Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR–Cas9 barcodes by scGESTALT

Bushra Raj [1,2]*, James A. Gagnon [1,2,3], Alexander F. Schier [1,2,4,5,6,7]

Lineage relationships among the large number of heterogeneous cell types generated during development are difficult to reconstruct in a high-throughput manner. We recently established a method, scGESTALT, that combines cumulative editing of a lineage barcode array by CRISPR–Cas9 with large-scale transcriptional profiling using droplet-based single-cell RNA sequencing (scRNA-seq). The technique generates edits in the barcode array over multiple timepoints using Cas9 and pools of single-guide RNAs (sgRNAs) introduced during early and late zebrafish embryonic development, which distinguishes it from similar Cas9 lineage-tracing methods. The recorded lineages are captured, along with thousands of cellular transcriptomes, to build lineage trees with hundreds of branches representing relationships among profiled cell types. Here, we provide details for (i) generating transgenic zebrafish; (ii) performing multi-timepoint barcode editing; (iii) building scRNA-seq libraries from brain tissue; and (iv) concurrently amplifying lineage barcodes from captured single cells. Generating transgenic lines takes 6 months, and performing barcode editing and generating single-cell libraries involve 7 d of hands-on time. scGESTALT provides a scalable platform to map lineage relationships between cell types in any system that permits genome editing during development, regeneration, or disease.

Introduction

Single-cell technologies have rapidly progressed to enable large-scale profiling of cell-to-cell heterogeneity in tissues during development, homeostasis, and disease[1–3]. It is now feasible to characterize cell identities based on transcriptional, epigenetic, and protein marker landscapes by capturing tens of thousands of single cells or nuclei in a short time frame and at a low cost. These methods rely on droplet microfluidics, high-density arrays, or combinatorial indexing for massively parallel experiments[4–17]. Among the various single-cell assays, scRNA-seq is widely used to categorize cell types using gene expression signatures in many tissues and animals, including humans, without requiring extensive prior knowledge of cell composition. Concomitant with technological advances, substantial efforts have been made to develop computational approaches to correct for batch effects, enable efficient data clustering, and compare datasets between different conditions or methods[18–21]. The resultant cell-type catalogs are revealing new insights into cellular heterogeneity, allowing de novo identification of rare cell types and gene markers, delineation of cellular dynamics via developmental trajectories and cell cycle states, and examination of changes in the structure of profiled cell types during disease[22–38].

Large-scale reconstruction of lineage relationships at single-cell resolution

Despite the ability to define cell types on a large scale, it has been challenging to determine the lineage relationships between cell types and to characterize the progressive steps in their specification[39–41]. Reconstruction of lineage trees that capture the patterns and timings of lineage segregation can provide information about the structural and functional organization of tissues and their dysregulation in congenital disease and cancer; this may eventually provide a means for translation into in vitro reprogramming and organoid systems. The realization of such lineage trees would require simultaneous measurements of cell-type identities and lineage relationships at high throughput.

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. [2]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. [3]Department of Biology, University of Utah, Salt Lake City, UT, USA. [4]Biozentrum, University of Basel, Basel, Switzerland. [5]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA. [7]Center for Brain Science, Harvard University, Cambridge, MA, USA. *e-mail: bushranraj@gmail.com

Lineage-tracing methods using barcoded viral libraries, Cre–loxP recombination, somatic mutations and transposon tagging have been linked with next-generation sequencing but have not been combined with simultaneous cell-type identification at single-cell resolution[42–48]. Coupling transcriptomic and lineage readouts with scRNA-seq would provide a powerful assay that could perform unbiased lineage tree reconstruction.

Lineage recording with CRISPR–Cas9 genome editing

The CRISPR–Cas9 nuclease system has become a widely used tool in genome engineering, with applications extending beyond gene editing to regulation of gene expression, epigenetic editing, base editing, knockout screening, in situ labeling, RNA tracking, protein localization, chromatin interaction profiling, and signal recording, among others[49–65]. Recently, CRISPR–Cas9 has been adapted for use in several lineage-tracing strategies, including GESTALT, LINNAEUS, MEMOIR, and ScarTrace[66–71]. In general, these methods rely on the introduction of Cas9-induced stochastic mutations (insertions or deletions during DNA repair following double-strand breaks) to predefined target sites in transgenes, and use the edited sequences as lineage barcodes for clonal tracing and lineage tree reconstruction.

We have established an approach, scGESTALT[72], which extends the CRISPR–Cas9 lineage recording capabilities of GESTALT[66] with large-scale transcriptional profiling using droplet-based scRNA-seq (inDrops)[8], to reconstruct developmental lineage trees, using zebrafish as a model system (Fig. 1). Our method makes use of combinatorial and cumulative additions of Cas9-induced mutations to a genomic CRISPR barcode array at multiple timepoints to write permanent records of lineage histories as cells divide[72]. The edited barcodes are expressed from a transgene, captured within cellular transcriptomes, and sequenced at single-cell resolution with inDrops, which enables the indexing of >12,000 cells/h into nanoliter-size droplets[73]. Using tools adapted from phylogenetics, branching lineage trees are generated, using patterns of mutations shared between barcodes, with the tips of the trees representing profiled cell types defined using associated scRNA-seq gene expression signatures.

Here, we explain in detail how to perform scGESTALT multi-timepoint barcode editing with transgenic zebrafish. We describe the generation of scRNA-seq transcriptome and lineage barcode libraries using the juvenile brain as an experimental system. Due to the inherent scalability of droplet-based scRNA-seq, scGESTALT can be used to profile thousands of cells from a single animal. Furthermore, scGESTALT can be adapted for use in any organism or in vitro model system that is compatible with genome editing and single-cell technologies to characterize lineages in development, regeneration, cellular reprogramming, and disease states.

Comparison with other lineage-tracing methods

To generate large-scale lineage trees or perform large-scale clonal analysis, lineage-tracing methods should possess high multiplexing capacities for unique and unambiguous labeling of ancestral cells. Approaches that incorporate high-throughput sequencing of diverse DNA or RNA lineage records are advantageous over the limited spectral labeling capacity of fluorescent reporters. For example, virally delivered DNA barcode libraries can, in principle, encode high sequence complexity. However, the relative representation of each barcode in the library and its successful recovery can confound data interpretation. If subsets of barcodes are overrepresented in the viral library, it will result in coalescence of independent clones marked by the same barcode into a single clone, and rare barcodes would be underrepresented and potentially missed during data analysis. Inefficiencies in viral transduction of certain cell types can also preclude its use in various tissues or organisms. Furthermore, as the barcodes are not mutable (i.e., cannot be edited during development), only clonal analysis can be performed. In this case, all descendants of an ancestrally marked founder cell can be identified (all cells in a clone inherit the same barcode sequence), but lineage relationships between the clonal descendants cannot be inferred. Another approach, known as Polylox barcoding, has been recently reported to potentially generate hundreds of thousands of in vivo barcode sequences (via Cre-induced recombination of a synthetic DNA cassette inserted into the genome); however, its compatibility with single-cell analysis and use for generating lineage trees have not been demonstrated[44]. Somatic mutations accumulated over an animal's lifespan can be harnessed to generate detailed lineage trees and do not require building transgenics or introducing barcoding reagents[43]. However, methods for retrieving somatic mutations from single cells involve whole-genome amplification and sequencing strategies that can introduce technical artifacts and are currently
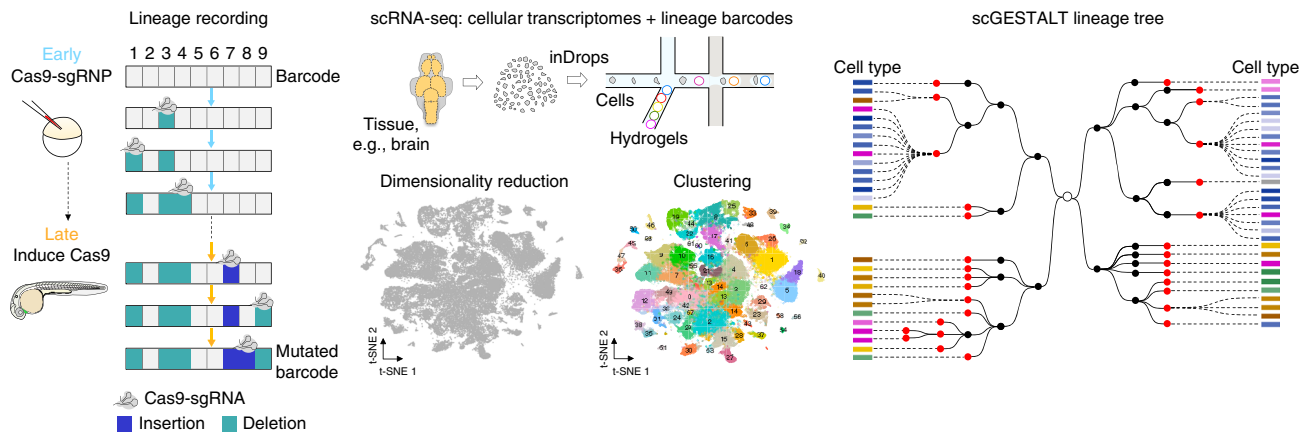
Fig. 1 | Simultaneous recovery of lineages and cell types at single-cell resolution using scGESTALT. Left, During early and late developmental time periods, CRISPR–Cas9-induced indels record cell lineages as mutated genomic barcode sequences. Middle, top, Tissues of interest, such as brain, are dissociated into a single-cell suspension and loaded into a microfluidics device (inDrops). Single cells are encapsulated in droplets and indexed using hydrogels (color-coded to indicate different cell-identifier primers) that are coated with oligodT primers. Polyadenylated cellular transcriptomes and scGESTALT lineage barcodes bind to the oligodT sequences and are simultaneously extracted from the same cells. Transcriptome libraries are sequenced to generate gene expression matrices for thousands of single cells. Middle, bottom left, Gene expression profiles are used to perform dimensionality reduction using principal component analysis and visualized in two dimensions on a t-distributed stochastic neighbor embedding (t-SNE) plot. Single cells are represented as gray dots on the plot shown. Middle, bottom right, A modularity-based clustering algorithm (Louvain) is used to cluster cells into discrete cell types using top principal components. A t-SNE plot of 58,492 cells from $n = 22$ animals is color-coded to show 63 distinct clusters that were identified from zebrafish juvenile brains[72]. Right, scGESTALT libraries are sequenced to obtain lineage barcodes of profiled single cells. The inDrops index sequences are used to match transcriptomes and lineage barcodes for the same cells. Cell-lineage trees are generated using maximum parsimony based on patterns of shared edits. Black and red nodes represent early and late barcode edits, respectively. Dashed lines connect profiled single cells to nodes on the tree. Cells connected to the same node are clonal (i.e., contain the same lineage barcode). Each cell is categorized into a discrete cell type (color-coded rectangles) based on prior transcriptional clustering analysis. Brown shades represent forebrain cell types, blue shades represent midbrain cell types, green shades represent hindbrain cell types, and pink shades represent progenitor cell types. This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25–08. Adapted with permission from Raj et al.[72], Springer Nature.

cost-prohibitive for analyzing large single-cell populations[43,74,75]. By contrast, scGESTALT barcodes can be recovered by targeted enrichment from a predefined genetic locus. Coupling scGESTALT to the inDrops scRNA-seq platform also provides several advantages, such as high scalability, low library preparation cost (currently ~$0.06 per cell), high single-cell capture rates (>70% of cells), low cell input requirements, no cell capture size bias, and simultaneous transcriptome profiling for cell type or cell state identification[73]. Recently, inDrops has also been coupled with Tol2 transposon-based barcoding (TracerSeq) for lineage tracing in zebrafish[37].

Two other methods have linked CRISPR–Cas9 lineage tracing with scRNA-seq in zebrafish[68,69]. These techniques generate barcode edits in a multi-copy transgenic reporter (in tandem or in distinct genomic loci) by injecting Cas9 protein or mRNA along with one sgRNA targeting a reporter-specific sequence. This restricts lineage analysis to early embryonic stages, as Cas9 editing terminates upon degradation of Cas9/sgRNA or mutation of all potential sites, and the observed number of distinct lineage barcodes is limited to ~1,000 (ref. [68]). Multi-copy integrations also require that sequences from all sites be recovered to obtain full lineage information, which can be challenging with sparse scRNA-seq data[69]. The scGESTALT barcode array consists of nine distinct CRISPR sites in tandem, each of which can be targeted by a different sgRNA. It also enables multi-timepoint editing by targeting subsets of the CRISPR sites during early and late embryogenesis, using a combination of injection and transgenic Cas9/sgRNA expression. Such a modular system enables longer periods of lineage recording and increases barcode diversity such that >10,000 barcodes are observed[72]. Furthermore, the use of single-copy barcode integrations facilitates unequivocal profiling of the complete barcode array extracted from each cell, without the need for inference of dropout events (i.e., partial loss of lineage information due to distribution of lineage recording sites across multiple loci).

