Systematic characterisation of perturbations in complex developing cell populations reveals mechanisms underpinning development and disease

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

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK
2 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
3 Department of Haematology, University of Cambridge, Cambridge, UK
4 Wellcome-Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
5 Medical Research Council Laboratory of Molecular Biology, Cambridge, UK
6 Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK
7 Present Address: Genentech, South San Francisco, CA, USA
8 Inst. Biomedicine, Dept. Microbiology and Immunology, Sahlgrenska Center for Cancer Research, University of Gothenburg, Gothenburg, Sweden
* corresponding authors: bg200@cam.ac.uk, marioni@ebi.ac.uk, carolina.guibentif@gu.se

Abstract

Perturbation studies using gene knockouts have become a key tool for understanding the roles of regulatory genes in development and disease. Here we systematically characterise the effects of the knockout of the key developmental regulators *T* and *Mixl1* in chimeric mouse embryos during gastrulation and organogenesis. We present a comprehensive and effective suite of tools for systematic characterisation of effects at the level of differential abundance of cell types, lineage development, and gene dysregulation. We apply this strategy to a novel chimera data set with *Mixl1* knockout, and reveal a disruption in epicardium development in the absence of *Mixl1*, characterized by lack of upregulation along the lineage of the key transcription factor *Tbx18* and Wnt regulator *Sfrp5.* Finally, we demonstrate the broad utility of our perturbSuite framework for testing progression and differential expression along lineages by applying it to published AML

patient data. Specifically, we show how different responses to therapy are reflected in changes in gene expression along the myeloid trajectory between healthy and AML patients.

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(Replogle et al. 2022), others have focused on smaller knockouts in more complex structures, such as mouse models(Jin et al. 2020) or organoids(Fleck et al. 2022).

T is the founding member of the T-box family of transcription factors and has a conserved fundamental role in early embryonic development, notably during the establishment of the anterior-posterior axis, and in primitive streak cell migration during posterior mesoderm specification [reviewed in (Naiche et al. 2005)]. The homeobox transcription factor *Mixl1* is essential for the patterning of axial mesendodermal structures during early embryogenesis (Hart et al. 2002)]. In vivo evidence demonstrates highly context-specific regulatory functions of these genes(Goering et al. 2003; Pereira et al. 2012; Jackson et al. 2010), and recent single-cell -based approaches have confirmed a differential lineage sensitivity to essential transcription factor knockout, namely in the context of chimera assays.

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 (Pijuan-Sala et al. 2019; Guibentif et al. 2021; Mittnenzweig et al. 2021)(Fig 1A). In the resulting chimeras, WT cells develop normally, establishing signalling gradients that are necessary for the embryo to develop, while the effects of the knockout can be observed by studying the descendants of the injected mutant cells. This approach can reveal cell-autonomous effects of essential gene knockouts. Using single-cell profiling as a readout, and taking advantage of developmental atlases as reference datasets, enables the study of such perturbations well beyond differences in organ contribution (Fig1B, Supp. Fig. 1AB). However, the impact on lineage progression, on gene expression, and the interaction of these effects remains challenging to assess.

A number of previous studies used chimeric mouse embryos to study the effects of essential gene knockouts of *Tal1(Pijuan-Sala et al. 2019)* , *T (Guibentif et al. 2021)*, *Gata1(Barile et al. 2021)*, and *Foxc1 and Foxc2* genes(Mittnenzweig et al. 2021). The single-cell readout of these studies presents specific challenges, related eg. to sampling bias and data interpretation. To address these challenges, we herein present perturbSuite, a novel set of computational tools specifically designed for the analysis of perturbations in complex developing cell populations, and therefore ideally suited to chimera studies. We validate the proposed analysis framework using chimeric *T/Brachyury* embryos presented earlier(Guibentif et al. 2021) and present novel *Mixl1*$^{-/-}$ chimeric mouse embryo data. Finally, we extend the application of perturbSuite to uncover mechanisms of response to therapy in the context of leukemia progression.

[Results

PerturbSuite leverages single-cell profiling to analyse developmental perturbations comprehensively

We considered two independent chimera datasets, both generated at Embryonic Day (E) 8.5 of mouse development: i) a previously published dataset that studied the cell autonomous function of *T/Brachyury*; ii) a newly generated dataset created to study the role of *Mixl1*. Additionally, we consider data from a control experiment, where wildtype (WT) cells were injected at the blastocyst stage to create a control chimera. In both chimeras, since the mutant cells were labelled with the fluorescent marker tdTomato, we define the progeny of injected cells as tomato positive cells (tdTom$^+$, Fig. 1AB). The wildtype (host) embryonic cells are referred to as tomato negative (tdTom$^-$) as they lack expression of a fluorescent marker.

Profiling the chimeric embryos using single-cell transcriptomics (scRNA-seq), in conjunction with reference single-cell atlases (Imaz-Rosshandler et al. 2023) of early development opens exciting possibilities for studying the effects of perturbations during cellular diversification (Fig1B, Supp. Fig. 1ABC) For instance, we can explore whether the perturbed cells are depleted or enriched in certain cell types or lineages as well as assessing how far along specific lineages perturbed cells are able to progress (Fig1CD). Studying perturbation in complex systems like embryonic chimeras requires attention to specific challenges. First, if a perturbation leads to a strong reduction of some cell types, then the un-affected cell types will appear as if they were enriched in the perturbed cell population, when compared to the non-perturbed population. To account for this bias we developed a subsampling-based

normalisation procedure for differential abundance testing (Fig. 1E). Second, to control for both batch and experimental effects, any analysis must be framed in terms of two controls: first, an internal control of unperturbed host cells from the same sample, and second, an external control, e.g. the wildtype (WT) chimeras introduced earlier. Since the generation of the chimera leads to transcriptional differences between the wildtype injected $tdTom^+$ cells and the $tdTom^-$ wildtype host cells, this needs to be accounted for (Supp. Fig. 1CD), in order to distinguish experimental from biological effects (Supp. Fig. 1CD). Therefore, at all levels of analysis we always utilised both external and internal controls.

PerturbSuite performs differential abundance (perturbSuite_DA) and differential expression (perturbSuite_DE) testing using the two levels of control introduced above, thereby accounting for both batch and experimental effects (Fig. 1F) and enabling the identification of dynamic shifts in lineage development (perturbSuite_kinetics; Fig. 1G) as well as differential expression of genes and activity of pathways (Fig. 1H).

As well as the new computational strategies, compared to previous work(Guibentif et al. 2021), we also use a substantially expanded reference atlas relative to previous work (e.g., Guibentif et al., 2021) that contains four-times as many cells and adds four additional timepoints between E8.5 and E9.5 that more completely represents early organogenesis (Fig. 1B and Supp. Fig. 1ABC). This allowed us to identify chimeric cells that were similar to cells present in later stage embryos (E8.75-E9.5) that were not sampled in the(Pijuan-Sala et al. 2019; Guibentif et al. 2021) reference dataset (Supp. Fig. 1D).

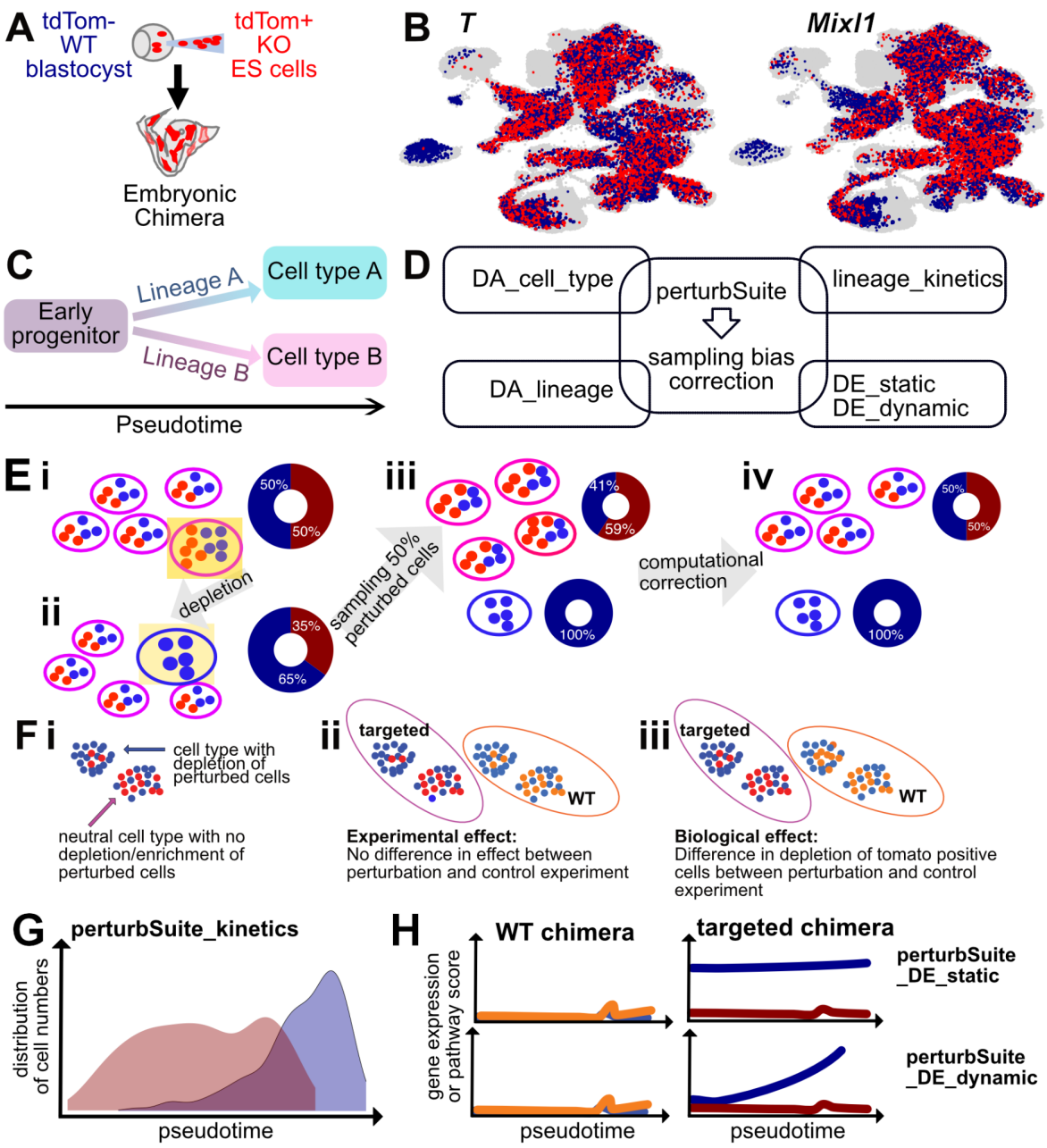


Fig. 1: An overview of Chimeras and perturbSuite. A) Chimeric embryos are created by injecting mutant cells into wildtype (WT) embryos at the blastocyst stage. B) Chimeric mouse embryos mapped to the extended mouse gastrulation atlas, tdTom$^+$ cells are in red, tomato negative ones in blue. C) Illustration of lineages. D) Overview of perturbSuite. E) PerturbSuite corrects for sampling bias; i) Example cell counts from an embryo with knockout cells in red, each cell type represented by a ellipsoid; ii) depletion of one cell type only iii) this depletion is unknown before data analysis, and therefore a fixed proportion of cells with fluorescent markers is sampled, leading to cell types seemingly enriched for knockout cells (iv) computational correction F) tdTom$^+$ cells (red) are underrepresented in one of the two cell types, ii) experimental effect: tdTom$^+$ cells are underrepresented for

the knockout chimeras only to a similar degree as for the WT chimeras, iii) biological effect G) Illustration of perturbSuite_kinetics. H) Static (top) and dynamic differential expression (bottom): changes along a trajectory are different for tdTom$^+$ versus tdTom$^-$ cells of the knockout chimera (right) and this difference is significantly larger than for the WT control experiment (left).

