# Unbiased profiling of multipotency landscapes reveals spatial modulators of clonal fate biases

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 of 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 fails to capture the spatiotemporal modulation of cell multipotency underlying morphogenesis. In this study we profile the multipotency landscape of the embryo, using *in vivo* single-cell clonal lineage tracing of mouse embryos traced from neurulation until mid-gestation, combined with a machine learning tool that categorizes individual clones into lineages based on shared transcriptional context. This revealed a previously unrecognized continuous, embryo-wide gradient of clonal fate biases, in which anatomical position and clonal composition are mutually predictive. Comparing clonal lineages revealed gene regulatory networks underlying the dynamic biasing of cells towards specific fates by spatial transcription factor programs. However, mosaic combinatorial perturbations targeting the Hedgehog pathway generated clones in which positional identity was mismatched with clonal composition, demonstrating that extrinsic signals can override the axial patterning system underlying clonal fate biases. Altogether, our work demonstrates an effective practical approach for dissecting mechanisms of lineage specification and has implications for stem cell engineering.

**Main Text:**

*Introduction*

Stochasticity during vertebrate embryonic development poses a challenge for understanding how the 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, each contributing just a part of the overall repertoire of the body's cell types (Fig. 1b). Consequently, the classic model of cell differentiation as a "tree-of-fates" represents a kind of aggregation of many individual clones at a given location. However, cell fate decisions and their underlying tree structure are thought to be highly influenced by anatomical position within 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 insights for identifying 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, especially when it comes to understanding how the proportions of cell types are established [9, 10]. Quantifying the spatiotemporal variation of clonal fate biases, and identifying potential biasing factors, are important steps 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 few 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. More fundamentally, the measurements of the cellular state that accompany tracing information are performed at the final timepoint. Consequently, molecular differences associated with 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 nature of embryonic development, for instance in the case of the neural crest [11, 12] and neuro-mesodermal progenitor cells [13], represents a partial solution to these challenges by enabling faithful differentiation trajectories to be constructed from lineage-aware gene expression measurements.

While traditionally, lineage tracing has been used to establishing the range of cell types that can arise from progenitor populations (i.e. potency), here we focus on quantitatively mapping the relative frequencies of transcriptional states occurring among the progeny of a cell (clonal fate

biases). Even for progenitor cells with identical lineage potency, such frequencies will vary due to the influence of external signals and stochastic differences in the progenitor state. Our aim is to characterize such fate variation and identify factors biasing fate realization in embryonic development. 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 study the diversity of fate biases occurring in ectodermal and mesodermal clones during mammalian embryonic development. This approach captured gradients of continuous clonal fate variation along anatomical axes as well as sequential fate restrictions over time, enabling computational inference of the regulatory gene programs likely responsible for modulation of fates in neural and mesodermal progenitor cells. This system of axial clonal patterning was also susceptible to disruption by multiplexed *in vivo* signaling receptor perturbations, suggesting that positional regulatory factors cooperate with extrinsic signaling during to ensure the correct clonal proportionality of cell types during organogenesis.

*Profiling of clonal behavior in mouse embryos*

To reconstruct joint patterns of transcriptional states and cell lineages in post-neurulation murine development, we performed ultrasound-guided nano-injections [14] into amniotic fluid to transduce mesoderm and ectoderm with heritable, transcribed genetic TREX barcodes [15, 16] (Fig. 1d). The injections were carried out once or twice serially between days E7.5 and E8.5 (Fig. 1e-f, Extended Data Fig. 1) to capture rapid developmental shifts during embryonic development. Trunk and cranial 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, Extended Data Fig. 2). 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 46-206 nL) such that only 10-40k cells would be labeled in each harvested body segment – an amount that could be taken through downstream processing steps without significant additional losses.

Our microinjection approach predominantly targets ectoderm [14-16], as well as some mesoderm. This is explained by the fact that the ultrasound guided injections shower the exterior of the early neurula with barcoded lentiviral particles, and the exposed cells residing in outermost positions have the highest chance to become labeled [14]. 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 [17]. Two days after the injection we saw the formation of large clones by whole mount analysis of embryos injected with low volumes of viral particles (Extended Data Fig. 3), and broad labeling thereafter (Extended Data Fig. 4), 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 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 neuroglial cells of the central and peripheral nervous system (Fig. 1i-j, Extended Data Fig. 5 and Extended Data Fig. 6). Most clones contained similar cells belonging to only one germ layer derivative (Fig. 1k-l; for cranial, see Extended Data Fig. 7). 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 (Fig. 1l). The patterns of clonal co-occurrence grouped similar cell types, recovering the expected consensus cell type lineage phylogeny (Fig. 1m and Extended Data Fig. 8). Reconstruction of such lineage structure from fixed barcode data, however, is known to be challenging [18]. 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 (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 [19] word embeddings to lineage tracing information (see Materials and Supplementary Note: Formal Description of clone2vec). 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 across the continuous expression space (Fig. 1p, Extended Data Fig. 9). By establishing this metric space, clone2vec enables exploration of recurrent clonal patterns and subsequent analysis of features associated with them. We characterized the performance of the clone2vec method using simulated data (see Supplementary Note: Simulations), as well as published datasets. We found that clone2vec embedding effectively reconstructs both discrete clusters of clones, as well as patterns of continuous clonal variation within clusters. Importantly, clone2vec performance remained robust under increasingly sparse sampling of cells (Supplementary Note: Simulations). We compared clone2vec analysis with the more traditional approach of representing clones in the space of cell types, where each clone is encoded as a vector of cell type abundances, followed by PCA dimensionality reduction. This embedding results in relatively poor reconstruction of the original clusters, with the advantages of clone2vec becoming more evident under sparser clonal coverage (see Supplementary Note: Simulations). Also, much like word embeddings, clonal embeddings are amenable to vector math operations (see Supplementary Note: Simulations) – for instance, multipotent neuromesodermal progenitor (NMP) clones are the vector sum of neuronal clones and mesodermal clones. Clone2vec re-analysis of published hematopoietic lineage tracing data [20] was able to recover previously described fate patterns with their associated genes, additionally uncovering more subtle continuous axes of variation (see Supplementary Note: Real Data). These features make clone2vec a useful tool for learning patterns of clonal behavior in the embryo.

Clone2vec analysis of our dataset revealed three major discrete compartments corresponding to derivatives of the neural ectoderm, non-neural ectoderm, and mesoderm (see Data Availability for an interactive visualization) and facilitated the assignment of clones into clonal clusters within

185 each compartment (Fig. 2a). The general structure of the clonal embedding reflected the observed
186 restricted nature of mesoderm, non-neural ectoderm, and neural ectoderm clones. However, two
187 clonal clusters exhibited neural-mesenchymal multipotency (Fig 2b, dotted box). Surveying the
188 specific combinations of epidermal, neuronal, and/or mesenchymal cell subtypes found in each
189 clonal cluster, revealed that nearby clusters often share many cell types, but differ chiefly in their
190 proportionality in which they are found, especially in mesenchymal clones (Fig. 2b). Reasoning
191 that adjacent clonal clusters represent points along a continuum of clonal behavior, we performed
192 principal components analysis (PCA) from discrete versus contiguous clonal clusters (Extended
193 Data Fig. 10a-b) and examined gene expression and cell fate distributions across a constructed
194 trajectory of clones following the top principal components (Extended Data Fig. 10c-d). The clones
195 in contiguous clusters exhibited a continuous gradient of gene expression (Extended Data Fig. 10e)
196 reflecting gradual transitions across mesenchymal cell types (Extended Data Fig. 10f-j).
197
198 To study the molecular drivers underlying continuous variation in clonal composition, we devised
199 a machine learning approach to identify genes whose expression, either in a specific cell type or
200 across all cell types, was strongly associated with clonal fate distributions. We first built a gradient
201 boosting model to predict the position of clones in the embedding based on cell type-stratified gene
202 expression in cells belonging to each clone (see Methods). We then scored gene importance by
203 quantifying each gene's predictive power using Shapley's values (measure of feature importance,
204 see Methods). Next, to exclude the most obvious predictors (cell type markers) we added an
205 additional categorical feature (cell type identity) to account for cell type-dependent effects.
206 Applying this approach to the trunk mesodermal clones revealed that many Hox transcription
207 factors were among the most important genes to predict clonal composition, suggesting that the
208 observed fate bias gradient was driven by clones arising from somites at different positions along
209 the body axis (Fig. 2c). We constructed a 'Hox score' to locate clones by aggregating their
210 expression levels of Hox genes (Fig. 2d, Extended Data Fig. 11a-d). Clone position along the
211 observed clone2vec fate gradient, and clone Hox score, were linearly related (Extended Data Fig.
212 11c). To validate that the clonal embedding captured the anterior-posterior axis of the embryo, we
213 separately dissected lower and middle trunk prior to sorting, confirming a striking difference in
214 their representation along the mesenchymal domain embedding (Fig 2d). We exploited this metric
215 to uncover several genes (beyond the Hox code) that are highly correlated with the body axis
216 (Extended Data Fig. 11e-h). This linear relationship was driven primarily by somite derived clones,
217 not other mesenchymal cells (Extended Data Fig. 11i-l) nor the cell cycle (Extended Data Fig.
218 11m). This analysis suggests that somite clonal composition and body position are mutually
219 predictive, and that knowledge of a few genes can reconstruct both, even though the underlying
220 cell types occur along the entire body axis. Additionally, the data suggests that application of these
221 models can help reveal new positional information genes.
222
223 To identify early biases during progenitor cell differentiation, we examined the predictors of
224 clone2vec embedding position of clones based exclusively on the expression profile of progenitor
225 cells (Extended Data Fig. 12a-b, e, please see embeddings of both neuronal and mesodermal
226 lineages). To annotate the effect of gene expression on fate distribution, we built an additional
227 model that predicts cell type composition of the clone based on the expression in progenitor cells
228 (see Methods and Extended Data Fig. 12c-d, f-g) and outlined programs biasing towards one or
229 another fate. Applying this approach revealed, for example, that during mesenchymal progenitor
230 cell differentiation, the most powerful predictor appeared to be the *Tbx3* gene rather than one of

the Hox genes (Extended Data Fig. 12d), showing that gene expression correlates of clonal behavior are distinct when only considering progenitor cells, versus all cells (including differentiated progeny). The same strategy was used on a publicly available dataset of *in vitro* hematopoiesis (see Supplementary Note: Real Data) illustrating that this approach allows for systematic analysis of gene expression associations with clonal behaviour.

Because the embryo develops from head to tail, caudal cells should exhibit greater developmental potential than their anterior counterparts at a given moment. Neuromesodermal progenitors (NMPs, clonal cluster 15 and 16), which are a cell population that contributes to secondary neurulation and somitogenesis during axial extension, were found near the tail [13, 21]. We noted that, while the posterior-most clones corresponding to NMPs exhibited the greatest multipotency (Fig. 2e), levels of clone transcriptional diversity (measured via Shannon entropy) did not always correspond with body position (compare feature plots of mesenchymal clonal embeddings in Fig. 2d versus 2e). For instance, tail clusters 15, 16, and 7 had rather similar Hox scores (bracket, Fig. 2d), but the most multipotent NMP cluster 16 was enriched with E7.5-labeled clones (Fig. 2f), while other tail clusters (15, 7) contained a mix of E7.5 and E8.5 clones, suggesting a temporal, rather than spatial, explanation for their divergent profiles.

To visualize the gradual posteriorizing of NMP-competent clones over time, 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. 2g-h). 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 injected samples (Fig. 2h), consistent with an expected coupling between the progressive spatial and temporal fate restrictions in the NMP lineage.

We were intrigued that clonal cluster 15 behaved somewhat like NMPs by generating neural crest cells and somitic derivatives, but not CNS cells (Fig. 2b, f). Consistent with this, we observed some sections with labeling of both crest-derived neurons and vertebrae but few CNS neurons (Fig. 2h). Together, this evidence suggests some NMPs can directly generate neural crest during secondary neurulation without contributing to neural tube epithelialization, demonstrating existence of a neural crest - mesodermal bipotent progenitor (NCC-MP, Fig. 2i). To demonstrate the existence of this population, we analyzed the tail mesenchyme of four additional embryo samples (Extended Data Fig. 13a-f). Indeed, we again recovered a clonal cluster corresponding to NCC-MPs (Extended Data Fig. 13g-j). Compositional analysis of tail clones revealed that NCC-MPs differ significantly from canonical NMP clones in the number of CNS cells, but both produce similar amounts of neural crest cells; as a result, these clonal clusters are marked by a clear bias towards either CNS or neural crest cells (Extended Data Fig. 13k-m). The NCC-MP population has been predicted by earlier work [22, 23] but not yet reported. Further studies of NCC-MP biology might prove relevant to the etiology of some *spina bifida* subtypes.

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

Cell fate determination involves coordinated induction of transcriptional programs and, often, suppression of alternate fates [11]. By delineating sets of clones that demonstrate distinct fate outcomes, we studied how context-dependent transcriptional programs subtly bias multipotency among otherwise similar progenitor cell populations.

To illustrate how clonal embedding enables fine analysis of how positional factors influence cell differentiation, we examined a group of trunk mesodermal lineages that build the axial skeleton with the help of Hox-dependent patterning [24-26]. In the transcriptional embedding, trunk mesenchymal populations form a large diffuse cluster (Fig. 1j), containing progenitor as well as early differentiated populations (Fig. 3a, top left). While standard trajectory inference analyses on such a diverse population would fail due to uncertainty, clone2vec neatly separated the mesenchymal cells into a set of partially overlapping lineages (Fig. 3a). Ordering these lineages according to their Hox score revealed their graded contribution to skeletal cells, dermis, meningeal fibroblasts, and perivascular cells as a function of body position (Fig. 3b and boxed insets in Fig. 3a, Extended Data Fig. 10-11). We noted that cartilage cells were present in many of these lineages and therefore may be a good model to study position-specific modulation of regulatory programs in the body. We refer to such spatial regulatory factors that introduce variation to otherwise similar cell differentiation events as "auxiliary programs".

To further understand the factors driving variation in multipotency profiles, compared two 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. 3a). To perform the analysis we extracted bone, cartilage and progenitor cells from these lineages, filtered out meningeal cells, and re-embedded them to compare their chondrogenic trajectories (Fig. 3c). Mapping lineage-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 3d, Extended Data Fig. 14), while the osteochondral lineage 2 is likely from anterior structures including early bone collar of ribs, which starts forming from perichondrial cells at E13.5 (Fig. 3e) [27, 28]. We observed a full spectrum of intermediate cell states at E13.5 due to continuous differentiation from mesenchymal progenitors to mature chondrocytes from both clonal clusters (Fig. 3f-g), suggesting a *bona fide* developmental trajectory in both cases, although osteochondral clones appear to maintain expression of progenitor markers such as *Meox1* for longer (Fig. 3g). Additionally, the presence of meningeal cells in Cluster 8 indicates that it is not simply a less-mature version of Cluster 2 (as might be predicted based on the asynchrony of somite development).

