Figures

Figure 1: **Detecting perturbed cell states as differentially abundant graph neighbourhoods** (A) Schematic of the Milo workflow. Randomly sampled k-NN graph vertices (single-cells) are refined using a nearest-to median-position approach [REF]. Neighbourhoods are defined on each index vertex (cell), and cells are quantified according to the experimental design to generate a counts table. Per-neighbourhood cell counts are modelled using a negative binomial GLM, and hypothesis testing is performed to determine differentially abundant neighbourhoods. (B) A force-directed layout of a k-NN graph representing a simulated continuous trajectory of cells sampled from 2 experimental conditions (top panel - A: purple, B: white, bottom panel - kernel density of cells in condition 'B'). (C) Hypothesis testing using Milo accurately and specifically detects differentially abundant neighbourhoods (FDR 1%). Red points denote DA neighbourhoods. (D) A force-directed layout of neighbourhoods, anchored using the layout in (B). Each point represents a neighbourhood, the size of the points corresponds to the numbers of cells in each neighbourhood. Lines between points represent the number of shared cells. The DA neighbourhoods (FDR 1%) are colour by log fold-change.

Figure 2: **Milo outperforms alternative differential abundance testing approaches** (A) An example simulated trajectory of cells drawn from 5 groups with cells assigned to either conditions 'A' (purple points) or 'B' (white points). Inset bar plot shows the number of cells (y-axis) assigned to each condition according to the group from which cells were sampled (x-axis). (B) Rankings of DA testing methods across a number of measures to determine performance. Each box is coloured by the ranking of the method for each method, where a rank of 1 indicates the best performance and 5 indicates the worst across 100 simulated data sets. Ranks are calculated from the mean value across 100 simulated data sets; mean values are shown. PPV: positive predictive value, NPV: negative predictive value, TPR: true positive rate, TNR: true negative rate, FOR: false omission rate, FNR: false negative rate, FPR: false positive rate, FDR: false discovery rate. (C) The Matthews correlation coefficient assesses the performance of each method by integrating across multiple performance measures. Box plots show the MCC across 100 independent simulations for each method.

Figure 3: **Milo efficiently scales to large data sets** (A) Run time (y-axis) of the Milo workflow from graph building to differential abundance testing. Each point represents a down-sampled dataset, denoted by shape. Coloured points show the total number of cells in the full dataset labelled by the elapsed system time (mins). (B) Total memory usage (y-axis) across the Milo workflow. Each point represents a down-sampled dataset, denoted by shape. Coloured points are the full datasets labelled with the total memory usage (megabytes).

Figure 4: **Milo identifies the decline of a fate-biased precursor in the ageing mouse thymus** (A) A UMAP of single thymic epithelial cells sampled from mice aged 1-52 weeks old. Points are labelled according to their designation in. (B) A UMAP, as in (A), showing the mouse age from which the single cells were derived. (C) A graph representation of the results from Milo differential abundance testing. Nodes are neighbourhoods, coloured by their log fold change across ages. Non-DA neighbourhoods (FDR 10%) are coloured white, and sizes correspond to the number of cells in a neighbourhood. Graph edges depict the number of cells shared between adjacent neighbourhoods. (D) A swarm plot illustrating the relationship between Milo results and those reported in Baran-Gale et al. Each panel shows the neighbourhoods containing cells from the reported clusters, as in (A), coloured by their log fold change across ages (x-axis). The left panels represent the age-related change in abundance of the cluster identified in Baran-Gale et al. Red denotes a reduction of the cluster with age, whilst blue arrows denote clusters that increased in abundance. (E) A heatmap of genes differentially expressed between DA neighbourhoods in (D). Each column is a neighbourhood and rows are differentially expressed genes (FDR 1%). The top panel denotes the neighbourhood DA log fold-change.

Figure 5: **Milo identifies the compositional disorder in cirrhotic liver** (A-B) UMAP embedding of 58358 cells from healthy (n = 5) and cirrhotic (n = 5) human livers. Cells are colored by cellular lineage (A) and injury condition (B) (C) Graph representation of neighbourhoods identified by Milo. Each node represents one neighbourhood and the layout is determined by the position of the neighbourhood index in the UMAP embedding of single cells. Neighbourhoods showing differential abundance at FDR 10% are colored by log-fold change in abundance between healthy and cirrhotic samples. Node size is proportional to the number of cells in each neighbourhood. Edge size is proportional to the number of cells shared by the connected neighbourhoods. (D) Beeswarm plot showing the distribution of log-fold change in abundance between conditions in neighbourhoods from different cell type clusters. DA neighbourhoods at FDR 10% are coloured. Cell types detected as DA through clustering by Ramachandran et al. (2019) are annotated in the left side bar. (E) UMAP embedding and graph representation of neighbourhoods of 7995 cells from endothelial lineage. (F) Heatmap showing average neighbourhood expression of genes differentially expressed between DA neighbourhoods in the endothelial lineage (572 genes). Expression values for each gene are scaled