

Supplementary Text and Tables for DURIAN: an integrative deconvolution and imputation method for robust signaling analysis of single-cell transcriptomics data

Matthew K. Karikomi^{1,3}, Peijie Zhou^{1,3}, and Qing Nie^{1,2,3,*}

¹Department of Mathematics, University of California, Irvine, Irvine, CA 92697

²Department of Cell and Developmental Biology, University of California, Irvine, Irvine, CA 92697

³NSF-Simons Center for Multiscale Cell Fate Research, University of California

*To whom correspondence should be addressed: Email (PZ): peijiez1@uci.edu, Email (QN): qnie@math.uci.edu

Supplemental Methods

DURIAN algorithm

Imputation of single-cell dropout using bulk (or pseudobulk) transcriptomic data in DURIAN includes the following steps: (1) data pre-processing; (2) deconvolution of bulk data; (3) update of the imputation objective; (3) imputation of the dropout reads. Below we describe each of these steps.

Data pre-processing

Prior to analysis, outlier single-cell and bulk libraries are removed, where outliers are defined as library sizes outside 2 standard deviations of mean (within each data type). Then, each bulk and single-cell sample is scaled to $1e6$ (CPM scaling). During each deconvolution step, genes are analyzed which are expressed in at least 5 cells or 0.5% of all cells (whichever is greater).

We recognize that, while the best case scenario for DURIAN usage involves experiments that produce paired bulk and single-cell data, users of

our tool may be forced to pair datasets from different sources. For well-characterized and/or well-differentiated systems, such as the adult pancreas, published data from different sources may be paired with minimal consideration toward pre-processing. For other tissues, such as samples from human atopic dermatitis (ie eczema), whole-organism involvement with dermatologic, immunological, and neurological disruption introduces constraints which prevent naive pairing of sequencing from separate studies.

We employ two alternative feature selection strategies (fig. S2), depending on the reproducibility of single cell labels, which DURIAN treats as prior-knowledge: 1) Seurat variance stabilized transformation and subsequent CPM scaling, both standard approaches, 2) Use of the DESeq2 differential gene expression model with the following design formula: $X = \text{modality} + \text{group}$, where "X" are the number of detected reads in the sample, "modality" is a nuisance variable corresponding to the original sequencing strategy (either bulk or single cell converted to pseudobulk), and "group" is a biological condition (eg disease/healthy) present in both bulk and single-cell data sets. In order to employ (2), the single cell data from each of N donors is averaged to create N pseudobulk samples, and these are analyzed alongside the corresponding bulk data. The differentially expressed genes corresponding to this model are those which predict the biological state of the sample irrespective of whether it was sequenced from a bulk or single cell library.

We devised strategy (2) as a viable alternative to more standard feature selection after discovering cell populations, described by some published single-cell studies, which could not be recovered via deconvolution of condition-matched bulk from other studies. As noted above, this is likely to be a common problem, especially when the single-cell data in question seeks to identify novel disease or developmental cell types, as illustrated in Fig. S2 below for fibroblast subpopulations (FB.1-3), originally identified by He et al (GSE147424). In Fig. S2a, the published cell labels are shown, in contrast to the Seurat clustering using only the 'sequencing-strategy agnostic' genes identified by strategy (2) in Fig. S2 on right. This figure shows that the existence of FB.1-3 may be dependent on experimental conditions not recapitulated in other studies (eg Suarez et al (GSE65832) which was paired with GSE147424 in Fig 4."

Deconvolution

DURIAN alternates between two basic "stages": imputation and deconvolution. Because the imputed data is treated as fully-observed during deconvolution, while the deconvolution (via the deconvolution map) is treated as fully-observed during imputation, these two stages are fully modular and mathematically independent from each other. To illustrate this point, we implement two alternative deconvolution approaches: an LDA-based topic model

(alg. 1), and non-negative least-squares (alg. 2). In addition to demonstrating the adaptability of our model, these two deconvolution approaches offer complementary benefits for single-cell reference data: the regression-based approach is based on the MuSiC deconvolution Wang et al. [2019] algorithm which automatically performs batch correction based on inter-sample variation in the reference data. However, this approach requires multiple biological donor subjects, which are not always available. In contrast, our implementation of LDA makes no assumptions about sample origin, and can be used in cases where only a single biological replicate is available, or when sample origin is unknown.

Deconvolution via Topic-Model

To allow scalable deconvolution of large data sets, we introduce an exact, distributed implementation of stwdLDA that runs faster than the originally-described model (introduced for processing of natural-language text) without sacrificing accuracy. Deconvolution of the bulk data in DURIAN fits a mixture model using a matched single-cell dataset. We define a semi-supervised generative model based on standard latent dirichlet allocation Blei et al. [2003].

In the deconvolution of bulk transcriptomic data, the composition of each sample corresponds to biological celltypes. In this context, the use of a known expression profile serves three purposes: first, it allows DURIAN to be utilized in modular fashion alongside state of the art unsupervised clustering methods, to identify celltypes of interest prior to imputation. Second the use of these celltype profiles imposes a constraint that makes deconvolution and imputation by DURIAN repeatable, in contrast to unsupervised topic models such as NMF (without a sparsity constraint) or standard LDA which suffer from non-identifiability Rodríguez and Walker [2014]; Hoyer [2004]; Theis et al. [2005]. The third benefit of using known celltype expression profiles is that deconvolution is more scaleable on multiple bulk samples, which can be deconvoluted in parallel due to D-separation between the individual bulk samples Hansen et al. [2013], allowing the celltype assignment of reads from distinct bulk samples to be sampled independently. We therefore refer to our implementation of Gibbs sampling for the stwdLDA model as distributed, static topic word distribution LDA (dsLDA).

dsLDA Likelihood and Posterior Distribution

The model for dsLDA is based on the bulk read likelihood, where each read is assigned a ‘celltype’. The entire bulk sample of reads is then a mixture of various celltypes and their ratios are the latent composition of the tissue, aka the “deconvolution”. We can now give the likelihood for a sample of bulk reads \mathbf{w}_m , when the single-cell signature for celltype k is taken as the

concentration parameter β_k on the bulk celltype gene distribution ϕ_k , and Φ is the $G \times K$ matrix representing the single cell references of K celltypes, over G genes:

$$p(\mathbf{w}_m, \mathbf{z}_m, \boldsymbol{\theta}_m | \Phi) = \prod_{i=1}^{N_m} p(w_{m,i} | \phi_{z_{m,i}}) p(z_{m,i} | \boldsymbol{\theta}_m) p(\boldsymbol{\theta}_m | \boldsymbol{\alpha}) \quad (1)$$

In order to perform the deconvolution of W , we sample the latent celltype associated with each read until the Brooks-Gelman criteria are satisfied Brooks and Gelman [1998] and take the MAP estimate of the $\hat{\Theta}$ as the median of the highest density interval of MCMC samples.

The celltype assignment, which is drawn from the latent celltype proportion matrix $\boldsymbol{\theta}_m$, is solely dependent on the local statistics underlying the mixture of celltypes in the current bulk sample: the count $n_m^{(k)}$ of reads in the current sample assigned to celltype k . The second part of the conditional expresses how inference of the celltype giving rise to the current read relies on shared information between samples. The likelihood $p(w_{m,i} = g)$ of a read given celltype k is:

$$p(w_{m,i} = g) = \frac{n_{k,-i}^{(g)} + \beta_{k,g}}{\sum_{g=1}^G n_{k,-i}^{(g)} + \beta_{k,g}} \quad (2)$$

Marginalization of $\boldsymbol{\theta}_m$ allows sampling of the posterior for \mathbf{z}_m via histogram representation of $p(\mathbf{w}_m, \mathbf{z}_m)$ Rohde and Cappé [2011], giving rise to the "collapsed" explicit representation of stwdLDA Hansen et al. [2013]:

$$p(z_{m,i} = k | \mathbf{z}_{-i}, \mathbf{w}) \propto \frac{n_{m,-i}^{(k)} + \alpha_k}{\left[\sum_{k=1}^K n_m^{(k)} + \alpha_k \right] - 1} \cdot \phi_{z_{m,i}} \quad (3)$$

Observation of ϕ_m (the single-cell gene expression profiles) makes the assignment vectors for any two bulk samples $\mathbf{z}_m, \mathbf{z}_{m'} : m \neq m'$ independent from each other. We exploit this so-called d-separation between bulk samples (alg. 1) by sampling all bulk samples simultaneously, greatly accelerating deconvolution compared to previous approaches Hansen et al. [2013].