PerturbSuite characterises impact of *T* knockout on embryonic development and reveals its role for limb development

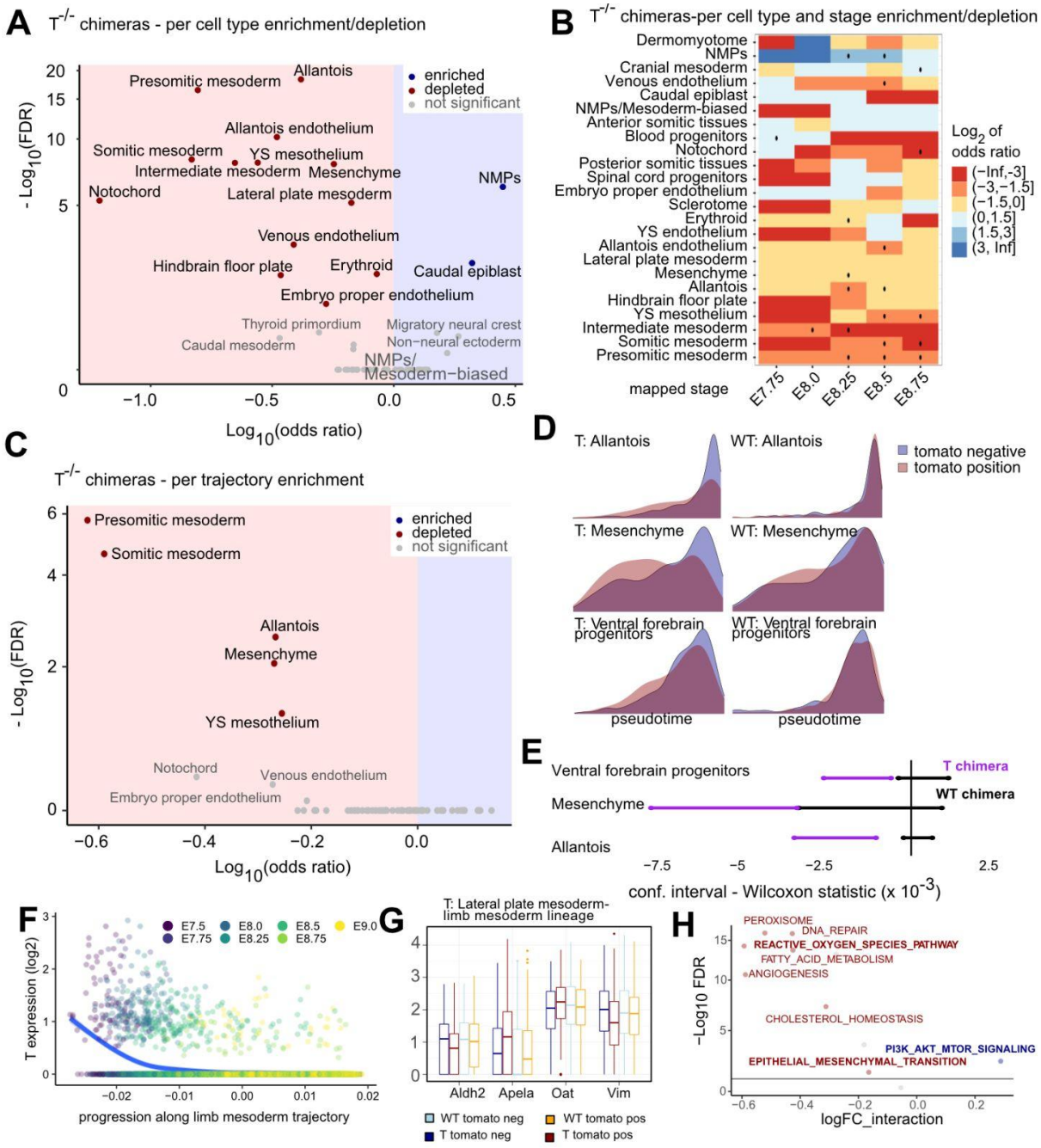


Fig. 2: PerturbSuite confirms existing results and reveals mechanistic insights into role of *T* in the development of limb mesoderm. A Overall differential abundance at the cell type level for *T*$^{-/-}$ chimera knockout cells with perturbSuite_DA_cell_type. The x-axis displays the log10(odds ratio) and the y-axis the -log10 FDR corrected p-value. Red indicates cell types that are depleted amongst the mutant cells, blue indicates cell types enriched amongst the mutant cells, with grey indicating no significant changes. B) Per-stage differential abundance at cell type level. Significance at an FDR-corrected (across both stages and cell types) p-value less than 0.1 is indicated by a dot. Significance depends on total cell numbers per mapped stage and mapped cell type across both the WT and *T*$^{-/-}$ chimeras. C) Differential abundance of WOT trajectories leading to the listed cell types at E9.25, for *T*$^{-/-}$ tomato knockout cells with perturb_DA_lineage. Axes and colours are

as in A. D) Pseudotime distributions of $T^{-/-}$ and WT tdTom^{+} (red) and tdTom^{-} (blue) cells. The left column displays the T mutant chimera and the right column indicates the WT chimera. E) Confidence interval for location parameter of Wilcoxon test for significance of delay along lineage for knockout cells with perturbSuite_kinetics. Confidence intervals not overlapping 0 show significance of delay compared to the internal control of the tdTom^{-} cells of the $T^{-/-}$ chimeras (purple bars). Confidence intervals for $T^{-/-}$ chimeras (purple bars) not overlapping with those for WT chimeras (black bars) mean significance relative to the external control of the WT chimeras. F) *T* is expressed only during the early stages of the limb mesoderm trajectory. Each dot represents a cell along the trajectory. The blue line shows smoothed expression (Loess curve). G) perturbSuite_DE_static applied to chimera lateral plate mesoderm cells along the limb mesoderm lineage identifies genes that are differentially expressed both within one chimera data set ($T^{-/-}$ chimera tdTom+ versus $T^{-/-}$ tdTom^{-}) and for which this effect is larger or in a different direction than in the WT chimera. H) Combining perturbSuite and MAYA reveals pathway deregulation for $T^{-/-}$ knockout cells at the lateral plate mesoderm stage of the limb mesoderm trajectory.

The cell type level analysis of $T^{-/-}$ chimeras with perturbSuite_DA_cell_type confirmed previous results(Guibentif et al. 2021), with depletion of Somitic Mesoderm, Presomitic Mesoderm, Intermediate Mesoderm and Notochord in the mutant cells and enrichment for NMPs. In addition, we observe a broad depletion effect on other mesodermally-derived lineages including endothelial cell types, namely Embryo proper, Venous and Allantois Endothelium, as well as Mesenchyme, Yolk sac mesothelium, and Allantois (Fig. 2A). We also observe enrichment of the early cell type Caudal Epiblast, suggesting a developmental delay of mutant cells. Additionally some cell types, such as Blood Progenitors and Cranial Mesoderm, show significant changes in relative abundance across different time points, (Fig. 2B). Additionally, we noted that accounting for the bias illustrated in Fig. 1E is important for avoiding false conclusions about cell types being enriched for knockout cells (Supp. Fig. 2A). Results from analysis using perturbSuite_DA_lineage are concordant with the cell type level results, with $T^{-/-}$ cells significantly depleted in trajectories towards Notochord, Somitic Mesoderm, Presomitic Mesoderm, Allantois and Mesenchyme (Fig.2C). perturbSuite_kinetics reveals a delay for both Mesenchyme and Allantois lineage development as a result of *T*

knockout (Fig. 2DE). Interestingly, 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 as well as a delay in its lineage development (Fig. 2DE). A role for *T* in cancers derived from mesenchymal tissues has been identified previously, where its overexpression is associated with changes in the epithelial-mesenchymal transition(Du et al. 2014; Hamilton et al. 2017; Huang et al. 2013). As previously noted, *T* knockdown severely disrupts the formation of posterior somitic tissues, see Figure 2A-C. Consistently, allantois cells, which arise from a posterior mesodermal precursor population, are also significantly depleted in *T* chimeras and tdTom^{+} *T* mutant cells are delayed in their progression along the inferred allantois trajectory (Figure 2DE). These results align with well known allantois defects that have been documented for *T* knockout mice and chimeric embryos(Gluecksohn-Schoenheimer 1944), giving confidence in the performance of PerturbSuite.

Finally, from the reference atlas we noted that *T* is expressed at early stages of the limb mesoderm trajectory (Fig. 2F), i.e. the lateral plate mesoderm cell type. Interestingly, perturbSuite_DE revealed new insights into limb development, as it uncovers a significant reduction of $T^{-/-}$ lateral plate mesodermal cells that normally form limb mesoderm. These findings are in line with a previous study that showed impaired forelimb bud formation in *T* mutant embryos(Liu et al. 2003). The underlying *T*-dependent mechanisms that disrupt forelimb formation, however, remain unclear(Papaioannou 2014) Using PerturbSuite, we identify that *Vim* and *Aldh2* are strongly downregulated (Fig. 2G), both compared to the tdTom^{-} host cells in the same embryos as well as the WT control experiment.

The downregulation of *Vim* suggests that *T* regulates epithelial-to-mesenchymal transition (EMT) along the limb mesoderm trajectory(Gros and Tabin 2014), which is supported by pathway analysis using MAYA(Landais and Vallot 2023) and perturbSuite_DE_static (Fig. 2h).

Aldh2 deficiency has been shown to increase oxidative stress(Szöcs et al. 2007; Wenzel et al. 2008; Choi, Tostes, and Webb 2011). Three pathways linked to this are also significantly downregulated: peroxisome, reactive oxygen species (ROS) and fatty acid metabolism.

A further interesting observation is the upregulation of Apela in $T^{-/-}$ cells (Fig. 2g), which activates PI3K/AKT/mTORC1 signalling(Ho et al. 2015) (Fig 2h), with loss of *Apela* being shown to cause defects in early mesodermal derivatives(Freyer et al. 2017).

A characterisation of *Mixl1* chimeras provides insights into epicardium development

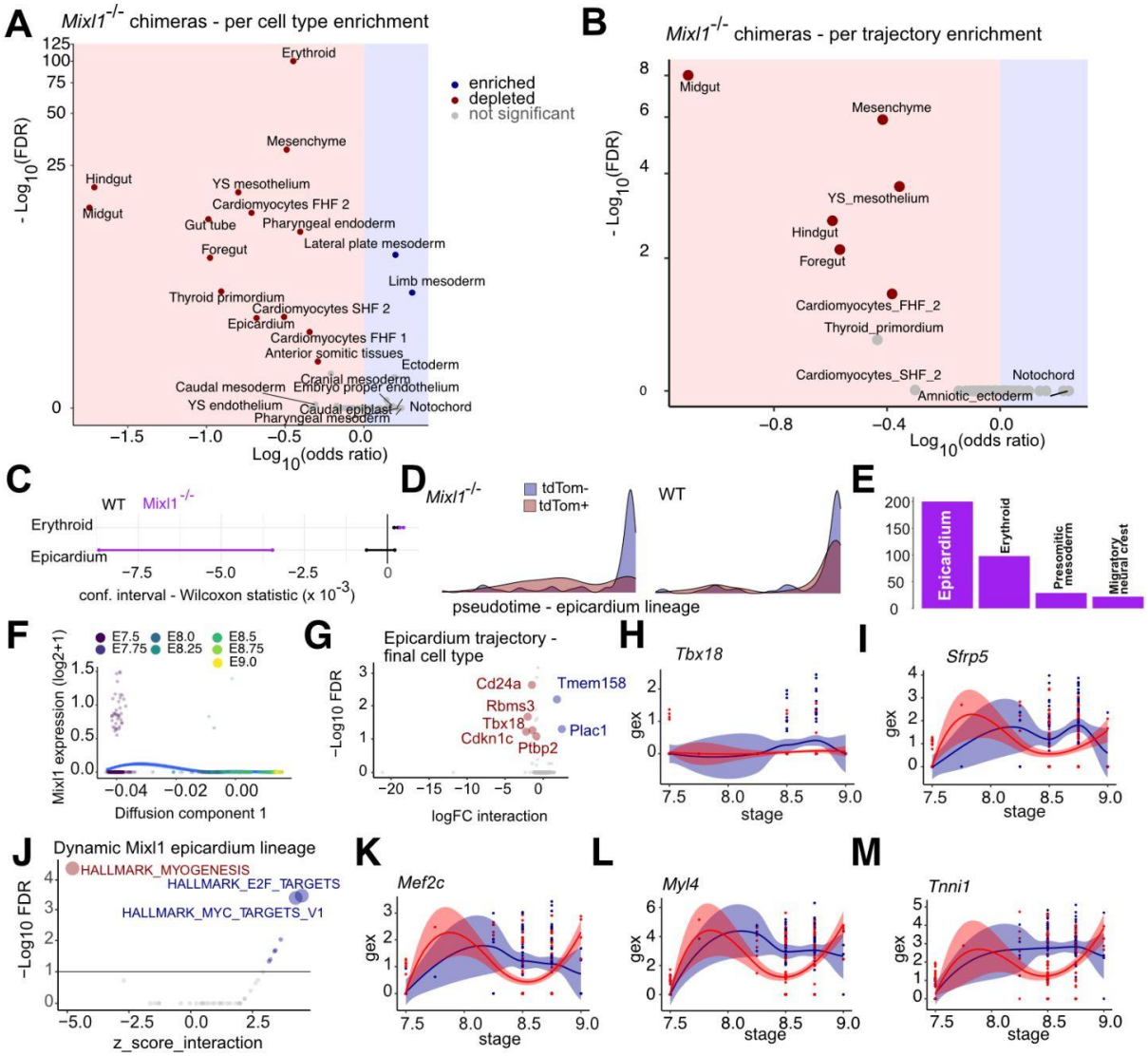


Fig. 3: Comprehensive analysis of Mixl1$^{-/-}$ chimeras using perturbSuite. A) Overall differential abundance at the cell type level for *Mixl1*$^{-/-}$ chimera knockout cells with perturbSuite_DA_cell_type. The x-axis displays the log10(odds ratio) and the y-axis the -log10 FDR corrected p-value. Red indicates cell types that are depleted amongst the mutant cells, blue indicates cell types enriched amongst the mutant cells, with grey indicating no significant changes. B) Differential abundance of WOT trajectories leading to the listed cell types at E9.25, for *T*$^{-/-}$ tomato knockout cells with perturb_DA_lineage. Axes and colours are as in A. C) Confidence interval for location parameter of Wilcoxon test for significance of delay along lineage for knockout cells with perturbSuite_kinetics. Confidence intervals not overlapping 0 show significance of delay compared to the internal control of the tdTom^{-} cells of the *T*$^{-/-}$ chimeras (purple bars). Confidence intervals for *T*$^{-/-}$ chimeras (purple bars) not overlapping with those for WT chimeras (black bars) mean significance relative to the external control of the WT chimeras. D) Comparing the distributions of *Mixl1*$^{-/-}$ and *WT*$^{-}$ tdTom^{+} and tdTom^{-} cells along pseudotime for the epicardium trajectory (left Mixl1 chimera; right WT chimera). E) Number of genes identified as dynamically DE with PerturbSuite_DE_dynamic, for the four lineages with the most such DE genes. F) Expression of *Mixl1* for epicardium lineage. The line is a curve showing smoothed expression (Loess curve).. G) PerturbSuite_DE_static for epicardium state. H) *Tbx18* expression in *Mixl1*$^{-/-}$ chimeras across mapped stages (tdTom^{+} cells in red, tdTom^{-} in blue). I) *Sfrp5* expression. The shaded areas are 95% confidence levels for the smoothed expression (curve). J) PerturbSuite_DE_dynamic at the pathway level identifies pathways for which *Mixl1* knockout leads to an increase (blue) or decrease (red) in activation along the progression of the lineage. K) *Mef2c,* L) *Myl4* and M) *Tnni1* expression.

