

Figure 1: Detecting perturbed cell states as differentially abundant graph neighbourhoods (A) Schematic of the Milo workflow. Neighbourhoods are defined on index cells, selected using a graph sampling algorithm. 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 KNN 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 graph representation of the results from Milo differential abundance testing. Nodes are neighbourhoods, coloured by their log fold-change. Non-DA neighbourhoods (FDR 1%) 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. The layout of nodes is determined by the position of the neighbourhood index cell in the force-directed embedding of single cells.

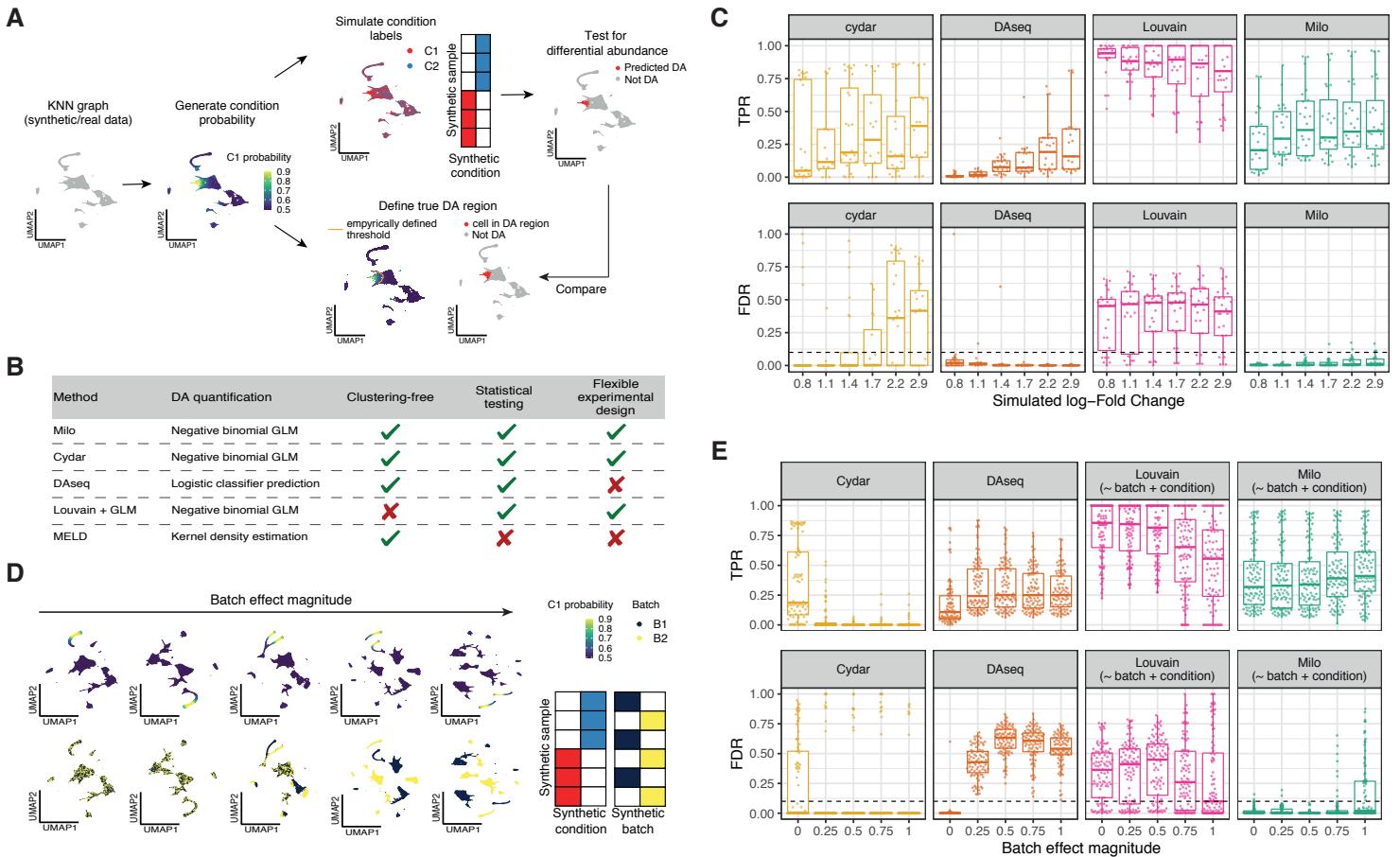


Figure 2: Milo outperforms alternative DA testing approaches and controls for false discoveries in the presence of batch effects. (A) Schematic of the strategy used to simulate ground truth regions of differential abundance to evaluate DA testing methods: on a KNN graph from real or simulated single-cell profiles we generate a smooth probability for each single-cell to be in Condition 1 ($P(C1)$), that is used to assign condition labels to cells (cells in each condition are randomly split into 3 synthetic replicates for testing). DA testing methods are run counting the cells in each synthetic condition. The cells assigned to a predicted DA region are compared to a ground truth DA region, defined by cells over an empirically determined threshold for each dataset topology. (B) A table outlining the characteristics of the methods compared to Milo (see Suppl Table 3 for a more extensive comparison). (C) True Positive Rate (TPR, top) and False Discovery Rate (FDR, bottom) for recovery of cells in simulated DA regions for DA populations with increasing simulated Fold-Change on the mouse gastrulation dataset. For each boxplot, results from 8 populations and 3 condition simulations per population are shown. Each panel represents a different DA method. (D) Example UMAP visualization of simulated batch effects of increasing magnitude. (E) True Positive Rate (TPR, top) and False Discovery Rate (FDR, bottom) for recovery of cells in simulated DA regions for DA populations with simulated batch effects of increasing magnitude. For each boxplot, results from 8 populations, simulated fold change > 1.5 and 3 condition simulations per population and fold change are shown (72 simulations per boxplot). Each panel represents a different DA method. The batch covariate is included in the test design for GLM based methods (~ batch + condition).