Limitations of scGESTALT

Several considerations have to be made when choosing scGESTALT for lineage recording. First, the edited barcode is recovered from fewer than 30% of profiled single cells, possibly due to low

expression level of the barcode, inefficient capture of the barcode transcript within droplets, or amplification bottlenecks during sequencing library preparation. If using this method for profiling lineages of rare cell types, an enrichment step (e.g., FACS) for populations of interest may be required to obtain sufficient lineage barcodes. Future optimizations using stronger promoter sequences to drive barcode expression or improved library preparation could address these issues. Alternatively, the barcode could be extracted from genomic DNA rather than the transcriptome[68]. Second, owing to the tandem arrangement of CRISPR sites on the barcode array, some lineage records can be erased by large deletions that span multiple sites. Third, barcode diversity is limited to ~10,000 unique labels, owing to relatively quick saturation of Cas9 editing activities. To follow fate specification of progenitors over longer timeframes (e.g., days or weeks) and generate lineage trees with more branches, the timing and duration of editing will need to be further modulated, and the diversity of barcodes will need to be increased to millions of unique labels. Finally, tissue dissociation results in loss of precise spatial information of profiled cells; thus, spatial distributions of clones and distantly related cells cannot be investigated.

Experimental design

The scGESTALT protocol is divided into four experimental sections: (i) generation of two transgenic zebrafish lines: one for barcode expression and the other for inducible Cas9 and ubiquitous sgRNA expression; (ii) performance of multi-timepoint barcode editing using a combination of injection and induction of barcoding reagents for early and late labeling of cells, respectively; (iii) transcriptome profiling using inDrops; and (iv) concurrent extraction of lineage barcodes from captured single cells. Transcriptome and scGESTALT libraries are sequenced separately, and gene expression and lineage information for each cell are matched using the inDrops cell identifier index.

Generating transgenics for barcoding

Two transgenic lines expressing different components of the scGESTALT barcoding strategy are established using the Tol2 transposon system (Fig. 2). In this system, a DNA sequence of interest, flanked by minimal *cis*-sequences from the left and right ends of the medaka Tol2 element, is randomly integrated into the genome in the presence of Tol2 transposase by a cut-and-paste mechanism[76]. Thus, a transgenic construct can be cloned in between the Tol2 ends, and the resulting plasmid is injected along with mRNA encoding Tol2 transposase into zebrafish embryos. Translated Tol2 protein will mediate excision and stable integration of the transgenic construct into the genome of embryonic cells during early development, including some cells that give rise to germ cells. The injected fish are outcrossed and screened for successful germline transmission of the Tol2 transgenic construct into the F_1 generation to establish transgenic animals. The lineage barcode array construct comprises a heat shock–inducible promoter driving ubiquitous DsRed transgene expression with the scGESTALT CRISPR array inserted into the 3′ untranslated region (UTR) upstream of the SV40 polyadenylation signal. In addition, the construct contains a heart-specific GFP reporter as a transgenesis marker (cmlc2:GFP). Upon heat shock, the wild-type or edited barcode is expressed as part of the DsRed mRNA and can be extracted from the cellular transcriptome. The barcode transgenic line is screened by qPCR to determine transgene copy number; single-copy integrants are identified and outcrossed to expand and maintain the line. A separate transgenic line is established using a Tol2 construct comprising a heat shock–inducible promoter driving Cas9-t2A-GFP and five independent U6 promoters driving five distinct sgRNAs targeting the barcode array. This line does not need to be surveyed for single-copy integration. The two lines are available upon request from the Schier or Gagnon labs (Reagents). Depending on the user's goal, it is possible to make changes to both of the above constructs and generate new transgenics. For example, cell- or tissue-specific promoters can be used to drive Cas9 or barcode expression. In addition, modifications can be made to integrate the constructs into other organisms or cell lines.

Early and late barcoding

To enable recording of lineages during early and late development, the scGESTALT barcode transgenic is crossed to the Cas9/sgRNA transgenic. Single-cell embryos are injected with Cas9 protein and a pool of sgRNAs that target the first four sites of the CRISPR array. This strategy initiates an 'early' round of Cas9 editing activity that introduces mutations at target sites 1–4 of the barcode. Upon zygotic genome activation, sgRNAs targeting sites 5–9 of the array are expressed from U6 promoters. Embryos are then heat-shocked at a time of choosing (e.g., 30 hours post fertilization (hpf)) to induce
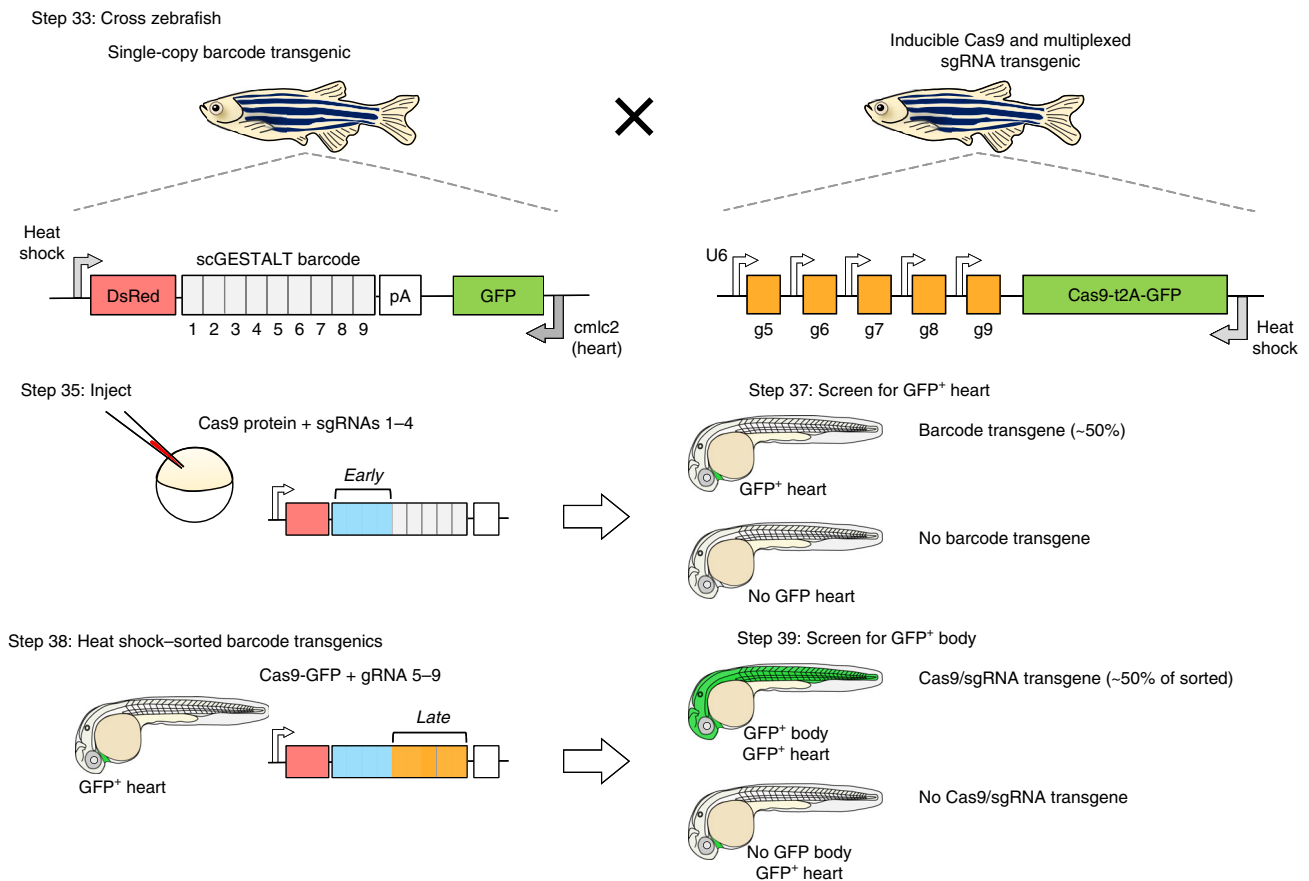
Fig. 2 | Strategy for a CRISPR–Cas9 system that enables early and late barcode editing. Zebrafish with single-copy heat shock promoter–driven scGESTALT barcodes (to promote ubiquitous barcode expression at the stage of interest) are crossed to zebrafish that express heat shock–inducible Cas9 and U6-driven sgRNAs 5–9. The barcode is cloned downstream of the DsRed coding sequence and upstream of the SV40 polyadenylation sequence (pA). The resulting embryos are injected with Cas9 protein and sgRNAs 1–4 at the one-cell stage (blue bars; early editing). The embryos are screened for GFP-positive heart transgenics (cmlc2 promoter drives heart-specific GFP expression) at 30 h post fertilization (hpf) to identify embryos containing the barcode transgene, and sorted embryos are heat-shocked to induce transgenic Cas9 for a second round of editing (orange bars; late editing). The embryos are screened again for ubiquitous GFP expression (Cas9 is linked to GFP with a t2A self-cleaving peptide), which indicates successful Cas9 transgene induction. Double-transgenic embryos are grown for downstream profiling, and heat-shocked at the time of interest (e.g., juvenile stage, 23–25 dpf) to induce expression of the edited barcode array before scRNA-seq analysis. Protocol steps for each stage are indicated. Adapted with permission from Raj et al.[72], Springer Nature.

ubiquitous expression of transgenic Cas9 and initiate barcode editing at target sites 5–9 of the array. The sequential activation of this 'early and late' editing strategy enables longer lineage recording and higher barcode diversity, and encodes the relative order of mutation patterns into the two halves of the array, thereby facilitating lineage tree reconstruction[72]. Edited double-transgenic embryos are selected by sequentially screening for strong heart-specific GFP expression to identify the barcode transgene and ubiquitous GFP expression to identify the Cas9/sgRNA transgene, and are grown until the user is ready to perform scRNA-seq (Fig. 2). Owing to the modular nature of the barcoding strategy, if the user is interested in tracing early lineage segregations, it is possible to barcode cells during early embryogenesis by injecting only Cas9 protein or mRNA together with all nine sgRNAs targeting the CRISPR array. The sequences for the nine sgRNA target sites are provided in Supplementary Table 1. Generally, the spectrum of barcode edits is dominated by deletions, which can be categorized into intrasite (a small deletion restricted to one target site) and intersite (larger deletions that span two or more sites). Insertions are observed at a low frequency. The combination of small indels and larger deletions in early and/or late editing generates a high overall diversity of lineage barcodes (<0.5% of lineage barcodes were found to overlap between $n = 8$ different embryos[72]). We direct the reader to ref. [72] for further information about the nature and frequency of barcode repair products.

Single-cell transcriptome profiling

scRNA-seq is performed to simultaneously extract edited lineage barcodes and cellular transcriptomes. For droplet-based scRNA-seq, samples are first dissociated into single-cell suspensions prior to loading on the device. We processed juvenile zebrafish brains to provide a snapshot of the heterogeneity of cell types and lineages in the vertebrate brain[72], and here provide details for brain dissociation and scRNA-seq using inDrops. However, the user can perform similar analyses using any tissue or developmental stage of interest for which a single-cell dissociation protocol that yields high cell viability is available. The inDrops protocol has been described in detail previously[73]. Briefly, it involves (i) obtaining a compatible microfluidic chip, (ii) synthesizing indexed hydrogel beads, (iii) encapsulating and indexing single cells in nanoliter-size droplets, and (iv) building transcriptome libraries for next-generation sequencing. As single cells are lysed during inDrops, polyadenylated cellular mRNA and edited scGESTALT lineage barcode mRNA are hybridized to oligodT primers on hydrogel beads (Fig. 3a). Following reverse transcription of hybridized mRNA, the cDNA is indexed with an inDrops cell identifier, subjected to second-strand synthesis and linearly amplified by in vitro transcription (IVT) (Fig. 3b). A fraction of the resulting amplified RNA is chemically fragmented for standard transcriptome library preparation and is converted to a DNA library by another round of reverse transcription followed by addition of Illumina-compatible sequencing adapters by PCR. The remaining unfragmented RNA is used for building scGESTALT libraries. scGESTALT can also be adapted for scRNA-seq with other droplet-based (e.g., Drop-seq[9] and 10X Genomics[10]) or plate/array-based methods. However, the steps for amplification and library construction of lineage barcodes as detailed below will have to be adjusted by the user on the basis of empirical data.

Targeted amplification of lineage barcodes

After IVT of indexed cDNA, a fraction of the full-length amplified RNA is reverse-transcribed using random hexamers. scGESTALT lineage barcodes are amplified from the resulting cDNA, using a nested PCR strategy (Fig. 3c). First, a larger fragment is enriched, using primers that flank the DsRed transgene and a universal inDrops adapter sequence located downstream of the cell identifier and unique molecular identifier (UMI, for counting the number of unique transcripts of a gene) sequences. Following purification, the PCR product is used to amplify a shorter lineage barcode fragment, using primers that flank the barcode array and a universal inDrops adapter sequence. Finally, Illumina-compatible DNA libraries are generated by following the same strategy for transcriptome libraries. If mixing multiple libraries in a sequencing run, it is critical that the same multiplexing primers be used for both transcriptome and lineage barcode libraries to ensure that these can later be matched to each other. In this protocol, we provide details for preparing 'V3' inDrops transcriptome and scGESTALT libraries, which are compatible with standard Illumina sequencing primers. If using 'V2' inDrops libraries, they will need to be sequenced with custom read and index primers as described previously[73].

Experimental controls

A negative control for editing involves no injection of Cas9 protein and sgRNAs 1–4, followed by no heat shock (i.e., no Cas9 transgene induction). In this case, no substantial editing of the barcode array at sites 1–9 should be observed by PCR amplification from genomic DNA or using next-generation sequencing from genomic DNA and/or scRNA-seq libraries. Similarly, injection of Cas9 protein alone (no sgRNAs 1–4), followed by no heat shock (no Cas9 transgene induction), should also result in little to no editing at sites 1–9. In some cases, low levels of editing may be observed at sites 5–9 due to cross-activity between injected Cas9 protein and U6-expressed sgRNAs 5–9. When no Cas9 protein and sgRNAs 1–4 are injected, but Cas9 is induced at 30 hpf by heat shock, editing is expected to be largely confined to sites 5–9. By contrast, performing two-timepoint editing by Cas9 protein and sgRNA 1–4 injection followed by heat-shock Cas9 induction, should result in edits observed throughout sites 1–9 of the barcode array.