Deconvolution via W-NNLS

We used the MuSiC implementation of weighted non-negative least-squares Wang et al. [2019] as a high-efficiency alternative to dsLDA. As in eq. 1 above, each column ϕ_k of the reference matrix Φ represents the relative expression levels of all genes in celltype k . The latent celltype proportions $\boldsymbol{\theta}_m$ in eq. 4 are solved via the Lawson-Hanson NNLS algorithm during iterative update of the latent scaling factors $\tilde{\boldsymbol{\alpha}}_m$ Wang et al. [2019]. Due to

its dependence on the variance across bulk samples, $\tilde{\alpha}$ inherently deals with batch effects Wang et al. [2019].

$$\tilde{\theta}_m = \underset{\alpha, \theta \geq 0}{\operatorname{argmin}} \left(\| \operatorname{diag}(\tilde{\alpha}_m)(\tilde{w}_m - \Phi \tilde{\theta}_m) \|_2^2 \right) \quad (4)$$

DURIAN Iterative Algorithm

Our approach alternates between fitting the dsLDA mixture model to compute the mtSCRABBLE objective and optimizing the mtSCRABBLE objective (D-step and the I-step, respectively). During the initial D-step, the unimputed scRNA-seq data used as the signature for the deconvolution of the bulk. The subsequent I-step utilizes A based on this deconvolution, and estimates the imputed scRNA-seq data \tilde{X} . Starting with the next D-step, all deconvolutions Θ_{i+1} utilize the imputed scRNA-seq \tilde{X}_i . During the I-step, the penalty parameter β controls how closely the current imputation is required to match the current deconvolution estimate (fig. S4).

Algorithm 1: DURIAN [dsLDA]

```

Initialize:  $\tilde{X} \leftarrow X$ ;
while  $\tilde{X}$  not converged do
  Update:  $\tilde{\Phi} \leftarrow \tilde{X}, W$ ;
  for each bulk sample  $j$  in parallel do
    Sample  $z_{m_j}$  with global static  $\tilde{\Phi}$ ;
    Update:  $N_{w_j}, N_j$ 
  end
  Update:  $N_{w_j} \leftarrow N_w + \sum_p (N_{w_j} - N_w)$ ;
  Update:  $\tilde{\Theta} \leftarrow N_j, N_w$ ;
  Update:  $\tilde{S} \leftarrow \tilde{\Theta}, W, \tilde{X}$ ;
  Optimize:  $\tilde{X}$ ;
end

```

Algorithm 2: DURIAN [MuSiC]

```

Initialize:  $\tilde{X} \leftarrow X$ ;
while  $\tilde{X}$  not converged do
  Update:  $\tilde{\Phi} \leftarrow \tilde{X}, W$ ;
  while  $\tilde{\alpha}$  not converged do
    Update:  $\tilde{\Theta} \leftarrow \tilde{X}, W, \tilde{\alpha}$ ;
    Optimize:  $\tilde{\alpha}$ ;
  end
  Update:  $\tilde{S} \leftarrow \tilde{\Theta}, W, \tilde{X}$ ;
  Optimize:  $\tilde{X}$ ;
end

```

Empirical Convergence Properties of DURIAN

The framework of DURIAN relies on an iterative scheme: the imputation stage treats the deconvolution map (reflecting the estimated celltype percentages of the bulk) as known, while deconvolution stage treats the current imputation estimate of the single-cell data as known. It thus remains to be shown that during successive DURIAN iterations, the deconvolution and imputation estimates are jointly improving. In Fig. S1, we show a 2-dimensional histogram of the scaled iteration-by-iteration mean error for down-sampled and splatter simulated runs in Figs. 2-3, for all replicates corresponding to the two highest dropout rates in each strategy. Because each simulated replicate contains completely different cells, and each down-sampled replicate contains a completely different subset of the original pancreatic data, we adopted the following scaling approach for both mean and L2 error: $|(y_t - y_T)/(y_1 - y_T)|$, where y_1, y_t, y_T represent the value of the statistic at initiation (unimputed data), at iteration t and at convergence, respectively. For each respective strategy (e.g. down-sampling, dsLDA deconvolution) the histogram in the top facet of Fig. S1a contains binned imputation error for all replicates and the histogram in the bottom facet directly below it contains the binned deconvolution error (RMSE) for the predicted celltype x bulk sample matrix vs known, for the same replicates. Regardless of synthetic data strategy or deconvolution method, these histograms suggest that as DURIAN approaches convergence, each successive round of imputation benefits from the previous round of deconvolution, and vice versa.

Synthetic Single-Cell Data

Synthetic single-cell data was generated with the splatter R package Zap-pia et al. [2017]. Detailed parameterization of the splatter graphical model is provided in our script `generate_splatter_k4_path_batchdrop.R` at author's github repository below.

Imputation Benchmarking

Pairwise Wilcox tests were performed on mean and L2 error statistics, with Benjamini-Hochberg adjustment for false-discovery.

Convergence Metrics

Data were scaled to relative values as described in Results. Outlier values at the 5 and 95 percentiles were removed before the respective continuous color scales were applied.

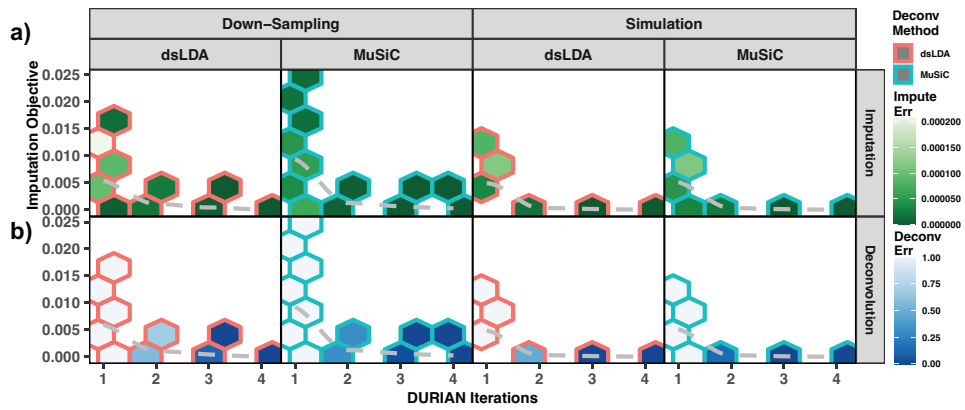


Figure S1: **Imputation and Deconvolution Error Mutually Improve as DURIAN Converges.** a-b) 2-Dimensional histograms of the scaled error (for imputation and deconvolution) for all replicates of the two highest dropout rates reported in figs 2-3. a) Scaled imputation error for DURIAN run with both dsLDA and NNLS (MuSiC) deconvolution strategies, with samples subdivided by synthetic data strategy: down-sampling and simulation. b) Scaled deconvolution error collected during the imputation runs plotted in (a).

Inference and Analysis of Cell Communication Networks

Inference of cell communication networks was performed as described in the CellChat Jin et al. [2021] documentation, available in the CellChat github repository. In particular, the threshold for permutation-based significance testing was left at the default value $p = 0.05$.

Data and Software

Published Software

Software for CellChat, SCRABBLE, URSM, DrImpute, and MuSiC were downloaded from the authors' respective github repositories. Because URSM is based on older Python 2.7 code, we provided updates to allow it to run on our cluster in a fork of the original github repo located in the "slurm/URSM" subdirectory of our github repository linked below. All computational experiments were run on the UCI HPC cluster using slurm.

Run Parameters of Published Software

SCRABBLE parameterization was explicitly noted in all figures and matched that of DURIAN for all common ADMM steps. DrImpute was run at its default parameterization. We adopted the following parameterization for URSM to roughly align its runtime (fig. S3) with other tested methods: number of EM iterations = 5, burn in length = 10, gibbs sample number = 10. Although URSM does not directly output imputed data, we took the scaled posterior gene expression profiles of each read for which the posterior expectation of the binary dropout indicator $A = 1$ Zhu et al. [2018]

Published Data

Mouse embryonic skin single cell and bulk sequencing Gupta et al. [2019]; Biggs et al. [2018] and previously published re-clustering labels Jin et al. [2021] was obtained from the NCBI GEO repository (GSE122043,GSE110459) and from private communication with the authors, respectively. Human adult pancreas single cell sequencing Baron et al. [2016] was obtained from the NCBI GEO repository (GSE84133). Human adult pancreas bulk sequencing Segerstolpe et al. [2016] was obtained from the Array Express repository (E-MTAB-5060). Human adult skin single cell and bulk sequencing He et al. [2020]; Suárez-Fariñas et al. [2015] were obtained from the NCBI GEO repository (GSE147424,GSE65832).