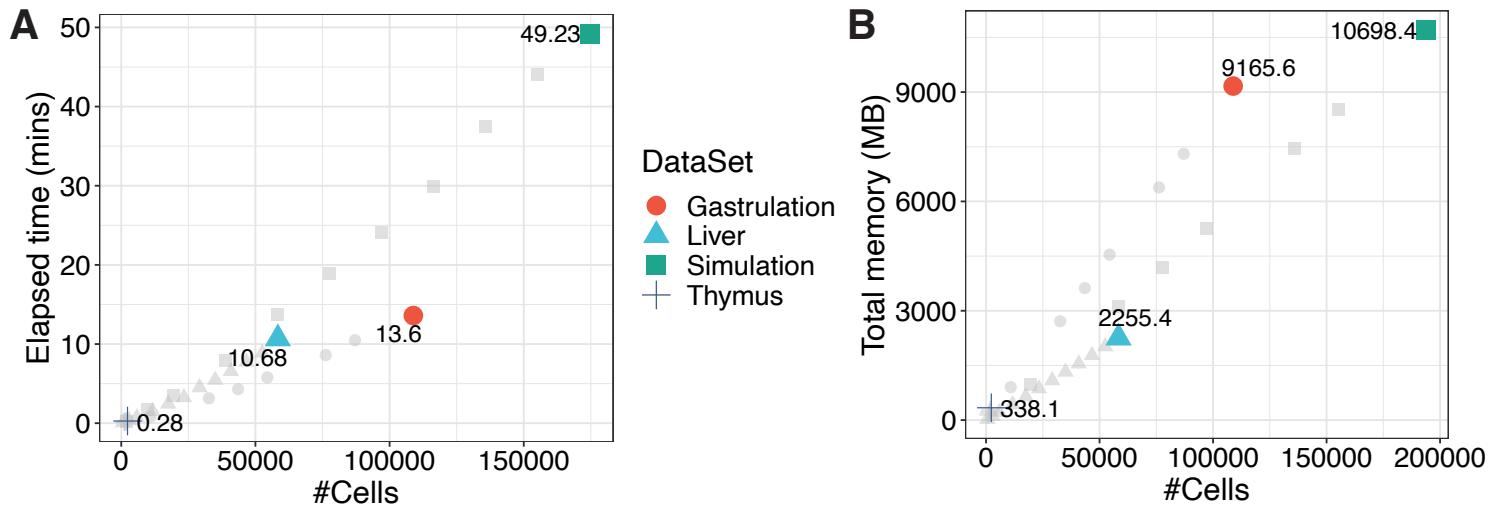


Figure 3: **Milo efficiently scales to large data sets** (A) Run time (y-axis) of the Milo workflow from graph building to differential abundance testing. Each point represents a down-sampled dataset, denoted by shape. Coloured points show the total number of cells in the full dataset labelled by the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