Focusing on chondrogenic differentiation in both lineages, we found that both executed a conserved transcriptional program, upregulating common chondrocyte gene markers (Fig. 3h and Extended Data Fig. 15a-d) including *Col11a1*, *Sox6*, *Col9a3*, *Acan*, *Comp*, *Fgfr3*, *Col11a2*, and *Matn3*. Expression of bone-related genes (Fig. 3i and Extended Data Fig. 15e-f) increased at the beginning of 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. 3j), aligning with existing models [29].

323 Despite being downregulated in the chondrogenic-only lineage (Extended Data Fig. 15 g-k), *Runx2*
324 expression significantly correlated with the bone-related signature even there (Fig. 3k), suggesting
325 that molecular regulators of mesenchymal *Runx2* expression drive a pro-chondrogenic progenitor
326 cell's osteogenic versus fibroblast potential in vivo.
327
328 Differential expression analysis of the clonal lineages with distinct *Runx2* activation implicated
329 TFs related to the Hox code, mesenchymal fate commitment, and skeletal development (Extended
330 Data Fig. 16a-h), leading us to question the hierarchy of observed TFs. Using SCENIC [30], we
331 inferred transcriptional factor activity and likely gene regulatory networks acting during cartilage
332 development in both clonal clusters (Fig. 3l-m and Extended Data Fig. 16i-j). The analysis
333 suggested an upstream regulatory role of Hox TFs, with *Hoxa10/Hoxd10* (lumbosacral)
334 coordinating the regulatory network in chondrogenic clones, and *Hoxa7/Hoxc8* (thoracic) in
335 osteochondral clones. *Hoxa7* was positioned directly upstream of *Runx2* expression in the
336 reconstructed gene regulatory network (Fig. 3m), which could explain the positional differences
337 in clone skeletogenic potential. Additionally, *Hoxc8* (Fig. 3m) was situated upstream of
338 mesenchymal TF *Prrx1* (via intermediates *Meis1* and *Pbx3*), which like *Meox1* exhibited sustained
339 expression specifically in the bone-forming lineage 2, suggesting that positional factors might
340 regulate the tempo of early chondrogenesis by maintaining a mesenchymal progenitor status
341 (*Meox1*/*Prrx1* expression). These roles are consistent with a direct regulatory function of Hox
342 genes in mesenchymal progenitor fate decisions, in agreement with observations at later
343 developmental stages [31].
344
345 Overall, combining high-throughput lineage tracing with trajectory inference on cell manifolds
346 [10] uncovered mechanisms of clonal fate biasing in the embryo, by decoupling cell type-specific
347 gene modules from spatial gene modules (Fig. 3n). Comparing lumbar and thoracic mesoderm
348 revealed 1) the core transcriptional program of axial pro-chondrogenic differentiation, 2) early
349 molecular indicators of osteochondrogenic biopotency, and 3) evidence that spatial programs drive
350 bone collar formation in thoracic cells, explaining why ribs mineralize faster than the posterior
351 vertebral cartilages. Untangling how spatial regulatory mechanisms are influenced by
352 heterochrony, and how this directs specific clonal dynamics such as differentiation tempo and
353 proliferative potential, remains a future challenge.
354
355 *Positional codes in facial mesenchyme revealed through clonal analysis.*

356 In contrast to the trunk, the anterior cranial region is largely devoid of Hox expression. While
357 previous studies revealed many TFs that are critical for establishing the facial prominences [32,
358 33], how spatial programs influence multipotency of facial cells at a molecular level remains
359 insufficiently understood. We applied our clonal variation approach to this problem, again
360 focusing on chondrocytes, reasoning that transcriptional variation within a single cell type would
361 accentuate position-specific programs by minimizing cell-type-related artefacts. The cranial
362 dataset captured a rich spectrum of clonal fate biases (Fig. 4a-b), covering populations contributing
363 to central and peripheral nervous system, dermis, meninges, and mesenchymal derivatives of the
364 musculoskeletal system which was segregated into the major lineages corresponding the major
365 facial prominences (Fig. 4b-d). Examining the spatial distributions of lineage-specific genes in
366 Stereo-seq data, we found that they consistently mapped to different facial subregions (Fig. 4c).
367 Differential expression analysis of specific chondrogenic signatures within all three major lineages
368 revealed key transcription factors, which were sufficient to reconstitute the expected positional

369 identity (Fig. 4e). We then confirmed the cartilage positional signatures (including previously
370 reported markers *Alx1/Asb4/Hoxa2* [34, 35]) were also observed in progenitor cells originating
371 from the same clones (Fig. 4e). The true anatomical position of these signatures was validated by
372 re-analysis of Stereo-seq data (Fig. 4f). This suggests that these signatures likely represent
373 position-specific programs modulating progenitor cell fate choices and plasticity. Indeed, fate
374 profiles of cranial clonal clusters revealed that Hox-negative BA1 (branchial arch 1) progenitor-
375 derived clones were highly multi-fated, containing fates from nearly all mesenchymal clusters,
376 whereas Hoxa2$^{+}$ clones from BA2 (branchial arch 2) lacked several mesenchymal cell types (Fig.
377 4d). CatBoost regression-based numeric estimation of gene importance for predicting clone
378 composition and location resulted a list of top factors including known spatial regulatory factors
379 (Fig. 4g). Overall, these results demonstrate how analysis of continuous clonal variation can
380 suggest candidates for positional programs within the developing embryo.
381
382 *Cranial neural crest generates neuronal and mesenchymal fates from partly overlapping but*
383 *mostly distinct progenitor pools.*
384
385 Beyond position-specific effects, continuous clonal variation can also capture temporal
386 dependencies in multipotency regulation. Cranial neural crest cells (CNCCs) emigrate in waves
387 from the dorsal neural fold but face different microenvironmental contexts depending on their
388 exact timing of delamination, which makes for a good system to study temporal modulation of cell
389 fate choices in vivo [36, 37]. As a population, the cranial neural crest is highly multipotent, giving
390 rise to facial skeletal and other mesenchymal structures as well as to the peripheral nervous system
391 (trigeminal and sympathetic ganglia, including the glial cells), but whether such broad
392 multipotency can be attributed to individual CNCCs remains unclear [37-40]. For instance, each
393 wave of CNCCs might carry specific fate biases; alternatively, multipotent CNCCs would be
394 biased by the post-migratory environment.
395
396 Examining clonal variation of CNCC lineages, we found that facial mesenchymal derivatives and
397 trigeminal neuroglial cell types were overwhelmingly separated into distinct clonal subpopulations
398 (clusters 6+19 versus clusters 0-1,3-5,7-10,14-16, Fig. 4d) and very few clones shared trigeminal
399 neuronal and mesenchymal fates. Furthermore, these two clonal subpopulations were
400 distinguished by sequential E7.5 + E8.5 double injection, with neuroglial cell types labeled
401 predominantly at later E8.5 injections (Fig. 4h). This result suggests that early and late-migrating
402 CNCC waves become differentially biased.
403
404 To test the temporal separation between mesenchymal and neuroglial trigeminal cell types, we
405 performed lineage tracing experiments in *Sox10-CreERT2/R26Confetti* embryos, which enable
406 clonal tracing based on multicolor labelling of stochastically recombined cell populations (Fig 4i).
407 Recombination was triggered by tamoxifen injection at slightly different developmental times
408 (~12h differences at E8.5 stage). The resulting clonal tracing showed that, although early labeled
409 clones contribute broadly to all facial tissues, later-labeled clones were confined to trigeminal
410 ganglion and outgoing nerves, revealing a systematic difference in the cell type composition
411 depending on the timing of recombination (Fig 4j). This result is consistent with a temporal model
412 in which different migratory waves of CNCC undergo differential fate biasing (Fig 4k). This
413 finding demonstrates how clonal variation can help to resolve subtle temporal effects on lineage
414 separation and fate biasing among nearly uniform progenitor cell populations.

415
416 *Perturb-TREX: Joint clonal tracing with multiplexed signaling receptor perturbations.*
417
418 Hypotheses about programs modulating progenitor fate must be eventually tested by experimental
419 perturbations. In this regard, the highly orchestrated developmental process poses a challenge, as
420 disruption of pathways critical for cell fate determination, even with conditional genetic
421 perturbations can result in major distortions of the tissue and lead to widespread indirect effects.
422 Mosaic perturbations of individual clones, however, are unlikely to cause such tissue distortions,
423 presenting a valuable opportunity to study the role of genes with strong phenotypes *in vivo*.
424
425 To examine the role of extrinsic signals in early trunk development, we combined TREX tracing
426 with multiplexed CRISPR/Cas9 gene editing of individual clones (Fig. 5 and Extended Data Fig.
427 17) [41]. Taking advantage of detailed scRNA-seq data on differentiating neural crest and other
428 cell lineages, we extended the barcoded TREX vectors with a library of guide RNAs targeting
429 receptors of signals associated with early cell fate bifurcation steps (Fig. 5a) [11]. These
430 multiplexed lentiviral libraries were injected into murine embryos at E7.5 or E8.5 and examined
431 with scRNA-seq and clone2vec at E13.5 (Fig. 5b). In part because of the known relationship
432 between skeletogenesis and Hedgehog signaling, we included two gRNAs targeting the Hedgehog
433 pathway opportunity to study the role of extrinsic signals during axial patterning of skeletal cell
434 subtypes. While, as expected, none of the embryos showed major developmental abnormalities,
435 significant fate deviations were observed for mosaic perturbations at the clonal level, resulting in
436 ectopic clonal clusters (Fig. 5c). These clusters were enriched for *Ptch1* crispant clones - a gene
437 encoding sonic hedgehog (SHH) receptor [42], and likewise, *Ptch1* gRNA$^{+}$ clones were
438 significantly enriched within the ectopic clusters (Fig. 5d), which represented some variation on
439 neural tube clones (clonal cluster/lineage 12), ventral mesenchyme (lineage 9), and somitic clones
440 (lineage 11).
441
442 Hierarchical clustering to determine lineage similarity was used to find points of comparison for
443 the ectopic clusters (Fig. 5e). For example, in the developing CNS, *Ptch1*-crispant clones were
444 enriched in progenitor populations (Fig. 5f-g) compared to similar CNS lineages. *Ptch1* crispant
445 clones also upregulated a unique gene expression signature (i.e. *Hhip, Fgf15, Ptch2, Plagl1)*
446 related to cell cycle and Hedgehog signaling (Fig. 5h), and significantly increased cell counts per
447 clone (Fig. 5i). Thus, *Ptch1* editing in neuroectodermal clones results in an accumulation of stalled
448 intermediate progenitors prior to their cell cycle exit and neuronal differentiation, and Ptch1-
449 mediated inhibition of Hedgehog signaling is needed for CNS neural cells to mature beyond a
450 transit-amplifying progenitor state.
451
452 The ectopic clone type resulting from elimination of *Ptch1* from mesoderm-derived mesenchyme
453 was not found in any other analyzed conditions (Fig. 5j-k). This distinct clone type was enriched
454 with vertebral chondrocytes and meningeal cells (much like the lumbar path 8 in Fig. 3) but was
455 specifically depleted of undifferentiated mesenchyme and fibroblasts (Fig 5j, compare to control
456 clusters in Fig 5k). This suggests that in contrast to the CNS, the gain-of-function of Hedgehog
457 signaling in mesoderm speeds up the tempo of differentiation, depleting the mesenchymal
458 progenitor pool.
459

460 Given that somites occur along the body axis, we were surprised that *Ptch1* editing resulted in a
461 single perturbed somite lineage and questioned whether this was driven by mis-patterning or
462 differential local sensitivity. We compared the Hox score of mesenchymal-biased lineages
463 including *Ptch1* gRNA-enriched clusters 9 and 11 (Fig. 5l), finding that they drastically differed
464 in positional identity, suggesting an effect of *Ptch1* editing on mesodermal axial patterning.
465 Indeed, analyzing Hox score per gRNA condition revealed a bimodal distribution of Ptch1 clones
466 but no other condition including EGFP controls. Cluster 9, with anterior identity (and found
467 primarily in middle trunk samples), was primarily biased towards duct mesenchymal cell types
468 forming from the ventral part of the somite (Extended Data Fig. 17e, see location of gene
469 expression score in Extended Data Fig. 14h) suggesting that *Ptch1* perturbation caused
470 ventralization of anterior but not posterior mesoderm. On the other hand, the somite Cluster 11
471 (posterior identity, found primarily in lower trunk sample) contained clones from a wide range of
472 Hox scores, some significantly anterior compared to nearest neighbor lineages from lumbar and
473 tail somites (Fig. 5l), suggesting that the axial pattern of clonal fate biases was disrupted.
474 Accordingly, Hox scores of *Ptch1* gRNA$^+$ clones were significantly less predictable than clones
475 bearing other gRNAs, when using any nearest neighbor clone as reference (Fig. 5n), and even
476 when using the same gRNA as reference (Fig. 5o).
477
478 Taken together, while a highly predictable embryo-wide gradient of clonal fate biases is observed
479 under natural conditions, this can be overridden by the activation of cell signaling pathways. Future
480 studies focusing on the lineage-specific interactions of *Ptch1* with cell differentiation programs
481 should be performed using detailed comparison of the regulatory programs acting differentially
482 across lineages (as in Fig. 3) coupled with combinatorial perturbations along a time series.
483
484 *Biasing of cell fate decisions in neural crest revealed via clonal analysis.*
485
486 We next focused on the trunk neural crest, which has been hotly debated as being either a mixed
487 population of restricted progenitors, or a population of highly multipotent stem cells [12, 36, 43-
488 45]. Clonal variation analysis of wildtype embryos revealed that all detected neural crest sub-
489 lineages were highly fate biased. This was observed for sympathetic, melanocyte, somatosensory,
490 enteric, and glial progenies (Fig. 6a-c). Importantly, none of the neural crest lineages were clonally
491 linked with mesenchyme (Fig. 6c), and sparingly few crest-mesenchymal bipotent clones were
492 observed. Thus, although vagal neural crest gives rise to both cardiac mesenchyme and enteric
493 neurons, this multipotency is more accurately attributed to the collective population, and not to
494 individual neural crest cells. Instead, our results suggest that prior to clonal expansion, labeled
495 neural crest progenitors acquire strong fate biases, corresponding to the biases described for
496 different migratory waves of neural crest [36, 46]. Such progenitor differences may be too subtle
497 to be observed from transcriptional analysis alone [47]. For instance, we find that the progenitor
498 state of the clones destined to give rise to the *Phox2b*$^+$ autonomic nervous system and enteric
499 nervous system within the developing gut, both map to the same overall transcriptional state (Fig.
500 6d-e). The knowledge of clonal identity enables a more powerful way of identifying biasing
501 programs by directly contrasting the progenitors with distinct fates. In this way we find that these
502 early progenitors show robust upregulation of a set of biasing genes, including *Hand2* - a
503 transcription factor known to be involved in neural crest enteric differentiation [48]. Since such
504 genes reflect early progenitor biasing mechanisms that appear prior to overt differentiation, their

