Supplemental methods

Numerical probabilistic model

The heuristic proceeds as follows: Initially, a low (not deep) cutting depth is set, namely *cut_level=0.5* (50% of total tree height). Next, the heuristic tries to reduce the cutting depth (*cut_level*) as much as possible by checking two conditions at every iteration of the inner loop (repeat_B): *condition1* requires the total number of clusters to be lower than 25% of the total cells. This condition avoids over-clustering and excessive fragmenting. On the other hand, *condition2* depends on a variable that is increased after every failed attempt in the inner loop (repeat_B). Specifically, *condition2* requires the largest cluster to be smaller than *max_cluster_size*cell_number*, where *cell_number* is the total number of cells and *max_cluster_size* is a variable whose initial value is set inversely proportional to the amount of cells (for instance *max_cluster_size=0.2* when *cell_number<1250* or *max_cluster_size=0.09* when *cell_number>5000*). The initial value of *max_cluster_size* is inversely proportional to the total cell number because higher cell numbers allow more fragmented clustering whilst preserving cell numbers in the clusters which are high enough to compute a smooth numerical model. Essentially, the two conditions work in synergy to avoid over-clustering (*condition1*) and under-clustering (*condition2*).

In a typical dataset, at the first attempt (with *cut_level=0.5), condition1* will be certainly satisfied (cut of low depth will yield few clusters) whereas *condition2* will be unsatisfied, which leads to an increase of the cutting depth. If the whole inner loop is completed without satisfying both conditions, *max_cluster_size* is increased, to relax the threshold guarding the under-clustering and facilitate the satisfiability of *condition2.*

TREE CUT PSEUDOCODE

repeat_A

repeat_B

perform tree cutting at *depth* = *cut_level* (%)

if *condition1* and *condition2* are satisfied **stop**

else

reduce *cut_level* (for instance *cut_level=cut_level+0*.01) and **stop** if *cut_level*>1

if exited from previous cycle and cut_level>0 then **done!**

else we have failed, *max_group_size* is increased, *cut_level* restored to 0.5 and start again from the main cycle.

Markers of CR cells

To calculate new markers for CR cells we selected genes that were i) markers for CR cells, using the 1.3M convoluted dataset (1,291 genes with Z-score>6, **Supplemental Table S2**) and ii) uniformly expressed within subtypes of CR cells, using the deconvoluted CR dataset. Genes without significant changes of expression (max fold change <1.5) between subtypes (CR1-CR8) of CR cells were labeled as uniformly expressed in CR cells. A total of 501 genes including *Reln* satisfied both requirements. However, restricting the intersection to strong CR makers showing at least 8-fold increased expression in CR cells and Z-score>15 resulted in six high confidence makers: *Reln* (Z-score=35), *Cacna2d2* (Z-score=30), *Eya2* (Z-score=21), *Tex15* (Z-score=19), *Cpeb1* (Z-score=17), *Vmn2r1* (Z-score=15).

Patient derived neural progenitor cells

Skin fibroblasts from two patients with Williams-Beuren (WB) and two with 7q11.23 microduplication (Dup7) syndrome were reprogrammed to induced pluripotent stem (iPS) cells by retroviral delivery of the pluripotency factors POU5F1, SOX2, KLF4 and MYC, at the Centre of Regenerative Medicine in Barcelona (CMR[B]). Individual iPS cells were picked to generate single clone colonies that were expanded and fully characterized (Martí et al. 2013). Briefly,

genomic stability was confirmed by karyotype; integration and silencing were verified by PCR and quantitative RT-PCR; pluripotency was demonstrated by Alkaline Phosphatase staining and expression of pluripotency markers by immunocytochemistry. Finally, the capacity to differentiate to mesoderm, ectoderm and endoderm germ layers both *in vitro* and *in vivo* was verified by embryoid bodies and teratoma formation followed by immunostaining. All iPS cells were deposited in the Stem Cell Bank repository of the Instituto de Salud Carlos III ([SWB]FiPS1-R4F-5, [SWB]FiPS-4F-5-6, [DUP7]FiPS-4F-3-1, [DUP7]FiPS4-R4F-2).

We generated neural progenitor cells from iPS cells following the Gibco protocol based on PSC Neural Induction Medium (NIM). Briefly, differentiation of iPS cell colonies was performed by seven days culture in NIM, followed by several passages of maturation in Complete Neural Expansion Medium (Neurobasal Medium, Advanced DMEM⁄F-12, and Neural Induction Supplement). Confirmation of expression of NPC markers was done by immunocytochemistry (Human Neural Stem Cell Immunocytochemistry Kit, Gibco). After 4-7 passages, NPC were detached with Accutase (Gibco) and resuspended. Single-cells were sorted in a BD Influx cell sorter to MARS-Seq plates (see below) for single-cell RNA sequencing (Flow Cytometry Core Facility, Univeritat Pompeu Fabra).