Data analysis

The data analysis consists of three major parts: (i) analysis of the transcriptome data for cell-type identification, (ii) analysis of the scGESTALT lineage barcodes for identification of editing events, and (iii) generation of a single-cell-resolved lineage tree with associated cell types. The inDrops.py pipeline (see 'Software' in the 'Equipment' section below) will generate a gene expression matrix that can be used as input for identification of cell types and gene markers using an analysis tool designed for processing sparse scRNA-seq data (e.g., Seurat[18,77]). The inDrops.py pipeline can also be used to obtain sequence reads for scGESTALT lineage barcodes that are sorted by the inDrops cell identifier index (i.e., to obtain lineage reads for single cells). Next, the sorted scGESTALT lineage barcode reads
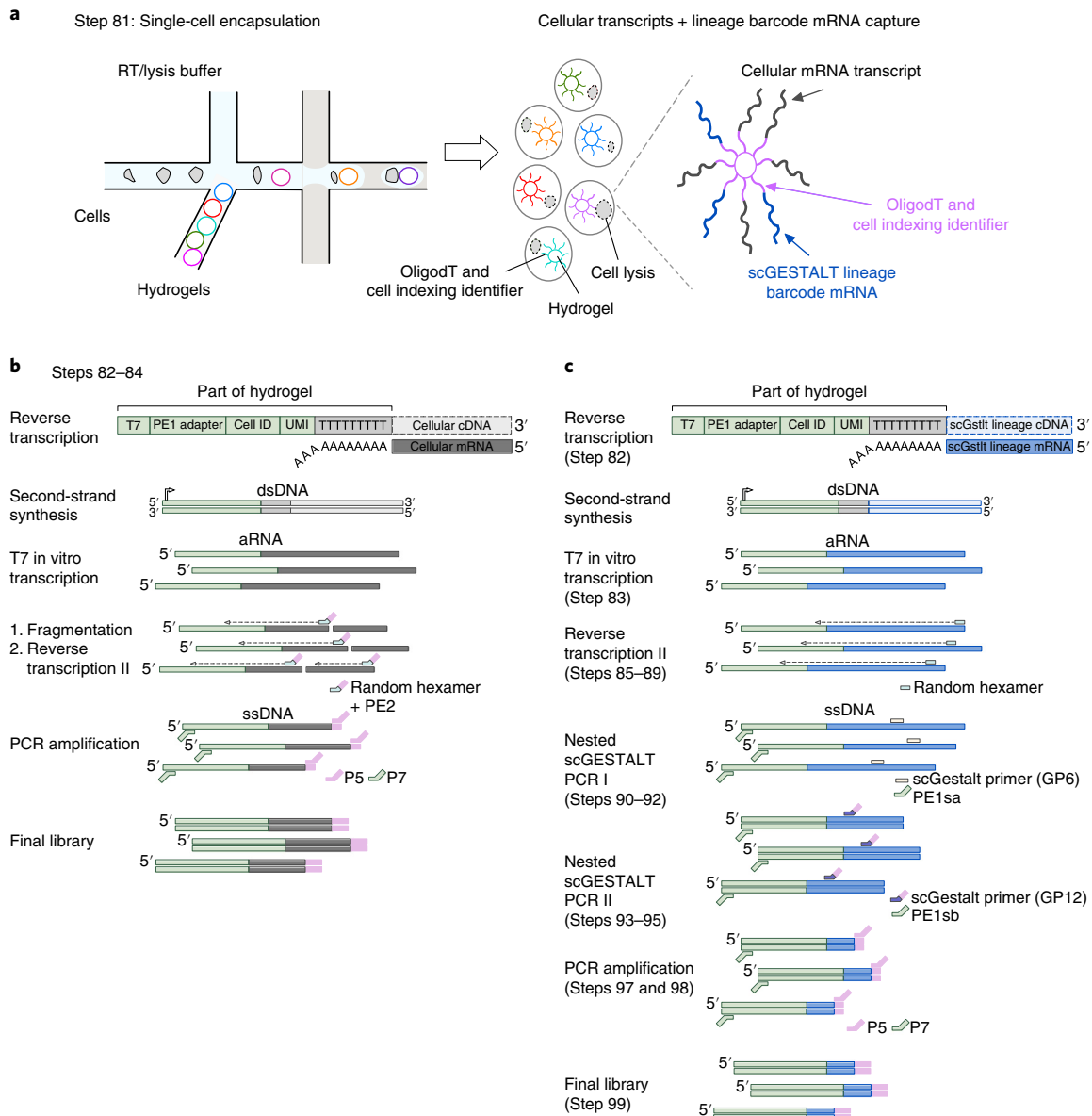
Fig. 3 | Transcriptome and scGESTALT library preparation overview. a, Single cells are encapsulated and indexed in droplets using the inDrops platform (left, Step 81). Upon lysis (center, cells shown with dotted lines), polyadenylated cellular transcripts and scGESTALT lineage barcodes hybridize to oligodT primers on hydrogels. Hydrogels are color-coded to indicate distinct indexing identifiers. Adapted with permission from Raj et al.[72], Springer Nature. **b**, Overview of steps for transcriptome library preparation from single cells (Steps 82–84). Hydrogels are coated with oligonucleotides containing a T7 promoter, an adapter sequence (PE1), a cell-indexing identifier (Cell ID), a unique molecular identifier (UMI), and an oligodT sequence. Polyadenylated cellular mRNA hybridizes to the oligodT sequence and is reverse-transcribed into a cellular cDNA sequence. Second-strand synthesis is carried out to generate double-stranded DNA (dsDNA) that is used for T7 IVT and generates linearly amplified RNA (aRNA). The aRNA is chemically fragmented and then subjected to a second round of reverse transcription using an oligonucleotide that contains a random hexamer and an adapter sequence (PE2). The resulting single-stranded DNA (ssDNA) is PCR-amplified with Illumina P5 and P7 adapter sequences containing an overlap with PE2 and PE1 sequences, and a limited number of cycles are used to generate the final sequencing-ready transcriptome library. ssDNA fragments that do not contain the PE1 adapter sequence will not be amplified. **c**, Overview of scGESTALT lineage barcode library preparation steps from single cells. Polyadenylated lineage barcode mRNAs, consisting of DsRed and edited scGESTALT CRISPR array (Fig. 2, Step 33), are expressed from a transgene upon heat shock of the animal before brain dissection and dissociation. The lineage barcode mRNA hybridizes to the oligodT sequence on the hydrogel and is reverse-transcribed into lineage barcode cDNA. Second-strand synthesis and T7 IVT are carried out in a manner similar to that for transcriptome library preparation (Step 82). However, the aRNA is not fragmented, in order to preserve the full lineage barcode sequence (Step 83, part of the reverse-transcribed product is not fragmented and is stored at −80 °C). Instead, reverse transcription is carried out using random hexamers to generate full-length ssDNA (Steps 85–89). The edited scGESTALT CRISPR array sequence is selectively enriched from the full-length ssDNA, using a nested PCR approach. The first round amplifies a longer piece of dsDNA, using a primer that hybridizes in the DsRed sequence (GP6) and one that overlaps the PE1 adapter sequence (PE1sa) (Steps 90–92). The resulting dsDNA is used in a second round to target the edited lineage barcode sequence, using a primer that binds upstream proximal to the start of the CRISPR lineage array sequence (GP12) and a primer that overlaps the adapter sequence (PE1sb) (Steps 93–95). The resulting dsDNA is PCR-amplified (Steps 97 and 98) in a manner similar to that used for transcriptome library preparation to generate the final sequencing-ready scGESTALT lineage barcode library (Step 99).

are further formatted, using the scGestaltPrepFunc.R script (Supplementary Software 1) and processed with the scGESTALT analysis pipeline (see 'Software' in the 'Equipment' section; also provided as Supplementary Software 2). This will generate a table containing the lineage barcode sequence for each cell as well as its inDrops cell identifier index sequence. Next, for each cell from which a lineage barcode was successfully recovered and processed, the matching transcriptome and hence cell identity is assigned using the Transcriptome-scGESTALTMatchPipe.R and MatchPipeFunc.R scripts (Supplementary Software 3 and 4). At this stage, a table will be generated (GestMaster.txt; see TestData in Supplementary Software 2) containing the scGESTALT lineage barcode, cell cluster identity (obtained using Seurat for cell-type classification), and the inDrops cell identifier. This can be used as input for drawing the overall multi-timepoint edited lineage tree using the scGESTALT analysis pipeline.

Materials

Biological materials

Zebrafish (*Danio rerio*) **!CAUTION** All research involving animals must be approved by the local institutional regulatory board and conducted in accordance with national and local regulations and guidelines. This procedure was approved by the Harvard University/Faculty of Arts & Sciences Standing (HU/FAS) Committee on the Use of Animals in Research & Teaching under protocol no. 25–08. The HU/FAS animal care and use program maintains full AAALAC accreditation, has an assurance on file with the Office of Laboratory Animal Welfare (OLAW; A3593-01), and is currently registered with the US Department of Agriculture.

- Wild-type strains (e.g., TL/AB, TL, or AB; available from ZIRC, http://www.zebrafish.org)
- Tg(hsp70:DsRed-barcodev7; myl7:eGFP) strain (allele no. a167; scGESTALT barcode transgenic; available upon request from the Schier or Gagnon laboratory)
- Tg(hsp70:Cas9-t2a-GFP; 5xU6:sgRNA) strain (allele no. a168; inducible Cas9 and U6-driven sgRNA transgenic; available upon request from the Schier or Gagnon laboratory)

Reagents

- Nuclease-free water (Life Technologies, cat. no. M5310-1L)
- NotI-HF (NEB; cat. no. R3189)
- E.Z.N.A. Cycle Pure Kit (Omega, cat. no. D6492-02)
- E.Z.N.A. Plasmid DNA Mini Kit I (Omega, cat. no. D6942-02)
- E.Z.N.A. Total RNA Kit I (Omega, cat. no. R6834-02)
- RNA Clean & Concentrator-25 (Zymo Research, cat. no. R1018)
- mMESSAGE mMACHINE SP6 (Thermo Fisher Scientific, cat. no. AM1340)
- Sodium hydroxide (NaOH; 1 M; Honeywell Research Chemicals, cat. no. 319511-1L)
- Tris–hydrochloric acid (HCl, pH 7.5; 1 M; VWR, cat. no. 75800-958)
- Tris–HCl (pH 7.0, 1 M; Thermo Fisher Scientific, cat. no. AM9851)
- Tris–HCl (pH 8.0, 1 M; Thermo Fisher Scientific, cat. no. 15568025)
- EnGen sgRNA Synthesis Kit, *S. pyogenes* (NEB, cat. no. E3322) ▲**CRITICAL** sgRNA oligonucleotide sequences (sgRNAs 1–4; see DNA oligonucleotide sequences below) are designed for use with this kit.
- EnGen Cas9 NLS, *S. pyogenes* (NEB, cat. no. M0646)
- iTaq Universal SYBR Green Supermix (Bio-Rad, cat. no. 1725120)
- Phusion High-Fidelity DNA Polymerase (NEB, cat. no. M0530) ▲**CRITICAL** Amplification of the scGESTALT library has been optimized using this enzyme.
- Q5 High-Fidelity DNA Polymerase (NEB, cat. no. M0491) ▲**CRITICAL** Amplification of the scGESTALT library has been optimized using this enzyme.
- KAPA 2× HiFi HotStart Ready Mix (KAPA Biosystems, cat. no. KK2601)
- dNTP mix (10 mM each; NEB, cat. no. N0447)
- 100-bp DNA ladder (NEB, cat. no. N3231)
- Papain Dissociation System (Worthington, cat. no. LK003150) ▲**CRITICAL** Brain dissociation has been optimized using this kit.
- Neurobasal Medium (Thermo Fisher Scientific, cat. no. 21103049) ▲**CRITICAL** Brain dissection and dissociation have been optimized using this medium.
- B27 supplement, serum free (50×; Thermo Fisher Scientific, cat. no. 17504044) ▲**CRITICAL** Brain dissection has been optimized using this supplement for high cell viability.
- DPBS, no calcium, no magnesium (Thermo Fisher Scientific, cat. no. 14190144)
- DBPS, calcium, magnesium, glucose, and pyruvate (Thermo Fisher Scientific, cat. no. 14287080)

- Trypan blue (0.4% (wt/vol); Thermo Fisher Scientific, cat. no. T10282)
- OptiPrep Density Gradient Medium (Sigma-Aldrich, cat. no. D1556-250ML)
- Agencourt AMPure XP magnetic beads (Beckman Coulter, cat. no. A63881)
- Random hexamers (50 µM; Fisher Scientific, cat. no. N8080127)
- PrimeScript Reverse Transcriptase (Takara Clontech, cat. no. 2680A)
- RNaseOUT Recombinant Ribonuclease Inhibitor (Fisher Scientific, cat. no. 10777019)
- Phenol red (0.5% (wt/vol); Sigma-Aldrich, cat. no. P0290)
- Methylene blue (0.1% (wt/vol); Aqua Solutions, cat. no. 5761-500ML)
- Instant Ocean Sea Salt (Instant Ocean, cat. no. SS3-50)
- Agarose (e.g., National Diagnostics, cat. no. EC-202)
- Sylgard 184 (Electron Microscopy Sciences, cat. no. 24236-10)
- Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q33230)
- Tricaine (Sigma-Aldrich, cat. no. E10521-50G) **! CAUTION** This chemical can cause skin irritation. Use gloves when handling.
- MiSeq Reagent Kit v2, 300-cycle kit (Illumina, cat. no. MS-102-2002)
- NextSeq 500/550 High Output v2 Kit, 75-cycle kit (Illumina, cat. no. FC-404-2005)
- PhiX Control v3 (Illumina, cat. no. FC-110-3001)
- Sodium bicarbonate buffer (pH 8.5, 1 M; Westnet, cat. no. BBB-2296)
- Trizma TRIS base (Sigma-Aldrich, cat. no. 77-86-1)
- EDTA (pH 8.0, 0.5 M; Sigma-Aldrich, cat. no. 324506-100ML)
- Agilent High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)

