Manuscript Click here to view linked References

A single cell and tissue-scale analysis suite resolves Mixl1's role in heart development

Magdalena E Strauss[1,2,3], Mai-Linh Nu Ton[4,5], Samantha Mason[4,5,6], Jaana Bagri[4,5], Luke TG Harland[4,5], Ivan Imaz-Rosshandler[4,7], Nicola K Wilson[4,5], Jennifer Nichols[8], Richard CV Tyser[5], Berthold Göttgens[4,5*], John C Marioni[1,9*], Carolina Guibentif[10*#]

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, CB10 1SD Cambridge, UK
2 Department of Mathematics and Statistics, University of Exeter, EX4 4PY Exeter, UK
3 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA Cambridge, UK
4 Department of Haematology, University of Cambridge, CB2 0AW Cambridge, UK
5 Cambridge Stem Cell Institute, University of Cambridge, CB2 0AW Cambridge, UK
6 Present Address: Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, EC1M 6AU London, UK
7 Present Address: Altos labs, San Diego, CA, USA.
8 MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, EH4 2XU Edinburgh, UK
9 Present Address: Genentech, South San Francisco, CA, USA
10 Inst. Biomedicine, Dept. Microbiology and Immunology, Sahlgrenska Center for Cancer Research, University of Gothenburg, 413 90 Gothenburg, Sweden

lead contact: carolina.guibentif@gu.se
* corresponding authors: bg200@cam.ac.uk, marioni@ebi.ac.uk, carolina.guibentif@gu.se

Summary

Perturbation studies using gene knockouts have become a key tool for understanding the roles of regulatory genes in development. However, large-scale studies dissecting the molecular role of development master regulators in every cell type throughout the embryo are technically challenging and scarce. Here we systematically characterise the knockout effects of the key developmental regulators *T/Brachyury* and *Mixl1* in gastrulation and early organogenesis using single-cell profiling of chimeric mouse embryos. For the analysis of this experimental data~~,~~ we present COSICC, an effective suite of statistical tools to characterise perturbation effects in complex developing cell populations. We gain ~~new~~ insights into *T*'s role in lateral plate mesoderm, in limb development, and in posterior intermediate mesoderm specification. Furthermore, we generate *Mixl1*$^{-/-}$ embryonic chimeras and reveal the role of this key

transcription factor in discrete mesoderm lineages, in particular concerning developmental dysregulation of the recently identified juxta-cardiac field.

Introduction

Using CRISPR knockouts in conjunction with a single-cell transcriptomic readout has become an important tool for understanding gene function. While a number of studies have focused on the development of experimental and analysis techniques for large scale screens, where up to thousands of genes are knocked down in cell lines[1], others have focused on smaller-scale knockout analysis in more complex settings, such as mouse models[2] or organoids[3].

Within the context of such highly informative lower throughput model systems, the generation of embryonic chimeras, where mutant cells are injected into wildtype (WT) embryos at the blastocyst stage, is a powerful tool to study the function of essential developmental transcription factors[4-6] (Figure 1A). In the resulting chimeras, WT cells develop normally, establishing signalling gradients necessary for the embryo to develop, while the effects of the knockout can be observed by studying the descendants of the injected mutant cells. Chimeras can therefore reveal the cell-autonomous role of essential genes, where a full knockout would lead to embryonic lethality and gross developmental malformations. Furthermore, single-cell profiling for chimeras enables the comprehensive study of knockout effects well beyond differences in organ contribution (Figure 1B, S1AB).

Although powerful, single-cell chimera studies are challenging to interpret, with sampling biases and other technical factors potentially confounding results. In particular, the systematic assessment of the impact of knockouts on lineage progression, on gene expression, and on the interaction of these effects remains challenging. To address these problems, we herein present COSICC (**CO**mparative analysis of **S**ingle-**C**ell-RNA-seq data for **C**omplex cell populations), a ~~novel~~ set of computational tools specifically designed for the analysis of perturbations in complex developing cell populations compared to control experiments, and therefore ideally suited to chimera studies. We applied this approach to dissect the role of two key gastrulation regulators, where conventional knockout embryos are embryonic lethal during organogenesis. We first validated COSICC using previously published data from chimeric *T/Brachyury*$^{-/-}$ embryos, where *T* had been shown to regulate the egress of posterior mesoderm in the late stages of gastrulation, with *T*$^{-/-}$ primitive streak cells contributing to an accumulated cell population with neuromesodermal progenitor transcriptional signature[4]. We then generated *Mixl1*$^{-/-}$ chimeric mouse embryos, which we analysed using single-cell RNA sequencing (scRNAseq) and revealed, through application of COSICC, the molecular and tissue-scale consequences of loss of function for this key developmental regulator at whole embryo scale. *Mixl1* is the only characterized mammalian member of the *Mix/Bix* transcription factor family, first identified in Xenopus as regulating mesendoderm development[7]. While *Mixl1*$^{-/-}$ mouse embryos show severe developmental defects and early embryonic lethality, in a previous chimera assay *Mixl1*$^{-/-}$ cells only showed contribution impairment to posterior gut tissues[8], while in the present study we characterize the molecular effects of *Mixl1* knockout throughout all embryonic lineages present at early organogenesis.

Results

COSICC leverages single-cell profiling to analyse developmental perturbations

We considered two independent chimera data sets, both generated at Embryonic Day (E) 8.5 of mouse development: i) a previously published data set that studied the cell autonomous function of *T* for validation of our analysis suite; and ii) a newly generated data set created to study the role of *Mixl1* ~~for new insights~~ in early organogenesis. Additionally, we used data from a control experiment, where wildtype (WT) cells were injected at the blastocyst stage to create control chimeras; this was also profiled at E8.5. In all chimeras, we define the progeny of injected cells as tdTomato positive cells (tdTom$^+$, Figure 1AB), since the injected cells have constitutive expression of the fluorescent marker tdTomato. The wildtype (host) cells lack tdTomato expression and are referred to as tdTomato negative (tdTom$^-$).

Profiling the chimeric embryos using single-cell RNA-sequencing (scRNA-seq), in conjunction with reference single-cell atlases of early development, opens exciting possibilities for studying the effects of perturbations during cellular diversification (Figure 1B, S1A-C). Importantly, in the mouse embryo dataset used as a reference framework for our analysis of embryo chimeras,

1 developmental trajectories were inferred by combining the Waddington Optimal Transport

2 (Waddington-OT) matrices[9] (Star Methods) with probabilistic mixture modelling. We can thus

3 explore whether the perturbed cells are depleted or enriched in certain cell types or in lineage

4 trajectories leading to each cell type, and assess the impact of the perturbation on developmental

5 progression along specific lineage trajectories (Figure 1C).

6

7 Our COSICC analysis framework (Figure 1D) evaluates knockout effects on cell type or cell group

8 abundance (COSICC_DA_group), on lineage trajectory abundance (COSICC_DA_lineage), and on

9 developmental delay along lineage trajectories (COSICC_kinetics). COSICC addresses two specific

10 challenges arising in the study of perturbations in complex systems. First, in an experiment where

11 a fixed total number of cells are collected, if the perturbation leads to a strong reduction of some

12 cell types or groups of cells, then the unaffected cell types/groups will appear as if they were

13 enriched in the perturbed cell population (Figure 1E). COSICC_DA_group and COSICC_DA_lineage

14 include a sampling bias correction approach to address this problem (Figure 1E, Star Methods),

15 based on sub-sampling with the assumption that most groups are not affected as a result of the

16 knockout. It should be noted that the same type of sampling bias also occurs in any perturbation

17 experiment with single-cell analysis readouts, where the number of sampled cells is pre-

18 determined based on technical constraints. Second, there are experimental effects inherent to a

19 chimera-based assay that may confound biological insights. All analyses were thus framed in

20 comparison with an external control data set, the WT chimeras introduced earlier (Figure S1CD).

21 Chimera generation may lead to transcriptional differences between tdTom^{+} and tdTom^{-} cells,

22 due to, for instance, to *ex vivo* culture applied to tdTom^{+} cells prior to injection, which can affect

their transcriptome even in the absence of genetic editing. COSICC therefore performs DA testing with reference to the external control, thus accounting for experimental effects (Figure 1F).

A framework to consider the external control is also included in COSICC_kinetics (Figure 1G), which identifies dynamic shifts in lineage development by testing whether developmental progression along a lineage trajectory has been delayed or accelerated for the knockout cells (Star Methods). Finally, we applied to each cell type a mixed effects model approach[10] to test for differential gene expression (DE), again contrasting to external controls, and also accounting for batch effects across different pools of chimeric mouse embryos (Figure S2, Star Methods).

In addition to the development of COSICC, our analysis framework uses a substantially extended reference atlas of mouse gastrulation[11] compared to the reference atlas used in previous work[6]. This extended mouse gastrulation atlas contains four additional timepoints between E8.5 and E9.5, four times as many cells as the previous reference, as well as updated cell type annotations, thereby more completely representing early mouse organogenesis (Figure 1B, S1A-C). Using this extended reference data set allowed us to identify cell type, stage and pseudotime of chimeric cells that, in spite of being collected at E8.5, were more similar to cells present in later stage embryos due to asynchrony in developmental progression within embryo litters[12], or because of faster progression along a developmental lineage trajectory (E8.75 and later, Figure S1D).

Characterisation of the impact of *T* knockout on embryonic development reveals its cell-autonomous role for intermediate mesoderm as well as limb development

1

2 We used all elements of COSICC (Figure 1D) to obtain a comprehensive view of the effects of *T*
3 knockout. First, the cell type level analysis of $T^{-/-}$ chimeras with COSICC_DA_group confirmed
4 previous results[4], with depletion of somitic mesoderm, presomitic mesoderm, intermediate
5 mesoderm and notochord in the mutant cells and enrichment for neuromesodermal progenitors
6 (NMPs; Figure 2A). In addition, we observed a broad depletion of other mesodermally-derived
7 cell types including endothelial cell types, namely embryo proper endothelium, venous
8 endothelium and allantois endothelium, as well as mesenchyme, yolk sac mesothelium, and
9 allantois. We also observed enrichment of caudal epiblast cells, which was only present at earlier
10 time points in wildtype embryos, suggesting a developmental delay of mutant cells and impaired
11 gastrulation as observed previously[4; 13]. Additionally, our pipeline facilitated assessment of time-
12 dependent cellular abundance changes, as seen for example for blood progenitors and cranial
13 mesoderm, which showed significant changes in relative abundance across different time points
14 (Figure 2B). We also noted that accounting for the bias illustrated in Figure 1E avoids incorrect
15 conclusions about cell types being enriched for knockout cells (Figure S3A).
16
17 Second, at the level of lineage trajectories, results from COSICC_DA_lineage were broadly
18 concordant with the cell type level results, with $T^{-/-}$ cells significantly depleted in trajectories
19 towards somitic mesoderm, presomitic mesoderm, allantois, mesenchyme and yolk sac
20 mesothelium (Figure 2C, Figure S3B).
21

Third, among the trajectories containing sufficient cells to assess lineage progression in both mutant and WT fraction (Figure S3C, Figure S4), COSICC_kinetics revealed a delay for both mesenchyme and allantois lineage development as a result of *T* knockout (Figure 2DE). We also observed a delay in the trajectory leading to an ectodermal cell type, the ventral forebrain progenitors. Interestingly, our analysis shows a depletion of the mesenchyme cell type as well as a delay in its lineage development (Figure 2DE), which links to ~~a~~ an association between overexpression of *T* and changes in epithelial-to-mesenchymal transition reported in cancer[14-16]. Further, the depletion of the allantois cell type and lineage trajectory, and the developmental delay along that lineage trajectory observed in tdTom^{+} $T^{-/-}$ cells (Figure 2A-D), are in line with well-documented allantois defects for *T* knockout mice and chimeric embryos[17], giving confidence in the performance of COSICC.