Supplemental Tables

Mean Error Benchmarks

Aggregate by method

	modelname	count	mean	sd
1	mtSCRABBLE 1,1e-06,1e-04	2.950E+02	4.681E+00	3.679E-01
2	DURIAN.MuSiC 1,1e-06,1e-04	3.000E+02	4.687E+00	3.677E-01
3	DURIAN.MuSiC 0.01,1e-05,1e-05	2.850E+02	4.699E+00	3.047E-01
4	mtSCRABBLE 0.01,1e-05,1e-05	2.860E+02	4.703E+00	3.044E-01
5	CMFImpute	2.520E+02	4.886E+00	1.725E-01
6	DURIAN.dsLDA 0.01,1e-05,1e-05	2.940E+02	4.925E+00	5.055E-01
7	DURIAN.dsLDA 1,1e-06,1e-04	2.950E+02	4.928E+00	5.798E-01
8	SCRABBLE 1,1e-06,1e-04	3.000E+02	5.048E+00	9.914E-01
9	DrImpute	2.830E+02	5.063E+00	6.479E-01
10	SCRABBLE 0.01,1e-05,1e-05	2.980E+02	5.157E+00	9.823E-01
11	dropout	3.000E+02	5.169E+00	3.696E-01
12	URSM	2.830E+02	6.330E+00	1.069E+00
13	G2S3	3.000E+02	6.378E+00	1.002E+00

Table S1: Stats for benchmark mean error (RMSE)

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	9.645E-01											
dropout	1.547E-25	6.357E-05										
DURIAN.dsLDA 0.01,1e-05,1e-05	9.045E-01	8.317E-03	3.249E-10									
DURIAN.dsLDA 1,1e-06,1e-04	7.351E-01	4.927E-04	2.581E-04	7.206E-01								
DURIAN.MuSiC 0.01,1e-05,1e-05	1.497E-14	7.684E-10	1.576E-45	5.297E-07	8.259E-04							
DURIAN.MuSiC 1,1e-06,1e-04	2.572E-05	7.343E-17	5.841E-40	5.922E-14	4.632E-10	6.626E-01						
G2S3	2.998E-90	5.872E-57	4.287E-80	3.863E-86	5.985E-87	6.743E-96	8.229E-98					
mtSCRABBLE 0.01,1e-05,1e-05	4.006E-14	1.454E-09	4.264E-45	9.209E-07	9.915E-04	8.982E-01	5.819E-01	5.997E-96				
mtSCRABBLE 1,1e-06,1e-04	8.432E-06	2.511E-17	3.037E-40	2.733E-14	3.207E-10	5.172E-01	9.059E-01	2.737E-97	4.517E-01			
SCRABBLE 0.01,1e-05,1e-05	1.558E-01	5.819E-01	3.255E-12	4.888E-01	3.265E-02	1.393E-07	9.869E-10	1.889E-54	2.958E-07	3.272E-10		
SCRABBLE 1,1e-06,1e-04	1.446E-03	1.084E-04	2.426E-19	1.560E-02	5.819E-01	2.637E-02	6.985E-04	5.483E-59	3.560E-02	3.636E-04	2.204E-03	
URSM	6.592E-88	7.988E-52	6.054E-68	6.195E-72	4.896E-74	3.691E-93	1.819E-95	4.110E-05	2.902E-93	9.083E-95	6.927E-52	1.651E-56

Table S2: Adjusted p-values for paired Wilcox tests on benchmark mean error (RMSE)

Down-Sampled

	strategy	modelname	count	mean	sd
1	Down-Sampling	CMFImpute	1.020E+02	4.914E+00	1.131E-01
2	Down-Sampling	DURIAN.MuSiC 0.01,1e-05,1e-05	1.350E+02	4.975E+00	1.709E-01
3	Down-Sampling	mtSCRABBLE 0.01,1e-05,1e-05	1.360E+02	4.977E+00	1.712E-01
4	Down-Sampling	DURIAN.MuSiC 1,1e-06,1e-04	1.500E+02	5.040E+00	1.087E-01
5	Down-Sampling	mtSCRABBLE 1,1e-06,1e-04	1.450E+02	5.040E+00	1.103E-01
6	Down-Sampling	DURIAN.dsLDA 0.01,1e-05,1e-05	1.440E+02	5.385E+00	2.851E-01
7	Down-Sampling	dropout	1.500E+02	5.430E+00	3.087E-01
8	Down-Sampling	DURIAN.dsLDA 1,1e-06,1e-04	1.500E+02	5.473E+00	2.115E-01
9	Down-Sampling	DrImpute	1.330E+02	5.651E+00	4.657E-01
10	Down-Sampling	SCRABBLE 1,1e-06,1e-04	1.500E+02	5.681E+00	1.072E+00
11	Down-Sampling	SCRABBLE 0.01,1e-05,1e-05	1.480E+02	5.785E+00	1.064E+00
12	Down-Sampling	G2S3	1.500E+02	6.978E+00	1.132E+00
13	Down-Sampling	URSM	1.330E+02	7.115E+00	1.124E+00

Table S3: Stats for benchmark mean error (RMSE) - Strategy:Down-Sampling

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04	URSM
DrImpute	1.294E-36												
dropout	5.035E-37	4.632E-05											
DURIAN.dsLDA 0.01,1e-05,1e-05	5.982E-36	2.931E-07	1.429E-01										
DURIAN.dsLDA 1,1e-06,1e-04	1.843E-39	2.324E-01	1.049E-03	9.483E-07									
DURIAN.MuSiC 0.01,1e-05,1e-05	4.769E-02	4.688E-38	2.290E-35	2.433E-33	2.198E-42								
DURIAN.MuSiC 1,1e-06,1e-04	2.595E-24	1.115E-41	3.193E-40	1.329E-38	3.166E-45	3.562E-07							
G2S3	8.908E-41	1.399E-27	7.322E-46	2.917E-46	6.369E-47	6.445E-47	8.420E-49						
mtSCRABBLE 0.01,1e-05,1e-05	4.006E-02	4.505E-38	3.084E-35	3.370E-33	1.448E-42	9.211E-01	6.947E-07	6.369E-47					
mtSCRABBLE 1,1e-06,1e-04	6.471E-24	7.132E-41	2.293E-39	9.042E-38	1.823E-44	4.304E-07	9.853E-01	2.858E-48	8.009E-07				
SCRABBLE 0.01,1e-05,1e-05	3.453E-23	3.424E-03	1.080E-01	2.568E-01	2.847E-03	1.328E-17	2.598E-09	5.656E-19	2.026E-17	3.364E-09			
SCRABBLE 1,1e-06,1e-04	1.482E-18	6.803E-07	4.331E-05	1.462E-04	1.645E-07	1.081E-12	1.923E-03	3.777E-21	1.788E-12	2.086E-03	9.195E-02		
URSM	6.823E-39	1.367E-30	4.618E-45	6.265E-45	1.088E-45	9.727E-45	1.245E-46	3.578E-02	7.199E-45	5.085E-46	1.926E-19	3.449E-21	

Table S4: Adjusted p-values for paired Wilcox tests on benchmark mean error (RMSE) - Strategy:Down-Sampling

	strategy	mean_dropout	modelname	count	mean	sd
1	Down-Sampling	0.76	DURIAN.MuSiC 0.01,1e-05,1e-05	3.900E+01	4.822E+00	2.633E-02
2	Down-Sampling	0.76	mtSCRABBLE 0.01,1e-05,1e-05	3.900E+01	4.822E+00	2.641E-02
3	Down-Sampling	0.76	CMFImpute	4.900E+01	4.840E+00	3.162E-02
4	Down-Sampling	0.76	DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	4.995E+00	3.424E-02
5	Down-Sampling	0.76	mtSCRABBLE 1,1e-06,1e-04	5.000E+01	4.995E+00	3.424E-02
6	Down-Sampling	0.76	DURIAN.dsLDA 0.01,1e-05,1e-05	4.800E+01	5.131E+00	4.510E-02
7	Down-Sampling	0.76	dropout	5.000E+01	5.140E+00	4.988E-02
8	Down-Sampling	0.76	DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	5.316E+00	4.896E-02
9	Down-Sampling	0.76	DrImpute	5.000E+01	5.435E+00	4.039E-01
10	Down-Sampling	0.76	SCRABBLE 1,1e-06,1e-04	5.000E+01	5.607E+00	1.119E+00
11	Down-Sampling	0.76	SCRABBLE 0.01,1e-05,1e-05	5.000E+01	5.722E+00	1.101E+00
12	Down-Sampling	0.76	G2S3	5.000E+01	6.877E+00	1.119E+00
13	Down-Sampling	0.76	URSM	5.000E+01	6.947E+00	1.087E+00

