Title: Unbiased profiling of multipotency landscapes reveals spatial modulators of clonal fate biases

Authors: Alek G Erickson[1]†, Sergey Isaev[2]†, Artem Artemov[2], Jingyan He[3], Bettina Semsch[3], Aliia Murtazina[1], Jia Sun[4], Katrin Mangold[3], Anthi Chalou[1], Jonas Frisen[3], Michael Ratz[3], Emma Andersson[3], Peter V. Kharchenko[5,6] * and Igor Adameyko[1,2] *

† - These authors contributed equally to this work
* - Corresponding authors. Email: igor.adameyko@ki.se and peter.kharchenko@post.harvard.edu

Affiliations:
[1] Department of Physiology and Pharmacology, Karolinska Institutet; 17177 Stockholm, Sweden.
[2] Department of Neuroimmunology, Center for Brain Research, Medical University Vienna; 1090 Vienna, Austria.
[3] Department of Cell and Molecular Biology, Karolinska Institutet; 17177 Stockholm, Sweden.
[4] Department of Comparative Medicine, Karolinska Institutet; 17177 Stockholm, Sweden.
[5] San Diego Institute of Science, Altos Labs; San Diego, CA, USA.
[6] Department of Biomedical Informatics, Harvard Medical School; Boston, MA 02115, USA.

Abstract: Embryogenesis is commonly viewed through a tree model of cell differentiation, which does not adequately represent the spatiotemporal modulation of cell multipotency underlying morphogenesis. Here we develop an integrated approach, combining *in vivo* single-cell high-throughput clonal lineage tracing with machine learning, to systematically decompose continuous spectra of clonal fate biases in mouse embryos traced from neurulation until mid-gestation. The reconstructed patterns of clonal variation uncovered gene programs driving dynamic positional biasing of clonal composition during axial skeletogenesis and peripheral neurogenesis. Mosaic combinatorial perturbations targeting multiple receptors, including the Hedgehog pathway, led to novel clone types, which has implications for engineering custom cell type assemblages from well-defined progenitors *in vivo*. Altogether, our work demonstrates an effective practical approach for interrogating programs guiding lineage specification.

Main Text:

Introduction

The stochastic nature of vertebrate embryonic development poses a key challenge in understanding how the entire organism robustly emerges from a single cell, through an orderly sequence of cell fate decisions and mitotic divisions (Fig. 1A) [1]. Each tissue, organ, and body part, is formed from a defined set of progenitor cells that generate progressively elaborate clonal structures, varying in composition according to their unique lineage histories (Fig. 1B). This natural variability is essential for the formation of the body plan and involves precise spatiotemporal

regulation of clonal structures via anisotropic proliferation and fate restrictions. Consequently, the classic model of cell differentiation as a "tree-of-fates" is a simplified representation, whereas the exact realization of such a tree is variably regulated in each location of the developing organism (Fig. 1C). Thus, a deeper understanding of organismal development requires massively parallel analysis of individual clonal fates in relation to instructive signals, timing, and locations.

Lineage tracing techniques have long been at the core of developmental biology, providing critical insights for identifying key progenitor populations and their differentiation potentials [2]. Earlier clonal tracing approaches, such as those that rely on dye injections or multiplexed imaging of fluorescent proteins, have generally been limited in scale and throughput. Recent advances in multiplexed genetic tracing methods, such as Brainbow [3], scGESTALT [4], PolyLoxExpress [5], CARLIN/DARLIN [6, 7], and LINNAEUS [8], enabled construction of high resolution lineage trees. These technologies have refined our understanding of cell differentiation pathways by revealing how cell progenitor fates vary systematically across biological contexts. Variability in cell fate can result from both intrinsic cellular biases and external environmental influences, though the molecular logic driving these variations remain incompletely understood [9, 10]. Identifying such biasing factors is an important step towards elucidating the genetic and epigenetic mechanisms that coordinate cell destinies in vivo.

Despite recent advancements in combining highly multiplexed tracing with analysis of cellular state, a number of practical challenges remain. One such difficulty is sparse coverage, where only a small fraction of cells within a tissue or organism are effectively labeled and recovered, making it difficult to reconstruct comprehensive lineage trees and potentially skewing the interpretation of the resulting lineages towards more proliferative cells. More fundamentally, the measurements of the cellular state that accompany tracing information are performed at the final timepoint. Consequently, molecular differences associated with the fate decisions cannot be observed directly and must be inferred under additional assumptions, such as ergodicity of the differentiation process or persistence of past progenitor signatures at the final time point. The asynchronous and inductive nature of mammalian development, however, in many cases allows to overcome these general challenges. For example, the timing of neural crest migration along different anterior-posterior coordinates results in an effective asynchrony that can be observed at a single developmental timepoint. Similarly, neuro-mesodermal progenitors persist over longer periods of time during development, generating somites and extending the body axis [11].

Here, we revisit the framework of clonal analysis using a cellular barcoding approach, combining highly multiplexed sparse tracing with unsupervised learning techniques and analysis of expression states, to investigate fate biasing across the cell type landscape of ectodermal and mesodermal derivatives during mammalian embryonic development. This approach exposed continuous clonal variations occurring along the body axes and sequential fate restrictions over time, enabling identification of auxiliary regulatory programs likely responsible for modulation of fates in multipotent progenitor populations. Finally, we incorporate multiplexed clonal perturbations to interrogate the potential role of different signaling mechanisms in modulating cell fate decisions during embryogenesis.

Profiling of clonal behavior in mouse embryos

95 To reconstruct joint patterns of transcriptional states and cell lineages in post-neurulation murine development, we performed ultrasound-guided nanoinjections [12] into amniotic fluid to transduce mesoderm and ectoderm with heritable, transcribed genetic TREX barcodes [13, 14] (Fig. 1D). The injections were carried out once or twice serially between days E7.5 and E8.5 (Fig. 1E-F, fig. S1) to capture rapid developmental shifts during embryonic development. Trunk and cranial
100 segments were harvested separately at E13.5 when most definitive cell types manifest, and labeled cells were analyzed by scRNA-seq for linking clonal information to each transcriptome (Fig. 1G-H, fig. S2). We optimized the study design to maximize the fraction of cells captured from each individual clone. To accomplish this, we injected embryos with low amounts of virus (in almost all cases 46-69nL) such that only 10-40k cells would be labeled in each harvested segment – an
105 amount that could be taken through downstream processing steps without significant additional losses.

Our microinjection approach predominantly targets ectoderm [12-14], as well as some mesoderm. This is explained by the fact that the ultrasound guided injections shower the exterior of the early
110 neurula with barcoded lentiviral particles, and the exposed cells residing in outermost positions have the highest chance to become labeled [12]. As expected, in our experiments, we did not recover traced endodermal derivatives. Following transduction of the embryos, retroviral integration in dividing cells takes approximately eight hours [15]. Two days after injection we saw formation of large clones by whole mount analysis of embryos injected with low volumes of viral
115 particles (fig. S3), and broad labeling thereafter (fig. S4), indicating that the tracer system activates swiftly. Thus, this method facilitates rapid, accurate, and broad lineage tracing of early ectodermal cells, as well as some mesodermal cells and axis-extending neuromesodermal progenitors. Single-cell RNA-seq analysis of eight control embryos (4 cranial, 7 trunk sections), recovered a total of 159,744 clonally barcoded cells. Specifically, we recovered 4411 E7.5-traced clones and 6661
120 E8.5-traced clones. Of those, 3442 E7.5-traced clones and 4722 E8.5-traced clones contained at least two cells. These informative clones encompassed 49% of all sequenced barcoded cells.

As expected, multicellular clones were distributed across the major embryonic ectodermal and mesodermal derivatives, including keratinocytes of the skin, various mesenchymal cells, and
125 neuroglial cells of the central and peripheral nervous system (Fig. 1I-J, fig. S5 and fig. S6). Most clones contained similar cells belonging to only one germ layer derivative (Fig. 1K-L; for cranial, see fig. S7). Multipotent clones spanning more than one major lineage were more abundant in the earlier developmental timepoint (1.7% in E8.5 vs. 5.2% in E7.5) indicating that early fate restrictions are already imposed in ectodermal and mesodermal cell lineages prior to neurulation
130 (Fig. 1L). The patterns of clonal co-occurrence grouped similar cell types, recovering the expected consensus cell type lineage phylogeny (Fig. 1M and fig. S8). Reconstruction of such lineage structure from fixed barcode data, however, is known to be challenging [16]. Moreover, the resulting phylogeny would represent a population average of cell fate decision structure, whereas the clones contributing to any given cell type can vary greatly in their overall lineage distribution
135 (Fig. 1B,K). We therefore reasoned that a new computational approach is needed to analyze the processes underlying the substantial diversity in clonal fate distributions.

Clone2vec recapitulates spatiotemporal patterns of multipotency in the embryo

To investigate variability among numerous sparsely sampled clonal populations, we developed a computational method, called clone2vec, which adapts word2vec [17] word embeddings to lineage tracing information. Clone2vec represents individual clones in a low-dimensional space that captures the variation of clonal fates (Fig. 1N-O). In such an embedding, neighboring clones exhibit similar cell fate distributions (Fig. 1P, fig. S9). We characterized the performance of the clone2vec method using simulated data, as well as a published lineage tracing dataset of clonal hematopoiesis [18] (see Supplementary Note). We found that clone2vec embedding effectively reconstructs both discrete clonal types as well as patterns of continuous clonal variation within discrete clusters. Clone2vec re-analysis of the hematopoietic tracing data was able to recover previously described fate patterns and uncovered more subtle continuous axes of variation. Importantly, clone2vec performance remained robust under increasingly sparse sampling of cells.

Unlike other published methods, such as CoSpar [19] and CellRank [20] this algorithm is designed to characterize clones' heterogeneity rather than assess the probabilities of fate selection during differentiation. The cluster-free approach of this method enables a shift away from a cell type-centric perspective, allowing for a nuanced description of complex patterns in clone distribution. Nevertheless, the output of clone2vec — comprising clonal embeddings and clusters of clones — can still be utilized in downstream analyses to investigate molecular determinants underlying differences in clonal behavior.

Clone2vec analysis of our dataset revealed continuous variation of clonal fates (Fig. 2A, see Data Availability for an interactive visualization) within the three major compartments corresponding to skin, neural ectoderm, and mesoderm clonal clusters (Fig. 2B). To illustrate the processes underlying continuous variation in clone development, we examined the genes associated with the first two principal components of clonal variation within the mesenchymal compartment. These included genes encoding transcription factors for specific mesenchymal cell types, progenitor markers, as well as known embryonic positional markers such as Hox genes (Fig. 2C). To validate that the clonal embedding captured the anterior-posterior axis of the embryo, we separately dissected lower and middle trunk prior to sorting, confirming a striking difference in their proportions along the vertical axis of the mesenchymal cluster embedding (Fig. 2D).

The first principal axis of the mesenchymal clonal variation, and the anteroposterior axial position of clones, were both associated with clones sharing both neural and mesenchymal progeny (Fig. 2E). We found that this was driven by derivatives of neuromesodermal progenitors (NMPs), which are a tailbud cell population that contributes to secondary neurulation and somitogenesis during axial extension [11, 21]. Our clonal data show that NMPs give rise to multipotent clones that contribute to both neural and mesenchymal cell types in a position- (Fig. 2D-F) and time- (Fig. 2G) dependent manner, with more posterior clones being more multipotent. To validate this, we serially sectioned littermate embryos double-injected with EGFP (at E7.5) and tdTomato (at E8.5) barcoded lentivirus, analyzing the transduced somite derivatives such as vertebrae and other cells types in every section along the anterior-posterior axis (Fig. 2H-I). The vertebrae were unlabeled in the mid-trunk but labeled in the posterior trunk, while cells in neural structures such as the neural tube and dorsal root ganglia showed both labels throughout the entire body. Labeling of somitic derivatives spread longer along the body axis in E7.5 injected embryos as compared to E8.5

185 injected samples, which demonstrated an expected coupling between the progressive spatial and temporal fate restrictions in the NMP lineage. Interestingly, an adjacent clonal cluster to NMPs featured clones from both E7.5 and E8.5 time points, producing a small number of neural crest cells and somitic derivatives, but not CNS cells (Fig. 2E, J). This shows evidence that NMPs can directly generate neural crest during secondary neurulation without contributing to neural tube
190 epithelialization, demonstrating existence of a neural crest - mesodermal bipotent progenitor (NCC-MP), which has been predicted by earlier work [22, 23], and may be relevant for *spina bifida* pathology.

