Cell composition analysis of bulk genomics using single-cell data

Amit Frishberg[1,7], Naama Peshes-Yaloz[1,7], Ofir Cohn[1], Diana Rosentul[1], Yael Steuerman[1], Liran Valadarsky[2], Gal Yankovitz[1], Michal Mandelboim[3,4], Fuad A. Iraqi[5], Ido Amit [2], Lior Mayo[1,6], Eran Bacharach [1,8]★ and Irit Gat-Viks [1,8]★

[1]School of Molecular Cell Biology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. [2]Department of Immunology, The Weizmann Institute of Science, Rehovot, Israel. [3]National Center for Influenza and Respiratory Viruses, Central Virology Laboratory, Sheba Medical Center at Tel HaShomer, Ramat-Gan, Israel. [4]Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [5]Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [6]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel. [7]These authors contributed equally: Amit Frishberg, Naama Peshes-Yaloz. [8]These authors jointly supervised this work: Eran Bacharach, Irit Gat-Viks. *e-mail: eranba@tauex.tau.ac.il; iritgv@post.tau.ac.il
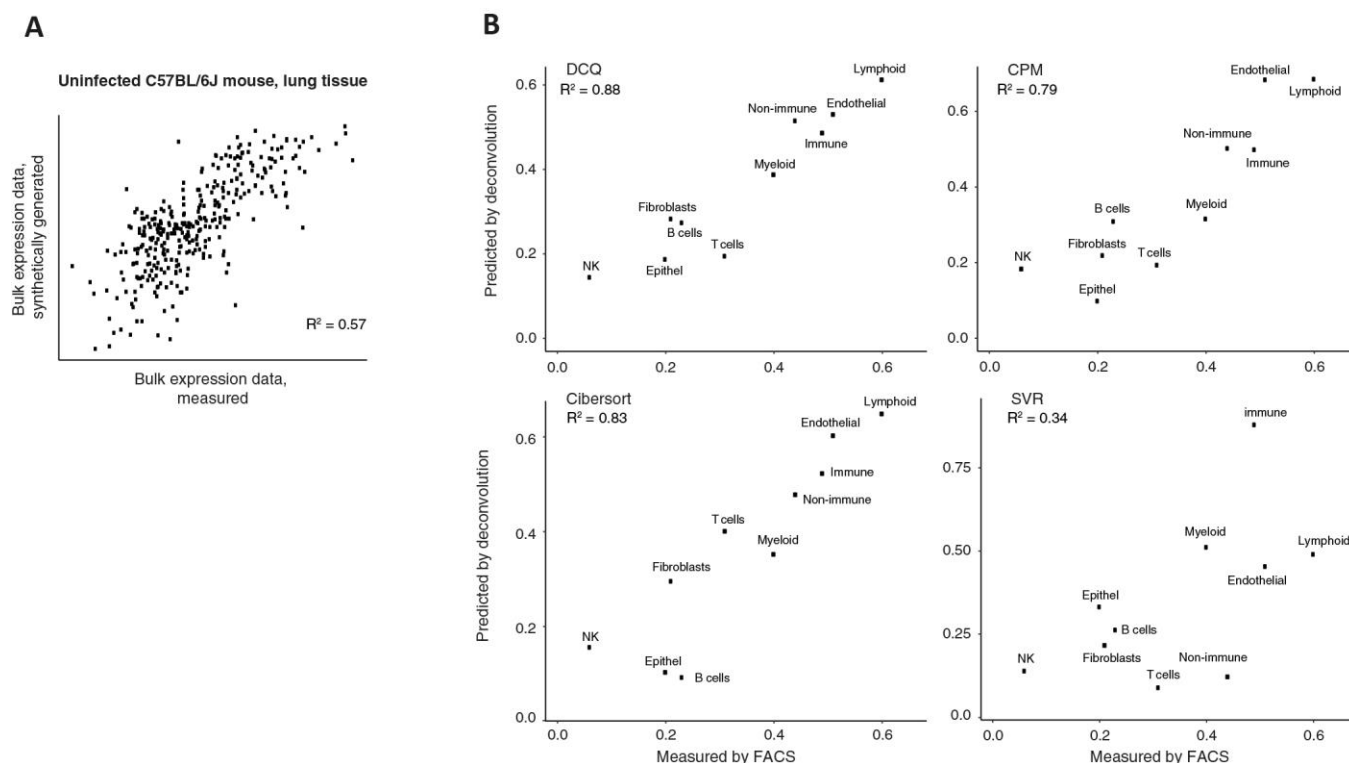
Supplementary Figure 1

Exploring the performance of deconvolution using untreated mice.

(**a**) Validation of synthetic data generation. To validate our simulation strategy, we compared measured bulk expression levels (*x* axis) and computationally-generated bulk expression levels (*y* axis) of untreated lungs from a C57BL/6J mouse (each dot is a single gene). Y axis: to synthetically-generate the bulk tissue, we applied eq. 1 (Supplementary Note 1) to calculate a weighted average of nine different cell types, using scRNA-seq profiles derived from the lungs of an uninfected C57BL/6J mouse and their known cell-type annotations[16]. The weighting is in accordance with the known fractions of the different cell types within the untreated lungs of C57BL/6J mice[22,32]. X axis: a bulk RNA-seq profile that was measured from the lungs of a naive C57BL/6J mouse ('untreated mice'; Supplementary Table 1). The scatter plot suggests a good match between measured and synthetically-generated bulk gene expression data, supporting the validity of our simulation approach. (**b**) Performance of deconvolution. We used lung tissues from naive (untreated) mice to demonstrate the robustness of deconvolution in the case of a limited cell-activation heterogeneity. To address this, deconvolution was applied using the following input data: (i) bulk RNA-seq profiles of the lung tissue, generated from five naive CC mice ('untreated mice'; Supplementary Table 1); and (ii) scRNA-seq reference data, derived from an uninfected mouse[16]. Using this input data, deconvolution was applied to predict cell-type abundance. Shown is a comparison between measured cell-type fractions in the naive lung tissue (data from refs. [22,32]; x axis) and deconvolution-predicted cell type quantities (averaged across mice; *y* axis), inferred by different methods (sub-panels; level of granularity=1 in all cases). Of note, all compared methods manifest a good match between measured and predicted cell-type quantities. This agrees well with the high accuracy of all deconvolution methods in the cell-type simulation (Fig. 2b).
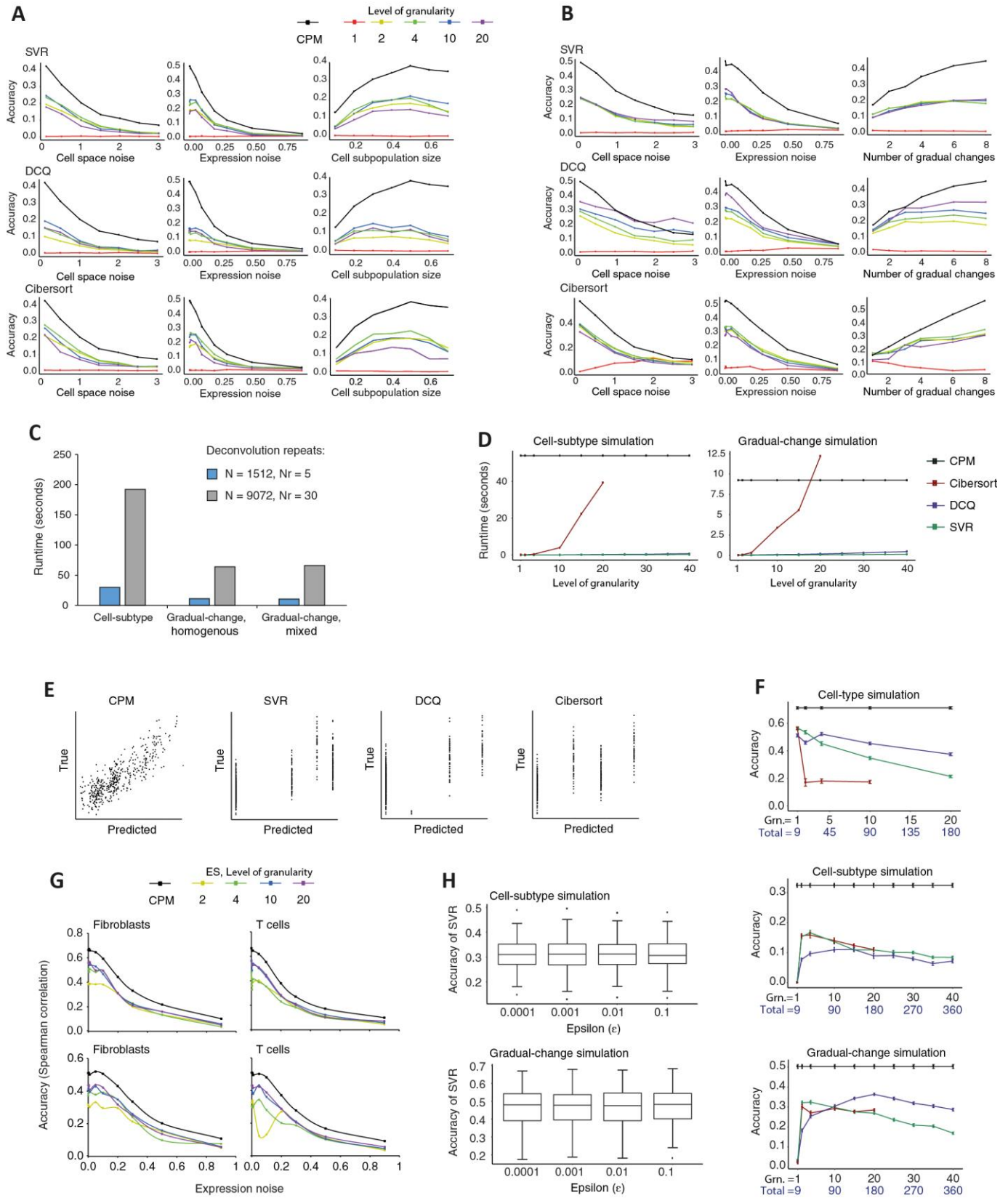
Supplementary Figure 2

Detailed analysis of CPM performance using relative bulk input data.

