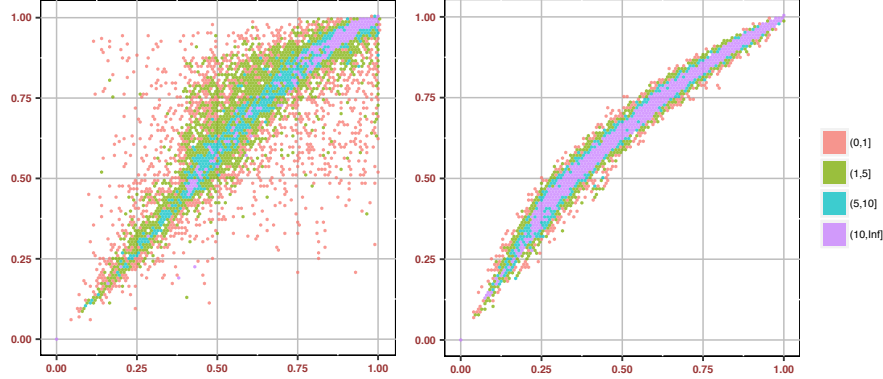
$PDD_K$  keeps decreasing and PDD keeps increasing. Our selection criterion ( $K = 5$ ) happens to choose  $K$  such that change between  $PDD_{K+1}$  and  $PDD_K$  is small while not inflating PDD. We generally obtain stable validity score and PDD simultaneously (Supplementary Figure S3). In addition, the randomized distance scheme helps by smoothing PDD (Supplementary Figure S4).

scDDboost may lose FDR control if  $K$  is not maintained at a sufficiently small value. Supplementary Figure S5 shows what happens as  $K$  increases in one case, other factors staying constant. In our simulation study, we note that the validity score method was

always conservative, and did not lead to overestimating  $K$ .



**Supplementary Figure S3:** PDD change under different number of subtypes  $K$  for dataset DEC-EC (GSE75748). We select  $K = 4$ , which also stabilize the PDD.



**Supplementary Figure S4:** PDD under  $K = 5$  vs.  $K = 6$  for dataset DEC-EC (GSE75748). PDD without randomization (left) vs. PDD with randomization (right). *scDDboost* gained robustness through averaging over randomized distances.

**2.3. Double Dirichlet Mixture.** In this section, we give proofs for the properties of DDM in Section 2.3 of the main paper. Using notation from the main paper, we have density functions:

$$p_{\pi}(\phi, \psi) = q_{\pi}(\Phi_{\pi}, \Psi_{\pi}) \prod_{b \in \pi} [p(\tilde{\phi}_b) p(\tilde{\psi}_b)]$$

with

$$q_{\pi}(\Phi_{\pi}, \Psi_{\pi}) = \frac{\Gamma(\sum_{b \in \pi} \beta_b)}{\prod_{b \in \pi} \Gamma(\beta_b)} \left[ \prod_{b \in \pi} \Phi_b^{\beta_b - 1} \right] 1[\Phi_{\pi} = \Psi_{\pi}]$$

and

$$p(\tilde{\phi}_b) = \frac{\Gamma(\sum_{k \in b} \alpha_k)}{\prod_{k \in b} \Gamma(\alpha_k)} \prod_{k \in b} \tilde{\phi}_k^{\alpha_k - 1}, \quad p(\tilde{\psi}_b) = \frac{\Gamma(\sum_{k \in b} \alpha_k)}{\prod_{k \in b} \Gamma(\alpha_k)} \prod_{k \in b} \tilde{\psi}_k^{\alpha_k - 1}.$$



**Supplementary Figure S5:** FDR control may be lost if  $K$  is too large. Calculations here use dataset EMTAB2805null (Table S2) and target 5% FDR, which is not controlled when  $K > 6$  But by the validity score method we would choose  $K$  much smaller, as that score stabilizes  $K \geq 3$ .

These serve as key components for proving DDM properties.

**PROOF OF PROPERTY 1.** When  $\phi$  and  $\psi$  only satisfy the coarsest constraint:  $\sum_{i=1}^K \phi_i = \sum_{i=1}^K \psi_i = 1$ , then  $\phi$  and  $\psi$  are independently Dirichlet distributed. Finer constraints will lead to dependency between  $\phi$  and  $\psi$  as there is a proper subset  $b$  of  $\pi$  such that  $\sum_{i \in b} \phi = \sum_{i \in b} \psi$ , which make  $P(\phi|\psi) \neq P(\phi)$ .  $\square$

**PROOF OF PROPERTY 2.** By the law of total expectation,  $E_\pi(\phi_k) = E_\pi(E_\pi((\phi_k|\Phi_b))) = E_\pi(E_{\tilde{\phi}_b}(\tilde{\phi}_k)) = E_{\tilde{\phi}_b}(\tilde{\phi}_k)E_\Phi(\Phi_b)$  where  $b$  is the block containing subtype index  $k$ . Since  $\tilde{\phi}_b \sim \text{Dirichlet}_{N(b)}[\alpha_b^1]$  and  $\Phi_\pi \sim \text{Dirichlet}_{N(\pi)}[\beta_\pi]$ , we have  $E_{\tilde{\phi}_b}(\tilde{\phi}_k) = \frac{\alpha_k^1}{\sum_{k' \in b} \alpha_{k'}^1}$ ,  $E_\Phi(\Phi_b) = \frac{\beta_b}{\sum_{b' \in \pi} \beta_{b'}}$  and  $E_\pi(\phi_k) = \frac{\alpha_k^1}{\sum_{k' \in b} \alpha_{k'}^1} \frac{\beta_b}{\sum_{b' \in \pi} \beta_{b'}}$ . The case for  $E_\pi(\psi_k)$  is similar.  $\square$

**PROOF OF PROPERTY 3.**  $t^1/t_\pi^1$  is independent of  $t^2/t_\pi^2$  conditioning on  $t_\pi^1$  and  $t_\pi^2$  by the neutrality property of Dirichlet distribution  $\square$

**PROOF OF PROPERTY 4.** For  $j = 1, 2$ , let  $T_b^j$  be the vector of  $t_k^j$  such that  $k \in b$ . Recall  $t_b^j = \sum_{k \in b} t_k^j$ . Without loss of generality, we consider the case condition  $j = 1$ . At the support of  $p_\pi$ , for different blocks,  $T_b^1|\tilde{\phi}_b$  are mutually independent. Then we have factorization:

$$p_\pi(t^1|t_\pi^1, y) = \prod_{b \in \pi} p(T_b^1|t_b^1, y)$$

and right hand side prior predictive function can be obtained by integrating out  $\tilde{\phi}_b$ . Namely

$$\begin{aligned} p(T_b^1 | t_b^1, y) &= \int_{\tilde{\phi}_b} p(T_b^1 | \tilde{\phi}_b) p(\tilde{\phi}_b) d\tilde{\phi}_b \\ &= \left\{ \left[ \frac{\Gamma(t_b^j + 1)}{\prod_{k \in b} \Gamma(t_k^j + 1)} \right] \left[ \frac{\Gamma(\sum_{k \in b} \alpha_k^j)}{\prod_{k \in b} \Gamma(\alpha_k^j)} \right] \left[ \frac{\prod_{k \in b} \Gamma(\alpha_k^j + t_k^j)}{\Gamma(t_b^j + \sum_{k \in b} \alpha_k^j)} \right] \right\} \end{aligned}$$

given the prior Dirichlet $[\alpha_b^1]$  of  $\tilde{\phi}_b$  and that  $p(T_b^1 | \tilde{\phi}_b)$  is a multinomial( $\tilde{\phi}_b$ ) distribution.  $\square$