Position-specific auxiliary transcriptional programs are superimposed on common developmental
195 *paths of cell lineages to regulate clonal composition*

Cell fate determination involves coordinated induction of transcriptional programs and, often, suppression of alternate fates [24]. By delineating sets of clones that demonstrate distinct fate outcomes, we studied how context-dependent transcriptional programs subtly bias multipotency
200 among otherwise similar progenitor cell populations. To illustrate how clonal embedding enables dissection of context-dependent differentiation, we examined a group of trunk mesodermal lineages that build the skeleton with the help of Hox-dependent patterning. In the transcriptional embedding, the mesenchymal populations form a large diffuse cluster (Fig. 1J), containing progenitor as well as early differentiated populations (Fig. 3A). In contrast, the analysis of clonal
205 clusters dissects the mesenchymal population into a set of multiple partially overlapping lineages (Fig. 3B). These clonally linked clusters reveal specific differentiation trajectories within a diffuse population where the standard approaches for trajectory analysis would fail due to uncertainty. The clonal clusters varied in their contributions to skeletal cells, dermis, meningeal fibroblasts, and perivascular cells (Fig. 3A-B and boxed insets). They also varied in the expression of Hox
210 genes (Fig. 2D), suggesting that clonal clusters capture position-specific modulations of the developmental programs necessary for formation of different body parts. We refer to such positional or signaling modulation that introduces variation in the course of otherwise similar clonal development and differentiation process as "auxiliary programs".

215 To understand the factors driving these discrete multipotency profiles, we compared two adjacent chondrogenic lineages – Cluster 8, containing cartilage/meninge clones (that did not engage in osteogenic differentiation, here referred to as 'chondrogenic'), and Cluster 2, containing 'osteochondral' cartilage/bone clones (Fig. 3C). We observed a full spectrum of intermediate cell states at E13.5 due to continuous differentiation from mesenchymal progenitors to mature
220 chondrocytes (Fig. 3D-E), suggesting a bona fide developmental trajectory in both cases.

Mapping cluster-specific cartilage gene expression signatures to publicly available Stereo-seq data suggested that chondrogenic lineage 8 represent the forming vertebrae in the lower trunk (Fig. 3F, fig. S10), while the osteochondral lineage 2 is likely from anterior structures including early bone
225 collar of ribs, which starts forming from perichondrial cells at E13.5 [25, 26]. Focusing on chondrogenic differentiation in both lineages, we found that both executed a conserved transcriptional program, upregulating common chondrocyte gene markers (Fig. 3G and fig. S11A-D) including *Col11a1*, *Sox6*, *Col9a3*, *Acan*, *Comp*, *Fgfr3*, *Col11a2*, and *Matn3*. Surprisingly, expression of bone-related genes (Fig. 3H and fig. S11E-F) increased at the beginning of
230 differentiation in both trajectories, even in the clonal cluster without bone, suggesting existence of

competitive programs underlying fate selection even when one of the fates is absent. Analysis of the bone-related genes with clear early activation in the osteochondral lineage revealed gene expression of the well-known TF *Runx2* in mesenchymal progenitors as the strongest indicator of overall clonal osteogenic potential (Fig. 3I), aligning with existing models [27]. Despite being downregulated in the chondrogenic-only lineage (fig. S11 G-K), *Runx2* expression significantly correlated with the bone-related signature even there (Fig. 3J), suggesting that molecular regulators of mesenchymal *Runx2* maintenance or repression drive clonal skeletal potential in vivo.

Differential expression analysis of the clonal lineages with distinct *Runx2* activation implicated TFs related to the Hox code, mesenchymal fate commitment, and skeletal development (fig. S12A-H), leading us to question the hierarchy of observed TFs. Using SCENIC [28], we inferred transcriptional factor activity and likely gene regulatory networks acting during cartilage development in both clonal clusters (Fig. 3K-L and fig. S12I-J). The analysis suggested an upstream regulatory role of Hox TFs, with *Hoxa10/Hoxd10* (lumbosacral) coordinating the regulatory network in chondrogenic clones, and *Hoxa7/Hoxc8* (thoracic) in osteochondral clones. *Hoxa7* was positioned directly upstream of *Runx2* expression in the reconstructed gene regulatory network (Fig. 3L), offering one potential explanation for the positional differences in clone skeletogenic potential. Additionally, Hoxc8 (fig. S12E) was positioned upstream of mesenchymal TF *Prrx1* (via intermediates *Meis1* and *Pbx3*), which exhibited sustained expression specifically in the bone-forming lineage, suggesting that positional factors might regulate the tempo of early chondrogenesis by maintaining a mesenchymal progenitor status (*Prrx1* expression). These roles are consistent with a known regulatory function of Hox genes in mesenchymal progenitor cell fate decisions [29].

Overall, the combination of two orthogonal techniques – trajectory inference on cell manifolds and lineage tracing [10] — enabled examination of developmental mechanisms used to guide specific realizations of possible cell fates to shape the embryo with high local precision. The example of the two mesodermal clusters illustrates how a core transcriptional program of pro-chondrogenic differentiation can be augmented by an auxiliary program such as the formation of a bone collar in the ribs by perichondrial cells. Such programs can be driven by position-specific or other local signals, modulating exact cell compositions (Fig. 3M). Auxiliary programs might alter the local tempo of cell type differentiation, proliferative potential and other aspects related to the specifics of tissue shaping. Distilling the core and auxiliary cell type differentiation programs should also aid in developing more efficient and parsimonious *ex vivo* differentiation methods needed for regenerative medicine.

Positional codes in facial mesenchyme revealed through clonal analysis.

In contrast to the trunk, the anterior cranial region is largely devoid of Hox expression. The logic of auxiliary programs guiding spatial patterning of the developing face remains insufficiently understood. We applied our clonal variation approach to this problem, focusing on chondrocytes, reasoning that assessing transcriptional variation within a single cell type would accentuate position-specific programs by minimizing cell-type-related artefacts. The cranial dataset captured a rich spectrum of clonal fates (Fig. 4A), covering populations contributing to central and peripheral nervous system, dermis, meninges, and mesenchymal derivatives of the musculoskeletal system which was segregated into the major lineages corresponding the major facial prominences (Fig. 4B-D). Examining the spatial distributions of lineage-specific genes in

Stereo-seq data, we found that they consistently mapped to different facial subregions (Fig. 4C). Differential expression analysis of specific chondrogenic signatures within all three major lineages revealed key transcription factors and signaling molecules, which were sufficient to reconstitute the expected positional identity (Fig. 4E). We then confirmed that these position-specific signatures (including previously reported markers Alx1/Asb4/Hoxa2 [30, 31]) also observed in other cell types, beyond chondrocytes, were originating from the same clones (Fig. 4E). The true anatomical position of these signatures was revealed by the analysis of Stereo-seq data (Fig. 4E). This suggests that these signatures likely represent position-specific programs modulating progenitor cell fate choices and plasticity. Indeed, fate profiles of cranial clone types revealed that Hox-negative BA1 progenitor-derived clones were highly multi-fated, containing fates from nearly all mesenchymal clusters, whereas Hoxa2^{+} clones from BA2 lacked several mesenchymal cell types (Fig. 4D). Overall, these results demonstrate how analysis of continuous clonal variation can suggest candidates for positional programs within the developing embryo.

Cranial neural crest generates neuronal and mesenchymal fates from partly overlapping and yet distinct progenitor pools.

Beyond position-specific effects, continuous clonal variation can also capture temporal dependencies in multipotency regulation. Cranial neural crest cells (CNCCs) emigrate in waves from the dorsal neural fold, but face different microenvironmental contexts depending on their exact timing of delamination, which makes for a good system to study temporal modulation of cell fate choices in vivo [32]. As a population, the cranial neural crest is highly multipotent, giving rise to facial skeletal and other mesenchymal structures as well as to the peripheral nervous system (trigeminal and sympathetic ganglia, including the glial cells), but whether such broad multipotency can be attributed to individual CNCCs remains unclear [33-36]. For instance, each wave of CNCCs might carry specific fate biases; alternatively, multipotent CNCCs would be biased by the post-migratory environment.

Examining clonal variation of CNCC lineages, we were surprised to find that facial mesenchymal derivatives and trigeminal neuroglial cell types were overwhelmingly separated into distinct clonal subpopulations (clusters 6+19 versus clusters 0-1,3-5,7-10,14-16, Fig. 4D) and very few clones shared trigeminal neuronal and mesenchymal fates. Furthermore, these two clonal subpopulations were distinguished by sequential E7.5 + E8.5 double injection, with neuroglial cell types labeled predominantly at later E8.5 injections. This result suggests that early and late-migrating CNCC waves become differentially fate-biased.

To test the temporal separation between mesenchymal and neuroglial trigeminal cell types, we performed lineage tracing experiments in *Sox10-CreERT2/R26Confetti* embryos, which enable clonal tracing based on multicolor labelling of stochastically recombined cell populations (Fig. 4G). Recombination was triggered by tamoxifen injection at slightly different developmental times (~12h differences at E8.5 stage). The resulting clonal tracing showed Confetti clones confined to trigeminal ganglion and outgoing nerves being distinct from the clones labelling the surrounding ectomesenchyme, revealing a systematic difference in the cell type composition depending on the precise timing of recombination. This result is consistent with a temporal model in which different migratory waves of CNCC undergo differential fate biasing (Fig. 4H). This finding demonstrates

how clonal variation can help to resolve subtle temporal effects on lineage separation and fate biasing among nearly uniform progenitor cell populations.

325 *Perturb-TREX: Joint clonal tracing with multiplexed signaling receptor perturbations.*

The programs modulating progenitor fate must be ultimately validated through perturbations. In this regard, the highly orchestrated developmental process poses a challenge, as disruption of pathways critical for cell fate determination, even with conditional genetic perturbations can result
330 in major distortions of the tissue and lead to widespread indirect effects. Mosaic perturbations of individual clones, however, are unlikely to cause such tissue distortions, presenting a valuable opportunity to study the role of genes with strong phenotypes *in vivo*.

To examine the role of extrinsic signals in early trunk development, we combined TREX tracing
335 with multiplexed CRISPR/Cas9 gene editing of individual clones (Fig. 5 and fig. S13) [37]. Taking advantage of detailed scRNA-seq data on differentiating neural crest and other cell lineages, we extended the barcoded TREX vectors with a library of guide RNAs targeting receptors of signals associated with early cell fate bifurcation steps (Fig. 5A) [24]. These multiplexed lentiviral libraries were injected into murine embryos at E7.5 or E8.5 and examined with scRNA-seq at
340 E13.5 (Fig. 5B). While, as expected, none of the embryos showed major developmental abnormalities, significant fate deviations were observed for mosaic perturbations at the clonal level (Fig. 5B-C). Such was the case for perturbations of *Ptch1* - a gene encoding sonic hedgehog (SHH) receptor [38] - where *Ptch1*-crispant clones showed massive shifts in clonal structures both in mesodermal (Fig. 5D) and neuroectodermal derivatives (Fig. 5E). In developing CNS, *Ptch1*-
345 crispant clones were enriched in progenitor populations (Fig. 5E), upregulated a unique gene expression signature (i.e. *Hhip, Fgf15, Ptch2, Plagl1)* related to cell cycle and Hedgehog signaling (Fig. 5F), and dramatically increased in cell count to hundreds of cells per clone (Fig. 5G). Even compared to similar clone types (Fig. 5H), the crispant clones were highly restricted to a progenitor-like signature (compare control clusters Fig. 5I-J to crispant clones in 5D-E). Thus,
350 *Ptch1* editing in neuroectodermal clones results in an accumulation of stalled intermediate progenitors prior to their cell cycle exit and neuronal differentiation, and Ptch1-mediated inhibition of SHH-related signaling is needed for CNS neural cells to mature beyond a transit-amplifying progenitor state.

355 The elimination of *Ptch1* from mesoderm-derived mesenchyme resulted in the formation of a clone type not found in any other analyzed conditions (Fig. 5B,D). This distinct clone type was enriched with vertebral chondrocytes and meningeal cells, while specifically depleted of undifferentiated mesenchyme and dermal fibroblasts (Fig. 5D, compare to control clusters in Fig. 5I). This suggests that restraining Hedgehog signaling is important for maintaining a pool of mesenchymal
360 progenitor cells. Other *Ptch1* crispant clones showed overproduction of duct mesenchyme (fig. S13) and increased clone size (Fig. 5G). Overall, the *Ptch1* perturbation caused mispatterning and potential ventralization of perturbed cells within early mesoderm. The ventralization-like phenotype is interesting as the clones with perturbed *Ptch1* become biased to the somitic fates forming predominantly in the ventral part of the somite. The mechanisms underlying the lineage-
365 specific interactions of *Ptch1* with cell differentiation programs can be further studied using clonal comparisons (as in Fig. 3) and combinatorial perturbations along a time series.