Table S5: Stats for benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.76

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	4.046E-17											
dropout	4.046E-17	3.506E-07										
DURIAN.dsLDA 0.01,1e-05,1e-05	1.225E-26	2.131E-08	3.728E-01									
DURIAN.dsLDA 1,1e-06,1e-04	4.046E-17	1.578E-02	4.631E-17	4.897E-17								
DURIAN.MuSiC 0.01,1e-05,1e-05	4.632E-03	1.373E-15	1.373E-15	6.247E-24	1.373E-15							
DURIAN.MuSiC 1,1e-06,1e-04	4.631E-17	3.674E-17	6.122E-17	9.968E-17	3.674E-17	1.373E-15						
G2S3	4.046E-17	1.853E-12	3.674E-17	4.789E-17	3.674E-17	1.373E-15	3.674E-17					
mtSCRABBLE 0.01,1e-05,1e-05	5.533E-03	1.373E-15	1.373E-15	6.247E-24	1.373E-15	9.010E-01	1.373E-15	1.373E-15				
mtSCRABBLE 1,1e-06,1e-04	4.631E-17	3.674E-17	6.122E-17	9.968E-17	3.674E-17	1.373E-15	8.775E-01	3.674E-17	1.373E-15			
SCRABBLE 0.01,1e-05,1e-05	1.850E-12	1.887E-01	8.775E-01	9.010E-01	1.253E-01	3.727E-13	1.321E-01	5.149E-07	3.884E-13	1.321E-01		
SCRABBLE 1,1e-06,1e-04	2.070E-12	1.731E-03	1.169E-02	1.656E-02	7.593E-04	1.594E-13	3.235E-01	1.094E-07	1.594E-13	2.983E-01		
URSM	4.046E-17	7.822E-14	3.674E-17	4.789E-17	3.674E-17	1.373E-15	3.674E-17	2.159E-01	1.373E-15	3.674E-17	2.374E-07	7.582E-08

Table S6: Adjusted p-values for paired Wilcoxon tests on benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.76

	strategy	mean	dropout	modelname	count	mean	sd
1	Down-Sampling	0.818		mtSCRABBLE 0.01,1e-05,1e-05	4.800E+01	4.911E+00	6.845E-02
2	Down-Sampling	0.818		DURIAN.MuSiC 0.01,1e-05,1e-05	4.800E+01	4.911E+00	6.850E-02
3	Down-Sampling	0.818		CMFImpute	4.700E+01	4.951E+00	7.024E-02
4	Down-Sampling	0.818		DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	4.999E+00	6.366E-02
5	Down-Sampling	0.818		mtSCRABBLE 1,1e-06,1e-04	5.000E+01	4.999E+00	6.366E-02
6	Down-Sampling	0.818		DURIAN.dsLDA 0.01,1e-05,1e-05	4.900E+01	5.285E+00	9.663E-02
7	Down-Sampling	0.818		dropout	5.000E+01	5.330E+00	9.642E-02
8	Down-Sampling	0.818		DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	5.388E+00	8.964E-02
9	Down-Sampling	0.818		DrImpute	5.000E+01	5.589E+00	3.931E-01
10	Down-Sampling	0.818		SCRABBLE 1,1e-06,1e-04	5.000E+01	5.653E+00	1.091E+00
11	Down-Sampling	0.818		SCRABBLE 0.01,1e-05,1e-05	5.000E+01	5.762E+00	1.074E+00
12	Down-Sampling	0.818		G2S3	5.000E+01	6.942E+00	1.129E+00
13	Down-Sampling	0.818		URSM	5.000E+01	7.031E+00	1.090E+00

Table S7: Stats for benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.818

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	1.207E-16											
dropout	2.185E-16	4.533E-04										
DURIAN.dsLDA 0.01,1e-05,1e-05	5.467E-22	8.092E-08	2.481E-03									
DURIAN.dsLDA 1,1e-06,1e-04	9.834E-17	4.968E-01	5.595E-05	6.372E-10								
DURIAN.MuSiC 0.01,1e-05,1e-05	3.435E-05	7.690E-17	9.929E-17	1.753E-22	6.446E-17							
DURIAN.MuSiC 1,1e-06,1e-04	7.042E-07	6.619E-17	9.929E-17	2.185E-16	6.302E-17	1.247E-12						
G2S3	7.144E-17	2.231E-11	5.010E-17	6.224E-17	5.010E-17	6.302E-17	5.010E-17					
mtSCRABBLE 0.01,1e-05,1e-05	2.918E-05	7.690E-17	9.929E-17	1.753E-22	6.446E-17	9.971E-01	1.064E-12	6.302E-17				
mtSCRABBLE 1,1e-06,1e-04	7.042E-07	6.619E-17	9.929E-17	2.185E-16	6.302E-17	1.247E-12	8.716E-01	5.010E-17	1.064E-12			
SCRABBLE 0.01,1e-05,1e-05	3.826E-07	5.810E-02	2.799E-01	4.831E-01	1.009E-01	4.126E-11	3.728E-04	3.079E-07	4.022E-11	3.728E-04		
SCRABBLE 1,1e-06,1e-04	1.649E-04	7.294E-04	1.844E-03	2.746E-03	1.213E-03	1.599E-09	2.652E-01	9.556E-08	1.599E-09	2.652E-01	2.635E-01	
URSM	7.144E-17	5.314E-13	5.010E-17	6.224E-17	5.010E-17	6.302E-17	5.010E-17	1.678E-01	6.302E-17	5.010E-17	1.323E-07	5.778E-08

Table S8: Adjusted p-values for paired Wilcoxon tests on benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.818

	strategy	mean_dropout	modelname	count	mean	sd
1	Down-Sampling	0.907	DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	5.127E+00	1.385E-01
2	Down-Sampling	0.907	mtSCRABBLE 1,1e-06,1e-04	4.500E+01	5.137E+00	1.416E-01
3	Down-Sampling	0.907	DURIAN.MuSiC 0.01,1e-05,1e-05	4.800E+01	5.163E+00	1.348E-01
4	Down-Sampling	0.907	mtSCRABBLE 0.01,1e-05,1e-05	4.900E+01	5.164E+00	1.327E-01
5	Down-Sampling	0.907	CMFImpute	7.000E+00	5.264E+00	1.389E-01
6	Down-Sampling	0.907	DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	5.716E+00	1.809E-01
7	Down-Sampling	0.907	DURIAN.dsLDA 0.01,1e-05,1e-05	4.700E+01	5.749E+00	1.652E-01
8	Down-Sampling	0.907	SCRABBLE 1,1e-06,1e-04	5.000E+01	5.783E+00	1.018E+00
9	Down-Sampling	0.907	dropout	5.000E+01	5.819E+00	1.641E-01
10	Down-Sampling	0.907	SCRABBLE 0.01,1e-05,1e-05	4.800E+01	5.875E+00	1.030E+00
11	Down-Sampling	0.907	DrImpute	3.300E+01	6.071E+00	3.837E-01
12	Down-Sampling	0.907	G2S3	5.000E+01	7.116E+00	1.157E+00
13	Down-Sampling	0.907	URSM	3.300E+01	7.497E+00	1.173E+00

Table S9: Stats for benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.907

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	7.200E-05											
dropout	4.101E-05	2.082E-02										
DURIAN.dsLDA 0.01,1e-05,1e-05	4.365E-05	1.178E-04	1.043E-02									
DURIAN.dsLDA 1,1e-06,1e-04	4.101E-05	7.074E-06	3.900E-04	9.671E-02								
DURIAN.MuSiC 0.01,1e-05,1e-05	3.223E-02	4.144E-22	7.952E-17	1.552E-24	1.905E-16							
DURIAN.MuSiC 1,1e-06,1e-04	9.382E-03	4.723E-14	4.878E-17	1.271E-16	7.952E-17	2.810E-02						
G2S3	4.101E-05	1.229E-05	1.275E-15	1.041E-15	4.076E-16	7.952E-17	4.878E-17					
mtSCRABBLE 0.01,1e-05,1e-05	3.104E-02	3.467E-22	6.142E-17	1.552E-24	1.340E-16	9.290E-01	1.991E-02	6.142E-17				
mtSCRABBLE 1,1e-06,1e-04	1.452E-02	1.590E-21	1.973E-16	1.552E-24	4.364E-16	8.900E-02	5.653E-01	1.941E-16	7.219E-02			
SCRABBLE 0.01,1e-05,1e-05	5.849E-01	7.313E-03	6.886E-03	1.164E-02	1.584E-02	3.446E-04	6.011E-06	4.651E-07	4.186E-04	1.865E-05		
SCRABBLE 1,1e-06,1e-04	9.903E-01	6.189E-04	2.698E-04	5.120E-04	6.806E-04	7.070E-03	4.554E-05	5.188E-08	7.620E-03	2.677E-04	4.734E-01	
URSM	7.200E-05	3.100E-09	5.889E-14	3.426E-21	5.889E-14	4.144E-22	4.723E-14	1.750E-02	3.467E-22	1.590E-21	1.040E-07	9.762E-08