We noted a striking effect of *T* knockout on the intermediate mesoderm cell-type, significant at the cell-type level as a whole (Figure 2A) and in particular for cells mapping to E8.0 and E8.25 time-points (Figure 2B). We further explored gene deregulation in this cell-type by performing DE analysis (Figure 2F, Table S1), and showed significant downregulation of the intermediate mesoderm master regulator *Osr1* as well as mesenchymal marker *Pdgfra*. Conversely, we see upregulation of epiblast genes *Pmaip1*[18] and *Tgif1*[19], caudal mesoderm marker *Cdx2*, along with deregulation of the Wnt pathway (with upregulation of *Wnt6* ligand as well as Wnt signaling inhibitor *Igfbp4*[20]). Metanephric mesenchyme was shown to derive from caudal mesoderm precursors expressing high levels of *T*, through an Osr1-expressing posterior intermediate mesoderm state[21]. With a drastic cell-autonomous effect of *T* knockout on intermediate

mesoderm cells, our data thus suggests that *T* plays a pivotal role in the regulation of posterior derivatives of intermediate mesoderm relevant for kidney development, likely through direct modulation of *Osr1* expression[22].

We also observed significant depletion of $T^{-/-}$ cells in lateral plate mesoderm (LPM). Since *T* is expressed in the LPM at early stages of the limb mesoderm trajectory (Figure S3D), we decided to investigate the impact of knocking out *T* on limb mesoderm development from LPM, given the impaired forelimb bud formation in *T* mutant embryos[23], where the underlying (*T*-dependent) mechanisms regulating forelimb formation has remained unclear[24]. We explored this further by performing DE analyses for both LPM and limb mesoderm (Figure 2G, Table S1). Interestingly we observed downregulation of *Cited1* in both LPM and limb mesoderm $T^{-/-}$ cells. While *Cited1* expression has been reported in defined limb mesoderm subsets[25; 26], its role in limb development remains unclear - our results suggest it may act downstream of *T* already within the LPM limb precursor cells. In LPM $T^{-/-}$ cells we also observed downregulation of *Hoxc5*, member of the Hox5 family directly implicated in forelimb development[27]. Finally, limb mesoderm $T^{-/-}$ cells also displayed downregulation of *Myc*, a key gene in limb development[28].

Taken together, our results indicate that beyond its previously documented role in orchestrating embryo axial elongation and somitogenesis, *T* also has important roles in cell-type specification within other mesodermal populations, specifically intermediate mesoderm and limb bud development.

Mixl1 knockout depletes epicardium and specific cardiac progenitor populations

Having validated COSICC on a previously published $T^{-/-}$ chimera data set, where we identified a cell-autonomous role of T for intermediate mesoderm specification and forelimb development, we proceeded to apply the same framework to a newly generated chimera data set. We decided to investigate the role of another gastrulation master regulator, the transcription factor *Mixl1*. Similarly to *T* knockout embryos, *Mixl1* mutants also show shortening of the anterior-posterior axis, however in addition they display disorganization of anterior structures (absence of a heart tube and abnormal headfolds), and severe gut developmental defects[8]. However, the molecular role of *Mixl1* within these cell populations has not been investigated. To systematically assess the embryo-wide effects of *Mixl1* knockout we generated $Mixl1^{-/-}$ chimeras and processed the resulting scRNA-seq data using COSICC. $Mixl1^{-/-}$ $tdTom^{+}$ mouse embryonic stem cells were generated using CRISPR/Cas9 to induce frameshift mutations in exon 2 of the *Mixl1* locus, thereby causing an early stop codon and functional inactivation of the gene and its homeobox domain (Figure S5A). Two independent $Mixl1^{-/-}$ $tdTom^{+}$ clones were injected into wildtype blastocysts to generate embryonic chimeras, and three independent pools of E8.5 $Mixl1^{-/-}$ chimeras were collected, containing 3 and 6 embryo chimeras from clone 1 respectively, and 4 embryo chimeras from clone 2. Single-cell suspensions were generated from each pool of embryos independently and cells were sorted by flow cytometry based on the tdTom reporter and analysed by scRNA-seq (see Star Methods).

Using COSICC_DA_group and COSICC_DA_lineage in *Mixl1*$^{-/-}$ chimeras we observed a broad effect on LPM and its derivatives: a depletion of cell types and lineages for cardiomyocytes of both first and second heart fields as well as epicardium and erythroid cells, and an increased representation of LPM and limb mesoderm (Figure 3A, Figure S5B,C). Collectively therefore, these observations amount to a depletion of splanchnic mesoderm with simultaneous enrichment of somatic mesoderm. Although *Mixl1*$^{-/-}$ mutants have been previously shown to have defective cardiac development[8; 29], the molecular role of *Mixl1* for cardiac cell type induction has remained elusive.

In addition to the impact on heart development, *Mixl1* knockout mice are known not to form a hindgut[8]. Consistently, in the *Mixl1*$^{-/-}$chimeras, hindgut and midgut were fully depleted at cell type level, while foregut, gut tube, pharyngeal endoderm and thyroid primordium were all partially, but substantially, depleted (Figure 3A, S5B).

Proceeding to investigate lineage trajectory development, note that COSICC_DA_lineage was not applicable to gut tube and pharyngeal endoderm, as there were too few cells at the final, E9.25, time point of the reference data set (Star Methods, Figure S3C, S4). Consistent with the *T*$^{-/-}$ chimera analysis, all cell types in the *Mixl1*$^{-/-}$ chimera that were depleted at the cell type level were also depleted at the lineage trajectory level (Figure 3B, S5C), in line with a general concordance between cell type differential abundance and changes in contribution to corresponding lineage trajectories. Finally, COSICC_kinetics revealed developmental delay for the epicardium lineage (Figure 3CD). In addition to the depletion of cardiac cell types (Figure 3A-

B) and substantial delay along the epicardium lineage trajectory (Figure 3CD), we observed a depletion of cells assigned to the mesenchyme cell type and lineage (Figure 3AB), a cell type that we found to be part of the epicardium trajectory (Figure S5D). Furthermore, we found a substantial number of differentially expressed genes for mesenchyme (Figure S5E, Table S1).

Prompted by the developmental delay in epicardium development, the impact on cardiac cell types, and the deregulation of the mesenchyme cell type on the epicardium trajectory together with recent reports suggesting previously unrecognised cardiac progenitor populations[30], we explored further the effects on cardiac development of *Mixl1* knockout. In particular, the gene deregulation of mesenchyme led us to investigate whether we could identify precursors of the epicardium within the population labelled as mesenchyme in the extended mouse gastrulation atlas. Indeed, sub-clustering the mesenchyme cell type (Figure 3E) revealed transcriptionally defined subsets, segregated by the expression of markers for a recently identified epicardial and cardiomyocyte progenitor population termed the juxta-cardiac field (JCF) [30]. This observation is consistent with the notion that cells annotated as mesenchyme may comprise JCF cells that would constitute putative progenitors for the cardiac cell types affected in *Mixl1*$^{-/-}$chimeras. Indeed, the average expression of JCF markers (JCF score, JCFS, see Methods) was downregulated in *Mixl1*$^{-/-}$ cells mapping to the mesenchyme cell type (Figure 3F, p-value< 2.2e^{-16} for Wilcoxon rank-sum test), and clusters with high JCFS were strongly depleted in *Mixl1*$^{-/-}$chimeras (Figure 3G). Altogether, this suggests that a previously unknown role for *Mixl1* in the onset of a JCF program, impacting the downstream development of cardiac cell types, including the epicardium.

Using COSICC, we can easily compare effects of different perturbations if they can be mapped to the same reference framework, in this case the extended mouse gastrulation atlas. As mentioned before, both *T* and *Mixl1* have been reported as major mesoderm regulators[31; 32; 4; 8]. Comparing embryo-wide effects of both knockouts, we observed that the two sets of chimeras generally had distinct phenotypes (Figure S6). Interestingly, LPM was depleted in *T*$^{-/-}$ chimeras and enriched in *Mixl1*$^{-/-}$ chimeras (Figure S6). Moreover, both chimeras displayed distinct defects in LPM derivatives, with a depletion of endothelial cells in *T*$^{-/-}$ chimeras compared to defective development of cardiac tissues in *Mixl1*$^{-/-}$ chimeras. This suggests independent roles of the two transcription factors in the development of populations derived from LPM. Alternatively, the population annotated as LPM in the atlas may also comprise heterogenous progenitors with restricted fates, where *Mixl1* and *T* regulate distinct gene networks.

Discussion

Recent years have seen an increasing number of large-scale perturbation experiments[33] in complex systems in *vivo*[2; 34] and in *vitro*[3], as well as a growing wealth of reference atlases being generated[35; 36], including in the context of international consortia[37; 38], along with the emergence of single-cell analysis of patient material[39; 40]. Principled analysis of large-scale complex perturbation data will be instrumental to harness such data sets in the context of increasingly granular single-cell reference atlases, leading to ~~new~~ biological and clinical insights.

Single-cell readouts for perturbation experiments in complex systems allow insights into molecular, cellular and tissue-scale processes across a wide range of cell types and lineages. Here we used scRNA-seq data generated from chimeras to study the role of the transcription factors *Mixl1* and *T* across all lineages in the developing mouse embryo. For this purpose, we developed a set of computational tools, COSICC, which enabled us to tackle the particular challenges of complex perturbation data. Our experimental and computational approach allowed the ~~novel~~ identification of affected lineage trajectories, such as limb mesoderm and intermediate mesoderm for *T*, and epicardium for *Mixl1*.

The COSICC framework is also relevant to applications outside of the specific use case of chimeric embryos in three ways. First, chimeric organoid systems, where mutant cells are cultured in the context of wildtype aggregates are emerging as promising alternatives to study genetic perturbations, including malignancies, *in vitro*[41-44]. The entire suite of COSICC will be a highly effective way of analysing single-cell studies in these types of systems. Second, the sampling bias correction step introduced for unbiased assessment of differential abundance at the level of cell type, cell group or lineage trajectory is widely applicable to the detection of the effect of CRISPR or drug perturbation on cell type composition for any experiment with a single-cell readout, e.g. in organoids, mouse models or non-model organisms[45], and has the potential to prevent future false positive results of overrepresentation of cell types in any such system. Third, while COSICC was designed for chimeric systems and development, where we see the strongest potential of the method and where there has been exciting ~~new~~ experimental developments with chimeric

organoids, we demonstrated the wider relevance of the proposed computational analysis methods to general perturbation and disease scRNA-seq data by applying it to a study of AML patient data[46] (~~Note S1 and ~~Figure S7).