Library preparation and sequencing

To construct single-cell libraries from poly(A)-tailed RNA, we applied massively parallel single-cell RNA sequencing (MARS-Seq) (Jaitin et al. 2014; Paul et al. 2015). Briefly, single cells were FACS isolated into 384-well plates, containing lysis buffer (0.2% Triton X-100 (Sigma-Aldrich); RNase inhibitor (Invitrogen)) and reverse-transcription (RT) primers. The RT primers contained the single-cell barcodes and unique molecular identifiers (UMIs) for subsequent de-multiplexing and correction for amplification biases, respectively. Single-cell lysates were denatured and immediately placed on ice. The RT reaction mix, containing SuperScript III reverse transcriptase (Invitrogen) was added to each sample. In the RT reaction, spike-in artificial transcripts (ERCC, Ambion) were included at a dilution of $1:16x10^6$ per cell. After RT, the cDNA was pooled using an automated pipeline (epMotion, Eppendorf). Unbound primers were eliminated by incubating

the cDNA with exonuclease I (NEB). A second pooling was performed through cleanup with SPRI magnetic beads (Beckman Coulter). Subsequently, pooled cDNAs were converted into double-stranded DNA with the Second Strand Synthesis enzyme (NEB), followed by clean-up and linear amplification by T7 *in vitro* transcription overnight. Afterwards, the DNA template was removed by Turbo DNase I (Ambion) and the RNA was purified with SPRI beads. Amplified RNA was chemically fragmented with Zn2+ (Ambion), then purified with SPRI beads. The fragmented RNA was ligated with ligation primers containing a pool barcode and partial Illumina Read1 sequencing adapter using T4 RNA ligase I (NEB). Ligated products were reverse-transcribed using the Affinity Script RT enzyme (Agilent Technologies) and a primer complementary to the ligated adapter, partial Read1. The cDNA was purified with SPRI beads. Libraries were completed through a PCR step using the KAPA Hifi Hotstart ReadyMix (Kapa Biosystems) and a forward primer that contains Illumina P5-Read1 sequence and the reverse primer containing the P7-Read2 sequence. The final library was purified with SPRI beads to remove excess primers. Library concentration and molecular size were determined with High Sensitivity DNA Chip (Agilent Technologies). The libraries consisted of 192 single-cell pools. Multiplexed pools (2) were run in one Illumina HiSeq 2500 Rapid two lane flow cell following the manufacturer's protocol. Primary data analysis was carried out with the standard Illumina pipeline. We produced 52 nt of transcript sequence reads.

Data processing

The MARS-Seq technique takes advantage of two-level indexing that allows the multiplexed sequencing of 192 cells per pool and multiple pools per sequencing lane. Sequencing was carried out as paired-end reads; wherein the first read contains the transcript sequence and the second read the cell barcode and UMI. Quality check of the generated reads was performed with the FastQC quality control suite. Samples that reached the quality standards were then processed to deconvolute the reads to single-cell level by de-multiplexing according to the cell and pool barcodes. Reads were filtered to remove poly(T) sequences. Reads were mapped with the RNA pipeline of the GEMTools 1.7.0 suite (Marco-Sola et al. 2012) using default parameters (6% of

mismatches, minimum of 80% matched bases, and minimum quality threshold of 26) and the genome references for human (GENCODE release 25, assembly GRCh38). Gene quantification was performed using UMI corrected transcript information to correct for amplification biases, collapsing read counts for reads mapping on a gene with the same UMI (allowing an edit distance up to 2 nt in UMI comparisons). Only unambiguously mapped reads were considered.

Data filtering

The analysis of spike-in control RNA content allowed us to identify empty wells and barcodes with more than 15% of reads mapping to spike-in artificial transcripts were discarded. In addition, cells with less than 60% of reads mapping on the reference genome or more than $2x10^6$ total reads were discarded (potential aggregates). Additionally, we also discarded cells presenting 1) low library complexity (<1,000 detected genes) 2) excessive library complexity (>10,000 genes detected) 3) excessive mitochondrial content (>20% of mitochondrial reads) 4) excessive ribosomal content (>10% of ribosomal reads) 5) excessive content of intronic reads (>20% of mapped reads). Overall, 73 cells did not satisfy these quality requirements and were discarded.

Simulated datasets

For data simulation we applied *Splatter* (Zappia et al. 2017) estimating parameters from NPC (*sim_NPC*) and a droplet-based experiment (2,520 random cells from: 1.3 Million Brain Cells from E18 Mice, 10x Genomics; *sim_10xG*). The datasets differed in the number of detected genes per cell, sparsity and heterogeneity. We recreated highly similar distributions of gene expression means and variances, cell library sizes and zeros counts as well as relationships of mean-variance and mean-zeros (**Fig. 2D** and **Fig. S4A**). We preserved the number of cells and genes as in the original dataset and defined groups of different proportions across multiple sequencing pools. The dimensions of the gene x cell matrices were 41,020 x 1,847 and 27,998 x 2,520 in sim_NPC and sim_10xG, respectively. Each tool has been applied on the complete dataset at the model-building step, before defining groups of proportions 1:1 (1x), 1:2 (2x) and 1:10 (10x). The number of DE genes between groups ranged from 18% to 30% of the total number of DE genes (around 47% of

total genes), being lowest at 10x and highest at 2x cases. While the composition of DE genes was similar in up-regulated and down-regulated genes, ratios of gene average means between groups could reach levels of expression magnitude up to twice as much as in sim_NPC. The datasets further differed in the proportion of outlier genes, which was around 1% in sim_10xG and ~2.5% in the sim_NPC.

ROC curves and pAUCs have been performed using the R package pROC (Robin et al. 2011). In all comparisons, only genes tested by all methods were considered. Genes were ranked by nominal *p*-values, which we used to define a score as 1-p, indicating the outcome of the prediction (DE or non-DE) for each tool. Predictions and true gene labels were assessed at different thresholds of these scores to compute relative specificity and sensitivity coordinates for ROC curves.

Supplemental references

Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, et al. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**: 776–779.

Marco-Sola S, Sammeth M, Guigó R, Ribeca P. 2012. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**: 1185–1188.

Martí M, Mulero L, Pardo C, Morera C, Carrió M, Laricchia-Robbio L, Esteban CR, Izpisua Belmonte JC. 2013. Characterization of pluripotent stem cells. *Nat Protoc* **8**: 223–253.

Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, et al. 2015. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**: 1663–1677.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**: 77.

Zappia L, Phipson B, Oshlack A. 2017. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol* **18**: 174.