Principles of cell fate decisions in neural crest revealed via clonal analysis.

370

Though *Ptch1* editing resulted in re-biasing of progenitors or striking disruptions of the lineage progression due to gain-of-function in the Hedgehog pathway, targeting other receptors might lead to tuning clonal compositions in a more subtle manner. We focused on the trunk neural crest, which has been hotly debated as being either a mixed population of restricted progenitors, or a population
375 of highly multipotent stem cells [39-42]. Clonal variation analysis of wildtype embryos revealed that all detected neural crest sub-lineages were highly fate biased. This was observed for sympathetic, melanocyte, somatosensory, enteric, and glial progenies (Fig. 6A-C). This result suggests that prior to clonal expansion, individual labeled neural crest progenitor cells acquire strong fate biases, corresponding to those previously ascribed to entire migratory waves of neural
380 crest [41, 43]. Such progenitor differences may be too subtle to be observed from transcriptional analysis alone [44]. For instance, we find that the progenitor state of the clones destined to give rise to the *Phox2b*$^{+}$ autonomic nervous system and enteric nervous system within the developing gut, both map to the same overall transcriptional state (Fig. 6D-E). The knowledge of clonal identity enables a more powerful way of identifying biasing programs by directly contrasting the
385 progenitors with distinct fates. In this way we find that these early progenitors show robust upregulation of a set of biasing genes, including *Hand2* - a transcription factor known to be involved in neural crest enteric differentiation [45]. Since such genes reflect early progenitor biasing mechanisms that appear prior to overt differentiation, their modulation maybe helpful for cell type generation protocols and applications intended to cure autonomic and enteric
390 dysfunctions, including Hirschsprung's disease [46].

We also examined positional distribution of the neural crest multipotency along the anterior-posterior and dorsoventral axes. Based on the *Hox* gene expression we find that along anterior-posterior axis an expected strong anterior bias of the enteric-biased clones [47], as well as paucity
395 of sympathetic clones the posterior sacral and post-sacral part of the body (Fig. 6F). The remaining clonal clusters showed broad distributions. To infer the distributions along the dorsoventral position, we examined the co-occurrence of neural crest and neuronal CNS fates within same clones. Since neural crest is formed in the most dorsal part of neural tube, increasing contribution of clones to CNS neurons would suggest a more ventral position - closer to the border between the
400 prospective crest domain and the future roof plate of the neural tube [48]. We find that that the clones biased towards somatosensory neurons and melanocytes shared much higher proportion of progeny with CNS neurons, suggesting that the originally transduced progenitor cells were positioned more ventrally (Fig. 6G). These results are consistent with previous spatiotemporal fate mapping of the dorsal neural tube in chick embryos [41].

405

Finally, to better understand the signals biasing neural crest clones, we examined the compositional changes associated with CRISPR/Cas9 perturbations of different receptors genes. We found significant shifts in clonal composition that can be described as a re-balancing of the autonomic vs somatosensory progenies in the neural crest lineage, suggesting that the corresponding receptors
410 and signaling pathways are indeed involved into defining proportions of fate outcomes (Fig. 6H-I). For instance, *Fzd2, Lrp2*, *Lgr4*-crispant clones had reductions in autonomic and enteric biases, with rebalanced increases in sensory and glial fates. Perturbation of *Fgfr1* showed exactly the opposite effects on clonal biases and composition. These results expand our understanding of the

role of signaling receptors in neural crest specification by demonstrating the existence of a sensory-autonomic clonal lineage bifurcation [24, 49] and its modulation by putative FGF and WNT signaling molecules. Together, the shifts observed in the clonal structures of neuroectodermal as well as mesodermal derivatives upon receptor perturbations demonstrate their role in biasing the clonal derivative spectra of individual multipotent progenitors towards specific fate outcomes.

Discussion

Context-dependent regulation of cell fates is a fundamental aspect of embryogenesis, essential for the proper formation of tissues. Tracing of clonal structures can reveal the spectrum of such fate modulation and provide important clues about the mechanisms involved. Embryogenesis, however, presents a particular challenge for the modern lineage tracing approaches, since the large number of generated cell types and the diverse position-specific nature of modulation demands both increased resolution and scale of tracing. A resolution that would enable biological discoveries has been difficult to achieve on an embryo scale, leading high-throughput tracing studies to focus on mostly focus on specific cell lineages such as hematopoiesis or brain development [13, 18].

Here, we presented an effective practical approach for studying mechanisms underlying fate modulation in an embryo, by integrating multiplexed *in vivo* tracing, computational analysis of transcriptional programs associated with clonal fate biases, and mosaic medium-throughput multiplexed perturbations. The scale of the study, in this case profiling thousands of individually barcoded ectodermal and mesodermal clones across developmental timepoints, is one enabling aspect.

In the past, low-throughput clonal lineage tracing analysis has been used extensively to successfully address hypotheses about cell multipotency and differentiation [40]. Although these studies have established the classic paradigms about lineage splits in embryogenesis, they were not sufficiently powered to investigate the continual variation of clonal fates. Even though different developing structures are often made from the same cell types [50, 51], these structures acquire distinct shapes and functions due to the fine regulation of cell fate proportions and spatial positioning of similar cell progenitor populations [52, 53], developing within the same cell lineages [54]. Insights into the continual variation of clonal structures help to explain how these morphogenesis-related differences gradually arise in a coordinated and context-aware manner.

Even state-of-the-art high-throughput lineage tracing methods still sample only a small fraction of cells in the embryo. The sparseness of the resulting clonal data represents a major challenge for the subsequent analysis. This has prompted us to develop a machine learning approach – clone2vec – which represents the second enabling aspect of our study. Starting with a cluster-free representation of expression state, clone2vec embeds sparse clonal data into a low dimensional space where clones with similar or compatible fate profiles are grouped together. This approach, as we show, is able to recover complex clonal fate patterns in a manner highly robust to dropouts (Supplementary Note), and also permits generation of new hypotheses. In addition to separating distinct progenitor lineages (*e.g.* mesenchymal and neural), clone2vec embeddings capture continuous clonal variation within these lineages, which is much less understood and is a central point of our study.

In contrast to tree representations, clonal embeddings enable analysis of gradual variation of clonal fates and inference of regulatory processes modulating clonal fates and proportions of cell types. In particular, we expect clones progressing within different spatial or temporal contexts have distinct but overlapping regulatory and transcriptional signatures [55]. A good example is the population of mesenchymal mesodermal derivatives which forms a large unstructured cluster of cells in the transcriptional embedding, posing a challenge for the traditional methods of trajectory reconstruction and other analysis. Lineage tracing and clone2vec analysis, however, effectively partitions this population into a series of clonally connected sub-clusters, representing variation of the fundamental differentiation process at different positional or temporal contexts (Fig. 3A). Furthermore, the analysis of transcriptional variation between clonal clusters can reveal more subtle "auxiliary programs" that regulate clonal structures. For instance, we showed an example of a positional-associated variation in skeletogenic mesenchyme, where additional programs do not appear to be essential for chondrogenic differentiation, but rather modulate secondary features such as a possibility of transiting into osteogenic fate or maturation dynamics of progenitor cells. Such approach should be useful for analyzing position-dependent modulation of skeletogenesis and progenitor potency in general in other contexts [55, 56].

In addition to resolving spatial modulation of clonal development, our analysis also highlighted some temporal variation in fate restrictions. In particular, our comparison of cranial neural crest cells tracing initiated at two successive developmental timepoints showed unanticipated temporal segregation of facial skeletogenic mesenchyme and trigeminal ganglion from the same progenitor population. This result shows that temporal windows shaping the potency of the cranial neural crest are essential for building the integrated head, where neural crest gives rise to the structures as different as peripheral innervation and facial bones [57]. Such separation of fate selection in time, as opposed to purely positional separation, offers advantages in craniofacial morphogenesis by adding a dynamic layer of control that complements positional cues [41, 58, 59]. Overall, the use of static barcodes introduced at precise developmental stages enabled comparison of clonal populations to find molecular drivers underlying progenitor cell properties. This can be contrasted with other methods employing dynamic barcodes [7], in which the timing of lineage bifurcations can be difficult to pinpoint, complicating efforts to link lineage decisions to the corresponding signals and developmental stages [8].

The coupling of clonal tracing with mosaic perturbations provides powerful means of validating mechanistic and causative relationships that underlie cell fate regulation in various spatiotemporal niches. Our focus in this study was on perturbations of extrinsic signaling axes, as prior studies and our own correlative analysis of natural clonal patterns indicated that they play a major role in modulating auxiliary programs and guiding progenitor fates. Indeed, we demonstrate that perturbations of specific extrinsic signals alter positional biases in the neural crest lineage. Furthermore, disrupting certain signaling axes, such as SHH (via *Ptch1*), results in formation of novel clone types with a progeny spectrum that was not encountered in wildtype embryos, suggesting a new opportunity to study plasticity and robustness of development. Since mosaic perturbations impact only a very small fraction of cells in the tissue, this approach can be used to evaluate the impact of signaling axes that would otherwise severely distort the tissue or lead to unviable states. In this way, mosaic multiplexed perturbations offer an effective strategy to study cell fate regulation in native tissue environments under both normal and diseased conditions.

In addition to insight into signaling landscapes and positional cues, high-throughput analysis of clonal biasing in the mammalian embryo uncovered previously unrecognized lineage relationships or fate restrictions. For instance, that NMPs – multipotent mesoderm-neuro competent cells [11, 21] - coexist in the tail region with mesoderm-restricted progenitors. Our results yielded transcriptional signatures that separate the E13.5 derivatives of NMPs from that of their fate-restricted neighbors (*i.e. Ran*, *Hoxb13*) that could be helpful in determining whether a cellular memory of NMP multipotency exists. This study also identified a population of murine NMPs that contribute neural crest cells but not spinal cord neurons (so-called NCC-MPs) [23]. While our results provide a gene signature to distinguish them (*i.e. Tspan18, Pde8a*), further studies will be needed to determine if neural crest cells derived from NCC-MPs bear functional differences from other sacral crest lineages.

On a larger developmental scale, our analysis reveals a picture of global fate restrictions at the stages of late gastrula and early neurula, which altogether opposes the idea of early plasticity of different germ layers as was suggested from classic experiments in vitro [60]. According to our data, the vast majority of clones does not span the competence of more than a single germ layer, albeit the small number of specific exceptions pertaining to cranial neural crest and NPMs. This does not rule out that cells from these stages might exhibit multipotency when subjected to different environments, as in the case of transplantation experiments that can reveal greater cell plasticity (as discussed by Zechner in [1]).

Overall, we hope the presented approach will be helpful for further studies aimed at systematically dissecting the spatial and temporal correlates of clonal variation in different cell systems. The current clonal atlas, including its interactive interface (browsable online at https://clones2cells.streamlit.app) should provide a useful resource for a wide range of questions related to the formation of boundaries between developing embryonic structures [61]. Similarly, clone2vec analysis should aid in the identification of fine-grained positional and temporal codes. While the Hox code and other crude positional systems provide a general framework for development of body axes and programming the large body domains [62], the fine-grained signals that guide specific cell behaviors within very local tissues are much harder to study due to the complexity and redundancy of embryonic signaling systems [63]. Our clone2vec approach can assist investigations of how fine-grained positional information informs clonal diversity within individual organs like the heart or lungs. This might shed light on other unresolved questions, for instance, the enigma of the anterior-posterior allocation of brain nuclei [64], or fine patterning of mesodermal populations building individual muscles [65]. On the medical side, the ability to combine clone2vec clonal analysis with mosaic perturbations enables multiplexing simulated conditions that mimic human genetic abnormalities, including neurodevelopmental disorders and pediatric cancers (i.e. neuroblastoma). This might help revealing clonal compensatory behavior in modelled congenital diseases, as well as in regeneration, potentially revealing clonal mechanisms of embryonic resilience [66].

