

1 *Milo*: differential abundance testing on single-cell data 2 using k-NN graphs

3 Emma Dann¹, Neil C. Henderson^{2,3}, Sarah A. Teichmann^{1,4}, Michael D. Morgan^{5,6†}, John C.
4 Marioni^{1,5,6†}

5 1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

6 2. Centre for Inflammation Research, The Queen's Medical Research Institute, University
7 of Edinburgh, Edinburgh, UK

8 3. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University
9 of Edinburgh, Crewe Road South, Edinburgh, UK

10 4. Theory of Condensed Matter Group, The Cavendish Laboratory, University of
11 Cambridge, Cambridge, UK

12 5. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-
13 EBI), Hinxton, Cambridge, UK

14 6. Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of
15 Cambridge, Cambridge, UK

16 † Correspondence: michael.morgan@cruk.cam.ac.uk and john.marioni@ebi.ac.uk

17 **Abstract:**

18 Single-cell omic protocols applied to disease, development or mechanistic studies can reveal
19 the emergence of aberrant cell states or changes in differentiation. These perturbations can
20 manifest as a shift in the abundance of cells associated with a biological condition. Current
21 computational workflows for comparative analyses typically use discrete clusters as input
22 when testing for differential abundance between experimental conditions. However, clusters
23 are not always an optimal representation of the biological manifold on which cells lie,
24 especially in the context of continuous differentiation trajectories. To overcome these barriers

25 to discovery, we present *Milo*, a flexible and scalable statistical framework that performs
26 differential abundance testing by assigning cells to partially overlapping neighbourhoods on a
27 k-nearest neighbour graph. Our method samples and refines neighbourhoods across the
28 graph and leverages the flexibility of generalized linear models, making it applicable to a wide
29 range of experimental settings. Using simulations, we show that *Milo* is both robust and
30 sensitive, and can reveal subtle but important cell state perturbations that are obscured by
31 discretizing cells into clusters. We illustrate the power of *Milo* by identifying the perturbed
32 differentiation during ageing of a lineage-biased thymic epithelial precursor state and by
33 uncovering extensive perturbation to multiple lineages in human cirrhotic liver. *Milo* is provided
34 as an open-source R software package with documentation and tutorials at
35 <https://github.com/MarioniLab/miloR>.

36

37 **Introduction:**

38 The advent and expansion of high-throughput and high-dimensional single-cell measurements
39 has empowered the discovery of specific cell-state changes associated with disease,
40 development and experimental perturbations. Perturbed cell states can be detected by
41 quantifying shifts in abundance of cell types in response to a biological insult. A common
42 analytical approach for quantitatively identifying such shifts is to ask whether the composition
43 of cells in predefined and discrete clusters differs between experimental conditions [1–5].
44 However, assigning single-cells to discrete clusters can be problematic, especially in the
45 context of continuous differentiation, developmental or stimulation trajectories, thus limiting
46 the power and resolution of such differential abundance (DA) testing strategies.

47

48 Alternative approaches for performing differential abundance testing without requiring clusters
49 to be defined have been proposed for high-throughput mass cytometry data [6]. For example,
50 *Cydar* constructs hyperspheres in the high-dimensional (protein) expression space and asks
51 whether the abundance of cells from different conditions varies in each hypersphere.

52 However, the construction of hyperspheres depends heavily upon the choice of input
53 parameters and upon data pre-processing. More recent developments have proposed
54 strategies for overcoming some of these limitations, but are themselves constrained to
55 pairwise comparisons, as implemented in *DAseq* [7], and thus lack flexibility.

56

57 To solve these challenges, we have developed a computational method that performs
58 differential abundance testing without relying on clustering cells into discrete groups. We make
59 use of a common data-structure that is embedded in many single-cell analyses: k-nearest
60 neighbour (k-NN) graphs. We model cellular states as overlapping neighbourhoods on such
61 a graph, which are then used as the basis for differential abundance testing. To account for
62 the non-independence of spatially overlapping neighbourhoods we build upon a previously
63 described strategy to control the spatial False Discovery Rate (FDR) [6].

64

65 Our method, which we call *Milo*, leverages the flexibility of generalized linear models (GLM),
66 thus allowing complex experimental designs. Moreover, by modelling cell states as
67 overlapping neighbourhoods, we are able to accurately pinpoint the perturbed cellular states,
68 enabling the identification of the underlying molecular programs. We demonstrate the power
69 of our approach by identifying perturbed cellular states from publicly available datasets in the
70 context of human liver cirrhosis and by uncovering a fate-biased progenitor in the ageing
71 murine thymus. Furthermore, we demonstrate the speed and scalability of our open-source
72 implementation of *Milo*, and demonstrate its superiority to alternative approaches.

73

74 **Results:**

75 *Modelling cell states as neighbourhoods on a k-NN graph*

76 We propose to model the differences in the abundance of cell states between experimental
77 conditions using graph neighbourhoods (Fig 1). Our computational approach allows
78 overlapping neighbouring regions, which alleviates the principal pitfalls of using discrete
79 clusters for differential abundance testing. We make use of a refined sampling implementation
80 [8], which leads to high coverage of the graph while simultaneously controlling the number of
81 neighbourhoods that need to be tested. For each neighbourhood we then perform hypothesis
82 testing between biological conditions to identify differentially abundant cell states whilst
83 controlling the FDR across the graph neighbourhoods.

84

85 Our method works on a k-NN graph that represents the high-dimensional relationships
86 between single-cells, a common scaffold for many single-cell analyses [1–4] (Fig 1A). The first
87 step in our method is to define a set of representative neighbourhoods on the k-NN graph,
88 where a neighbourhood is defined as the group of cells that are connected to an index cell by
89 an edge in the graph. Consequently, we need to sample a subset of single-cells to use as
90 neighbourhood indices. Adopting a purely random sampling approach means that the number
91 of neighbourhoods required to sample a fixed proportion of cells scales linearly with the total
92 number of index cells (Supp Fig 1B). This leads to an increased multiple testing burden, with
93 the potential to reduce statistical power. To solve this problem we have implemented a refined
94 sampling scheme (Fig 1A) [8]. Concretely, we perform an initial sparse sampling, without
95 replacement, of single-cells and compute the k nearest neighbors for each sampled cell. We
96 then calculate the median position of each set of nearest neighbors and find the nearest cell
97 to this median position. These adjacent cells become the set of indices from which we compute
98 the final set of neighbourhoods. This procedure has three main advantages: (1) fewer, yet
99 more representative, neighbourhoods are selected, as initial random samplings from dense
100 regions of the k-NN graph will often converge to the same index cell (Supp Fig 1A), (2) the
101 representative neighbourhoods include more cells on average (Suppl Fig 1B) and (3)
102 neighbourhood selection is more robust across initializations (Supp Fig 1C).

103

104 Next, we count the numbers of cells present in each neighbourhood (per experimental sample)
105 and use these for differential abundance testing between conditions. To incorporate complex
106 experimental designs (e.g., the presence of multiple conditions) we test for differences in
107 abundance using a Negative Binomial GLM framework [9,10]. By doing this, we can borrow
108 information across neighbourhoods, allowing robust estimation of dispersion parameters. We
109 also employ a quasi-likelihood F-statistic [11] for comparing different hypotheses, which has
110 been shown to be powerful in single-cell differential testing [12]. To account for multiple
111 hypothesis testing we use a weighted FDR procedure [13] that accounts for the spatial overlap
112 of neighbourhoods as initially introduced in *Cydar* [6]. We adapt this procedure for a k-NN
113 graph, and weight each hypothesis test P-value by the reciprocal of the kth nearest neighbour
114 distance.