COSICC is a highly flexible framework. We used mutual nearest neighbours[47] to map chimera data to the reference data set, for consistency with the batch correction method previously applied to our reference data set[11]. However, COSICC may be used with any mapping[48-50], cell lineage estimation[51], and pseudotime/cell ordering method. Instead of pseudotime, COSICC could alternatively be combined with the transcriptional rank of the individual embryo in the reference data set to which the chimera cell is most similar[5]. COSICC is a cell-specific coordinate-based approach, locating each individual cell within a framework of precomputed scores on a reference data set (e.g. the extended mouse gastrulation data set). This includes lineage scores (e.g. Waddington-OT) and lower-dimensional representations (e.g. diffusion maps). This coordinate-based approach, where each cell is associated with its own coordinates, differs from neighbourhood-based approaches that look for small sets of similar cells between a query and a reference data set[52; 36]. This cell-specific coordinate-based strategy maintains the high granularity of the single-cell approach and provides a location of the cell in terms of the normal development of a reference atlas, along a specific direction (e.g., lineage trajectory).

The application of COSICC to the $T^{-/-}$ chimera data set validated our method by reproducing results from previous analyses[4]. Due to better cell type identification thanks to the extended reference atlas, we detected additional significant effects of *T* knockout on notochord (Figure

2A), strongly supported by the literature[31; 32; 17]. Finally, our chimera-based experimental approach followed by single-cell transcriptional readout and systematic analysis using the COSICC workflow enabled the dissection of the cell-autonomous role of the *T* transcription factor within discrete mesodermal populations, shedding light on previously unknown molecular roles of *T* in intermediate mesoderm and limb development.

We then generated an additional ~~new~~ chimera data set targeting another key gastrulation regulator, *Mixl1*. In addition to the expected strong depletion of definitive endoderm tissues[8], we noted a marked impact on cardiac lineages, in particular a depletion of specific groups of precursor cells to epicardium. Indeed, *Mixl1*-KO mice show severe cardiac malformations[29], and epicardium plays an important role in coordinating the development of the other heart tissues, namely through modulation of the *Wnt* pathway[53]. Furthermore, *Mixl1* has recently been implicated in an LPM inducing regulatory transcription factor network conserved in chordates[54], but based on the studied enhancer region this previous study did not extend to mammals. Our observations of depleted splanchnic mesoderm (cardiac tissues) and enrichment of somatic mesoderm (limb) suggests that *Mixl1* may also have a major role in balancing lineage propensity within LPM precursors in mammalian embryos. Finally, our systematic analysis also allowed us to refine cell typing in the reference data set. We found that the general impact of *Mixl1* knockout on cardiac development is reflected by the failure of *Mixl1* knockout cells to develop a JCF signature. This observation has led to identification of cell subsets with a high JCF signature within the mesenchyme cell type in the extended mouse gastrulation atlas.

Resource Availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Carolina Guibentif (carolina.guibentif@gu.se).

Materials availability

Wildtype tdTomato and mutant *Mixl1*$^{-/-}$ mESC lines are available upon request.

Data and code availability

Single-cell RNA-seq data for the *Mixl1* chimeras have been deposited at Arrayexpress and are publicly available. Accession numbers are listed in the key resources table.

The COSICC R package (https://github.com/MarioniLab/COSICC) and all original analysis code in R (https://github.com/MarioniLab/analysis_chimera_data) have been deposited on Github and are publicly available (https://doi.org/10.5281/zenodo.15008255 [55]). Count data for *T*$^{-/-}$ and WT chimeras was obtained from[4], via the *MouseGastrulationData* Bioconductor package[56]. Reference embryo data, including dimensionality reduction and Waddington-OT matrices was obtained from a previous study[11].

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Limitations of the study

All our computational analysis steps, and in particular finer qualitative analyses require sufficient cell numbers of perturbed and wildtype cells mapping to the specific cell-types/lineage trajectories (see Figure S3C and Figure S4). It should be noted that this is not a problem of the COSICC method per se, as any method based on the concepts of rigorous statistical inference will detect significant differential abundance or expression only in the presence of sufficient evidence, i.e. a large enough sample size. In fact, this general problem can be limiting for the analysis of rare populations, or in settings where there is limited access to sufficient cell numbers. *Mixl1* and *T* knockout lead to strong depletion of several cell types, which means that for the respective lineage trajectories we were not able to use COSICC_kinetics to identify potential lags in development or test for differential gene expression, as this would have required the presence of both mutant and WT cells mapping to those cell types. In that sense the present study informs future work aiming at dissecting the molecular role of these master-regulators in the development of cell types found here to be completely depleted at E8.5, for instance to design experiments where embryos are collected at earlier time-points in order to capture the affected developmental precursor of interest.

Acknowledgements

We thank William Mansfield at and the Wellcome-MRC Cambridge Stem Cell Institute animal facility for blastocyst injections, the Flow Cytometry Core Facility at the Cambridge Institute for Medical Research for cell sorting, Katarzyna Kania and the CRUK-CI genomics core for preparing the 10X libraries and for sequencing and Jonathan Griffiths for comments on the manuscript.

M.E.S. is supported by the Wellcome Trust (220442/Z/20/Z). C.G. was funded by the Swedish Research Council (2017-06278) and by a Swedish Childhood Cancer Fund position grant (TJ2021-0009). M.-L.N.T. was funded by a Herchel Smith PhD Fellowship in Science. S.M. was funded by a British Heart Foundation PhD Studentship. J. B. was funded by a Wellcome Trust PhD studentship. L.T.G.H. was funded by a Wellcome Early-Career Award (226309/Z/22/Z). R.T. was funded by the British Heart Foundation Cambridge Centre of Research Excellence (RE/18/1/34212) and core support from Wellcome to the Wellcome-MRC Cambridge Stem Cell Institute. Work in the Göttgens group is supported by Wellcome, Blood Cancer UK, MRC and CRUK, and by core support grants from Wellcome to the Wellcome-MRC Cambridge Stem Cell Institute. This work was funded as part of a Wellcome grant (220379/B/20/Z) awarded to B.G. and J.C.M. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions

M.E.S. analysed the *Mixl1* embryonic chimera data set, wrote all code, implemented COSICC and applied it to all data sets presented; M.-L.N.T. performed *Mixl1* chimeric embryo dissections and generated the scRNA-seq data set; C.G. generated the *Mixl1* knockout ESC lines; S.M. genotyped *Mixl1* ESC knockout clones; J.B. assisted with embryo dissections; N.W. assisted in scRNAseq data

generation; M.E.S, C.G. and J.C.M. interpreted the results with input from L.T.G.H., I.I-R., M.-L.N.T.,R.T. and B.G.; I.I-R. provided processed data and Waddington-OT analysis for the extended mouse gastrulation atlas; M.E.S. wrote the first draft of the manuscript; M.E.S., C.G., L.T.G.H., M.-L.N.T., R.T., B.G., and J.C.M edited the manuscript; J.N., B.G., J.C.M. and C.G supervised the study.

Declaration of interests

J.C.M. has been an employee of Genentech since September 2022, and I.I.-R. is an employee of Altos labs. The remaining authors declare no competing interests.

Figure Legends

Figure 1: **Overview of chimeric mouse embryos and COSICC framework**. A) Chimeric embryos were created by injecting mutant tdTom$^+$ cells into wildtype tdTom$^-$ (WT) embryos at the blastocyst stage. B) Chimeric mouse embryos mapped to the extended mouse gastrulation atlas (reference atlas cells in grey, tdTom$^+$ cells in red, tdTom$^-$ cells in blue). C) Illustration of lineages. D) Overview of COSICC. E) Correction of sampling bias; i) **Top**: Cell counts from an embryo without perturbation effect with tdTom$^+$ cells in red and tdTom$^-$ cells in blue. Middle: depletion of one cell type only; this depletion is unknown before data analysis, therefore a fixed proportion of cells with fluorescent markers is sampled, leading to cell types seemingly enriched for knockout

1 cells. **Bottom**: computational correction. F) **Top:** no change in cell type abundance for either WT

2 or targeted chimeras. **Middle**: reduction in abundance of tdTom^{+} cells for targeted, but no (or

3 less) reduction for WT chimeras. **Bottom**: Similar levels of reduction in cell type abundance in

4 tdTom^{+} cells for WT and targeted chimeras. G) Illustration of COSICC_kinetics output (i) and

5 statistical assessment of perturbation induced delay (Wilcoxon rank-sum test, Methods, ii). See

6 also Figure S1, S7.

7

8 **Figure 2: COSICC confirms existing results and reveals mechanistic insights into the role of *T* in**

9 **the development of discrete mesodermal populations.** A) DA at the cell type level for $T^{-/-}$

10 chimera knockout cells (COSICC_DA_group). B) Per-stage DA at cell type level. Significance at

11 FDR<0.1 is indicated by a dot. Cell types found to be significantly enriched or depleted (A) and

12 related cell types are shown. C) DA of Waddington-OT trajectories leading to the listed cell types

13 at E9.25 (COSICC_DA_lineage). D) Pseudotime distributions of tdTom^{+} (red) and tdTom^{-} (blue)

14 cells within $T^{-/-}$ and WT chimeras. E) Confidence interval for location parameter of Wilcoxon rank-

15 sum test (COSICC_kinetics). Confidence intervals for $T^{-/-}$ chimeras (purple bars) overlapping

16 neither 0 nor the confidence intervals for WT chimeras (black bars) imply significant

17 developmental delay (Figure 1G,ii). F) Volcano plot illustrating statically differentially expressed

18 genes for intermediate mesoderm. G) Volcano plot illustrating statically differentially expressed

19 genes for LPM (i) and limb mesoderm (ii). F-G: Imprinted genes, taken from Supplementary Data

20 1 of Santini et al. [57], are in purple. The x-axis represents the log2-fold change for differential

21 expression contrasted with the WT chimeras. YS: Yolk Sac. A-C: odds ratio and FDR values

22 calculated using Fisher's exact test (see Star Methods). See Table S2 for cell numbers for each

1 stage (mapped from the extended mouse gastrulation atlas), cell type and sample. See also Figure
2 S3, S4, S6.
3
4 **Figure 3: COSICC reveals mechanistic insights into epicardium development.** A) DA at the cell
5 type level for *Mixl1*$^{-/-}$ chimera knockout cells (COSICC_DA_group). B) COSICC_DA_lineage. C)
6 Confidence interval for location parameter of Wilcoxon rank-sum test (COSICC_kinetics). D)
7 Pseudotime distributions tdTom+ (red) and tdTom- cells (blue) within *Mixl1*$^{-/-}$ and WT chimeras
8 for the epicardium lineage. E) Sub-clustering of mesenchyme. i) Uniform Manifold Approximation
9 and Projection (UMAP) coordinates recomputed for the cells annotated as epicardium or
10 mesenchyme in the reference data set. (ii) UMAP coloured by JCF score (JCFS). (iii) density plot
11 of JCFS split by cluster. Clusters in (i) were labelled according to their levels of JCFS. Cluster Mes_4
12 mostly comprised the cells labelled as epicardium in the reference data set. F) Distribution of
13 JCFS for *Mixl1* and WT chimeras across all cells from the mesenchyme cell type shows a depletion
14 of JCF signature for the *Mixl1*$^{-/-}$ cells. G) COSICC_DA_group reveals strong depletion for *Mixl1*$^{-/-}$
15 cells for the clusters with high JCFS. To perform the normalizing step (Figure 1E), we ran
16 COSICC_DA_group across all cell types with mesenchyme replaced by the new subclusters, and
17 plotted results for the mesenchyme subclusters. YS: Yolk Sac. See Table S2 with n values
18 corresponding to the number of cells for each cell type and sample. See also Figure S3, S4, S5, S6.
19
20

21 # Star Methods

1

2 ## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

3 ### Mouse embryonic stem cell lines

4 Mouse embryonic stem cell (mESC) lines were generated in-house and have not been
5 authenticated. They were expanded under the 2i+LIF conditions[58], in a humidified incubator at
6 37°C and 7% CO2, and routinely tested negative for mycoplasma infection. A male, karyotypically
7 normal, tdTomato-expressing mESC line[6] was validated through ability to contribute to all
8 embryo structures in a morula aggregation assay. It was then targeted to disrupt the *Mixl1* locus
9 using the CRISPR/Cas9 system (see Method Details). Two mutant clones were used to generate
10 *Mixl1*$^{-/-}$ embryonic chimeras.