PROOF OF PROPERTY 5.  $t_\pi^1$  and  $t_\pi^2$ , given the condition label  $y$ , are independent and identically distributed with  $t_\pi^1 | \Phi \sim \text{multinomial}(\Phi)$ . Thus

$$\begin{aligned} p_\pi(t_\pi^1, t_\pi^2 | y) &= \int_{\Phi} p(t_\pi^1 | \Phi) p(t_\pi^2 | \Phi) p(\Phi) d\Phi \\ &= \left[ \frac{\Gamma(n_1 + 1) \Gamma(n_2 + 1)}{\prod_{b \in \pi} \Gamma(t_b^1 + 1) \Gamma(t_b^2 + 1)} \right] \left[ \frac{\Gamma(\sum_{b \in \pi} \beta_b)}{\prod_{b \in \pi} \Gamma(\beta_b)} \right] \left[ \frac{\prod_{b \in \pi} \Gamma(\beta_b + t_b^1 + t_b^2)}{\Gamma(n_1 + n_2 + \sum_{b \in \pi} \beta_b)} \right]. \end{aligned}$$

As prior of  $\Phi$  is Dirichlet $[\beta]$  and  $n_j = \sum_{b \in \pi} t_b^j$  for  $j = 1, 2$ .  $\square$

To prove Property 6, we need a fact about dimensionality of the intersection of two  $A_\pi$ 's.

LEMMA 1. If  $\pi_2$  is not a refinement of  $\pi_1$  then  $A_{\pi_1} \cap A_{\pi_2}$  is a lower dimensional subset of  $A_{\pi_2}$ .

PROOF OF LEMMA 1. To formalize the problem in linear algebra, we consider the vector space  $R^{2K}$ , and define a map from block to vector in  $R^K$ :  $g(b) = v_b$ , where  $i$ th component of  $v_b$  is 1 if  $i \in b$  and 0 otherwise.

Let  $V_1, V_2$  denote the orthogonal space of  $\phi - \psi$  when  $(\phi, \psi) \in \cap A_{\pi_2}, A_{\pi_2}, A_{\pi_1}$ . Notice that  $\dim(A_{\pi_1} \cap A_{\pi_2}) = \dim(\phi - \psi) + \dim(\psi) = K - \dim(V_1) + K - 1 = 2K - \dim(V_1) - 1$ ,  $\dim(A_{\pi_2}) = 2K - \dim(V_2) - 1$ ,  $\dim(V_2) = N(\pi_2)$ . Assuming  $\pi_1 = \{b_1^1, \dots, b_s^1\}$ , and  $\pi_2 = \{b_1^2, \dots, b_t^2\}$ . The corresponding vectors are  $v_1^1, \dots, v_s^1$  and  $v_1^2, \dots, v_t^2$ . We claim there must be a  $b_i^1 \in \pi_1$  whose corresponding vector  $v_i^1$  is linear independent with  $v_1^2, \dots, v_t^2$ . If not, for every  $v_i^1$  there exists  $\alpha_1^i, \dots, \alpha_t^i$  such that

$$(4) \quad v_i^1 = \sum_{j=1}^t \alpha_j^i v_j^2$$

If  $b_j^2 \cap b_i^1 \neq \emptyset$ , then  $v_i^1 v_j^2 > 0$  and we multiply  $v_j^2$  on both sides of (4), we obtain  $v_i^1 v_j^2 = \alpha_j^i (v_j^2)^2$ , as  $v_p^2 v_q^2 = 0$  if  $p \neq q$ . This implies  $\alpha_j^i > 0$ . Consider  $x = g(b_j^2 \setminus b_i^1)$ . We have  $x v_i^1 = 0$  and multiply  $x$  on both sides of (4) to obtain  $\alpha_j^i v_j^2 x = 0$ . Thus  $x$  must be the zero vector and  $b_j^2 \setminus b_i^1 = \emptyset$ , which implies  $b_j^2 \subset b_i^1$ . That is to say when  $b_j^2 \cap b_i^1 \neq \emptyset$ ,  $b_j^2$  must

be a subset of  $b_i^1$ . So  $b_i^1$  is the union of some blocks in  $\pi_2$ . This implies  $\pi_2$  is a refinement of  $\pi_1$ , which is a contradiction. Consequently, there exists  $b \in \pi_1$  whose  $v_b$  is linear independent with  $v_{b'}, \forall b' \in \pi_2$ . Thus the  $\dim(V_1)$  is at least  $N(\pi_2) + 1, \dim(A_{\pi_1} \cap A_{\pi_2}) < \dim(A_{\pi_2})$ .  $\square$

PROOF OF PROPERTY 6. For any  $\pi$ ,  $P(A_\pi, |y, z) = \sum_{\tilde{\pi} \in \Pi} \int_{A_\pi} \omega_{\tilde{\pi}}^{\text{post}} d\phi d\psi$ , notice the support of  $\omega_{\tilde{\pi}}^{\text{post}}$  is  $A_{\tilde{\pi}}$ . By Lemma 1, we know if  $\tilde{\pi}$  does not refine  $\pi$ , then  $\int_{A_\pi} \omega_{\tilde{\pi}}^{\text{post}} d\phi d\psi$  is an integral on lower dimension set and vanishes. if  $\tilde{\pi}$  refines  $\pi$ , then  $\int_{A_\pi} \omega_{\tilde{\pi}}^{\text{post}} d\phi d\psi = \int_{A_{\tilde{\pi}}} \omega_{\tilde{\pi}}^{\text{post}} d\phi d\psi = \omega_{\tilde{\pi}}^{\text{post}}$ . We have  $P(A_\pi, |y, z) = \sum_{\tilde{\pi} \in \Pi} \omega_{\tilde{\pi}}^{\text{post}} 1[\tilde{\pi} \text{ refines } \pi]$ .  $\square$