505 modulation may be helpful for cell type generation protocols and applications intended to cure
506 autonomic and enteric dysfunctions, including Hirschsprung's disease [49].
507
508 We also examined positional distribution of the neural crest multipotency along the anterior-
509 posterior and dorsoventral axes. Based on the *Hox* gene expression we find that along anterior-
510 posterior axis an expected strong anterior bias of the enteric-biased clones [50], as well as paucity
511 of sympathetic clones the posterior sacral and post-sacral part of the body (Fig. 6f). The remaining
512 clonal clusters showed broad distributions. To infer the distributions along the dorsoventral
513 position, we examined the co-occurrence of neural crest and neuronal CNS fates within the same
514 clones. Since neural crest is formed in the most dorsal part of neural tube, increasing contribution
515 of clones to CNS neurons would suggest a more ventral position - closer to the border between the
516 prospective crest domain and the future roof plate of the neural tube [48]. We find that the clones
517 biased towards somatosensory neurons and melanocytes shared a much higher proportion of
518 progeny with CNS neurons, suggesting that the originally transduced progenitor cells were
519 positioned more ventrally (Fig. 6g). These results are consistent with previous spatiotemporal fate
520 mapping of the dorsal neural tube in chick embryos [36].
521
522 Finally, to better understand the signals biasing neural crest clones, we examined the compositional
523 changes associated with CRISPR/Cas9 perturbations of different receptor genes. Our previous
524 work [11] suggested that neural crest cells undergo transcriptional bifurcation that determines
525 autonomic or sensory fates. We hypothesized this bifurcation point would be sensitive to the
526 signaling environment, leading to reciprocal effects on autonomic versus sensory clonal biases.
527 Receptor perturbation caused significant and opposing shifts in clonal composition of autonomic
528 vs somatosensory progenies in the neural crest lineage, suggesting that the corresponding receptors
529 and signaling pathways are indeed involved into defining proportions of fate outcomes (Figure 6h-
530 i). For instance, *Fzd2, Lrp2*, *Lgr4*-crispant clones had reductions in autonomic and enteric biases,
531 accompanied by increases in sensory and glial fates. The perturbation of *Fgfr1* showed exactly the
532 opposite effects on clonal biases and composition. These results expand our understanding of the
533 role of signaling receptors in neural crest specification by demonstrating the existence of a sensory-
534 autonomic clonal lineage bifurcation [11, 51] and its modulation by putative FGF and WNT
535 signaling molecules.
536
537 **Discussion**
538
539 Context-dependent regulation of cell fates is a fundamental aspect of embryogenesis, essential for
540 the proper formation of tissues. Tracing of clonal structures can reveal the spectrum of such fate
541 modulation and provide important clues about the mechanisms involved. Embryogenesis,
542 however, presents a particular challenge for modern lineage tracing approaches, since the large
543 number of generated cell types and the diverse position-specific nature of modulation demands
544 both increased resolution and scale of tracing. A resolution that would enable biological
545 discoveries has been difficult to achieve on an embryo scale, leaving high-throughput tracing
546 studies to focus on specific cell lineages such as hematopoiesis or forebrain development [15, 20].
547
548 Here, we presented an effective practical approach for studying mechanisms underlying fate
549 modulation in an embryo, by integrating multiplexed *in vivo* tracing, computational analysis of
550 transcriptional programs associated with clonal fate biases, and mosaic medium-throughput

551 multiplexed perturbations. The scale of the study, in this case profiling thousands of individually
552 barcoded ectodermal and mesodermal clones across developmental timepoints, is one enabling
553 aspect that allows study of relative levels of continuous versus discrete clonal variation across cell
554 lineages and spatial domains.
555
556 In the past, low-throughput clonal lineage tracing analysis has been used extensively to
557 successfully address hypotheses about cell multipotency and differentiation [44]. Although these
558 studies have established the classic paradigms about lineage splits in embryogenesis, they were
559 not sufficiently powered to investigate the continual variation of clonal fates. Even though
560 different developing structures are often made from the same cell types [52, 53], these structures
561 acquire distinct shapes and functions due to the fine regulation of cell fate proportions and spatial
562 positioning of similar cell progenitor populations [54, 55], developing within the same cell lineages
563 [56]. Insights into the continual variation of clonal structures help to explain how these
564 morphogenesis-related differences gradually arise in a coordinated and context-aware manner.
565
566 Even state-of-the-art high-throughput lineage tracing methods still sample only a small fraction of
567 cells in the embryo. The sparseness of the resulting clonal data represents a major challenge for
568 subsequent analysis. This has prompted us to develop a machine learning approach – clone2vec –
569 which represents the second enabling aspect of our study. Starting with a continuous representation
570 of expression state, clone2vec embeds sparse clonal data into a low dimensional space where
571 clones with similar or compatible fate profiles are grouped together. In contrast to the methods
572 aimed at analysis of factors associated with specific cell fate choices (CoSpar [57] or CellRank
573 [58]), clone2vec aims to provide an effective overview of clonal variation, enabling exploration of
574 recurrent patterns and subsequent analysis of associated molecular features. The continuous
575 representation of the expression space enables a shift away from a cell type-centric perspective of
576 clones' classification, allowing for a nuanced description of state patterns in clone distribution.
577 This approach, as we show, can recover complex clonal fate patterns in a manner highly robust to
578 dropouts (Supplementary Note Simulations), and permits generation of new hypotheses. In
579 addition to separating distinct progenitor lineages (*e.g.* mesenchymal and neural), clone2vec
580 embeddings capture continuous clonal variation within these lineages, which is much less
581 understood and is a central point of our study.
582
583 In contrast to tree representations [5-7, 18], clonal embeddings enable analysis of gradual variation
584 of clonal fates and inference of regulatory processes modulating clonal fates and proportions of
585 cell types across spatial locations. Clones within different spatiotemporal contexts display distinct
586 but overlapping regulatory and transcriptional signatures [59]. A good example is the population
587 of mesenchymal mesodermal derivatives which forms a large unstructured cluster of cells in the
588 transcriptional embedding, posing a challenge for the traditional methods of trajectory
589 reconstruction and other analysis. Lineage tracing and clone2vec analysis, however, effectively
590 partitions this population into a series of clonally connected sub-clusters, representing variation of
591 the fundamental differentiation process at different positional or temporal contexts (Fig. 3a). Such
592 nuanced clonal patterns cannot be easily discerned using traditional approaches such as
593 hierarchical clustering (Extended Data Figure 11n). Furthermore, the analysis of transcriptional
594 variation between clonal clusters can reveal more subtle "auxiliary programs" that regulate clonal
595 structures. For instance, we observed position-dependent variation in skeletogenic mesenchyme,
596 where additional programs do not appear to be essential for chondrogenic differentiation, but rather

modulate secondary features such as a probability of transiting into osteogenic fate or maturation dynamics of progenitor cells. Such an approach should be useful for analyzing position-dependent modulation of progenitor potency in other contexts such as during tissue regeneration [60]. Further, distilling core versus auxiliary differentiation programs for more cell types might aid optimization of efficient and parsimonious *ex vivo* differentiation methods needed for regenerative medicine.

In addition to resolving spatial modulation of clonal development, we also observed heterochrony of clonal fate biases by introducing static barcodes at precise developmental stages. This can be contrasted with other methods employing dynamic barcodes [7, 8], 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. Our comparison of cranial neural crest cells traced from two successive time points showed temporal segregation of facial skeletogenic mesenchyme and trigeminal ganglion from the cranial neural crest, which was noted in chick [40] but not yet in mammals, to our knowledge. This finding suggests that temporal windows shaping cell potency are essential for building an integrated head, as neural crest gives rise to structures as distinct as peripheral nerves and facial bones [61]. The diversity and range of facial shapes produced during vertebrate craniofacial morphogenesis likely benefits from an added layer of dynamic temporal control that complements positional cues [62, 63].

The coupling of clonal lineage 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 Hedgehog (via *Ptch1*), results in formation of ectopic clonal clusters in which cell differentiation appears accelerated or retarded, 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 [13, 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 classical models assume that neural crest arises from the neuroepithelium, our findings suggest that a subset of NMPs gives rise to neural crest independently of the CNS, which might represent a novel mechanism of neural crest induction. This implies that axial progenitor pools are more diversified than previously thought, and their regulation will influence the balance between CNS, neural crest, and mesodermal cells during secondary neurulation and body extension. While our results provide

643 a gene signature to distinguish them (*i.e. Tspan18, Pde8a*), further studies will be needed to
644 determine if neural crest cells derived from NCC-MPs bear functional differences from other sacral
645 crest cells, and how differences in NCC-MP behavior affect spinal cord development and function.
646
647 On a larger developmental scale, our analysis reveals a picture of global fate restrictions at the
648 stages of late gastrula and early neurula, which altogether opposes the idea of early plasticity of
649 different germ layers as was suggested from classic experiments in vitro [64]. According to our
650 data, most clones do not span the competence of more than a single germ layer, except for the
651 NMPs and a handful of exceptional clones from the cranial neural crest. This does not rule out that
652 cells from these stages might exhibit different multipotency profiles when subjected to different
653 environments, as in the case of transplantation experiments that can reveal greater cell plasticity
654 [1, 40].
655
656 Overall, we hope the presented approach will be helpful for further studies aimed at systematically
657 dissecting the spatial and temporal correlates of clonal variation in different cell systems. The
658 current clonal atlas, including its interactive interface (browsable online at
659 https://clones2cells.streamlit.app) should provide a useful resource for a wide range of questions
660 related to the formation of boundaries between developing embryonic structures [65]. Similarly,
661 clone2vec analysis should aid in the identification of fine-grained positional and temporal codes.
662 While the Hox code and other crude positional systems provide a general framework for the
663 development of body axes and programming the large body domains [25], the fine-grained signals
664 that guide specific cell behaviors within very local tissues are much harder to study due to the
665 complexity and redundancy of embryonic signaling systems [66]. Clone2vec may assist
666 investigations of how fine-grained positional information informs cell diversity within individual
667 organs like the heart or lungs. This might shed light on other unresolved questions, for instance,
668 the enigma of the anterior-posterior allocation of brain nuclei [67], or finer patterning of
669 mesodermal populations building individual muscles [68]. Further, joining clone2vec clonal
670 analysis with mosaic perturbations enables multiplexing simulated conditions that mimic human
671 genetic abnormalities, including neurodevelopmental disorders and pediatric cancers (i.e.
672 neuroblastoma). This might help reveal clonal compensatory behavior in modelled congenital
673 diseases, as well as in regeneration, potentially revealing clonal mechanisms of embryonic
674 resilience [69].
675
676 Despite the potential, we realize the limitations of the study. The presented multiplexed tracing
677 and perturbation approach can be widely applied for focused analysis of specific tissues and
678 compartments, however, the potential to refine these techniques is substantial. Current limitations
679 include the constrained developmental time frames and specific cell types that can be effectively
680 targeted using existing barcoding techniques. Inference of differentiation paths or the state of
681 progenitor cells based on one measured endpoint poses well-known fundamental challenges,
682 particularly when the underlying process is asynchronous and one cannot observe, for example,
683 both progenitor and differentiated cells from the same clone in the snapshot provided by scRNA-
684 seq data. While we show that successive barcoding rounds can help, incorporating dynamic
685 barcoding strategies such as CRISPR scarring [5-8, 70] is likely to improve path reconstruction,
686 and systematic variation of experimental endpoints might assist progenitor state inference.
687
688

**Acknowledgements:**

*General acknowledgements*
We thank Olga Kharchenko for the lovely illustrations. We thank Ruslan Soldatov for the insightful discussions. We thank Daniel Wies, Adrian Martinez Martin, and Polina Kameneva for their technical assistance. We acknowledge the National Genomics Infrastructure, Biomedicum Flow Core, Biomedicum Imaging Core, Advanced Light Microscopy Core at SciLifeLab, and Biomedicum Comparative Medicine for access to their facilities and services during the project.

*Funding*
AGE was supported by the National Institute of Dental and Craniofacial Research via the Ruth L. Kirschstein National Research Service Award (F32DE029662). IA and SE were supported by the ERC Synergy grant (KILL-OR-DIFFERENTIATE), Swedish Research Council, Paradifference Foundation, Bertil Hallsten Research Foundation, Cancer Foundation in Sweden, Knut and Alice Wallenberg Foundation, Austrian Science Fund. P.V.K. was founded by the NSF-14-532-CAREER grant.

*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.

## Methods

*CloneID and gRNA-CloneID library construction*
Barcoded lentiviral libraries were generated based on Ratz et al [15]. Briefly, a lentiviral backbone was constructed containing human elongation factor 1 alpha (EF1a) promoter driving expression of either tandem dimer tomato (tdTomato) or green fluorescent protein fused to histone 2b for nuclear localization (H2B-EGFP). For perturbation libraries, a gRNA scaffold using capture sequence "cs1" within the hairpin loop was inserted to LV-U6-tdTomato-NLS constructs using Gibson assembly. Then, the resulting vector was digested using BsmBI and separately annealed gRNA oligonucleotide sequences (which were chosen from the validated sequences in the Broad Institute's Brie CRISPR knockout library) were cloned into the gRNA scaffold by ligation. The resulting vectors were amplified by transformation by heat shock into ONE-SHOT STBL3 competent cells, isolated using a miniprep, and sequence-verified by Sanger sequencing. Next, control backbones (or equimolar mixtures of gRNA-containing backbones) were used for insertion

735 of an amplified random 30N oligonucleotide library via Gibson assembly. The library was further
736 amplified via electroporation of Endura electrocompetent cells. The barcoded plasmids were then
737 subjected to third-generation lentiviral packaging with a VSV-G envelope and concentrated by
738 ultracentrifugation by KI Virus Tech facility or by GEGTech (Paris, France). The resulting ultra-
739 concentrated viruses had high functional titers greater than $10^9$ TU/ml.
740
741
742
743 *Animal experimentation*
744 All experiments were approved by the Swedish Animal Welfare Body Jordbruksverket De
745 regionala Djurförsöksetiska nämnderna and Stockholms Djurförsöksetiska nämnd, permit number
746 N15907-2019/18314-2021. Animals from non-genetically modified mouse lines CD-1 and
747 C57BL/6 were purchased from Charles River Laboratories and Janvier. Animals were kept in an
748 SPF animal facility with standardized conditions (24 °C, 12:12-h light–dark cycle, normal
749 humidity, food and water ad libitum). CD1 or Cas9-GFP mice were placed together into time-
750 controlled matings, and the next day the female was plug checked – if found positive, this was
751 counted as embryonic day (E) 0.5.
752
753 *Ultrasound guided injections (single injections)*
754 Injections for lentiviral gene delivery were performed according to previously published protocols
755 [14]. A glass capillary was pulled, and the tip was grinded to create an ultra-sharp needle. At
756 embryonic day E7.5-E8.5, the females were anesthetized with isoflurane (4% induction,
757 maintained at 1.5-2.5%) and placed on a heated plate for surgery. Depth of anesthesia was tested
758 through pinching the paw. Eye gel was applied to prevent dehydration. Pain relief e.g.
759 Buprenorphine 0.05-0.1 mg/kg was administered via subcutaneous injection shortly before the first
760 incision. All instruments were autoclaved and disinfected with chlorhexidine alcohol before
761 surgery. The skin was disinfected with chlorhexidine alcohol, and the hair on the abdomen was
762 removed with a shaver or hair removal agent such as Veet. A small incision was placed in the
763 abdomen and the uterus was exposed and pulled through an elastic bottom of a modified petri dish
764 with PBS. The uterus was kept moist with warm NaCl solution (37C). Embryos were visualized
765 using a Vevo2100 ultrasound imaging system. Using the glass capillary, lentivirus (from 23-206
766 nL) was injected through the uterine wall into the amniotic cavity. After the embryos were injected,
767 the uterus was brought back into the abdomen and the incisions in the muscle wall were sewn
768 together with absorbent suture thread, and the skin was closed with suture thread or staples. The
769 mice were taken off isoflurane and allowed to wake up in their own cage on a heating pad under
770 supervision. Postoperative pain relief was given in the form of a subcutaneous injection of
771 Buprenorphine (0.05-0.1 mg/kg). For embryo retrieval, the mother was euthanized via CO2
772 overdose and cervical dislocation, and the embryos were collected at E13.5 and euthanized via
773 decapitation.
774
775 *Ultrasound-guided double injections*
776 To determine temporal aspects of general fate restrictions in vivo, two sequential uterine injections
777 in pregnant CD1 mice were performed. A first injection of 206 nL (corresponding to ~313000
778 predicted unique CloneIDs) pMR532-LV-EGFP-bc lentivirus was done on mouse embryos from
779 embryonic day 7 (TS11). A second injection of 69nL (corresponding to ~72,000 predicted unique
780 CloneIDs) of pMR671-LV-TOM-bc was performed at an interval of 12-24 hours after the first.