11

12 ### Mouse embryos

13 All procedures were performed in strict accordance with UK Home Office regulations for animal
14 research. Chimeric mouse embryos were generated as described in the project licence number
15 PPL 70/8406, following procedures detailed in Guibentif et al[4]: animals used in this study were 6-
16 10 week-old females, maintained on a lighting regime of 14 hours light and 10 hours darkness
17 with food and water supplied ad libitum. For chimera generation, E3.5 blastocysts were derived
18 from wildtype C57BL/6 matings, and after injection of the mutant cells, the resulting chimeric
19 embryos were transferred to C57BL/6 recipient females at 0.5 days of pseudopregnancy
20 following mating with vasectomised males. All chimeric embryos were collected at E8.5 (see
21 details below).

22

METHOD DETAILS

Mixl1$^{-/-}$ embryo chimera data generation

All procedures involving mouse embryos were performed in strict accordance with the UK Home Office regulations for animal research under the project license number PPL 70/8406.

TdTomato-expressing mouse embryonic stem cells (ESC) were derived as previously described[6]. Briefly, ESC lines were derived from E3.5 blastocysts obtained by crossing a male ROSA26tdTomato (Jax Labs – 007905) with a wildtype C57BL/6 female, expanded under the 2i+LIF conditions[59] and transiently transfected with a Cre-IRES-GFP plasmid[60] using Lipofectamine 3000 Transfection Reagent (ThermoFisher Scientific, #L3000008) according to manufacturer's instructions. A tdTomato-positive, male, karyotypically normal line, competent for chimera generation as assessed using morula aggregation assay, was selected for targeting *Mixl1*. Two pairs of guides were designed to induce large deletions in Exon 2 of the *Mixl1* locus using the http://crispr.mit.edu tool (guide pair 1: AAGCGGCGCCTTCTGCGAAC and TGCTGGGGCGCGAGAGTCGT; guide pair 2: TTGCGGCGCTGTGGCGCCGA and CGCTCCCGCAAGTGGATGTC) and were cloned into the pX458 plasmid (Addgene, #48138) as previously described[60]. The obtained plasmids were then used to transfect the cells and single transfected clones were expanded and assessed for Cas9-induced mutations. Genomic DNA was isolated by incubating cell pellets in 0.1 mg/ml of Proteinase K (Sigma, #03115828001) in TE buffer at 50°C for 2 hours, followed by 5 min at 99°C. The sequence flanking the guide-targeted sites was amplified from the genomic DNA by polymerase chain reaction (PCR) in a Biometra T3000 Thermocycler (30 sec at 98°C; 30 cycles of 10 sec at 98°C, 20 sec at 58°C, 20 sec at 72°C; and elongation for 7 min at 72°C) using the Phusion

High-Fidelity DNA Polymerase (NEB, #M0530S) according to the manufacturer’s instructions. To assess the CRISPR/Cas9 targeting, primers to amplify the targeted regions and including Nextera overhangs were used (target pair 1: F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTATTCCCGCGGCGTCT; R-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCGAGCTGAACGACGT; target pair 2: F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAGCTCCAGTTCGCAGA and R-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAGTTTGCAGTCTAGAACC), allowing library preparation with the Nextera XT Kit (Illumina, #15052163), and sequencing was performed using the Illumina MiSeq system according to manufacturer’s instructions. One ESC clone from each guide pair transfection showing a frameshift mutation resulting in the functional inactivation of the *Mixl1* locus was selected for injection into C57BL/6 E3.5 blastocysts (Figure S5A). Chimeric embryos were harvested at E8.5, dissected, and single-cell suspensions were generated by TrypLE Express dissociation reagent (Thermo Fisher Scientific) incubation for 7-10 minutes at 37°C under agitation. Single-cell suspensions were sorted into tdTom+ and tdTom- samples using a BD Influx sorter with DAPI at 1µg/ml (Sigma) as a viability stain for subsequent 10X scRNA-seq library preparation (version 3 chemistry), and sequencing using the Illumina Novaseq6000 platform in one full S1 flowcell. In total, three independent pools of chimeric embryos were processed, two pools generated with a clone successfully targeted with guide pair 1 (containing respectively 3 and 6 embryos), and one clone successfully targeted with guide pair 2 (comprising 4 embryos). In the chimera cell pools, while all tdTom^{+} cells (progeny of the injected ESC lines) are male, the tdTom^{-} cells (progeny of host blastocysts) comprise a mix of both male and female cells. Thus to avoid sex bias in the results, Xist and Y chromosome genes were excluded for the highly variable

1 genes during data processing, quality control (QC) and mapping (see section below on Data
2 processing, QC and mapping to extended mouse gastrulation atlas) as in previous studies[4].
3 COSICC also accounts for this potential sex bias, as it incorporates an external control (WT
4 chimeras) with a similar sex bias (male injected cells, mixed host cells), so that perturbation
5 effects identified are not driven by expression of sex genes.

6 This resulted in 17393 tdTom$^-$ and 18754 tdTom$^+$ cells that passed quality control (see "Data
7 processing, QC and mapping to extended mouse gastrulation atlas" below), with an average
8 of 2574 genes and 10727 UMIs detected per cell. To exclude transcriptional effects intrinsic
9 to the chimera assay, chimeric embryos were generated by injecting the parental tdTom$^+$ *Mixl1*$^{+/+}$
10 (WT) line into C57BL/6 E3.5 blastocysts and processed as for the *Mixl1*$^{-/-}$ samples.

11

12 **QUANTIFICATION AND STATISTICAL ANALYSIS**

13 All the analyses and statistical tests were performed in R(v. 4.1.1). An R package was created to
14 facilitate application of COSICC (https://github.com/MarioniLab/COSICC). All n values applied in
15 the statistical analyses detailed below are listed in Table S2.

16

17 ***Data processing, QC and mapping to extended mouse gastrulation atlas***

18 For the *Mixl1*$^{-/-}$ chimeras, raw data were processed with Cell Ranger (v. 6.01, 10X Genomics)
19 without mapping intronic reads, with reads mapped to the mm10 genome and counted with
20 GRCm38.92 annotation to ensure consistency with the reference atlas[11] and the existing *T*$^{-/-}$
21 chimera data set[4]. For the *T*$^{-/-}$ chimera we used processed reads[4], for the *Mixl1*$^{-/-}$ we followed the

1 quality control steps detailed in Guibentif et al. [4] and
2 https://github.com/MarioniLab/EmbryoTimecourse2018 using R (v. 4.1.1).
3

4 ***COSICC overview***

5 COSICC provides a comprehensive picture of knockout effects in complex organisms (Figure 1D).
6 First, we investigated whether knockout causes differential abundance at the level of cell types,
7 compared to changes seen between tdTom$^+$ and tdTom$^-$ cells from the WT chimeras
8 (COSICC_DA_group, Figure 1F, 2A, 3A). Second, for cell types represented at E9.25, we identified
9 the cells comprising the inferred developmental trajectories ending in the respective cell types
10 (henceforth called lineage trajectories), and tested whether a knockout would lead to abundance
11 or depletion of cells for the lineage trajectory, again compared to changes in the WT chimeras
12 (COSICC_DA_lineage, Fig2C, 3B). Third, for lineage trajectories that were not severely depleted
13 for knockout cells, we tested whether the progression along the lineage trajectory was delayed
14 or accelerated (COSICC_kinetics, Figure 1G, 2DE, 3CD, S7DE). Finally, we applied a mixed effects
15 model approach[10] to test for differential gene expression for each cell type individually, again
16 contrasted with external controls (WT chimeras, Figure 2FG, S5E).
17

18 ***COSICC_DA_group: group-based DA testing***

19 COSICC_DA_group tests whether for a distinct group of cells (such as a cell type or sub-celltype)
20 there is significant depletion or enrichment of tdTom$^+$ cells compared to tdTom$^-$ in the knockout
21 chimeras, compared to the difference between tdTom$^+$ and tdTom$^-$ in the WT chimeras (Figure
22 1F).

1

2 Group-based DA testing (COSICC_DA_group) was performed as follows:

3 First, we discarded groups with less than 30 WT $tdTom^+$ cells assigned to them. For each other
4 group c the following ratios were computed:

5 $r_{c,target}$ = number of $tdTom^+$ cells in knockout chimeras mapping to c/number of $tdTom^-$ cells in
6 knockout chimeras mapping to c

7 $r_{c,control}$ = number of $tdTom^+$ cells in WT chimeras mapping to c/number of $tdTom^-$ cells in WT
8 chimeras mapping to c

9

10 As illustrated in Figure 1E, the data generation was affected by an experimental bias affecting
11 compositional data analysis: groups that are not depleted will appear to be over-represented.
12 We corrected for this sampling bias as follows (Figure 1F):

13 First we computed medians across all groups: m_{target} = $median_c(r_{c,target})$ and $m_{control}$ =
14 $median_c(r_{c,control})$, where $median_c$ refers to the median across all groups.

15 Then we sampled for the knockout chimeras (without replacement) n_{target} cells among the $tdTom^+$
16 cells and n_{target} x m_{target} cells among the $tdTom^-$ cells, where n_{target} = min(number of $tdTom^+$ cells
17 in knockout chimera mapping to the group, number of $tdTom^-$ cells in knockout chimera mapping
18 to the group/m_{target}). Under the assumption that most groups are not affected by depletion or
19 enrichment as a result of the knockout (and that therefore the median ratio m_{target} is not
20 affected), this normalisation corrects for the fact that given the depletion of larger groups, an
21 approximately equal number of $tdTom^+$ and $tdTom^-$ cells overall will lead to spurious enrichment

results for other groups, if not corrected. We proceeded analogously for the WT injected control chimeras.

After the sampling step described in the previous paragraph, we performed, for each group, a Fisher's exact test to determine the odds ratio and p-value of a chimera cell mapping to a particular group being significantly more or significantly less likely to be tdTom$^+$ than tdTom$^-$. We repeated the sampling and the Fisher's exact test 100 times to avoid dependence of the magnitude and significance of the inferred effect on the randomly chosen sample. For downstream analysis we used the median odds ratio and the median p-value, obtained as median across 100 subsamples.