To systematically assess the embryo-wide effects of *Mixl1* knockout we generated *Mixl1*$^{-/-}$ chimeras and processed the resulting scRNAseq data using perturbSuite. *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 homeobox domain. Two independent *Mixl1*$^{-/-}$ tdTom+ clones were injected into wildtype blastocysts to generate embryonic chimeras. Three independent pools of E8.5 *Mixl1*$^{-/-}$ chimeras were singularized and sorted by flow cytometry based on the tdTom reporter and analysed by scRNAseq (see Methods).

Using DA_cell_type and DA_lineage in *Mixl1*$^{-/-}$ chimeras we observed a broad effect on lateral plate mesoderm 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 lateral plate and limb mesoderm (Fig. 3A, Supp. Fig. 3A). We also observe depletion of splanchnic mesoderm (cardiac tissues) and enrichment of somatic mesoderm (limb, Fig 3A). Although *Mixl1*$^{-/-}$ mutants have been previously shown to have defective cardiac development(Pulina et al. 2014), the molecular role of *Mixl1* for cardiac cell type induction has not been determined.

Mixl1 knockout mice do not form a hindgut(Hart et al. 2002). Consistently, in the *Mixl1*$^{-/-}$ chimeras,, Hindgut and Midgut are fully depleted at cell type level, while Foregut, Gut tube, Pharyngeal endoderm and Thyroid primordium are all partially, but substantially, depleted (Fig. 3A). Note that perturbSuite_DA_lineage was not used for Gut tube and Pharyngeal endoderm, as they are below our set frequency threshold at the final time point of E9.25 (Methods, Supp. Fig. 3E,5B). All of the other cell types mentioned above were also depleted at the lineage level (Fig. 3B), in line with a general concordance between cell type differential abundance and changes in contribution to corresponding lineage trajectories (Supp Fig 2D).

Next, we explored whether there were any alterations in the dynamics of progression along trajectories using perturbSuite_kinetics. Of the depleted cell types with sufficient numbers of cells mapping to the lineage (Supp. Fig. 3E, Supp. Fig. 5B), only Epicardium and Erythroid show significantly altered dynamics. We observe a large delay along the Epicardium lineage in *Mixl1*$^{-/-}$ chimeras (Fig. 3CD) in line with the depletion of this cell type, and a very small acceleration towards later erythroid stages for those *Mixl1*$^{-/-}$ cells that are able to enter the erythroid lineage (Fig. 3C), while the contribution to erythroid cell type states was reduced (Fig. 3A).

The delay in the epicardium lineage (Fig. 3CD) and large number of dynamically DE genes (perturbSuite_DE_dynamic, Fig. 3E, Supplemental table 1) suggests an important role for

Mixl1 in epicardium development prompting further analysis. AApplying perturbSuite_DE_static revealed significant differences for *Mixl1*$^{-/-}$ cells both relative to tdTom^{-} host cells in the same embryos and to the control experiment (Fig. 3G, Fig. 1F), even though *Mixl1* is lowly expressed in the reference atlas and only at early stages of the epicardium trajectory (Fig. 3F). Importantly, we observed downregulation of *Tbx18*, which is a known regulator of the normal structural and functional development of the epicardium(S.-P. Wu et al. 2013). Our results suggest that expression of *Mixl1* in lateral plate mesoderm is required for downstream development of epicardium and *Tbx18* upregulation (Supp. Fig. 3B).

Applying perturbSuite_DE_dynmaic revealed 199 significantly dynamically DE genes for the epicardium lineage (Fig. 3E Supp. Fig. 3C). In particular, we found significant differential dynamics for *Sfrp5* (Fig. 3I), which is expressed in a set of cardiac tissues, including epicardium, where it has been suggested to modulate *Wnt* signalling (Fig. 3J)(Fujii et al. 2017).

To better understand the relevance and mechanisms of the dynamically DE genes, we combined perturbSuite_DE_dynamic with pathway analysis (Fig.3J), and identified genes that are both members of a de-regulated pathway and were also identified as dynamically differentially expressed. For the strongly downregulated Myogenesis pathway (Fig.3J), *Mef2c* (Fig. 3K), *Myl4* (Fig. 3L), and *Tnni1* (Fig. 3M) had high weight for the pathway score and were also dynamically downregulated on a per-gene basis (Fig 3J, Supp Fig. 3D). This suggests that in the absence of *Mixl1*, the lateral plate mesoderm common precursor population, which gives rise to epicardium and other cardiac lineages, has a decreased potential towards a cardiomyocyte fate, consistent with the observed depletion of the Cardiomyocyte cell types (Fig. 2A, Supp. Fig. 3A). Furthermore, the hallmark *E2f* and *Myc* target gene sets are most strongly upregulated (Fig.3J), indicating a persistent dividing progenitor state in mutant cells

(that lack the ability to differentiate), consistent with increased abundance of the lateral plate mesoderm cell type (Fig. 3A).

Using perturbSuite, 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. Both *T* and *Mixl1* have been reported as major mesoderm regulators(Guibentif et al. 2021; Hart et al. 2002; Chesley 1935; Beddington, Rashbass, and Wilson 1992). We found that in general, and in particular for a number of specific cell types such as Notochord, Node, Caudal Mesoderm and Caudal Epiblast, *Mixl1* expression is downregulated in $T^{-/-}$ chimeras (Supp. Fig. 4AB), albeit with small differences in absolute terms, given the generally low expression of *Mixl1* at E8.5. This would suggest a role of *Mixl1* downstream of *T* in the development of these cell types. However, we find that these two chimeras generally have non-overlapping phenotypes (Supp Fig 4C), and that they also differ largely in terms of the lineages displaying dynamically DE genes (Supp. Fig. 4D).

Interestingly, Lateral Plate Mesoderm (LPM) is depleted in $T^{-/-}$ chimeras and enriched in $Mixl1^{-/-}$ chimeras (Supp. Fig4C). Moreover, both chimeras display non-overlapping defects in LPM derivatives, with a depletion of endothelial cells in $T^{-/-}$ chimeras and defective development of cardiac tissues in $Mixl1^{-/-}$ chimeras.This suggests independent roles of the two transcription factors in the development of populations derived from lateral plate mesoderm. Another possibility is that the population annotated as lateral plate mesoderm in the atlas may comprise heterogenous progenitors with restricted fates, where Mixl1 and T regulate distinct gene networks.

Dynamic trajectory analysis in AML

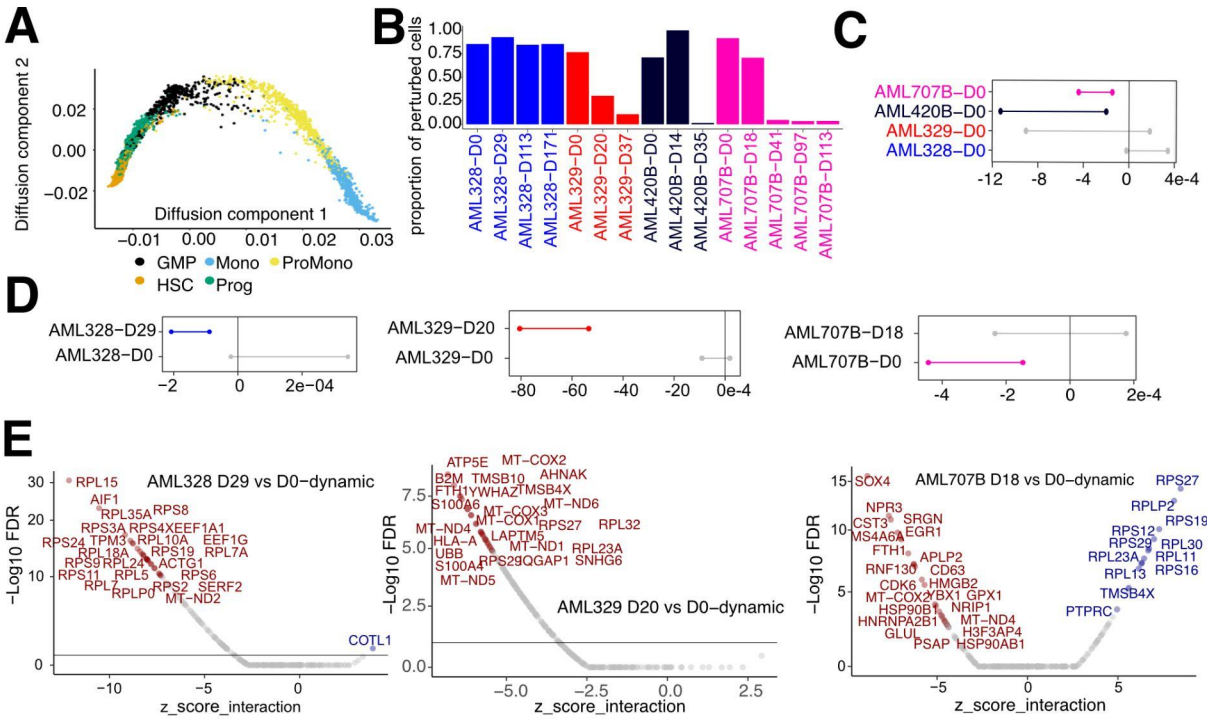


Figure 4: Insights into progression of AML. A) Diffusion components for myeloid trajectory in reference data set (bone marrow samples from healthy donors), coloured by cell type, as annotated in van Galen et al.(van Galen et al. 2019) Each dot is a cell. B) Proportion of cells along the myeloid trajectory that are malignant. We plot this for four patients (indicated in different colours) for whom the proportion of malignant cells at D0 is less than 95%. The threshold of 95% is needed so that there are both normal and malignant cells for comparisons.. C) Wilcoxon test comparing pseudotime along the myeloid trajectory between malignant and normal cells for the different donors shows a significant delay for patients AML-707B and AML-420B (perturbSuite_kinetics). The plot shows the confidence intervals for location parameter of Wilcoxon test for significance of delay along myeloid trajectory for malignant cells compared to normal cells. Confidence intervals below and not including zero mean that there is a significant delay along the myeloid trajectory for the malignant compared to the normal cells. D) PerturbSuite_kinetics identifies changes in the delay or acceleration along the trajectory between malignant and normal cells for D0 versus the first time point after diagnosis. AML-420B was excluded from this analysis, as less than 5% of the cells were normal at the first time point after diagnosis. E) Differences in dynamic gene expression changes for the first time point after diagnosis

compared to D0, restricted to myeloid-trajectory genes (genes are associated with development along the myeloid trajectory, but not with differences between donors).

In addition to the study of the developing embryo, single-cell analysis and transcriptional trajectories may also provide insight into the molecular mechanisms that underpin other processes, such as cancer progression. To illustrate the potential of PerturbSuite in this context we applied it to data generated as part of van Galen et al.'s single-cell RNA-seq study of acute myeloid leukemia (AML)(van Galen et al. 2019). In this study, the authors used single-cell genotyping along with a machine learning classifier to discriminate mutated from WT cells within individual tumour samples.

We used perturbSuite to gain insight into changes in myeloid development of malignant cells as a result of treatment. The correspondence between this dataset and the chimera data sets discussed in previous sections is that the patient data from day x after diagnosis corresponds to the targeted chimeras (*T*$^{-/-}$ or *Mixl1*$^{-/-}$ chimeras), while the data from the day of diagnosis corresponds to the WT chimeras. Patient cells classified as malignant correspond to the tdTom^{+} cells, while patient cells classified as WT correspond to tdTom^{-} (Supp. Fig. 6).