Table S10: Adjusted p-values for paired Wilcoxon tests on benchmark mean error (RMSE) - Strategy:Down-Sampling, Mean Dropout:0.907

Simulated

	strategy	modelname	count	mean	sd
1	Simulation	DURIAN.MuSiC 1,1e-06,1e-04	1.500E+02	4.334E+00	9.188E-02
2	Simulation	mtSCRABBLE 1,1e-06,1e-04	1.500E+02	4.334E+00	9.186E-02
3	Simulation	DURIAN.dsLDA 1,1e-06,1e-04	1.450E+02	4.363E+00	9.276E-02
4	Simulation	SCRABBLE 1,1e-06,1e-04	1.500E+02	4.416E+00	1.302E-01
5	Simulation	DURIAN.MuSiC 0.01,1e-05,1e-05	1.500E+02	4.451E+00	1.409E-01
6	Simulation	mtSCRABBLE 0.01,1e-05,1e-05	1.500E+02	4.455E+00	1.430E-01
7	Simulation	DURIAN.dsLDA 0.01,1e-05,1e-05	1.500E+02	4.483E+00	1.511E-01
8	Simulation	SCRABBLE 0.01,1e-05,1e-05	1.500E+02	4.538E+00	1.688E-01
9	Simulation	DrImpute	1.500E+02	4.542E+00	1.412E-01
10	Simulation	CMFImpute	1.500E+02	4.867E+00	2.013E-01
11	Simulation	dropout	1.500E+02	4.908E+00	2.042E-01
12	Simulation	URSM	1.500E+02	5.635E+00	8.879E-02
13	Simulation	G2S3	1.500E+02	5.778E+00	8.052E-02

Table S11: Stats for benchmark mean error (RMSE) - Strategy:Simulation

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	9.290E-31											
dropout	6.693E-03	1.653E-34										
DURIAN.dsLDA 0.01,1e-05,1e-05	3.070E-37	2.012E-04	4.649E-41									
DURIAN.dsLDA 1,1e-06,1e-04	1.687E-47	2.814E-27	2.825E-48	7.438E-13								
DURIAN.MuSiC 0.01,1e-05,1e-05	3.999E-41	9.899E-08	2.960E-44	7.930E-02	3.192E-08							
DURIAN.MuSiC 1,1e-06,1e-04	7.491E-49	1.999E-33	1.195E-49	1.345E-19	2.946E-03	3.006E-14						
G2S3	4.210E-50	4.210E-50	4.210E-50	4.210E-50	2.382E-49	4.210E-50	4.210E-50					
mtSCRABBLE 0.01,1e-05,1e-05	1.193E-40	2.955E-07	4.755E-44	1.166E-01	1.516E-08	8.113E-01	1.112E-14	4.210E-50				
mtSCRABBLE 1,1e-06,1e-04	7.491E-49	2.014E-33	1.195E-49	1.384E-19	3.037E-03	3.238E-14	9.136E-01	4.210E-50	1.149E-14			
SCRABBLE 0.01,1e-05,1e-05	1.121E-32	5.834E-01	1.854E-36	4.585E-03	1.976E-23	4.550E-06	3.130E-29	4.210E-50	1.403E-05	3.203E-29		
SCRABBLE 1,1e-06,1e-04	6.912E-45	8.993E-15	4.765E-47	1.896E-04	4.165E-05	4.391E-02	9.093E-11	4.210E-50	2.693E-02	9.342E-11	2.982E-12	
URSM	4.210E-50	4.210E-50	4.210E-50	4.210E-50	2.382E-49	4.210E-50	4.210E-50	4.366E-31	4.210E-50	4.210E-50	4.210E-50	4.210E-50

Table S12: Adjusted p-values for paired Wilcoxon tests on benchmark mean error (RMSE) - Strategy:Simulation

L2 Error Benchmarks

Aggregate by method

	modelname	count	mean	sd
1	DURIAN.dsLDA 0.01,1e-05,1e-05	2.940E+02	5.964E+00	6.762E-01
2	mtSCRABBLE 0.01,1e-05,1e-05	2.860E+02	5.975E+00	6.956E-01
3	DURIAN.MuSiC 0.01,1e-05,1e-05	2.850E+02	5.981E+00	7.006E-01
4	SCRABBLE 1,1e-06,1e-04	3.000E+02	6.159E+00	1.482E+00
5	SCRABBLE 0.01,1e-05,1e-05	2.980E+02	6.193E+00	1.463E+00
6	DURIAN.dsLDA 1,1e-06,1e-04	2.950E+02	6.316E+00	7.393E-01
7	mtSCRABBLE 1,1e-06,1e-04	2.950E+02	6.429E+00	8.466E-01
8	DURIAN.MuSiC 1,1e-06,1e-04	3.000E+02	6.438E+00	8.430E-01
9	DrImpute	2.830E+02	6.581E+00	4.297E-01
10	G2S3	3.000E+02	6.649E+00	3.186E-01
11	CMFImpute	2.520E+02	7.123E+00	3.635E-01
12	dropout	3.000E+02	7.133E+00	4.004E-01
13	URSM	2.830E+02	7.479E+00	3.537E-01

Table S13: Stats for benchmark L2 error (L2 norm)

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.145E-37											
dropout	6.671E-01	7.294E-41										
DURIAN.dsLDA 0.01,1e-05,1e-05	6.435E-62	3.416E-21	1.037E-63									
DURIAN.dsLDA 1,1e-06,1e-04	1.719E-34	2.016E-04	9.655E-39	1.597E-18								
DURIAN.MuSiC 0.01,1e-05,1e-05	5.744E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18							
DURIAN.MuSiC 1,1e-06,1e-04	5.770E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22						
G2S3	4.258E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02					
mtSCRABBLE 0.01,1e-05,1e-05	1.926E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26				
mtSCRABBLE 1,1e-06,1e-04	3.124E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21			
SCRABBLE 0.01,1e-05,1e-05	9.582E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01		
SCRABBLE 1,1e-06,1e-04	5.083E-01	1.000E+00	1.073E-02	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.310E-06	
URSM	2.862E-21	8.659E-65	6.024E-28	5.271E-94	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.869E-28	6.024E-28

Table S14: Adjusted p-values for paired Wilcox tests on benchmark L2 error (L2 norm)

Down-Sampled

	strategy	modelname	count	mean	sd
1	Down-Sampling	DrImpute	1.330E+02	6.199E+00	1.523E-01
2	Down-Sampling	G2S3	1.500E+02	6.505E+00	3.732E-01
3	Down-Sampling	DURIAN.dsLDA 0.01,1e-05,1e-05	1.440E+02	6.622E+00	2.671E-01
4	Down-Sampling	mtSCRABBLE 0.01,1e-05,1e-05	1.360E+02	6.641E+00	4.043E-01
5	Down-Sampling	DURIAN.MuSiC 0.01,1e-05,1e-05	1.350E+02	6.659E+00	3.945E-01
6	Down-Sampling	CMFImpute	1.020E+02	6.726E+00	8.114E-02
7	Down-Sampling	dropout	1.500E+02	6.749E+00	8.337E-02
8	Down-Sampling	DURIAN.dsLDA 1,1e-06,1e-04	1.500E+02	7.012E+00	2.877E-01
9	Down-Sampling	URSM	1.330E+02	7.106E+00	4.615E-02
10	Down-Sampling	DURIAN.MuSiC 1,1e-06,1e-04	1.500E+02	7.256E+00	2.788E-01
11	Down-Sampling	mtSCRABBLE 1,1e-06,1e-04	1.450E+02	7.265E+00	2.742E-01
12	Down-Sampling	SCRABBLE 1,1e-06,1e-04	1.500E+02	7.637E+00	2.854E-02
13	Down-Sampling	SCRABBLE 0.01,1e-05,1e-05	1.480E+02	7.663E+00	2.276E-02