The steps described above were applied to each group. FDR correction[61] was subsequently applied across groups.

We applied COSICC_DA not only to cell types, but to a combination of cell type and the embryonic stage of a chimera cell (obtained from mapping to the reference atlas, see the section *Data processing, QC and mapping to extended mouse gastrulation atlas*), as described above with cell type and stage combinations replacing groups.

COSICC_DA_lineage: lineage trajectory-based DA testing

COSICC_DA_lineage tests whether tdTom$^+$ chimera cells have an increased or decreased predisposition to develop into a particular cell type at a later stage (Figure 1C). COSICC_DA_lineage adapts COSICC_DA_group for lineage trajectory-based rather than cell type-

based comparison by using the Waddington-OT method[9] to obtain probabilities of a cell being part of a lineage trajectory. Waddington-OT, a method based on optimal transport, computes a matrix of transition probabilities linking cells at a specific time point to those at the next time point. Applying transition matrices repeatedly allows the computation of normalised fate matrices assigning to each cell and each lineage trajectory the probability that the cell is part of the specific lineage trajectory. For transitions between each pair of consecutive states, we used the Waddington-OT transition matrices for the reference atlas computed for publication of the extended mouse gastrulation atlas[11]. To compute fate matrices from the transition matrices, we used the command line interface as in Section 5 of the Waddington-OT tutorial (https://broadinstitute.github.io/wot/tutorial/).

We included lineage trajectories whose respective terminal cell type contains more than 100 cells at E9.25 in the extended gastrulation atlas, and performed 10 iterations of the following:

> For each cell from the chimeric embryos a lineage trajectory was sampled according to the probability distribution given by the Waddington-OT fate matrix. Given the sampled lineage trajectory for each cell (a fixed state and not a probability of being in a state), we applied COSICC_DA_group, performing the sampling of cells for sampling bias correction and Fisher tests as for COSICC_DA_group 30 times for each sampling of lineage trajectories based on the Waddington-OT fate matrix. We performed FDR correction across lineage trajectories.

We then computed the medians across odds-ratios and FDR-corrected p-values across the 10 repeats of sampling lineage trajectories based on the Waddington-OT fate matrix.

1

2 ***COSICC_kinetics: testing for acceleration or delay of knockout cells along a trajectory***

3 To determine whether there is a delay or acceleration for tdTom^{+} cells along a trajectory, we
4 combined Waddington-OT fate matrices with statistical modelling and pseudotime.

5 Cells were assigned to lineage trajectories using a mixture model of skewed t-distributions[62] with
6 two or three components (the optimal number of components was determined by the Bayesian
7 information criterion[63]. For each mapped stage, we proceeded as follows.

8 First we preselected those cells that had higher normalised Waddington-OT scores than random
9 normalised Waddington-OT fate probabilities for the lineage trajectory (assuming a uniform
10 random distribution), i.e. higher than the inverse of the total number of cells at their respective
11 stage. Then, we fitted the mixture model and assigned to the lineage trajectory those cells that
12 were in the cluster of the highest Waddington-OT fate probabilities. For each stage we discarded
13 cell types that did not constitute at least 10% of the cells for the lineage trajectory at that stage,
14 to mitigate any potential mistakes from the algorithm used to map the chimera data to the
15 reference data set influencing the trajectory. From the time point onward from which the final
16 cell type was the most frequent cell type in a lineage trajectory we excluded all other cell types
17 from the lineage trajectory.

18 For the reference data set, the extended mouse gastrulation atlas, temporal genes were selected
19 by testing their association with embryonic time using ANOVA (reference-based temporal genes).
20 Genes for which the residuals from the ANOVA regression were associated with batch, i.e. genes
21 correlated with time to different degrees for different batches, were removed from the list of
22 reference-based temporal genes. Then we computed diffusion maps for the reference atlas

cells[64; 65] based on the reference-based temporal genes to obtain a more fine-grained resolution of temporal development. We identified pseudotime as the diffusion component most correlated with actual time, to avoid focusing on components not related to time, but to e.g. spatial location. As the starting point for the trajectories we generally used E7.5, as most chimeric tdTom$^+$ and tdTom$^-$ cells mapped between E7.75 and E9.0 (Figure S1D). COSICC_kinetics was applied to all lineage trajectories that were not severely depleted for tdTom$^+$ chimera cells, i.e. for which the odds ratio of (tdTom$^+$ cells for knockout chimera/tdTom$^-$ cells for knockout chimeras)/(tdTom$^+$ cells for WT chimera/tdTom$^-$ cells for WT chimeras) > 0.05.

Pseudotimes were assigned to chimera cells by assigning to the chimera cell the average of the pseudotimes of the 10 atlas cells to which the chimera cell was most correlated, where for the computation of correlation we used the reference-based temporal genes only. This ensured that the pseudotime mapping was based on genes that differ across stages in the reference data set and excluded genes whose expression depended on batch.

We used the Wilcoxon rank-sum test[66] to identify trajectories whose median was more strongly shifted in tdTom$^+$ cells for the knockout chimeras than for the WT chimeras (Figure 1G, 2E, 3C, S7DE). This identified trajectories for which the difference between tdTom$^+$ and tdTom$^-$ cells for the knockout chimeras was significant and the 95% confidence intervals for the location parameter (the difference in pseudotime median for tdTom$^+$ versus tdTom$^-$ cells) of the respective Wilcoxon tests were non-overlapping between WT and knockout chimeras.

DE testing for chimera data

DE tests were performed for each cell type separately. To test for DE between tdTom$^+$ and tdTom$^-$ cells in the knockout chimeras, contrasted with the same change in the WT chimeras, while also accounting for batch effects across the different pools of chimeras, we used the following negative binomial mixed effects model, applying the NEBULA[10] approach:

$$y_{ij} \sim NB(\mu_{ij} = \pi_j \cdot (\alpha_i + b_{1,i} \cdot m_j + b_{2,i} \cdot t_j + b_{3,i} \cdot m_j \cdot t_j + log(w_i | p_j), \phi_i)$$

where y_{ij} is the expression of gene *i* in cell *j*, m_j is binary with $m_j = 1$ if the cell is marked by the fluorescent marker (i.e. a tdTom$^+$ cell), $t_j = 1$ if the cell is part of the knockout chimeras (as opposed to the control chimeras), π_j is the size-factor of cell j (computed using the scran[67] Bioconductor package as part of normalisation) and ϕ_i is the dispersion parameter for the negative Binomial distribution for gene *i*. w_i is a random effect at the level of the pool p_j containing cell j. We then test whether $b_{3,i}$ is significantly different from 0, as this means a significant difference between the knockout and WT chimeras.

Treatment of batch effects across pools of chimeras

Chimeras were obtained in pools of several embryos (see Section *Mixl1$^{-/-}$ embryo chimera data generation*), without information concerning individual embryos. Therefore, we were not able to model the effect of inter-embryo variation. We checked consistency across pools for DA testing (Figure S2), reduced the potential impact of batch effects across pools for COSICC_DA_kinetics by computing pseudotimes based on reference-based temporal genes (see Section *COSICC_kinetics: testing for acceleration or delay of knockout cells along a trajectory*), and explicitly modelled the batch structure for DE testing using mixed effects models.

1 ***JCF scores***

2 JCF scores were obtained by averaging the expression levels of JCF genes
3 (https://crukci.shinyapps.io/heartAtlas/, cluster marker genes for me5) obtained from the single-
4 cell atlas of the mouse embryonic heart[30].

5

6 ***Sub-clustering of mesenchyme cell type***

7 We used Louvain clustering[68] on a nearest neighbour graph with k=20 neighbours. To enable
8 detection of genes differentially expressed across sub-clusters without inflation of the false
9 positive rate, we used only 50% of the cells in the clustering algorithm and assigned the remaining
10 cells to the clusters using support vector machines[69]. We then applied COSICC_DA_group to the
11 subgroups of the mesenchyme cell type in the same way as to cell types.

12

13
Supplemental Information

14 **Document S1.** Figures S1–S7 ~~and Note S1~~.

15 **Table S1.** Results from analysis of $T^{-/-}$ and $Mixl1^{-/-}$ data sets.

16 **Table S2.** All n values in statistical analyses relevant for Figures 2 and 3.

17

18

19
References

20

21 1. Replogle, J.M., Saunders, R.A., Pogson, A.N., Hussmann, J.A., Lenail, A., Guna, A.,
22 Mascibroda, L., Wagner, E.J., Adelman, K., Lithwick-Yanai, G., et al. (2022). Mapping information-

1 rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell *185*, 2559-+.
2 10.1016/j.cell.2022.05.013.

3 2. Jin, X., Simmons, S.K., Guo, A., Shetty, A.S., Ko, M., Nguyen, L., Jokhi, V., Robinson, E.,
4 Oyler, P., Curry, N., et al. (2020). In vivo Perturb-Seq reveals neuronal and glial abnormalities
5 associated with autism risk genes. Science *370*. 10.1126/science.aaz6063.

6 3. Fleck, J.S., Jansen, S.M.J., Wollny, D., Zenk, F., Seimiya, M., Jain, A., Okamoto, R., Santel,
7 M., He, Z., Camp, J.G., and Treutlein, B. (2023). Inferring and perturbing cell fate regulomes in
8 human brain organoids. Nature *621*, 365-372. 10.1038/s41586-022-05279-8.

9 4. Guibentif, C., Griffiths, J.A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V.,
10 Gottgens, B., and Marioni, J.C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo.
11 Dev Cell *56*, 141-153 e146. 10.1016/j.devcel.2020.11.013.

12 5. Mittnenzweig, M., Mayshar, Y., Cheng, S., Ben-Yair, R., Hadas, R., Rais, Y., Chomsky, E.,
13 Reines, N., Uzonyi, A., Lumerman, L., et al. (2021). A single-embryo, single-cell time-resolved
14 model for mouse gastrulation. Cell *184*, 2825-2842 e2822. 10.1016/j.cell.2021.04.004.

15 6. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J.,
16 Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of
17 mouse gastrulation and early organogenesis. Nature *566*, 490-495. 10.1038/s41586-019-0933-9.

18 7. Pereira, L.A., Wong, M.S., Mei Lim, S., Stanley, E.G., and Elefanty, A.G. (2012). The Mix
19 family of homeobox genes--key regulators of mesendoderm formation during vertebrate
20 development. Dev Biol *367*, 163-177. 10.1016/j.ydbio.2012.04.033.

21 8. Hart, A.H., Hartley, L., Sourris, K., Stadler, E.S., Li, R.L., Stanley, E.G., Tam, P.P.L., Elefanty,
22 A.G., and Robb, L. (2002). is required for axial mesendoderm morphogenesis and patterning in
23 the murine embryo. Development *129*, 3597-3608.

24 9. Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J.,
25 Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression
26 Identifies Developmental Trajectories in Reprogramming. Cell *176*, 1517.
27 10.1016/j.cell.2019.02.026.

28 10. He, L., Davila-Velderrain, J., Sumida, T.S., Hafler, D.A., Kellis, M., and Kulminski, A.M.
29 (2021). NEBULA is a fast negative binomial mixed model for differential or co-expression analysis
30 of large- scale multi-subject single-cell data. Commun Biol *4*. ARTN 629 10.1038/s42003-021-
31 02146-6.