Plasmids

- pCS2-zT2TP (available from the K. Kawakami laboratory, National Institute of Genetics, Mishima, Japan[78])
- pTol2-hspDRv7_scGstlt (Addgene, plasmid ID 108870)
- pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA (Addgene, plasmid ID 108871)

DNA oligonucleotide sequences (purified by standard desalting; can be custom ordered from IDT)

Name	Sequence	Purpose	Step
qPCRctrl F	5′-TCAGTCAACCATTCAGTGGCCCAT-3′	PCR-amplify ultraconserved genomic region	23
qPCRctrl R	5′-CAGGAAAGGGAATGCAGGGTTTGT-3′	PCR-amplify ultraconserved genomic region	23
qPCRdsRed F	5′-GAGCGCGTGATGAACTTCGAGG-3′	PCR-amplify dsRed transgenic region	23
qPCRdsRed R	5′-CAGCCCATAGTCTTCTTCTGCATTACG-3′	PCR-amplify dsRed transgenic region	23
sgRNA1	5′TTCTAATACGACTCACTATAGACAGCAGTATCATCGACTAGTTTTAGAGCTAGA-3′	Synthesize sgRNA1 for CRISPR–Cas9 lineage barcoding by injection	27
sgRNA2	5′-TTCTAATACGACTCACTATAGAGAGCGCGCTCGTCGACTAGTTTTAGAGCTAGA-3′	Synthesize sgRNA2 for CRISPR–Cas9 lineage barcoding by injection	27
sgRNA3	5′-TTCTAATACGACTCACTATAGTCAGCAGTACTACTGACGAGTTTTAGAGCTAGA-3′	Synthesize sgRNA3 for CRISPR–Cas9 lineage barcoding by injection	27
sgRNA4	5′-TTCTAATACGACTCACTATAGACAGCAGTGTGTGAGTCTAGTTTTAGAGCTAGA-3′	Synthesize sgRNA4 for CRISPR–Cas9 lineage barcoding by injection	27
scGESTALT F	5′-TCGAGCTCAAGCTTCGG-3′	PCR-amplify scGESTALT fragment for assessing editing efficiency	43
scGESTALT R	5′-CTGCCATTTGTCTCGAGGTC-3′	PCR-amplify scGESTALT fragment for assessing editing efficiency	43
inDrops_GP6	5′-GAGGACTACACCATCGTGGAG-3′	PCR-amplify long scGESTALT fragment from inDrops library	90
inDrops_PE1Sa	5′-CTCTTTCCCTACACGACGCTGGGTGTCGGGTGCAG-3′	PCR-amplify long scGESTALT fragment from inDrops library	90
inDrops_GP12	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNNNNNTCGAGCTCAAGCTTCGGAC-3′, where N stands for a random base	PCR-amplify short scGESTALT fragment from inDrops library	93

Table continued

(continued)

Name	Sequence	Purpose	Step
inDrops_PE1Sb	5'- CTCTTTCCCTACACGACGCT-3'	PCR-amplify short scGESTALT fragment from inDrops library	93
R1-PCRix	5'-AATGATACGGCGACCACCGAGATCTA CACXXXXXXTCGTCGGCAGCGTC-3', where XXXXXX is an index sequence for multiplexing (Supplementary Table 2)	PCR-amplify transcriptome or scGESTALT final sequencing library	97
R2-PCR	5'-CAAGCAGAAGACGGCATACGAGATGG GTGTCGGGTGCAG-3'	PCR-amplify transcriptome or scGESTALT final sequencing library	97

Equipment

- 35-μm Cell strainer (VWR, cat. no. 352235)
- 20-μm Cell strainer (Sysmex, cat. no. 04-004-2325)
- 1.5-ml Microcentrifuge tubes (USA Scientific, cat. no. 1415-2600)
- 1.5-ml Microcentrifuge tubes (USA Scientific, cat. no. 1415-2500)
- 0.2-ml PCR tubes (VWR, cat. no. 53509-304)
- 50-ml Tubes (VWR, cat. no. 352070)
- 15-ml Tubes (VWR, cat. no. 352096)
- 5-ml Sterile serological pipettes (VWR, cat. no. 89130-896)
- NanoDrop spectrophotometer (Thermo Scientific, cat. no. ND-2000)
- PCR and qPCR machines (Bio-Rad, model no. C-1000 and CFX96 Touch)
- DNA electrophoresis system (Bio-Rad, cat. no. 1640301)
- Benchtop centrifuge for 1.5-ml tubes (Eppendorf, cat. no. 54124)
- Refrigerated centrifuge for 1.5-ml tubes (Eppendorf, cat. no. 5424R)
- Minifuge (Benchmark, cat. no. C1008-C)
- Water bath (VWR, cat. no. 89501-460)
- Heat block for 1.5-ml tubes (VWR, cat. no. 10153-348)
- Vortex (Scientific Industries, cat. no. SI-0236)
- Magnetic rack for 0.2-ml PCR strips (Permagen, cat. no. MSR1224)
- 2100 Electrophoresis Bioanalyzer instrument (Agilent, cat. no. G2939AA)
- Laser-based micropipette puller (Sutter Instrument, cat. no. P-2000)
- Microinjection glass capillaries (World Precision Instruments, cat. no. TW100F-4)
- Fine forceps (Fine Science Tools, cat. no. 11254-20)
- Fine scissors (Fine Science Tools, cat. no. 14040-10)
- Microloader tips for glass capillaries (Eppendorf, cat. no. 930001007)
- Injection molds (Adaptive Science Tools, model no. TU-1)
- Pneumatic PicoPump (World Precision Instruments, model no. PV820)
- Piconozzle Kit v2 (World Precision Instruments, cat. no. 5430-ALL)
- Manipulator, magnetic stand, and iron plate (Narishige, model. no. M-152, GJ-1, IP)
- Treadlite II Foot Switch (Linemaster, model no. T-91-S)
- Micrometer for calibration (AmScope, cat. no. MR100)
- Dissection microscope with light source (Zeiss, model no. Stemi 2000)
- Fluorescence microscope with GFP filter (Zeiss)
- 10-cm Petri dishes (VWR, cat. no. 25384-342)
- Insect pins (Fine Science Tools, cat. no. 26002-15)
- Breeding tanks and 0.5-, 1-, and 2-L tanks (Techniplast, custom order)
- 95% O_2:5% CO_2 gas tank (Airgas, cat. no. X02OX95C2003102)
- Gas regulator (VWR, cat. no. 55850-392)
- Tygon tubing for bubbling gas (Component Supply, cat. no. E-3603)
- Hemocytometer (Bulldog Bio, cat. no. DHCN420)
- Qubit fluorometer (Thermo Fisher Scientific, cat. no. Q33226)
- Qubit assay tubes (Thermo Fisher Scientific, cat. no. Q32856)

Software
inDrops pipeline (inDrops.py: https://github.com/indrops/indrops)
scGESTALT pipeline (https://github.com/aaronmck/SC_GESTALT)
R (https://www.r-project.org)
RStudio (https://www.rstudio.com)
Seurat (https://satijalab.org/seurat/)

Reagent setup
Zebrafish
Adult fish are maintained at a density of eight to ten fish per 2-L tank with a 14-h light:10-h dark cycle at 28.5 °C. The wild-type TL/AB strain is bred in-house by crossing the wild-type TL strain with the wild-type AB strain and is used in all experiments described in this protocol. Animals used for breeding are typically 3–18 months old. **! CAUTION** This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25–08.

Phenol red
Dilute 0.5% (wt/vol) phenol red stock 1:5 with nuclease-free water. Use this diluted stock for injections, and store at room temperature (23–26 °C) for up to 1 year.

Blue water
Combine 20 L of water in a carboy with 20 ml of 0.1% (wt/vol) methylene blue and 5 g of Instant Ocean sea salt, and adjust the pH to 7.0 with sodium bicarbonate buffer. Store at room temperature for up to 1 month.

50 mM NaOH
Combine 19 ml of nuclease-free water with 1 ml of 1 M NaOH. Store at room temperature for up to 3 months.

1 M Tris base
Add 24.228 g of TRIS base to 200 ml of water. Store at room temperature for at least 1 year.

MESAB, 25×
Combine 489.5 ml of water with 2 g of Tricaine powder and 10.5 ml of 1 M Tris base, and adjust the pH to 7. Make 10-ml aliquots and store at −20 °C for up to 1 year.

B27 supplement
Upon receiving the product, thaw the solution completely at room temperature. Make 500-μl aliquots and store at −20 °C for up to 1 year.

RNA elution buffer
Combine 9.9 ml of nuclease-free water, 100 μl of 1 M Tris–HCl (pH 7.0), and 2 μl of 0.5 M EDTA (pH 8.0). Store 1-ml aliquots at −20 °C for up to 1 year.

DNA elution buffer
Combine 9.9 ml of nuclease-free water, 100 μl of 1 M Tris–HCl (pH 8.0), and 2 μl of 0.5 M EDTA (pH 8.0). Store 1-ml aliquots at −20 °C for up to 1 year.

R1-PCRix/R2-PCR primer mix
For each variant of R1-PCRix, combine 25 μl of 10 μM R1-PCRix primer and 25 μl of 10 μM R2-PCR primer in DNA elution buffer. Store primer mixes at −20 °C for up to 1 year. See Supplementary Table 1 for a list of 24 R1-PCRix primer sequences.

Equipment setup
Preparation of injection needles
Pull glass capillaries for injection as described previously[79].

Embryo microinjection dish

Boil 3 g of agarose in 100 ml of blue water in a microwave until the agarose dissolves. Let it cool slightly on the bench, and pour 25–30 ml per 10-cm Petri dish. Gently layer an injection mold (pre-wetted with water to avoid introducing bubbles) on top of the agarose. Let the agarose solidify and remove the injection mold. The microinjection dish can be stored at 4 °C for 2 weeks.

Microinjection setup

Configure and connect the Pneumatic PicoPump, piconozzle, manipulator, magnetic stand, iron plate, and foot switch according to the manufacturer's instructions. Connect the Pneumatic Pico-Pump to a nitrogen tank according to the manufacturer's instructions.

Preparation of 95% O$_2$:5% CO$_2$ gas tank

Fit the gas tank with the gas regulator and Tygon tubing according to the manufacturer's instructions.

Preparation of Sylgard dissection dishes

Mix 45 ml of Sylgard elastomer base with 4.5 ml of Sylgard elastomer curing agent (10:1 ratio; both reagents provided in Sylgard 184 kit) in a 50-ml tube. Mix thoroughly by inverting several times and transferring back and forth between two 50-ml tubes. Pour ~10 ml of the mix into each 10-cm Petri dish. Gently tap the dish to force air bubbles to rise to the top. Allow to cure and set overnight. The dishes can be stored at room temperature for at least 1 year.

inDrops microfluidics platform setup

Detailed instructions for setting up the inDrops platform have been provided previously[73]. Note that the system should be primed and ready for cell barcoding before starting single-cell preparation (Step 48). Alternatively, if working in a pair, one person should get the device ready while the other is preparing cells for encapsulation.

Procedure

Tol2 transposase mRNA synthesis ● Timing 1 d

1 Linearize 5 μg of pCS2-zT2TP plasmid using NotI-HF enzyme according to the manufacturer's instructions.
2 Purify the linearized plasmid with a mini quick-spin DNA column (from the E.Z.N.A. Cycle Pure Kit) according to the manufacturer's protocol and elute in 50 μl of nuclease-free water.
3 In vitro–transcribe Tol2 mRNA using the mMESSAGE mMACHINE SP6 Kit. Assemble the reaction below (20 μl in total) and follow the manufacturer's protocol.

Component	Amount (μl)	Final concentration
2× NTP/CAP	10	1×
10× buffer	2	1×
Linear, purified pCS2-zT2TP	6	~20–30 ng/μl
Enzyme mix	2	
Total	20	

4 Purify Tol2 mRNA using a mini quick-spin RNA column (from the E.Z.N.A. Total RNA Kit I) according to the manufacturer's protocol and elute in 50 μl of nuclease-free water. Measure the mRNA concentration using a NanoDrop spectrophotometer. Successful IVT reactions will yield mRNA at concentrations of 300–500 ng/μl. Dilute Tol2 mRNA to 100 ng/μl, using nuclease-free water.
   ■ PAUSE POINT Store the mRNA in 3-μl aliquots at −80 °C for up to 1 year.
5 Purify DNA from minipreps of pTol2-hspDRv7_scGstlt and pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA plasmids using a plasmid purification kit (from the E.Z.N.A. Plasmid DNA Mini Kit I). To remove residual RNase A, purify ~2 μg of miniprep DNA with a mini quick-spin DNA column (e.g., E.Z.N.A. Cycle Pure Kit) according to the manufacturer's protocol and elute in 30 μl of nuclease-free water. Measure DNA concentrations using the NanoDrop spectrophotometer. Dilute each of the plasmids to 25 ng/μl, using nuclease-free water.
   ■ PAUSE POINT Store diluted plasmids at −20 °C for at least 1 year.