PROOF OF THEOREM 3. Recall the DDM prior:  $p(\phi, \psi) = \sum_{\pi \in \Pi} p_\pi(\phi, \psi)$ . By Bayes's rule  $p(\phi, \psi | y, z) \propto p(\phi, \psi, y, z) = \sum_{\pi \in \Pi} p(y, z | \phi, \psi) p_\pi(\phi, \psi) \omega_\pi$  and the 1-1 map from  $(\phi, \psi)$  to  $(\tilde{\phi}, \tilde{\psi}, \Phi)$ , we have

$$p(y, z | \phi, \psi) p_\pi(\phi, \psi) = p(y, z | \tilde{\phi}, \tilde{\psi}, \Phi_\pi) p(\tilde{\phi}) p(\tilde{\psi}) p(\Phi_\pi)$$

when  $(\phi, \psi) \in A_\pi$ . Let us denote right hand side of the above equation as  $U_\pi$ , then

$$U_\pi = \omega_\pi A_1 A_2 A_3 \prod_{k=1}^K (\tilde{\phi}_k)^{t_k^1 + \alpha_k^1} (\tilde{\psi}_k)^{t_k^2 + \alpha_k^2} \prod_{b \in \pi} (\Phi_b)^{t_b^1 + t_b^2 + \beta_b},$$

where  $A_1$  is the product of normalizing terms from multinomial distribution of  $z^1$  and  $z^2$ ,  $A_1 = \frac{\Gamma(n_1+1)\Gamma(n_2+1)}{\prod_{j=1}^2 \prod_{k=1}^K \Gamma(t_k^j+1)}$ , and  $A_2$  is the product of normalizing terms from Dirichlet distribution of  $\tilde{\phi}$  and  $\tilde{\psi}$ ,  $A_2 = \frac{\Gamma(\sum_{k=1}^K \alpha_k^1+1)\Gamma(\sum_{k=1}^K \alpha_k^2+1)}{\prod_{j=1}^2 \prod_{k=1}^K \Gamma(\alpha_k^j+1)}$ , and  $A_3$  is the normalizing term from

Dirichlet distribution of  $\Phi_\pi$ ,  $A_3 = \frac{\Gamma(\sum_{b \in \pi} \beta_b+1)}{\prod_{b \in \pi} \Gamma(\beta_b+1)}$ . Looking at the indices of  $\tilde{\phi}, \tilde{\psi}$  and  $\Phi$ , we can decompose  $U_\pi$  as a product of three Dirichlet densities with a normalizing term. Namely  $U_\pi = C_\pi f_1 f_2 f_3$ , where  $f_1 \sim \text{Dirichlet}[\alpha^1 + t^1]$ ,  $f_2 \sim \text{Dirichlet}[\alpha^2 + t^2]$  and  $f_3 \sim \text{Dirichlet}[\beta + t^1 + t^2]$ . Considering the normalizing factors for densities  $f_1, f_2$  and  $f_3$ , and multiplying them with  $A_1, A_2$  and  $A_3$ , we have  $C_\pi = p_\pi(t^1 | t_\pi^1, y) p_\pi(t^2 | t_\pi^2, y) p_\pi(t_\pi^1, t_\pi^2 | y) \omega_\pi$ . Consequently, we have

$$(\phi, \psi) | y, z \sim \text{DDM} \left[ \omega^{\text{post}} = (\omega_\pi^{\text{post}}), \alpha^1 + t^1, \alpha^2 + t^2 \right] \quad \text{and} \\ \omega_\pi^{\text{post}} \propto p_\pi(t^1 | t_\pi^1, y) p_\pi(t^2 | t_\pi^2, y) p_\pi(t_\pi^1, t_\pi^2 | y) \omega_\pi.$$

Notice in DDM, we restrict  $\beta = \alpha^1 + \alpha^2$ .

$\square$

### 3. Numerical Experiments.

3.1. *Synthetic Data.* Here we look more closely at the synthetic data from splatter, a state-of-the-art simulation framework to generate scRNA-seq data (?). Gene expression of cells within the same group is simulated through a multi-layer model. The layers consider many factors, including outliers and dropout, library size (sequencing depth), and trended gene-wise dispersion. Internally there are Gamma and Poisson elements, but the layering induces marginal distributions that are more complex than the Negative Binomial. Parameters for splatter’s multi-layer model may be set from empirical data sets, and we take the default settings in our experiments. For generating data across multiple subtypes, splatter requires a pair  $(\theta, \gamma)$  of location and scale parameters governing the intensity of DE for those signature genes that distinguish the subtypes. (These parameters are also denoted **de.facLoc** and **de.facScale**.) We follow the discussions at: <https://github.com/Oshlack/splatter-paper/blob/master/analysis/clustering.Rmd> and <https://github.com/Oshlack/splatter-paper/blob/master/analysis/simulations.Rmd>. We pick the two default settings for  $(\theta, \gamma)$ , as splatter argues they gave authentic simulated data under some criterion. We notice that the evidence of a true change is weak on these settings, leading to low power for all methods. So we amplify the differential genes with two additional settings:  $(-0.1, 0.3)$ [default] to  $(-0.1, 1)$  and also  $(0.1, 0.4)$ [default] to  $(0.3, 0.5)$ . These choices enable better resolution for methods comparison.

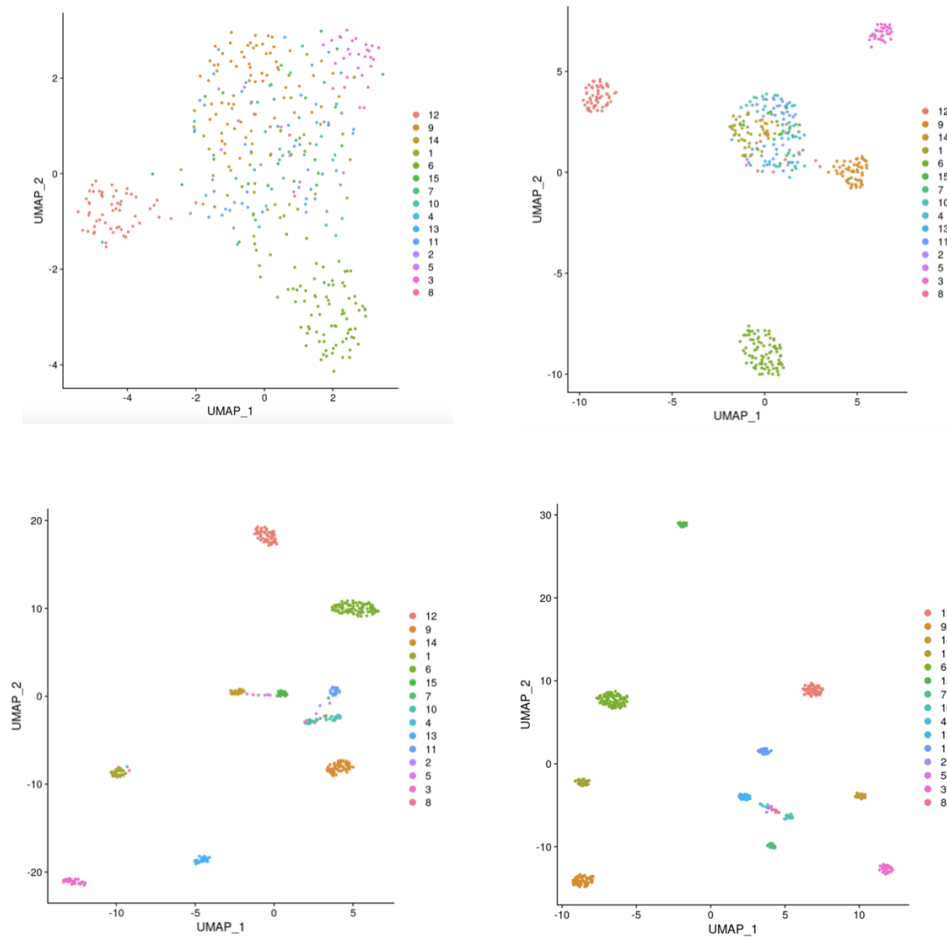
In summary, splatter take three parameters, the number of subtypes  $K$  and a pair of the location and scale parameters for multi-group simulation. To evaluate performance relative to the number of subtypes, we consider  $K \in \{3, 7, 12\}$ . For the location/scale pair, we use two settings suggested by splatter, which are  $(0.1, 0.4)$  and  $(-0.1, 0.3)$  and two cases with stronger signal:  $(0.3, 0.5)$  and  $(-0.1, 1)$ .

