A pluripotent stem cell atlas of multilineage differentiation reveals *TMEM88* as a developmental regulator of mammalian blood pressure

Sophie Shen[1], Tessa Werner[1], Samuel W. Lukowski[1], Stacey Andersen[1,2], Yuliangzi Sun[1], Woo Jun Shim[1], Dalia Mizikovsky[1], Sakurako Kobayashi[1], Jennifer Outhwaite[1], Han Sheng Chiu[1], Xiaoli Chen[1], Gavin Chapman[3,4], Ella M. M. A. Martin[3], Di Xia[1,2], Duy Pham[1], Zezhuo Su[5,6], Daniel Kim[7], Pengyi Yang[7,8], Men Chee Tan[1,9], Enakshi Sinniah[1], Qiongyi Zhao[1], Sumedha Negi[1], Meredith A. Redd[1], Joseph E. Powell[10,11], Sally L. Dunwoodie[3,4,12], Patrick P. L. Tam[13,14], Mikael Bodén[15], Joshua W. K. Ho[5,6], Quan Nguyen[1], Nathan J. Palpant[1,*8]

[1] Institute for Molecular Bioscience, The University of Queensland, St Lucia QLD 4072, Australia

[2] Genome Innovation Hub, The University of Queensland, St Lucia QLD 4072, Australia

[3] Victor Chang Cardiac Research Institute, Darlinghurst NSW 2010, Australia

[4] School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW 2052, Australia

[5] School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

[6] Laboratory of Data Discovery for Health Limited ($D^2$4H), Hong Kong Science Park, Hong Kong SAR, China

[7] Computational Systems Biology Group, Children's Medical Research Institute, University of Sydney, Westmead NSW 2145, Australia

[8] Charles Perkins Centre, School of Mathematics and Statistics, University of Sydney, Camperdown NSW 2006, Australia

[9] Queensland Facility for Advanced Genome Editing, The University of Queensland, St Lucia QLD 4072, Australia

[10] Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Darlinghurst NSW 2010, Australia

[11] University of New South Wales, Cellular Genomics Futures Institute, Sydney NSW 2052, Australia

[12] Faculty of Science, University of New South Wales, Sydney, NSW 2052, Australia

[13] Embryology Research Unit, Children's Medical Research Institute, University of Sydney, Westmead NSW 2145, Australia

[14] School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown NSW 2006, Australia

[15] School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia QLD 4072, Australia

[*] Corresponding Author

Corresponding Author
Associate Professor Nathan Palpant
Institute for Molecular Bioscience
The University of Queensland
Brisbane, Australia, 4072
Email: n.palpant@uq.edu.au
Phone: +61 0439241069

SUMMARY

Pluripotent stem cells provide a scalable approach to analyse molecular regulation of cell differentiation across multiple developmental lineage trajectories. In this study, we engineered barcoded iPSCs to generate an atlas of multilineage differentiation from pluripotency, encompassing a time-course of WNT-induced differentiation perturbed using modulators of WNT, BMP, and VEGF signalling. Computational mapping of *in vitro* cell types to *in vivo* developmental lineages revealed a diversity of iPSC-derived cell types comprising mesendoderm lineage cell types including lateral plate and paraxial mesoderm, neural crest, and primitive gut. Coupling this atlas of *in vitro* differentiation with Summary data-based Mendelian Randomisation analysis of human complex traits, we identify the WNT-inhibitor protein *TMEM88* as a putative regulator of mesendodermal cell types governing development of diverse cardiovascular and anthropometric traits. Using genetic loss of function models, we show that *TMEM88* is required for differentiation of diverse endoderm and mesoderm cell lineages *in vitro* and that *TMEM88* knockout *in vivo* results in a significant dysregulation of arterial blood pressure. This study provides an atlas of multilineage iPSC differentiation coupled with new molecular, computational, and statistical genetic tools to dissect genetic determinants of mammalian developmental physiology.

KEYWORDS

Human pluripotent stem cell, CHIR-99021, single-cell, multiplexing, barcode, iPSC, TRIAGE, complex trait genetics

INTRODUCTION

Human pluripotent stem cell (hPSC) models are ideally suited to deconstruct early human developmental processes in a scalable and cell type-specific manner. Significant investment and sector growth in stem cell technologies highlight their importance in developmental biology research, industry drug discovery pipelines, disease modelling, and regenerative medicine[1]. In contrast to *in vivo* cell atlases covering numerous developmental stages across diverse species[2-4], atlases of *in vitro* differentiation are not widely available. As a consequence, mechanisms governing cell differentiation and fate specification across cell lineages remain inadequately understood[5-7], creating barriers in guiding the differentiation of hPSCs into cell types to study the mechanism of cell fate decisions[8-10]. Strategies to perturb and map sequential stages of multi-lineage cell differentiation decisions are needed to understand and optimise methods for deriving cell types that model and inform mechanisms of development and disease *in vivo*.