Despite the potential, we realize the limitations of the study. The presented multiplexed tracing and perturbation approach can be widely applied for focused analysis of specific tissues and compartments, however, the potential to refine these techniques is substantial. Current limitations include the constrained developmental time frames and specific cell types that can be effectively

targeted using existing barcoding techniques. Inference of differentiation paths or the state of progenitor cells based on the measured end point poses well-known fundamental challenges, particularly when the underlying process is asynchronous and one cannot observe, for example, both progenitor and differentiated cells from the same clone in the snapshot provided by scRNA-seq data. While we show that successive barcoding rounds can help, incorporation of dynamic barcoding strategies [5], such as CRISPR scarring [6-8, 67] could improve path reconstruction, though not progenitor state inference.

References and Notes:

1. Zechner, C., E. Nerli, and C. Norden, *Stochasticity and determinism in cell fate decisions.* Development, 2020. **147**(14).
2. Kretzschmar, K. and F.M. Watt, *Lineage tracing.* Cell, 2012. **148**(1-2): p. 33-45.
3. Livet, J., et al., *Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system.* Nature, 2007. **450**(7166): p. 56-62.
4. McKenna, A., et al., *Whole-organism lineage tracing by combinatorial and cumulative genome editing.* Science (New York, N.Y.), 2016. **353**(6298): p. aaf7907-aaf7907.
5. Pei, W., et al., *Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding.* Cell Stem Cell, 2020. **27**(3): p. 383-395 e8.
6. Bowling, S., et al., *An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells.* Cell, 2020. **181**(6): p. 1410-1422 e27.
7. Li, L., et al., *A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells.* Cell, 2023. **186**(23): p. 5183-5199 e22.
8. Spanjaard, B., et al., *Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars.* Nature Biotechnology, 2018. **36**(5): p. 469-473.
9. Raj, B., et al., *Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain.* Nature Biotechnology, 2018. **36**(5): p. 442-450.
10. Wagner, D.E. and A.M. Klein, *Lineage tracing meets single-cell omics: opportunities and challenges.* Nature Reviews Genetics, 2020. **21**(7): p. 410-427.
11. Tzouanacou, E., et al., *Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis.* Dev Cell, 2009. **17**(3): p. 365-76.
12. Mangold, K., et al., *Highly efficient manipulation of nervous system gene expression with NEPTUNE.* Cell Rep Methods, 2021. **1**(4).
13. Ratz, M., et al., *Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics.* Nature Neuroscience, 2022. **25**(3): p. 285-294.
14. Kameneva, P., et al., *Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin.* Nat Genet, 2021.
15. Roe, T., et al., *Integration of murine leukemia virus DNA depends on mitosis.* EMBO J, 1993. **12**(5): p. 2099-108.
16. Weinreb, C. and A.M. Klein, *Lineage reconstruction from clonal correlations.* Proc Natl Acad Sci U S A, 2020. **117**(29): p. 17041-17048.

17. Mikolov, T., et al., *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781, 2013.
18. Weinreb, C., et al., *Lineage tracing on transcriptional landscapes links state to fate during differentiation.* Science, 2020. **367**(6479).
19. Wang, S.W., et al., *CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information.* Nat Biotechnol, 2022. **40**(7): p. 1066-1074.
20. Weiler, P., et al., *CellRank 2: unified fate mapping in multiview single-cell data.* Nat Methods, 2024. **21**(7): p. 1196-1205.
21. Henrique, D., et al., *Neuromesodermal progenitors and the making of the spinal cord.* Development, 2015. **142**(17): p. 2864-2875.
22. Frith, T.J., et al., *Human axial progenitors generate trunk neural crest cells in vitro.* Elife, 2018. **7**.
23. Lukoseviciute, M., S. Mayes, and T. Sauka-Spengler, *Neuromesodermal progenitor origin of trunk neural crest in vivo.* bioRxiv, 2021: p. 2021.02.10.430513.
24. Soldatov, R., et al., *Spatiotemporal structure of cell fate decisions in murine neural crest.* Science, 2019. **364**(6444).
25. Flaherty, K. and J.T. Richtsmeier, *It's about Time: Ossification Center Formation in C57BL/6 Mice from E12–E16.* J Dev Biol, 2018. **6**(4).
26. Maes, C., et al., *Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels.* Dev Cell, 2010. **19**(2): p. 329-44.
27. Loebel, C., et al., *In vitro osteogenic potential of human mesenchymal stem cells is predicted by Runx2/Sox9 ratio.* Tissue Eng Part A, 2015. **21**(1-2): p. 115-23.
28. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering.* Nature Methods, 2017. **14**(11): p. 1083-1086.
29. Bradaschia-Correa, V., et al., *Hox gene expression determines cell fate of adult periosteal stem/progenitor cells.* Sci Rep, 2019. **9**(1): p. 5043.
30. Gu, R., et al., *Single-cell transcriptomic signatures and gene regulatory networks modulated by Wls in mammalian midline facial formation and clefts.* Development, 2022. **149**(14).
31. Kessler, S., et al., *A multiple super-enhancer region establishes inter-TAD interactions and controls Hoxa function in cranial neural crest.* Nat Commun, 2023. **14**(1): p. 3242.
32. Krispin, S., et al., *Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest.* Development, 2010. **137**(4): p. 585-595.
33. Kaucka, M., et al., *Analysis of neural crest-derived clones reveals novel aspects of facial development.* Sci Adv, 2016. **2**(8): p. e1600060.
34. Dupin, E., et al., *The issue of the multipotency of the neural crest cells.* Dev Biol, 2018. **444 Suppl 1**: p. S47-S59.
35. Weston, J.A. and J.P. Thiery, *Pentimento: Neural Crest and the origin of mesectoderm.* Developmental Biology, 2015. **401**(1): p. 37-61.
36. Baker, C.V., et al., *Early- and late-migrating cranial neural crest cell populations have equivalent developmental potential in vivo.* Development, 1997. **124**(16): p. 3077-3087.
37. Platt, Randall J., et al., *CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling.* Cell, 2014. **159**(2): p. 440-455.
38. Cohen, M., et al., *Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms.* Nature Communications, 2015. **6**(1): p. 6709.

39. Bronner-Fraser, M. and S.E. Fraser, *Cell lineage analysis reveals multipotency of some avian neural crest cells.* Nature, 1988. **335**(6186): p. 161-4.
40. Baggiolini, A., et al., *Premigratory and migratory neural crest cells are multipotent in vivo.* Cell Stem Cell, 2015. **16**(3): p. 314-22.
41. Krispin, S., et al., *Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest.* Development, 2010. **137**(4): p. 585-95.
42. Erickson, A.G., P. Kameneva, and I. Adameyko, *The transcriptional portraits of the neural crest at the individual cell level.* Semin Cell Dev Biol, 2023. **138**: p. 68-80.
43. Tang, W., et al., *Clonal analysis and dynamic imaging identify multipotency of individual Gallus gallus caudal hindbrain neural crest cells toward cardiac and enteric fates.* Nat Commun, 2021. **12**(1): p. 1894.
44. Yu, V.W.C., et al., *Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells.* Cell, 2017. **168**(5): p. 944-945.
45. D'Autréaux, F., et al., *Hand2 is necessary for terminal differentiation of enteric neurons from crest-derived precursors but not for their migration into the gut or for formation of glia.* Development, 2007. **134**(12): p. 2237-2249.
46. Montalva, L., et al., *Hirschsprung disease.* Nature Reviews Disease Primers, 2023. **9**(1): p. 54.
47. Kam, M.K.M. and V.C.H. Lui, *Roles of Hoxb5 in the development of vagal and trunk neural crest cells.* Development, Growth & Differentiation, 2015. **57**(2): p. 158-168.
48. Krispin, S., E. Nitzan, and C. Kalcheim, *The dorsal neural tube: a dynamic setting for cell fate decisions.* Dev Neurobiol, 2010. **70**(12): p. 796-812.
49. Sutton, G., R.N. Kelsh, and S. Scholpp, *Review: The Role of Wnt/β-Catenin Signalling in Neural Crest Development in Zebrafish.* Frontiers in Cell and Developmental Biology, 2021. **9**.
50. Noden, D.M. and P.A. Trainor, *Relations and interactions between cranial mesoderm and neural crest populations.* Journal of Anatomy, 2005. **207**(5): p. 575-601.
51. Wang, M., et al., *Distinct Gene Regulatory Dynamics Drive Skeletogenic Cell Fate Convergence During Vertebrate Embryogenesis.* bioRxiv, 2024: p. 2024.03.26.586769.
52. Kaucka, M., et al., *Oriented clonal cell dynamics enables accurate growth and shaping of vertebrate cartilage.* eLife, 2017. **6**: p. e25902.
53. Kaucka, M., et al., *Signals from the brain and olfactory epithelium control shaping of the mammalian nasal capsule cartilage.* eLife, 2018. **7**: p. e34465.
54. Hirasawa, T. and S. Kuratani, *Evolution of the vertebrate skeleton: morphology, embryology, and development.* Zoological Lett, 2015. **1**: p. 2.
55. Bastide, S., et al., *TATTOO-seq delineates spatial and cell type–specific regulatory programs in the developing limb.* Science Advances. **8**(50): p. eadd0695.
56. Otsuki, L. and E.M. Tanaka, *Positional Memory in Vertebrate Regeneration: A Century's Insights from the Salamander Limb.* Cold Spring Harb Perspect Biol, 2022. **14**(6).
57. Adameyko, I. and K. Fried, *The Nervous System Orchestrates and Integrates Craniofacial Development: A Review.* Frontiers in Physiology, 2016. **7**(49).
58. Usui, K. and M. Tokita, *Creating diversity in mammalian facial morphology: a review of potential developmental mechanisms.* EvoDevo, 2018. **9**(1): p. 15.
59. Chen, Y.C. and N. Konstantinides, *Integration of Spatial and Temporal Patterning in the Invertebrate and Vertebrate Nervous System.* Front Neurosci, 2022. **16**: p. 854422.

60. Waddington, C.H. and J. Taylor, *Conversion of Presumptive Ectoderm to Mesoderm in the Chick.* Journal of Experimental Biology, 1937. **14**(3): p. 335-339.
61. Dahmann, C., A.C. Oates, and M. Brand, *Boundary formation and maintenance in tissue development.* Nature Reviews Genetics, 2011. **12**(1): p. 43-55.
62. Carrasco, A.E. and S.L. López, *The puzzle of Hox genes.* Int J Dev Biol, 1994. **38**(4): p. 558-64.
63. Tkačik, G. and T. Gregor, *The many bits of positional information.* Development, 2021. **148**(2): p. dev176065.
64. Lipovsek, M. and R.J. Wingate, *Conserved and divergent development of brainstem vestibular and auditory nuclei.* Elife, 2018. **7**.
65. Grimaldi, A. and S. Tajbakhsh, *Diversity in cranial muscles: Origins and developmental programs.* Curr Opin Cell Biol, 2021. **73**: p. 110-116.
66. Solini, G.E., et al., *Xenopus embryos show a compensatory response following perturbation of the Notch signaling pathway.* Developmental Biology, 2020. **460**(2): p. 99-107.
67. Alemany, A., et al., *Whole-organism clone tracing using single-cell sequencing.* Nature, 2018. **556**(7699): p. 108-112.
68. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data.* Cell Systems, 2019. **8**(4): p. 281-291.e9.
69. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.
70. Ahlmann-Eltze, C. and W. Huber, *Comparison of transformations for single-cell RNA-seq data.* Nat Methods, 2023. **20**(5): p. 665-672.
71. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony.* Nature Methods, 2019. **16**(12): p. 1289-1296.
72. Traag, V.A., L. Waltman, and N.J. van Eck, *From Louvain to Leiden: guaranteeing well-connected communities.* Scientific Reports, 2019. **9**(1): p. 5233.
73. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction.* arXiv preprint arXiv:1802.03426, 2018.
74. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat Biotechnol, 2015. **33**(5): p. 495-502.
75. Kang, J.B., et al., *Efficient and precise single-cell reference atlas mapping with Symphony.* Nat Commun, 2021. **12**(1): p. 5890.
76. Faure, L., et al., *scFates: a scalable python package for advanced pseudotime and bifurcation analysis from single-cell data.* Bioinformatics, 2023. **39**(1).
77. Haghverdi, L., et al., *Diffusion pseudotime robustly reconstructs lineage branching.* Nat Methods, 2016. **13**(10): p. 845-8.
78. Alpert, A., et al., *Alignment of single-cell trajectories to compare cellular expression dynamics.* Nat Methods, 2018. **15**(4): p. 267-270.
79. Giorgino, T., *Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.* Journal of Statistical Software, 2009. **31**(7): p. 1 - 24.
80. Chen, A., et al., *Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays.* Cell, 2022. **185**(10): p. 1777-1792.e21.
81. Büttner, M., et al., *scCODA is a Bayesian model for compositional single-cell data analysis.* Nat Commun, 2021. **12**(1): p. 6876.