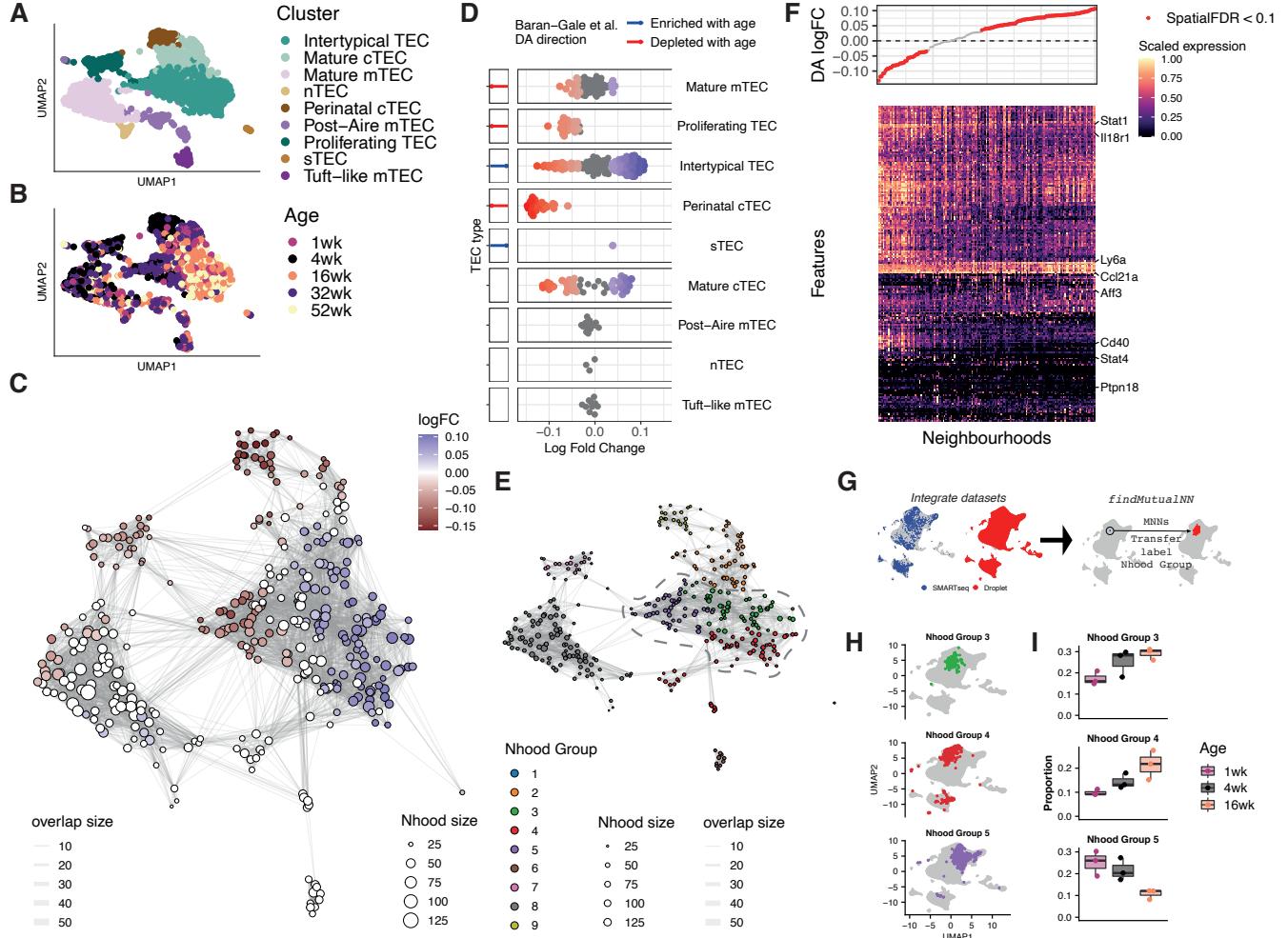


Figure 4: Milo identifies the decline of a fate-biased precursor in the ageing mouse thymus. (A-B) A UMAP of single thymic epithelial cells sampled from mice aged 1-52 weeks old. Points are labelled according to their annotation in Baran-Gale et al. 2020 (A) and mouse age (B) (C) A graph representation 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. The layout of nodes is determined by the position of the neighbourhood index cell in the UMAP embedding of single cells. (D) Beeswarm plot showing the distribution of log-fold change across age in neighbourhoods containing cells from different cell type clusters. DA neighbourhoods at FDR 10% are coloured. Cell types detected as DA through clustering by Baran-Gale et al. [3] are annotated in the left side bar. (E) Neighbourhood grouping, overlaid on the neighbourhood graph as in (C). Colours denote assignment of neighbourhoods to discrete groups using Louvain clustering. The region encircled by the dashed line denotes neighbourhood groups that correspond to the Intertypical TEC subpopulation. (F) A heatmap of genes differentially expressed between DA neighbourhoods in the Intertypical TEC cluster. Each column is a neighbourhood and rows are differentially expressed genes (FDR 5%). Expression values for each gene are scaled between 0 and 1. The top panel denotes the neighbourhood DA log fold-change. (G) Schematic showing the label transfer of neighbourhood group labels to an independent droplet scRNA-seq dataset of ageing mouse TEC from Baran-Gale et al. (2020). Datasets are integrated, then mutual nearest neighbours (MNNs) are identified, and neighbourhood group labels are transferred to cells in the droplet scRNA-seq dataset. (H) A joint UMAP embedding of SMARTseq and droplet scRNA-seq datasets, highlighting label-transferred neighbourhood groups, as coloured in (E). (I) Boxplots showing the proportions of cells from the droplet scRNA-seq dataset for each neighbourhood group across ages, providing independent validation of the results in (C). Individual points represent replicates (n=3 mice per age).