115

116 Although the GLM framework allows the incorporation of nuisance covariates, to maximize the
117 power of DA testing, confounding effects should be minimized prior to graph building, for
118 example by applying an appropriate batch integration (practical considerations and
119 demonstrations on how to account for batch effects can be found in the Supplementary Notes
120 and Suppl.Fig.2-3).

121

122 To illustrate the *Milo* workflow we generated a simulated trajectory [14] composed of cells
123 sampled from two experimental conditions ('A' and 'B'; Fig1B). Cells in a defined
124 subpopulation of this trajectory were simulated to be more abundant in the 'B' condition (Fig
125 1B); this region of differential abundance is not defined as a distinct cluster by widely-used
126 clustering algorithms (Supp Fig 4). However, applying *Milo* to these simulated data specifically
127 detects that this region contains different abundances of cells from the two conditions (Fig 1C-
128 D).

129

130 *Milo out-performs existing methods for differential abundance testing*

131 To illustrate the power and accuracy of *Milo* we first simulated 100 independent continuous
132 trajectories, each consisting of 2000 single-cells, and assigned cells equally to one of 2
133 conditions: 'A' or 'B'. To simulate a subpopulation of differential abundance we sampled 90%
134 of cells in a specific region of each trajectory from condition 'B'. Moreover, we assigned cells
135 to one of 3 replicates per condition, thus mimicking a common experimental design. These
136 simulated data sets provide a ground truth against which the performance of differential
137 abundance testing approaches can be compared (Fig 2A).

138

139 As well as *Milo*, we applied two methods designed for differential abundance testing using
140 single-cell data to these simulated datasets: *Cydar*, originally designed to model differential
141 abundance in mass cytometry data [6], and *DAseq*, which utilises a logistic classifier to predict
142 which cells are from single-cell DA subpopulations represented by a reduced dimensional
143 space [7] (Fig 2B). In addition, we applied the current best-practise analysis strategy for single-
144 cell analysis: graph-clustering followed by differential abundance testing of clusters between
145 conditions. For this approach we applied 2 commonly used community detection algorithms:
146 Louvain [15] and Walktrap [16]. We modelled the differential abundance of clusters from these
147 algorithms using a NB GLM, as implemented in *edgeR* [9]. To ensure comparability between
148 methods we used the same reduced dimensional space as the input for all methods and the
149 same parameter values, where these were shared, e.g. the value of 'k' for k-NN graph building.
150 Where parameters were specific to a method, we made use of the recommended practise by
151 the method developers to select an appropriate value (Supp Table 1).

152

153 For each simulated dataset we computed the confusion matrix of each method against the
154 ground truth, and calculated a number of common summary statistics (Fig 2C, Supp Fig 5),
155 enabling an assessment of how well each method performs across a variety of metrics. To
156 generate a single value for comparing methods, and integrate across the four categories of a

157 confusion matrix, we also calculated the Matthews correlation coefficient (MCC) [17] (Fig 2D).
158 The MCC takes values between 1 (highly consistent) and -1 (highly inconsistent), thus
159 providing an intuitive assessment of method performance. Across all 100 simulations we found
160 that *Milo* out-performed all other methods, including both clustering methods, demonstrating
161 the additional gains of modelling cell states using overlapping neighbourhoods (Fig 2C, Supp
162 Fig 5). This was confirmed when examining the MCC, where we observed that *Milo* yielded
163 the highest median correlation (0.85) and lowest variance (Fig 2D). Conversely, the clustering-
164 based methods resulted in highly variable MCC values, illustrating the sensitivity of these
165 approaches to the input data set. In sum, our simulations demonstrate that *Milo* circumvents
166 a common bottleneck in single-cell analyses: the need to perform iterative rounds of
167 community detection to achieve an optimal clustering prior to differential abundance testing.

168

169

170 *Milo is fast and scalable*

171 The benchmarking dataset is fairly typical in size for current single-cell experiments. However,
172 moving forward, the number of cells assayed is likely to increase with advances in
173 experimental sample multiplexing [18,19]. As such, we tested the scalability of the *Milo*
174 workflow, and profiled the memory usage across multiple steps. For this we ran *Milo* on 3
175 published datasets of differing sizes from ~2000 to ~130,000 cells, representing differences
176 in both biological and experimental complexity [2–4], as well as a dataset of 200,000 simulated
177 single-cells from a linear trajectory (see Methods). Using these 4 data sets we measured the
178 amount of time required to execute the *Milo* workflow from graph-building through to
179 differential abundance testing (Fig 3A). In parallel, we profiled the amount of memory used
180 across the entire workflow (Fig 3B) and at each defined step (Supp Fig 6). Notably, the amount
181 of time taken increased linearly with the total size of the data set (Fig 3A), which for a large
182 set of 200k cells was less than 90 minutes. Moreover, the total memory usage across all steps

183 of the *Milo* workflow scaled primarily with the size of the input dataset (Fig 3B), indicating that
184 the complexity and composition of the single-cells largely determines the memory
185 requirements (Supp Fig 6). Importantly, these memory requirements are within the resources
186 of common desktop computers (i.e. <16GB). This benchmarking analysis demonstrates that
187 *Milo* is able to perform differential abundance analysis in large and complex datasets at a
188 scale and speed that is feasible on a desktop computer.

189

190 *Milo identifies the decline of a fate-biased epithelial precursor in the ageing mouse thymus*

191 To demonstrate the utility of *Milo* in a real-world setting we applied it to a single-cell RNA-seq
192 dataset of mouse thymic epithelial cells (TEC) sampled across the first year of mouse life,
193 which were previously clustered into 9 distinct TEC subtypes (Fig 4A) [3]. These data,
194 generated using plate-based SMART-seq2, consist of 2327 single-cells equally sampled from
195 mice at 5 different ages: 1, 4, 16, 32 and 52 weeks old (Fig 4B). Moreover, the experimental
196 design included 5 replicate experimental samples of cells for each age. The goal of the study
197 was to identify TEC subtypes that change in frequency during natural ageing.

198

199 To this end, we first constructed a k-NN graph, before assigning cells to 363 neighbourhoods,
200 which were then used to test for differential abundance of TEC states across time. At a 10%
201 FDR, we identified 217 DA neighbourhoods (112 showed a decreased abundance with age,
202 105 an increased abundance with age) spanning multiple TEC states (Fig 4C). We compared
203 our results to those generated in the original publication, which demonstrated that we were
204 able to identify the same DA states (Fig 4D), including changes in the abundance of the 'sTEC'
205 population, which consisted of just 24 cells. Moreover, whilst we recovered the previously
206 reported accumulation of Intertypical TEC with age, we also identified a novel subset of these
207 cells that were depleted with age (Fig 4C-D).