For the frequencies of each subtypes, we have different constraint under different  $K$ . Let  $(\phi_i, \psi_i)$  be the frequencies of subtype  $i$  in condition 1 and 2. For  $K = 3$ , there is only one trivial constraint,  $\phi_1 + \phi_2 + \phi_3 = 1 = \psi_1 + \psi_2 + \psi_3$ . For  $K = 7$ , we have  $\phi_1 + \phi_2 = \psi_1 + \psi_2, \phi_3 + \phi_4 + \phi_5 = \psi_3 + \psi_4 + \psi_5, \phi_6 + \phi_7 = \psi_6 + \psi_7$ . And when  $K = 15$ , we have  $\sum_{i=1}^2 \phi_i = \sum_{i=1}^2 \psi_i, \sum_{i=3}^5 \phi_i = \sum_{i=3}^5 \psi_i, \sum_{i=6}^{10} \phi_i = \sum_{i=6}^{10} \psi_i, \sum_{i=11}^{15} \phi_i = \sum_{i=11}^{15} \psi_i$ .

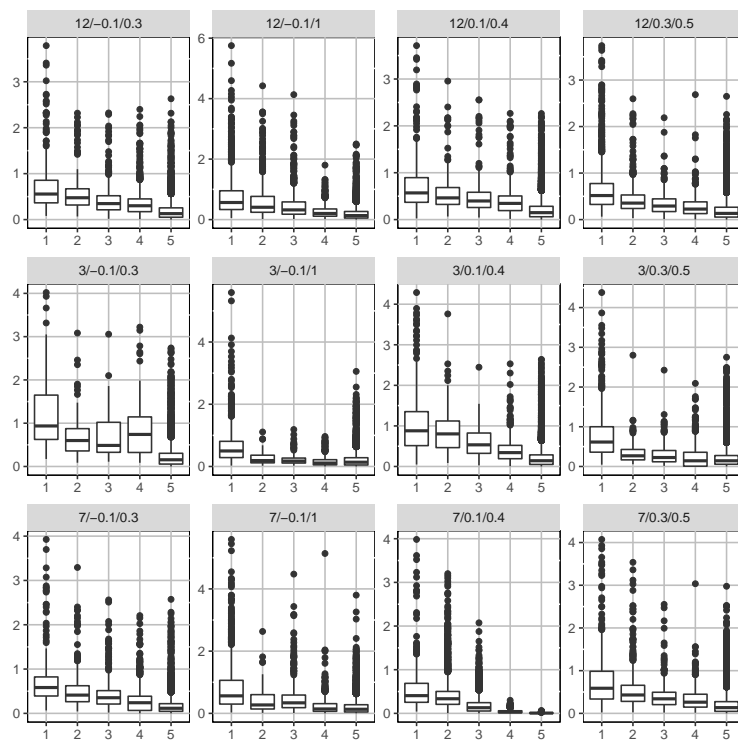
UMAP plots provide a global (transcriptomic) view of changes between splatter-simulated subtypes: Supplementary Figures S6, summarize the case for  $K = 15$  subtypes. Parameter choices provide for some strong and some very weak differences.

Supplementary Figure S7 confirms that the probability scores from scDDboost are associated with simpler measures on the magnitude of distributional change. These plots compare equivalence probabilities to empirical Wasserstein distance (Dobrushin (1972)).

The main paper shows performance of scDDboost on simulated data in terms of false positive rates and power at specific thresholds. Alternatively, we may view it as more of a prediction problem and report receiver-operating-characteristic (ROC), as in Supplementary Figure S8. Each sub-figure corresponds to one of the parameter settings and reports an average over ten replicates under the same parameters setting. scDDboost tends to outperform other methods.



**Supplementary Figure S6:** UMAP of transcripts under different parameters for simulated data. Different parameters resulted in different degree of separation of subtypes. Top left the hyper-parameters are  $(-0.1, 0.3)$ , top right is  $(0.1, 0.4)$ . We can see many subtypes are nested with each other. at the Bottom from left to right are  $(0.3, 0.5)$  and  $(-0.1, 1)$ , where we have stronger signal and subtypes are more separated. We have  $K = 15$  subtypes with cells in each subtype marked with a common color.



**Supplementary Figure S7:**  $P(ED_g|X, y)$  given by scDDboost (horizontal) versus empirical Wasserstein distance (vertical). Genes associated with boxes from left to right having  $P(ED_g|X, y)$  range from 0 - 0.2, 0.2 - 0.4, 0.4 - 0.6, 0.6 - 0.8, 0.8 - 1. Simulation cases have splatter parameters in the format: number of clusters / location / scale





**Supplementary Figure S8:** Roc curve of the 12 simulation settings, under each setting, TPR and FPR are averaged over ten replicates, generally scDDboost performs better than other methods

3.2. *Empirical Study.* In this section, we provide details of the empirical datasets and also demonstrate consistency to Wasserstein distance on one dataset FUCCI ([Leng et al., 2015](#)).