Zebrafish transgene microinjection ● Timing 3 h

6   The day before performing microinjections, set up six to ten overnight tanks of wild-type zebrafish for mating as described previously[78].

   ! CAUTION All vertebrate animal work must be performed in accordance with relevant guidelines and regulations.

7   On the morning of microinjections, assemble reagent mixes (3 µl in total) as follows for each plasmid. Vortex to mix, centrifuge briefly (2,000g, room temperature, 2 s), and hold on ice.

Component	Amount (µl)	Final concentration
pTol2-hspDRv7_scGstlt OR pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA plasmid (25 ng/µl)	1.5	12.5 ng/µl
Tol2 mRNA (100 ng/µl)	1.0	33.3 ng/µl
Phenol red (0.1% (wt/vol))	0.5	0.017%
Total	3	

8   Stagger the mating of the fish to enable several rounds of injections to be performed at the one-cell stage (~20–30 min after fertilization). For example, mating can be initiated in two tanks at once, followed by ~35-min intervals between each subsequent set of mating pairs.

9   While fish are mating, prepare an injection needle by breaking the tip of a pulled glass capillary with fine forceps. Load 1.5 µl of the pTol2-hspDRv7_scGstlt injection mix (Step 7) into the needle with a microloader tip and attach it to the piconozzle and manipulator. Set the pressure to 20 p.s.i. and calibrate the needle with a micrometer as described previously[79].

10 Collect fertilized eggs into Petri dishes with blue water. Transfer ~100 one-cell-stage embryos to the agarose ramp in the injection dish (distribute the embryos across four rows of the ramp), and add enough blue water to just cover the embryos. Pierce the chorion with the calibrated needle, move the needle into the cytoplasm, and inject 1 nl of the reagent mix. Repeat this step until all embryos are injected. Transfer the injected embryos with blue water to a new Petri dish and incubate overnight at 28.5 °C. Keep some uninjected control embryos to assess the health of the clutch.

11 Repeat Steps 8–10 for the pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA injection mix (Step 7).

12 At the end of the day, remove any unfertilized (embryos will appear as if they are still at the single-cell stage rather than having undergone multiple rounds of cell division) or damaged (embryos are squashed or have become deformed from the injection procedure) embryos from the injected batches.

Sorting and rearing transgenic fish ● Timing 1 d of sorting, 2–3 months of rearing

13 Heat-shock fish injected with pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA at 24 hpf. Transfer 20–30 healthy-looking embryos to a 1.5-ml microcentrifuge tube and add 500 µl of blue water. Use as many tubes as required, depending on how many embryos survived injection (20–30 embryos per tube). Incubate the tubes in a heat block set to 37 °C for 30 min. Transfer heat-shocked embryos back to Petri dishes with blue water and incubate at 28.5 °C until ready to sort (Step 14). Alternatively, one can perform the heat shock in 50-ml tubes (with 5 ml of blue water) in a water bath at 37 °C for 30 min.

14 GFP expression starts to become apparent ~2 h after heat shock and becomes bright enough for easy screening and sorting ~4 h after heat shock. Place the Petri dish from Step 13 on the stage of a fluorescence microscope fitted with a GFP filter. Swirl the Petri dish to bring the embryos to the center and turn on the fluorescent light. Look through the eyepiece to identify fish that exhibit mosaic GFP fluorescence throughout the body. Using a plastic or glass transfer pipette, transfer the brightest GFP-positive fish to a new Petri dish containing 20 ml of blue water. Incubate at 28.5 °C until 5 d post fertilization (dpf) and then transfer to 2-L tanks (eight to ten larvae per tank) in a zebrafish facility to raise to adulthood.

   ▲ CRITICAL STEP Seeding tanks at a low density will enable fish to grow and reach sexual maturity faster.

15 Screen and sort fish injected with pTol2-hspDRv7_scGstlt at 30 hpf. Check for bright GFP fluorescence in the heart, using a fluorescence microscope. Using a plastic or glass transfer pipette, transfer GFP-positive fish to a new Petri dish containing 20 ml of blue water. Incubate at 28.5 °C until 5 dpf and then transfer to tanks (eight to ten larvae per tank) in a zebrafish facility to raise to adulthood.

Screening for germline transmission ● Timing 2 d, then 2–3 months to rear

16 Identify adult founder fish with germline integration by outcrossing individual injected F_0 animals to wild-type fish of the opposite sex. Collect embryos the next morning in blue water and incubate overnight at 28.5 °C. Keep potential founders separate until screening of clutches is complete.

17 Screen clutches from pTol2-hsp70l:Cas9-t2A-GFP, 5xU6:sgRNA F_0 fish by heat-shocking embryos at 24 hpf as described in Steps 13 and 14. The screened F_1 embryos are henceforth referred to as the 'hsp:Cas9, U6:sgRNA' transgenic line.

18 Screen clutches from pTol2-hspDRv7_scGstlt F_0 fish by GFP heart expression at 30 hpf as described in Step 15. The screened F_1 embryos are henceforth referred to as the 'scGESTALT' transgenic line.

19 Incubate sorted F_1 embryos at 28.5 °C until 5 dpf and then transfer to 2-L tanks (eight to ten larvae per tank) in a zebrafish facility to raise to adulthood.
   ■ PAUSE POINT Maintain stable lines by outcrossing. Adults can be maintained for up to 2 years.

Determination of copy number of scGESTALT barcode ● Timing 1 d

20 Fin-clip adult F_1 pTol2-hspDRv7_scGstlt fish. Anesthetize fish in batches of three to four with 100 ml of 1× MESAB and snip a small section of the tail fin, using fine scissors. Transfer the tail fin to a PCR strip tube and hold on ice. Transfer each fin-clipped fish to an individual 0.5-L tank to recover and number them uniquely. Repeat this step until all F_1 fish have been fin-clipped.

21 Prepare genomic DNA using a modified HotSHOT method[80]. Add 50 μl of 50 mM NaOH to each fin clip and incubate at 95 °C for 20 min in a thermocycler.

22 Add 5 μl of 1 M Tris–HCl, pH 7.5, and vortex to mix. Centrifuge briefly (2,000g, room temperature, 2 s).
   ■ PAUSE POINT Genomic DNA can be stored at 4 °C for at least 3 months.

23 Set up qPCR reactions in triplicate. For each genomic DNA sample, set up qPCR reactions with two sets of primers, one for DsRed (to determine copy number of scGESTALT barcode) and another for an ultraconserved control region that is known to be present in two copies[80]. Thus, each sample will have a total of six reactions. In addition, set up a qPCR with reference standards using genomic DNA from fish with no DsRed integration and from fish with known single- and/or double-copy DsRed integrations.

Component	Amount (μl)	Final concentration
Bio-Rad iTaq mix	10	1×
Primer (F+R mix) (10 μM)	0.8	0.4 μM
Nuclease-free water	7.2	
Genomic DNA (1:5 dilution)	2	
Total	20	

Primer pairs: (1) qPCRctrl F+R; (2) qPCRdsRed F+R (see DNA oligonucleotide sequences in Reagents). F, forward; R, reverse.

24 Run qPCR as follows:

Cycle number	Denature	Anneal/extend	Final
1	95 °C, 3 min		
2–36	95 °C, 10 s	55 °C, 30 s (with fluorescence readout)	
37			8 °C, hold

25 Calculate copy number with the equations below:

$$\Delta(\mathrm{Ct}) = (\mathrm{Ct\ for\ dsRed\ PCR}) - (\mathrm{Ct\ for\ cons\_ctrl\ PCR}).$$

– Perform calculations for all unknown samples and positive and negative reference standards.

$$\Delta\Delta\mathrm{Ct} = \Delta(\mathrm{Ct, sample}) - \Delta(\mathrm{Ct, reference\ standard}).$$

– Perform calculations for all unknown samples and positive and negative reference standards. $\Delta(\mathrm{Ct},$ sample) is the $\Delta(\mathrm{Ct})$ for each sample and positive and negative reference standards calculated using

the first equation. Use the $\Delta(Ct)$ of the known single-copy reference standard as $\Delta(Ct, \text{reference standard})$. Thus, $\Delta\Delta Ct$ for the single-copy reference standard will be 0.

$$\text{Copy number in the sample} = 2^{--\Delta\Delta Ct}.$$

- Perform calculations for all unknown samples and positive and negative reference standards. The value for the known single-copy reference will be 1. The values for the known negative and double-copy reference standards should be ~0 and ~2, respectively.

See Anticipated results and the Supplementary Data for examples of single-copy calculations.

26 Keep single-copy scGESTALT F_1 animals for all downstream experiments. Identify females and maintain them in separate 1-L tanks for the barcoding experiments described below.

■**PAUSE POINT** Maintain stable lines by outcrossing single-copy F_1 animals. Adults can be maintained for up to 2 years.

scGESTALT sgRNA synthesis ● Timing 2 h

27 Generate sgRNAs targeting sites 1–4 of the scGESTALT barcode using the EnGen sgRNA Synthesis Kit according to the manufacturer's guidelines. Oligonucleotide sequences are provided in the Reagents list (see DNA oligonucleotide sequences).

28 Purify sgRNAs with an RNA Clean & Concentrator-25 Kit according to the manufacturer's guidelines and elute with 50 µl of nuclease-free water.

29 Quantify the sgRNAs using a NanoDrop spectrophotometer and dilute each sgRNA to ~350 ng/µl with nuclease-free water. Combine equal volumes (e.g., 5 µl) of sgRNAs 1–4 to make an equimolar pool.

▲**CRITICAL STEP** Avoid exposing sgRNA pools to freeze–thaw cycles. Each aliquot is sufficient for one injection mix.

■**PAUSE POINT** Store the sgRNA pool in 1.5-µl aliquots at −80 °C for up to 1 year.

scGESTALT 'early' barcode editing ● Timing 2 h

30 The day before performing microinjections, set up four to six overnight tanks with one scGESTALT F_1 female and one hsp:Cas9, U6:sgRNA F_1 male per tank for mating.

▲**CRITICAL STEP** Male hsp:Cas9, U6:sgRNA must be used in the 'early' and 'late' barcoding experiments to avoid maternal contribution of Cas9 mRNA and sgRNAs, and to enable barcode editing at later developmental timepoints. Female scGESTALT animals must be verified to have single-copy transgene integrations.

31 On the morning of the microinjections, assemble the following reagent mix (3.3 µl in total). Vortex to mix and centrifuge briefly (2,000g, room temperature, 2 s).

Component	Amount (µl)	Final concentration
Engen Cas9 NLS buffer	1	
Phenol red (0.1% (wt/vol))	0.3	0.009%
sgRNA pool (~350 ng/µl each; from Step 29)	1	~106 ng/µl
EnGen Cas9 NLS protein	1	
Total	3.3	

32 Incubate the mixture at room temperature for 5 min to enable Cas9–sgRNA complexes to assemble. Hold the mix on ice until ready to start injections.

33 Cross scGESTALT and hsp:Cas9, U6:sgRNA fish as described in Step 8.

34 While the fish are mating, prepare and calibrate the needles for microinjection as described in Step 9. Load 1.5 µl of the Cas9–sgRNA injection mix into the needle.

35 Collect fertilized eggs into Petri dishes with blue water and arrange one-cell-stage embryos for injections as described in Step 10. Inject 1.5 nl of Cas9–sgRNA mix through the chorion.

36 Transfer the injected embryos with blue water to a new Petri dish and incubate overnight at 28.5 °C. Keep some uninjected control embryos to assess the health of the clutch. At the end of the day, remove any unfertilized or damaged embryos from the injected batches.

scGESTALT 'late' barcode editing ● Timing 6 h

37 At 30 hpf, screen embryos from Step 36 for heart GFP expression as described in Step 15 to identify those carrying the scGESTALT transgene (~50% of injected progeny).

38 Heat-shock the sorted embryos for 30 min at 37 °C as described in Step 13.

39 Screen the embryos 2–4 h after heat shock for ubiquitous GFP expression as described in Step 14 to identify those carrying the hsp:Cas9, U6:sgRNA transgene.
    ? TROUBLESHOOTING

Rearing double-transgenic fish ● Timing ~25 d

40 Incubate sorted double-transgenic embryos from Step 39 (~25% of injected progeny) at 28.5 °C until 5 dpf and then transfer to 2-L tanks (eight to ten larvae per tank) in a zebrafish facility to raise to juveniles (23–25 dpf).
    ■**PAUSE POINT** The rearing time is user-dependent. Animals can be used in experiments at earlier stages or can be grown to adults.

Identification of barcode-edited juveniles ● Timing 1 d

41 The day before performing droplet encapsulation for scRNA-seq, identify individual juveniles from Step 40 with edited scGESTALT barcodes. Fin-clip the juveniles as described in Step 20. Transfer each fin-clipped fish to an individual 0.5-L tank to recover and number them uniquely.