Acknowledgements:
We thank Olga Kharchenko for the lovely illustrations, and acknowledge Ruslan Soldatov for the insightful discussions. We thank Adrian Martinez Martin and Polina Kameneva for the technical assistance. We acknowledge the National Genomics Infrastructure, Biomedicum Flow Core, Biomedicum Imaging Core, and Comparative Medicine for access to their facilities and services during the project.

Funding:
National Institute of Dental and Craniofacial Research via the Ruth L. Kirschstein National Research Service Award F32DE029662 (AGE).
ERC Synergy grant KILL-OR-DIFFERENTIATE (IA, SE, AA)
Swedish Research Council (IA, AGE, EA, JF)
Paradifference Foundation (IA)
Bertil Hallsten Research Foundation (IA)
Cancer Foundation in Sweden (IA)
Knut and Alice Wallenberg Foundation (EA, IA, AGE)
Austrian Science Fund (IA, SE, AA)
NSF-14-532-CAREER grant (PVK)

Author contributions:
Conceptualization: IA, PVK, AGE, SI
Investigation: IA, PVK, AGE, SI, MR, JF, EA, BS, JH, JF
Validation: AGE, SI, AM, BS, EA, JH
Computational investigation and analysis: SI, AGE, MR, PVK
Writing – original draft: IA, PVK, AGE, SI, EA, AC, BS, JF, MR, AM
Funding and resources acquisition: IA, EA, AGE, JF, PVK
Supervision and project administration: IA, EA, AGE, JF, PVK

Competing Interests:
The authors declare that they have no competing interests.

Data and materials availability:
All code that was used in the current study is readily available on Github: (https://github.com/kharchenkolab/clonal-atlas-paper/). The clone2vec Python package is also available on Github: (https://github.com/kharchenkolab/scLiTr). Additional documentation for the scLiTr package can be found here: (https://sclitr.readthedocs.io/). The modified TREX pipeline is available here: (https://github.com/serjisa/TREX.modified). Also, the code for the clones2cells app is available via this link: (https://github.com/serjisa/clones2cells_app). Raw data has been submitted to the NCBI Gene Expression Omnibus database (GEO) and is accessible via the following link: (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE269395). Data has also been uploaded to Zenodo: (https://zenodo.org/records/11476029). The data can be browsed via the clones2cells app via this link: (https://clones2cells.streamlit.app). Alternatively, browse the datasets on our CELLXGENE single-cell dataset repository: (https://adameykolab.hifo.meduniwien.ac.at/cellxgene_public/filecrawl/2024_Erickson_Isaev)

Supplementary Materials:

Materials and Methods

Supplementary Note

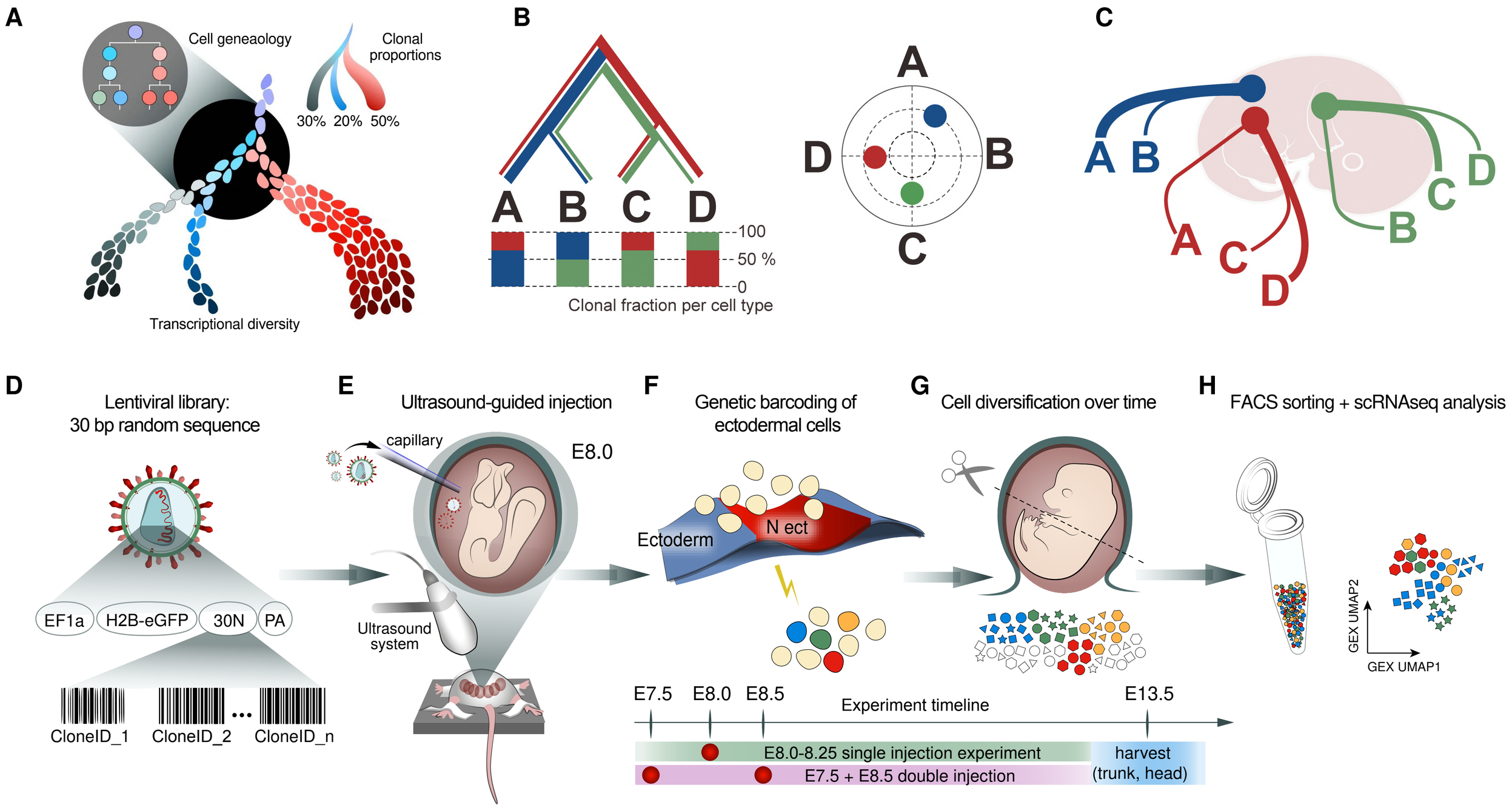
Figs. S1 to S14

References (*68–81*)

Figure 1

Clonal composition is driven by proliferation and differentiation

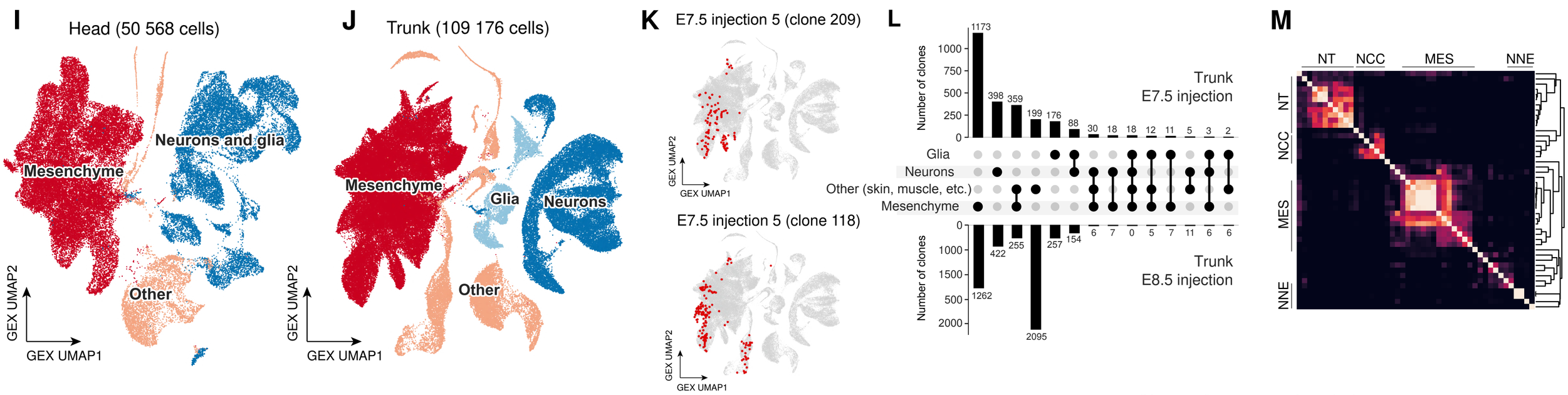
Clonal superposition drives overall lineage tree construction



Single-cell transcriptomics analysis of barcoded body regions

Frequency of clonal behaviors observed in the dataset

Cell type lineage correlation



clone2vec framework

Clones from one clonal cluster occupy the same transcriptional niches

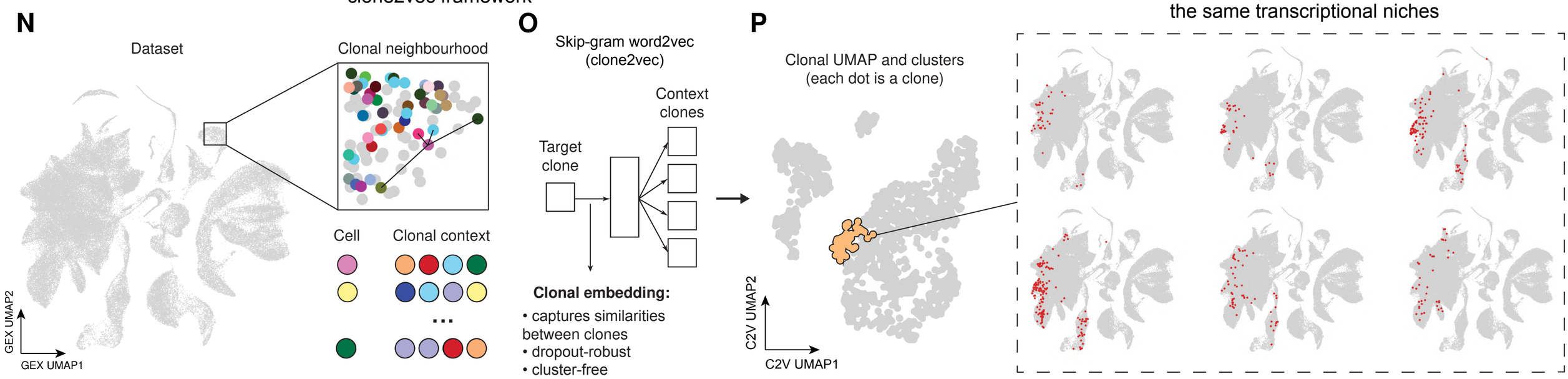


Figure 1. Machine learning-based profiling of clonal behavior in mouse embryos.
(A) Scheme of a single clone developing over time via mitotic divisions and cell maturation, determining the final clonal composition. **(B)** Left, collective multipotency model, where partially overlapping fate sub-spectrums, from populations of clones, generate a fate spectrum (representing the repertoire of lineage-connected cell type combinations). Right, radar plot of the three clones according to their clonal fate composition. The position of the dots indicates the bias of clones towards specific fates (such as cell type "A" and "B"). **(C)** Anatomical position of the theoretical clones shown in (B). **(D)** Clonal barcoding system. **(E)** Experimental setup. Ultrasound is used to visualize the nanoinjector and embryo while injecting lentivirus to the amniotic cavity, once or twice, between the stages of E7.5-E8.5. **(F)** Infection of the exposed virus-accessible cells in neural and non-neural ectoderm (uninfected cells beige, infected cells colorful). **(G)** Clones are allowed to develop until E13.5 when tissue is harvested from head and trunk body regions (each clone color-coded; each cell type shaped differently). **(H)** Barcoded cells are analyzed by single cell transcriptomics for clone reconstruction. **(I-J)** UMAP embedding of the barcoded mouse embryonic E13.5 cells from head (I) and trunk (J), color coded by the major cell types retrieved. **(K)** Examples of individual clones with partially overlapping compositions. **(L)** Upset plot of major types of trunk clones resulting from injection at E7.5 (top) versus E8.5 (bottom). **(M)** Pearson's correlation matrix based on clone sharing pairwise across all retrieved cell types. **(N)** Schematic of clonal context neighborhood. **(O)** Skip-gram production from clonal neighborhoods using word2vec architecture. **(P)** Example of a clonal cluster within the resulting clone2vec UMAP embedding. (right) Example of similar clones constituting a clonal cluster.