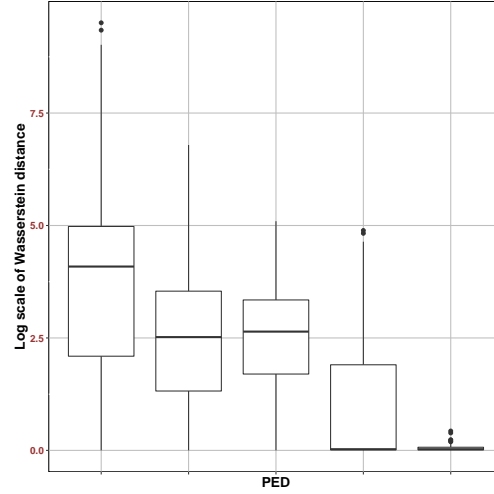
**Data sets** Details for the datasets used in the empirical studies with the estimated number of subtypes  $K$  are shown in Supplementary Table S1.

Data set	Conditions	Number of cells/condition	Organism	Ref	K
GSE94383	0 min unstim vs 75min stim	186,145	human	( <a href="#">Lane et al., 2017</a> )	9
GSE48968-GPL13112	BMDC (2h LPS stimulation) vs 6h LPS	96,96	mouse	( <a href="#">Shalek et al., 2014</a> )	4
GSE52529	T0 vs T72	69,74	human	( <a href="#">Trapnell et al., 2014</a> )	7
GSE74596	NKT1 vs NTK2	46,68	mouse	( <a href="#">Engel et al., 2016</a> )	7
EMTAB2805	G1 vs G2M	95,96	mouse	( <a href="#">Buettner et al., 2015</a> )	6
GSE71585-GPL13112	Gad2tdTpositive vs Cux2tdTnegative	80,140	mouse	( <a href="#">Tasic et al., 2016</a> )	4
GSE64016	G1 vs G2	91,76	human	( <a href="#">Leng et al., 2015</a> )	6
GSE79102	patient1 vs patient2	51, 89	human	( <a href="#">Kiselev et al., 2017</a> )	4
GSE45719	16-cell stage blastomere vs mid blastocyst cell	50, 60	mouse	( <a href="#">Deng et al., 2014</a> )	4
GSE63818	Primordial Germ Cells, developmental stage: 7 week gestation vs Somatic Cells, developmental stage: 7 week gestation	40,26	mouse	( <a href="#">Guo et al., 2015</a> )	6
GSE75748	DEC vs EC	64, 64	human	( <a href="#">Chu et al., 2016</a> )	5
GSE84465	neoplastic cells vs non-neoplastic cells	1000, 1000	human	( <a href="#">Darmanis et al., 2017</a> )	9

SUPPLEMENTARY TABLE S1  
Datasets used for empirical study

For the first 11 datasets in Supplementary Table S1 we use all the cells within that condition under same batch. The last one is the largest dataset we explored containing 3589 cells and comparing neoplastic cells (1091 cells) vs non-neoplastic cells (2498 cells). We randomly sampled 1000 cells from each condition, because it takes a lot of time for DESeq and scDD to compute when using all the samples and we conjecture that 1000 cells each condition would be enough to represent the heterogeneity. We run the comparison on those subsamples instead and found DESeq identified significantly smaller numbers of positives than other methods. It is intuitive that we are more likely to encounter subtle changes when we have large samples, and only considering mean shifts would have limited power.

We also observed consistent distributional change measurements by scDDboost and Wasserstein distance (Supplementary Figure S9).



**Supplementary Figure S9:**  $P(ED_g|X, y)$  given by scDDboost versus empirical Wasserstein distance. Genes associated with boxes from left to right having  $P(ED_g|X, y)$  range from 0 - 0.2, 0.2 - 0.4, 0.4 - 0.6, 0.6 - 0.8, 0.8 - 1, data used: FUCCI

Datasets used for generating the Null cases are shown in Supplementary Table S2.

Data set	Conditions	Number of cells/condition	Organism
GSE63818null	7 week gestation	20,20	mouse
GSE75748null	DEC	32, 32	human
GSE94383null	T0	93, 93	human
GSE48968-GPL13112null	BMDC (2h LPS stimulation)	48,48	mouse
GSE74596null	NKT1	23,23	mouse
EMTAB2805null	G1	48,48	mouse
GSE71585-GPL13112null	Gad2tdTpositive	40,40	mouse
GSE64016null	G1	46,45	human
GSE79102null	patient1	26, 25	human

SUPPLEMENTARY TABLE S2

*Datasets used for null cases, as cells are coming from same biological condition, there should not be any differential distributed genes, any positive call is false positive*

3.3. *Bursting parameters.* Relevant material is reported in the main paper. This empty subsection is included to maintain labeling consistency.

3.4. *Time complexity.* Relevant material is reported in the main paper. This empty subsection is included to maintain labeling consistency.

3.5. *Diagnostics.*

3.5.1. *Negative binomial assumption.* We deploy a Pearson-type goodness-of-fit test (Yin and Ma (2013)) to check the adequacy of the negative binomial (NB) assumption in cell clusters from two data sets (EMTAB2805 and GSE45719). Supplementary Figure S10 presents the cumulative distribution of test p-values under each subtype, and indicates relatively little lack of fit. Among genes that demonstrate non-NB behavior, we find that most are handled the same way by scDDboost as by the standard comparison methods in terms of DD calls at 5% FDR. Supplementary Figures S11, and S12 display data from genes that show non-NB behavior and also are uniquely identified by scDDboost as being DD.

3.5.2. *Constant shape assumption: Likelihood ratio test.* As implemented, scDDboost treats the negative binomial mixture components as having constant shape within each gene. Here we investigate this assumption. Let  $\sigma^k$  be the shape parameter from subtype  $k$ . We have null hypothesis

$$H_0 : \sigma^1 = \sigma^2 = \dots = \sigma^K,$$

the test statistic is  $2(l_1 - l_0)$ , where  $l_1$  is the log likelihood maximized over  $r$  and  $p$  given the data.  $l_0$  is the log likelihood maximized under the  $H_0$  constraint. We use simulation to verify the validity of the test. We simulate 500 random variables that are NB(3,0.5) distributed and another 500 random variables with distribution NB(3,0.7). Then test statistic and corresponding p value is recorded. We repeat the procedure for 100 times. Figure S13 shows that the p value is roughly uniform distributed. We then use likelihood ratio test to identify genes against constant shape assumption and we investigate the positives among those genes. Only a few genes uniquely called positive by scDDboost against constant shape assumption.

3.5.3. *Clustering.* How sensitive are scDDboost results to the particular clustering methodology? To investigate, we applied the sc3 clustering method in place of the default in several numerical experiments. Following the original negative control experiment (main Figure 7), we applied scDDboost on each of five random splits of a number of data sets. In this case, we used five data sets: EMTAB2805, GSE64016, GSE93818, GSE74596, GSE71585. We observed zero positives by scDDboost at a 5% FDR call rate. We also checked scDDboost with sc3 in the simulation study (compare to Figures 4, 6, main). Here 5 replicate data sets were drawn in each setting. Supplementary Figure S14 shows that FDR is well controlled for the null genes.