(**a**,**b**) Performance of CPM for cell-subtype simulation (**a**) and gradual-change simulation (**b**). Accuracy of inferring the correct cell heterogeneity was calculated using the Pearson correlation coefficient, between predicted and ground-truth cell abundance. The calculated accuracy is plotted against varying data parameters, for alternative deconvolution methods: SVR (Top), DCQ (middle) and Cibersort (bottom). CPM accuracy is color coded in black; other colors represent the number of single-cell groups (the level of granularity) within each cell type. (**c,d**) Running time. (**c**) Running time of the CPM algorithm (*y* axis; seconds) for different simulation types (*x* axis) across various numbers of deconvolution repeats (color coded). Reported is the running time per sample, averaged across 100 samples. (**d**) Running time (*y* axis; seconds) of different deconvolution methods (color coded) for different granularity levels of the reference data (*x* axis). Shown is running time for the cell-subtype simulation (left) and gradual change simulation (right). Reported is the running time per sample, averaged across 100 input bulk samples. (**e**) Predictions in the gradual changes simulation. Shown is a comparison between predicted (x axis) and true (y axis) T-cell abundance for a representative synthetic relative bulk profile using four different methods (sub-panels). CPM successfully captures the continuity in the gradual-change simulation. In contrast, the compared methods are limited to one inferred value for each of the four cell groups. The alternative methods were applied with a reference dataset that was generated using granularity of 4 cell groups. (**f**) Analysis of granularity. Performance of the different deconvolution methods (*y* axis) across varying granularity of the reference data that is given as input to the alternative methods (*x* axis), for the cell-type simulation (top), cell-subtype simulation (middle) and gradual-change simulation (bottom). X axis, top: the granularity of the reference data, defined as the numbers of single-cell groups within each cell type (each of these groups is represented with a single reference profile). X axis, bottom: the total number of reference profiles. In the cell-type-level simulation (top), granularity=1 attains the best performance. In the two other simulations (middle and bottom), a higher granularity allows better sensitivity but also leads to scalability issues. Data are mean± stdev over 100 synthetic bulk profiles. Abbreviations: Grn, Granularity. (**g**) Performance assessment in comparison with the enrichment scheme. Accuracy (y axis) is shown across varying expression noise levels (*x* axis) for different methods: CPM (black) and an enrichment scheme ('ES'; see implementation in **Supplementary Note 1**) that was applied using reference data of varying granularity (2 to 20; color coded). Results are shown for fibroblasts (left) and T cells (right), building on a single-cell-type design that was tailored to provide an unbiased comparison with ES. The accuracy was evaluated through both parametric (Pearson, top) and a-parametric (Spearman, bottom) correlation coefficients to avoid bias due to potential non-linear relationships of enrichment scores[31]. (**h**) Assessment of SVR performance. Box plots of the accuracy of SVR (y axis) across 100 cell-subtype simulations (top) and 100 gradual-change simulations (bottom) across different ε values (x axis). Each box represents 0.25-0.75 percentiles; whiskers show 95% confidence interval; horizontal lines represent the median. The plots indicate that the selected value of ε does not have a substantial effect on the overall performance.
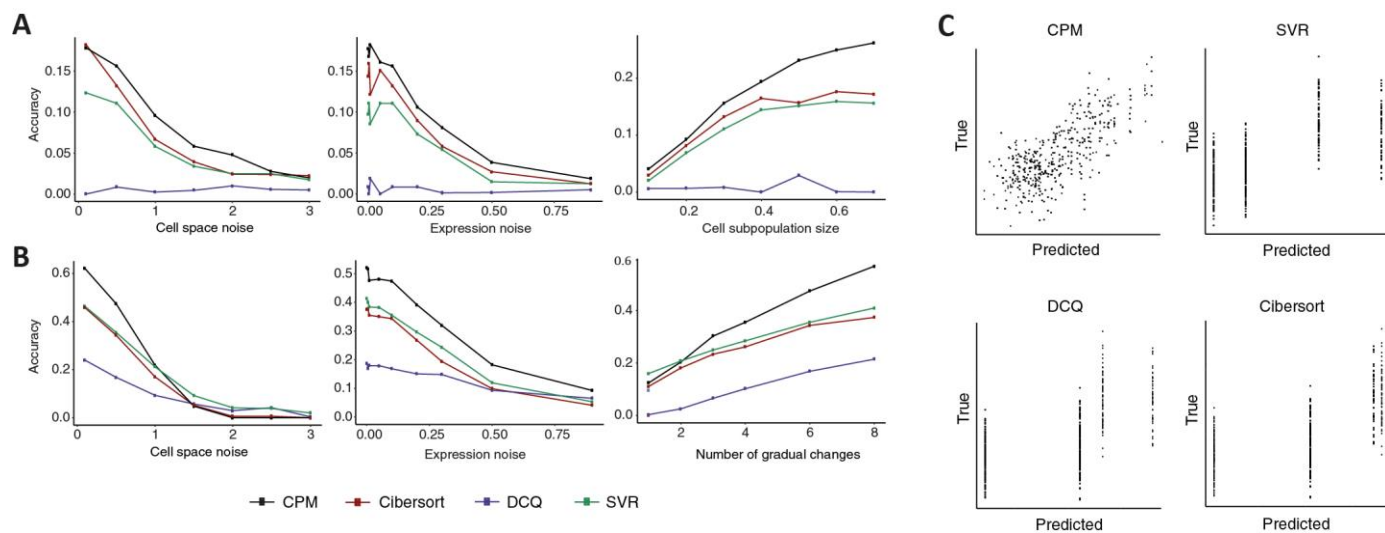
Supplementary Figure 3

CPM performance assessment for absolute bulk input data.