Table S15: Stats for benchmark L2 error (L2 norm) - Strategy:Down-Sampling

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.145E-37											
dropout	6.671E-01	7.294E-41										
DURIAN.dsLDA 0.01,1e-05,1e-05	6.435E-62	3.416E-21	1.037E-63									
DURIAN.dsLDA 1,1e-06,1e-04	1.719E-34	2.016E-04	9.655E-39	1.597E-18								
DURIAN.MuSiC 0.01,1e-05,1e-05	5.744E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18							
DURIAN.MuSiC 1,1e-06,1e-04	5.770E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22						
G2S3	4.258E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02					
mtSCRABBLE 0.01,1e-05,1e-05	1.926E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26				
mtSCRABBLE 1,1e-06,1e-04	3.124E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21			
SCRABBLE 0.01,1e-05,1e-05	9.582E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01		
SCRABBLE 1,1e-06,1e-04	5.083E-01	1.000E+00	1.073E-02	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.310E-06	
URSM	2.862E-21	8.659E-65	6.024E-28	5.271E-94	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.869E-28	6.024E-28

Table S16: Adjusted p-values for paired Wilcox tests on benchmark L2 error (L2 norm) - Strategy:Down-Sampling

	strategy	modelname	count	mean	sd
1	Down-Sampling	DrImpute	5.000E+01	6.129E+00	4.490E-02
2	Down-Sampling	G2S3	5.000E+01	6.595E+00	4.101E-01
3	Down-Sampling	CMFImpute	4.900E+01	6.655E+00	2.165E-02
4	Down-Sampling	dropout	5.000E+01	6.658E+00	2.365E-02
5	Down-Sampling	DURIAN.dsLDA 0.01,1e-05,1e-05	4.800E+01	6.910E+00	4.763E-02
6	Down-Sampling	mtSCRABBLE 0.01,1e-05,1e-05	3.900E+01	7.057E+00	5.639E-02
7	Down-Sampling	DURIAN.MuSiC 0.01,1e-05,1e-05	3.900E+01	7.063E+00	5.596E-02
8	Down-Sampling	URSM	5.000E+01	7.115E+00	4.209E-02
9	Down-Sampling	DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	7.300E+00	5.523E-02
10	Down-Sampling	mtSCRABBLE 1,1e-06,1e-04	5.000E+01	7.523E+00	6.119E-02
11	Down-Sampling	DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	7.526E+00	6.125E-02
12	Down-Sampling	SCRABBLE 1,1e-06,1e-04	5.000E+01	7.608E+00	1.509E-02
13	Down-Sampling	SCRABBLE 0.01,1e-05,1e-05	5.000E+01	7.638E+00	1.212E-02

Table S17: Stats for benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.76

	DrImpute	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.384E-37												
dropout	6.687E-01	7.294E-41											
DURIAN.dsLDA 0.01,1e-05,1e-05	4.275E-62	3.416E-21	1.037E-63										
DURIAN.dsLDA 1,1e-06,1e-04	1.933E-34	2.016E-04	9.655E-39	1.597E-18									
DURIAN.MuSiC 0.01,1e-05,1e-05	4.680E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18								
DURIAN.MuSiC 1,1e-06,1e-04	5.978E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22							
G2S3	3.034E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02						
mtSCRABBLE 0.01,1e-05,1e-05	1.570E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26					
mtSCRABBLE 1,1e-06,1e-04	3.231E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21				
SCRABBLE 0.01,1e-05,1e-05	9.586E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01			
SCRABBLE 1,1e-06,1e-04	5.102E-01	1.000E-00	1.073E-02	1.000E-00	1.000E-00	1.000E-00	1.000E-00	1.000E-00	1.000E-00	1.000E-00	1.310E-06		
URSM	1.742E-21	8.659E-65	6.024E-28	5.271E-94	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.860E-28	6.024E-28	

Table S18: Adjusted p-values for paired Wilcox tests on benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.76

	strategy	modelname	count	mean	sd
1	Down-Sampling	DrImpute	5.000E+01	6.108E+00	6.281E-02
2	Down-Sampling	G2S3	5.000E+01	6.446E+00	4.214E-01
3	Down-Sampling	DURIAN.dsLDA 0.01,1e-05,1e-05	4.900E+01	6.670E+00	5.810E-02
4	Down-Sampling	dropout	5.000E+01	6.743E+00	2.459E-02
5	Down-Sampling	CMFImpute	4.700E+01	6.773E+00	1.908E-02
6	Down-Sampling	mtSCRABBLE 0.01,1e-05,1e-05	4.800E+01	6.824E+00	7.099E-02
7	Down-Sampling	DURIAN.MuSiC 0.01,1e-05,1e-05	4.800E+01	6.832E+00	6.954E-02
8	Down-Sampling	DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	7.104E+00	6.301E-02
9	Down-Sampling	URSM	5.000E+01	7.110E+00	4.377E-02
10	Down-Sampling	mtSCRABBLE 1,1e-06,1e-04	5.000E+01	7.347E+00	7.409E-02
11	Down-Sampling	DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	7.350E+00	7.416E-02
12	Down-Sampling	SCRABBLE 1,1e-06,1e-04	5.000E+01	7.637E+00	1.480E-02
13	Down-Sampling	SCRABBLE 0.01,1e-05,1e-05	5.000E+01	7.664E+00	1.071E-02

Table S19: Stats for benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.818

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.384E-37											
dropout	6.687E-01	7.294E-41										
DURIAN.dsLDA 0.01,1e-05,1e-05	4.275E-62	3.416E-21	1.037E-63									
DURIAN.dsLDA 1,1e-06,1e-04	1.933E-34	2.016E-04	9.655E-39	1.597E-18								
DURIAN.MuSiC 0.01,1e-05,1e-05	4.680E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18							
DURIAN.MuSiC 1,1e-06,1e-04	5.978E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22						
G2S3	3.034E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02					
mtSCRABBLE 0.01,1e-05,1e-05	1.570E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26				
mtSCRABBLE 1,1e-06,1e-04	3.231E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21			
SCRABBLE 0.01,1e-05,1e-05	9.586E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01		
SCRABBLE 1,1e-06,1e-04	5.102E-01	1.000E+00	1.073E-02	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.310E-06	
URSM	1.742E-21	8.659E-65	6.024E-28	5.271E-94	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.869E-28	6.024E-28

Table S20: Adjusted p-values for paired Wilcox tests on benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.818

	strategy	modelname	count	mean	sd
1	Down-Sampling	mtSCRABBLE 0.01,1e-05,1e-05	4.900E+01	6.129E+00	1.026E-01
2	Down-Sampling	DURIAN.MuSiC 0.01,1e-05,1e-05	4.800E+01	6.157E+00	1.120E-01
3	Down-Sampling	DURIAN.dsLDA 0.01,1e-05,1e-05	4.700E+01	6.277E+00	7.139E-02
4	Down-Sampling	DrImpute	3.300E+01	6.444E+00	6.474E-02
5	Down-Sampling	G2S3	5.000E+01	6.472E+00	2.546E-01
6	Down-Sampling	DURIAN.dsLDA 1,1e-06,1e-04	5.000E+01	6.633E+00	7.251E-02
7	Down-Sampling	dropout	5.000E+01	6.847E+00	4.066E-02
8	Down-Sampling	mtSCRABBLE 1,1e-06,1e-04	4.500E+01	6.889E+00	9.106E-02
9	Down-Sampling	DURIAN.MuSiC 1,1e-06,1e-04	5.000E+01	6.891E+00	8.762E-02
10	Down-Sampling	CMFImpute	7.000E+00	6.942E+00	1.384E-02
11	Down-Sampling	URSM	3.300E+01	7.086E+00	5.090E-02
12	Down-Sampling	SCRABBLE 1,1e-06,1e-04	5.000E+01	7.667E+00	1.539E-02
13	Down-Sampling	SCRABBLE 0.01,1e-05,1e-05	4.800E+01	7.687E+00	1.020E-02

Table S21: Stats for benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.907