**4. Posterior consistency.** In this section, we prove Theorem 4 and we discuss a case when condition (12) fails. The density of DDM is computed by product or ratio over



**Supplementary Figure S10:** Cumulative distribution functions for subtype specific unadjusted p values of goodness of fit for NB distribution at two data sets(left is EMTAB2805, right is GSE45719), different colors are correspond to different subtypes

several gamma functions. We use a crucial lemma which gives us an approximation to the gamma function, namely

**LEMMA 2.** For  $x \geq 1$ ,  $\frac{x^{x-c}}{e^{x-1}} \leq \Gamma(x) \leq \frac{x^{x-1/2}}{e^{x-1}}$ , where  $c = 0.577215\dots$  is the Euler-Mascheroni constant.

**PROOF OF LEMMA 2.** By (Li and ping Chen, 2007), we have  $\frac{x^{x-c}}{e^{x-1}} \leq \Gamma(x) \leq \frac{x^{x-1/2}}{e^{x-1}}$  for  $x > 1$  and now we added the case when  $x = 1, \Gamma(x) = 1$  so that both sides will include



**Supplementary Figure S11:** Data set EMTAB2805, violin plot of the 57 genes having evidence against negative binomial distribution assumption and uniquely called positive under `scDDboost`

the equality case. □

We have another lemma.

**LEMMA 3.** *If  $(\phi, \psi) \in A_{\pi_1} \cap A_{\pi_2}$ , follow the conditions in Theorem 1 then*

$$\frac{\omega_{\pi_1}^{post}}{\omega_{\pi_2}^{post}} \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad \text{if } N(\pi_1) < N(\pi_2)$$

**PROOF OF LEMMA 3.** Recall  $\omega_{\pi}^{post} \propto p_{\pi}(t^1 | t_{\pi}^1, y) p_{\pi}(t^2 | t_{\pi}^2, y) p_{\pi}(t_{\pi}^1, t_{\pi}^2 | y) \omega_{\pi}$  and  $\text{RHS} = g(\pi, \alpha, \beta, n_1, n_2) f(\pi, t^1, t^2, \alpha, \beta)$  and  $\frac{\omega_{\pi_1}^{post}}{\omega_{\pi_2}^{post}} = \frac{g(\pi_1, \alpha, \beta, n_1, n_2)}{g(\pi_2, \alpha, \beta, n_1, n_2)} \frac{f(\pi_1, t^1, t^2, \alpha, \beta)}{f(\pi_2, t^1, t^2, \alpha, \beta)}$  where

$$g(\pi, t^1, t^2, \alpha, \beta) = \left[ \prod_{j=1}^2 \prod_{b \in \pi} \prod_{k \in b} \frac{\Gamma(\sum_{k \in b} \alpha_k^j)}{\Gamma(\alpha_k^j)} \right] \frac{\Gamma(n_1 + 1) \Gamma(n_2 + 1)}{\prod_{b \in \pi} \Gamma(\beta_b)} \frac{\Gamma(\sum_{b \in \pi} \beta_b)}{\Gamma(n_1 + n_2 + \sum_{b \in \pi} \beta_b)}$$

$$f(\pi, t^1, t^2, \alpha, \beta) = \left[ \prod_{j=1}^2 \prod_{b \in \pi} \prod_{k \in b} \frac{1}{\Gamma(t_k^j + 1)} \frac{\prod_{k \in b} \Gamma(\alpha_k^j + t_k^j)}{\Gamma(t_b^j + \sum_{k \in b} \alpha_k^j)} \right] \prod_{b \in \pi} \Gamma(\beta_b + t_b^1 + t_b^2)$$



**Supplementary Figure S12:** Data set GSE45719, violin plot of the 13 genes having evidence against negative binomial distribution assumption and uniquely called positive under scDDboost

For notational simplicity, we use the abbreviation  $g(\pi), f(\pi)$  to substitute

$g(\pi, \alpha, \beta, n_1, n_2), f(\pi, t^1, t^2, \alpha, \beta)$ . We take log on  $\frac{\omega_{\pi_1}^{\text{post}}}{\omega_{\pi_2}^{\text{post}}}$ , denote it as LR.  $\text{LR} = \log g(\pi_1) - \log g(\pi_2) + \log f(\pi_1) - \log f(\pi_2)$ . Denote  $C(\pi_1, \pi_2, \alpha, \beta) = \log g(\pi_1) - \log g(\pi_2)$ ,  $C(\pi_1, \pi_2, \alpha, \beta)$  does not change with sample size  $n_1, n_2$  and is a constant determined by partition  $\pi_1, \pi_2$  and hyper parameters  $\alpha, \beta$ . For further convenience of notation let  $h(x) = \log \Gamma(x)$  and  $\gamma_b^j = \sum_{k \in b} \alpha_k^j$ . Denote  $R(\pi_1, \pi_2, t^1, t^2, \alpha, \beta) = \log f(\pi_1) - \log f(\pi_2)$ . And removing the common part of  $f(\pi_1)$  and  $f(\pi_2)$ , we have

$$R(\pi_1, \pi_2, t^1, t^2, \alpha, \beta) = d(\pi_1, t^1, t^2, \alpha, \beta) - d(\pi_2, t^1, t^2, \alpha, \beta)$$

where

$$d(\pi, t^1, t^2, \alpha, \beta) = \sum_{b \in \pi} h(\beta_b + t_b^1 + t_b^2) - \sum_{j=1}^2 \sum_{b \in \pi} h(t_b^j + \gamma_b^j)$$

Recall  $\beta_b = \gamma_b^1 + \gamma_b^2$  and from Lemma 2,  $(x - c) \log(x) - x + 1 \leq h(x) \leq (x - 1/2) \log(x) - x + 1$  we have

(5)

$$d(\pi, t^1, t^2, \alpha, \beta) \geq \sum_{b \in \pi} (\beta_b + t_b^1 + t_b^2 - c) \log(\beta_b + t_b^1 + t_b^2) - \sum_{j=1}^2 \sum_{b \in \pi} (t_b^j + \gamma_b^j - 1/2) \log(t_b^j + \gamma_b^j) + N(\pi)$$

(6)

$$d(\pi, t^1, t^2, \alpha, \beta) \leq \sum_{b \in \pi} (\beta_b + t_b^1 + t_b^2 - 1/2) \log(\beta_b + t_b^1 + t_b^2) - \sum_{j=1}^2 \sum_{b \in \pi} (t_b^j + \gamma_b^j - c) \log(t_b^j + \gamma_b^j) + N(\pi)$$