781 There was no observable scar tissue formation on the day of the second injection. The ultrasound-
782 guided injection itself took place as described in the above section. The embryos survived without
783 visible developmental delays after the single or double injections. Embryos were euthanized by
784 decapitation at E13.5, and cells from the most highly labeled embryos were harvested for
785 downstream analysis as described above, while the remaining littermates were fixed overnight in
786 paraformaldehyde at 4C for histological analysis of cell labeling distribution. During and after the
787 second injection, the mice were given the following pain relief protocol: 5 mg/kg of the long-
788 acting anesthetic Meloxicam, 2 mg/kg of the local anesthetic Bupivacaine, and 0.1 mg
789 Buprenorphine. We tested this protocol for pain relief after double injections 24 hours apart.
790 Together with the veterinarian, we used 3 females to evaluate the effectiveness and impact of the
791 protocol. We monitored the females for pain and recovery from the procedure, assessed the healing
792 of the surgical wound, and checked the status of the fetuses toward the end of the procedure. All
793 females recovered quickly after the first and second injection (within 10 min of placing the cage
794 on a warm pad, the females were walking around the cage and eating normally). The animals
795 showed minimal signs of discomfort or pain after the procedure, on the injection days as well as
796 the day after.
797
798 *Single cell dissociation*
799 The embryos were removed from the uterus, sacrificed, and placed into a petri dish containing
800 Hank's Balanced Salt Solution (HBSS) on ice. Embryos were screened for expression of tdTomato
801 or EGFP using a benchtop fluorescence stereoscope. Limbs, liver, and the ventral skin were
802 removed. Body regions, primarily lower trunk, middle trunk, and cranial regions, were dissected
803 separately and placed into a small Petri dish containing 2.5 mL of a cell dissociation buffer
804 consisting of 2mg/mL Collagenase P (Roche, Sigma-Aldrich), 1X TrypLE, and 1uL/mL DNASe
805 I, fully dissolved in HBSS (without phenol red, including $Ca^{2+}$ and $Mg^{2+}$). Tissue was chopped
806 finely with scalpel and forceps under control of stereomicroscope and transferred to 15mL falcon
807 tubes. The tissue pieces were triturated 10X using a transfer pipette, and then 10X using a 1mL
808 pipette. Samples were then incubated for 20 minutes at 37 C in a preheated shaking incubator set
809 at 150 RPM, with the tubes tilted to ~60 degrees. During this time every 6-8 minutes, the samples
810 were vortexed and triturated using a 1mL pipette. After the tissue clumps disappeared, 12.5 mL of
811 2% Fetal Bovine Serum (FBS) diluted in ice-cold HBSS (without phenol red, without $Ca^{2+}$ and
812 $Mg^{2+}$, H6648 SIGMA) was added. The samples were centrifuged in a cold centrifuge (4C) for 15
813 minutes at 500 g. The supernatant was carefully removed and the pellet was resuspended in 1mL
814 HBSS + 2% FBS and kept on ice thereafter. After centrifugation and resuspension, the cells were
815 filtered into round-bottom polypropylene test tubes (Falcon) using a cell strainer with a 40
816 micrometer woven polyethylene terephthalate mesh (pluriSelect pluriStrainer SKU 43-10040-40)
817 for downstream cell sorting.
818
819 *FACS (Fluorescence activated cell sorting)*
820 FACSDiva software (BD), connected to FACSAria Fusion or FACSAria III flow cytometers (BD)
821 equipped with 100 micrometer nozzle and 1.5 neutral density filter, were used to isolate the
822 fluorescently labeled cells. During sorting the samples were refrigerated to 4 C and regularly
823 agitated by spinning. The event rate was kept below 4000 events/second. Debris, dead cells and
824 doublets were excluded by gating forward-scatter and side-scatter area versus width. Log EGFP
825 and tdTomato fluorescence was detected by excitation with 488 nm and 561 nm lasers, with
826 emissions filtered by 530/30 nm and 610/20 nm bandpass filters, respectively. FP-negative cells

827 served as an internal negative control for FACS sorting. Sorting precision mode was set to "four-
828 way purity". FP-positive cells were sorted directly into Low-binding tubes containing 50uL of
829 HBSS + 2% FBS. For each sample, all the sorted $FP^+$ cells were centrifuged in a cooled centrifuge
830 at 500g for 5 min and resuspended for cell counting with a hemocytometer. From FACS the
831 proportion of labeled cells was ~5% on average. Because tracing every cell in the embryo across
832 organogenesis would not be feasible (because of the massive number of cells in a single E13.5
833 embryo) we prioritized measuring fewer but more "complete" clones. We observe consistent cell
834 types and clone types across embryos, indicating that a broad range of accessible progenitor types
835 was sampled in all experiments.
836
837 *Single cell transcriptomics*
838 Cells were loaded into the Chromium machine (aiming for ~10,000 cells per reaction) using the
839 V3.1 single-cell reagent kit (10X Genomics). Following GEM creation and cell lysis, cDNA was
840 synthesized by reverse transcription, then amplified by PCR for 12 cycles as per the manufacturer′s
841 protocol (10X Genomics). A portion of the amplified cDNA was used to generate Gene expression
842 libraries and then indexed using the Single Index Plate T Set A. CloneID enrichment from cDNA
843 libraries was additionally performed based on [15] to increase the retrieval of the barcodes by ~5-
844 10%. Briefly, two nested PCR reactions were performed using Q5 error-correcting DNA
845 polymerase to amplify the CloneID-containing region, and a third reaction was used for sample
846 indexing using the Single Index Plate T Set A. In between each reaction, amplicons were purified
847 by 0.8X AMPure beads using a DynaMag Magnetic separator. Indexed GE and CloneID-
848 enrichment libraries were quantified using BioAnalyzer and/or Qubit, and then pooled prior to
849 Illumina sequencing on a NovaSeq 6000 S2-200 v1.5 flow cell with read setup 28-8-0-166.
850
851 *Confetti lineage tracing experiment*
852 Transgenic mouse line R26R-Confetti [37, 54], homozygous for the confetti allele, was crossed
853 with Sox10-CreERT2 or PLP-CreERT2 mice. At embryonic day E9 they were injected with 0.02
854 mg/g tamoxifen (TM) intraperitoneally. Approximately 42 h post TM injection, the pregnant mice
855 were sacrificed and embryos still in the uterus were collected in PBS on ice. Then, under the stereo
856 microscope (ZEISS Discovery V8) and in PBS, the muscle layers of the uterus were ripped
857 carefully with micro-dissecting forceps for the embryos to be extracted. Cre-positive R26R-
858 Confetti embryos were identified on the basis of multicolor fluorescence using a stereo microscope
859 fluorescence adapter. E10.5 R26R-Confetti $Cre^+$ mouse embryos were processed to be analyzed
860 histologically or in whole mount. Embryos were fixed in 4% paraformaldehyde (PFA) solution in
861 phosphate buffer saline (PBS) overnight at 4C on rotation. Afterwards, fixed samples headed for
862 histological analysis were washed three times in 1x PBS for 5 min each and were immersed in
863 30% sucrose in PBS overnight at 4oC on rotation. The next day, they were embedded in Optimal
864 Cutting Temperature (OCT)-compound on dry ice and stored at -20C until cryosectioning. An
865 NX70-Epredia cryostat was used to cut sections with a 14µm thickness which were serially
866 collected onto Superfrost® Plus slides (Thermo Scientific, J1800AMNZ/ground 90°). Prior to
867 storing at -20C, they were left to dry on the bench for 2 hours. Fixed embryos headed for whole
868 mount analysis were washed in PBS before immersed in CUBIC-Reagent I clearing solution for 1
869 day at 37 C, gently shaking. Then, they were carefully washed with PBS at room temperature (RT)
870 prior to being immersed in CUBIC Reagent II clearing solution at RT for approximately 6 h. The
871 embryos should be totally transparent in order to be placed onto glass bottom dishes (MatTek
872 Corporation) for imaging. The imaging media is CUBIC-Reagent II. This procedure was also used

873 for the whole mount imaging short-term E8.5 lentiviral transductions in E10.5 embryos, except
874 the samples were imaged using a light sheet microscope (Zeiss Z.1) equipped with a 5X objective.
875 Zen Black software was used during acquisition, and manual segmentation in Bitplane Imaris
876 software (v9) was used to reveal the clonal labeling patterns beneath the skin.
877
878 *Immunohistochemistry*
879 For histological analysis, tissue sections on slides were submerged in Target Retrieval Solution
880 (DAKO, S1699) for antigen retrieval and brought to boiling point using a microwave oven. The
881 solution was then left to cool down for 40-60 min. Sections were washed in PBS containing 0.1%
882 Tween-20 (PBST). A hydrophobic barrier was created with Super PAP Pen (Invitrogen, 008899)
883 surrounding the tissue. The sections were incubated in a humidified chamber overnight at 4 C with
884 primary antibodies diluted in PBST. The next day, sections were washed in PBST and incubated
885 with secondary antibodies diluted in PBST for 45-90 min. Afterwards, slides were washed with
886 PBST, excessive liquid was removed and the slides were mounted using Mowiol (Merck, 81381)
887 as mounting medium. Primary antibodies that were used against mouse at the indicated dilutions:
888 goat anti-Sox10 (1:800, R&D systems, AF2864), mouse anti-neurofilament medium chain (2H3)
889 (1:100, DSHB, AB_2314897), goat anti-CD31 (1:400, R&D systems, AF3628), mouse anti-
890 Myosin Heavy Chain (MyHC) (1:50, R&D systems, MAB4470), mouse anti-Tyrosine
891 Hydroxylase (TH) (1:1000, Sigma-Aldrich, T2928), chicken anti-GFP (1:500, Aves Labs, GFP-
892 1020), rabbit anti-tRFP (1:500, Invitrogen, PA5-79658), rabbit anti-Sox9 (1:500, Millipore).
893 Primary antibodies were detected with Alexa Fluor 405-, 488-, 555- and 647-conjugated secondary
894 antibodies produced in donkey against mouse, rabbit, goat or chicken antibodies (1:1,000,
895 Molecular Probes, Thermo Fisher Scientific).
896
897 *Microscopy and image analysis*
898 Confocal microscopes were used for the imaging of either tissue sections or whole mount samples.
899 Whole mount imaging of cleared E13.5 transduced embryos was performed using light sheet
900 microscopy on a Zeiss Z.1 microscope using a 5X objective. For the Sox10-CREERT2 Confetti
901 labeling, the ZEISS LSM 980 with Airyscan 2 confocal was used due to its capacity to effectively
902 distinguish the overlapping emissions between GFP and YFP. The images were acquired at a 20x
903 objective and 2056x2056 resolution with the use of a z-stack. The imaging of the sections from
904 lentiviral-injected samples was conducted at ZEISS LSM 800 and 800 Airy with a 20x objective
905 and 1024x1024 resolution. ImageJ 1.53t software was used for the analysis of tissue sections. The
906 different color combinations in Sox10-CREERT2 Confetti sections were counted manually with
907 the "Cell Counter" plugin. The determination of DRG units was based on the number of cells in
908 each DRG per embryo. The lentiviral-injected labeled cells in tissue sections were counted with
909 the Cell Counter plugin in compartments of interest like DRG and NT. Bitplane Imaris version
910 10.0 software was used for the analysis of whole Sox10-CREERT2 Confetti embryos. The four
911 channels were first masked based on the "Surface" to remove background, and a second surface
912 mask was used to distinguish double positive cells.
913
914 *scRNA-seq Data Preprocessing and Clone Identification*
915 In the course of the experiments, one to three libraries were generated for each sample: (1) 10x v3
916 Gene Expression, and (2-3) viral barcode-containing cDNA libraries (two in the case of double
917 injection). Detailed descriptions of each sample, corresponding libraries, and FP sequences are
918 provided in the Github repository of this project (please see Code Availability Statements). Reads

from experiments (1) and (2) were aligned to a chimeric mm39 and FP genome reference using Cell Ranger 7.0.1. Subsequently, Gene Expression matrices underwent additional filtering based on sequencing depth, UMI distribution per cell, genes per cell, and the percentage of mitochondrial expression per cell. Specific thresholds are detailed in the Github repository (Code Availability Statement). Clonal barcodes were identified using an algorithm suggested in [13] with modifications available at https://github.com/serjisa/TREX.modified. Firstly, barcode sequence corrections were performed individually for each cell, enhancing robustness against artificial merging but slightly increasing clonal dropout rates. Secondly, cells were grouped into clones only if they shared identical sets of viral barcodes, precluding the merging of cells with different barcodes. Clonal identification was performed simultaneously on all samples corresponding to the same mouse. Doublet filtering was conducted based on gene expression and clonal information before grouping cells into clones. Scrublet v.0.2.3 [71] with manual thresholds was employed for manual doublet filtering. Additionally, cells with two or more viral barcodes, where such combinations existed in only one cell, were considered doublets. Despite these measures, a few clusters likely representing doublets based on marker expression were observed in the final embedding; these were retained in the analysis and labeled as "Doublets." Reference gene expression embeddings were constructed using scanpy v.1.9.3 [72] in several steps. Gene expression embeddings from all control samples (both head and trunk) were created to identify major cell type domains (Mesenchyme, Neurons, and Other cells). Genes present in fewer than five cells were filtered out. Subsequently, counts per cell were linearly transformed to sum to 10,000 counts per cell, followed by log-transformation of (counts + 1) [73]. The top 3,000 highly variable genes were selected, and PCA was performed on gene-wise scaled expressions. Using the top 30 PCs, dataset horizontal integration was performed with the harmonypy v.0.0.5 package [74]. These corrected PCs were used to construct a kNN graph based on Euclidean distances, followed by community detection with the Leiden algorithm [75] (leidenalg v.0.9.1) and visualization with UMAP [76] (umap-learn v.0.5.3). These processing steps were repeated for each of the larger cell subclusters in the head and trunk individually, with an additional step to mitigate the effects of the cell cycle on clustering results. Cell cycle phases were assigned to each cell as described in [77], and a "reference" was built with cells in interphase. This reference was used for label and coordinate transfer using Symphony [78] (symphonypy v.0.2). The joint latent representation post-assignment was used for further analysis.