32 11. Imaz-Rosshandler, I., Rode, C., Guibentif, C., Harland, L.T.G., Ton, M.L.N., Dhapola, P.,
33 Keitley, D., Argelaguet, R., Calero-Nieto, F.J., Nichols, J., et al. (2024). Tracking early mammalian
34 organogenesis - prediction and validation of differentiation trajectories at whole organism scale.
35 Development *151*. ARTN dev201867 10.1242/dev.201867.

1 12. Lawson, K.A., and Wilson, V. (2016). 3 - A Revised Staging of Mouse Development Before
2 Organogenesis. In Kaufman's Atlas of Mouse Development Supplement, R. Baldock, J. Bard, D.R.
3 Davidson, and G. Morriss-Kay, eds. (Academic Press), pp. 51-64. https://doi.org/10.1016/B978-
4 0-12-800043-4.00003-8.

5 13. Wilson, V., Manson, L., Skarnes, W.C., and Beddington, R.S.P. (1995). The T-Gene Is
6 Necessary for Normal Mesodermal Morphogenetic Cell Movements during Gastrulation.
7 Development *121*, 877-886. 10.1242/dev.121.3.877.

8 14. Du, R., Wu, S.S., Lv, X.N., Fang, H.H., Wu, S.D., and Kang, J.B. (2014). Overexpression of
9 brachyury contributes to tumor metastasis by inducing epithelial-mesenchymal transition in
10 hepatocellular carcinoma. J Exp Clin Canc Res *33*. ARTN 105 10.1186/s13046-014-0105-6.

11 15. Hamilton, D.H., David, J.M., Dominguez, C., and Palena, C. (2017). Development of Cancer
12 Vaccines Targeting Brachyury, a Transcription Factor Associated with Tumor Epithelial-
13 Mesenchymal Transition. Cells Tissues Organs *203*, 128-138. 10.1159/000446495.

14 16. Huang, B., Cohen, J.R., Fernando, R.I., Hamilton, D.H., Litzinger, M.T., Hodge, J.W., and
15 Palena, C. (2013). The embryonic transcription factor Brachyury blocks cell cycle progression and
16 mediates tumor resistance to conventional antitumor therapies. Cell Death Dis *4*. ARTN e682
17 10.1038/cddis.2013.208.

18 17. Gluecksohn-Schoenheimer, S. (1944). The development of normal and homozygous
19 brachy (T/T) mouse embryos in the extraembryonic coelom of the chick. P Natl Acad Sci USA *30*,
20 134-140. DOI 10.1073/pnas.30.6.134.

21 18. Ihry, R.J., Salick, M.R., Ho, D.J., Sondey, M., Kommineni, S., Paula, S., Raymond, J., Henry,
22 B., Frias, E., Wang, Q., et al. (2019). Genome-Scale CRISPR Screens Identify Human Pluripotency-
23 Specific Genes. Cell Reports *27*, 616-+. 10.1016/j.celrep.2019.03.043.

24 19. Powers, S.E., Taniguchi, K., Yen, W.W., Melhuish, T.A., Shen, J., Walsh, C.A., Sutherland,
25 A.E., and Wotton, D. (2010). Tgif1 and Tgif2 regulate Nodal signaling and are required for
26 gastrulation. Development *137*, 249-259. 10.1242/dev.040782.

27 20. Zhu, W.D., Shiojima, I., Ito, Y., Li, Z., Ikeda, H., Yoshida, M., Naito, A.T., Nishi, J., Ueno, H.,
28 Umezawa, A., et al. (2014). IGFBP-4 is an inhibitor of canonical Wnt signalling required for
29 cardiogenesis (vol 454, pg 345, 2008). Nature *506*, 254-254. 10.1038/nature13003.

30 21. Taguchi, A., Kaku, Y., Ohmori, T., Sharmin, S., Ogawa, M., Sasaki, H., and Nishinakamura,
31 R. (2014). Redefining the In Vivo Origin of Metanephric Nephron Progenitors Enables Generation
32 of Complex Kidney Structures from Pluripotent Stem Cells. Cell Stem Cell *14*, 53-67.
33 10.1016/j.stem.2013.11.010.

34 22. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B.,
35 Macura, K., and Herrmann, B.G. (2017). Antagonistic Activities of Sox2 and Brachyury Control the

Fate Choice of Neuro-Mesodermal Progenitors. Developmental Cell *42*, 514-+. 10.1016/j.devcel.2017.07.021.

23. Liu, C.Q., Nakamura, E., Knezevic, V., Hunter, S., Thompson, K., and Mackem, S. (2003). A role for the mesenchymal T-box gene Brachyury in AER formation during limb development. Development *130*, 1327-1337. 10.1242/dev.00354.

24. Papaioannou, V.E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. Development *141*, 3819-3833. 10.1242/dev.104471.

25. Dunwoodie, S.L., Rodriguez, T.A., and Beddington, R.S.P. (1998). Msg1 and Mrg1, founding members of a gene family, show distinct patterns of gene expression during mouse embryogenesis. Mech Develop *72*, 27-40. Doi 10.1016/S0925-4773(98)00011-2.

26. Zhang, B., He, P., Lawrence, J.E.G., Wang, S.Y., Tuck, E., Williams, B.A., Roberts, K., Kleshchevnikov, V., Mamanova, L., Bolt, L., et al. (2023). A human embryonic limb cell atlas resolved in space and time. Nature. 10.1038/s41586-023-06806-x.

27. Xu, B., Hrycaj, S.M., McIntyre, D.C., Baker, N.C., Takeuchi, J.K., Jeannotte, L., Gaber, Z.B., Novitch, B.G., and Wellik, D.M. (2013). Hox5 interacts with Plzf to restrict Shh expression in the developing forelimb. P Natl Acad Sci USA *110*, 19438-19443. 10.1073/pnas.1315075110.

28. Ota, S., Zhou, Z.Q., Keene, D.R., Knoepfler, P., and Hurlin, P.J. (2007). Activities of N-Myc in the developing limb link control of skeletal size with digit separation. Development *134*, 1583-1592. 10.1242/dev.000703.

29. Pulina, M.V., Sahr, K.E., Nowotschin, S., Baron, M.H., and Hadjantonakis, A.K. (2014). A Conditional Mutant Allele for Analysis of Mixl1 Function in the Mouse. Genesis *52*, 417-423. 10.1002/dvg.22768.

30. Tyser, R.C.V., Ibarra-Soria, X., McDole, K., Jayaram, S.A., Godwin, J., van den Brand, T.A.H., Miranda, A.M.A., Scialdone, A., Keller, P.J., Marioni, J.C., and Srinivas, S. (2021). Characterization of a common progenitor pool of the epicardium and myocardium. Science *371*, 1011-+. ARTN eabb2986 10.1126/science.abb2986.

31. Beddington, R.S.P., Rashbass, P., and Wilson, V. (1992). Brachyury - a Gene Affecting Mouse Gastrulation and Early Organogenesis. Development, 157-165. https://doi.org/10.1242/dev.116.Supplement.157

32. Chesley, P. (1935). Development of the short-tailed mutant in the house mouse. J Exp Zool *70*, 429-459. DOI 10.1002/jez.1400700306.

33. Replogle, J.M., Norman, T.M., Xu, A., Hussmann, J.A., Chen, J., Cogan, J.Z., Meer, E.J., Terry, J.M., Riordan, D.P., Srinivas, N., et al. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nature Biotechnology *38*, 954-+. 10.1038/s41587-020-0470-y.

1 34. Saunders, L.M., Srivatsan, S.R., Duran, M., Dorrity, M.W., Ewing, B., Linbo, T.H., Shendure,
2 J., Raible, D.W., Moens, C.B., Kimelman, D., and Trapnell, C. (2023). Embryo-scale reverse
3 genetics at single-cell resolution. Nature *623*, 782-+. 10.1038/s41586-023-06720-2.

4 35. Cao, J.Y., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A.,
5 Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene
6 expression. Science *370*, 808-+. ARTN eaba7721 10.1126/science.aba7721.

7 36. Ton, M.L.N., Keitley, D., Theeuwes, B., Guibentif, C., Ahnfelt-Ronne, J., Andreassen, T.K.,
8 Calero-Nieto, F.J., Imaz-Rosshandler, I., Pijuan-Sala, B., Nichols, J., et al. (2023). An atlas of rabbit
9 development as a model for single-cell comparative genomics. Nature Cell Biology *25*, 1061-+.
10 10.1038/s41556-023-01174-0.

11 37. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B.,
12 Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. Elife *6*.
13 10.7554/eLife.27041.

14 38. Tabula Sapiens, C., Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R.,
15 Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ,
16 single-cell transcriptomic atlas of humans. Science *376*, eabl4896. 10.1126/science.abl4896.

17 39. Ushijima, T., Clark, S.J., and Tan, P. (2021). Mapping genomic and epigenomic evolution
18 in cancer ecosystems. Science *373*, 1474-1479. 10.1126/science.abh1645.

19 40. Zhang, Y., Wang, D., Peng, M., Tang, L., Ouyang, J., Xiong, F., Guo, C., Tang, Y., Zhou, Y.,
20 Liao, Q., et al. (2021). Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res *40*, 81.
21 10.1186/s13046-021-01874-1.

22 41. Chen, S., Giannakou, A., Wyman, S., Gruzas, J., Golas, J., Zhong, W., Loreth, C., Sridharan,
23 L., Yamin, T.T., Damelin, M., and Geles, K.G. (2018). Cancer-associated fibroblasts suppress SOX2-
24 induced dysplasia in a lung squamous cancer coculture. Proc Natl Acad Sci U S A *115*, E11671-
25 E11680. 10.1073/pnas.1803718115.

26 42. Khan, A.O., Rodriguez-Romera, A., Reyat, J.S., Olijnik, A.A., Colombo, M., Wang, G., Wen,
27 W.X., Sousos, N., Murphy, L.C., Grygielska, B., et al. (2023). Human Bone Marrow Organoids for
28 Disease Modeling, Discovery, and Validation of Therapeutic Targets in Hematologic Malignancies.
29 Cancer Discov *13*, 364-385. 10.1158/2159-8290.CD-22-0199.

30 43. Patikas, N., Ansari, R., and Metzakopian, E. (2023). Single-cell transcriptomics identifies
31 perturbed molecular pathways in midbrain organoids using alpha-synuclein triplication
32 Parkinson's disease patient-derived iPSCs. Neurosci Res *195*, 13-28.
33 10.1016/j.neures.2023.06.001.

34 44. Popova, G., Soliman, S.S., Kim, C.N., Keefe, M.G., Hennick, K.M., Jain, S., Li, T., Tejera, D.,
35 Shin, D., Chhun, B.B., et al. (2021). Human microglia states are conserved across experimental

1 models and regulate neural stem cell responses in chimeric organoids. Cell Stem Cell *28*, 2153-
2 2166 e2156. 10.1016/j.stem.2021.08.015.