Plots in **a**, **b** and **c** are depicted as in Fig. 2c,d and Supplementary Fig. 2e, respectively, but for the case of absolute bulk input data.
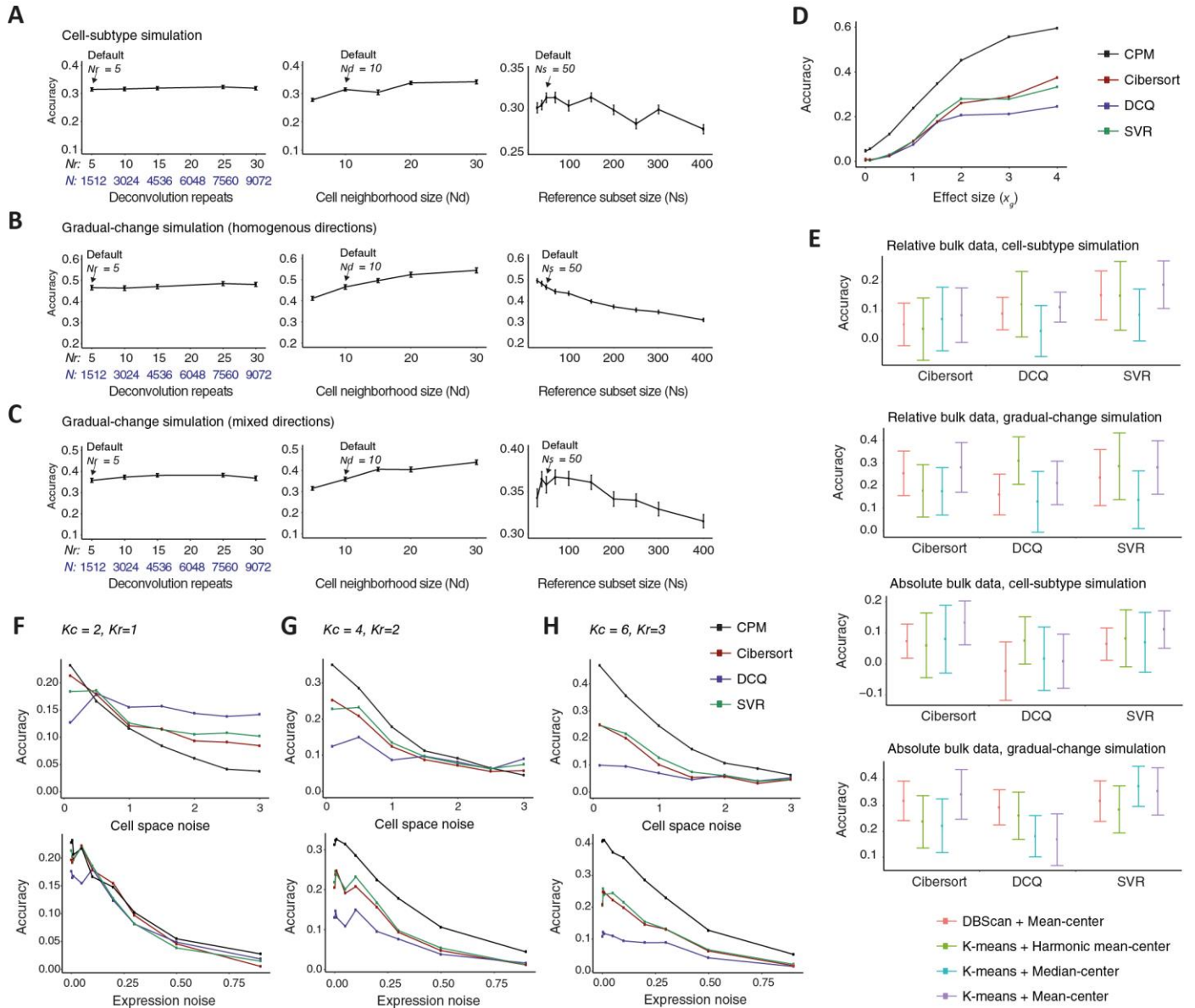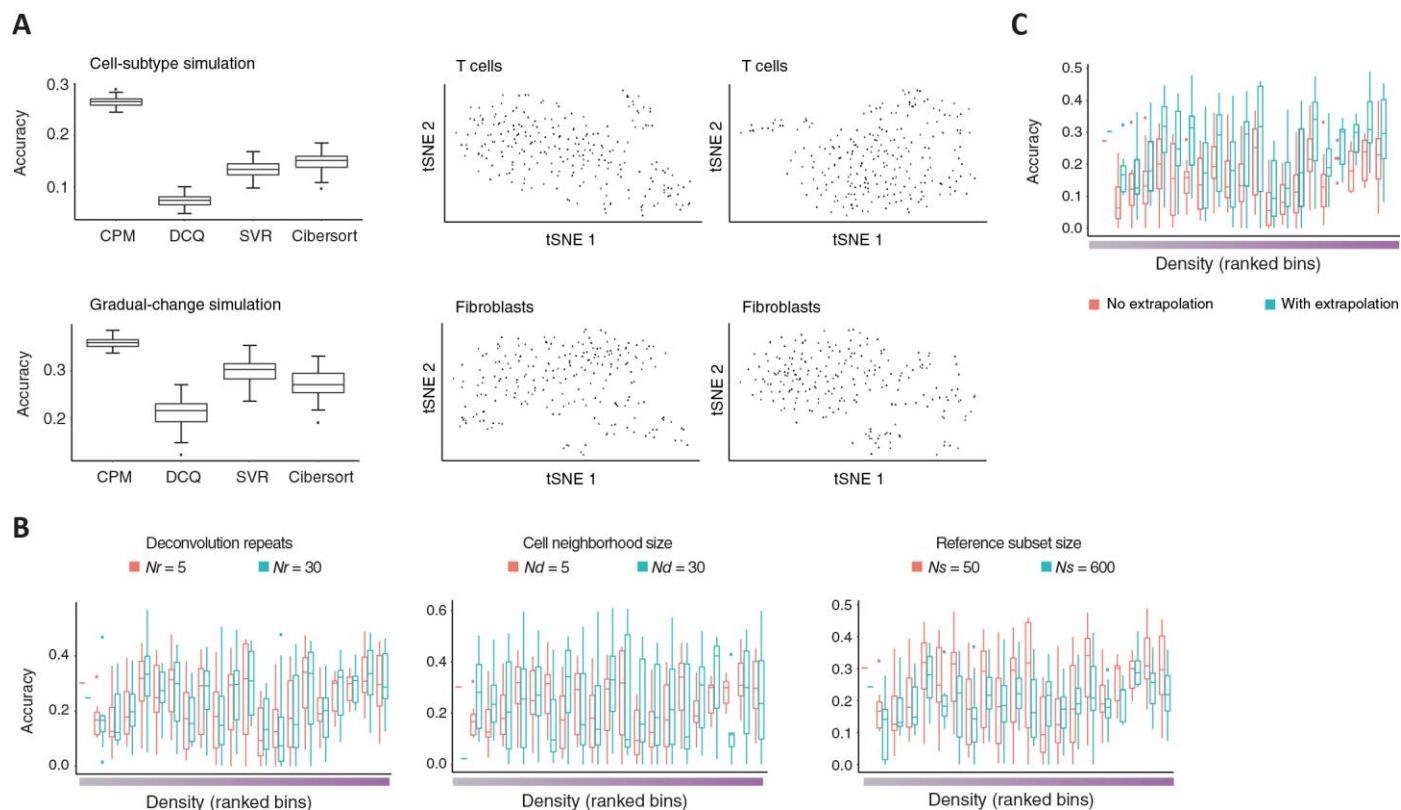
Supplementary Figure 4

CPM performance assessment using additional parameter settings.

(**a-c**) Performance across different CPM parameters. Shown is CPM accuracy (*y* axis) for different numbers of deconvolution repeats (left), cell neighborhood sizes (middle) and reference subset sizes (right) (*x* axis), for the case of cell subtype simulation (**a**), and gradual-change simulation (**b,c**), in which the various cell-state trajectories carry the same directions (**b**), or mixed directions (**c**), of gradual alternations. Default parameter settings are marked by arrows. Data are mean ± s.d. over 100 synthetic bulk profiles. The plots suggest that CPM is generally robust to the number of repeats and may gain from an increased cell neighborhood size. In contrast, CPM attains high accuracy only when the reference subset size (*Ns* parameter) is relatively small. (**d**) Performance of deconvolution using alternative effect sizes. Shown is the accuracy (*y* axis) of different deconvolution methods (color-coded) when applied on gradual-change simulation with different levels of effect size (denoted x_g in Supplementary Note 1; *x* axis). Unlike CPM, the alternative deconvolution methods were applied on a reference datasets with a granularity of 4. A constant effect size $x_g=1$ was used as a default value in all other analyses. (**e**) Performance of deconvolution using four alternative reference-reconstruction methods. Shown is the accuracy of different deconvolution methods when using different reference reconstruction methods (see Supplementary Note 1 for details): DBSCAN with mean-center approach (as devised in previous studies[10,11]; red), and *K*-means with mean-centers (purple), median-centers (blue), and harmonic-mean (green). *K*-means was applied with *K*=4 (that is, granularity=4) within each cell type. Data are mean ± s.d. over 100 synthetic bulk profiles. Overall, all reference-reconstruction methods attained similar performance. (**f,g,h**) CPM performance for mixed increase/decrease gradual-change directions. Plots are depicted as in Fig. 2d, but for synthetic data in