208

209 We have previously shown that Intertypical TEC represent an adult progenitor of medullary
210 TEC (mTEC) [3]. To understand the function of the novel sub-state of Intertypical TEC
211 identified using *Milo* we performed marker gene expression identification between the
212 Intertypical TEC in neighbourhoods enriched or depleted in younger mice (FDR 1%; Fig4E).
213 This analysis indicated that the cells from younger mice up-regulated multiple cytokine
214 response genes (e.g. *Stat1*, *Stat4*, *Aff3*) (Fig 4E), illustrated by the enriched Gene Ontology
215 term GO:0034097 'response to cytokine' (enrichment adjusted p-value= 2.48×10^{-3}). Cytokine
216 signalling is key to mTEC differentiation [20,21], indicating that these TEC from younger mice
217 might be differentiating more efficiently to the mTEC lineage. The discovery that medullary-
218 biased Intertypical TEC are less abundant with age was corroborated by our original study
219 utilising a much larger data set of ~90,000 single-cells coupled with lineage-tracing [3].
220 Therefore, these analyses demonstrate the sensitivity of *Milo* by identifying that a mTEC
221 progenitor state is depleted with age, a finding that was not resolved using clustering
222 approaches.

223

224 *Milo identifies compositional disorder in cirrhotic human liver*

225 To demonstrate the applicability of our method in multiple biological contexts, we next applied
226 *Milo* to a large dataset of hepatic cells isolated from 5 healthy and 5 cirrhotic human livers [2].
227 The original study assigned cells to multiple lineages, including immune, endothelial and
228 mesenchymal cells (Fig 5A-B). A key goal of the study was to ask whether different cell types
229 were differentially abundant between experimental samples taken from healthy and cirrhotic
230 tissue. In the original study, cells from each lineage were sub-clustered and these sub-clusters
231 were interrogated using a Poisson GLM to identify whether there were differential contributions
232 from cirrhotic and healthy donors.

233

234 To explore whether more subtle differences could be detected, we applied *Milo* analysis to
235 2696 neighbourhoods spanning the k-NN graph and identified 1404 neighbourhoods with
236 differential abundance (10% FDR; Fig 5C). To assess performance, we compared DA results
237 with those from the compositional analysis performed by Ramachandran et al. [2]. *Milo*
238 recovered DA neighbourhoods in all clusters identified as differentially abundant between
239 cirrhotic or uninjured tissue in the original study (Fig 5D).

240

241 Moreover, *Milo* identified multiple groups of neighbourhoods within the same pre-defined sub-
242 clusters that showed opposing directions of differential abundance between the control and
243 cirrhotic liver experimental samples (Fig 5D). In other words, within a sub-cluster, some
244 neighbourhoods were enriched for control experimental samples whilst others were enriched
245 for disease experimental samples. These patterns, exemplified by the T cell (2) and the
246 endothelial (5) compartments were obscured in the previous study due to the reliance on pre-
247 clustering (Fig 5D).

248

249 To further explore the biological meaning of these neighbourhoods, we first focused on the
250 hepatic endothelial cells, where we resolved disease specific subpopulations at higher
251 resolution than was possible by clustering-based analysis (Fig. 5D). *Milo* identified a gradient
252 of changes in neighbourhood abundance across this compartment, suggestive of a continuous
253 transition between healthy and diseased cell states (Fig 5E). To identify gene expression
254 signatures associated with this change, we performed differential expression analysis
255 between cells in DA neighbourhoods with positive and negative log fold changes, identifying
256 83 differentially expressed genes (FDR 10%; Methods) (Fig5F). In the cirrhosis-enriched
257 neighbourhoods, we recovered over-expression of known markers of scar-associated
258 endothelium, including *ACKR1*, *PLVAP* and *VWA1* (Fig. 5F) [2]. We also recovered over-
259 expression of genes associated with regulation of leukocyte recruitment, confirming the
260 validated immunomodulatory phenotype displayed by scar-associated tissue (Supp Fig 7A)
261 [22]. In addition, cirrhotic endothelium displays down-regulation of genes involved in response

262 to infection, endocytosis and immune complex clearance, including *FCN2*, *FCN3*, and
263 *FCGR2B* (Supp Fig 7B), which have been suggested as an additional component of cirrhosis-
264 associated immune dysfunction [23,24].

265

266 *Milo* also identified strong DA between healthy and cirrhotic cells in lineages that were
267 unexplored in the original study, such as the cholangiocyte compartment (Fig 5D).
268 Cholangiocytes are epithelial cells that line a three-dimensional network of bile ducts known
269 as the biliary tree, and cholangiocyte proliferation can be induced by a broad range of liver
270 injuries, in a process termed the ductular reaction [25]. However, the gene signatures
271 associated with this process in human cirrhosis are largely unexplored. Indeed, *Milo* recovered
272 an enrichment of disease-specific cholangiocytes (Supp Fig 7C-D), and differential gene
273 expression analysis detected strong over-expression of genes associated with calcium
274 signalling (Supp Fig 7E-F), a signalling pathway frequently dysregulated in liver disease and
275 a potential target for clinical intervention [26,27].

276

277 These analyses demonstrate the potential of using DA subpopulations detected by *Milo* to
278 recover known and novel signatures of disease-specific cell states.

279

280 **Discussion:**

281 Given the increasing number of complex single-cell datasets where multiple conditions are
282 assayed [18,19], *Milo* tackles a key problem: robustly determining sets of cells that are
283 differentially abundant between conditions without relying on pre-existing sets of clusters.
284 Moreover, *Milo* is fully interoperable with established single-cell analysis workflows and is
285 implemented as an open-source R software package [28] with documentation and tutorials at
286 <https://github.com/MarioniLab/miloR>.

287

288 The definition of neighbourhoods, as implemented in *Milo*, overcomes the main limitations of
289 standard-of-practice clustering-based DA analysis, whilst utilising a common data-structure in
290 single-cell analysis - graphs. A strength of our approach is that it is applicable to a wide range
291 of datasets with vastly different topologies, including gradual state transitions, thus removing
292 the need for time-consuming iterative sub-clustering and identifying subtle differences in
293 differential abundance that would otherwise be obscured (Fig 5D).

294

295 Recently, other clustering-free methods have been proposed to detect compositional
296 differences between experimental conditions [7,29]. However, these are most suitable for
297 pairwise comparisons between two biological conditions, and cannot be easily extended to
298 detect changes across continuous conditions (age, time points) or multifactorial conditions. By
299 modelling cell counts with a NB GLM, *Milo* can incorporate arbitrarily complex experimental
300 designs as demonstrated by our application of *Milo* to detect compositional changes in the
301 ageing mouse thymus (Fig 4) and across early embryonic development in mice (Supp Fig 3).
302 Moreover, we show how nuisance technical covariates can be included in the GLM model to
303 increase the power of DA testing in the presence of batch effects (Supp Fig 2-3).

304

305 Although we have addressed several important challenges, *Milo* is not free of limitations.
306 Firstly, the testing framework requires a replicated experimental design to estimate the
307 dispersions of counts for each condition. Whilst this is not strictly a limitation of *Milo*, it reflects
308 the importance of properly replicated experimental design in single-cell experiments. A
309 potential solution would be to use a mixed effects model utilising random
310 intercepts. Secondly, the detection of DA subpopulations by *Milo* requires a k-NN graph that
311 reflects the true cell-cell similarities in the phenotypic manifold; a limitation shared with all DA
312 methods that work on graphs or reduced dimensional spaces [30]. Additionally, while *Milo* can
313 account for artefacts such as batch effects during DA testing, we show that optimal results are
314 achieved when batch correction is performed prior to graph construction (Suppl. Note 2, Suppl.
315 Fig. 1, Suppl. Fig. 2). Thirdly, cells in a single neighbourhood do not necessarily represent a

316 unique biological subpopulation; a cellular state might span multiple neighbourhoods.
317 Accordingly, we search for marker genes of DA states by aggregating cells in adjacent and
318 concordantly DA neighbourhoods (Fig. 4E, 5F). One challenge of this approach is that rare
319 cell states may be represented by a small subset of neighbourhoods, thus making
320 identification of marker genes challenging. To overcome this problem one can either choose
321 a smaller value of k or alternatively construct a graph on cells from a particular lineage of
322 interest.