To implement perturbSuite, we used the samples from healthy bone-marrow as a reference in lieu of the extended mouse gastrulation atlas for the chimera data sets. We identified genes that are associated with development along the myeloid trajectory, but not with differences between donors (myeloid-trajectory genes). To examine progression, we took advantage of the annotated cell types of this dataset, which correspond to particular stages along the myeloid trajectory. Then we calculated diffusion maps and computed pseudotime, where annotated cell types follow the expected progression along component 1, from the hematopoietic stem cell (HSC) to the mature cell type (monocyte) (Fig. 4A).

Next, we looked at how patient samples differ from each other at the time of diagnosis and in terms of their changes after diagnosis. For this, we focused on patient samples that include multiple time points and for which at D0 (diagnosis) less than 95% of the cells along the myeloid trajectory were classified as malignant (Fig. 4B). This ensures a sufficient cell number for both malignant and normal cells at D0 and after treatment, allowing enough power for comparison. We applied the Wilcoxon test to the patient samples to test delay along the myeloid trajectory for malignant cells compared to their WT counterparts. At D0, AML420B and AML707B both show a significant delay, which means that malignant cells are behind normal cells for the myeloid trajectory and are more stem cell-like (Fig. 4C).

While the original publication showed that the presence of a specific cell type composition in progenitors is strongly related to worse prognosis (Fig. 4ef in van Galen et al.(van Galen et al. 2019)), perturbSuite allows us to study the differences between the location on the myeloid trajectory between malignant and normal cells of the same patient and also to follow the composition over time and test for significance of changes in the location of cells along the myeloid trajectory. To this end, we used perturbSuite_kinetics to assess delay along the myeloid trajectory of the first time-point sampled after day 0. We examined the 3 patients from Fig. 4B for whom the percentage of malignant cells in the myeloid trajectory is between 5% and 95% at the first time point after D0 (Fig. 4D).

Interestingly, for AML328 where there is no induction therapy after diagnosis, the malignant cells are now delayed compared to the normal ones, while there was no delay at D0. For AML707B and AML329, where induction therapy was applied after D0, we observe different

patterns. In AML707B the initial delay is no longer present at the next time point, while in AML329 the malignant cells show a further delay following induction therapy. These differences can be attributed to the diverse mutational landscape of these patients(van Galen et al. 2019), different disease stages as well as heterogeneity of response to therapy. As shown in Fig. 4B, the proportion of malignant cells does not decrease for AML328 over the course of time, while it does for AML329 and AML707B.

We then applied perturbSuite_DE to examine dynamic gene expression effects (Fig 1f) between D0 and the first time point after diagnosis, based on the myeloid-trajectory genes identified above. In both AML328 and AML329, the differentiation delay of mutant cells in the first time-point sampled after day 0 is accompanied by a decrease, along the differentiation, of ribosomal protein gene expression compared to normal cells in the same patient (Fig. 4E, Supplemental table 2). In fact, for AML328, 41 of the 85 significantly down-regulated myeloid-trajectory genes are part of the KEGG ribosome pathway, while only 22 out of the 179 not significantly downregulated myeloid-trajectory genes overall are part of the pathway (p=8e-10, Fisher's exact test). For AML329, 54 of the 140 down-regulated genes are part of the ribosome pathway, compared to 9 out of the 124 not downregulated genes(p=10e-10, Fisher's exact test). Interestingly we observe the opposite effect in AML 707B, where the initial delay is no longer present after induction therapy. Here ribosomal protein gene expression increases along differentiation in mutant cells (12 out of 16 upregulated genes are ribosomal, 51 of 248 not upregulated genes, p=1e-5). Of note, these cells also show dysregulation of genes previously implicated in AML development, namely *SOX4(H. Zhang et al. 2013)*, *PTPRC*/CD45(Saint-Paul et al. 2016) and *EGR1.(Le Beau et al. 1993)*

Discussion

Single-cell readouts for perturbation experiments in complex systems allow insights into molecular processes across a wide range of cell types and lineages. Here we studied the role of the transcription factors *Mixl1* and *T/Brachyury* across all lineages in the developing mouse embryo, by studying scRNA-seq data generated from chimeras. In addition, we developed a set of computational tools, perturbSuite, to tackle the particular challenges of complex perturbation data. This allowed us to identify novel affected lineages, such as limb mesoderm for *T* and *Mixl1*, and epicardium for *Mixl1*, thereby gaining new insights into the role of those transcription factors. We also demonstrated the wider relevance of the proposed analysis methods to general perturbation and disease scRNAseq data by applying it to a study of AML patient data(van Galen et al. 2019).

PerturbSuite is a method to infer perturbation effects in complex systems that is independent of the choice of method used to map the data to the reference data set. Here we used mutual nearest neighbours(Haghverdi et al. 2018) for consistency with our reference set(Imaz-Rosshandler et al. 2023). PerturbSuite is flexible enough to use any batch mapping(Hao et al. 2021; Kiselev, Yiu, and Hemberg 2018; Lotfollahi et al. 2022), cell lineage estimation(Lange et al. 2022), and pseudotime/cell ordering method. However, the identification of temporal genes and transfer of pseudotime to perturbation datasets from the reference data is an important and required part of peturbSuite_kinetics and perturbSuite_DE.

PerturbSuite 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. WOT) 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 similar groups (neighbourhoods) of cells between a query and a reference data set(Ton et al. 2023; Dann et al. 2022) . 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 normal development of a reference atlas, along a specific direction (lineage).

In a similar approach to perturbSuite_kinetics, Mittnenzweig et al.(Mittnenzweig et al. 2021) compared chimera cells in terms of the transcriptional rank of the individual embryo in their reference data set to which the chimera cell is most similar. PerturbSuite_kinetics and perturbSuite_DE_lineage may also be used with transcriptional rank instead of pseudotime, allowing for systematic testing of differences to both the host cells and the WT chimera experiments.

Implementation of perturbSuite on the $T^{-/-}$ chimera dataset validated our method by reproducing results from previous analyses(Guibentif et al. 2021), with strong effects of *T* depletion on the NMP population and somitic mesoderm. Using the extended mouse gastrulation atlas(Imaz-Rosshandler et al. 2023) as a reference provided more statistical power, and more effects reached statistical significance in agreement with the literature, for instance in the case of Notochord(Beddington, Rashbass, and Wilson 1992; Chesley 1935) and Allantois(Gluecksohn-Schoenheimer 1944). In addition, mesenchyme was depleted in these chimeras, and showed a delay in lineage development. *T* has been reported to promote EMT in tumour cells(Hamilton et al. 2017; Huang et al. 2013; Palena et al. 2011; Du et al. 2014), while *T* knockdown was shown to reduce EMT in NMPs(Dias et al. 2020), impacting the

development of posterior somitic tissues[5]. The extension of the reference dataset to E9.5 also allowed us to analyse the transcriptional trajectories towards lineages developing beyond E8.5. This was the case of limb tissues, where our results suggest that *T* also regulates the EMT pathway required for development of limb mesoderm from lateral plate mesoderm(Gros and Tabin 2014).

We then applied PerturbSuite to a novel chimera dataset targeting another key mesoderm regulator, *Mixl1*. In addition to an expected strong depletion of definitive endoderm tissues(Hart et al. 2002), we note a marked impact on cardiac lineages, with the Epicardium cell type showing most dysregulated gene expression. *Mixl1*-KO mice show severe cardiac malformations(Pulina et al. 2014), and epicardium plays an important role in coordinating the development of the other heart tissues, namely through modulation of the *Wnt* pathway(Fujii et al. 2017). Our results suggest this could be regulated by *Mixl1* through activation of *Sfrp5*, which is depleted in the epicardium lineage for *Mixl1*$^{-/-}$ chimera cells. *Mixl1* has been recently implicated in a lateral plate mesoderm inducing regulatory transcription factor network conserved in chordates(Prummel et al. 2019), but based on the studied enhancer region this did not extend to mammals. Our observations of depleted splanchnic mesoderm (cardiac tissues) and enrichment of somatic mesoderm (limb) suggests that *Mixl1* may still have a major role in balancing lineage propensity within lateral plate mesoderm precursors in the mouse. The comparison of both chimeras indicates that *T* and *Mixl1* may co-regulate early mesoderm formation; however, they seem to belong to independent regulatory networks when it comes to mesoderm downstream diversification.

In the adult, oncogenic mutations also lead to molecular perturbations in a minority population within a complex cellular system. When it comes to therapeutic intervention,

strategies need to consider the different responses in perturbed cancer cells vs the remaining healthy populations. In the present study we extend the application of perturbSuite to a published AML dataset(van Galen et al. 2019) and characterized individual patient samples in terms of differentiation blockage of leukemia cells and response to treatment. We observed that delayed differentiation was associated with a decreased expression of protein synthesis machinery along the myeloid trajectory. This is relevant since translation control is emerging as a potential therapeutic target for AML(Hein et al. 2017; Sjövall et al. 2023; J. Wu et al. 2020).

With the increasing number of large-scale perturbation experiments(Replogle et al. 2020) in complex systems in-vivo(Jin et al. 2020; Saunders et al. 2022) and in vitro(Fleck et al. 2022) as well as the emergence of single-cell analysis to patient material(Y. Zhang et al. 2021), principled analysis of large-scale complex perturbation data will allow the identification of lineages of interest, leading to new biological and clinical insights.

Methods

Embryo Chimera data

$T^{-/-}$ and wildtype chimera data

Count data for $T^{-/-}$ and wildtype chimeras was obtained from a previous study(Guibentif et al. 2021), via the *MouseGastrulationData* Bioconductor package(Jonathan Griffiths 2021).

Reference data

Processed data including dimensionality reduction and WOT scores for the extended mouse gastrulation data was obtained from the authors of that study(Imaz-Rosshandler et al.

2023). Processed data for the earlier mouse gastrulation atlas was obtained via the Bioconductor package(Jonathan Griffiths 2021).

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

All procedures involving mouse embryos were performed in strict accordance to 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(Pijuan-Sala et al. 2019). 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(Ying et al. 2008) and transiently transfected with a Cre-IRES-GFP plasmid(Wray et al. 2011) 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(Ran et al. 2013). 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. Chimaeric 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). This resulted in 17393 $tdTom^-$ and 18754 $tdTom^+$ cells that passed quality control (see "Data

processing, QC and mapping to extended mouse gastrulation atlas" below). To exclude transcriptional effects intrinsic to the chimera assay, chimaeric embryos were generated by injecting the parental tdTom^{+} *Mixl1*$^{+/+}$ (WT) line into C57BL/6 E3.5 blastocysts and processed as for the *Mixl1*$^{-/-}$ samples.

AML data

Annotated AML data(van Galen et al. 2019) was obtained through GEO: GSE116256.

Computational methods

PerturbSuite provides a comprehensive picture of knockout effects in complex organisms by studying them at four levels (Fig. 1D). First, it investigates whether knockout causes differential abundance at the level of cell types, compared to changes seen between tdTom^{+} and tdTom^{-} cells from the WT chimeras (perturbSuite_DA_cell_type, Fig. 1F, 2A, 3A). Second, for cell types represented at E9.25, we identify the cells comprising the inferred developmental trajectories ending in the respective cell types (henceforth called lineage), and our method determines whether a knockout will lead to abundance or depletion of cells for the lineage, again compared to changes in the WT chimeras. (perturbSuite_DA_lineage, Fig2C, 3B). Third, for lineages that are not severely depleted at the lineage level, our suite of methods also determines whether the progression along the lineage is delayed or accelerated (perturbSuite_kinetics)Fig. 1G,Fig 2DE, 3CD, 4CD). Finally, we determine for which lineages knockout cells are affected by differences in gene expression (perturbSuite_DE, Fig 1H, 2FG, 3E-M, 4E, Supp. Fig. 3C, Supp. Fig. 4D), and identify trajectory-specific affected genes featuring putative molecular mechanisms downstream of the knockout.

Data processing, QC and mapping to extended mouse gastrulation atlas

For the *Mixl1*$^{-/-}$ chimeras, raw data were processed with CellRanger without mapping intronic reads, with reads mapped to the mm10 genome and counted with GRCm38.92 annotation, including tdTomato sequence, to ensure consistency with the reference atlas(Imaz-Rosshandler et al. 2023) and the existing *T*$^{-/-}$ chimera data set(Guibentif et al. 2021). For the *T*$^{-/-}$ chimera we used processed reads(Guibentif et al. 2021), for the Mixl1$^{-/-}$ we followed the quality control steps detailed in Guibentif et al. and https://github.com/MarioniLab/EmbryoTimecourse2018.