Figure 2

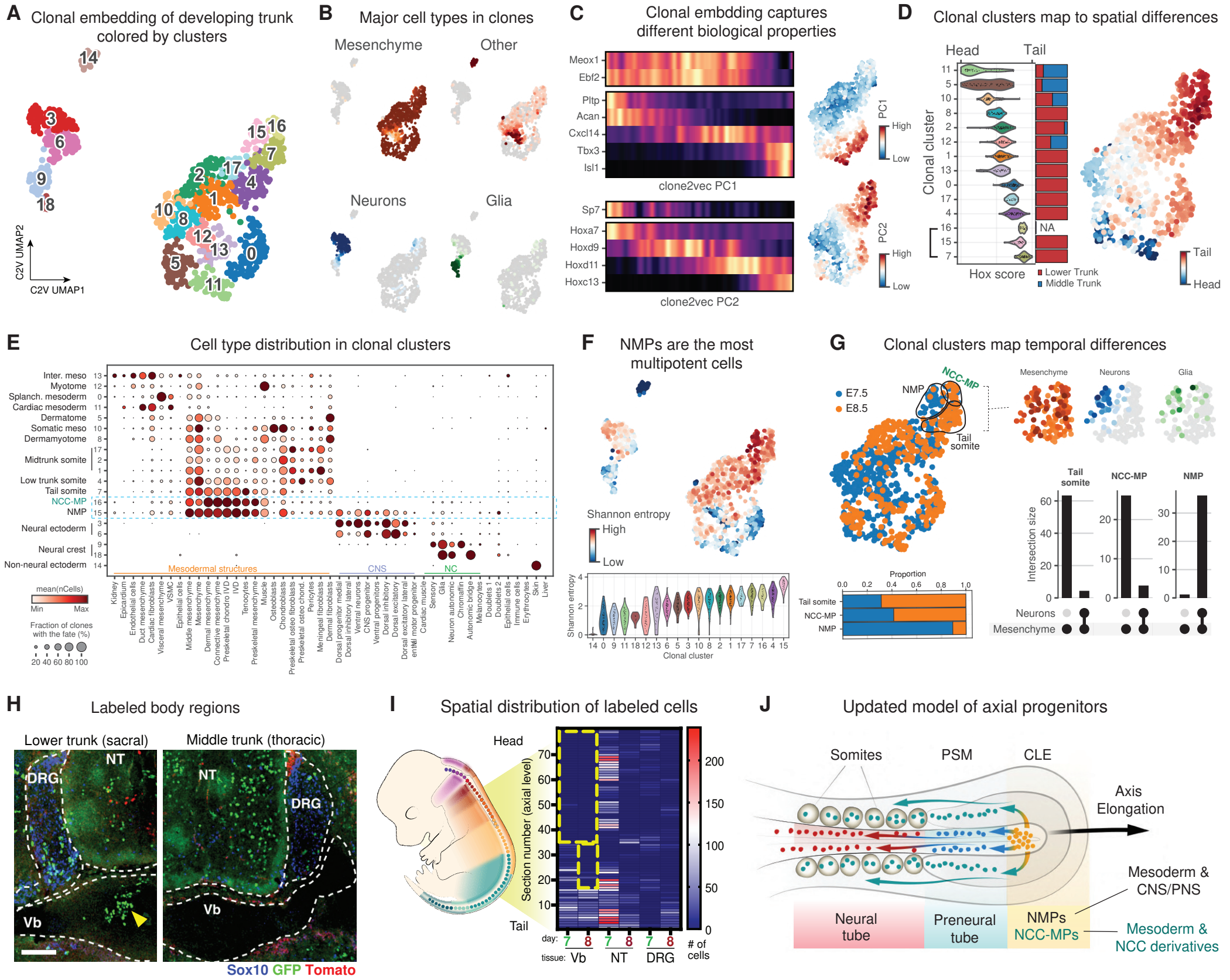

A
Clonal embedding of developing trunk colored by clusters
C2V UMAP2
C2V UMAP1
B
Major cell types in clones
Mesenchyme
Other
Neurons
Glia
C
Clonal embdding captures different biological properties
Meox1
Ebf2
Pltp
Acan
Cxcl14
Tbx3
Isl1
clone2vec PC1
Sp7
Hoxa7
Hoxd9
Hoxd11
Hoxc13
clone2vec PC2
PC1
PC2
High
Low
D
Clonal clusters map to spatial differences
Head
Tail
Clonal cluster
Hox score
NA
Lower Trunk
Middle Trunk
Tail
Head
E
Cell type distribution in clonal clusters
Inter. meso
Myotome
Splanch. mesoderm
Cardiac mesoderm
Dermatome
Somatic meso
Dermamyotome
Midtrunk somite
Low trunk somite
Tail somite
NCC-MP
NMP
Neural ectoderm
Neural crest
Non-neural ectoderm
Mesodermal structures
CNS
NC
mean(nCells)
Min
Max
Fraction of clones with the fate (%)
20 40 60 80 100
F
NMPs are the most multipotent cells
Shannon entropy
High
Low
Clonal cluster
G
Clonal clusters map temporal differences
E7.5
E8.5
NMP
NCC-MP
Tail somite
Mesenchyme
Neurons
Glia
Proportion
Intersection size
H
Labeled body regions
Lower trunk (sacral)
Middle trunk (thoracic)
NT
DRG
Vb
Sox10 GFP Tomato
I
Spatial distribution of labeled cells
Head
Tail
Section number (axial level)
day:
tissue:
Vb
NT
DRG
# of cells
J
Updated model of axial progenitors
Somites
PSM
CLE
Axis Elongation
Mesoderm & CNS/PNS
Mesoderm & NCC derivatives
Neural tube
Preneural tube
NMPs
NCC-MPs

Figure 2. Clone2vec captures spatiotemporal patterns of multipotency in the embryo.
(A-G) Analysis of the Clone2vec UMAP embedding of E13.5 barcoded trunk cells. **(A)** Embedding color-coded by clonal clusters (clustering resolution 2). **(B)** Feature plot color-coded by major cell types. **(C)** Heatmap showing the distribution of genes representing the top principal components of variation across mesenchymal clones in clone2vec space. **(D)** Left, distribution of each clonal cluster Hox gene expression profile, plotted per clone against the known anteroposterior position of embryonic Hox domains (x-axis) and compared to ground-truth manually dissected body regions (horizontal bar plot). Right, feature plot of each mesenchymal clone's Hox score. **(E)** Dot plot showing cell type composition among clonal clusters. Color represents average cells in a fate per clone within that cluster; dot size represents the fraction of the clones in the cluster containing that fate. **(F)** Top, feature plot showing the number of fates per clone. Bottom, violin plot showing the distribution of Shannon entropy per clone within each clonal cluster. **(G)** Top left, mesenchyme embedding color-coded by injection time. Insets highlight the multipotent neuronal-mesenchymal clones. Bottom left, the composition of each clonal cluster by each injection time. Bottom right, upset plot quantifying the number of multipotent versus restricted neuronal/mesenchymal clones per clonal cluster. **(H)** Confocal micrographs of a dual injected (GFP, at E7.5; Tomato, at E8.5) E13.5 embryo at sacral (left) and thoracic (right) axial levels, stained for Sox10 (blue), GFP (green) and Tomato (red). Scale bar = 100 µm. **(I)** Heat map displaying several labeled GFP^+ and $Tomato^+$ cells within the vertebra, neural tube, and sensory ganglia per section, ordered by axial position. **(J)** Schematic of axial extension model featuring canonical as well as NCC-biased NMP clones (termed as NCC-MP, or "neural crest cell and mesodermal progenitor") in the caudal lateral epiblast. NMP; neuromesodermal progenitor. NCC; neural crest cell. UMAP; uniform manifold approximation projection.

Figure 3

Heterogeneity of cell type content within clones containing mesenchyme

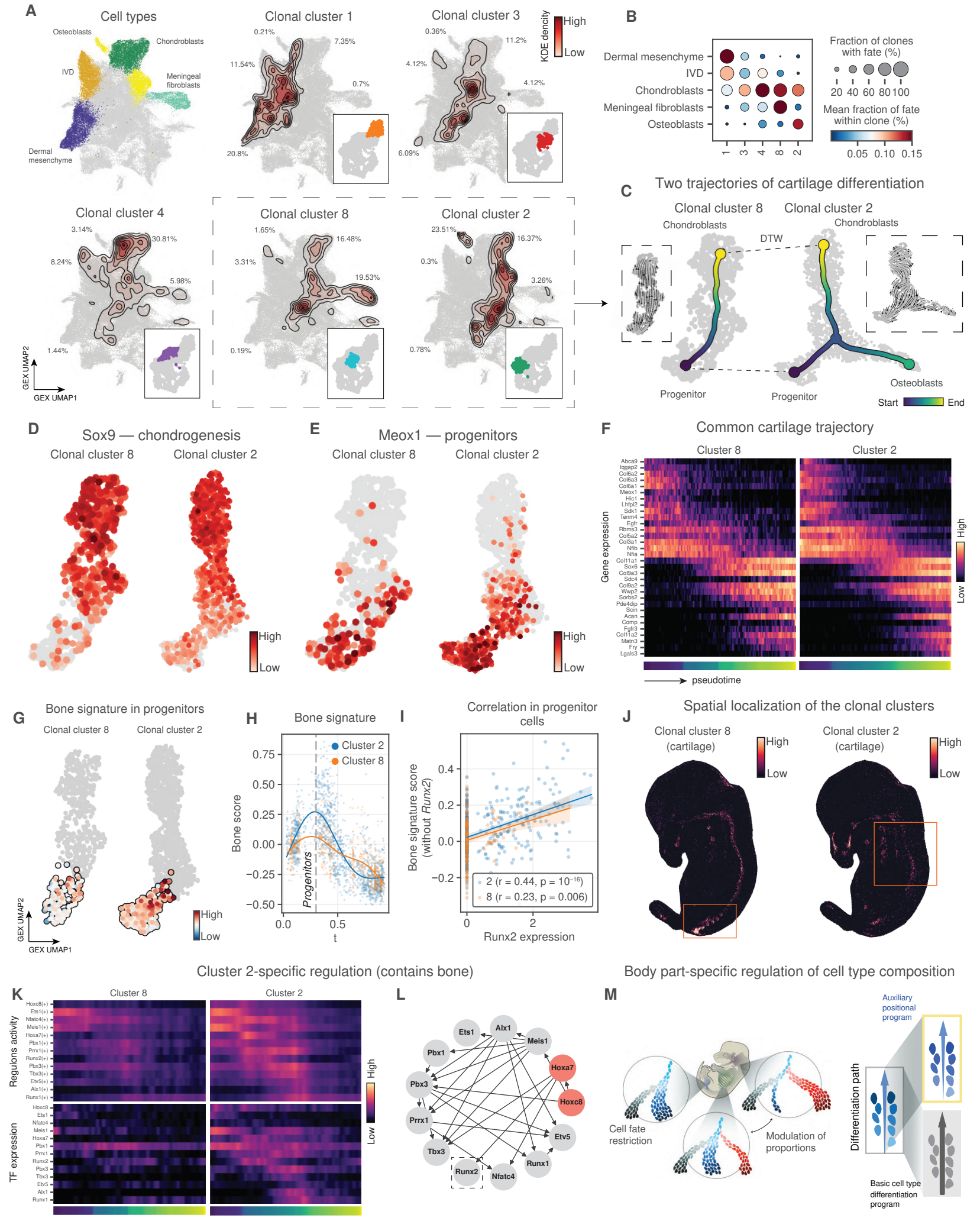


Figure 3. Position-specific programs are superimposed on common paths of cell maturation. **(A)** Density plots of cells from each clonal cluster (clustering resolution 1) showing the proportions dedicated to each selected mesenchymal cell type (color-coded in the top left cell embedding). Insets mark the position on the clone2vec embedding. **(B)** A dot plot summarizing the fate distribution of each clone type shown in (A). **(C)** Trajectory inference of chondrocyte-restricted (clone type 8) and osteochondral bipotent (clone type 2) cells (dark blue representing the origin). Dotted insets show RNA velocity arrows denoting the directionality of cell maturation. **(D-E)** Feature plots displaying *Sox9* (D) and Meox1 (E) expression per cell within the selected chondrocyte-restricted (left) and bipotent osteochondral (right) clone types. **(F)** Heatmap displaying common gene expression pattern along the chondrogenic trajectories of the restricted (left) and bipotent (right) clone types. **(G)** Expression of bone-related genes in progenitors (pseudotime < 0.3) cells. **(H)** Scatterplot of gene expression of the bone-related gene signature, as a function of pseudotime, in cells belonging to the chondral versus osteochondral clone types. **(I)** Correlation between *Runx2* and bone-related genes' expression in progenitor cells from these two clone types. **(J)** E13.5 mouse Stereoseq feature plot of gene scores derived from transcriptional profiles of chondrocytes within chondral (left) versus osteochondral (right) clone types. **(K)** Heatmap displaying patterns of regulon activity (top) and transcription factor expression (bottom) unique to osteochondrogenic clones. **(L)** Gene regulatory networks predicted within this lineage. **(M)** Model of positional regulation of clone development, in which differentiation driven by a core set of genes is influenced by position-specific auxiliary programs.