*Lineage Relationships and clone2vec Algorithm*

For plotting correlational relationships between cell types, Pearson correlation-based hierarchical clustering was employed. To create clonal latent embeddings, only clones with at least 10 cells were selected. Cells from these clones, termed "whitelisted cells," were used to find 15 nearest neighbors among the whitelisted cells based on batch-corrected Harmony coordinates using the pynndescent v.0.5.7 package. Each whitelisted cell and its 15 nearest neighbors formed a "bag of clones," consisting of 15 pairs of context for the target clone. Using a Skip-Gram architecture [19], a neural network was trained to predict clonal labels of context cells based on the clonal label of the target cell. Clonal labels were one-hot-encoded, and a fully connected neural network with one hidden layer (size = 10) was trained with a negative log-likelihood loss. The matrix describing the linear transformation between the input and hidden layers was used as the resulting embedding, capturing similarities between clones. This embedding enabled the construction of a kNN graph of clones, community detection, and UMAP visualization. Clonal embeddings were constructed for both E7.5 and E8.5 clones simultaneously. Cells with multiple clonal labels from E7.5 and

E8.5 were duplicated prior to "bag of clones" construction to count them twice with different clonal labels. For per-clone gene expression analysis, the expression of each cell within a clone was averaged to obtain the final per-clone expression. For embedding exploratory purposes, PCA was performed on the clone2vec embedding space, and gene expression was used to identify associations with clone2vec PCs. Clonal cluster names were manually assigned by inspection of the cell states and cell types that were present in them, and literature searches for embryonic progenitor cell types that would give rise to each of those cell type ensembles. This information was used in combination with analysis of the gene expression patterns indicating anatomical position, and spatial validation in mouse embryos by mapping clone type-specific gene signatures onto publicly available Stereo-seq datasets.

*Gene Importance Analysis*

To identify genes associated with clonal composition, we built models to predict either clone2vec latent embedding coordinates or cell type proportions within each clone. For this purpose, we used a gradient boosting model, as its applicability for gene regulatory network (GRN) inference has been demonstrated [79]. Specifically, we employed the relatively overfitting-robust CatBoost algorithm [80], following the strategy described below. First, we created pseudo-bulk profiles for each clone within each cell type to mitigate learning of cell type–specific gene expression. To further reduce this effect, we included cell type as a categorical predictor alongside gene expression. Clones were then split into two subsets: a training set (80% of clones) and a validation set (20%) to monitor overfitting during training. The validation set was selected using a geometric sketching approach [81], choosing clones that best represent the clone2vec space. CatBoost regression (for latent coordinate prediction) or classification (for clone composition prediction) was applied using MultiRMSE or MultiCrossEntropy loss functions on both the training and validation sets. Training was stopped if the validation loss did not improve for more than 100 iterations, and the best model was selected for further analysis. The final model was used to estimate feature importance for each prediction across the dataset using Shapley values [82, 83]. To quantify the overall importance of each gene in clonal behavior, we averaged the L2-norm of the SHAP vectors from the clone2vec model across all generated pseudo-bulks. To investigate gene associations with cell types, we extracted SHAP value modules from the cell type model and assigned them the sign of their correlation with gene expression. This sign correction allowed us to distinguish between positive and negative markers of cell type proportions. Thus, the clone2vec model captures the overall variance in clonal behavior, while the cell type model provides a mechanistic interpretation of gene effects at the cell type level. Both models were trained using the same train/validation split for consistency. For visualization, sign-corrected SHAP magnitudes from each cell type can be min-max normalized with respect to their sign.

*Cartilage Trajectory Analysis*

To reconstruct differentiation trajectories within individual clonal clusters, cells were re-embedded, and clusters linked to the cartilage/bone trajectory were selected from clonal clusters 2 and 8 (clustering resolution = 1, see Extended Data Fig. 18 for parameter sweeps of clustering resolution). Velocity analysis with UniTVelo (v.0.2.5.2) confirmed the differentiation directionality within the selected cells. Bifurcation points were identified using the scFates package (v.1.0.7) [84]. Pseudotime was inferred using the diffusion pseudotime approach, and cells with pseudotime value less than 0.3 were considered as progenitors [85]. Alignment between

two differentiation trajectories was performed with the DTW Python package (v.1.3.0) [86, 87]. The significance of the association between gene expression and pseudotime was determined using GAM fitting and LRT, comparing models with and without pseudotime-associated splines. Differences in associations with pseudotime across trajectories were identified by comparing GAMs with and without clonal cluster assignment as a covariate. Genes significantly associated with pseudotime ($p < 10^{-25}$) and not differing between clonal clusters (FDR > 0.05) were considered "common", at the same time genes with significant lineage association (FDR $< 10^{-5}$) and fold-change > 1 between lineages were considered as "unique". Bone-related signature was gotten by differential expression analysis between cartilage and bone in clonal cluster 2 (top-20 highly expressed in cartilage trajectory genes with FDR $< 10^{-20}$ and logFC > 4) using Student's t-test. Analysis of "early" genes was performed with standard scFates pipeline [84]. Transcription factor (TF) activity inference was conducted using pyscenic (v.0.12.1) [30] with arboreto (v.0.1.6). Differential activity analysis between clonal clusters followed the same procedure. Graph of the potential regulation was gotten from cisTarget step of pyscenic protocol. In all cases signature scores were calculated with sc.tl.score_genes scanpy function.

*Spatial Validation of Gene Expression Signatures*

Stereo-seq data from [88] was used to identify the spatial localization of gene expression signatures. Gene signatures were identified using Student's t-test between groups of cells (e.g., between different clonal clusters or within the same cell type across different clonal clusters) with a minimum log2FC of 2. Expression of gene signatures within one cell type was visualized as the product of gene and cell type signatures. Hox-score was calculated as a min-max normalized weighted sum of expressions of the following Hox-genes:

$$\textit{Hox score}_{\text{raw}} = (1 \times \text{expr}(\textit{Hoxb4}) + 2 \times \text{expr}(\textit{Hoxb5}) + 3 \times \text{expr}(\textit{Hoxc6}) + 4 \times \text{expr}(\textit{Hoxa7}) + 5 \times \text{expr}(\textit{Hoxc8}) + 6 \times \text{expr}(\textit{Hoxd9}) + 7 \times \text{expr}(\textit{Hoxc10}) + 8 \times \text{expr}(\textit{Hoxd11}) + 9 \times \text{expr}(\textit{Hoxc11}) + 10 \times \text{expr}(\textit{Hoxc12}) + 11 \times \text{expr}(\textit{Hoxc13})) / 66;$$

$$\textit{Hox score} = (\textit{Hox score}_{\text{raw}} - \min(\textit{Hox score}_{\text{raw}})) / (\max(\textit{Hox score}_{\text{raw}}) - \min(\textit{Hox score}_{\text{raw}})).$$

*Perturbation Analysis*

Perturbational screening experiments were mapped onto the control reference via Symphony, as previously described (separately for cycling and non-cycling cells). Each clone from the perturbational experiment was assigned to a gRNA based on the most abundant gRNA combination within the clone. Clones with multiple gRNAs were excluded from further analysis. New clonal embeddings were constructed as described in the Lineage Relationships and Clone2vec Algorithm section. Differential compositional testing between clones in different clonal clusters was conducted using scCODA [89] to analyze clone composition in the entire clonal embedding rather than individual cells within different cell types. An independent embedding focusing on NC-related clones (with at least three cells) was created using both control and perturbed cells. Compositional testing was performed for perturbations with multiple replicates using beta regression and LRT with FDR multiple testing correction.

## **Code 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)

## Data Availability

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/15397247). 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)

## References


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. Soldatov, R., et al., *Spatiotemporal structure of cell fate decisions in murine neural crest.* Science, 2019. **364**(6444).
12. 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.
13. Tzouanacou, E., et al., *Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis.* Dev Cell, 2009. **17**(3): p. 365-76.

14. Mangold, K., et al., *Highly efficient manipulation of nervous system gene expression with NEPTUNE.* Cell Rep Methods, 2021. **1**(4).
15. 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.
16. Kameneva, P., et al., *Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin.* Nat Genet, 2021.
17. Roe, T., et al., *Integration of murine leukemia virus DNA depends on mitosis.* EMBO J, 1993. **12**(5): p. 2099-108.
18. Weinreb, C. and A.M. Klein, *Lineage reconstruction from clonal correlations.* Proc Natl Acad Sci U S A, 2020. **117**(29): p. 17041-17048.
19. Mikolov, T., et al., *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781, 2013.
20. Weinreb, C., et al., *Lineage tracing on transcriptional landscapes links state to fate during differentiation.* Science, 2020. **367**(6479).
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. Hall, B.K. and T. Miyake, *All for one and one for all: condensations and the initiation of skeletal development.* Bioessays, 2000. **22**(2): p. 138-47.
25. Carrasco, A.E. and S.L. López, *The puzzle of Hox genes.* Int J Dev Biol, 1994. **38**(4): p. 558-64.
26. McIntyre, D.C., et al., *Hox patterning of the vertebrate rib cage.* Development, 2007. **134**(16): p. 2981-2989.
27. 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).
28. 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.
29. 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.
30. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering.* Nature Methods, 2017. **14**(11): p. 1083-1086.
31. Bradaschia-Correa, V., et al., *Hox gene expression determines cell fate of adult periosteal stem/progenitor cells.* Sci Rep, 2019. **9**(1): p. 5043.
32. Depew, M.J., et al., *Reassessing the Dlx code: the genetic regulation of branchial arch skeletal pattern and development.* J Anat, 2005. **207**(5): p. 501-61.
33. Xu, P., et al., *Fox proteins are modular competency factors for facial cartilage and tooth specification.* Development, 2018. **145**(12).
34. 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).
35. 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.

1148 36. Krispin, S., et al., *Evidence for a dynamic spatiotemporal fate map and early fate*
1149 *restrictions of premigratory avian neural crest.* Development, 2010. **137**(4): p. 585-95.
1150 37. Kaucka, M., et al., *Analysis of neural crest-derived clones reveals novel aspects of facial*
1151 *development.* Sci Adv, 2016. **2**(8): p. e1600060.
1152 38. Dupin, E., et al., *The issue of the multipotency of the neural crest cells.* Dev Biol, 2018.
1153 **444 Suppl 1**: p. S47-S59.
1154 39. Weston, J.A. and J.P. Thiery, *Pentimento: Neural Crest and the origin of mesectoderm.*
1155 Developmental Biology, 2015. **401**(1): p. 37-61.
1156 40. Baker, C.V., et al., *Early- and late-migrating cranial neural crest cell populations have*
1157 *equivalent developmental potential in vivo.* Development, 1997. **124**(16): p. 3077-3087.
1158 41. Platt, Randall J., et al., *CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer*
1159 *Modeling.* Cell, 2014. **159**(2): p. 440-455.
1160 42. Cohen, M., et al., *Ptch1 and Gli regulate Shh signalling dynamics via multiple*
1161 *mechanisms.* Nature Communications, 2015. **6**(1): p. 6709.
1162 43. Bronner-Fraser, M. and S.E. Fraser, *Cell lineage analysis reveals multipotency of some*
1163 *avian neural crest cells.* Nature, 1988. **335**(6186): p. 161-4.
1164 44. Baggiolini, A., et al., *Premigratory and migratory neural crest cells are multipotent in*
1165 *vivo.* Cell Stem Cell, 2015. **16**(3): p. 314-22.
1166 45. Krispin, S., E. Nitzan, and C. Kalcheim, *The dorsal neural tube: a dynamic setting for*
1167 *cell fate decisions.* Dev Neurobiol, 2010. **70**(12): p. 796-812.
1168 46. Tang, W., et al., *Clonal analysis and dynamic imaging identify multipotency of individual*
1169 *Gallus gallus caudal hindbrain neural crest cells toward cardiac and enteric fates.* Nat
1170 Commun, 2021. **12**(1): p. 1894.
1171 47. Yu, V.W.C., et al., *Epigenetic Memory Underlies Cell-Autonomous Heterogeneous*
1172 *Behavior of Hematopoietic Stem Cells.* Cell, 2017. **168**(5): p. 944-945.
1173 48. D'Autréaux, F., et al., *Hand2 is necessary for terminal differentiation of enteric neurons*
1174 *from crest-derived precursors but not for their migration into the gut or for formation of*
1175 *glia.* Development, 2007. **134**(12): p. 2237-2249.
1176 49. Montalva, L., et al., *Hirschsprung disease.* Nature Reviews Disease Primers, 2023. **9**(1):
1177 p. 54.
1178 50. Kam, M.K.M. and V.C.H. Lui, *Roles of Hoxb5 in the development of vagal and trunk*
1179 *neural crest cells.* Development, Growth & Differentiation, 2015. **57**(2): p. 158-168.
1180 51. Sutton, G., R.N. Kelsh, and S. Scholpp, *Review: The Role of Wnt/β-Catenin Signalling in*
1181 *Neural Crest Development in Zebrafish.* Frontiers in Cell and Developmental Biology,
1182 2021. **9**.
1183 52. Noden, D.M. and P.A. Trainor, *Relations and interactions between cranial mesoderm*
1184 *and neural crest populations.* Journal of Anatomy, 2005. **207**(5): p. 575-601.
1185 53. Wang, M., et al., *Distinct Gene Regulatory Dynamics Drive Skeletogenic Cell Fate*
1186 *Convergence During Vertebrate Embryogenesis.* bioRxiv, 2024: p. 2024.03.26.586769.
1187 54. Kaucka, M., et al., *Oriented clonal cell dynamics enables accurate growth and shaping of*
1188 *vertebrate cartilage.* eLife, 2017. **6**: p. e25902.
1189 55. Kaucka, M., et al., *Signals from the brain and olfactory epithelium control shaping of the*
1190 *mammalian nasal capsule cartilage.* Elife, 2018. **7**.
1191 56. Hirasawa, T. and S. Kuratani, *Evolution of the vertebrate skeleton: morphology,*
1192 *embryology, and development.* Zoological Lett, 2015. **1**: p. 2.