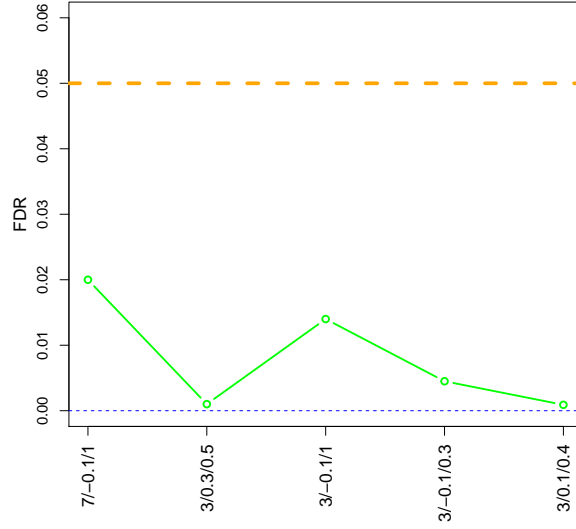
**Supplementary Figure S13:** cumulative distribution for p values of LRT

$$\begin{aligned} \text{RHS of (4)} &= \Sigma_b \left[ (t_b^1 + \gamma_b^1) \log\left(1 + \frac{t_b^2 + \gamma_b^2}{t_b^1 + \gamma_b^1}\right) + (t_b^2 + \gamma_b^2) \log\left(1 + \frac{t_b^1 + \gamma_b^1}{t_b^2 + \gamma_b^2}\right) \right. \\ &\quad \left. + (1 - c) \log(\beta_b + t_b^1 + t_b^2) - 1/2 \left( \log\left(1 + \frac{t_b^2 + \gamma_b^2}{t_b^1 + \gamma_b^1}\right) + \log\left(1 + \frac{t_b^1 + \gamma_b^1}{t_b^2 + \gamma_b^2}\right) \right) \right] + N(\pi) \end{aligned}$$

By Taylor expansion at  $x = 1$ ,  $\log(x + 1) = \log(2) + 1/2(x - 1) - 1/8(x - 1)^2 + g(\xi)(x - 1)^3$ , where  $g(\xi)$  is the reminder term of form  $\frac{1}{3(1+\xi)^3}$  for  $0 < \xi < x$ . For a fixed  $n_1, n_2$ , we have

$$\begin{aligned} \text{RHS of (4)} &= (n_1 + n_2) \log(2) - \Sigma_{b \in \pi} (1/8(X_b^1 + X_b^2) \\ &\quad + g(\xi_b)(Y_b^1 + Y_b^2)) + T(\pi) + N(\pi) \end{aligned}$$





**Supplementary Figure S14:** FDR for simulation data sets, when using scDDboost with sc3 clustering

where  $X_b^1 = \frac{(t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)^2}{t_b^1 + \gamma_b^1}$ ,  $X_b^2 = \frac{(t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)^2}{t_b^2 + \gamma_b^2}$ ,  $Y_b^1 = \frac{(t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)^3}{(t_b^1 + \gamma_b^1)^2}$ ,  $Y_b^2 = \frac{(t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)^3}{(t_b^2 + \gamma_b^2)^2}$   
and  $T(\pi) = \sum_{b \in \pi} [(1 - c) \log(\beta_b + t_b^1 + t_b^2) - 1/2(\log(1 + \frac{t_b^2 + \gamma_b^2}{t_b^1 + \gamma_b^1}) + \log(1 + \frac{t_b^1 + \gamma_b^1}{t_b^2 + \gamma_b^2}))]$   
Similarly

$$\begin{aligned} \text{RHS of (5)} &= (n_1 + n_2) \log(2) - \sum_{b \in \pi} (1/8(X_b^1 + X_b^2) \\ &\quad + g(\xi_b)(Y_b^1 + Y_b^2)) + U(\pi) + N(\pi) \end{aligned}$$

$$U(\pi) = \sum_{b \in \pi} [(2c - 1/2) \log(\beta_b + t_b^1 + t_b^2) - c(\log(1 + \frac{t_b^2 + \gamma_b^2}{t_b^1 + \gamma_b^1}) + \log(1 + \frac{t_b^1 + \gamma_b^1}{t_b^2 + \gamma_b^2}))]$$

Using above inequalities, we have

$$\begin{aligned} R(\pi_1, \pi_2, t^1, t^2, \alpha, \beta) &\leq U(\pi_1) - T(\pi_2) - 1/8(\sum_{b \in \pi_1} (X_b^1 + X_b^2) - \sum_{b \in \pi_2} (X_b^1 + X_b^2)) \\ &\quad + \sum_{b \in \pi_1} g(\xi_b)(Y_b^1 + Y_b^2) - \sum_{b \in \pi_2} g(\xi_b)(Y_b^1 + Y_b^2) \end{aligned}$$

$Y_b^j = \frac{((t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)/\sqrt{n})^3/\sqrt{n}}{((t_b^j + \gamma_b^j)/n)^2}$ , by LLN the denominator goes to a constant and by CLT

in the numerator  $(t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)/\sqrt{n} \rightarrow (t_b^1 - t_b^2)/\sqrt{n} \rightarrow \sqrt{n}[(t_b^1/n - \Phi_b) - (t_b^2/n - \Psi_b)]$ , which converges to a normally distributed random variable when  $\Phi_b = \Psi_b$ . So  $Y_b^j$  is  $o_p(1)$ . Similarly,  $X_b^j = \frac{((t_b^1 - t_b^2 + \gamma_b^1 - \gamma_b^2)/\sqrt{n})^2}{t_b^j + \gamma_b^j/n}$  is asymptotically gamma ( $\chi$ -square) distributed.  $g(\xi_b)$  has bounded variance,  $U(\pi_1) - T(\pi_2) = -\log(n)$  if  $N(\pi_2) < N(\pi_1)$  as  $\log(\beta_b + t_b^1 + t_b^2) - \log(\beta_{b'} + t_{b'}^1 + t_{b'}^2) = \log(\frac{\beta_b + t_b^1 + t_b^2}{n}) - \log(\frac{\beta_{b'} + t_{b'}^1 + t_{b'}^2}{n}) \rightarrow O(1)$  a.s., which completes the proof.

□

PROOF OF THEOREM 4. Recall  $\sum_{\pi \in \Pi} \omega_{\pi}^{\text{post}} = 1$  and  $P(A_{\pi}|y, z) = \sum_{\tilde{\pi} \in \Pi} \omega_{\tilde{\pi}}^{\text{post}} 1[\tilde{\pi} \text{ refines } \pi]$ . If  $(\phi, \psi) \notin Q$ , for all the  $A_{\pi}$  covers  $(\phi, \psi)$  there is one finest  $\pi^*$  with the largest  $N(\pi^*)$  and every other  $\pi$  that  $(\phi, \psi) \in A_{\pi}$  is coarser than  $\pi^*$ . Theorem 4 now follows by Lemma 3.

□

Under some choices of  $(\phi, \psi)$ , condition (12) could fail.