Consistently with batch-correction that had been used to integrate samples for the reference data (extended mouse gastrulation atlas(Imaz-Rosshandler et al. 2023)), chimera data were mapped to the extended mouse gastrulation atlas using the mutual nearest neighbours(Haghverdi et al. 2018), with the underlying PCA space computed based on the reference data only. Before this, the extended mouse gastrulation atlas was subset to 10,000 cells per time point, as in (Guibentif et al. 2021), to avoid biases in the mapping arising from different numbers of cells at different time points. The mapped cell type of a cell from a chimera sample is the most frequent cell type among its 10 nearest neighbours in the extended mouse gastrulation atlas. UMAP(McInnes et al. 2018) coordinates for the low-dimensional representation for visualisation, WOT scores(Schiebinger et al. 2019) for the extended mouse gastrulation atlas were taken from (Imaz-Rosshandler et al. 2023). Each chimeric cell was assigned the WOT scores of its nearest neighbour among the atlas cells, i.e. each cell was assigned a WOT for each lineage.

perturbSuite_DA_cell_type: Cell type-based differential abundance testing

Cell type-based differential abundance testing determines whether there is significant depletion or enrichment for $tdTom^+$ cells compared to $tdTom^-$ in the knockout chimeras, compared to the difference between $tdTom^+$ and $tdTom^-$ in the WT chimeras (Fig. 1F). PerturbSuite_DA accounts for biases in differential abundance testing using ideas similar to TMM normalisation for differential expression testing for bulk RNA-seq data(Robinson and Oshlack 2010).

Cell type-based differential abundance testing was performed as follows:

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

$r_{c,target}$ = number of $tdTom^+$ cells in knockout chimeras mapping to the cell type divided by the number of $tdTom^-$ cells in knockout chimera mapping to the cell type

$r_{c,control}$ = number of $tdTom^+$ cells in WT chimeras mapping to the cell type divided by the number of $tdTom^-$ cells in WT injected chimera mapping to the cell type

Then we performed normalisation to account for the sampling biases (Fig. 1E) as follows:

First we computed m_{target} = median($r_{c,target}$) and $m_{control}$ = median($r_{c,control}$).

From the knockout chimera we sampled (without replacement) n_{target} cells among the tdTom+ cells and n_{target} x m_{target} cells among the tdTom- cells, where n_{target} = min(number of $tdTom^+$ cells in knockout chimera mapping to the cell type, number of $tdTom^-$ cells in knockout chimera mapping to the cell type/m_{target}). Under the assumption that most cell types are not affected by depletion or enrichment in consequence of the knockout (and that therefore the median ration m_{target} is not affected), this normalisation accounts for the fact that given the

depletion of larger cell types, an approximately equal number of tdTom$^+$ and tdTom$^-$ cells overall will lead to spurious enrichment results for other cell types, if not corrected. Similarly, we sampled from the WT injected control chimeras.

After sampling, for each cell type we performed a Fisher's exact test to determine the odds ratio and p-value of a chimera cell mapping to a particular cell type being significantly more or significantly less likely to be tdTom$^+$ than tdTom$^-$. We repeated the sampling and the Fisher 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. FDR correction(Benjamini and Hochberg 1995) was subsequently applied to the p-values.

We applied perturbSuite_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*), in exactly the way described above with cell type+stage combinations replacing cell types.

perturbSuite_DA_lineage: lineage-based differential abundance testing

This is a summary test to detect whether tdTom$^+$ chimera cells have an increased or decreased predisposition to develop into a particular cell type at a later stage (Fig 1C).

The test is similar to perturbSuite_DA_cell_type, but needs to be adapted as we now use lineage scores, i.e. probabilities of a cell being part of a lineage, for each cell rather than cell types (definite assignments to a cell type).

We considered lineages whose final cell types present with more than 100 cells at E9.25 in the extended gastrulation atlas. The following was repeated 10 times:

> For each cell from the chimeric embryos a lineage was sampled according to the WOT scores. Given the sampled lineage for each cell (a fixed state and not a probability of

being in a state), we can now apply perturbSuite_DA_cell_type, performing the sampling of cells and Fisher tests were performed 30 times as for perturbSuite_DA_cell_type.

After repeating the above 10 times, we computed the medians across odds-ratios and FDR-corrected p-values across the 10 repeats.

perturbSuite_kinetics: Testing for acceleration or delay of knockout cells along a trajectory

To determine whether there is a delay or acceleration for tdTom$^+$ cells along a trajectory, we combined WOT scores with statistical modelling and pseudotime.

Cells were selected as being part of a lineage using a mixture model of skewed t-distributions[32] with two or three components (the optimal number was determined by the Bayesian information criterion) to identify for each mapped stage which cells belong to the lineage. First we preselected those cells that have higher normalised WOT scores than random, i.e. higher than the inverse of the total number of cells at the stage. Then, we fitted the mixture model and selected from among the cells that passed the first threshold of having WOT scores above random those that were in the cluster of the highest WOT scores. For each stage we discarded cell types that did not constitute at least 10% of the cells for the lineage at that stage, to avoid mistakes from the mapping algorithm to influence the trajectory. From the time point onward from which the final cell type is the most frequent cell type in that trajectory we excluded other cell types from the trajectory.

After the assignment of cells to lineages, temporal genes were selected by testing their association with embryonic time. Then a linear regression model was used to obtain residuals for the association of gene expression with embryonic time for these temporal genes. Genes for which the residuals from that regression were associated with batch, i.e. genes correlated

with time to different degrees for different batches, were removed from the list of temporal genes. Then we used diffusion maps[30] to obtain a more fine-grained resolution of temporal development, identifying 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 (Supp. Fig. 1C). Starting from E7.5 allowed for enough power resulting from a sufficient number of mapping cells, as well as sensitivity to lineage delay.

This method was applied to all lineages 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 (tdTom$^+$ and tdTom$^-$) by assigning to the chimera cell the average of the pseudotimes of the 10 atlas cells the chimera cell is most correlated to, where for the computation of the correlation expression we only used genes that were identified as temporal genes in the previous step using atlas cells only. This ensured the pseudotime mapping was based on genes that differ across stages in the reference data set and excluded genes whose expression depends on batch.

We used the Wilcoxon rank sum test (Mann and Whitney 1947) to identify trajectories whose median was more strongly shifted in tdTom$^+$ cells for the knockout chimeras than for the WT chimeras (Fig. 1G, Fig. 2DE, Fig. 3CD, Fig. 4CD). This identified trajectories for which the difference between tdTom$^+$ and tdTom$^-$ cells for the knockout chimera was signficant and the 95% confidence intervals for the location parameter (the difference in pseudotime

median of the distribution of the tdTom$^+$ and tdTom$^-$ cells) of the Wilcoxon test were non-overlapping between WT and knockout chimeras (Fig. 2E, 3C,Fig. 4CD).

perturbSuite_DE_static: Lineage-specific static differential expression test

The idea of testing for testing for significance of differences between tdTom$^+$ and tdTom$^-$ cells for each lineage was extended to changes not only in pseudotimes between the tdTom$^+$ and tdTom$^-$ cells (perturbSuite_kinetics), but to differences in gene expression between tdTom$^+$ and tdTom$^-$ cells specific to the knockout chimeras as opposed to the WT chimera experiment (Fig 1H). To investigate this, for each gene *j* identified as relevant to the development of the specific lineage (based on the atlas cells), the following mixed model was used:

$$t_i \sim \beta_j \cdot x_{ij} + \alpha_j + \gamma_j \cdot x_{ij} \mid k_i \quad (1)$$

where t_i=1 if cell *i* is tdTom$^+$ (t_i=0 if cell *i* is tdTom$^-$), x_{ij} is the expression level of gene *j* in cell *i* and and k_i is an indicator of whether the cell is from the knockout (*T* or *Mixl1*) chimera (k_i=1) or the WT chimera (k_i=0). The parameters β_j , α_j and γ_j are inferred. α_j is the intercept parameter, the level of gene expression independent of the tomato status of a cell or its membership in the knockout or the WT chimera. β_j corresponds to the added level of gene expression (increase if β_j > 0, decrease if β_j < 0) resulting from the cell being tdTom$^+$ rather than tdTom$^-$. Finally, γ_j is the value that needs to be added to β_j (negative or positive), if the cell is from the knockout rather than the WT chimera. Thus, testing for whether γ_j is significantly different from 0 informs about whether the tomato status (tdTom$^+$ versus tdTom$^-$) is better explained by a specific parameters for knockout ($\beta_j + \gamma_j$) and WT chimeras (β_j), rather than by one shared parameter. Thereby we were able to infer whether the relation

between the expression of gene *i* and the tomato status was different for the target compared to the WT chimeras.

To ensure that the effect was significant not only compared to the WT chimeras, put also per se for tdTom$^+$ versus tdTom$^-$ cells of the knockout chimera itself, the following model was used to test for $\beta_j \neq 0$, where *i* was now restricted to the knockout chimera cells, i.e. the model is only used for the knockout chimeras.

$$t_i \sim \beta_j \cdot x_{ij} + \alpha_j \quad (2)$$

In summary, if $\gamma_j \neq 0$ for model (1) and for $\beta_j \neq 0$ for model (2) (significantly), then the gene is differentially expressed for tdTom$^+$ compared to tdTom$^-$ cells for the knockout chimera (β_j , internal control), and that difference is significantly larger or smaller than the analogous difference for the WT chimera experiment (γ_j , external control).

To avoid confusing differential abundance of different cell types or lineages along the lineage for tdTom$^+$ versus tdTom$^-$ cells (and therefore spuriously detect differential expression of cell type or lineage markers between tdTom$^+$ and tdTom$^-$ cells), we restricted perturbSuiteDE_static to testing differential expression of one cell type alone one particular lineage (Fig. 2G, 3H). We also applied perturbSuite_DE_static to pathway scores inferred using MAYA(Landais and Vallot 2023) instead of genes (Fig. 2H).

3.2.3 perturbSuite_DE_dynamic: Lineage-specific dynamic differential expression test

Static differential expression was then extended to dynamic effects along pseudotime (Fig. 1H). The following model was used

$$t_i \sim \beta_j \cdot (x_{ij} \cdot d_i) + \alpha_j + \gamma_j \cdot (x_{ij} \cdot d_i) \mid k_i \quad (3)$$

where t_i=1 if cell *i* is tdTom$^+$ (t_i=0 if cell *i* is tdTom$^-$), x_{ij} is the expression level of gene *j* in cell *i*, d_i is the pseudotime of cell *i*, and and k_j is an indicator of whether the cell is from the knockout chimera (k_i = 1) or the WT chimera (k_i = 0). The difference between models (1) and (3) is that model (3) test the change (increase or decrease) of gene expression along the lineage using the interaction between pseudotime and gene expression ($x_{ij} \cdot d_i$), rather than test for differences in static gene expression.

Similar to perturbSuite_DE_static, we also checked with perturbSuite_DE_dynamic that the effect is significant for the knockout chimeras per se, and not only compared to the WT.

$$t_i \sim \beta_j \cdot (x_{ij} \cdot d_i) + \alpha_j + \gamma_j \cdot (x_{ij} \cdot d_i) \mid k_i \quad (4)$$

In summary, if $\gamma_j \neq 0$ for model (3) and for $\beta_j \neq 0$ for model (4) (significantly), then the way that the gene changes its expression along the trajectory is different for tdTom$^+$ compared to tdTom$^-$ cells for the knockout chimera (β_j , internal control), and that difference is significantly larger or smaller than the analogous difference for the WT chimera experiment (γ_j , external control).

As perturbSuite_DE_dynamic compares how expression changes along the trajectory rather than, there are, unlike with perturbSuite_DE_static, no confounding effects with differential kinetics along the lineage or differential abundance of cell type along the lineage. We also combined perturbSuite_DE_dynamic with MAYA pathway scores (Fig. 3J).

Data and code availability

Raw sequencing data for the Mixl1 chimeras has been submitted to the Arrayexpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-13409.

Code is available on GitHub (https://github.com/MarioniLab/perturbSuite).

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 and Richard Tyser 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. Work in the Gottgens 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.