1193 57. Wang, S.W., et al., *CoSpar identifies early cell fate biases from single-cell transcriptomic*
1194 *and lineage information.* Nat Biotechnol, 2022. **40**(7): p. 1066-1074.
1195 58. Weiler, P., et al., *CellRank 2: unified fate mapping in multiview single-cell data.* Nat
1196 Methods, 2024. **21**(7): p. 1196-1205.
1197 59. Bastide, S., et al., *TATTOO-seq delineates spatial and cell type–specific regulatory*
1198 *programs in the developing limb.* Science Advances. **8**(50): p. eadd0695.
1199 60. Otsuki, L. and E.M. Tanaka, *Positional Memory in Vertebrate Regeneration: A Century's*
1200 *Insights from the Salamander Limb.* Cold Spring Harb Perspect Biol, 2022. **14**(6).
1201 61. Adameyko, I. and K. Fried, *The Nervous System Orchestrates and Integrates*
1202 *Craniofacial Development: A Review.* Frontiers in Physiology, 2016. **7**(49).
1203 62. Usui, K. and M. Tokita, *Creating diversity in mammalian facial morphology: a review of*
1204 *potential developmental mechanisms.* EvoDevo, 2018. **9**(1): p. 15.
1205 63. Chen, Y.C. and N. Konstantinides, *Integration of Spatial and Temporal Patterning in the*
1206 *Invertebrate and Vertebrate Nervous System.* Front Neurosci, 2022. **16**: p. 854422.
1207 64. Waddington, C.H. and J. Taylor, *Conversion of Presumptive Ectoderm to Mesoderm in*
1208 *the Chick.* Journal of Experimental Biology, 1937. **14**(3): p. 335-339.
1209 65. Dahmann, C., A.C. Oates, and M. Brand, *Boundary formation and maintenance in tissue*
1210 *development.* Nature Reviews Genetics, 2011. **12**(1): p. 43-55.
1211 66. Tkačik, G. and T. Gregor, *The many bits of positional information.* Development, 2021.
1212 **148**(2): p. dev176065.
1213 67. Lipovsek, M. and R.J. Wingate, *Conserved and divergent development of brainstem*
1214 *vestibular and auditory nuclei.* Elife, 2018. **7**.
1215 68. Grimaldi, A. and S. Tajbakhsh, *Diversity in cranial muscles: Origins and developmental*
1216 *programs.* Curr Opin Cell Biol, 2021. **73**: p. 110-116.
1217 69. Solini, G.E., et al., *Xenopus embryos show a compensatory response following*
1218 *perturbation of the Notch signaling pathway.* Developmental Biology, 2020. **460**(2): p.
1219 99-107.
1220 70. Alemany, A., et al., *Whole-organism clone tracing using single-cell sequencing.* Nature,
1221 2018. **556**(7699): p. 108-112.
1222 71. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell*
1223 *Doublets in Single-Cell Transcriptomic Data.* Cell Systems, 2019. **8**(4): p. 281-291.e9.
1224 72. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression*
1225 *data analysis.* Genome Biology, 2018. **19**(1): p. 15.
1226 73. Ahlmann-Eltze, C. and W. Huber, *Comparison of transformations for single-cell RNA-*
1227 *seq data.* Nat Methods, 2023. **20**(5): p. 665-672.
1228 74. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with*
1229 *Harmony.* Nature Methods, 2019. **16**(12): p. 1289-1296.
1230 75. Traag, V.A., L. Waltman, and N.J. van Eck, *From Louvain to Leiden: guaranteeing well-*
1231 *connected communities.* Scientific Reports, 2019. **9**(1): p. 5233.
1232 76. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and*
1233 *projection for dimension reduction.* arXiv preprint arXiv:1802.03426, 2018.
1234 77. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat
1235 Biotechnol, 2015. **33**(5): p. 495-502.
1236 78. Kang, J.B., et al., *Efficient and precise single-cell reference atlas mapping with*
1237 *Symphony.* Nat Commun, 2021. **12**(1): p. 5890.

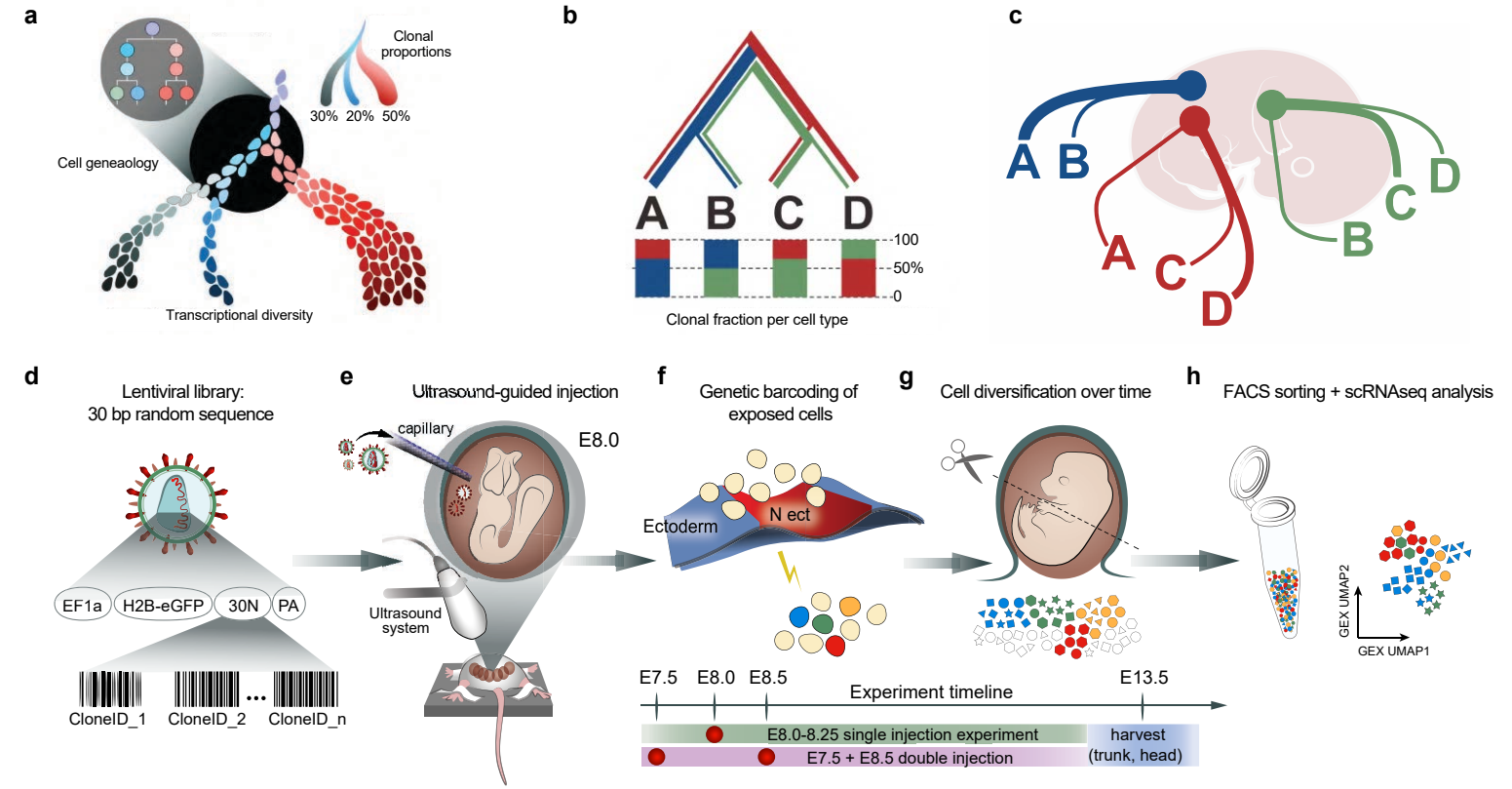
79. Moerman, T., et al., *GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks.* Bioinformatics, 2019. **35**(12): p. 2159-2161.
80. Dorogush, A., *CatBoost: gradient boosting with categorical features support*, V. Ershov, Editor. 2018: Arxiv. p. arXiv:1810.11363.
81. Hie, B., et al., *Geometric Sketching Compactly Summarizes the Single-Cell Transcriptomic Landscape.* Cell Syst, 2019. **8**(6): p. 483-493 e7.
82. Lundberg, S., *Consistent Individualized Feature Attribution for Tree Ensembles*, G. Erion, Editor. 2019. p. arXiv:1802.03888.
83. Lundberg, S., *A unified approach to interpreting model predictions*, S.-I. Lee, Editor. 2017: arXiv. p. arXiv:1705.07874.
84. Faure, L., et al., *scFates: a scalable python package for advanced pseudotime and bifurcation analysis from single-cell data.* Bioinformatics, 2023. **39**(1).
85. Haghverdi, L., et al., *Diffusion pseudotime robustly reconstructs lineage branching.* Nat Methods, 2016. **13**(10): p. 845-8.
86. Giorgino, T., *Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.* Journal of Statistical Software, 2009. **31**(7): p. 1 - 24.
87. Alpert, A., et al., *Alignment of single-cell trajectories to compare cellular expression dynamics.* Nat Methods, 2018. **15**(4): p. 267-270.
88. Chen, A., et al., *Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays.* Cell, 2022. **185**(10): p. 1777-1792.e21.
89. Büttner, M., et al., *scCODA is a Bayesian model for compositional single-cell data analysis.* Nat Commun, 2021. **12**(1): p. 6876.

Figure 1

Clonal composition is driven by proliferation and differentiation

Lineage trees are superpositions of clonal fate biases

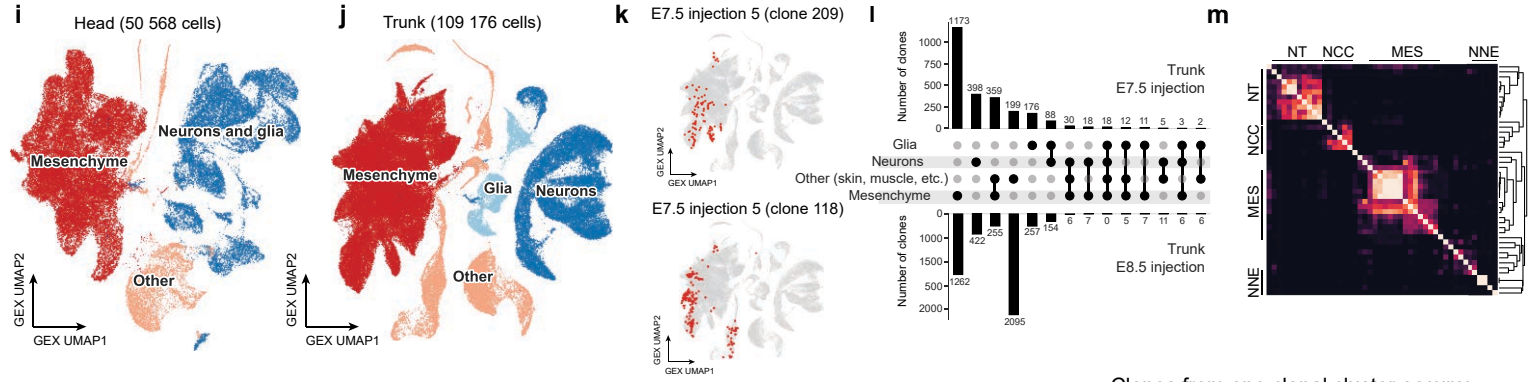
How does anatomical location shape clonal fate biases?




Single-cell transcriptomics analysis of barcoded body regions

Frequency of clonal behaviors observed in the dataset

Cell type lineage correlation




Clone2vec learns clonal relationships from transcriptional context

Clones from one clonal cluster occupy similar transcriptional niches

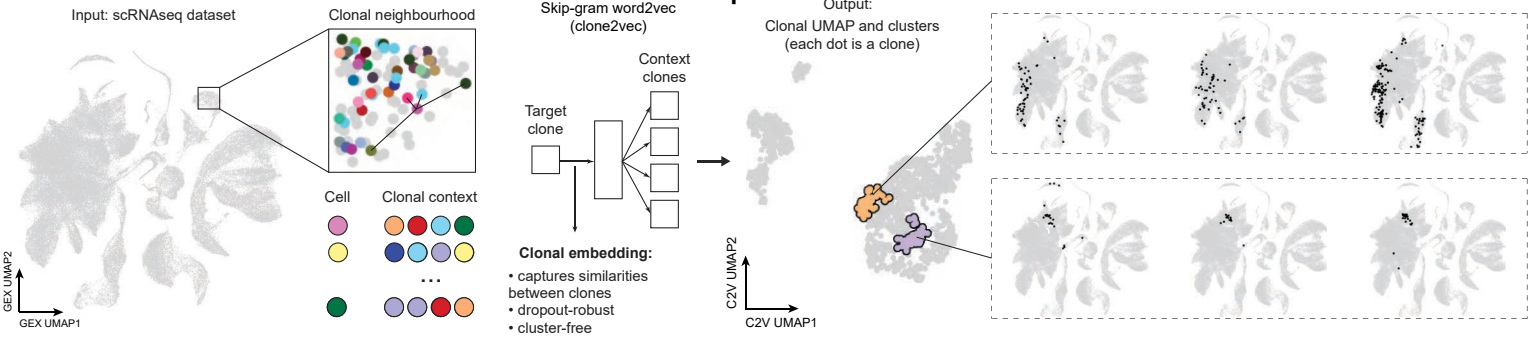


**Figure 1. Machine learning-based profiling of clonal behavior in mouse embryos.**

**(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 a population of three clones, collectively generate a fate spectrum (full repertoire of cell types).
**(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 into 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) as well as some mesoderm.
**(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