Figure 4

Mapping of craniofacial clonal cluster signatures to transcriptional and anatomical space

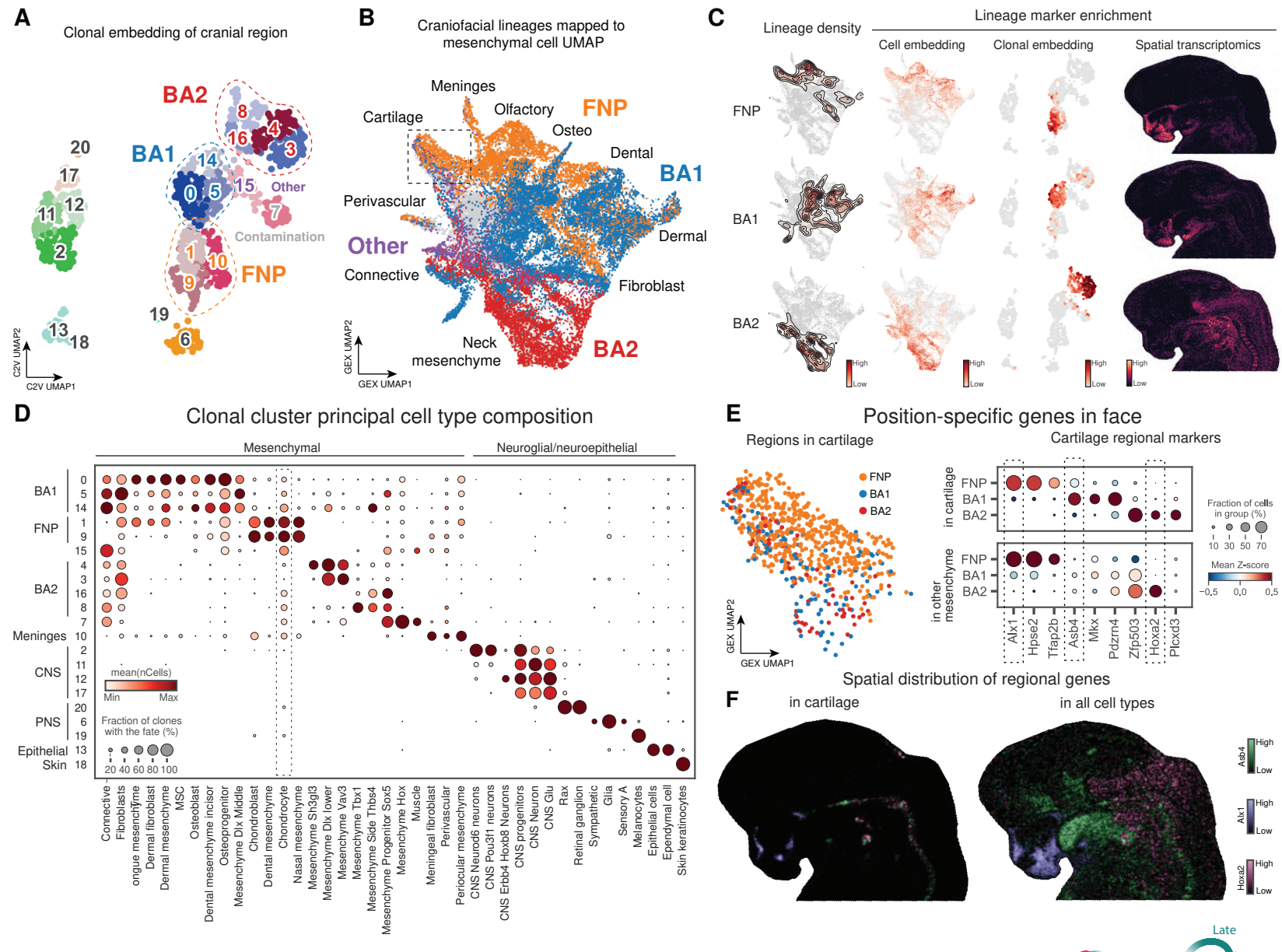
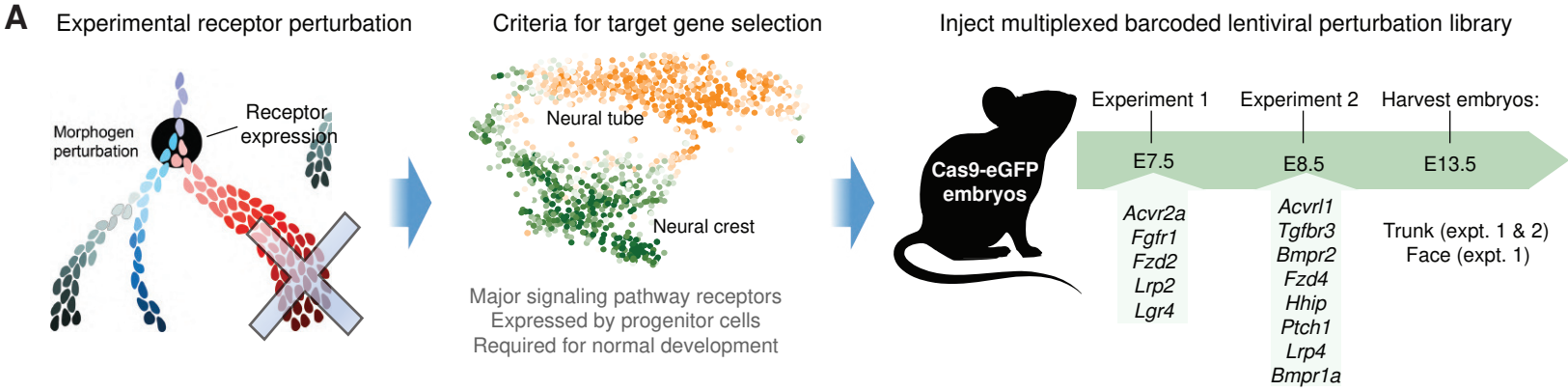


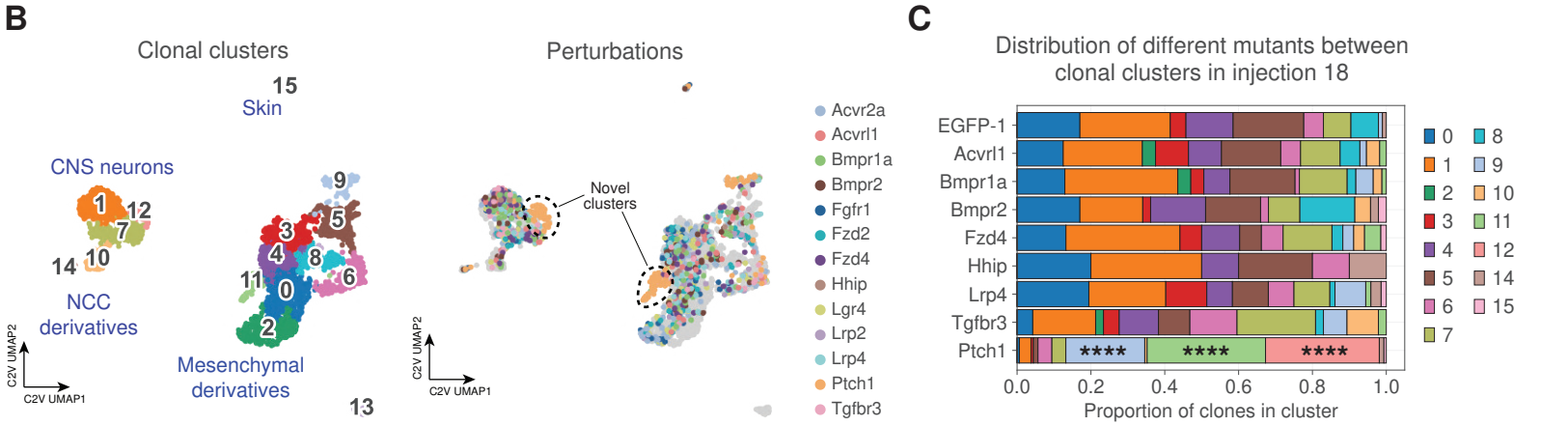
Figure 4. Collective multipotency of the cranial neural crest cell population.
(A) Clone2vec UMAP embedding of E13.5 barcoded craniofacial cells, color-coded by clonal clusters (clustering resolution 2). The dotted lines indicate the major discovered mesenchymal lineages. **(B)** Cell UMAP embedding of E13.5 barcoded craniofacial mesenchymal cells, color coded by the major lineages. Dotted box indicates the position of cartilage cells (later analyzed in (E). **(C)** Mapping of nose (top), jaws (middle), and neck (bottom)-related clonal lineages by density plots for these lineage (left) and feature plots of aggregated lineage-specific gene expression scores on cell embeddings (second row), clonal embeddings (third row) and E13.5 Stereoseq mouse data (right). **(D)** Dot plot summarizing the fate distribution of each clone type in (A). Dotted line highlights the chondrocyte cluster. **(E)** Top left; isolated chondrocyte cell type colored by lineage. Right; Differential expression analysis of chondrocytes revealed lineage-enriched genes within mesenchymal cells (top dot plot) and chondrocytes (lower dot plot). **(F)** Expression of *Alx1*, *Asb4*, *Hoxa2* in cartilage (left) and in general (right). **(G)** Embedding in (a) color-coded by timing of injection. **(H-I)** Confocal micrographs of Sox10-CreERT2; R26R-Confetti embryos. Left; Sagittal section through the head of the embryo traced from E8.5 to E13.5. ERBB3 is labeled in magenta and 2H3 is labeled in dark blue. Cyan, yellow, green, and red colors derive from the confetti fluorescent protein labeling. Bottom left; an inset of trigeminal ganglia (outlined with the dotted line) showing rare double-positive GFP/RFP cells. Right panels; maximum intensity projection of whole mounts of Confetti-traced E10.5-E11.0 embryos traced from E8.5-E9. The proportion of observed tracing in the jaw versus trigeminal ganglia is quantified (I). **(J)** Model of evolving fate biases in neural crest based on the timing of emigration from the neural tube.