Authors' contributions

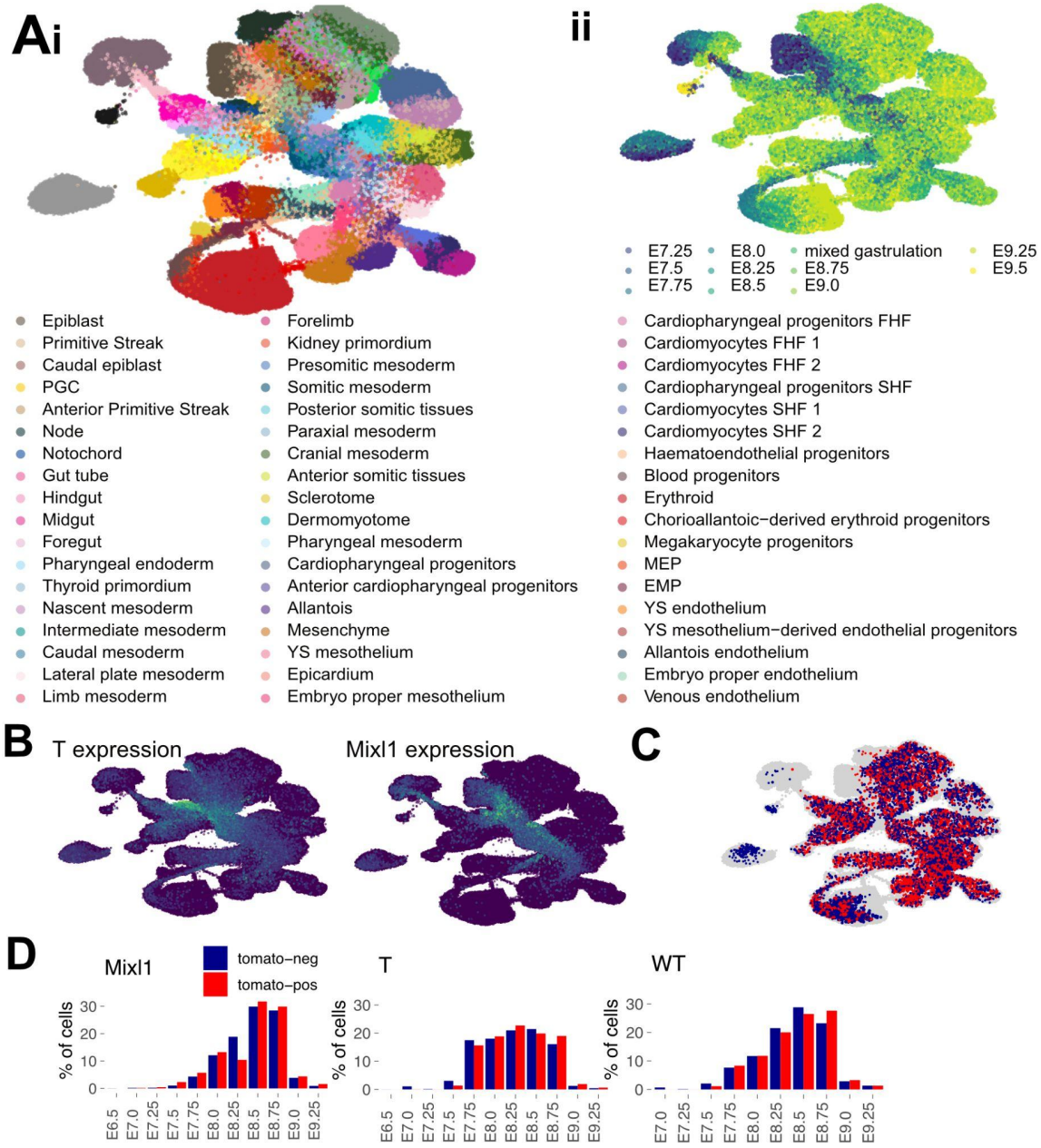
M.E.S. analysed the *Mixl1* embryonic chimera dataset, wrote all code, implemented perturbSuite and applied it to all datasets presented; M.-L.N.T. performed *Mixl1* chimeric embryo dissections and generated the scRNAseq dataset; C.G. generated the *Mixl1* knockout ESC lines; S.M. genotyped *Mixl1* ESC 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.H., I.I-R., M.-L.N.T. and B.G.; I.I-R. provided processed data and WOT analysis for the extended mouse gastrulation atlas; M.E.S. wrote the first draft of the

manuscript; M.E.S., C.G., L.H., M.-L.N.T., B.G., and J.C.M edited the manuscript; J.N., B.G., J.C.M. and C.G supervised the study.

Competing interests

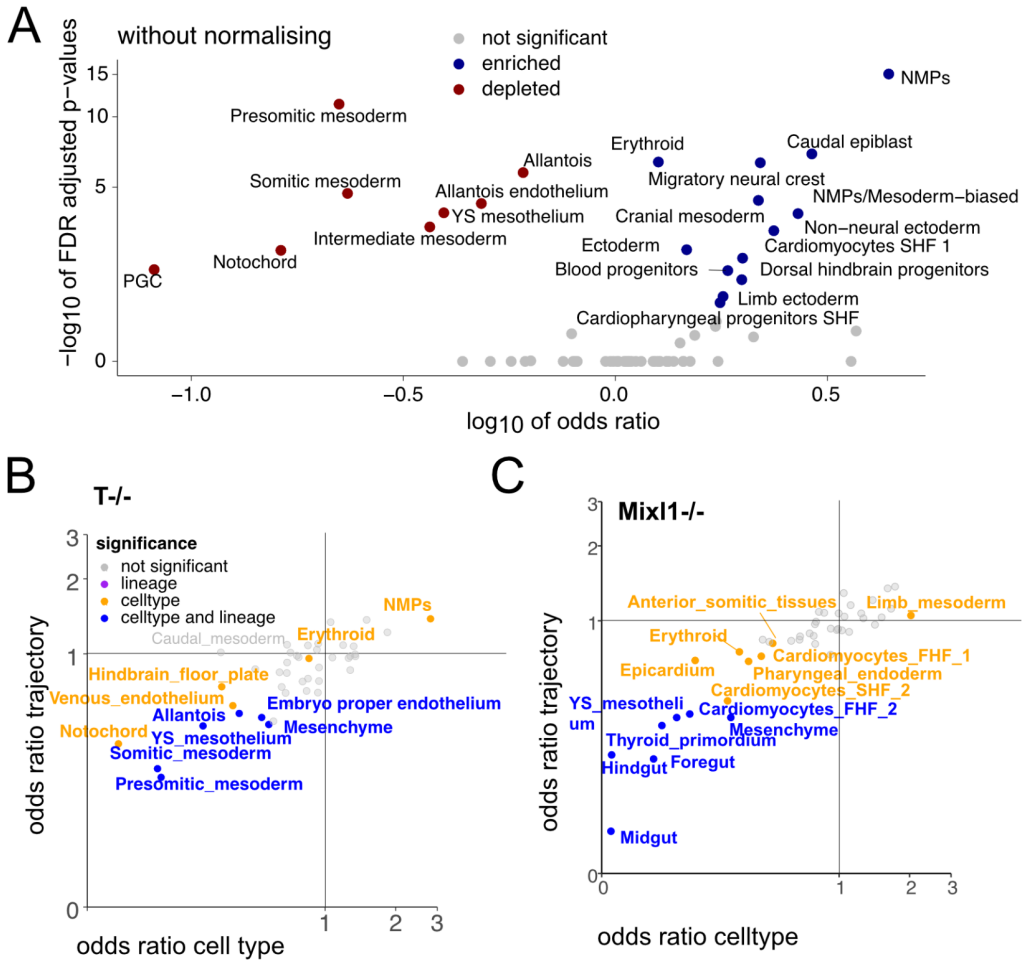
J.C.M. has been an employee of Genentech since September 2022. The remaining authors declare no competing interests.

Supplementary Figures

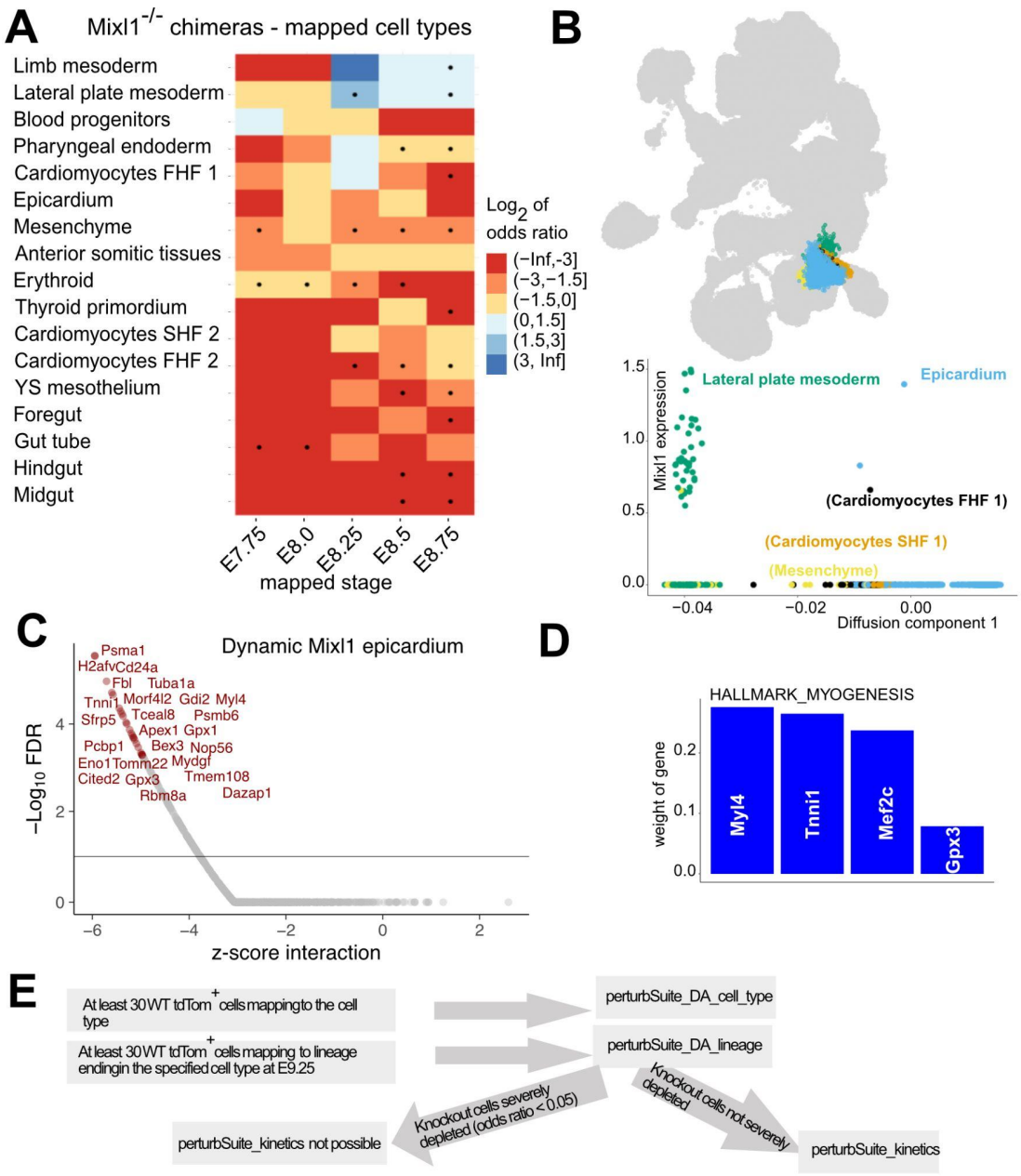


Supplementary Figure 1: Modelling based on the extended mouse gastrulation atlas. A) UMAP representation of extended mouse gastrulation atlas, coloured by cell types(Imaz-Rosshandler et al. 2023) (i) and stages (ii) B) Expression of *T* and *Mixl1* in the extended mouse gastrulation atlas. C) Mapping of the WT chimera to the extended mouse gastrulation data set shows differences between tdTom^{+} (red) and tdTom^{-} (blue) . D) Mapping of chimeras harvested at E8.5 to different stage in the extended mouse gastrulation atlas. The actual time of collection is E8.5, but some cells are delayed in development (earlier than E8.5), and others

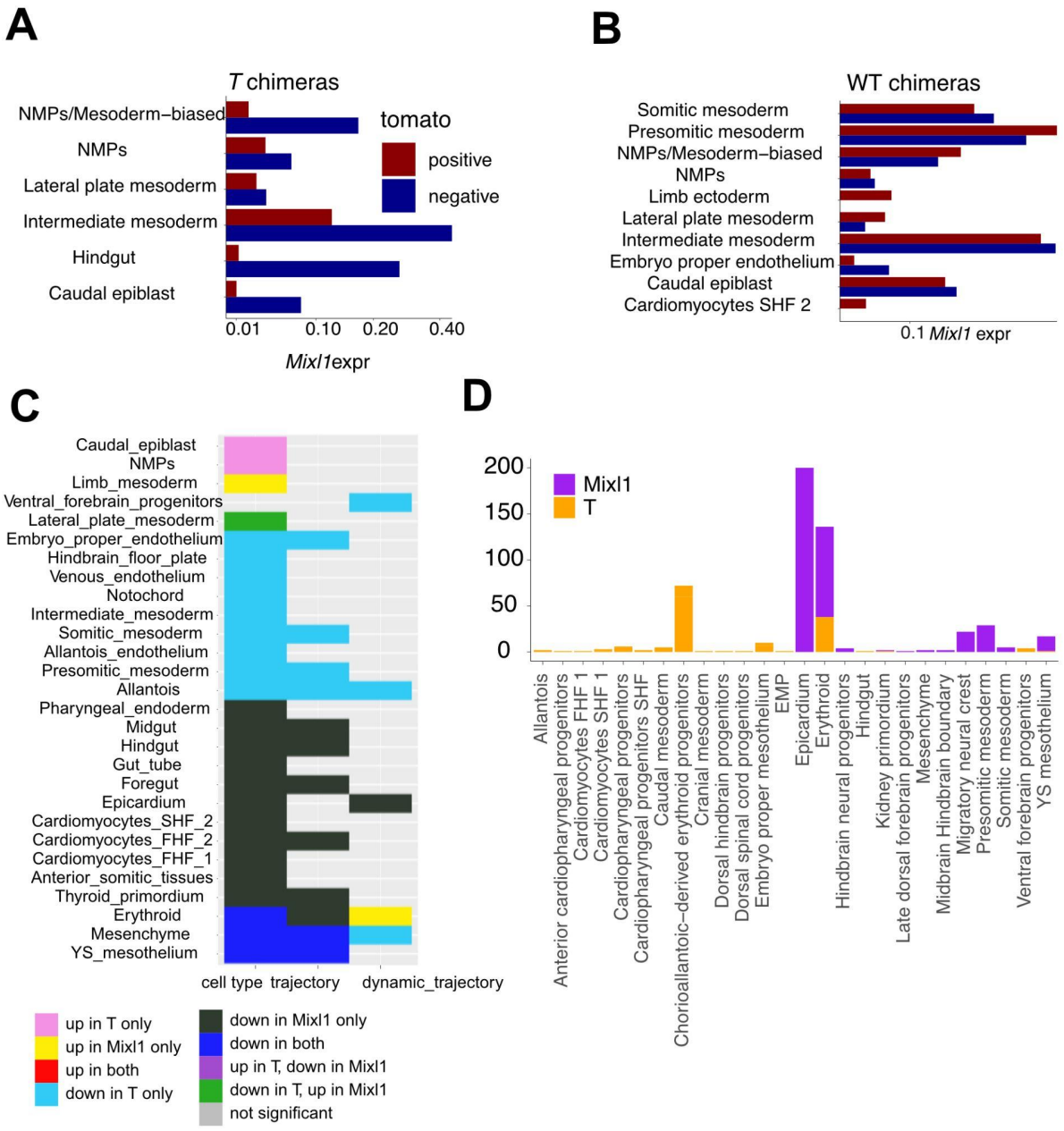
advanced.