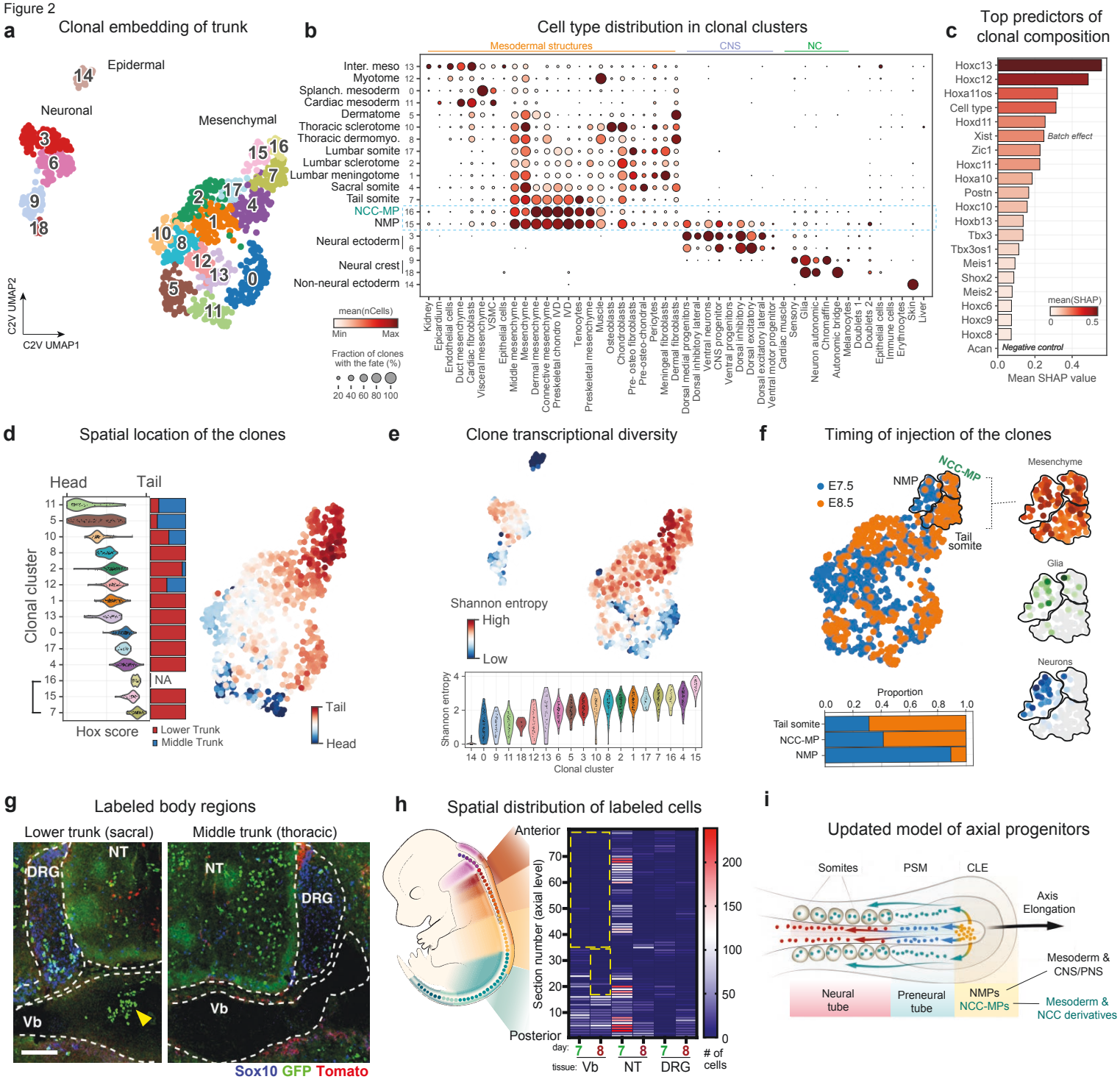

a
Clonal embedding of trunk
Epidermal
Neuronal
Mesenchymal
C2V UMAP2
C2V UMAP1
b
Cell type distribution in clonal clusters
Mesodermal structures
CNS
NC
Inter. meso
Myotome
Splanch. mesoderm
Cardiac mesoderm
Dermatome
Thoracic sclerotome
Thoracic dermomyo.
Lumbar somite
Lumbar sclerotome
Lumbar meningotome
Sacral somite
Tail somite
NCC-MP
NMP
Neural ectoderm
Neural crest
Non-neural ectoderm
mean(nCells)
Min
Max
Fraction of clones with the fate (%)
20 40 60 80 100
c
Top predictors of clonal composition
Hoxc13
Hoxc12
Hoxa11os
Cell type
Hoxd11
Xist
Batch effect
Zic1
Hoxc11
Hoxa10
Postn
Hoxc10
Hoxb13
Tbx3
Tbx3os1
Meis1
Shox2
Meis2
Hoxc6
Hoxc9
Hoxc8
Acan
Negative control
mean(SHAP)
Mean SHAP value
d
Spatial location of the clones
Head
Tail
Clonal cluster
Hox score
NA
Lower Trunk
Middle Trunk
e
Clone transcriptional diversity
Shannon entropy
High
Low
Clonal cluster
f
Timing of injection of the clones
E7.5
E8.5
NCC-MP
NMP
Tail somite
Mesenchyme
Glia
Neurons
Proportion
g
Labeled body regions
Lower trunk (sacral)
Middle trunk (thoracic)
NT
DRG
Vb
Sox10 GFP Tomato
h
Spatial distribution of labeled cells
Anterior
Posterior
Section number (axial level)
day:
tissue: Vb NT DRG
# of cells
i
Updated model of axial progenitors
Somites
PSM
CLE
Axis Elongation
Mesoderm & CNS/PNS
Neural tube
Preneural tube
NMPs
NCC-MPs
Mesoderm & NCC derivatives

**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), labeled according to the fate biases predominantly observed in each of the three main super-clusters (corresponding to neural ectoderm, non-neural ectoderm, and mesodermal derivatives).
**(b)** Dot plot showing cell type composition among clonal clusters. Color represents average cells within each fate, per clone, within that cluster; dot size represents the fraction of the clones in the cluster containing that fate.
**(c)** Elbow plot showing the top twenty most important genes for predicting cell type composition of clones using the CatBoost model, color coded by the Shapley value (feature importance) for each gene for that model.
**(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)** Top, feature plot showing the number of fates per clone. Bottom, violin plot showing the distribution of Shannon entropy (here used as a metric for transcriptional diversity) per clone within each clonal cluster.
**(f)** Top left, mesenchyme embedding color-coded by injection time. Insets highlight the multipotent neuronal-mesenchymal clones in tail mesenchyme (see Extended Data Figures for validation of NCC-MP clones). Bottom: the composition of each clonal cluster by each injection time.
**(g)** 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. Vb; vertebrae. DRG; dorsal root ganglion. NT; neural tube.
**(h)** Heat map displaying the labeled $GFP^+$ and $Tomato^+$ cells within the vertebra, neural tube, and sensory ganglia per section, ordered by axial position. Dotted box highlights the absence of labeled vertebral chondrocytes from anterior sections.
**(i)** 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

Fate biases of trunk mesenchyme clonal clusters (spatially-distinct somite-derived lineages)

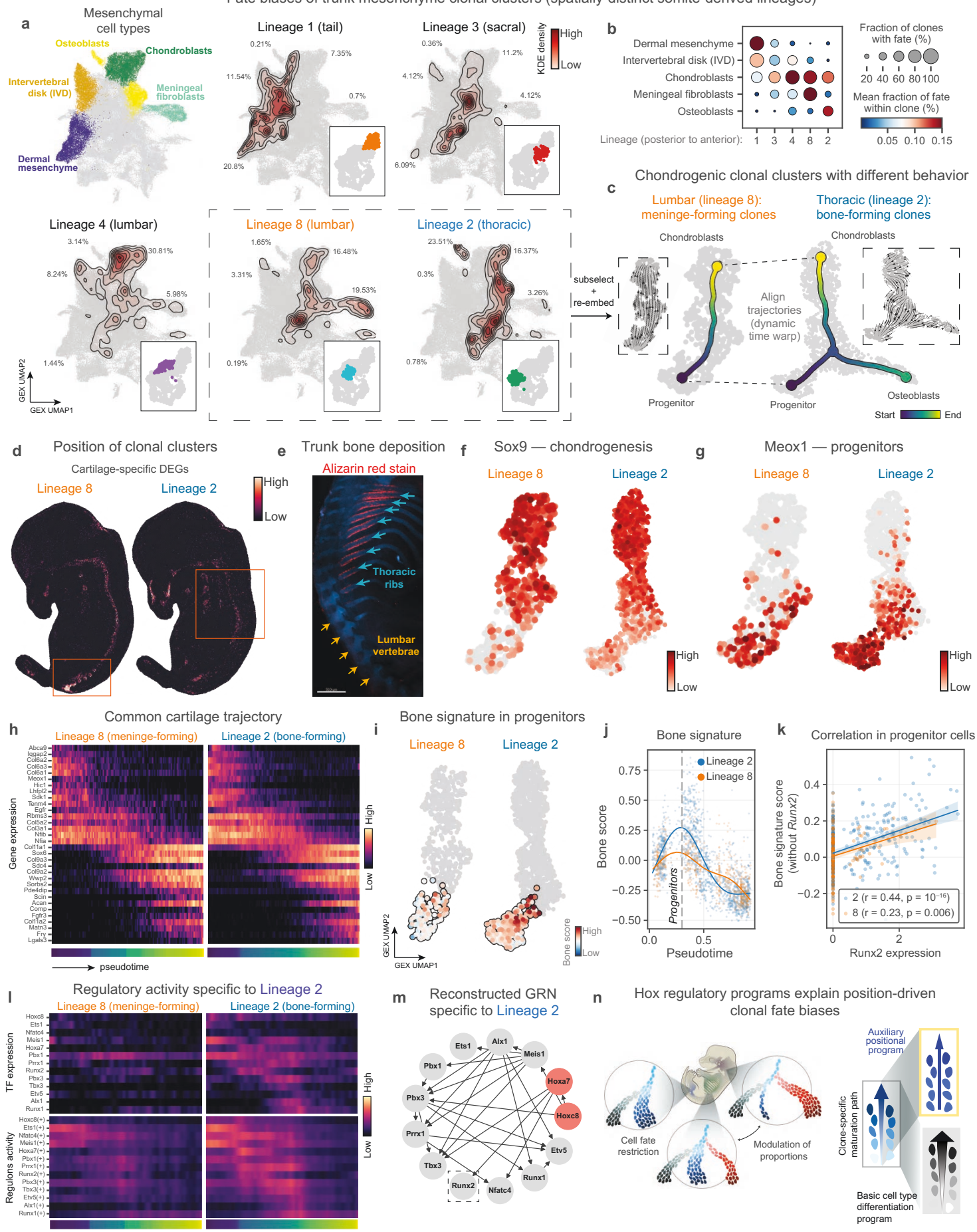


**Figure 3. Position-specific programs are superimposed on common paths of cell maturation.**

**(a)** Kernel density estimate (KDE) 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 of the clonal cluster on the mesenchymal portion of the clone2vec embedding (Figure 2a).
**(b)** A dot plot summarizing the fate distribution of each lineage shown in (a), ordered by their position along the body axis.
**(c)** Trajectory inference and alignment of the chondrocyte-meningeal clone type (lineage 8) and the osteochondral clone type (lineage 2) cells by dynamic time warp (DTW). Dark blue represents the origin of each trajectory (mesenchymal progenitors). Dotted insets show RNA velocity arrows denoting the directionality of cell maturation.
**(d)** E13.5 mouse Stereoseq feature plot of gene scores derived from transcriptional profiles of chondrocytes within fibro-chondral (left) versus osteochondral (right) clone types.
**(e)** Confocal fluorescence micrograph of a skeletal preparation of the trunk of an E14.5 embryo. Blue indicates cartilage; red indicates mineralized bone tissue.
**(f-g)** Feature plots displaying *Sox9* (f) and Meox1 (g) expression per cell within the selected chondrocyte-restricted (left) and bipotent osteochondral (right) clone types.
**(h)** Heatmap displaying common gene expression pattern along the chondrogenic trajectories of the restricted (left) and bipotent (right) clone types.
**(i)** Expression of bone-related genes in progenitors (pseudotime < 0.3) cells.
**(j)** 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.
**(k)** Correlation between *Runx2* and bone-related genes' expression in progenitor cells from these two clone types.
**(l)** Heatmap displaying patterns of regulon activity (top) and transcription factor expression (bottom) unique to osteochondrogenic clones.
**(m)** Gene regulatory networks predicted within this lineage.
**(n)** Model of positional regulation of clone development, in which cell differentiation dynamics (driven by a core set of cell-type-specific genes) is under the influence of position-specific auxiliary programs (such as Hox genes).

Figure 4

# Mapping of craniofacial clonal cluster signatures to transcriptional and anatomical space

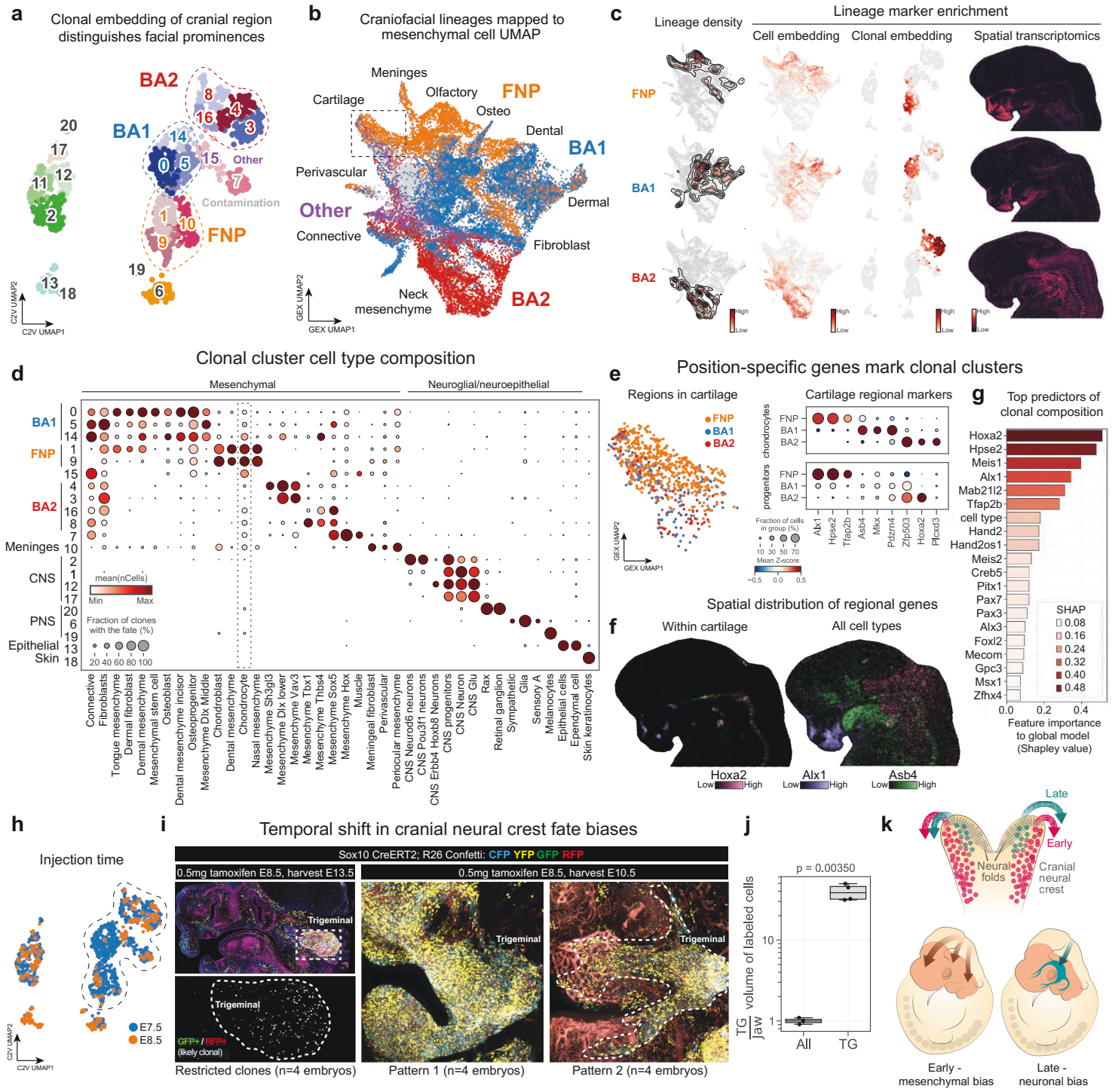


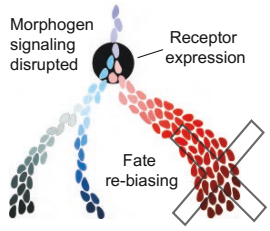
**Figure 4. Collective multipotency of the cranial neural crest cell population.**

**(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 each 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)** Elbow plot showing the Shapley value for the top 20 most important features (overall cell type ranked seventh) in a CatBoost model trained to predict cell type composition and position of facial clones.
**(h)** Embedding in (a) color-coded by timing of injection.
**(i-j)** 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 in (j).
**(k)** Model of evolving fate biases in neural crest based on the timing of emigration from the neural tube.