323

324 Following the generation of reference single-cell atlases for multiple organisms and tissues,
325 an increasing number of studies now focus on quantifying how cell populations are perturbed
326 in disease, ageing, and development, using, for example, large scaled pooled CRISPR
327 screens [31–33]. We envision that *Milo* will see use in all of these contexts. By leveraging a
328 cell-cell similarity structure, *Milo* is also applicable to single-cell assays other than scRNA-seq,
329 including multi-omic assays [34–38]. Thus, *Milo* has the potential to facilitate the discovery of
330 fundamental biological and medically important processes across multiple layers of molecular
331 regulation when they are assayed at single-cell resolution.

332

333 **Methods:**

334 *Milo*

335 *Milo* detects sets of cells that are differentially abundant between conditions by modelling
336 counts of cells in neighbourhoods on a k -NN graph. The workflow includes the following steps:

337

338 **(A) Construction of the k -NN graph:** Similar to many other tasks in single-cell analysis, *Milo*
339 uses a k -NN graph computed based on similarities in gene expression space as a
340 representation of the phenotypic manifold in which cells lie. We assume that the k -NN graph
341 is a faithful representation of the single cell phenotypes. Therefore, any batch effect should be

342 corrected prior to graph building to maximize the power of DA testing. In addition, nuisance
343 covariates can be incorporated in the experimental design of the NB GLM framework (practical
344 considerations and demonstrations on how to account for batch effects can be found in the
345 Supplementary Notes and Supp Fig 2-3). Throughout this paper, we build the k-NN graph
346 based on similarity in reduced principal component (PC) space.

347

348 **(B) Definition of cell neighbourhoods:** We define the neighbourhood n_j of cell j as the group
349 of cells that are connected to j by an edge in the graph. We refer to j as the index of the
350 neighbourhood. In order to define a representative subset of neighbourhoods that span the
351 whole k-NN graph, we implement a previously adopted algorithm to sample the index cells in
352 a graph [8,39] (See Supplementary Note 1.1.2 for a detailed description).

353

354 **(C) Counting cells in neighbourhoods:** For each neighbourhood we count the number of
355 cells from each experimental sample, constructing a neighbourhood x experimental sample
356 count matrix.

357

358 **(D) Testing for differential abundance in neighbourhoods:** To test for differential
359 abundance, we analyse neighbourhood counts using the quasi-likelihood (QL) method in
360 *edgeR*, similarly to the implementation in *Cydar* [6]. We fit a NB GLM to the counts for each
361 neighbourhood and use the QL F-test with a specified contrast to compute a P value for each
362 neighbourhood. Details of the statistical framework are provided in Supplementary Note 1.1.3

363

364 **(E) Controlling the Spatial FDR in neighbourhoods:** To control for multiple testing, we
365 adapt the Spatial FDR method introduced by *Cydar* [6]. The Spatial FDR can be interpreted
366 as the proportion of the union of neighbourhoods that is occupied by false-positive
367 neighbourhoods. To control the spatial FDR in the k-NN graph, we apply a weighted version
368 of the Benjamini-Hochberg (BH) method, where P values are weighted by the reciprocal of the

369 neighbourhood connectivity. As a measure of neighbourhood connectivity, we use the
370 Euclidean distance to the k-th nearest neighbour of the index cell for each neighbourhood.

371

372 A full description of *Milo* can be found in Supplementary Notes.

373

374 *Visualization of DA neighbourhoods*

375 To visualize results from differential analysis on neighbourhoods, we construct an abstracted
376 graph, where nodes represent neighbourhoods and edges represent the number of cells in
377 common between neighbourhoods. The size of nodes represents the number of cells in the
378 neighbourhood. The position of nodes is determined by the position of the sampled index cell
379 in the single-cell UMAP, to allow qualitative comparison with the single cell embedding.

380

381 *Mouse thymus analysis*

382 Single-cell data are available from ArrayExpress (accession E-MTAB-8560), additional meta-
383 data were acquired from Baran-Gale *et al.* [3] including cluster identity and highly variable
384 genes (HVGs). The dataset consists of 2327 single thymic epithelial cells that passed QC (see
385 [3] for details). Log-normalized gene expression values were used as input, along with 4906
386 HVGs, to estimate the first 50 principal components using a randomized PCA implemented in
387 the R package `irlba`, the first 40 of which were used for k-NN graph building (k=21) and
388 UMAP embedding. The refined sampling, using an initial random sampling of 30% of cells,
389 identified 363 neighbourhoods. Differential abundance testing used the mouse age as a linear
390 predictor variable, thus log fold changes are interpreted as the per-week linear change in
391 neighbourhood abundance. Neighbourhood cluster identity was assigned by taking the most
392 abundant cluster identity amongst neighbourhood cells.

393

394 Differential expression (DE) testing was performed on cells within neighbourhoods containing
395 a majority of cells from the Intertypical TEC cluster. This subset of neighbourhoods was
396 aggregated into 2 groups based on similarity of log fold change direction and number of
397 overlapping cells (≥ 10 cells). DE testing was performed comparing the log normalized gene
398 expression of neighbourhood cells between the more and less abundant groups using a linear
399 model implemented in the Bioconductor [40,41] package `limma` [42], using 1% FDR. Gene
400 Ontology Biological Process term analysis was performed on the 448 DE genes (adj. P-value
401 < 0.01) using the R package `enrichR` [43].

402

403 *Liver cirrhosis analysis*

404 The dataset including cell type annotations was downloaded from
405 <https://datashare.is.ed.ac.uk/handle/10283/3433> (GEO accession: GSE136103 [2]). The
406 dataset comprises 58358 cells, obtained from 5 healthy and 5 cirrhotic liver samples.
407 Following the pre-processing steps from the original publication, dimensionality reduction with
408 PCA was performed on the 3000 top highly variable genes (HVGs), calculated using
409 `modelGeneVar` and `getTopHVGs` from the R package `scrn` [44], and the top 11 PCs were
410 retained for k-NN graph building and UMAP embedding. Refined sampling identified 2676
411 neighbourhoods ($k=30$, initial proportion of sampled cells = 0.05). We run *Milo* to test for DA
412 between cirrhotic and healthy experimental samples. To assign cell type annotations to
413 neighbourhoods, we take the most frequent annotation between cells in each neighbourhood.
414 Neighbourhoods are generally homogeneous, with an average of 80% of cells belonging to
415 the most abundant cell type label.

416

417 For the focused analysis on the endothelial and cholangiocyte lineages, DE testing was
418 performed on the subset of neighbourhoods from the selected lineage. Neighbourhoods were
419 aggregated into 2 groups based on similarity of log fold change direction. DE testing was

420 performed summing the gene expression counts for each experimental sample and
421 neighbourhood group between the more and less abundant groups using the quasi-likelihood
422 test implemented in `edgeR` [9], using 10% FDR. GO term analysis was performed on the
423 significant DE genes using the R package `clusterProfiler` [45].