which gradual alterations in some cell types were reversed. (**f**) Gradual alterations in two cell types; reversed trajectory in one cell type (Kc=2, Kr=1). (**g**) Gradual alterations in four cell types; reversed trajectory in two cell types (Kc=4, Kr=2). (**h**) Gradual alterations in six cell types, reversed trajectory in three cell types (Kc=6, Kr=3). Accuracy (*y* axis) is shown against different cell space noise (top) or expression noise (bottom) (*x* axis). The alternative methods were applied with a reference dataset that was generated using granularity of 4 cell groups.
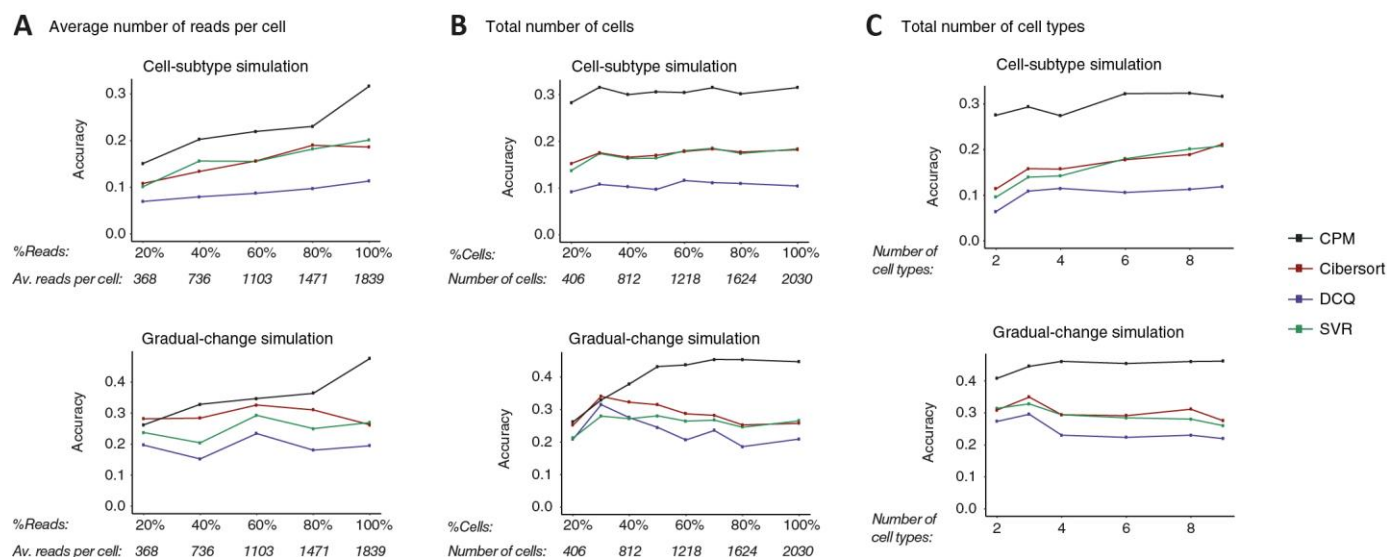
Supplementary Figure 5

The impact of cell-state space on the performance of CPM.

(**a**) CPM performance for different cell-state-space diagrams. Left: Distributions of the accuracy (*y* axis) of different deconvolution methods (*x* axis) across n=100 analyses that differ only in their input cell-state-space diagram. Each of the cell-state-space diagrams was generated by a separate tSNE run that was applied on the reference scRNA-seq data (in all cases dimension reduction was applied on the 101 generic response genes). All alternative methods were applied with granularity=4. Each box represents 0.25-0.75 percentiles; whiskers show 95% confidence interval; horizontal lines represent the median. Right: Exemplified are two cell-state-space diagrams of the same cell type (top: T cells, bottom: fibroblasts) that were generated through separate tSNE runs and were used in this performance analysis. Overall, the high accuracy of CPM remained high when using different tSNE-derived cell-state-space diagrams (left), despite marked differences between different diagrams of the same cell type (right). (**b**) The impact of the cell density on CPM performance. Shown is the distribution of the accuracy measure (*y* axis) across different cell-density levels within the cell-state space (ranked binned densities; *x* axis). Results are shown for different CPM parameters (color coded): the number of deconvolution repeats (left), cell neighborhood sizes (middle) and reference subset sizes (right). n=100 synthetic profiles. Each box represents 0.25-0.75 percentiles; whiskers show 95% confidence interval; horizontal lines represent the median. The plots suggest that the accuracy of CPM is generally independent of the local density across the cell-state space. (**c**) The impact of extrapolation. Shown is the distribution of the accuracy measure (*y* axis) across different cell-density levels within the cell-state space (ranked binned densities; *x* axis). Results are shown for the accuracy of CPM in the presence (light blue) or absence (orange) of extrapolation. n=100 synthetic profiles. Each box represents 0.25-0.75 percentiles; whiskers show 95% confidence interval; horizontal lines represent the median. The plot demonstrates the advantage of extrapolation and that a similar CPM accuracy was attained in regions that differ in their density.

Supplementary Figure 6

The impact of the quality of scRNA-seq data on CPM performance.

Shown is accuracy (y axis) for different parameters of the reference scRNA-seq data: averaged numbers of reads per cell (**a**), numbers of reference cells (**b**) and numbers of cell types within the reference dataset (**c**) (x axis). The different reference datasets were generated by eliminating some cells or reads from the original scRNAseq. Yet, all synthetic bulk data collections were still generated using the original (complete) scRNA-seq data. Results are shown for the case of cell subtype simulation (top) and gradual-change simulation (bottom). Plots are depicted as in Fig. 2c,d. The alternative methods were applied with a reference dataset that was generated using granularity of 4 cell groups.

A — Mouse #8, Mouse #38

CPM-inferred cell abundance vs Ranked reference single cells

B — Influenza virus

1. Infect individuals — 48h p.i. — I1, I2 ... Im

2. Measure body weight — %Weight loss / Individuals

3. Measure lungs, bulk response profiles

Reference data: lungs scRNA-seq & cell activation states

4. CPM for each mouse

For each cell type (e.g. T cells)

5. Calculate cell-to-phenotype (C-to-P) correlations

Relative cell abundance — I1, Im, I2 / Activation state — x

For each state x:
%weight loss vs Cell abundance — C-to-P corr. = r_x

C-to-P corr. — r_x, 0 / Activation state — x

C — Different cell-state-space
D — Absolute bulk data
E — Averaged data
F — Replicate reference mouse
G — Using published bulk data

(T cells / Fibroblasts / MPS rows)

Cell-to-phenotype correlation / Activation state (ranked bins)

H — Permuted / Real

Test statistics, gradual-change test
P= 4.9·10⁻⁶, 1.5·10⁻¹¹⁸, 6.7·10⁻¹⁷
Fibroblasts, T cells, MPS

LR, stepwise-change test
P= 3.1·10⁻⁴, 0.098, 9.5·10⁻¹⁰
Fibroblasts, T cells, MPS

I — Granularity = 4, Granularity = 10, Granularity = 20

DCQ / SVR / Cibersort

Cell-to-phenotype correlation / Activation state (ranked bins)

J — T cells / Fibroblasts

Cell-to-phenotype correlation / Activation state (ranked bins)

Influenza-infected reference
Uninfected reference

Supplementary Figure 7

CPM analyses of influenza virus infection.