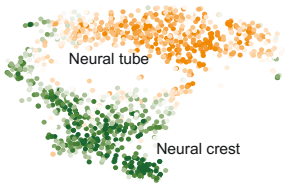
Figure 5

## Perturb-TREX experimental design

**a** Experimental receptor perturbation

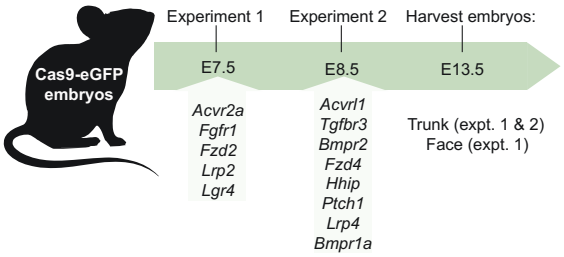



Criteria for target gene selection



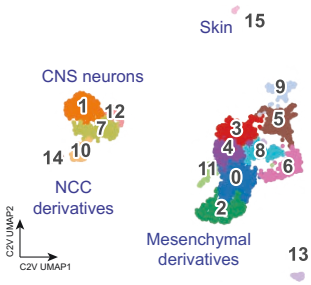

Major signaling pathway receptors
Expressed by progenitor cells
Required for normal development

Inject multiplexed barcoded lentiviral perturbation library

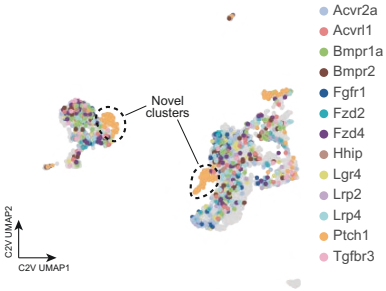



## Analysis of the perturbed clonal landscapes in trunk

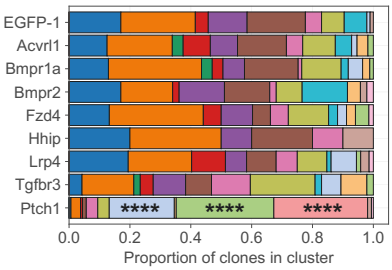
**b** Clonal clusters (lineages)



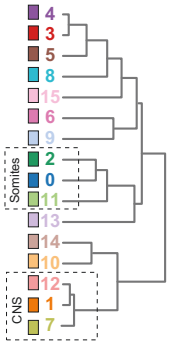

**c** Perturbations




**d** Distribution of crispants among lineages

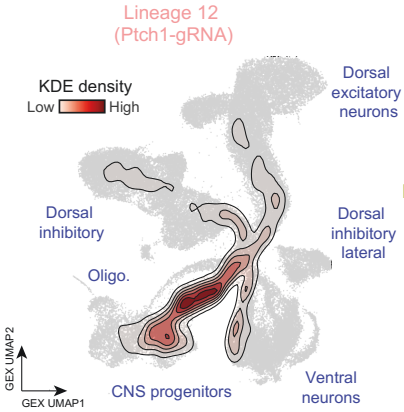



**e** Similarity of clonal clusters (via fate biases)

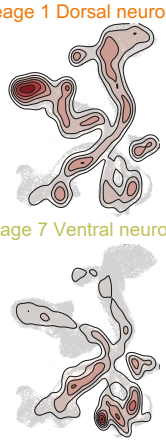



## Ptch1-gRNA arrests CNS cell maturation, resulting in larger clones

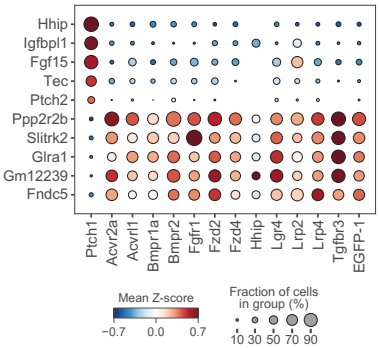
**f** Trunk CNS neurons embedding



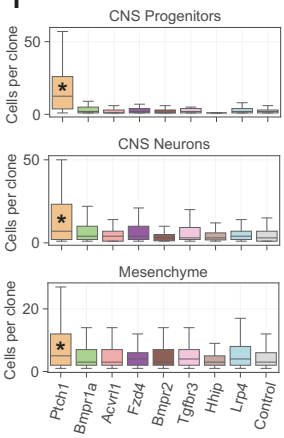

**g** Lineage 1 Dorsal neurons / Lineage 7 Ventral neurons



**h** Top-10 DEGs in CNS progenitors

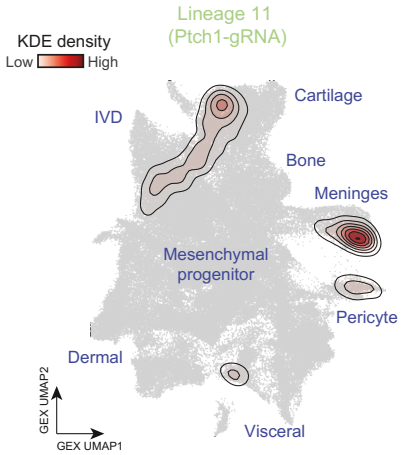



**i** Effect of gRNA on clone size

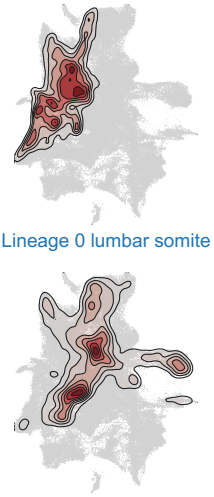



## Ptch1-gRNA disrupts the somitic fate gradient

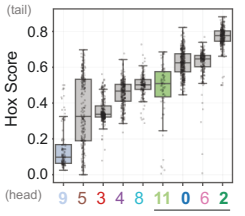
**j** Trunk mesenchyme embedding



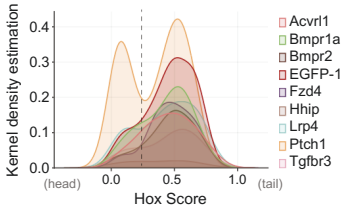

**k** Lineage 2 tail somite / Lineage 0 lumbar somite



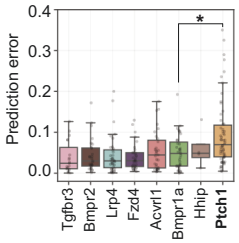
**l** Lineage positional identity




**m** gRNA effect on mesenchyme patterning




**n** Hox 'mismatch score'




**o** Prediction of Hox score from clone2vec (lower trunk)

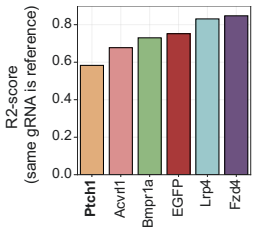


**Figure 5. Novel clonal behaviors caused by genetic perturbations.**

**(a)** Experimental design.
**(b)** Clone2vec UMAP embedding of E13.5 control and perturbed barcoded trunk cells, labeled and color-coded by clusters (clustering resolution 1).
**(c)** Same embedding as in (b), instead color-coded by perturbation. Dotted lines indicate clonal clusters highly enriched with Ptch1 gRNA+ clones. Note cluster 9 is also highly enriched with Ptch1 gRNA+ clones.
**(d)** Bar chart showing the proportion of clone types constituting each gRNA+ population, statistical significance was calculated with scCODA.

**(e)** Hierarchical clustering of clone types based on Euclidean distance between centroids in clone2vec space. Dotted boxes indicate the novel clone types.

**(f-g)** Density plot of the Ptch1-enriched neuronal clone type (f), and the two clone types most closely related to it (g), based on the clustering in (e).
**(h)** Dot plot displaying the top differentially expressed genes by the Ptch1 gRNA+ CNS neural progenitors.
**(i)** 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]).
**(j-k)** Density plot of the novel Ptch1-enriched mesenchymal clone type (j) and of the two clone types most closely related to it (k), based on the clustering in (e).

**(l)** Bar and whisker plot showing positional identity of each clone type, quantified by the Hox genes present in each clone. Higher score indicates a position closer to the tail. The label for each clone type is color coded according to the color scheme in (b) and (e). Blue and green bars represent Ptch1-gRNA-enriched mesenchymal clone types (9 and 11, respectively).

**(m)** Kernel density estimation plot showing the distribution of Hox scores per clone across different perturbation conditions. Note the bi-modal distribution in Ptch1 crispant clones.
**(n)** Error rate in predicting Hox score from nearest-neighbor clones (from clone2vec space). Low values indicate that the clones of a given condition consistently have a similar expression profile for Hox genes relative to neighboring clones, whereas high values indicate clones having different positional identities compared to cells with similar fate biases.

**(o)** R2-score (proportion of variance) for each gRNA, reflecting the Hox score predictability of gRNA+ clones, based on nearest neighbor clones (from clone2vec space) bearing the same gRNA. High score indicates high predictability (low variance) in positional identity within the condition, for clones with similar clonal fate biases. Note the low score (high variance) within Ptch1 clones.

Figure 6

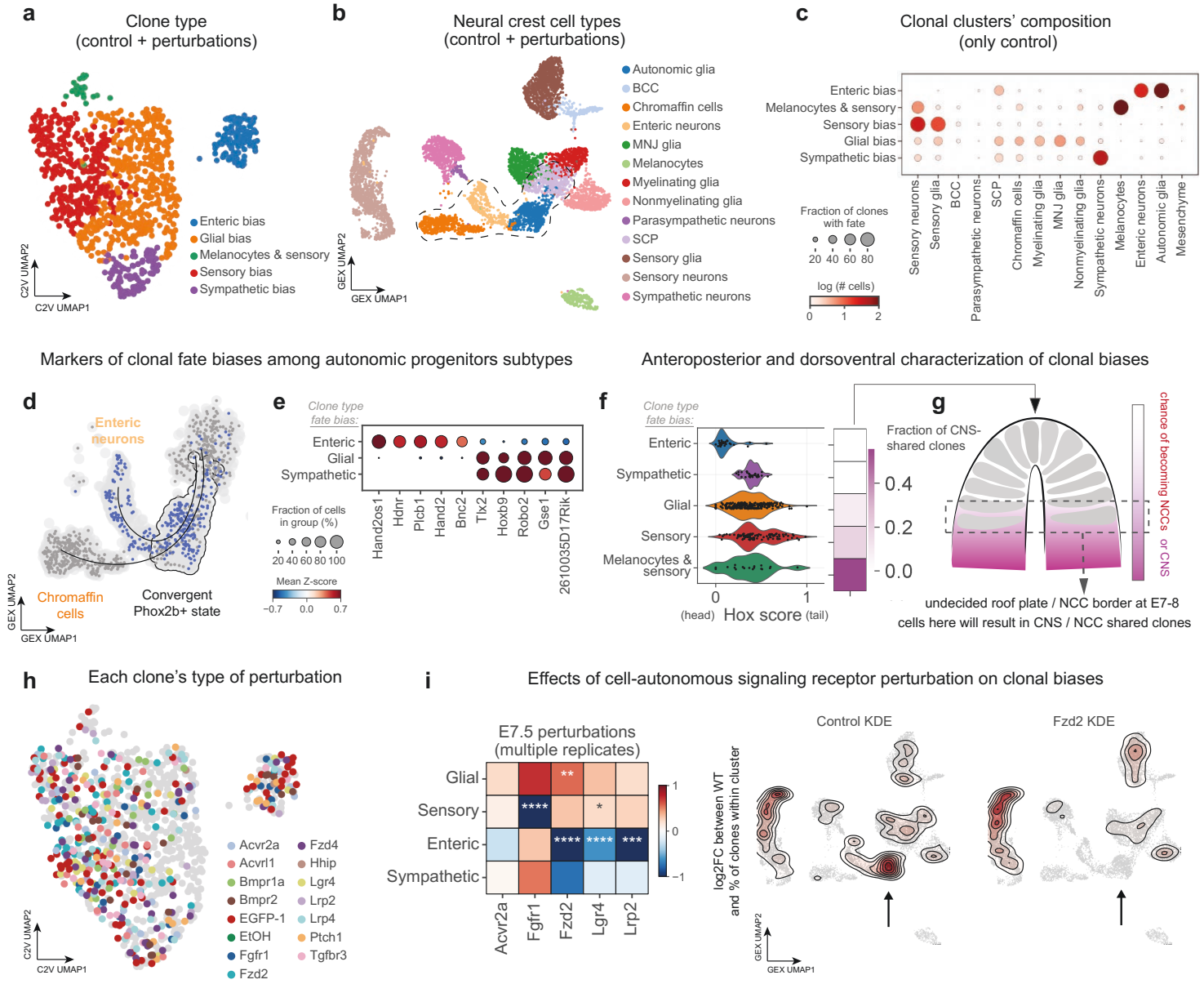

a
Clone type
(control + perturbations)
Enteric bias
Glial bias
Melanocytes & sensory
Sensory bias
Sympathetic bias
C2V UMAP2
C2V UMAP1
b
Neural crest cell types
(control + perturbations)
Autonomic glia
BCC
Chromaffin cells
Enteric neurons
MNJ glia
Melanocytes
Myelinating glia
Nonmyelinating glia
Parasympathetic neurons
SCP
Sensory glia
Sensory neurons
Sympathetic neurons
GEX UMAP2
GEX UMAP1
c
Clonal clusters' composition
(only control)
Enteric bias
Melanocytes & sensory
Sensory bias
Glial bias
Sympathetic bias
Sensory neurons
Sensory glia
BCC
Parasympathetic neurons
SCP
Chromaffin cells
Myelinating glia
MNJ glia
Nonmyelinating glia
Sympathetic neurons
Melanocytes
Enteric neurons
Autonomic glia
Mesenchyme
Fraction of clones with fate
20 40 60 80
log (# cells)
0 1 2
Markers of clonal fate biases among autonomic progenitors subtypes
d
Enteric neurons
Chromaffin cells
Convergent Phox2b+ state
GEX UMAP2
GEX UMAP1
e
Clone type fate bias:
Enteric
Glial
Sympathetic
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
Clone type fate bias:
Enteric
Sympathetic
Glial
Sensory
Melanocytes & sensory
0
1
(head) Hox score (tail)
0.4
0.2
0.0
g
Fraction of CNS-shared clones
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
Sensory
Enteric
Sympathetic
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 neural tube development in which dorsal NCC progenitors emigrate prior to ventral cells. This differential timing spent in the tube could affect the number of cell divisions, resulting in different proportions of CNS cells in the clone). This model results from the plotted data in panel f showing the fraction of neural crest clones shared with CNS.

**(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.