3 45. Pérez-Posada, A., García-Castro, H., Emili, E., Vanni, V., Arias-Baldrich, C., Frölich, S., van
4 Heeringen, S.J., Kenny, N., and Solana, J. (2025). The Regulatory Logic of Planarian Stem Cell
5 Differentiation. bioRxiv, 2024.2008.2023.608747. 10.1101/2024.08.23.608747.

6 46. van Galen, P., Hovestadt, V., Wadsworth, M.H., Hughes, T.K., Griffin, G.K., Battaglia, S.,
7 Verga, J.A., Stephansky, J., Pastika, T.J., Story, J.L., et al. (2019). Single-Cell RNA-Seq Reveals AML
8 Hierarchies Relevant to Disease Progression and Immunity. Cell *176*, 1265-+.
9 10.1016/j.cell.2019.01.031.

10 47. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-
11 cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol
12 *36*, 421-427. 10.1038/nbt.4091.

13 48. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J.,
14 Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data.
15 Cell *184*, 3573-3587 e3529. 10.1016/j.cell.2021.04.048.

16 49. Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq
17 data across data sets. Nat Methods *15*, 359-362. 10.1038/nmeth.4644.

18 50. Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., Buttner, M., Wagenstetter,
19 M., Avsec, Z., Gayoso, A., Yosef, N., Interlandi, M., et al. (2022). Mapping single-cell data to
20 reference atlases by transfer learning. Nat Biotechnol *40*, 121-130. 10.1038/s41587-021-01001-
21 7.

22 51. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M.,
23 Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat
24 Methods *19*, 159-170. 10.1038/s41592-021-01346-6.

25 52. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D., and Marioni, J.C. (2022).
26 Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat
27 Biotechnol *40*, 245-253. 10.1038/s41587-021-01033-z.

28 53. Fujii, M., Sakaguchi, A., Kamata, R., Nagao, M., Kikuchi, Y., Evans, S.M., Yoshizumi, M.,
29 Shimono, A., Saga, Y., and Kokubo, H. (2017). 5 identifies murine cardiac progenitors for all
30 myocardial structures except for the right ventricle. Nature Communications *8*. ARTN 14664
31 10.1038/ncomms14664.

32 54. Prummel, K.D., Hess, C., Nieuwenhuize, S., Parker, H.J., Rogers, K.W., Kozmikova, I.,
33 Racioppi, C., Brombacher, E.C., Czarkwiani, A., Knapp, D., et al. (2019). A conserved regulatory
34 program initiates lateral plate mesoderm emergence across chordates. Nat Commun *10*, 3857.
35 10.1038/s41467-019-11561-7.

1 55. Strauss, M., Ton, M.-L.N., Mason, S., Bagri, J., Imaz-Rosshandler, I., Harland, L.T.G., Wilson,
2 N., Nichols, J., Tyser, R., Goettgens, B., et al. (2025). Software and processed data sets
3 accompanying the publication A single cell and tissue-scale analysis suite resolves Mixl1's role in
4 heart development. Zenodo. 10.5281/zenodo.15008255.

5 56. Griffiths, J., and Lun, A. (2019). Single-Cell Transcriptomics Data across Mouse
6 Gastrulation and Early Organogenesis. Bioconductor. 10.18129/B9.bioc.MouseGastrulationData.

7 57. Santini, L., Halbritter, F., Titz-Teixeira, F., Suzuki, T., Asami, M., Ma, X., Ramesmayer, J.,
8 Lackner, A., Warr, N., Pauler, F., et al. (2021). Genomic imprinting in mouse blastocysts is
9 predominantly associated with H3K27me3. Nature Communications *12*. ARTN 3804
10 10.1038/s41467-021-23510-4.

11 58. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and
12 Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.
13 10.1038/nature06968.

14 59. Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A.
15 (2011). Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency
16 network and increases embryonic stem cell resistance to differentiation. Nat Cell Biol *13*, 838-
17 845. 10.1038/ncb2267.

18 60. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome
19 engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308. 10.1038/nprot.2013.143.

20 61. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical
21 and Powerful Approach to Multiple Testing. J R Stat Soc B *57*, 289-300. DOI 10.1111/j.2517-
22 6161.1995.tb02031.x.

23 62. Prates, M.O., Cabral, C.R.B., and Lachos, V.H. (2013). mixsmsn: Fitting Finite Mixture of
24 Scale Mixture of Skew-Normal Distributions. J Stat Softw *54*, 1-20. 10.18637/jss.v054.i12

25 63. Schwarz, G. (1978). Estimating the Dimension of a Model. Ann. Statist. *6*, 461-464.
26 10.1214/aos/1176344136.

27 64. Coifman, R.R., and Lafon, S. (2006). Diffusion maps. Appl Comput Harmon A *21*, 5-30.
28 10.1016/j.acha.2006.04.006.

29 65. Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional
30 single-cell analysis of differentiation data. Bioinformatics *31*, 2989-2998.
31 10.1093/bioinformatics/btv325.

32 66. Mann, H.B., and Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables
33 is Stochastically Larger than the Other. Ann. Math. Statist. *18*, 50-60. 10.1214/aoms/1177730491.

1 67. Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level
2 analysis of single-cell RNA-seq data with Bioconductor. F1000Res *5*, 2122.
3 10.12688/f1000research.9501.2.

4 68. Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of
5 communities in large networks. J Stat Mech-Theory E. Artn P10008 10.1088/1742-
6 5468/2008/10/P10008.

7 69. Zhang, J.M., Kamath, G.M., and Tse, D.N. (2019). Valid Post-clustering Differential Analysis
8 for Single-Cell RNA-Seq. Cell Syst *9*, 383-392 e386. 10.1016/j.cels.2019.07.012.

9

10

CellPress

KEY RESOURCES TABLE

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Proteinase K	Sigma	Sigma
Phusion High-Fidelity DNA Polymerase	NEB	M0530S
Lipofectamine 3000 Transfection Reagent	ThermoFisher Scientific	L3000008
TrypLE Express dissociation reagent	Thermo Fisher Scientific	12604013
Critical commercial assays		
Nextera XT Kit	Illumina	15052163
Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3	10X Genomics	PN-1000075
Chromium Chip B Single Cell Kit	10X Genomics	PN-1000073
Deposited data		
E8.5 *T*$^{-/-}$ chimeras scRNAseq data	Guibentif et al. [4]	ArrayExpress: E-MTAB-8811
E8.5 WT chimeras scRNAseq data	Guibentif et al. [4]	ArrayExpress: E-MTAB-8812
extended mouse gastrulation scRNAseq raw data	Imaz-Rosshandler et al. [11]	ArrayExpress E-MTAB-11763
extended mouse gastrulation scRNAseq processed data, including metadata and WOT trajectory inference	Imaz-Rosshandler et al. [11]	https://marionilab.github.io/ExtendedMouseAtlas/
E8.5 *Mixl1*$^{-/-}$ chimeras scRNAseq data	This paper	ArrayExpress E-MTAB-13409
GRCm38 Mouse reference genome Build 38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001635.20/
Annotated AML data	Van Galen et al. [46]	GEO: GSE116256
Experimental models: Cell lines		
TdTomato-expressing mouse embryonic stem cells	Pijuan-Sala et al. [6]	N/A
Experimental models: Organisms/strains		
C57BL/6 wild type mice	Charles River	C57BL/6J (JAX™ Mice Strain)
Oligonucleotides		
Mixl1-targetting guide pair 1 guide 1: AAGCGGCGCCTTCTGCGAAC	This paper	N/A
Mixl1-targetting guide pair 1 guide 2: TGCTGGGGCGCGAGAGTCGT	This paper	N/A
Mixl1-targetting guide pair 2 guide 1: TTGCGGCGCTGTGGCGCCGA	This paper	N/A
Mixl1-targetting guide pair 2 guide 2: CGCTCCCGCAAGTGGATGTC	This paper	N/A

primer including Nextera overhangs for guide pair 1: F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTATTCCCGCGGCGTCT	This paper	N/A
primer including Nextera overhangs for guide pair 1: R-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCGAGCTGAACGACGT	This paper	N/A
primer including Nextera overhangs for guide pair 2: F-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAGCTCCAGTTCGCAGA	This paper	N/A
primer including Nextera overhangs for guide pair 2: R-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAGTTTGCAGTCTAGAACC	This paper	N/A
Recombinant DNA		
pX458 plasmid	Ran et al. [60]	Addgene, #48138
Software and algorithms		
MouseGastrulationData Bioconductor package	Griffiths [56]	https://bioconductor.org/packages/release/data/experiment/html/MouseGastrulationData.html
Cell Ranger v. 6.01	10x Genomics	N/A
R v.4.1.1	The R Project	https://www.r-project.org/
COSICC code	This paper	https://github.com/MarioniLab/analysis_chimera_data https://doi.org/10.5281/zenodo.15008255
COSICC package	This paper	https://github.com/MarioniLab/COSICC https://doi.org/10.5281/zenodo.15008255
Other		
Novaseq6000 sequencing platform	Illumina	N/A