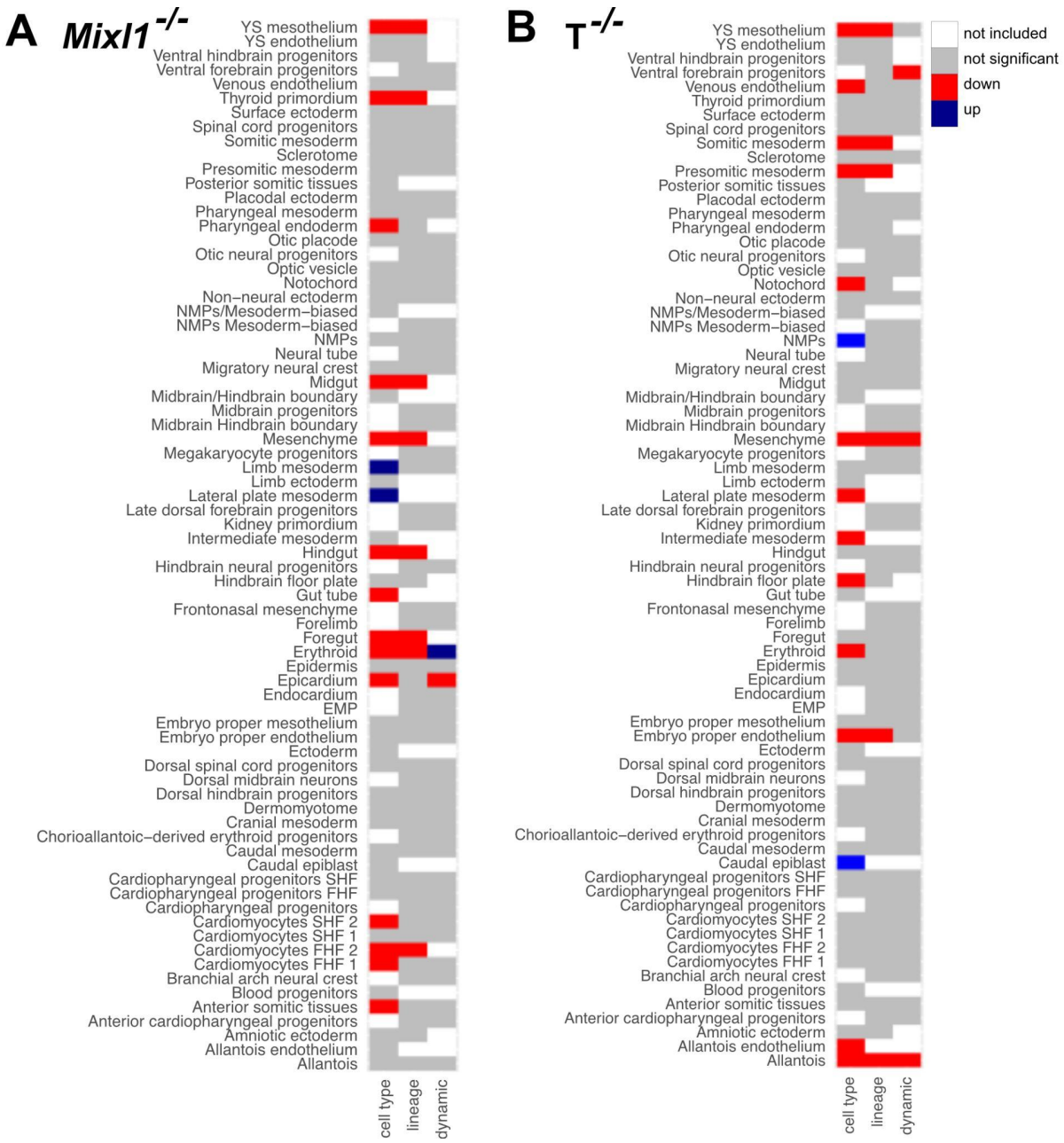
Supplementary Figure 2: **Supplementary figures for differential abundance analysis.** A) Differential abundance analysis without normalisation does not account for sampling bias and leads to more cell types showing up as enriched. B and C) Comparing lineage-based to cell type-based enrichment and depletion of $T^{-/-}$(B)/Mixl1$^{-/-}$(C) knockout cells. Colours indicate significance of cell type or lineage.



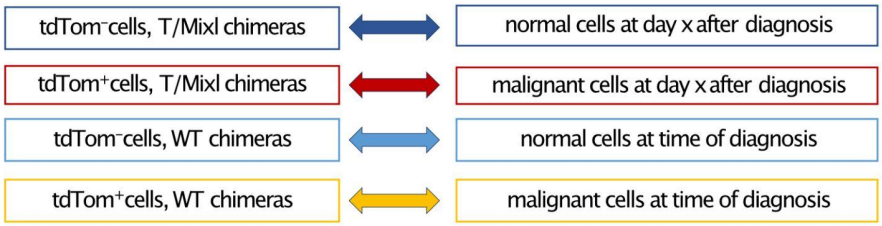
Supplementary Fig 3: Additional illustrations for differential expression analysis. A) Per mapped stage and cell type enrichment and depletion for *Mixl1*$^{-/-}$ chimeras. B) Top: Illustration of limb mesoderm lineage on the UMAP representation of the reference atlas. Cells that are not part of the lineage are in grey. There is a small number of cells labelled cardiomyocytes, which could result from a small number of cells with mislabelled cell types in the reference atlas or a small extent of lineage misassignments from the WOT scores. Bottom: Expression of *Mixl1* along the epicardium lineage from E7.5 in the extended mouse gastrulation atlas, coloured by cell type. C) Dynamic gene expression changes for *Mixl1*$^{-/-}$ tdTom+ cells for the epicardium lineage. The genes significant for both the internal tdTom^{+} versus tdTom^{-} comparison with the *Mixl1*$^{-/-}$ chimeras, as well as compared to the control experiment are marked in red. D) Weights found by MAYA for the Myogeneis hallmark set. E) Schematic of applicability of perturbSuite_DA_cell_type, perturbSuite_DA_lineage and perturbSuite_kinetics.



Supplementary Figure 4: Across-chimera comparison. A) *Mixl1* expression in $T^{-/-}$ chimeras, for cell types with average expression of at least 0.033 for either the tdTom^{+} or the tdTom^{-} cells (top) and at least 50 tdTom^{+} cells, and across different mapped time points (bottom). B) *Mixl1* expression in WT chimeras, with cut-offs for expression and cell numbers as in A. C) Comparing effects between $T^{-/-}$ and *Mixl1*$^{-/-}$, for differential abundance of cell type (left), differential abundance of lineage (middle), differential speed of progression along the trajectory. D) Number of genes identified as dynamically DE for each lineage for $T^{-/-}$ and *Mixl1*$^{-/-}$ chimeras, excluding lineages without any significantly dynamically DE genes.



Supplementary Figure 5: Expansion of Supp. Fig. 4c, including all cell types and lineages. Grey colour signifies that a test was performed, but there was no significant enrichment or depletion. White signifies that the test could not be done, because of small cell numbers in control chimeras to which the E8.5 chimeras map (left column-cell type), small numbers of cells of the cell type at the E9.25 state (middle column-lineage), or small numbers of cells of the cell type at the E9.25 or complete or near-complete depletion of the lineage for the knockout cells (right column-dynamic).



Supplementary Figure 6: Correspondence between chimera and AML data sets.
PerturbSuite_kinetics and perturbSuite_DE_dynamic were applied to the AML data set using internal and external controls similar to the chimera set-up. The figure illustrates which subset of the AML data corresponds to which part of the chimear data for this analysis.

References

Barile, Melania, Ivan Imaz-Rosshandler, Isabella Inzani, Shila Ghazanfar, Jennifer Nichols, John C. Marioni, Carolina Guibentif, and Berthold Göttgens. 2021. "Coordinated Changes in Gene Expression Kinetics Underlie Both Mouse and Human Erythroid Maturation." *Genome Biology* 22 (1): 197.

Beddington, R. S., P. Rashbass, and V. Wilson. 1992. "Brachyury--a Gene Affecting Mouse Gastrulation and Early Organogenesis." *Development . Supplement*, 157–65.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 57 (1): 289–300.

Chesley, Paul. 1935. "Development of the Short-Tailed Mutant in the House Mouse." *The Journal of Experimental Zoology* 70 (3): 429–59.

Choi, Hyehun, Rita C. Tostes, and R. Clinton Webb. 2011. "Mitochondrial Aldehyde Dehydrogenase Prevents ROS-Induced Vascular Contraction in Angiotensin-II Hypertensive Mice." *Journal of the American Society of Hypertension: JASH* 5 (3): 154–60.

Dann, Emma, Neil C. Henderson, Sarah A. Teichmann, Michael D. Morgan, and John C. Marioni. 2022. "Differential Abundance Testing on Single-Cell Data Using K-Nearest Neighbor Graphs." *Nature Biotechnology* 40 (2): 245–53.

Du, Rui, Shanshan Wu, Xiaoning Lv, Henghu Fang, Sudong Wu, and Jingbo Kang. 2014. "Overexpression of Brachyury Contributes to Tumor Metastasis by Inducing Epithelial-Mesenchymal Transition in Hepatocellular Carcinoma." *Journal of Experimental & Clinical Cancer Research: CR* 33 (1): 105.

Fleck, Jonas Simon, Sophie Martina Johanna Jansen, Damian Wollny, Fides Zenk, Makiko Seimiya, Akanksha Jain, Ryoko Okamoto, et al. 2022. "Inferring and Perturbing Cell Fate Regulomes in Human Brain Organoids." *Nature*, October. https://doi.org/10.1038/s41586-022-05279-8.

Freyer, Laina, Chih-Wei Hsu, Sonja Nowotschin, Andrea Pauli, Junji Ishida, Keiji Kuba, Akiyoshi Fukamizu, et al. 2017. "Loss of Apela Peptide in Mice Causes Low Penetrance Embryonic Lethality and Defects in Early Mesodermal Derivatives." *Cell Reports* 20 (9): 2116–30.

Fujii, Masayuki, Akane Sakaguchi, Ryo Kamata, Masataka Nagao, Yutaka Kikuchi, Silvia M. Evans, Masao Yoshizumi, Akihiko Shimono, Yumiko Saga, and Hiroki Kokubo. 2017. "Sfrp5 Identifies Murine Cardiac Progenitors for All Myocardial Structures except for the Right Ventricle." *Nature Communications* 8 (March): 14664.

Galen, Peter van, Volker Hovestadt, Marc H. Wadsworth Ii, Travis K. Hughes, Gabriel K. Griffin, Sofia Battaglia, Julia A. Verga, et al. 2019. "Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity." *Cell* 176 (6): 1265–81.e24.

Gluecksohn-Schoenheimer, S. 1944. "The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick." *Proceedings of the National Academy of Sciences of the United States of America* 30 (6): 134–40.

Goering, Lisa M., Kazuyuki Hoshijima, Barbara Hug, Brent Bisgrove, Andreas Kispert, and David Jonah Grunwald. 2003. “An Interacting Network of T-Box Genes Directs Gene Expression and Fate in the Zebrafish Mesoderm.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (16): 9410–15.

Gros, Jerome, and Clifford J. Tabin. 2014. “Vertebrate Limb Bud Formation Is Initiated by Localized Epithelial-to-Mesenchymal Transition.” *Science* 343 (6176): 1253–56.