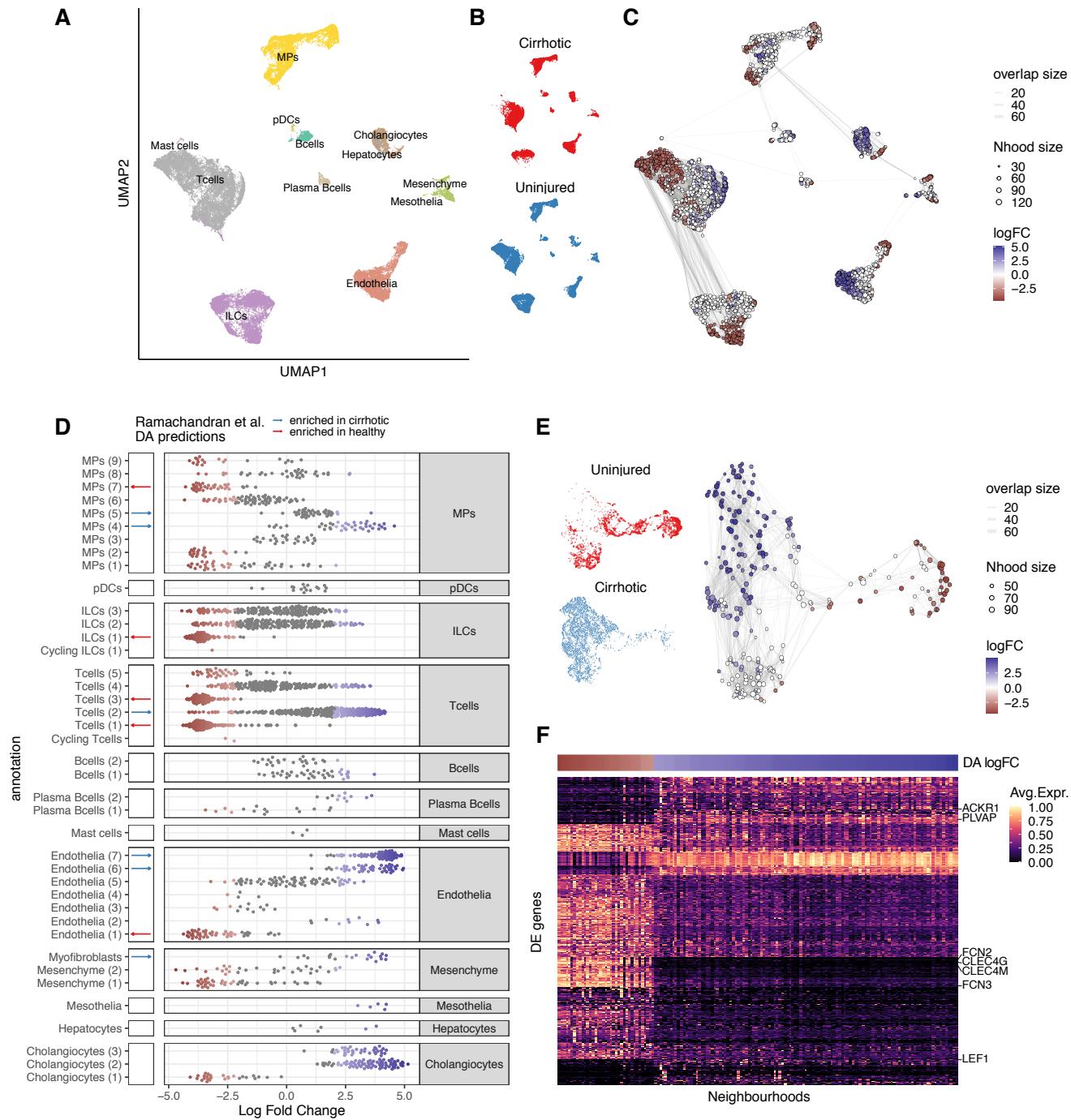


Figure 5: Milo identifies the compositional disorder in cirrhotic liver (A-B) UMAP embedding of 58358 cells from healthy ($n = 5$) and cirrhotic ($n = 5$) human livers. Cells are colored by cellular lineage (A) and injury condition (B) (C) Graph representation of neighbourhoods identified by Milo. Nodes are neighbourhoods, coloured by their log fold change between cirrhotic and healthy samples. 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. The layout of nodes is determined by the position of the neighbourhood index cell in the UMAP embedding of single cells. (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. [2] are annotated in the left side bar. (E) UMAP embedding and graph representation of neighbourhoods of 7995 cells from endothelial lineage, coloured by DA log-fold change. (F) Heatmap showing average neighbourhood expression of genes differentially expressed between DA neighbourhoods in the endothelial lineage (788 genes). Expression values for each gene are scaled between 0 and 1. The top panel denotes the neighbourhood DA log fold-change.