(**a**) Demonstration of differences between cell population maps of different influenza virus-infected mice. Shown is the inferred relative abundance of MPS cells (*y* axis) for each individual reference single cell (*x* axis), for two representative individuals (#8 and #38, as indicated in Fig. 3a,b). Shown are average and standard deviations of each reference cell over the *N* bootstrap reference subsets. Reference cells are ranked according to their averaged relative cell abundance. The analysis demonstrates inter-individual variation in cell population maps. (**b**) Experimental design: analysis of cellular heterogeneity during *in-vivo* influenza virus infection. Bulk profiles of transcriptional response to influenza virus infection in lungs (2 days p.i., versus PBS-treated) were generated across the panel of 38 CC mice. These profiles were then integrated with reference scRNA-seq data of the same experimental design, to infer the cellular heterogeneity of each bulk sample. CPM utilized the trajectory of cell-activation states as the cell-state space (brownish scale). For each activation cell state, we further calculated the correlation (across individuals) between the measured percentage of body weight loss and the inferred relative cell abundance (e.g., for cell state *x*, cell-to-phenotype correlation is r_x). (**c**) Cell-to-phenotype correlations along the activation trajectory are robust to different cell-state-space solutions. Cell-to-phenotype correlation coefficients for the 38 CC mice (*y* axis) are presented for the indicated cell types (panels) across cell activation state bins (ranked bins; x axis). CPM analysis was repeated 10 times, each repeat was applied using a different tSNE-generated cell-state-space input. For each cell-state-space input, we first applied CPM and then used the inferred quantities to calculate the cell-to-phenotype correlation coefficients; next, coefficients were averaged over cells from each activation-state bin (denoted 'averaged coefficients'). Shown are mean and standard deviation of averaged-coefficients across the n=10 repeats. (**d,e,f,g**) The robustness of variation in cell-to-phenotype correlations along the cell activation trajectory. Cell-to-phenotype Pearson correlation coefficients across the 38 infected CC mice (*y* axis), averaged by the cells in each activation state bins (*x* axis), presented for T cells (top, n=378 cells), fibroblasts (middle, n=375 cells) and MPS cells (bottom, n=103 cells). Data are mean ± s.d. over cells. The original analysis (as in Fig. 3c) was modified in several ways: (**d**) usage of absolute bulk profiles of infected mice (rather than bulk relative profiles); (**e**) usage of the averaged profile of each recombinant inbred strain (rather than profiles of individual mice); (**f**) usage of single cell reference data from a replicate mouse; and (**g**) using single-cell and bulk data from different experimental settings, generated in different labs: scRNA-seq data at 2 days post influenza infection[16] and published microarray bulk data at 4 days post influenza infection across the pre-CC mice[21]. (**h**) A significant gradual change in cell-to-phenotype correlations over the trajectory of cell activation states. Top: Gradual-change test. Fibroblasts, T cells and MPS demonstrate monotonic (ever-increasing) cell-to-phenotype correlations over the trajectory of cell activation states (red; test statistics values), whereas reshuffling-based data (n=100 permutations) demonstrate a poor gradual change (black box plots of the test statistics values). P-values were calculated based on comparison between the reshuffling-based and observed test statistics values (a gradual-change test, Supplementary Note 1). Bottom: Stepwise-change test. Fibroblasts, T cells and MPS demonstrate gradual change in cell-to-phenotype correlations over the activation trajectory (red; LR score, a stepwise-change test), whereas reshuffling-based data (n=100 permutations) demonstrate a poor stepwise change (black box plots of LR scores). P-values were calculated based on comparison between the reshuffling-based and observed LR scores (a stepwise change test, Supplementary Note 1). Each box represents 0.25-0.75 percentiles of reshuffling-based scores; whiskers show 95% confidence interval; horizontal lines represent the median. (**i**) Analysis of cell-to-phenotype correlations in MPS cells using alternative methods. The entire experimental design was applied as illustrated in plot B for CPM, but using the alternative deconvolution methods. Results are shown as in Fig. 3c but instead of using CPM, were generated with alternative deconvolution methods (DCQ: top, SVR: middle, and Cibersort: bottom) and different granularity levels (granularity=4,10 and 20 in left, middle and right panels, respectively). Shown are CPM-inferred relative MPS abundance values (y axis), averaged over cells from each activation state bin (x axis). n=103 MPS cells; error bars, standard deviations. The plots indicate a lack of consistency among different levels of granularity and different methods (in particular, decrease in DCQ with granularity=4; increase in SVR with granularity=4,10 and Cibersort with granularity=10; and an absence of trend when using DCQ with granularity=10,20, SVR with granularity=20 and Cibersort with granularity=4,20). (**j**) Analysis of cell-to-phenotype correlations using alternative reference data. The entire experimental design was applied as illustrated in plot B, using two optional reference datasets and their associated cell-state space: (1) a relevant reference that was derived from an influenza virus-infected mouse and its associated trajectory[16] (black); and (2) an irrelevant reference data, derived from the lungs of an uninfected C57BL/6J mouse[16] together with its trajectory (red). Shown are CPM-inferred relative abundance values (y axis), averaged over cells from each activation state bin (x axis), for T cells (top, n=378) and fibroblasts (bottom, n=375). Error bars, standard deviations. The wider trajectory of the infected mice stems from its broad heterogeneity of activation states. Plots are presented as described in Fig. 3C**.** The lack of an increasing trend obtained when using an irrelevant reference data highlights the importance of using reference data that can represent the cellular heterogeneity within the bulk tissue under study.
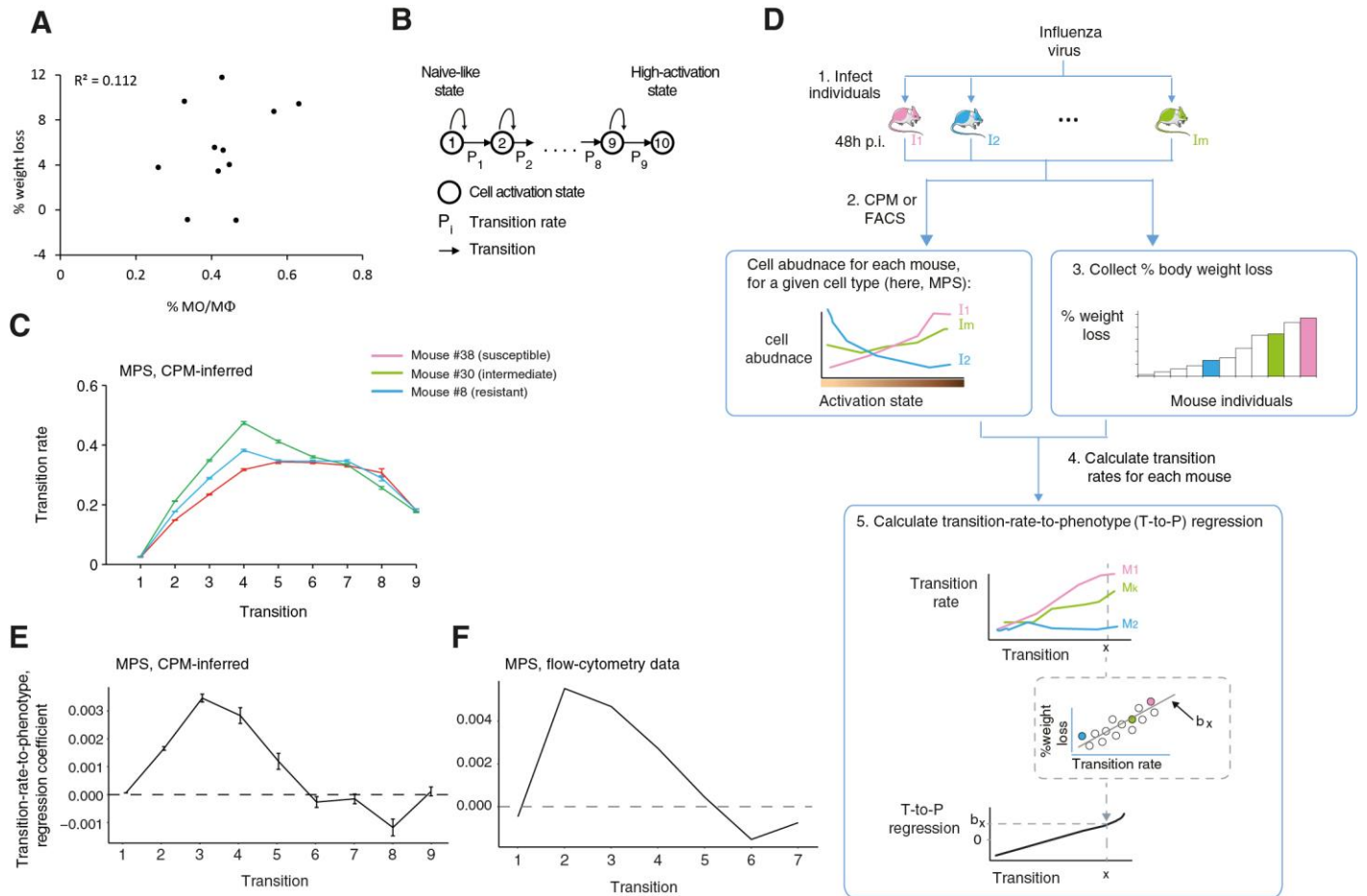
Supplementary Figure 8

Inferring dynamics with a Markov model.

(**a**) Poor correlations between the percentage of MO/MΦs in the entire tested cell population and the body weight loss. The percentage of body weight loss of infected mouse individuals (*y* axis) as a function of their percentage of MO/MΦs (including all cell states, as measured by flow cytometry; *x* axis). Relying on this overall poor correlation, we anticipated that the inter-individual variation in cell-state distribution (Figs. 3c and 4c,e) stems from inter-individual variation in temporal cell-state dynamics. One such model is examined in panels **b**–**e**. (**b**) A Markov model. The analysis takes as input (i) the absolute cell abundance levels along the activation cell-state trajectory, for each infected mouse; and (ii) absolute cell abundance levels in an uninfected tissue. We assume a model of several discrete cell states over the activation process. Based on a Markovian assumption we calculate (for each mouse) the 'transition rates' between cell states, as described in the Methods. (**c**) Demonstration of differences between transition rates of MPS from different mice. Shown is the inferred transition rates (*y* axis) calculated based on CPM-inferred cell abundance of MPS for each transition between consecutive activation states (*x* axis), for the three representative individuals that are highlighted in Fig. 3a,b (with the same color coding). Data are mean ± s.e.m. of transition rates over 100 repeats of transition-rate calculations; each such calculation was performed with cell quantities that were randomly sampled from the inferred distribution of cell quantities across the *N* deconvolution repeats. (**d**) Experimental design. For each mouse, cell abundance was collected (either using CPM or flow cytometry) along the various cell activation states. For each transition between consecutive activation states, we further calculated the transition rate, for each mouse individual (see panel **b**). This data allows calculation of regression (across individuals) between the measured clinical outcome and the inferred transition rate, in each state transition along the activation process ('transition-rate-to-phenotype regression'). For instance, for a given cell state *x*, the transition-rate-to-phenotype regression coefficient is b_x. (**e**,**f**) Transition-rate-to-phenotype relations. Transition-rate-to-phenotype regression coefficients (*y* axis) for different transitions along the cell-state trajectory (*x* axis) of the MPS population. Shown are coefficients that were calculated based on absolute cell abundance levels that were inferred using CPM (**e**), or measured for the population of MPS using flow cytometry (**f**). Of note, transition-rate-to-phenotype coefficients based on direct flow cytometry measurements (**f**) closely matched those of CPM predictions (**e**). In **e**, data are mean ± s.e.m. of regression coefficients over 100 repeats of transition-rate calculations that were generated by random sampling.