424

425 *Mouse gastrulation data*

426 The raw count matrix and batch corrected PCA matrix were downloaded via the R package
427 `MouseGastrulationData` [46]. To construct the uncorrected k-NN graph, raw counts
428 were log transformed and PCs were computed on the 5000 top variable genes. Refined
429 sampling identified 11895 neighbourhoods in the uncorrected graph and 8451
430 neighbourhoods in the MNN corrected graph ($k = 30$, initial proportion of sampled cells = 0.1).

431

432 *Scalability analysis*

433 We assessed the scalability of *Milo* by profiling the time taken to execute the workflow, starting
434 with the k-NN graph building step and concluding with the differential abundance testing. We
435 simulated a dataset of 200000 single-cells using the `dynTOY` package implemented in R [14].
436 With this large simulation we down-sampled to specific proportions, ranging from 1 to 100%,
437 and recorded the elapsed system time to complete the *Milo* workflow using the `system.time`
438 function in R [28]. In addition, we performed an equivalent analysis using the published data-
439 sets included in this manuscript: mouse thymus [3], human liver [2], and mouse gastrulation
440 [4]. All timings are reported in minutes.

441

442 To assess the memory usage of the *Milo* workflow we made use of the `Rprof` function in R
443 to record the total amount of memory used at each step. We followed the same approach as

444 above, down-sampling simulated and published datasets from 1 to 100% of the total cell
445 numbers. All memory usage is reported in megabytes (MB).

446

447 For both the system timing and memory usage we ran the simulated and published datasets
448 down-sampling analyses on a single node of the high-performance computing (HPC) cluster
449 at the Cancer Research UK - Cambridge Institute. Each node has 2x Intel Xeon E5-2698
450 2.20Ghz processors with 40 cores per node and 384GB DDR4 memory; cluster jobs were run
451 using a single core.

452

453 *Batch effect analysis*

454 **Simulated data:** we simulated a dataset representing a continuous trajectory of 5000 cells
455 using the R package `dynTOY` [14]. We assigned cells to one of 6 experimental samples and
456 samples to one of 2 conditions. In a specific region of the trajectory we assigned 80% of cells
457 to condition 'B' and 20% to condition 'A', simulating differential abundance between conditions.
458 A batch effect was incorporated by adding a gaussian random vector to the expression profiles
459 of all cells in two of six samples. We performed batch correction using the MNN method, as
460 implemented in the R package `batchelor` by the function `fastMNN`, using default
461 parameters [47]. Refined sampling identified 298 neighbourhoods in the uncorrected graph
462 and 317 neighbourhoods in the MNN corrected graph ($k = 10$, initial proportion of sampled
463 cells = 0.1). We ran *Milo* to test for DA between conditions, with and without accounting for
464 the batch effect in the experimental design (`design = ~ condition` or `design = ~`
465 `batch + condition`).

466

467 **Mouse gastrulation data:** We ran *Milo* to test for DA with 3 alternative experimental designs:
468 (A) test for DA across developmental time points (`design = ~ time point`), (B) test for
469 DA across developmental time points, accounting for the sequencing batch (`design = ~`

470 seq. batch + time point) and (C) test for DA between sequencing batches (design =
471 ~ seq. batch).

472

473 References:

- 474 1. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-
475 cell RNA-seq data. *Nat Rev Genet.* 2019;20: 273–282. doi:10.1038/s41576-018-0088-9
- 476 2. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT,
477 et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature.*
478 2019;575: 512–518. doi:10.1038/s41586-019-1631-3
- 479 3. Baran-Gale J, Morgan MD, Maio S, Dhalla F, Calvo-Asensio I, Deadman ME, et al.
480 Ageing compromises mouse thymus function and remodels epithelial cell
481 differentiation. *eLife.* 2020;9. doi:10.7554/eLife.56221
- 482 4. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al.
483 A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature.*
484 2019;566: 490–495. doi:10.1038/s41586-019-0933-9
- 485 5. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell
486 survey of the small intestinal epithelium. *Nature.* 2017;551: 333–339.
487 doi:10.1038/nature24489
- 488 6. Lun ATL, Richard AC, Marioni JC. Testing for differential abundance in mass cytometry
489 data. *Nat Methods.* 2017;14: 707–709. doi:10.1038/nmeth.4295
- 490 7. Zhao J, Jaffe A, Li H, Lindenbaum O, Sefik E, Jackson R, et al. Detection of
491 differentially abundant cell subpopulations discriminates biological states in scRNA-seq
492 data. *Bioinformatics;* 2019 Jul. doi:10.1101/711929

- 493 8. Gut G, Tadmor MD, Pe'er D, Pelkmans L, Liberali P. Trajectories of cell-cycle
494 progression from fixed cell populations. *Nat Methods*. 2015;12: 951–954.
495 doi:10.1038/nmeth.3545
- 496 9. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
497 expression analysis of digital gene expression data. *Bioinformatics*. 2010;26: 139–140.
498 doi:10.1093/bioinformatics/btp616
- 499 10. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-
500 Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:
501 4288–4297. doi:10.1093/nar/gks042
- 502 11. Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Detecting Differential Expression in
503 RNA-sequence Data Using Quasi-likelihood with Shrunken Dispersion Estimates. *Stat*
504 *Appl Genet Mol Biol*. 2012;11. doi:10.1515/1544-6115.1826
- 505 12. Sonesson C, Robinson MD. Bias, robustness and scalability in single-cell differential
506 expression analysis. *Nat Methods*. 2018;15: 255–261. doi:10.1038/nmeth.4612
- 507 13. Benjamini Y, Hochberg Y. Multiple Hypotheses Testing with Weights. *Scand J Stat*.
508 1997;24: 407–418. doi:10.1111/1467-9469.00072
- 509 14. Cannoodt R, Saelens W, Deconinck L, Saeys Y. dyngen: a multi-modal simulator for
510 spearheading new single-cell omics analyses. *Bioinformatics*; 2020 Feb.
511 doi:10.1101/2020.02.06.936971
- 512 15. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in
513 large networks. *J Stat Mech Theory Exp*. 2008;2008: P10008. doi:10.1088/1742-
514 5468/2008/10/P10008
- 515 16. Pons P, Latapy M. Computing communities in large networks using random walks (long
516 version). *ArXiv Phys E-Prints*. 2005.

- 517 17. Matthews BW. Comparison of the predicted and observed secondary structure of T4
518 phage lysozyme. *Biochim Biophys Acta BBA - Protein Struct.* 1975;405: 442–451.
519 doi:10.1016/0005-2795(75)90109-9
- 520 18. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM, et al. Cell
521 Hashing with barcoded antibodies enables multiplexing and doublet detection for single
522 cell genomics. *Genome Biol.* 2018;19. doi:10.1186/s13059-018-1603-1
- 523 19. McGinnis CS, Patterson DM, Winkler J, Conrad DN, Hein MY, Srivastava V, et al.
524 MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged
525 indices. *Nat Methods.* 2019;16: 619–626. doi:10.1038/s41592-019-0433-8
- 526 20. Akiyama T, Shimo Y, Yanai H, Qin J, Ohshima D, Maruyama Y, et al. The Tumor
527 Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the
528 Thymic Medullary Microenvironment and Self-Tolerance. *Immunity.* 2008;29: 423–437.
529 doi:10.1016/j.immuni.2008.06.015
- 530 21. Hikosaka Y, Nitta T, Ohigashi I, Yano K, Ishimaru N, Hayashi Y, et al. The Cytokine
531 RANKL Produced by Positively Selected Thymocytes Fosters Medullary Thymic
532 Epithelial Cells that Express Autoimmune Regulator. *Immunity.* 2008;29: 438–450.
533 doi:10.1016/j.immuni.2008.06.018
- 534 22. Wilkinson AL, Qurashi M, Shetty S. The Role of Sinusoidal Endothelial Cells in the Axis
535 of Inflammation and Cancer Within the Liver. *Front Physiol.* 2020;11.
536 doi:10.3389/fphys.2020.00990
- 537 23. Foldi I, Tornai T, Tornai D, Sipeki N, Vitalis Z, Tornai I, et al. Lectin-complement
538 pathway molecules are decreased in patients with cirrhosis and constitute the risk of
539 bacterial infections. *Liver Int.* 2017;37: 1023–1031. doi:10.1111/liv.13368