Guibentif, Carolina, Jonathan A. Griffiths, Ivan Imaz-Rosshandler, Shila Ghazanfar, Jennifer Nichols, Valerie Wilson, Berthold Göttgens, and John C. Marioni. 2021. “Diverse Routes toward Early Somites in the Mouse Embryo.” *Developmental Cell* 56 (1): 141–53.e6.

Haghverdi, Laleh, Aaron T. L. Lun, Michael D. Morgan, and John C. Marioni. 2018. “Batch Effects in Single-Cell RNA-Sequencing Data Are Corrected by Matching Mutual Nearest Neighbors.” *Nature Biotechnology* 36 (5): 421–27.

Hamilton, Duane H., Justin M. David, Charli Dominguez, and Claudia Palena. 2017. “Development of Cancer Vaccines Targeting Brachyury, a Transcription Factor Associated with Tumor Epithelial-Mesenchymal Transition.” *Cells, Tissues, Organs* 203 (2): 128–38.

Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck 3rd, Shiwei Zheng, Andrew Butler, Maddie J. Lee, et al. 2021. “Integrated Analysis of Multimodal Single-Cell Data.” *Cell* 184 (13): 3573–87.e29.

Hart, Adam H., Lynne Hartley, Koula Sourris, Elizabeth S. Stadler, Ruili Li, Edouard G. Stanley, Patrick P. L. Tam, Andrew G. Elefanty, and Lorraine Robb. 2002. “Mixl1 Is Required for Axial Mesendoderm Morphogenesis and Patterning in the Murine Embryo.” *Development* 129 (15): 3597–3608.

Hein, Nadine, Donald P. Cameron, Katherine M. Hannan, Nhu-Y N. Nguyen, Chun Yew Fong, Jirawas Sornkom, Meaghan Wall, et al. 2017. “Inhibition of Pol I Transcription Treats Murine and Human AML by Targeting the Leukemia-Initiating Cell Population.” *Blood* 129 (21): 2882–95.

Ho, Lena, Shawn Y. X. Tan, Sheena Wee, Yixuan Wu, Sam J. C. Tan, Navin B. Ramakrishna, Serene C. Chng, et al. 2015. “ELABELA Is an Endogenous Growth Factor That Sustains hESC Self-Renewal via the PI3K/AKT Pathway.” *Cell Stem Cell* 17 (4): 435–47.

Huang, B., J. R. Cohen, R. I. Fernando, D. H. Hamilton, M. T. Litzinger, J. W. Hodge, and C. Palena. 2013. “The Embryonic Transcription Factor Brachyury Blocks Cell Cycle Progression and Mediates Tumor Resistance to Conventional Antitumor Therapies.” *Cell Death & Disease* 4 (6): e682.

Imaz-Rosshandler, Ivan, Christina Rode, Carolina Guibentif, Mai-Linh N. Ton, Parashar Dhapola, Daniel Keitley, Ricard Argelaguet, et al. 2023. “Tracking Early Mammalian Organogenesis - Prediction and Validation of Differentiation Trajectories at Whole Organism Scale.” *bioRxiv*. https://doi.org/10.1101/2023.03.17.532833.

Jackson, Steven A., Jacqueline Schiesser, Edouard G. Stanley, and Andrew G. Elefanty. 2010. “Differentiating Embryonic Stem Cells Pass through ‘Temporal Windows’ That Mark Responsiveness to Exogenous and Paracrine Mesendoderm Inducing Signals.” *PloS One* 5 (5): e10706.

Jin, Xin, Sean K. Simmons, Amy Guo, Ashwin S. Shetty, Michelle Ko, Lan Nguyen, Vahbiz Jokhi, et al. 2020. “In Vivo Perturb-Seq Reveals Neuronal and Glial Abnormalities Associated with Autism Risk Genes.” *Science* 370 (6520). https://doi.org/10.1126/science.aaz6063.

Jonathan Griffiths, Aaron Lun. 2021. *Single-Cell -Omics Data across Mouse Gastrulation and Early Organogenesis* (version 1.14.0). Bioconductor package. https://bioconductor.org/packages/release/data/experiment/html/MouseGastrulationData.html.

Kiselev, Vladimir Yu, Andrew Yiu, and Martin Hemberg. 2018. “Scmap: Projection of Single-Cell RNA-Seq Data across Data Sets.” *Nature Methods* 15 (5): 359–62.

Landais, Yuna, and Céline Vallot. 2023. “Multi-Modal Quantification of Pathway Activity with MAYA.” *Nature Communications* 14 (1): 1668.

Lange, Marius, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, et al. 2022. “CellRank for Directed Single-Cell Fate Mapping.” *Nature Methods* 19 (2): 159–70.

Le Beau, M. M., R. Espinosa 3rd, W. L. Neuman, W. Stock, D. Roulston, R. A. Larson, M. Keinanen, and C. A. Westbrook. 1993. “Cytogenetic and Molecular Delineation of the Smallest Commonly Deleted Region of Chromosome 5 in Malignant Myeloid Diseases.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (12): 5484–88.

Liu, Chunqiao, Eiichiro Nakamura, Vladimir Knezevic, Sherrie Hunter, Katherine Thompson, and Susan Mackem. 2003. “A Role for the Mesenchymal T-Box Gene Brachyury in AER Formation during Limb Development.” *Development* 130 (7): 1327–37.

Lotfollahi, Mohammad, Mohsen Naghipourfar, Malte D. Luecken, Matin Khajavi, Maren Büttner, Marco

Wagenstetter, Žiga Avsec, et al. 2022. “Mapping Single-Cell Data to Reference Atlases by Transfer Learning.” *Nature Biotechnology* 40 (1): 121–30.
Mann, H. B., and D. R. Whitney. 1947. “On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other.” *The Annals of Mathematical Statistics* 18 (1): 50–60.
McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. “UMAP: Uniform Manifold Approximation and Projection.” *Journal of Open Source Software* 3 (29): 861.
Mittnenzweig, Markus, Yoav Mayshar, Saifeng Cheng, Raz Ben-Yair, Ron Hadas, Yoach Rais, Elad Chomsky, et al. 2021. “A Single-Embryo, Single-Cell Time-Resolved Model for Mouse Gastrulation.” *Cell* 184 (11): 2825–42.e22.
Naiche, L. A., Zachary Harrelson, Robert G. Kelly, and Virginia E. Papaioannou. 2005. “T-Box Genes in Vertebrate Development.” *Annual Review of Genetics* 39: 219–39.
Palena, Claudia, Romaine I. Fernando, Mary T. Litzinger, Duane H. Hamilton, Bruce Huang, and Jeffrey Schlom. 2011. “Strategies to Target Molecules That Control the Acquisition of a Mesenchymal-like Phenotype by Carcinoma Cells.” *Experimental Biology and Medicine* 236 (5): 537–45.
Papaioannou, Virginia E. 2014. “The T-Box Gene Family: Emerging Roles in Development, Stem Cells and Cancer.” *Development* 141 (20): 3819–33.
Pereira, Lloyd A., Michael S. Wong, Sue Mei Lim, Edouard G. Stanley, and Andrew G. Elefanty. 2012. “The Mix Family of Homeobox Genes--Key Regulators of Mesendoderm Formation during Vertebrate Development.” *Developmental Biology* 367 (2): 163–77.
Pijuan-Sala, Blanca, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, et al. 2019. “A Single-Cell Molecular Map of Mouse Gastrulation and Early Organogenesis.” *Nature* 566 (7745): 490–95.
Prummel, Karin D., Christopher Hess, Susan Nieuwenhuize, Hugo J. Parker, Katherine W. Rogers, Iryna Kozmikova, Claudia Racioppi, et al. 2019. “A Conserved Regulatory Program Initiates Lateral Plate Mesoderm Emergence across Chordates.” *Nature Communications* 10 (1): 3857.
Pulina, Maria V., Kenneth E. Sahr, Sonja Nowotschin, Margaret H. Baron, and Anna-Katerina Hadjantonakis. 2014. “A Conditional Mutant Allele for Analysis of Mixl1 Function in the Mouse.” *Genesis* 52 (5): 417–23.
Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. 2013. “Genome Engineering Using the CRISPR-Cas9 System.” *Nature Protocols* 8 (11): 2281–2308.
Replogle, Joseph M., Thomas M. Norman, Albert Xu, Jeffrey A. Hussmann, Jin Chen, J. Zachery Cogan, Elliott J. Meer, et al. 2020. “Combinatorial Single-Cell CRISPR Screens by Direct Guide RNA Capture and Targeted Sequencing.” *Nature Biotechnology* 38 (8): 954–61.
Replogle, Joseph M., Reuben A. Saunders, Angela N. Pogson, Jeffrey A. Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, et al. 2022. “Mapping Information-Rich Genotype-Phenotype Landscapes with Genome-Scale Perturb-Seq.” *Cell* 185 (14): 2559–75.e28.
Robinson, Mark D., and Alicia Oshlack. 2010. “A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data.” *Genome Biology* 11 (3): R25.
Saint-Paul, Laetitia, Chi-Hung Nguyen, Anne Buffière, Jean-Paul Pais de Barros, Arlette Hammann, Corinne Landras-Guetta, Rodolphe Filomenko, et al. 2016. “CD45 Phosphatase Is Crucial for Human and Murine Acute Myeloid Leukemia Maintenance through Its Localization in Lipid Rafts.” *Oncotarget* 7 (40): 64785–97.
Saunders, Lauren M., Sanjay R. Srivatsan, Madeleine Duran, Michael W. Dorrity, Brent Ewing, Tor Linbo, Jay Shendure, et al. 2022. “Deep Molecular, Cellular and Temporal Phenotyping of Developmental Perturbations at Whole Organism Scale.” *bioRxiv*. https://doi.org/10.1101/2022.08.04.502764.
Schiebinger, Geoffrey, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, et al. 2019. “Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming.” *Cell* 176 (4): 928–43.e22.
Sjövall, Daniel, Sudip Ghosh, Jenny Hansson, Carolina Guibentif, and Pekka Jaako. 2023. “Defective Ribosome Assembly Impairs Leukemia Stem Cell Function in a Murine Model of Acute Myeloid Leukemia.” *bioRxiv*. https://doi.org/10.1101/2023.01.22.525120.
Szöcs, Katalin, Bernard Lassègue, Philip Wenzel, Maria Wendt, Andreas Daiber, Matthias Oelze, Thomas Meinertz, Thomas Münzel, and Stephan Baldus. 2007. “Increased Superoxide Production in Nitrate Tolerance Is Associated with NAD(P)H Oxidase and Aldehyde Dehydrogenase 2 Downregulation.” *Journal of Molecular and Cellular Cardiology* 42 (6): 1111–18.
Ton, Mai-Linh Nu, Daniel Keitley, Bart Theeuwes, Carolina Guibentif, Jonas Ahnfelt-Rønne, Thomas Kjærgaard Andreassen, Fernando J. Calero-Nieto, et al. 2023. “An Atlas of Rabbit Development as a Model for Single-Cell Comparative Genomics.” *Nature Cell Biology* 25 (7): 1061–72.

Wenzel, Philip, Johanna Müller, Sarah Zurmeyer, Swenja Schuhmacher, Eberhard Schulz, Matthias Oelze, Andrea Pautz, et al. 2008. “ALDH-2 Deficiency Increases Cardiovascular Oxidative Stress--Evidence for Indirect Antioxidative Properties.” *Biochemical and Biophysical Research Communications* 367 (1): 137–43.

Wray, Jason, Tüzer Kalkan, Sandra Gomez-Lopez, Dominik Eckardt, Andrew Cook, Rolf Kemler, and Austin Smith. 2011. “Inhibition of Glycogen Synthase Kinase-3 Alleviates Tcf3 Repression of the Pluripotency Network and Increases Embryonic Stem Cell Resistance to Differentiation.” *Nature Cell Biology* 13 (7): 838–45.

Wu, Junqing, Yanyu Xiao, Jie Sun, Huiyu Sun, Haide Chen, Yuanyuan Zhu, Huarui Fu, et al. 2020. “A Single-Cell Survey of Cellular Hierarchy in Acute Myeloid Leukemia.” *Journal of Hematology & Oncology* 13 (1): 128.

Wu, San-Pin, Xiu-Rong Dong, Jenna N. Regan, Chang Su, and Mark W. Majesky. 2013. “Tbx18 Regulates Development of the Epicardium and Coronary Vessels.” *Developmental Biology* 383 (2): 307–20.

Ying, Qi-Long, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. 2008. “The Ground State of Embryonic Stem Cell Self-Renewal.” *Nature* 453 (7194): 519–23.

Zhang, Hong, Meritxell Alberich-Jorda, Giovanni Amabile, Henry Yang, Philipp B. Staber, Annalisa Di Ruscio, Robert S. Welner, et al. 2013. “Sox4 Is a Key Oncogenic Target in C/EBPα Mutant Acute Myeloid Leukemia.” *Cancer Cell* 24 (5): 575–88.

Zhang, Yijie, Dan Wang, Miao Peng, Le Tang, Jiawei Ouyang, Fang Xiong, Can Guo, et al. 2021. “Single-cell RNA Sequencing in Cancer Research.” *Journal of Experimental & Clinical Cancer Research: CR* 40 (1): 1–17.