42 Prepare genomic DNA as described in Steps 21 and 22.

43 Set up PCR reactions (10 μl) as follows:

Component	Amount (μl)	Final concentration
5× HF Phusion buffer	2	1×
10 mM dNTP	0.2	0.2 mM
Nuclease-free water	6.2	
scGESTALT primer mix (10 μM)	0.5	0.5 mM
Phusion polymerase	0.1	
Genomic DNA	1	
Total	10	

44 Run the following program on a thermocycler:

Cycle number	Denature	Anneal	Extend	Final
1	98 °C, 30 s			
2–34	98 °C, 10 s	63 °C, 20 s	72 °C, 20 s	
35			72 °C, 3 min	
36				8 °C, hold

45 Load the PCR products on 1% agarose gel, along with a 100-bp DNA ladder according to standard procedures.

46 Run the gel and visualize the bands. Successfully edited scGESTALT barcodes will appear as smears ranging from ~120 to 250 bp. Generally, >80% of injected embryos are efficiently edited. See Anticipated results for examples of successful editing.
    ? TROUBLESHOOTING

47 The next day, select two to four juveniles showing sufficient barcode editing for inDrops experiments and keep each juvenile in a separate 0.5-L tank overnight.

Preparation of single-cell suspension of whole brain ● Timing 2.5 h

48 The next day, pre-warm 40 ml of zebrafish facility water in a water bath at 37 °C in a 50-ml tube.

49 Heat-shock one 23- to 25-dpf juvenile selected in Step 47 to induce expression of the barcode array. Incubate the animal in a 50-ml tube with 20 ml of pre-warmed fish facility water at 37 °C for 40 min. This will promote expression of scGESTALT mRNA barcodes just before inDrops

encapsulation. While the fish is being heat-shocked, proceed with preparation of dissociation solutions as described below (Steps 50–54).

50 Transfer 25 ml of Neurobasal Medium to a 50-ml tube. Add 500 μl of B27 supplement. Oxygenate the solution by passing 95% O_2:5% CO_2 gas through tubing attached to a sterile 5-ml serological pipette for 2 min (Supplementary Fig. 1). Keep the Neurobasal/B27 medium on ice.

51 Prepare papain solution for dissociation. Add 5 ml of Neurobasal Medium to Vial 2 (papain) of the Papain Dissociation System. Add 0.5 ml of Neurobasal Medium to Vial 3 (DNase). Gently resuspend the DNase and transfer 250 μl to Vial 2. Mix the papain/DNase solution by inverting the vial several times.

52 Oxygenate the papain/DNase mixture by passing 95% O_2:5% CO_2 gas through tubing attached to a sterile 5-ml serological pipette for 2 min (Supplementary Fig. 1 and Supplementary Video 1).
▲ CRITICAL STEP Do not bubble gas directly into the solution, as this will cause frothing and protein denaturation, and reduce papain activity. Instead, slowly move the pipette in a circular motion above the liquid's surface, being careful not to cause the liquid to spill over (adjust the gas pressure, if necessary).

53 Prepare ovomucoid protease inhibitor (Vial 4) as described by the manufacturer.

54 Incubate the papain/DNase mixture in a 34 °C water bath while preparing the single-cell brain suspension as described below (Steps 55–61).

55 Anesthetize the heat-shocked juvenile from Step 49 by adding 800 μl of 25× MESAB.

56 Add 1 ml of DPBS (containing calcium, magnesium, glucose, and pyruvate) to a 1.5-ml microcentrifuge tube and hold on ice.

57 Add 10 ml of ice-cold Neurobasal/B27 medium to a 10-cm Sylgard dish. Add 400 μl of 25× MESAB.

58 Transfer the anesthetized fish to the Sylgard dish (Supplementary Fig. 2). Pin the fish just posterior of the head, in the middle of the trunk and near the tail, using three insect pins.

59 Using a pair of fine forceps, remove the jaw, eyes, heart, and gut tissues. Pierce the skin on top of the head and gently peel it back to expose the brain. Using fine forceps, gently scoop the brain out, taking care not to lose part of the hindbrain in the process. Rip the brain tissue into six to eight small pieces using two pairs of fine forceps. Typically, dissection takes ~5 min per brain.

60 Transfer the brain tissue pieces to the 1.5-ml microcentrifuge tube with DPBS from Step 56.

61 Briefly spin down the tube using a benchtop minicentrifuge (2,000g, room temperature, 2 s). Discard the DPBS and add 1 ml of ice-cold DPBS (no calcium, no magnesium). Keep the tube at room temperature.

62 Remove the papain/DNase mixture from the water bath (Step 54). Transfer 1 ml of the papain/DNase solution to a new 1.5-ml tube and add 1 ml of Neurobasal Medium. This dilutes the papain to 10 units/ml. Note that the concentration of papain can be adjusted, if needed (generally 10–20 units/ml is equally effective for dissociation).

63 Discard the DPBS from the brain tissue pieces (Step 61) and add 1 ml of the diluted papain/DNase mixture from Step 62. Transfer the entire contents of the 1.5-ml microcentrifuge tube to a 15-ml tube, and oxygenate the solution as described in Step 52.
▲ CRITICAL STEP Reoxygenation maintains cell viability during dissociation.

64 Incubate the brain tissue pieces in a 34 °C water bath for 20 min.

65 Gently triturate the tissue ten times with a p1000 pipette set to 800 μl.
▲ CRITICAL STEP Do not triturate too vigorously; avoid introducing bubbles. Generally, pipetting at a rate of 1 s per up-and-down motion of the pipette is desirable.

66 Reoxygenate the solution for 1 min as described in Step 52. Incubate in a 34 °C water bath for an additional 5–6 min.

67 Gently triturate the tissue 15–20 times with a p1000 pipette as described in Step 65.
▲ CRITICAL STEP Typically, this range of trituration is sufficient for efficient tissue dissociation. However, depending on the strength of the papain, if large pieces of tissues are still visible, it may be necessary to perform additional pipetting or to increase the incubation time by 5–10 min. If a longer incubation at 34 °C is necessary, repeat Step 66 just before re-incubation.

68 Transfer the entire contents of the 15-ml tube to a 1.5-ml microcentrifuge tube. Spin at 500g for 4 min at 4 °C.

69 While the tube is spinning, prepare inhibitor solution as follows. Transfer 0.9 ml of Earl's Balanced Salt Solution (Vial 1) to a 15-ml tube. Add 100 μl of resuspended ovomucoid protease inhibitor from Step 53. Add 50 μl of resuspended DNase (Vial 3) from Step 51. Oxygenate the mix as described in Step 52 and keep on ice.

70 Following centrifugation in Step 68, discard the papain supernatant. A clear cell pellet should be visible. Resuspend the pellet with 1 ml of the prepared inhibitor solution.

71 Spin at 500*g* for 4 min at 4 °C.

72 Discard the supernatant and gently resuspend the pellet with 1 ml of ice-cold DPBS (no calcium, no magnesium).

73 Sequentially filter the resuspension first through a 35-μm cell strainer, followed by a 20-μm cell strainer. Perform all steps on ice with prechilled tubes.

74 Spin at 500*g* for 4 min at 4 °C.

75 Discard the supernatant and add 500 μl of ice-cold DBPS (no calcium, no magnesium). Do not pipette the pellet. This step is to remove any residual debris.

76 Spin at 500*g* for 4 min at 4 °C.

77 Discard the supernatant and resuspend the pellet in 400 μl of ice-cold DPBS (no calcium, no magnesium). Keep on ice.

78 Determine the cell concentration and viability. Add 10 μl of the cell suspension to 10 μl of trypan blue in a 1.5-ml microcentrifuge tube. Mix by gentle pipetting and load 10 μl onto a hemocytometer. Count the cells and determine the concentration as described by the manufacturer. Cell viability should be >70% (ideally >80%).

    ▲ **CRITICAL STEP** High cell viability is required to proceed with scRNA-seq experiments. Poor viability will result in fewer cells passing data quality filtering steps and can introduce more ambient RNA (i.e., non-cell-associated RNA, such as free-floating RNA from dead or dying cells) into the cell suspension, compromising single-cell measurements.

    ? TROUBLESHOOTING

79 Dilute the cells to ~200,000 cells/ml with DPBS (no calcium, no magnesium).

80 Add 300 μl of 36% (vol/vol) OptiPrep to 300 μl of diluted cells and gently resuspend the cell suspension. The final cell concentration is 100,000 cells/ml in 18% (vol/vol) OptiPrep/DPBS. Proceed immediately to inDrops cell encapsulation.

inDrops single-cell encapsulation ● Timing 30–60 min

81 Run the inDrops device as described previously[73]. Depending on the desired number of cells to be profiled, cells can be collected at a rate of 10,000–20,000 cells/h. Generally, single-cell transcriptomes are obtained for ~70% of cells introduced into the device.

Reverse transcription in droplets ● Timing 3 h

82 Proceed to reverse transcription as described previously[73]. If >3,000 cells were captured, split the emulsion into fractions containing ~3,000 cells.

    ■ **PAUSE POINT** Libraries can be stored after reverse transcription at −80 °C for at least 3 months.

inDrops transcriptome library preparation ● Timing 2 d

83 Generate single-cell transcriptome libraries as described previously[73], with the following modification. After IVT (for linear amplification) of the second-strand synthesis product, purify the reaction mixture with a 1.3× volume of AMPure beads according to the manufacturer's instructions, and elute with 25 μl of RNA elution buffer.

84 Use 9 μl of the purified post-IVT product to proceed with fragmentation of amplified RNA and complete library preparation as described previously[73].

    ■ **PAUSE POINT** Store the remaining unfragmented RNA at −80 °C for up to 3 months.

scGESTALT library preparation: reverse transcription ● Timing 2.5 h

85 Assemble the following reaction mixture (11 μl in total) on ice using unfragmented RNA from Step 84:

Component	Amount (μl)	Final concentration
Unfragmented RNA	5	
50 μM random hexamer	1.5	3.75 μM
10 mM dNTP	1	0.5 mM
Nuclease-free water	3.5	
Total	11	

86 Vortex the mixture and briefly centrifuge (2,000*g*, room temperature, 2 s). Incubate the reaction at 70 °C (lid at 105 °C) for 3 min and immediately cool on ice.

87 Add the following reverse transcription reagents on ice and mix by pipetting (20 µl in total). Briefly spin down (2,000g, room temperature, 2 s).

Component	Amount (µl)	Final concentration
Mixture from Step 86	11	
5× PrimeScript buffer	4	1×
Nuclease-free water	3.5	
RNaseOut (40 U/µl)	1	2 U/µl
PrimeScript RT enzyme (200 U/µl)	0.5	5 U/µl
Total	20	

88 Incubate the reaction mixture in a thermocycler at 30 °C for 10 min, followed by 42 °C for 1 h, and then 70 °C for 15 min.

89 Purify the reverse transcription product with a 1.2× volume of AMPure beads (24 µl) and elute with 20 µl of DNA elution buffer.
    ■ **PAUSE POINT** The purified cDNA can be stored at −20 °C for at least 3 months.

scGESTALT library preparation: PCR round 1 ● Timing 1.5 h

90 Set up a PCR reaction (25 µl) for targeted amplification of scGESTALT barcodes as follows:

Component	Amount (µl)	Final concentration
5× Q5 reaction buffer	5	1×
10 mM dNTP	0.5	0.2 mM
10 µM inDrops_GP6	1.25	0.5 µM
10 µM inDrops_PE1Sa	1.25	0.5 µM
cDNA from Step 89	7.5	
Q5 DNA polymerase	0.25	
Nuclease-free water	4.25	
Q5 enhancer	5	
Total	25	

▲ **CRITICAL STEP** This PCR amplifies a longer scGESTALT fragment, using a forward primer that binds to the DsRed transgene upstream of the scGESTALT barcode. We have found that using a nested PCR strategy improves the specificity of amplification.

91 Run the following program on a thermocycler:

Cycle number	Denature	Anneal	Extend	Final
1	98 °C, 30 s			
2–16	98 °C, 10 s	61 °C, 25 s	72 °C, 30 s	
17			72 °C, 2 min	
18				8 °C, hold

92 Add 25 µl of nuclease-free water to the PCR product and purify with a 0.6× volume of AMPure beads (30 µl). Elute the product with 20 µl of DNA elution buffer.
    ■ **PAUSE POINT** The purified PCR product can be stored at −20 °C for at least 1 week.

scGESTALT library preparation: PCR round 2 ● Timing 1.5 h

93 Set up a second PCR reaction (25 µl) for targeted amplification of scGESTALT barcodes as follows:

Component	Amount (µl)	Final concentration
5× HF Phusion reaction buffer	5	1×
10 mM dNTP	0.5	0.2 mM
		Table continued

(continued)

Component	Amount (µl)	Final concentration
10 µM inDrops_GP12	1.25	0.5 µM
10 µM inDrops_PE1Sb	1.25	0.5 µM
PCR product from Step 92	8	
Phusion DNA polymerase	0.25	
Nuclease-free water	8.75	
Total	25	

▲ **CRITICAL STEP** This PCR amplifies the desired scGESTALT fragment, using a forward primer that binds the 5′ end of the scGESTALT barcode sequence. The primer contains 9×N sequences, where N is a random base, to introduce sequence variety.

94 Run the following program on a thermocycler:

Cycle number	Denature	Anneal	Extend	Final
1	98 °C, 30 s			
2–9	98 °C, 10 s	60 °C, 25 s	72 °C, 30 s	
16			72 °C, 2 min	
17				8 °C, hold

95 Add 25 µl of nuclease-free water to the PCR product and purify with a 0.6× volume of AMPure beads (30 µl). Elute the product with 15 µl of DNA elution buffer.
■ **PAUSE POINT** The purified PCR product can be stored at −20 °C for at least 1 week.