	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.384E-37											
dropout	6.687E-01	7.294E-41										
DURIAN.dsLDA 0.01,1e-05,1e-05	4.275E-62	3.416E-21	1.037E-63									
DURIAN.dsLDA 1,1e-06,1e-04	1.933E-34	2.016E-04	9.655E-39	1.597E-18								
DURIAN.MuSiC 0.01,1e-05,1e-05	4.680E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18							
DURIAN.MuSiC 1,1e-06,1e-04	5.978E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22						
G2S3	3.034E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02					
mtSCRABBLE 0.01,1e-05,1e-05	1.570E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26				
mtSCRABBLE 1,1e-06,1e-04	3.231E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21			
SCRABBLE 0.01,1e-05,1e-05	9.586E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01		
SCRABBLE 1,1e-06,1e-04	5.102E-01	1.000E+00	1.073E-02	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.000E+00	1.310E-06	
URSM	1.742E-21	8.659E-65	6.024E-28	5.271E-94	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.869E-28	6.024E-28

Table S22: Adjusted p-values for paired Wilcox tests on benchmark L2 error (L2 norm) - Strategy:Down-Sampling, Mean Dropout:0.907

Simulated

	strategy	modelname	count	mean	sd
1	Simulation	SCRABBLE 1,1e-06,1e-04	1.500E+02	4.681E+00	6.180E-02
2	Simulation	SCRABBLE 0.01,1e-05,1e-05	1.500E+02	4.744E+00	8.526E-02
3	Simulation	DURIAN.dsLDA 0.01,1e-05,1e-05	1.500E+02	5.333E+00	1.108E-01
4	Simulation	mtSCRABBLE 0.01,1e-05,1e-05	1.500E+02	5.371E+00	7.508E-02
5	Simulation	DURIAN.MuSiC 0.01,1e-05,1e-05	1.500E+02	5.372E+00	7.679E-02
6	Simulation	DURIAN.dsLDA 1,1e-06,1e-04	1.450E+02	5.595E+00	4.794E-02
7	Simulation	mtSCRABBLE 1,1e-06,1e-04	1.500E+02	5.620E+00	4.163E-02
8	Simulation	DURIAN.MuSiC 1,1e-06,1e-04	1.500E+02	5.620E+00	4.164E-02
9	Simulation	G2S3	1.500E+02	6.793E+00	1.502E-01
10	Simulation	DrImpute	1.500E+02	6.920E+00	2.885E-01
11	Simulation	CMFImpute	1.500E+02	7.393E+00	1.923E-01
12	Simulation	dropout	1.500E+02	7.516E+00	1.373E-01
13	Simulation	URSM	1.500E+02	7.809E+00	2.623E-02

Table S23: Stats for benchmark L2 error (L2 norm) - Strategy:Simulation

	DrImpute	CMFImpute	DrImpute	dropout	DURIAN.dsLDA 0.01,1e-05,1e-05	DURIAN.dsLDA 1,1e-06,1e-04	DURIAN.MuSiC 0.01,1e-05,1e-05	DURIAN.MuSiC 1,1e-06,1e-04	G2S3	mtSCRABBLE 0.01,1e-05,1e-05	mtSCRABBLE 1,1e-06,1e-04	SCRABBLE 0.01,1e-05,1e-05	SCRABBLE 1,1e-06,1e-04
DrImpute	2.145E-37												
dropout	6.671E-01	7.294E-41											
DURIAN.dsLDA 0.01,1e-05,1e-05	6.435E-62	3.416E-21	1.037E-63										
DURIAN.dsLDA 1,1e-06,1e-04	1.719E-34	2.016E-04	9.655E-39	1.597E-18									
DURIAN.MuSiC 0.01,1e-05,1e-05	5.744E-53	6.552E-23	1.649E-52	2.595E-01	7.977E-18								
DURIAN.MuSiC 1,1e-06,1e-04	5.770E-16	2.784E-02	2.122E-17	2.304E-23	6.544E-05	3.709E-22							
G2S3	4.258E-36	2.638E-02	1.284E-32	2.109E-27	1.481E-04	3.350E-25	7.135E-02						
mtSCRABBLE 0.01,1e-05,1e-05	1.926E-53	8.019E-24	4.012E-53	3.389E-01	7.190E-18	9.905E-01	3.684E-22	2.758E-26					
mtSCRABBLE 1,1e-06,1e-04	3.124E-16	1.856E-02	6.999E-18	1.175E-22	9.422E-05	1.054E-21	9.905E-01	4.578E-02	1.054E-21				
SCRABBLE 0.01,1e-05,1e-05	9.582E-01	9.905E-01	6.307E-02	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01	9.905E-01			
SCRABBLE 1,1e-06,1e-04	5.083E-01	1.900E-00	1.073E-02	1.900E-00	1.900E-00	1.900E-00	1.900E-00	1.900E-00	1.900E-00	1.900E-00	1.310E-06		
URSM	2.862E-21	8.659E-65	6.024E-28	5.271E-04	4.161E-55	1.718E-86	1.532E-45	3.800E-86	3.422E-87	1.487E-44	3.869E-28	6.024E-28	

Table S24: Adjusted p-values for paired Wilcoxon tests on benchmark L2 error (L2 norm) - Strategy:Simulation

References

Xuran Wang, Jihwan Park, Katalin Susztak, Nancy R. Zhang, and Mingyao Li. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nature Communications*, 10(1):1–9, January 2019. ISSN 2041-1723. doi: 10.1038/s41467-018-08023-x. URL <https://www.nature.com/articles/s41467-018-08023-x>.

David M. Blei, Michael I. Jordan, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(1):993–1022, 2003. ISSN 1533-7928. URL <http://www.jmlr.org/papers/v3/blei03a.html>.

Carlos E. Rodríguez and Stephen G. Walker. Label Switching in Bayesian Mixture Models: Deterministic Relabeling Strategies. *Jour-*

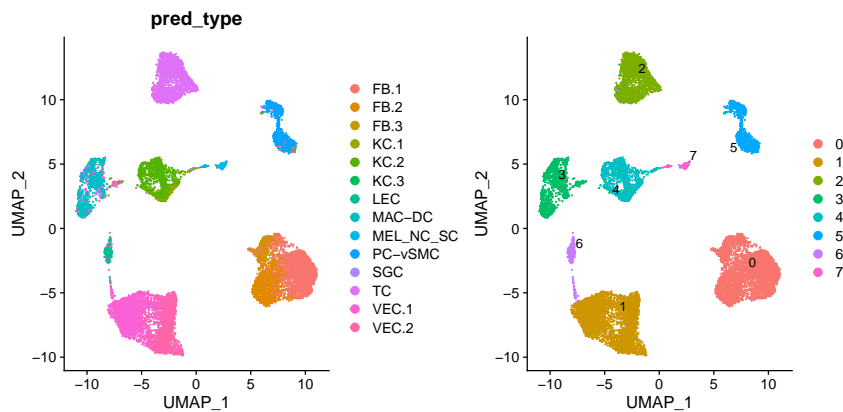


Figure S2: Clustering of Provided Labels vs De-Novo via Feature Selection Strategy 2. a) Provided cluster labels. b) Clustering based on selected features.

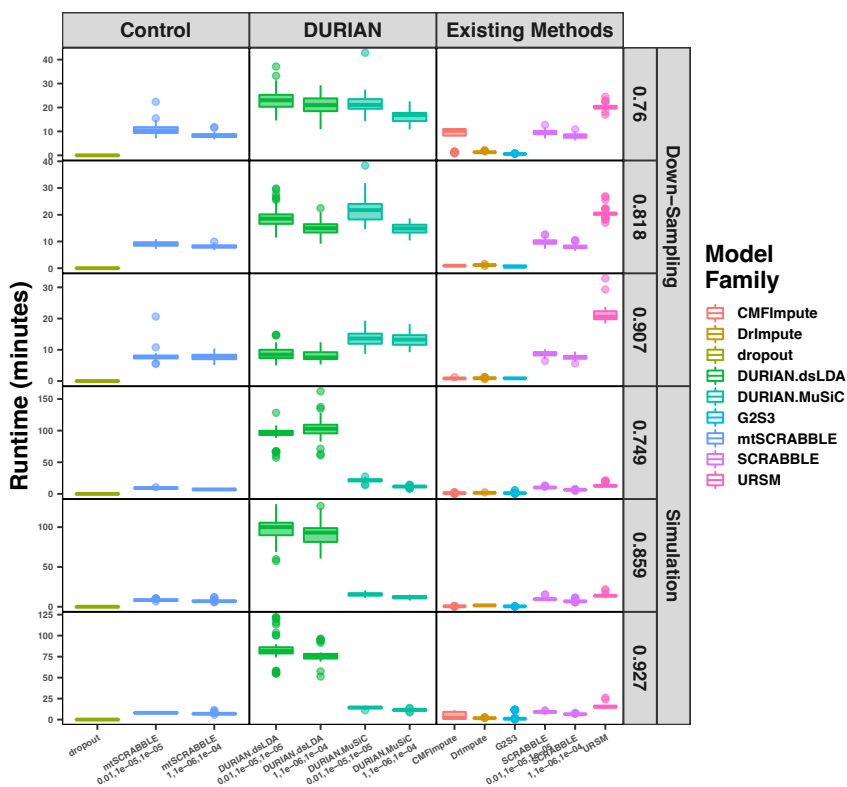


Figure S3: Benchmark Runtimes of DURIAN vs Other Methods.