Figure 1

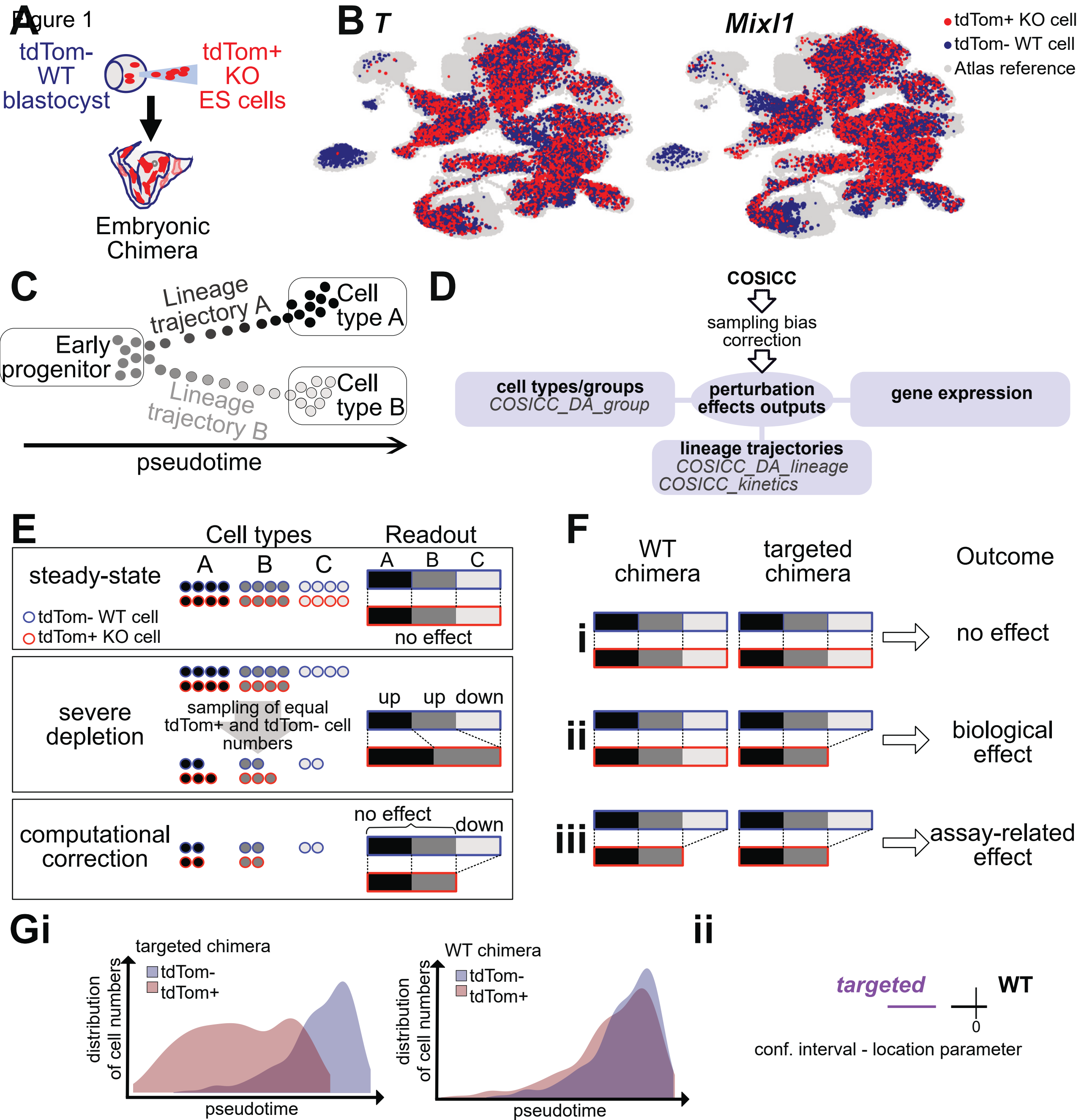


Figure 2

A

$T^{-/-}$ chimeras - per cell type DA

enriched
depleted
not significant

Allantois
Presomitic mesoderm
Allantois endothelium
Somitic mesoderm
YS mesothelium
Intermediate mesoderm
Mesenchyme
NMP
Notochord
Lateral plate mesoderm
Caudal Epiblast
Venous endothelium
Hindbrain floor plate
Erythroid
Embryo proper endothelium

- Log_{10}(FDR)
20
15
10
5
0

Log_{10}(odds ratio)
−1.0
−0.5
0.0
0.5

B

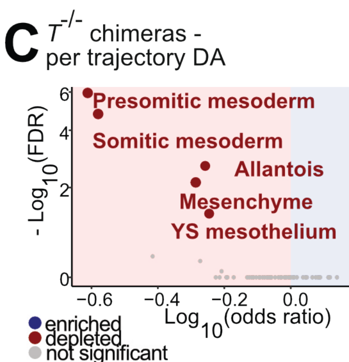
$T^{-/-}$ chimeras-per cell type and stage DA

Dermomyotome
NMPs
Cranial mesoderm
Venous endothelium
Caudal epiblast
NMPs/Mesoderm-biased
Anterior somitic tissues
Blood progenitors
Notochord
Posterior somitic tissues
Spinal cord progenitors
Embryo proper endothelium
Sclerotome
Erythroid
YS endothelium
Allantois endothelium
Lateral plate mesoderm
Mesenchyme
Allantois
Hindbrain floor plate
YS mesothelium
Intermediate mesoderm
Somitic mesoderm
Presomitic mesoderm

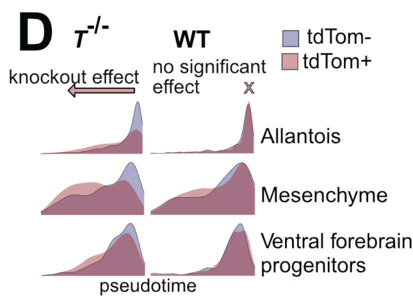
E7.75
E8.0
E8.25
E8.5
E8.75

Log_2 of odds ratio
(−Inf,−3]
(−3,−1.5]
(−1.5,0]
(0,1.5]
(1.5,3]
(3, Inf]

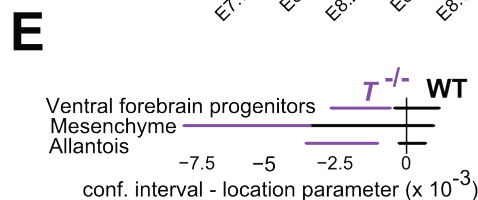
C



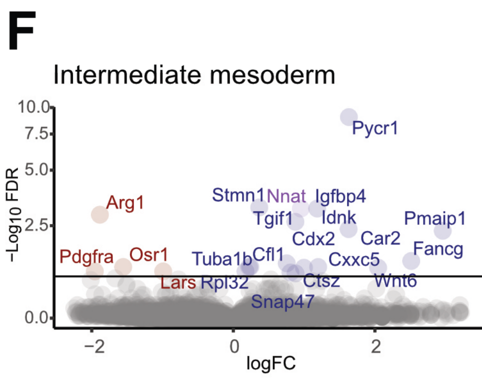
D



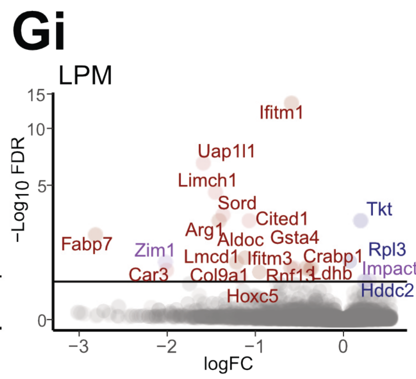
E



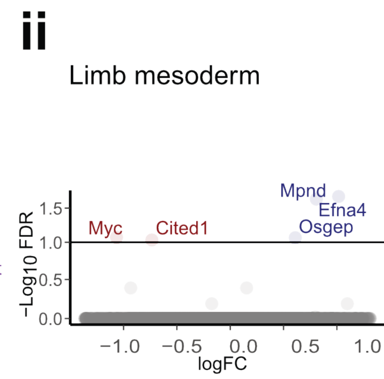
F



Gi



ii



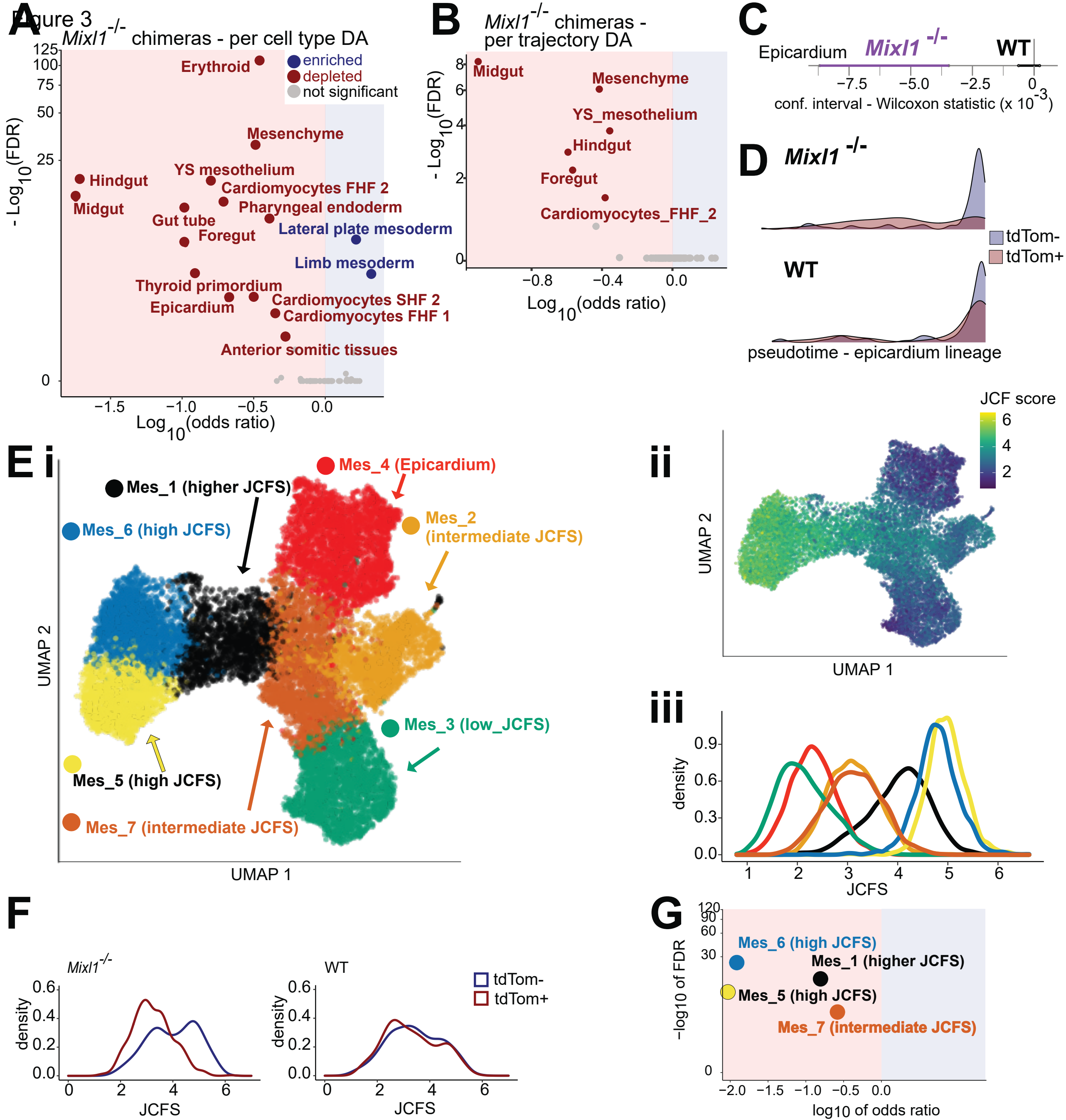


Figure 3
A
Mixl1 -/- chimeras - per cell type DA
enriched
depleted
not significant
Erythroid
Mesenchyme
Hindgut
Midgut
YS mesothelium
Cardiomyocytes FHF 2
Pharyngeal endoderm
Gut tube
Foregut
Lateral plate mesoderm
Limb mesoderm
Thyroid primordium
Epicardium
Cardiomyocytes SHF 2
Cardiomyocytes FHF 1
Anterior somitic tissues
- Log10(FDR)
Log10(odds ratio)
B
Mixl1 -/- chimeras - per trajectory DA
Midgut
Mesenchyme
YS_mesothelium
Hindgut
Foregut
Cardiomyocytes_FHF_2
- Log10(FDR)
Log10(odds ratio)
C
Epicardium
Mixl1 -/-
WT
conf. interval - Wilcoxon statistic (x 10-3)
D
Mixl1 -/-
WT
tdTom-
tdTom+
pseudotime - epicardium lineage
E i
Mes_4 (Epicardium)
Mes_1 (higher JCFS)
Mes_6 (high JCFS)
Mes_2 (intermediate JCFS)
Mes_3 (low_JCFS)
Mes_5 (high JCFS)
Mes_7 (intermediate JCFS)
UMAP 2
UMAP 1
ii
JCF score
UMAP 2
UMAP 1
iii
density
JCFS
F
Mixl1 -/-
WT
tdTom-
tdTom+
density
JCFS
G
Mes_6 (high JCFS)
Mes_1 (higher JCFS)
Mes_5 (high JCFS)
Mes_7 (intermediate JCFS)
-log10 of FDR
log10 of odds ratio