96 Check the concentration of 1 µl of eluate, using a Qubit fluorometer according to the manufacturer's guidelines. Generally, concentrations range from 1 to 2.5 ng/µl. If concentrations are substantially higher, repeat Steps 93 and 94 with lower DNA input and/or fewer PCR cycles.

scGESTALT library preparation: final PCR ● Timing 1.5 h

97 Set up a PCR reaction (25 µl) for final library amplification as follows:

Component	Amount (µl)	Final concentration
2× KAPA HiFi HotStart Ready Mix	12.5	1×
Eluate from Step 95 (1:10 dilution)	5	
R1-PCRix/R2-PCR primer mix (5 µM)	2.5	0.5 µM
Nuclease-free water	5	
Total	25	

▲ **CRITICAL STEP** If multiplexing libraries, use different R1-PCRix/R2-PCR primer mix variants. For each scGESTALT library, ensure that the same primer pairs as those used for building the transcriptome libraries are used. For example, if multiplexing two libraries (e.g., A and B) in the same sequencing run, use R1-PCRix1/R2-PCR (for library A) and R1-PCRix2/R2-PCR (for library B) for both transcriptome and scGESTALT sequencing.

98 Run the following program on a thermocycler:

Cycle number	Denature	Anneal	Extend	Final
1	98 °C, 30 s			
2–3	98 °C, 20 s	55 °C, 30 s	72 °C, 30 s	
4–9/11[a]	98 °C, 20 s	65 °C, 30 s	72 °C, 30 s	
12			72 °C, 2 min	
13				8 °C, hold

[a]The number of cycles performed at the annealing temperature of 65 °C typically ranges from 6 to 8. This number can also be determined empirically by qPCR as described previously[70].

99 Add 25 µl of nuclease-free water to the PCR product and purify with a 0.6× volume of AMPure beads (30 µl). Elute the product with 25 µl of RNA elution buffer. The eluate is the final sequencing-ready library.

■ PAUSE POINT Libraries can be stored at −20 °C for at least a month.

100 Check the concentration of 1 µl of eluate using a Qubit fluorometer. Generally, concentrations range from 2.5 to 6 ng/µl.

scGESTALT library preparation: library quality check ● Timing 2 h

101 Use the Agilent High Sensitivity DNA Kit with a Bioanalyzer, according to the manufacturer's protocol, to determine library size with 1 µl of eluate from Step 99. See Anticipated results for examples of successful libraries.

102 Determine the molar concentration of the library. Dilute the library to 10 nM. If sequencing multiple libraries together, combine equal volumes of libraries (each at 10 nM) to be multiplexed in 1.5 ml of low-adhesion microcentrifuge tubes.

■ PAUSE POINT Libraries can be stored at −20 °C for at least a month or at −80 °C for a longer period.

Sequencing transcriptome and scGESTALT libraries ● Timing 1 d

103 Sequence transcriptome libraries using a NextSeq 500/550 High Output v2 Kit (75-cycle kit) with standard Illumina primers.

 Read 1 (transcript read) is 61 cycles.
 Index 1 (part 1 of inDrops cell barcode) is eight cycles.
 Index 2 (library index) is eight cycles.
 Read 2 (part 2 of inDrops cell barcode (8 bp), followed by 6-bp UMI) is 14 cycles.

104 Sequence scGESTALT libraries using the MiSeq Reagent Kit v2 (300-cycle kit) and 20% PhiX spike-in with standard Illumina primers.

 Read 1 (scGESTALT lineage barcode) is 250 cycles.
 Index 1 (part 1 of inDrops cell barcode) is eight cycles.
 Index 2 (library index) is eight cycles.
 Read 2 (part 2 of inDrops cell barcode (8 bp), followed by 6-bp UMI) is 14 cycles.

Sequencing data processing ● Timing 1 d

105 Process raw transcriptome reads as described previously[73]. The inDrops bioinformatics pipeline (inDrops.py) is available at https://github.com/indrops/indrops. Bowtie v.1.1.1 (https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.1/) is used with parameter −e 200; UMI quantification is used with parameter −u 2.

106 Process scGESTALT reads in parallel, using the inDrops.py script, and stop just prior to the transcriptome-mapping step. Modify the Trimmomatic settings to LEADING: '10'; SLIDING-WINDOW: '4:5'; MINLEN: '16'. This will generate files with inDrops cell identifiers and the scGESTALT lineage barcode sequence. The data can be further processed using the scGestaltPrepFunc.R script in Supplementary Software 1 or a similar custom script.

107 Process the resulting files, using the pipeline available at https://github.com/aaronmck/SC_GESTALT (Supplementary Software 2). This will generate tables containing the lineage barcode sequence for each cell.

Downstream data processing ● Timing 3 d

108 Using Seurat (or the related scRNA-seq analysis tool), identify cell clusters (i.e., different cell types) using the transcriptome scRNA-seq data (for a tutorial, refer to https://satijalab.org/seurat/). In brief, gene expression profiles are used to perform dimensionality reduction using principal component analysis with highly variable genes. A modularity-based clustering algorithm (Louvain) is used to cluster cells into discrete cell types using top principal components. The result is visualized in two dimensions on a t-distributed stochastic neighbor embedding (t-SNE) plot (Fig. 1). For further details on how clustering is performed for the zebrafish juvenile brain dataset, we refer the reader to the Methods section of ref. [72]. See Anticipated results for examples of clustering analyses from zebrafish brain scRNA-seq.

109 Match transcriptome (from Step 108) and lineage barcodes (from Step 107) for each cell, using the inDrops cell identifier index sequence (Supplementary Software 3 and 4), and generate lineage trees
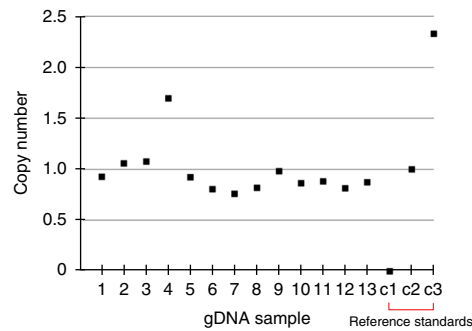
Fig. 4 | Barcode copy-number determination. Copy number of scGESTALT barcode integration was determined using qPCR. Samples 1–13 represent different genomic DNA samples ($n = 13$ animals). Reference standards: c1 is a negative control ($n = 1$ animal); c2 is known to contain one copy ($n = 1$ animal); and c3 is known to contain two copies ($n = 1$ animal). This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25–08.
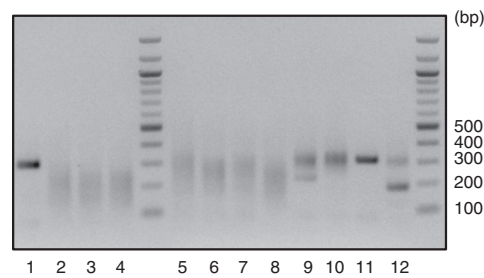

Fig. 5 | scGESTALT barcode editing. scGESTALT barcode zebrafish were crossed to zebrafish that express heat shock–inducible Cas9 and U6-driven sgRNAs 5–9. The resulting embryos were injected with Cas9 protein and sgRNAs 1–4 at the one-cell stage. Embryos were heat-shocked at 30 hpf to induce transgenic Cas9 for a late round of editing. Double-transgenic (scGESTALT$^+$, hsp:Cas9$^+$; lanes 2–8, $n = 7$ embryos) and single-transgenic embryos (scGESTALT$^+$, hsp:Cas9$^-$; lanes 9–12, $n = 4$ embryos) were identified by screening for GFP expression. The gel shows PCR results of amplifying the scGESTALT barcode (unedited = ~300 bp). Large smear patterns (120–250 bp) are observed in early- and late-edited embryos (lanes 2–8), whereas embryos that were only mutated at sites 1–4 display less editing (lanes 9–12). The band at ~200 bp in lane 12 probably represents large deletion(s) between sites 1 and 4 that occurred early in development and were inherited by most cells. Note that samples with such dominant large deletions should not be used for downstream experiments and analyses, as they are likely to have low barcode diversity. Sample in lane 11 was probably not efficiently injected. Lane 1 represents a control embryo, which was injected with Cas9 protein only (no sgRNAs 1–4, $n = 1$ embryo) and was not heat-shocked. As expected, the barcode is not edited in this case. This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25–08.

(based on maximum parsimony), using the pipeline available at https://github.com/aaronmck/SC_GESTALT. See Anticipated results for an example lineage tree.

Troubleshooting

Extensive troubleshooting pointers for running the inDrops device and preparing inDrops transcriptome libraries have been provided previously[73].

There are three main points in the scGESTALT protocol at which issues may arise.

Inefficient Cas9 transgene induction (Step 39)

After heat shock of the embryos at 30 hpf, GFP expression is used as a proxy for successful Cas9 transgene induction. If weak or no GFP is observed, it indicates that heat shock was not successful. Test the heat-shock conditions empirically, and, if necessary, increase the time of heat shock by 10 min. Do not heat-shock for >45 min, as this can result in large deletions in the lineage barcode array. Alternatively, screen F1 transgenic animals to identify those with robust GFP expression following heat shock.
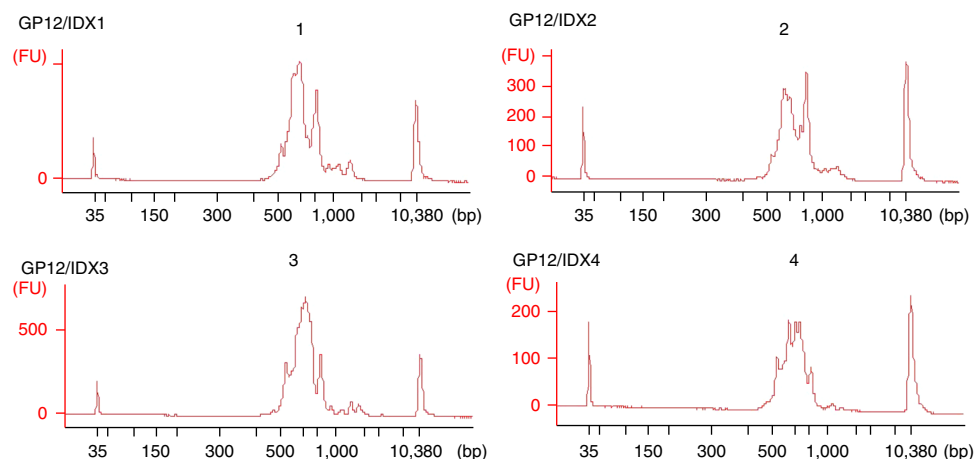
Fig. 6 | Bioanalyzer electropherograms of scGESTALT sequencing libraries. Traces were obtained from 23- to 25-dpf juvenile zebrafish brains ($n = 1$ animal). The single-cell emulsion (Step 82) from one inDrops collection was split into four fractions of ~3,000 cells each, and the resulting libraries were indexed uniquely. The average size of the libraries is 600–650 bp. Peaks at 35 and 10,380 bp represent gel migration markers. This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25–08.

No or inefficient editing of the barcode (Step 46)

This usually indicates problems with the editing reagents. Ensure that the Cas9 protein used for injection is active by testing it for successful editing of endogenous genes, e.g., the *tyr* gene encoding tyrosinase results in loss of pigmentation by 48 hpf when mutated (see ref. [81] for details and *tyr* sgRNA sequences). If the Cas9 is active, then the sgRNA pool is probably compromised. Remake the sgRNAs, aliquot in single-use amounts, and store at −80 °C for up to 1 year. Do not freeze/thaw sgRNAs, and discard each aliquot after use.

Low cell viability (Step 78)

High-quality single-cell preparations are critical for obtaining successful single-cell transcriptome and scGESTALT libraries. If cell viability is <70%, repeat dissociation with another sample. Ensure that all media are properly oxygenated (this is the most critical part of the protocol), especially during the papain incubation at 34 °C (Steps 63 and 64). Do not incubate tissue for >30 min at 34 °C and do not triturate the sample vigorously after papain digestion. Gently pipette cells during DBPS washes and the final resuspension.