Supplementary Table 1

CC strain	Age	Initial weight	Treatment	Data	Weight loss (%)
111A	8.1	18.2	Influenza infection	RNA-seq	4.4
111A	8.1	19.5	Influenza infection	RNA-seq	1.5
1488A	9.6	23.4	Influenza infection	RNA-seq	4.7
1912A	9.6	24	Influenza infection	RNA-seq	1.7
1912A	9.6	22.7	Influenza infection	RNA-seq	2.6
2126A	8.9	12.8	Influenza infection	RNA-seq	3.1
2126A	8.9	14.9	Influenza infection	RNA-seq	-0.3
21B	8.1	15.5	Influenza infection	RNA-seq	1.9
21B	8.1	15.1	Influenza infection	RNA-seq	0.0
2513A	9.6	15.2	Influenza infection	RNA-seq	3.9
2513A	9.6	18	Influenza infection	RNA-seq	1.7
2750A	9.1	24.2	Influenza infection	RNA-seq	0.8
2750A	9.1	20.1	Influenza infection	RNA-seq	2.5
3348A	9.1	18.6	Influenza infection	RNA-seq	11.3
3348A	9.1	18.5	Influenza infection	RNA-seq	6.5
3912A	8.1	19	Influenza infection	RNA-seq	4.2
3912A	8.1	20.5	Influenza infection	RNA-seq	4.4
4438A	8.4	24.5	Influenza infection	RNA-seq	4.1
4438A	8.4	20.6	Influenza infection	RNA-seq	4.4
5000A	9.3	23.9	Influenza infection	RNA-seq	6.3
5000A	9.3	22.3	Influenza infection	RNA-seq	9.0
5001A	8.4	23.2	Influenza infection	RNA-seq	0.9
5003A	9.7	18	Influenza infection	RNA-seq	1.1
5003A	9.7	17.2	Influenza infection	RNA-seq	-0.9
5004A	8.7	19.6	Influenza infection	RNA-seq	2.0
5004A	8.7	18	Influenza infection	RNA-seq	-0.6
5010A	8.3	19.7	Influenza infection	RNA-seq	8.6
5010A	8.3	19.8	Influenza infection	RNA-seq	4.5
5021A	8.9	19.6	Influenza infection	RNA-seq	5.1
5021A	8.9	18.5	Influenza infection	RNA-seq	6.5
5022A	8.7	19.7	Influenza infection	RNA-seq	2.0
5022A	8.7	20.2	Influenza infection	RNA-seq	4.0
5023A	9.4	19.6	Influenza infection	RNA-seq	5.6
5023A	9.4	17	Influenza infection	RNA-seq	1.8
57B	8.7	23.6	Influenza infection	RNA-seq	4.7
57B	8.7	23.8	Influenza infection	RNA-seq	1.7
72A	9.0	20.1	Influenza infection	RNA-seq	6.0
72A	9.0	21.4	Influenza infection	RNA-seq	5.6
111A	8.1	17.1	PBS	RNA-seq	0.6
1912A	9.7	22.1	PBS	RNA-seq	-0.7
2126A	8.9	14.3	PBS	RNA-seq	2.1
2126A	8.9	15	PBS	RNA-seq	0.7
21B	8.9	18.8	PBS	RNA-seq	1.6
21B	8.9	18.2	PBS	RNA-seq	-0.5
21B	8.1	22.1	PBS	RNA-seq	-0.9
2513A	9.6	16.7	PBS	RNA-seq	0.6
2513A	9.6	16.2	PBS	RNA-seq	-0.9
2750A	9.0	19.9	PBS	RNA-seq	3.0
2750A	9.0	24.7	PBS	RNA-seq	6.1
3912A	8.1	19.5	PBS	RNA-seq	1.5
3912A	8.1	19.6	PBS	RNA-seq	3.6
4438A	8.3	18.4	PBS	RNA-seq	-0.1
4438A	8.3	17.5	PBS	RNA-seq	-0.1
5000A	9.3	19.9	PBS	RNA-seq	8.5
5000A	9.3	22.5	PBS	RNA-seq	5.8
5001A	8.4	19.2	PBS	RNA-seq	1.0
5001A	8.4	19.5	PBS	RNA-seq	1.0
5003A	9.7	18.7	PBS	RNA-seq	-0.1
5003A	9.7	18.2	PBS	RNA-seq	3.3
3348A	9.1	19.3	PBS	RNA-seq	2.1
3348A	9.1	18.4	PBS	RNA-seq	0.5

CC strain	Age	Initial weight	Treatment	Data	Weight loss (%)
5004A	8.7	20.5	PBS	RNA-seq	0.5
5010A	8.3	19.1	PBS	RNA-seq	3.1
5021A	8.9	18.8	PBS	RNA-seq	1.1
5022A	9.3	21.9	PBS	RNA-seq	1.8
5022A	9.3	19.7	PBS	RNA-seq	1.5
5022A	8.7	16.7	PBS	RNA-seq	2.4
5023A	9.3	18.7	PBS	RNA-seq	1.1
57B	8.1	22.6	PBS	RNA-seq	0.0
57B	8.1	19	PBS	RNA-seq	0.0
72A	9.0	22.1	PBS	RNA-seq	6.8
72A	9.0	19.7	PBS	RNA-seq	3.0
2126A	8.4	17.7	Untreated	RNA-seq	Irrelevant
111A	9.1	17.6	Untreated	RNA-seq	Irrelevant
1912A	8.7	21	Untreated	RNA-seq	Irrelevant
2750A	8.9	19.6	Untreated	RNA-seq	Irrelevant
5023A	8.1	14.5	Untreated	RNA-seq	Irrelevant
C57BL/6J*	7.6	18.4	Untreated	RNA-seq	Irrelevant
72A	7.9	18.6	Influenza infection	FACS	3.8
72A	9.4	19.5	Influenza infection	FACS	11.8
72A	9.4	20.7	Influenza infection	FACS	9.7
3912A	8.9	22.7	Influenza infection	FACS	-0.9
3912A	8.9	20.2	Influenza infection	FACS	3.5
4438A	9.0	20.5	Influenza infection	FACS	8.8
4438A	9.9	23.5	Influenza infection	FACS	-0.9
5001A	9.1	18	Influenza infection	FACS	9.4
5001A	9.3	19.7	Influenza infection	FACS	5.6
5023A	8.6	15	Influenza infection	FACS	5.3
5023A	8.6	14.8	Influenza infection	FACS	4.1
72A	7.9	19.5	PBS	FACS	0.5
3912A	8.9	20.4	PBS	FACS	0.0
4438A	9.1	26.9	PBS	FACS	5.2
5001A	9.1	20	PBS	FACS	1.5
5023A	8.6	14.3	PBS	FACS	3.5

Table S1. **Description of the animals under study**. Shown are the Collaborative Cross (CC) strain identifiers (Column A), mouse age (weeks, Column B) and weight (in grams; Column C) of each mouse before the treatment. The experimental procedure—treatment applied and subsequent measurements—are indicated in Columns D and E, respectively. The clinical outcome, namely the percentage of body weight loss, is indicated in Columns F. * The C57BL/6J strain is one of the CC founder strains.

SUPPLEMENTARY NOTE 1

The contribution of the CPM algorithm

CPM, the algorithm developed in this study, utilizes the SVR deconvolution approach, since (i) SVR accuracy has been previously demonstrated within the Cibersort framework[1], and (ii) SVR can be applied on both absolute and relative bulk genomics data (unlike alternative approaches, such as the DCQ framework[2]). A substantial challenge in the application of deconvolution methods to single-cell reference data is in maintaining robustness with a large number of reference profiles. Although deconvolution methods are generally robust and practical with a relatively small number of reference profiles, scaling to a large reference collection is problematic: most methods can handle only a few dozen of reference profiles, while DCQ, which is scalable to 200 profiles, was designed specifically to relative bulk data[3]. Another challenge is the biased representation of cell states within the reference data. Here we make two contributions that make the CPM methodology accurate and scalable: (1) To solve the scaling issue, we applied SVR multiple times, each time on a different subset of the reference profiles, and then aggregated the inferred values into a final abundance prediction for each reference profile. This avoids the need to utilize many reference profiles simultaneously and to improve robustness with bootstrapping. (2) To avoid biases due to imbalance of reference single cells over the cell-state space, we ensured that the cell subsets used in the SVR are uniformly distributed over the cell-state space.