**Supplementary Figure S15:** Four subtypes of cells, simplexes of  $(\phi, \psi)$  satisfying different constraints.

In Supplementary Figure S15, there are four subtypes, the rectangle with magenta boundary is a simplex  $A_{\pi_1} = \{(\phi, \psi) : \phi_1 + \phi_2 = \psi_1 + \psi_2\}$ , the rectangle with blue boundary is another simplex  $A_{\pi_2} = \{(\phi, \psi) : \phi_1 + \phi_3 = \psi_1 + \psi_3\}$ . The green line refers to  $A_{\pi_3} = \{(\phi, \psi) : \phi_1 = \psi_1, \phi_2 = \psi_2\}$ , the yellow line refers to  $A_{\pi_4} = \{(\phi, \psi) : \phi_1 = \psi_1, \phi_3 = \psi_3\}$ . the purple line refers to  $O = \{(\phi, \psi) : \phi_1 + \phi_2 = \psi_1 + \psi_2, \phi_1 + \phi_3 = \psi_1 + \psi_3\}$ , which is the intersection of  $A_{\pi_1}$  and  $A_{\pi_2}$ , and finally the black dot which is the intersection of those three lines refers to the simplex with finest partitions,  $\phi_i = \psi_i, \forall i = 1, \dots, 4$ . When  $(\phi, \psi)$  is from the purple line except the black dot, condition (12) would fail as there is not a finest  $\pi^*$  of  $H(\phi, \psi)$ . This may be of theoretical interest, but the practical implications of this finding are negligible as further computations have demonstrated.

## References.

- BUETTNER, F., NATARAJAN, K. N., CASALE, F. P., PROSERPIO, V., SCIALDONE, A., THEIS, F. J., TEICHMANN, S. A., MARI-  
ONI, J. C. and STEGLE, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing  
data reveals hidden subpopulations of cells. *Nature Biotechnology* **33** 155 EP -.
- CHU, L.-F., LENG, N., ZHANG, J., HOU, Z., MAMOTT, D., VEREIDE, D. T., CHOI, J., KENDZIORSKI, C., STEWART, R. and  
THOMSON, J. A. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation  
to definitive endoderm. *Genome Biology* **17** 173. .

- DAHL, D. B. (2009). Modal clustering in a class of product partition models. *Bayesian Anal.* **4** 243–264.
- DARMANIS, S., SLOAN, S. A., CROOTE, D., MIGNARDI, M., CHERNIKOVA, S., SAMGHABABI, P., ZHANG, Y., NEFF, N., KOWARSKY, M., CANEDA, C., LI, G., CHANG, S. D., CONNOLLY, I. D., LI, Y., BARRES, B. A., GEPHART, M. H. and QUAKE, S. R. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell reports* **21** 1399–1410.
- DENG, Q., RAMSKÖLD, D., REINIUS, B. and SANDBERG, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science* **343** 193–196.
- DOBRUSHIN, R. L. (1972). Asymptotic behavior of Gibbsian distributions for lattice systems and its dependence on the form of the volume. *Theoretical and Mathematical Physics* **12** 699–711.
- ENGEL, I., SEUMOIS, G., CHAVEZ, L., SAMANIEGO-CASTRUITA, D., WHITE, B., CHAWLA, A., MOCK, D., VIJAYANAND, P. and KRONENBERG, M. (2016). Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. *Nature Immunology* **17** 728 EP -.
- GUO, F., YAN, L., GUO, H., LI, L., HU, B., ZHAO, Y., YONG, J., HU, Y., WANG, X., WEI, Y., WANG, W., LI, R., YAN, J., ZHI, X., ZHANG, Y., JIN, H., ZHANG, W., HOU, Y., ZHU, P., LI, J., ZHANG, L., LIU, S., REN, Y., ZHU, X., WEN, L., GAO, Y. Q., TANG, F. and QIAO, J. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* **161** 1437–1452.
- JARA, A., HANSON, T., QUINTANA, F., MÜLLER, P. and ROSNER, G. (2011). DPpackage: Bayesian Semi- and Nonparametric Modeling in R. *Journal of Statistical Software* **40** 1–30.
- KISELEV, V. Y., KIRSCHNER, K., SCHAUB, M. T., ANDREWS, T., YIU, A., CHANDRA, T., NATARAJAN, K. N., REIK, W., BARAHONA, M., GREEN, A. R. and HEMBERG, M. (2017). SC3: consensus clustering of single-cell RNA-seq data. *Nature Methods* **14** 483 EP -.
- LANE, K., VAN VALEN, D., DEFELICE, M. M., MACKLIN, D. N., KUDO, T., JAIMOVICH, A., CARR, A., MEYER, T., PE'ER, D., BOUTET, S. C. and COVERT, M. W. (2017). Measuring Signaling and RNA-Seq in the Same Cell Links Gene Expression to Dynamic Patterns of NF-B Activation. *Cell Systems* **4** 458–469.e5.
- LENG, N., DAWSON, J. A., THOMSON, J. A., RUOTTI, V., RISSMAN, A. I., SMITS, B. M. G., HAAG, J. D., GOULD, M. N., STEWART, R. M. and KENDZIORSKI, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29** 1035–1043.
- LENG, N., CHU, L.-F., BARRY, C., LI, Y., CHOI, J., LI, X., JIANG, P., STEWART, R. M., THOMSON, J. A. and KENDZIORSKI, C. (2015). Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments. *Nature Methods* **12** 947 EP -.
- LI, X. and PING CHEN, C. (2007). Inequalities for the gamma function. In (2007), Art. 28. [ONLINE: <http://ijipam.vu.edu.au/article.php?sid=842>].
- RAY, S. and TURI, R. H. (2000). Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation.
- SHALEK, A. K., SATIJA, R., SHUGA, J., TROMBETTA, J. J., GENNERT, D., LU, D., CHEN, P., GERTNER, R. S., GAUBLomme, J. T., YOSEF, N., SCHWARTZ, S., FOWLER, B., WEAVER, S., WANG, J., WANG, X., DING, R., RAYCHOWDHURY, R., FRIEDMAN, N., HACHOEN, N., PARK, H., MAY, A. P. and REGEV, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510** 363 EP -.
- TASIC, B., MENON, V., NGUYEN, T. N., KIM, T. K., JARSKY, T., YAO, Z., LEVI, B., GRAY, L. T., SORESENSEN, S. A., DOLBEARE, T., BERTAGNOLLI, D., GOLDY, J., SHAPOVALOVA, N., PARRY, S., LEE, C., SMITH, K., BERNARD, A., MADISEN, L., SUNKIN, S. M., HAWRYLYCZ, M., KOCH, C. and ZENG, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* **19** 335 EP -.
- TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S., MORSE, M., LENNON, N. J., LIVAK, K. J., MIKKELSEN, T. S. and RINN, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32** 381–386.
- YIN, G. and MA, Y. (2013). Pearson-type goodness-of-fit test with bootstrap maximum likelihood estimation. *Electronic journal of statistics* **7** 412–427.