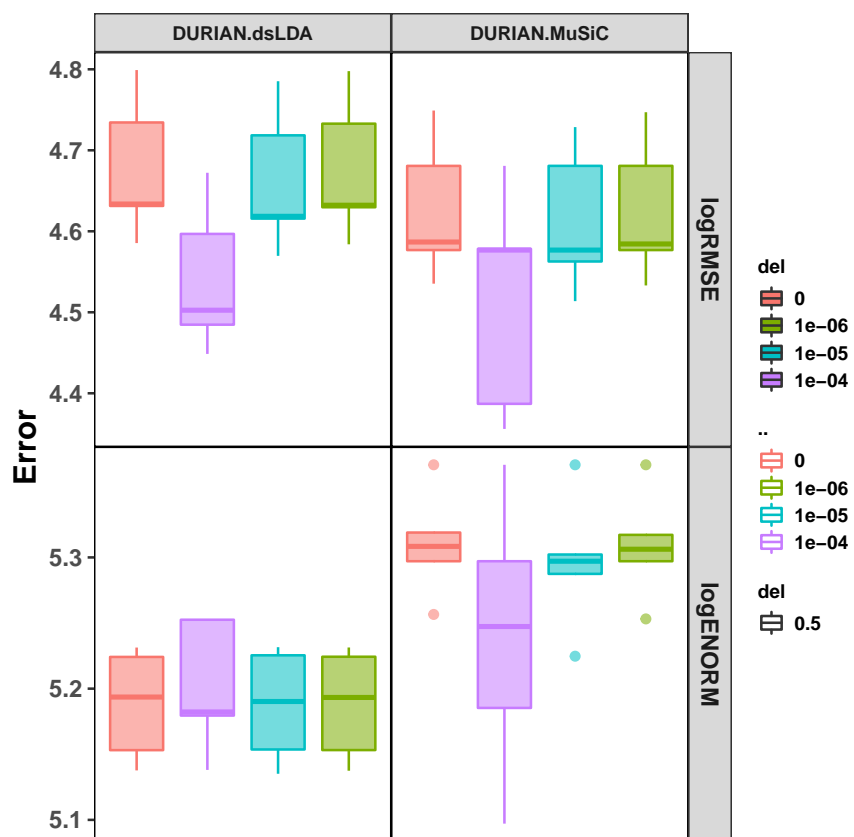


Figure S4: Variation of DURIAN deconvolution parameter.

- nal of Computational and Graphical Statistics*, 23(1):25–45, January 2014. ISSN 1061-8600. doi: 10.1080/10618600.2012.735624. URL <https://doi.org/10.1080/10618600.2012.735624>.
- Patrik O. Hoyer. Non-negative Matrix Factorization with Sparseness Constraints. *Journal of Machine Learning Research*, 5(Nov):1457–1469, 2004. ISSN 1533-7928. URL <https://www.jmlr.org/papers/v5/hoyer04a.html>.
- Fabian J. Theis, Kurt Stadlthanner, and Toshihisa Tanaka. First results on uniqueness of sparse non-negative matrix factorization. In *2005 13th European Signal Processing Conference*, pages 1–4, September 2005.
- Joshua A. Hansen, Eric K. Ringger, and Kevin D. Seppi. Probabilistic Explicit Topic Modeling Using Wikipedia. In Iryna Gurevych, Chris Biemann, and Torsten Zesch, editors, *Language Processing and Knowledge in the Web*, Lecture Notes in Computer Science, pages 69–82, Berlin, Heidelberg, 2013. Springer. ISBN 978-3-642-40722-2.
- Stephen P. Brooks and Andrew Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, December 1998. ISSN 1061-8600. doi: 10.1080/10618600.1998.10474787. URL <https://amstat.tandfonline.com/doi/abs/10.1080/10618600.1998.10474787>. Publisher: Taylor & Francis.
- David Rohde and Olivier Cappé. Online maximum-likelihood estimation for latent factor models. In *2011 IEEE Statistical Signal Processing Workshop (SSP)*, pages 565–568, June 2011. doi: 10.1109/SSP.2011.5967760. ISSN: 2373-0803.
- Luke Zappia, Belinda Phipson, and Alicia Oshlack. Splatter: simulation of single-cell RNA sequencing data. *Genome Biology*, 18(1):174, September 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1305-0. URL <https://doi.org/10.1186/s13059-017-1305-0>.
- Suoqin Jin, Christian F. Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V. Plikus, and Qing Nie. Inference and analysis of cell-cell communication using CellChat. *Nature Communications*, 12(1):1088, February 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-21246-9. URL <https://www.nature.com/articles/s41467-021-21246-9>. Number: 1 Publisher: Nature Publishing Group.
- Lingxue Zhu, Jing Lei, Bernie Devlin, and Kathryn Roeder. A unified statistical framework for single cell and bulk RNA sequencing data. *The Annals of Applied Statistics*, 12(1):609–632, March

2018. ISSN 1932-6157, 1941-7330. doi: 10.1214/17-AOAS1110. URL <https://projecteuclid.org/euclid.aos/1520564486>.
- Khusali Gupta, Jonathan Levinsohn, George Linderman, Demeng Chen, Thomas Yang Sun, Danni Dong, M. Mark Taketo, Marcus Bosenberg, Yuval Kluger, Keith Choate, and Peggy Myung. Single-Cell Analysis Reveals a Hair Follicle Dermal Niche Molecular Differentiation Trajectory that Begins Prior to Morphogenesis. *Developmental Cell*, 48(1):17–31.e6, January 2019. ISSN 15345807. doi: 10.1016/j.devcel.2018.11.032. URL <https://linkinghub.elsevier.com/retrieve/pii/S1534580718309882>.
- Leah C Biggs, Otto JM Mäkelä, Satu-Marja Myllymäki, Rishi Das Roy, Katja Närhi, Johanna Pispa, Tuija Mustonen, and Marja L Mikkola. Hair follicle dermal condensation forms via Fgf20 primed cell cycle exit, cell motility, and aggregation. *eLife*, 7:e36468, July 2018. ISSN 2050-084X. doi: 10.7554/eLife.36468. URL <https://doi.org/10.7554/eLife.36468>. Publisher: eLife Sciences Publications, Ltd.
- Maayan Baron, Adrian Veres, Samuel L. Wolock, Aubrey L. Faust, Renaud Gaujoux, Amedeo Vetere, Jennifer Hyoje Ryu, Bridget K. Wagner, Shai S. Shen-Orr, Allon M. Klein, Douglas A. Melton, and Itai Yanai. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Systems*, 3(4):346–360.e4, October 2016. ISSN 2405-4712. doi: 10.1016/j.cels.2016.08.011. URL <https://www.sciencedirect.com/science/article/pii/S2405471216302666>.
- Åsa Segerstolpe, Athanasia Palasantza, Pernilla Eliasson, Eva-Marie Andersson, Anne-Christine Andréasson, Xiaoyan Sun, Simone Picelli, Alan Sabirsh, Maryam Clausen, Magnus K Bjursell, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell metabolism*, 24(4):593–607, 2016.
- Helen He, Hemant Suryawanshi, Pavel Morozov, Jesús Gay-Mimbrera, Ester Del Duca, Hyun Je Kim, Naoya Kameyama, Yeriel Estrada, Evan Der, James G. Krueger, Juan Ruano, Thomas Tuschl, and Emma Guttman-Yassky. Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *Journal of Allergy and Clinical Immunology*, 145(6):1615–1628, June 2020. ISSN 0091-6749. doi: 10.1016/j.jaci.2020.01.042. URL <https://www.sciencedirect.com/science/article/pii/S0091674920301822>.
- Mayte Suárez-Fariñas, Benjamin Ungar, Joel Correa da Rosa, David A Ewald, Mariya Rozenblit, Juana Gonzalez, Hui Xu, Xiuzhong Zheng, Xianguyu Peng, Yeriel D Estrada, et al. Rna sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms

with potential therapeutic implications. *Journal of Allergy and Clinical Immunology*, 135(5):1218–1227, 2015.