Synthetic data generation and the accuracy score

A single 'synthetic data collection' consisted of 100 input bulk profiles, where each profile is generated by mimicking the heterogeneity of cells within a biological complex tissue. Gene expression of a gene j in a synthetic bulk profile k, denoted z_{jk}, includes a mix of isolated single cells:

$$z_{jk} = \sum_{i=1..L} f_{ik} \cdot b_{ij} + \varepsilon_{ik} \quad [1]$$

where b_{ij} denotes the gene expression value of gene j in the reference single cell i, L is the number of single cells in the reference data, and f_{ik} is the fraction of reference cell i in sample k, formalized as:

$$f_{ik} = \frac{c_{ik}}{\sum_{i=1..L} c_{ik}} + \eta_{ik} \quad [2]$$

where c_{ik} is the quantity of cell i in sample k. We further introduce noise level ε_{ik} and η_{ik} to the values of z_{jk} and f_{ik}, respectively, by sampling from a normal distribution with a zero mean and a variance that is proportional (by a factor γ_g) to the expected value (namely, the average of z_{jk} or f_{ik} across the 100 profiles). The level of the '*expression noise*' in our synthetic data is therefore determined by the proportion factor γ_g. To avoid biases due to imbalance of cell densities, all synthetic bulk profiles were generated as a mixture of pre-selected

L=378 single cells (derived from the infected mouse at 2 days p.i.) that were uniformly distributed over the cell-state space.

Next, using our basic simulation setting (Eqs. 1-2), we generated three types of simulations. All three simulation types rely on the partition of single cells into nine cell types, and mainly differ in their cell quantity (c_{ik}) values. In the first simulation type—'cell-type simulation'—cells of different types attained different quantities, whereas all cells within each cell type attained the same cell quantity value. The two other simulations, in contrast, are focused on intra-cell-type heterogeneity: either an intra-cell-type gradual change in the quantity of cells throughout a certain trajectory of cell states (the 'gradual-change simulation'), or alternatively, an intra-cell-type changes in quantities of a selected cell subpopulation (the 'cell-subtype simulation').

Specifically, in the cell-type simulations, the quantity of each cell *i* within selected cell types was set to c_{ik}=1+ e_c (e_c is denoted the '*effect size*' of the cell type simulation), whereas the quantity of the remaining cells was set to c_{ik}=1. In the case of the gradual-change simulation, for each cell type, the position of single cells along the activation-state trajectory were used as the cell-state space (see details in the 'reference single-cell data' section). Cell quantities were generated based on this cell-state space in four steps. In the first step, we select the *Kc* cell types within which we simulate gradual change in cell quantities. Among these selected cell types, we further randomly selected *Kr* cell types whose activation trajectory was reversed. The second step introduced a normally-distributed noise in the positions of all cells with a variance that is proportional to the standard deviation of cell positions, using a proportion factor γ_p. The level of the '*cell space noise*' in our synthetic data is therefore determined by γ_p. In the third step, we standardized the positions from the second step to lie in [0,1]. In the fourth step, cell quantities (c_{ik}) for the selected cell types were calculated using an exponential function $f(x_g) = b^{x_g}$ in which *b* is the positions from the third step and the exponent x_g is the 'effect size' of the gradual-change simulation. The value of c_{ik} in all remaining (unchanged) cells in all samples was the mean standardized score of their trajectory.

For the cell-subtype simulation, for each cell type, we used the first two principle components as the cell-state space, and further introduced a cell space noise to the positions of cells within this space, as described above. Cell quantities within each given cell type were generated as follows: we first selected a single reference cell and then used its neighboring cells, assuming an Euclidian cell space, as the selected cell subtype (using a certain 'cell subpopulation size' parameter). Next, the quantity of each cell *i* within the selected subset was set to c_{ik}=1+ e_s where e_s is the magnitude of cell quantity changes, denoted the 'effect size' of the cell-subtype simulation. The default value of c_{ik} in all remaining (unchanged) cells in all samples was set to 1.

Whereas the data generation above refers to absolute expression values, we also generated relative-expression synthetic data collections. This was done by generating additional 100 control synthetic profiles for each synthetic collection (in which c_{ik} is always the mean trajectory standardized score or 1 for either gradual-change simulation and cell-subtype simulation, respectively), and then calculating the differential expression values of each synthetic data profile.

For each simulation setting we generated a single synthetic data collection and analyzed the performance of each method on this collection. In particular, for each synthetic profile collection, an 'accuracy' metric was calculated as the Pearson correlation between the actual fractions f_{ik} and predicted fractions across all L reference single cells and 100 bulk profiles. Overall we generated 1671 synthetic data collection, for three simulations settings (cell-type, cell-subtype and gradual-change simulations), seven different levels of expression noise (γ_g ranges from 0.001 to 0.9), seven levels of cell space noise (γ_p ranges from 0.1 to 3), seven cell subtypes in the cell-subtype simulations (fractions ranging between 0.1 to 0.7 of cells) assuming either absolute or relative bulk data, nine combinations of Kc and Kr, and 5 effect size levels in the gradual-change simulation. The effect sizes of the cell type and cell subtype simulations (e_c and e_s) were fixed to 0.5 as the range of effects parallels the abovementioned changes in noise factors. In the cell-type simulation, the effect was added to cells within six arbitrarily-selected cell types. Our default set of parameter is: cell space noise = 0.5, expression noise = 0.1, cell subpopulation size = 0.3, Kc = 6 and Kr=0, assuming relative bulk data, and effect size x_g= 1 (corresponding to a linear function). In all cases, we report only the particular parameters that are different from this default setting.

Generating reference data for the compared methods

Since the compared deconvolution methods rely on a relatively small number of input reference profiles, the reference data was constructed using the input scRNA-seq profiles. In particular, we implemented the methodology that was previously applied in scRNA-seq-based deconvolution studies[4,5]: single cells of each cell type were partitioned into cell groups using DBSCAN clustering (using ε=3 and minPts=15; as previously described[6]). Then, the center of each such group was used as a reference profile; the 'mean center' (namely, the averaged cell profiles) of each cell group was used as its center. We refer to this approach as the 'DBSCAN + mean center method'. To generalize the reference-construction approach to a pre-selected size of reference dataset, we clustered the cells within each cell type using K-means clustering (instead of DBSCAN) and then identified the center of each cell group (the clustering relies on the cell-state space). K therefore specifies the number of single cell groups within each cell type and is referred to as the 'level of granularity'. We applied three alternative center-identification methods: (1) the 'mean-center', as detailed above; (2) the 'median-center', which is the median vector across cells; and (3) the 'harmonic-center', defined as the in verse of the mean of the in verse of the values. In all settings, gene marker selection was done by selecting a set of genes that minimizes the condition number. We refer to these approaches as the 'K-means + mean-/median-/harmonic-center' methods. All reported results, except from **Supplementary Fig. 4E**, use K-means clustering followed by the mean-center statistics. Specifically, each of the compared methods was analyzed using a variety of granularity (K) values.

Synthetic data analysis demonstrates the tradeoff between complexity and scalability

The lower performance of the alternative deconvolution methods (compared to CPM; e.g., **Figure 2CD**) could be due to either a lack of scalability to a large number of reference profiles, or simply due to the cell-state complexity within the bulk cell population. Examination of the performance across varying levels of granularity indicated that the existing methods compromise between scaling issues and cell-state complexity: as the number of reference profiles increases, the ability to accurately estimate the cell-state complexity

increases; however, further increase in the number of reference profiles leads to decreased accuracy due to scaling issues (**Supplementary Fig. 2F**).

We further observed that the high performance of CPM depended not only on its 'scalability', but also on its ability to handle a high cell-state complexity: analysis of an "enrichment scheme", which analyzes each reference profile independently and thereby is scalable to a large reference collection, resulted in substantially lower accuracy compared to CPM (**Supplementary Fig. 2G**). In particular, the enrichment scheme was implemented as previously descried[7]. In brief, we used the single-sample GSEA (ssGSEA[8]) method to determine the enrichment of gene markers associated with each reference profile within the top-ranked genes of the complex tissue. The set of gene markers associated with each reference profile was defined as the *Ng* top ANOVA-scored genes. Importantly, the comparison of CPM to ssGSEA was applied on the cell-subtype simulations, but using only a single cell type in each synthetic data collection (a 'single-cell-type design'). Current enrichment-based methods differ in their post-processing step that typically compensates between cell types but does not compensate between cell states of the same cell type[7]; our single-cell-type design therefore provides a broad comparison to the different enrichment-based approaches regardless their particular post-processing compensation strategy.

Synthetic data analysis of different sequencing depths

To explore whether the quality of scRNA-seq data can derive improved performance, we further analyzed the impact of the number of single cells as well as the sequencing depth per cell. As the quality of scRNA-seq increased, the performance of CPM also increased; specifically, the impact of data quality on CPM appeared to be more substantial than that on alternative methods (**Supplementary Fig. 6AB**). In addition, the absence of an entire cell type from the reference dataset was also evaluated (**Supplementary Fig. 6C**), suggesting that CPM is more robust to missing cell types than the existing deconvolution methods.

Pre-processing of bulk RNA-Seq data

Reads alignment and transcript quantification were performed as described earlier[2], with several modifications. In brief, reads were aligned using the HISAT aligner[9] to the mouse reference genome (NCBI 37, mm9). Reads that were mapped to multiple positions were excluded. Recorded are those reads that were mapped to mouse gene exons (using the UCSC transcript annotation). Expression levels were then calculated and normalized by the total number of mapped reads per experiment, using HOMMER[10]. The absolute bulk profile of an infected mouse are the HOMMER-normalized expression values of the relevant lung sample.