Figure 5

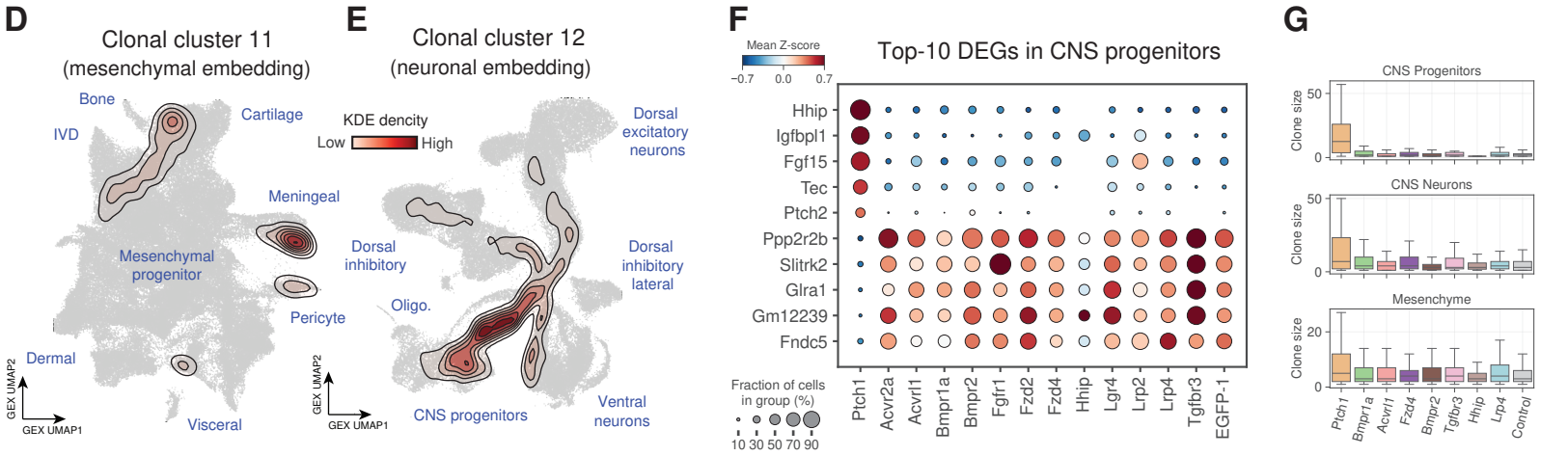
Perturb-TREX experimental design



Analysis of the perturbed clonal landscapes in trunk



Analysis of novel clone types formed by Ptch1-gRNA cells



Control clone types most similar to the novel clone types

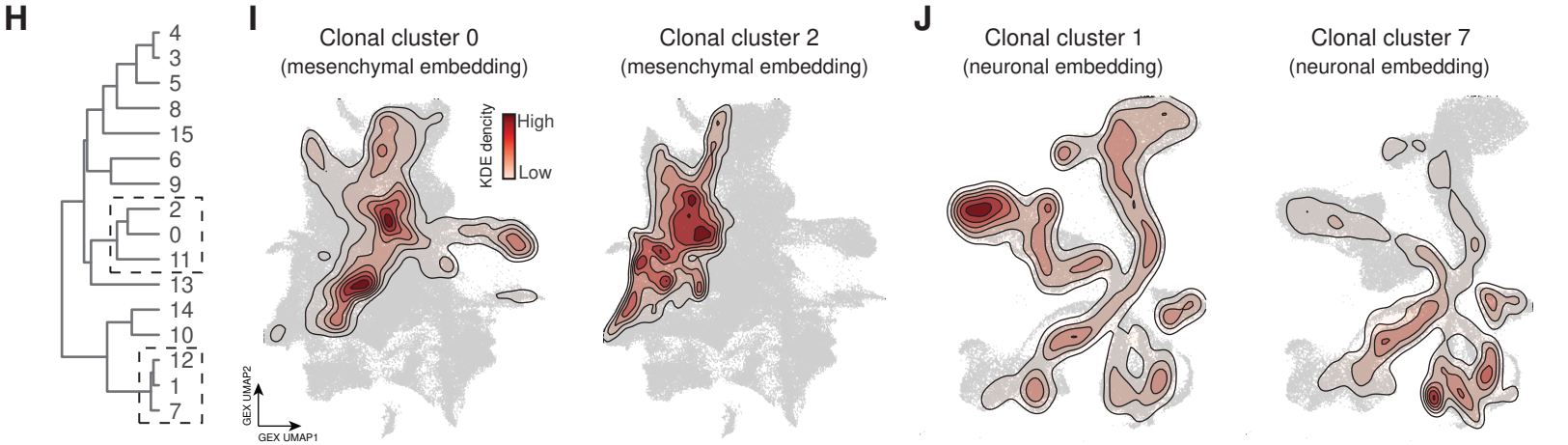


Figure 5. Novel clonal behaviors caused by genetic perturbations.
(A) Experimental design. **(B)** Clone2vec UMAP embedding of E13.5 control and perturbed barcoded trunk cells, color-coded by clusters (left, clustering resolution 1) and by perturbation (right). Dotted lines indicate clonal clusters highly enriched with Ptch1 gRNA$^+$ clones. **(C)** Bar chart showing the proportion of clone types constituting each gRNA$^+$ population, statistical significance was calculated with scCODA **(D-E)** Density plot of the Ptch1-enriched mesenchymal (D) and neuronal (E) clone types. **(F)** Dot plot displaying the top differentially expressed genes by the Ptch1 gRNA$^+$ CNS neural progenitors. **(G)** Box and whisker plot quantifying the distribution of the number of cells per clone within each gRNA$^+$ population (line = median; box = IQR; whiskers = [Q1 - 1.5IQR; Q3 + 1.5IQR]). **(H)** Hierarchical clustering of clone types based on Euclidean distance between centroids in clone2vec space. Dotted boxes indicate the novel clone types. **(I-J)** Clone types most closely related to the mesenchymal (I) and neuronal (J) Ptch1 novel clone type.

Figure 6

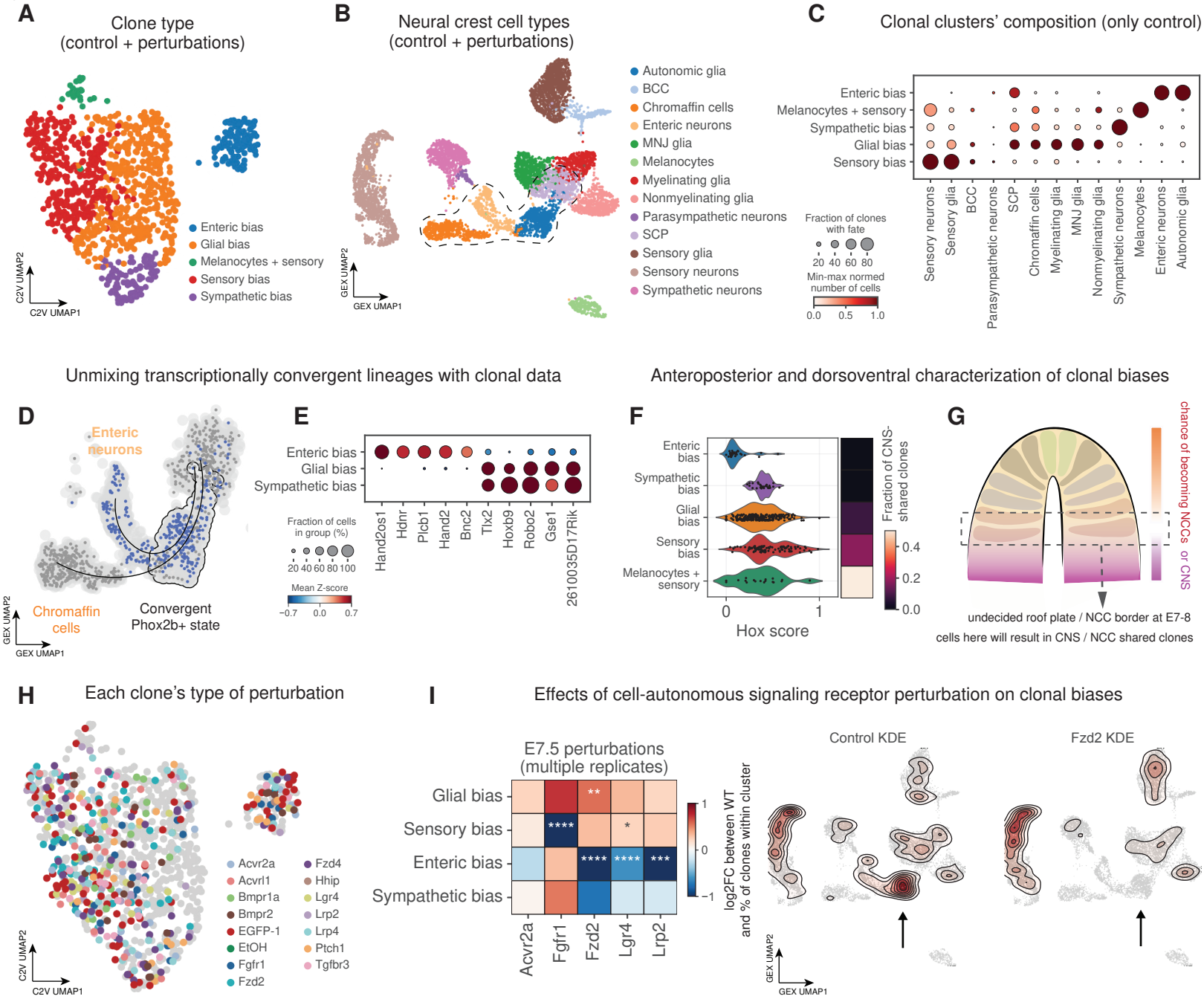

A
Clone type
(control + perturbations)
Enteric bias
Glial bias
Melanocytes + sensory
Sensory bias
Sympathetic bias
C2V UMAP2
C2V UMAP1
B
Neural crest cell types
(control + perturbations)
Autonomic glia
BCC
Chromaffin cells
Enteric neurons
MNJ glia
Melanocytes
Myelinating glia
Nonmyelinating glia
Parasympathetic neurons
SCP
Sensory glia
Sensory neurons
Sympathetic neurons
GEX UMAP2
GEX UMAP1
C
Clonal clusters' composition (only control)
Enteric bias
Melanocytes + sensory
Sympathetic bias
Glial bias
Sensory bias
Sensory neurons
Sensory glia
BCC
Parasympathetic neurons
SCP
Chromaffin cells
Myelinating glia
MNJ glia
Nonmyelinating glia
Sympathetic neurons
Melanocytes
Enteric neurons
Autonomic glia
Fraction of clones with fate
20 40 60 80
Min-max normed number of cells
0.0 0.5 1.0
Unmixing transcriptionally convergent lineages with clonal data
D
Enteric neurons
Chromaffin cells
Convergent Phox2b+ state
GEX UMAP2
GEX UMAP1
E
Enteric bias
Glial bias
Sympathetic bias
Hand2os1
Hdnr
Plcb1
Hand2
Bnc2
Tlx2
Hoxb9
Robo2
Gse1
2610035D17Rik
Fraction of cells in group (%)
20 40 60 80 100
Mean Z-score
−0.7 0.0 0.7
Anteroposterior and dorsoventral characterization of clonal biases
F
Enteric bias
Sympathetic bias
Glial bias
Sensory bias
Melanocytes + sensory
0
1
Hox score
Fraction of CNS-shared clones
0.4
0.2
0.0
G
chance of becoming NCCs or CNS
undecided roof plate / NCC border at E7-8
cells here will result in CNS / NCC shared clones
H
Each clone's type of perturbation
Acvr2a
Acvrl1
Bmpr1a
Bmpr2
EGFP-1
EtOH
Fgfr1
Fzd2
Fzd4
Hhip
Lgr4
Lrp2
Lrp4
Ptch1
Tgfbr3
C2V UMAP2
C2V UMAP1
I
Effects of cell-autonomous signaling receptor perturbation on clonal biases
E7.5 perturbations
(multiple replicates)
Glial bias
Sensory bias
Enteric bias
Sympathetic bias
Acvr2a
Fgfr1
Fzd2
Lgr4
Lrp2
**
****
*
****
****
***
1
0
−1
log2FC between WT and % of clones within cluster
Control KDE
Fzd2 KDE
GEX UMAP2
GEX UMAP1

Figure 6. Multipotency landscape of trunk neural crest in normal and perturbed conditions.

(A) Clone2vec UMAP embedding of E13.5 control and perturbed barcoded trunk neural crest-related cells (with more than 2 cells), color-coded by clone types. **(B)** UMAP cell embedding of the same group of cells shown in (A), color coded by the cell type. Dotted line shows enteric and chromaffin trajectories. **(C)** Dot plot showing cell type distribution of each clone type. Dot size corresponds to the proportion of clones including a given fate; Dot color corresponds to the average number of cells with that fate per clone. **(D)** Cell UMAP of the diverging enteric and chromaffin trajectories. Highlighted region indicates the common Phox2b$^+$ glial progenitor state. **(E)** Dot plot showing gene expression profiles specific to enteric versus glial and sympathetic lineages within that shared state. **(F)** Predicted anteroposterior positions based on Hox gene expression across all clones within each clone type (left) compared to the proportion of clones shared with CNS cell types shown as a vertical heatmap (right). **(G)** Model of early CNS/NCC shared clone development based on asymmetric cell divisions before NCC delamination. **(H)** Embedding as in (A) color-coded by perturbation. **(I)** Left; heat map showing fate distributions shifted by each perturbation. Beta-regression LRT-based p-values with FDR multiple testing correction was used for significance estimation. Right; Density plots of control (left) and gRNA$^+$ (right) clones on the NCC cell UMAP embedding. Arrows indicate the specific loss of enteric fates.