- 540 24. Ganesan LP, Kim J, Wu Y, Mohanty S, Phillips GS, Birmingham DJ, et al. FcγRIIb on
541 Liver Sinusoidal Endothelium Clears Small Immune Complexes. *J Immunol.* 2012;189:
542 4981–4988. doi:10.4049/jimmunol.1202017
- 543 25. Sato K, Marzioni M, Meng F, Francis H, Glaser S, Alpini G. Ductular Reaction in Liver
544 Diseases: Pathological Mechanisms and Translational Significances: Liver Injury and
545 Regeneration. *Hepatology.* 2019;69: 420–430. doi:10.1002/hep.30150
- 546 26. Oliva-Vilarnau N, Hankeova S, Vorrink SU, Mkrтчian S, Andersson ER, Lauschke VM.
547 Calcium Signaling in Liver Injury and Regeneration. *Front Med.* 2018;5.
548 doi:10.3389/fmed.2018.00192
- 549 27. Rodrigues M, Gomes D, Nathanson M. Calcium Signaling in Cholangiocytes: Methods,
550 Mechanisms, and Effects. *Int J Mol Sci.* 2018;19: 3913. doi:10.3390/ijms19123913
- 551 28. R Core Team. R: A Language and Environment for Statistical Computing. Vienna,
552 Austria: R Foundation for Statistical Computing; 2017. Available: [https://www.R-](https://www.R-project.org)
553 [project.org](https://www.R-project.org)
- 554 29. Burkhardt DB, Stanley JS, Tong A, Perdigoto AL, Gigante SA, Herold KC, et al.
555 Quantifying the effect of experimental perturbations in single-cell RNA-sequencing data
556 using graph signal processing. *Bioinformatics;* 2019 Jan. doi:10.1101/532846
- 557 30. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al.
558 Orchestrating single-cell analysis with Bioconductor. *Nat Methods.* 2019 [cited 19 Nov
559 2020]. doi:10.1038/s41592-019-0654-x
- 560 31. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq:
561 Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic
562 Screens. *Cell.* 2016;167: 1853-1866.e17. doi:10.1016/j.cell.2016.11.038

- 563 32. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al.
564 Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*.
565 2017;14: 297–301. doi:10.1038/nmeth.4177
- 566 33. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting
567 Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*.
568 2016;167: 1883-1896.e15. doi:10.1016/j.cell.2016.11.039
- 569 34. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK,
570 Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single
571 cells. *Nat Methods*. 2017;14: 865–868. doi:10.1038/nmeth.4380
- 572 35. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, et al. Joint
573 profiling of chromatin accessibility and gene expression in thousands of single cells.
574 *Science*. 2018;361: 1380–1385. doi:10.1126/science.aau0730
- 575 36. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and
576 chromatin accessibility in the same cell. *Nat Biotechnol*. 2019;37: 1452–1457.
577 doi:10.1038/s41587-019-0290-0
- 578 37. Zhu C, Yu M, Huang H, Juric I, Abnoui A, Hu R, et al. An ultra high-throughput method
579 for single-cell joint analysis of open chromatin and transcriptome. *Nat Struct Mol Biol*.
580 2019;26: 1063–1070. doi:10.1038/s41594-019-0323-x
- 581 38. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential
582 Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*. 2020;183: 1103-
583 1116.e20. doi:10.1016/j.cell.2020.09.056
- 584 39. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, et al. Wishbone
585 identifies bifurcating developmental trajectories from single-cell data. *Nat Biotechnol*.
586 2016;34: 637–645. doi:10.1038/nbt.3569

- 587 40. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al.
588 Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*.
589 2015;12: 115–121. doi:10.1038/nmeth.3252
- 590 41. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al.
591 Bioconductor: open software development for computational biology and
592 bioinformatics. *Genome Biol*. 2004;5: R80. doi:10.1186/gb-2004-5-10-r80
- 593 42. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
594 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*.
595 2015;43: e47–e47. doi:10.1093/nar/gkv007
- 596 43. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr:
597 a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids*
598 *Res*. 2016;44: W90–W97. doi:10.1093/nar/gkw377
- 599 44. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of
600 single-cell RNA-seq data with Bioconductor. *F1000Research*. 2016;5: 2122.
601 doi:10.12688/f1000research.9501.2
- 602 45. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing
603 Biological Themes Among Gene Clusters. *OMICS J Integr Biol*. 2012;16: 284–287.
604 doi:10.1089/omi.2011.0118
- 605 46. Griffiths J, Lun A. MouseGastrulationData: Single-Cell Transcriptomics Data across
606 Mouse Gastrulation and Early Organogenesis. 2020. Available:
607 <https://github.com/MarioniLab/MouseGastrulationData>
- 608 47. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-
609 sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*.
610 2018;36: 421–427. doi:10.1038/nbt.4091

611

612 Code and data availability:

613 *Milo* is implemented as an open-source package in R: <https://github.com/MarioniLab/miloR>.

614 Code used to generate figures and perform analyses can be found at
615 https://github.com/MarioniLab/milo_analysis_2020.

616

617 Acknowledgments & Funding:

618 We thank Shila Ghazanfar for feedback on the method, Natsuhiko Kumasaka for comments
619 on the manuscript, Chenqu Suo, Veronika Kedlian and Rasa Elmentaite for feedback on the
620 software package. JCM acknowledges core funding from EMBL and core funding from Cancer
621 Research UK (C9545/A29580), which supports MDM. ED and SAT acknowledge Wellcome
622 Sanger core funding (WT206194). NCH is supported by a Wellcome Trust Senior Research
623 Fellowship in Clinical Science (ref. 219542/Z/19/Z), Medical Research Council, and a Chan
624 Zuckerberg Initiative Seed Network Grant.

625

626 Author contributions:

627 ED, MDM & JCM conceived the method idea. ED & MDM developed the method, wrote the
628 code and performed analyses. ED, MDM, SAT & NCH interpreted the results. ED, MDM, SAT,
629 NCH & JCM wrote and approved the manuscript. MDM & JCM oversaw the project.