Additional analyses support the gradual change predicted by the CPM method

To support the observed stepwise change in cell-to-phenotype correlations (**Figure 3C**), we performed several analyses. First, additional analyses showed that the same gradual changes also appeared using alternative techniques by which the trajectory was defined (**Supplementary Fig. 7C**), for absolute input profiles (**Supplementary Fig. 7D)**, when analysing the average prediction of each genetic background (**Supplementary Fig. 7E**), and using reference data from another mouse (**Supplementary Fig. 7F**).

Second, we analyzed microarray gene expression data from a public repository (GEO accession number GSE30506), consisting of bulk profiles of the lung tissue at 4 days after influenza virus infection, across a collection of 44 pre-CC mice with extreme weight-loss phenotypes. In this analysis, CPM combined bulk and single-cell expression datasets that were generated by different labs and experimental settings: bulk pre-CC profiles at 4 days p.i.[11] and scRNA-seq profiles that were generated at 2 days p.i.[12]. Comparison with weight loss data at 4 days post influenza infection across the pre-CCs[11] confirmed the gradual change in cell-to-phenotype correlations (**Supplementary Fig. 7G**).

Third, to further confirm the gradual change in cell-to-phenotype coefficients over the trajectory, we designed and applied two statistical tests: a gradual-change test and a stepwise-change test. (i) *Gradual-change testing.* To assess gradual (ever-increasing) changes, we calculated the average cell-to-phenotype coefficients through a 50-cells sliding window along the antiviral trajectory. Each window was assigned a "+" sign if its average coefficient was the maximal compared to all its predecessors and a "-" sign otherwise. The test statistics was defined as the percentage of windows carrying a "+" sign. Statistical significance was evaluated by repeating this procedure 100 times, each time using permuted data that was generated by shuffling the position of cells along the trajectory; each reported P-value was then calculated based on the approximated distribution of permuted test statistics. (ii) *Stepwise-change testing.* We observe that the cell-to-phenotype coefficients are negative in low-antiviral state and positive in high-antiviral state, indicating a transition in the level of the cell-to-phenotypes coefficients during the progression of cells along the activation trajectory. Our null hypothesis is therefore a standard model of a one-step transition, and the alternative hypothesis is a stepwise transition. In accordance, the cell-state trajectory was divided into either two intervals (the null hypothesis) or ten equal consecutive intervals (the alternative hypothesis), and Gaussian parameters were fitted for the cell-to-phenotype coefficients of cells within each such interval. We then calculate a likelihood ratio (LR) score as the ratio between the goodness of fit of the two models - namely, the ratio between the maximal likelihood value when using a ten-Gaussian model versus the maximal likelihood among all 2-bins models that differ in their particular division cutoff. To estimate statistical significance, we repeated this procedure 100 times with randomly shuffled positions of cells along the activation trajectory, and then calculated a P-value for the observed LR score based on the approximated distribution of shuffling-based LR scores.

Overall, the data supported gradual changes in the levels of cell-to-phenotype correlations over the activation process ($p < 10^{-5}$, gradual-change test; **Supplementary Fig. 7H, top**) and a better fit of the stepwise model compared to the one-step model ($p < 0.1$, stepwise-change test; **Supplementary Fig. 7H, bottom**).

Finally, we asked whether the gradual-change trend is also revealed by alternative deconvolution methods or by using an unrelated reference dataset. **Supplementary Fig. 7I** shows that using each of the alternative deconvolution methods, the increasing trend in cell-to-phenotype correlation cannot be determine due to the lack of consistency among different levels of granularity (e.g., Cibersort captured the trend with a granularity of 10 but have missed the trend with a granularity of 4 and 20). Similarly, the lack of a trend when using an uninfected-reference dataset (**Supplementary Fig. 7J**) exemplifies the importance of using reference and bulk data derived from a similar experimental setting. These findings therefore highlight the advantage of a CPM

model that is based on a continuous space of cell states, and which is constructed using reference data of a similar protocol.

Analysis of deconvolution using naive mice

We used lung tissues from naive (untreated) mice to demonstrate the robustness of deconvolution on a very different dataset: whereas the infected lung tissue harbors substantial cell-activation heterogeneity within each cell type, the naive lungs typically harbor discrete cell types with a limited cell-activation heterogeneity within each cell type[12]. To test deconvolution in this case, we generated RNA-seq data of the lung tissue from five naive CC mice (**Supplementary Table 1**), and applied CPM and alternative deconvolution methods on this bulk data. As a reference data we used scRNA-seq profiles of an uninfected mouse (2075 cells that were partitioned into nine cell types[12]; data from GEO accession number GSE107947). Since we focused on inter-cell-type heterogeneity, granularity=1 (a single group for each cell type) was used to construct the reference data for the alternative methods. Using comparison with known cell-type fractions in naive mice (**Supplementary Fig. 1B**), we find strong support for the accuracy of all compared methods in predicting the quantities of discrete cell types.

Cell population maps, inferred by CPM, are a valuable resource for future investigations

CPM can be used not only to predict the cell composition by their lineage association and state, but also to utilize those predictions for future investigations. For instance, these cell-state-specific quantities may be used to calculate cell-state-specific expression within a complex tissue, as previously described for the case of cell types[13]. As another example, **Supplementary Figure 8** demonstrates how temporal dynamics can be inferred from cell-state-specific quantities. In particular, our simplified cell-state transition model suggested relationships between cell-state transition rates and physiological outcomes. Further studies are needed in order to combine additional key features of this dynamic process (such as a potential 'lag' in onset time), which may contribute to inter-individual variation in influenza-infection outcomes. While this study demonstrates a simplified stochastic model, the interpretation of CPM-reconstructed cell population maps should be further enhanced by advanced utilization of complex neighborhood structures (such as cyclic and bifurcating trajectories), as well as by continuous stochastic modeling, as suggested in previous studies[14].

Furthermore, extensive single-cell resource catalogues such as the Human Cell Atlas[15] are currently being accumulated, suggesting that CPM may soon become applicable to analyze both archived and newly generated bulk profiles without requiring expertise in single-cell technologies. For instance, application of CPM to the growing number of genomic datasets, such as TCGA[16] and GTeX[17], should provide a resource for studying the genetic basis of cellular heterogeneity and its relationships with disease outcome. This may motivate the development of a new generation of personalized predictive tools and risk factor identification methodologies[18] that are especially designed to exploit complex patterns of inter-individual variation in the cellular population structure.

References

1   Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

2   Altboum, Z. *et al.* Digital cell quantification identifies global immune cell dynamics during influenza infection. *Molecular systems biology* **10**, 720-720, doi:10.1002/msb.134947 (2014).

3   Frishberg, A., Brodt, A., Steuerman, Y. & Gat-Viks, I. ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data. *Bioinformatics* **32**, 3842-3843, doi:10.1093/bioinformatics/btw535 (2016).

4   Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e1624, doi:10.1016/j.cell.2017.10.044 (2017).

5   Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189, doi:10.1126/science.aad0501 (2016).

6   Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans. Database Syst.* **42**, 19:11–19:21, doi:10.1145/3068335 (2017).

7   Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* **18**, 220, doi:10.1186/s13059-017-1349-1 (2017).

8   Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7, doi:10.1186/1471-2105-14-7 (2013).

9   Kim, D., Langmead, B. & Salzberg, S. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).

10   Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576-589, doi:https://doi.org/10.1016/j.molcel.2010.05.004 (2010).

11   Bottomly, D. *et al.* Expression quantitative trait Loci for extreme host response to influenza a in pre-collaborative cross mice. *G3 (Bethesda, Md.)* **2**, 213-221, doi:10.1534/g3.111.001800 (2012).

12   Steuerman, Y. *et al.* Dissection of influenza infection in vivo by single-cell RNA sequencing. *Cell Systems* **6**, 679-691, doi:10.1016/j.cels.2018.05.008 (2018).

13   Shen-Orr, S. S. *et al.* Cell type–specific gene expression differences in complex tissues. *Nature Methods* **7**, 287, doi:10.1038/nmeth.1439 (2010).

14   Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331, doi:10.1038/nature21350 (2017).

15   Regev, A. *et al.* The Human Cell Atlas. *eLife* **6**, e27041, doi:10.7554/eLife.27041 (2017).

16   Hutter, C. & Zenklusen, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **173**, 283-285, doi:10.1016/j.cell.2018.03.042 (2018).

17   Kaul, R. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nature Genetics* **49**, 1664–1670, doi:10.1038/ng.3969 (2017).

18   West, M., Ginsburg, G. S., Huang, A. T. & Nevins, J. R. Embracing the complexity of genomic data for personalized medicine. *Genome Res* **16**, 559-566, doi:doi: 10.1101/gr.3851306 (2006).