Timing

Zebrafish transgenesis
Steps 1–5, Tol2 mRNA synthesis: 1 d
Steps 6–12, microinjection: 3 h
Steps 13–15, screening and rearing fish: 2–3 months
Steps 16–19, screening for germline transmission: 2 d, wait for adults 2–3 months
Steps 20–26, copy-number determination: 1 d
scGESTALT barcode editing
Steps 27–29, sgRNA synthesis: 2 h
Steps 30–36, early-timepoint Cas9 editing: 2 h
Steps 37–39, late-timepoint Cas9 editing: 6 h
Step 40, rearing double-transgenic fish: ~25 d
Steps 41–47, identification of edited fish: 1 d
inDrops transcriptome and lineage profiling
Steps 48–80, single-cell processing of brain tissue: 2.5 h
Step 81, cell encapsulation: 30–60 min
Step 82, reverse transcription: 3 h
Steps 83 and 84, transcriptome library preparation: 2 d
Steps 85–89, scGESTALT reverse transcription: 2.5 h
Steps 90–92, scGESTALT PCR round 1: 1.5 h
Steps 93–96, scGESTALT PCR round 2: 1.5 h
Steps 97–100, scGESTALT final library PCR: 1.5 h
Steps 101 and 102, checking of library with Bioanalyzer: 2 h

Sequencing and raw data processing
Steps 103 and 104, MiSeq and NextSeq runs: 1 d
Steps 105–107, processing of raw data: 1 d
Downstream data processing
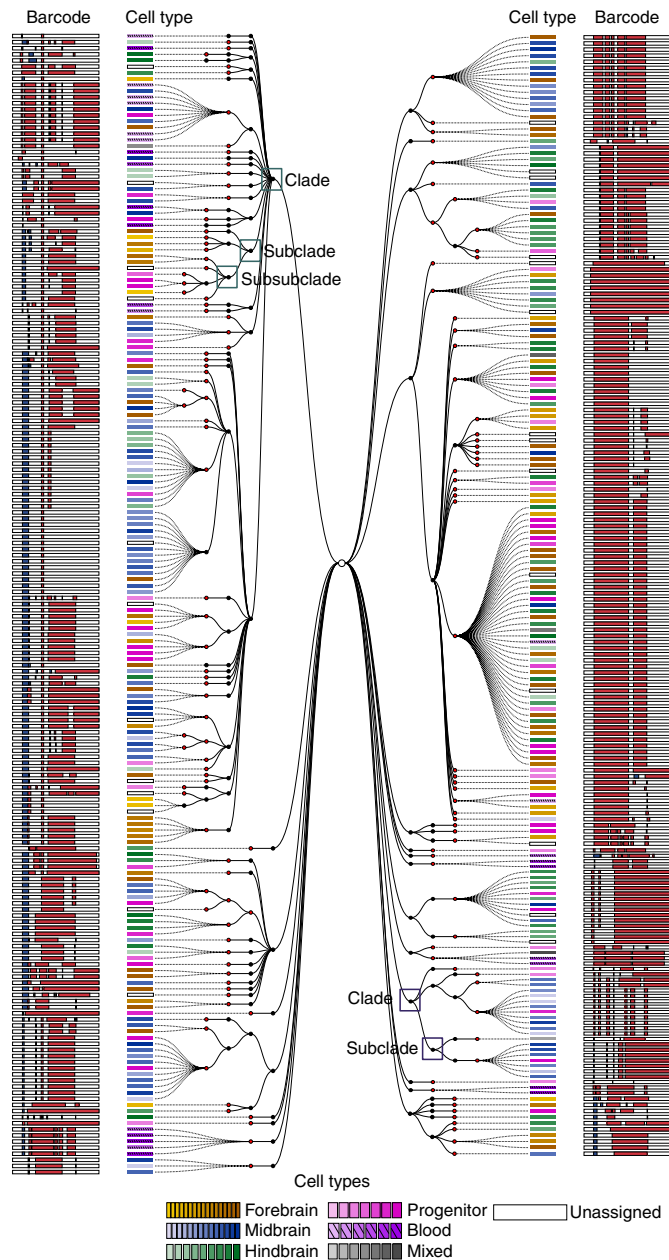Steps 108 and 109, clustering analysis: 3 d (preliminary analysis)

Fig. 7 | Zebrafish brain lineage tree generated using scGESTALT. An example of a reconstructed lineage tree from a single juvenile zebrafish brain. 376 edited barcodes were recovered from single cells using inDrops. A cell-lineage tree was generated from the barcodes based on shared edits using a maximum-parsimony approach. Black nodes represent early barcode edits (Cas9 and sgRNA injection at the one-cell stage, Step 35); red nodes represent late edits (heat shock–induced Cas9 transgene expression, Step 38). Dashed lines join single cells to terminal nodes (representing the final edited barcode sequence) on the tree. Distinct cell types (identified from simultaneous transcriptome capture and cell-clustering analyses) are color-coded as indicated in the legend. The edited barcode for each cell is shown as a white bar with deletions (red) and insertions (blue). Examples of clades and subclades are indicated on the tree. A clade on the tree represents all lineage barcodes that share at least one common edit, and subclades that branch from the original clade contain increasingly restricted subsets of barcodes that contain the previous edit(s) as well as additional shared edits. Adapted with permission from Raj et al.[72], Springer Nature. This procedure was approved by the HU/FAS Committee on the Use of Animals in Research & Teaching under protocol no. 25-08.

Anticipated results

The protocol described herein enables the user to apply CRISPR–Cas9 for cumulative, combinatorial, and heritable genomic barcode editing at multiple timepoints to generate permanent records of cell lineages during development. As the mutated barcodes are also expressed as mRNA, both lineage information and cell-type identity can be simultaneously extracted from single cells at high throughput using droplet-based scRNA-seq. Cell relationships at the level of gene expression and lineages can then be represented on lineage trees, which can be explored to identify the timing and patterns of lineage segregation within or between tissue regions and cellular subtypes.

Below, we describe results obtained at various stages of the protocol that can be used to gauge the experiment's success.

Barcode copy-number determination by qPCR (Step 25)

The number of integrations of the scGESTALT lineage barcode in the transgenic F_1 generation can be determined by qPCR. Figure 4 shows representative qPCR results from this assay. Animals with a calculated copy number of ≤1 (those with <0.6 are not considered integrants) are considered to be single integrants and can be used for downstream experiments and further expanded to maintain single-copy transgenic lines. Raw data for the calculations are included in the Supplementary Data.

Identification of successful barcode editing (Step 46)

Genomic DNA from edited fish is extracted, and lineage barcodes are PCR-amplified and visualized on an agarose gel. As shown in Fig. 5, band patterns that appear as smears ranging from 120 to 250 bp are indicative of efficiently edited barcodes. Depending on whether barcodes were edited at only 'early' timepoints by injections, only 'late' timepoints by Cas9 induction, or at both 'early' and 'late' timepoints by combination of Cas9 injection and induction, the gel band patterns have slight distinctions. If the pattern is dominated by short bands at ~100–150 bp, it is advisable to avoid using the corresponding animal in scRNA-seq experiments, as it is likely that this individual's barcode edits will be predominantly large deletions, reducing the diversity of edits and the resolution of the lineage trees. In addition, because Cas9-induced mutations are stochastic, gel band patterns in animals that were edited in a similar manner will be variable.

Bioanalyzer electropherograms of successful scGESTALT libraries (Step 101)

The profiles of amplified lineage barcode libraries are shown in Fig. 6. A successful library typically ranges in size between 500 and 800 bp, with an average of 600–650 bp. Some larger fragments may be observed, but these are generally not problematic for sequencing. Alternatively, the ratio of AMPure beads at Steps 95 and 99 can be increased (e.g., to 0.8× or 1×) to remove such fragments. If fragments are considerably smaller than this range, it may indicate that the barcode contains large deletions. Refer to the previous section to determine how to avoid using animals with undesirable barcode mutation patterns.

Detailed discussion of results from inDrops transcriptome libraries has been provided before[73]. It is worth noting that zebrafish neural cells have low RNA content, with estimates in the 1- to 2-pg range[23]. This is still sufficient for scRNA-seq profiling, and Bioanalyzer electropherograms display similar profiles, as previously described for inDrops libraries from other cell types[73].

Clustering analyses and lineage trees (Steps 108 and 109)

The digital gene expression matrix output (rows contain gene counts, and columns represent single cells) from the inDrops processing pipeline is used with any relevant scRNA-seq computational tool to identify cell types. We will use Seurat[18] as an example for the rest of this section. First, the matrix is column-normalized and log-transformed (see tutorial at https://satijalab.org/seurat/). Next, low-quality single-cell libraries are filtered out from the dataset. The metrics for what constitutes 'low quality' will vary between cell types, tissues, dissociation protocols, and sequencing depth. For zebrafish brain data, we filtered cells with <500 expressed genes or >9% mitochondrial content[72]. Potential cell doublets or multiplets should also be filtered out (e.g., by identifying cells with high numbers of gene counts and UMIs that are outliers of a normal distribution). Next, highly variable genes are identified and used for principal component analysis. The top principal components are used for dimensionality reduction and for grouping all cells into distinct clusters (using a Louvain modularity algorithm). Clusters are expected to be supported by cells from multiple biological replicates; otherwise they may indicate technical artifacts in the data. The resulting clusters can be

compared to each other to identify differentially expressed genes, which may serve as gene markers for individual cell types. Clusters can be classified using the identified gene markers and validated with in situ hybridizations.

To generate scGESTALT lineage trees, the inDrops index sequences are used to match transcriptomes and lineage barcodes for the same cells. Cell-lineage trees are generated using maximum parsimony (adapted from phylogenetics) based on patterns of shared edits (for more details, see refs [66,72]). To represent early and late editing, the tree is anchored with edits that occur at sites 1–4 of the barcode array (early editing from injections at the one-cell stage) and is then extended with additional edits accumulated at sites 5–9 (late editing from heat shock–induced Cas9 activity). A clade on the tree represents all lineage barcodes that share at least one common edit, and subclades that branch from the original clade contain increasingly restricted subsets of barcodes that contain the previous edit(s) as well as additional shared edits (Figs. 1 and 7). Individual cells (identified by their inDrops index) containing each of the recovered barcodes are connected to the tips of the tree (i.e., the terminal lineage barcode sequence). Cells connected to the same terminal node are clonal (i.e., contain the same lineage barcode). Cell annotations (e.g., cell type and spatial/regional origin) can be color-coded and added to the tree to explore various lineage relationships to, for example, determine the relationship between cells belonging to the same cluster or cells from different regions of the tissue.

Reporting Summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

References

1. Tanay, A. & Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
2. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).
3. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: recording the past and predicting the future. *Science* **358**, 69–75 (2017).
4. Han, X. et al. Mapping the mouse cell atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).
5. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
6. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
7. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
8. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
9. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
10. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
11. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
12. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
13. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
14. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
15. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
16. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
17. Mulqueen, R. M. et al. Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428–431 (2018).
18. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
19. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
20. Kiselev, V. Y., Yiu, A. & Hemberg, M. scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359–362 (2018).

21. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

22. Shekhar, K. et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323.e30 (2016).

23. Pandey, S., Shekhar, K., Regev, A. & Schier, A. F. Comprehensive identification and spatial mapping of Habenular neuronal types using single-cell RNA-seq. *Curr. Biol.* **28**, 1052–1065.e7 (2018).

24. Cusanovich, D. A. et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).

25. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).

26. Halpern, K. B. et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356 (2017).

27. La Manno, G. et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016).

28. Da Mi et al. Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018).

29. Tusi, B. K. et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018).

30. Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).

31. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* **3**, 346–360.e4 (2016).

32. Haber, A. L. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

33. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).

34. Hochgerner, H., Zeisel, A., Lönnerberg, P. & Linnarsson, S. Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nat. Neurosci.* **21**, 290–299 (2018).

35. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

36. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

37. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).

38. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).

39. Spanjaard, B. & Junker, J. P. Methods for lineage tracing on the organism-wide level. *Curr. Opin. Cell Biol.* **49**, 16–21 (2017).

40. Woodworth, M. B., Girskis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017).

41. Ma, J., Shen, Z., Yu, Y.-C. & Shi, S.-H. Neural lineage tracing in the mammalian brain. *Curr. Opin. Neurobiol.* **50**, 7–16 (2018).

42. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).

43. Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).

44. Pei, W. et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).

45. Fuentealba, L. C. et al. Embryonic origin of postnatal neural stem cells. *Cell* **161**, 1644–1655 (2015).

46. Harwell, C. C. et al. Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* **87**, 999–1007 (2015).

47. Mayer, C. et al. Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* **87**, 989–998 (2015).

48. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).

49. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).

50. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* **10**, 973–976 (2013).

51. Maeder, M. L. et al. CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).

52. Cheng, A. W. et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).

53. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. USA* **112**, 11870–11875 (2015).

54. Liu, X. et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell* **170**, 1028–1043.e19 (2017).

55. Liao, H.-K. et al. In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. *Cell* **171**, 1495–1507.e15 (2017).
56. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
57. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
58. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
59. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
60. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
61. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
62. Nelles, D. A. et al. Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell* **165**, 488–496 (2016).
63. Mikuni, T., Nishiyama, J., Sun, Y., Kamasawa, N. & Yasuda, R. High-throughput, high-resolution mapping of protein localization in mammalian brain by in vivo genome editing. *Cell* **165**, 1803–1817 (2016).
64. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016).
65. Chow, R. D. et al. AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glio-blastoma. *Nat. Neurosci.* **20**, 1329–1341 (2017).
66. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
67. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
68. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108 (2018).
69. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
70. Kalhor, R., Mali, P. & Church, G. M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017).
71. Schmidt, S. T., Zimmerman, S. M., Wang, J., Kim, S. K. & Quake, S. R. Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* **6**, 936–942 (2017).
72. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
73. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
74. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
75. Bae, T. et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2018).
76. Kawakami, K. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* **8** (Suppl. 1), S7 (2007).
77. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
78. Suster, M. L., Abe, G., Schouw, A. & Kawakami, K. Transposon-mediated BAC transgenesis in zebrafish. *Nat. Protoc.* **6**, 1998–2021 (2011).
79. Fisher, S. et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305 (2006).
80. Pan, Y. A. et al. Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **140**, 2835–2846 (2013).
81. Yin, L. et al. Multiplex conditional mutagenesis using transgenic expression of Cas9 and sgRNAs. *Genetics* **200**, 431–441 (2015).

Acknowledgements

Author contributions

B.R. and J.A.G. developed and optimized the scGESTALT protocols and analyzed the data. B.R. wrote the manuscript with edits by A.F.S. and J.A.G.

Data availability

Figure 4 has associated raw data (Supplementary Data). There is no restriction on data availability. scGESTALT computational scripts and analysis pipeline are available at https://github.com/aaronmck/SC_GESTALT and are included as Supplementary Software 2 with this protocol.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-018-0058-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.R.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 23 October 2018

Related links
Key references using this protocol
Raj, B. et al. *Nat. Biotechnol.* **36**, 442–450 (2018): https://doi.org/10.1038/nbt.4103
McKenna, A. et al. *Science* **353**, aaf7907 (2016): https://doi.org/10.1126/science.aaf7907