630

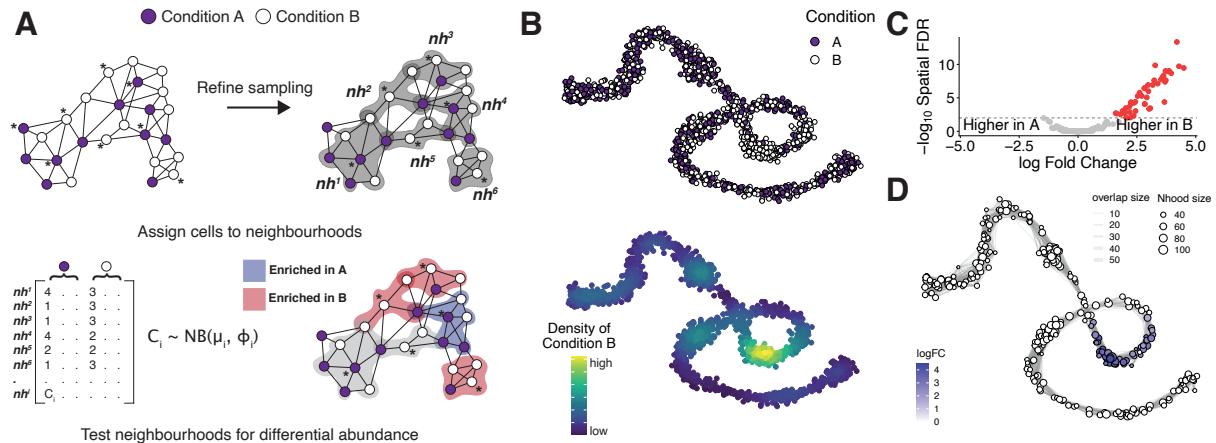


Figure 1: Detecting perturbed cell states as differentially abundant graph neighbourhoods (A) Schematic of the Milo workflow. Neighbourhoods are defined on index cells, selected using a graph sampling algorithm. Cells are quantified according to the experimental design to generate a counts table. Per-neighbourhood 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 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 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.

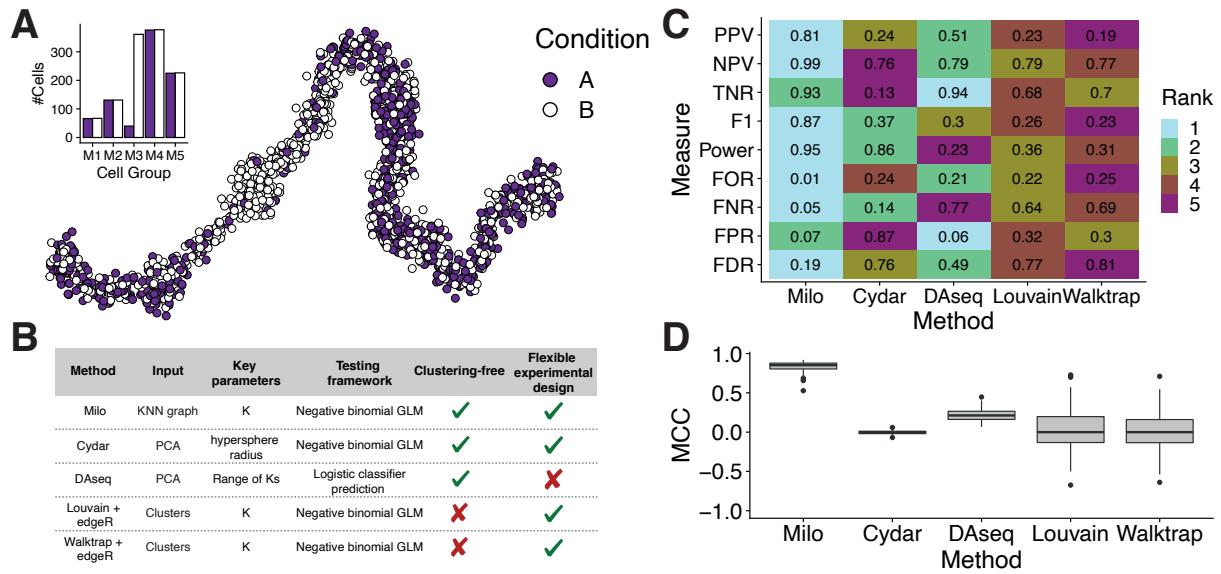


Figure 2: **Milo outperforms alternative differential abundance testing approaches** (A) An example simulated trajectory of cells drawn from 5 groups with cells assigned to either conditions ‘A’ (purple points) or ‘B’ (white points). Inset bar plot shows the number of cells (y-axis) assigned to each condition according to the group from which cells were sampled (x-axis). (B) A table describing the different methods compared to Milo, along with the input, key parameters and an overview of the testing framework for each. (C) Rankings of DA testing methods across a number of measures to determine performance. Each box is coloured by the ranking of each measure 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, TNR: true negative rate, F1: F1 score, FOR: false omission rate, FNR: false negative rate, FPR: false positive rate, FDR: false discovery rate. (D) 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.

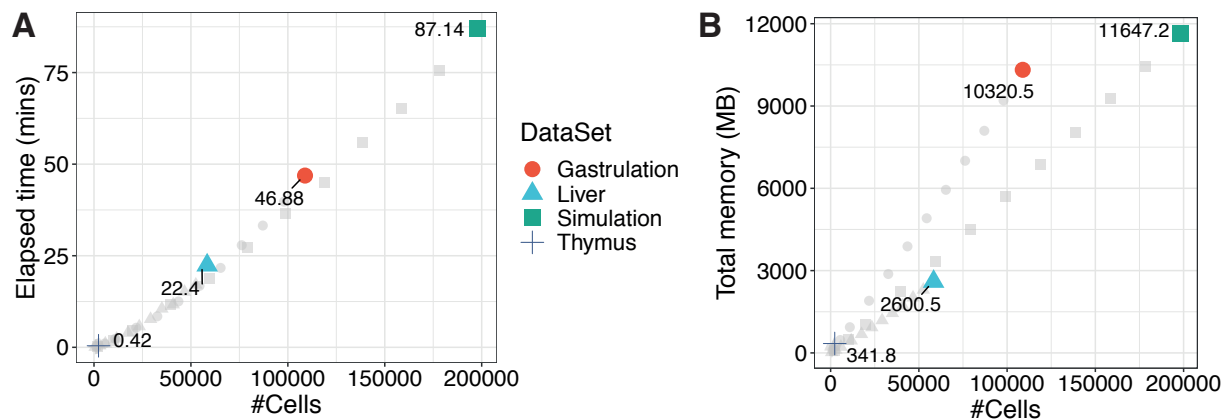


Figure 3: **Milo efficiently scales to large data sets** (A) Run time (y-axis) of the Milo workflow from graph building to differential abundance testing. Each point represents a down-sampled dataset, denoted by shape. Coloured points show the total number of cells in the full dataset labelled by the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

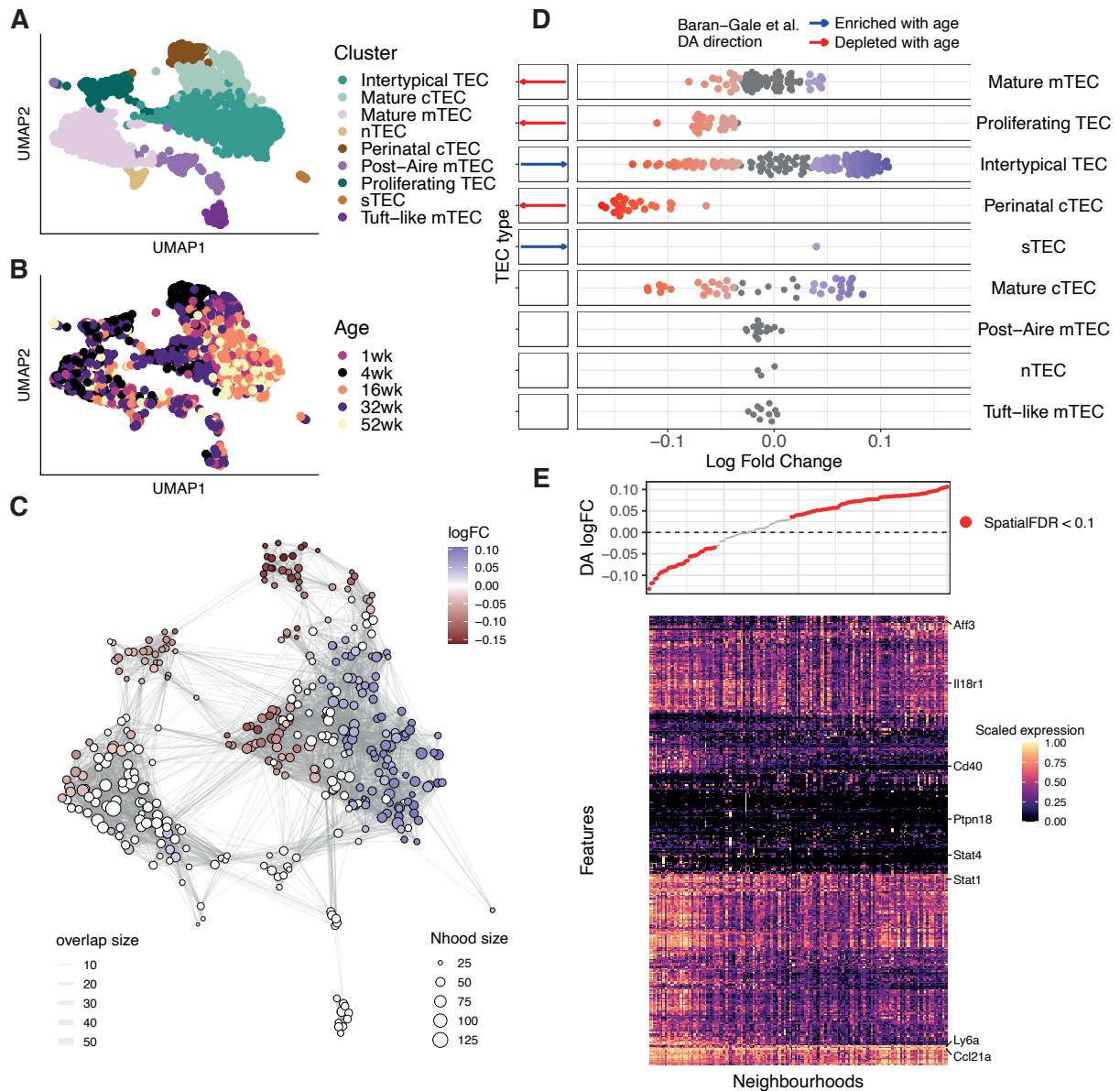


Figure 4: Milo identifies the decline of a fate-biased precursor in the ageing mouse thymus (A-B) A UMAP of single thymic epithelial cells sampled from mice aged 1-52 weeks old. Points are labelled according to their annotation in Baran-Gale et al. 2020 (A) and mouse age (B) (C) A graph representation 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. (2020) are annotated in the left side bar. (E) 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 1%). Expression values for each gene are scaled between 0 and 1. The top panel denotes the neighbourhood DA log fold-change.

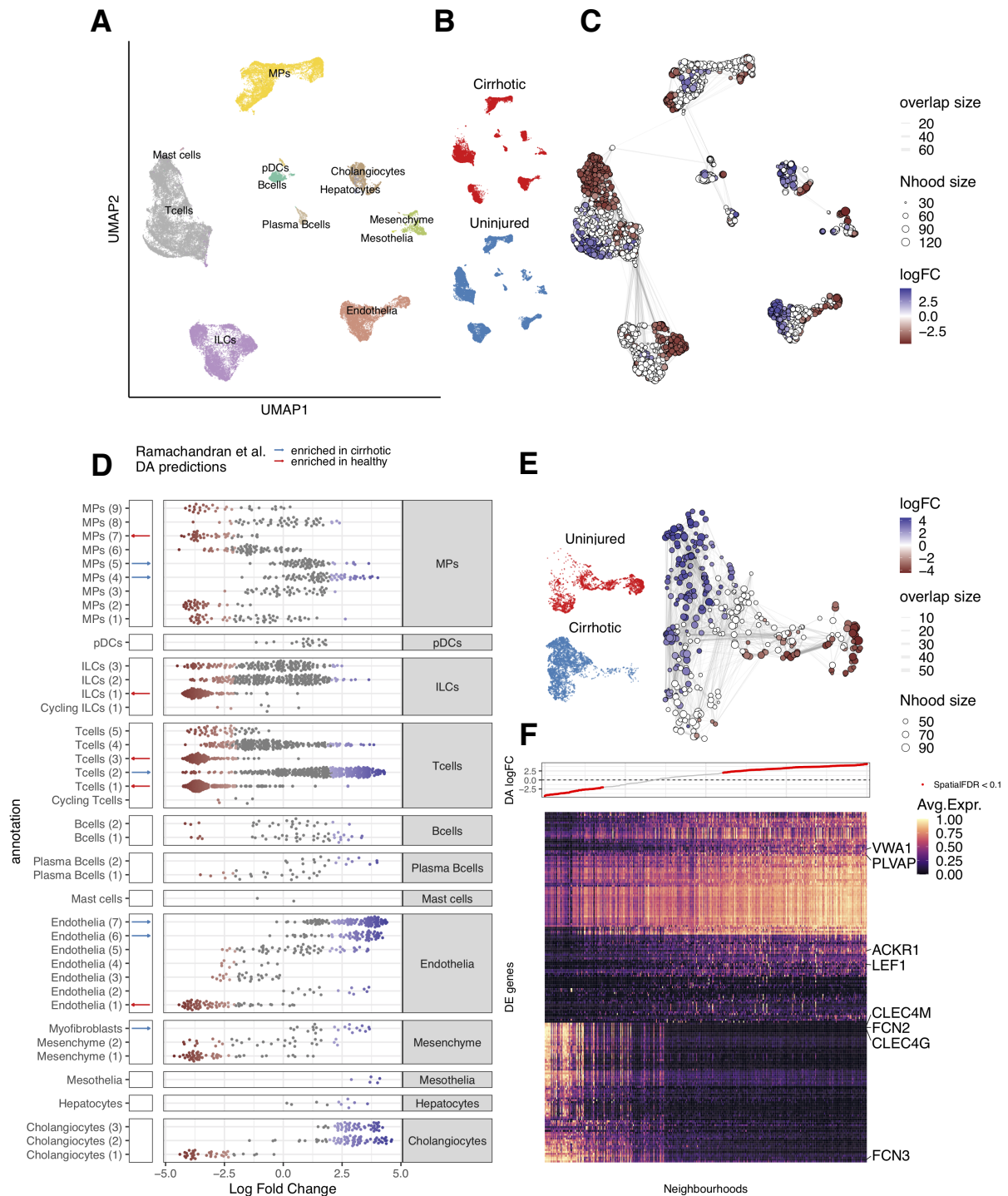


Figure 5: Milo identifies the compositional disorder in cirrhotic liver (A-B) UMAP embedding of 58358 cells from healthy ($n = 5$) and cirrhotic ($n = 5$) human livers. Cells are colored by cellular lineage (A) and injury condition (B) (C) Graph representation of neighbourhoods identified by Milo. Nodes 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. (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 between 0 and 1. The top panel denotes the neighbourhood DA log fold-change.