This study presents a multi-scale genetic analysis workflow across *in vitro* and *in vivo* biology to accelerate knowledge gain into mechanisms of development. We first develop a multiplexing method for single-cell RNA sequencing (scRNA-seq), enabled by genomic integration of transcribed barcodes in human induced pluripotent stem cells (iPSCs). Using barcoded iPSCs, we perform a multiplexed perturbation study to characterise the diversification of cell types during multi-lineage differentiation from pluripotency *in vitro*. We evaluate this *in vitro* differentiation atlas using a suite of computational methods to extract the temporal and developmental regulatory mechanisms of differentiating cell lineages.

Translating insights from *in vitro* iPSCs into physiological mechanisms of development or disease remains a significant challenge because *in vitro* models lack complex tissue and organismal phenotypes. We address this limitation by drawing on human complex trait genetic data. Genome-wide association studies (GWAS) and databases mapping genetic diversity to transcriptional variation across tissues (GTEx) reveal how genetic changes influence phenotypic variation in humans. By comparing genetic regulators of *in vitro* cell diversification with GWAS data from UK Biobank, we identify and functionally validate the transmembrane protein, *TMEM88*, as a genetic regulator of cardiovascular blood pressure. Collectively, this study provides new cell-based experimental tools, data resources, and computational workflows that leverage the power of atlas-level multilineage *in vitro* cell diversification to discover genetic determinants of developmental physiology.

RESULTS

Derivation of barcoded iPSCs for multiplexed single-cell RNA-seq

Cell barcoding strategies enable sample multiplexing and economise the effort and costs of data generation by scRNA-seq analysis. To enhance sample multiplexing capabilities in human induced pluripotent stem cells (hiPSCs), we designed a cell barcoding method to enable multiplexed single-cell analysis of isogenic human cell lines (**Figure 1a**). 10,000 15-base pair barcodes were randomly generated using an even 25% probability for the presence of each A, C, T and G nucleotides. After excluding barcodes starting or ending with a stop codon or containing repetitive runs of 4 or more nucleotides, we selected 18 barcodes, requiring a minimum Hamming distance of 5 nucleotides between each pair of barcodes (**Table 1**). Using this design, we built barcode plasmids into the AAVS1-2A-Puro gene targeting cassette[11,12] to enable expression of a barcoded GFP transcript downstream of the ubiquitously expressed CAG at the

human *AAVS1* locus after CRISPR-Cas9 integration (**Supplementary File 1**). Transcribed barcodes can thus be captured by any single-cell RNA-seq protocol and amplified in a separate sequencing library to allow each read from pooled experimental samples to be assigned back to its sample of origin without the need for exogenous labelling before sample pooling (**Figure 1a**). With thousands of potential barcode combinations, the extent of multiplexing is only limited by the scale of cell capture technologies. We engineered 18 isogenic hiPSC lines with unique barcodes, each of which passed quality testing for karyotype, pluripotency, and purity (**Figure 1b-e**).

We performed a pilot scRNA-seq experiment pooling all 18 barcoding lines as undifferentiated iPSCs to validate homogeneity of their transcriptomic profiles and to confirm that the transcribed barcodes could be used to easily assign a sample of origin for each cell (**Figure 1f**). From a total of 6,711 cells captured, 19.4% had no reads mapped to any genomic barcode, however 95.0% of these cells with no genomic barcode reads also did not pass quality control on the basis of their low total read count, gene count, and high mitochondrial RNA content (**Table S1** and **Figure S1a**). The remainder of cells had reads mapped to between one and seven genomic barcodes, where on average, cells were mapped to two barcodes, with just one read mapped to the lower "expressed" barcode (**Table S1** and **Figure 1g-h**). Dimensionality reduction analysis shows that transcriptomes are sufficiently homogenous to randomly distribute across UMAP space (**Figure 1f),** and cells from each line reveal comparable reads per cell and expression of pluripotency markers *(POU5F1)* (OCT4) and *SOX2* (**Figure 1i** and **Figure S1a**). We additionally used cell hashing antibodies[13] to label four samples (BC07, BC08, BC12, BC13) to benchmark the accuracy of our method against an established technology. 882 cells were mapped to this subset of genomic barcodes, while 1104 cells had cell hashing (HTO) reads mapped, with an 83.2% agreement between the genomic barcoding and cell hashing sample assignment (**Figure 1j** & **Table S1**). We also observed comparable transcriptomic quality between the two methods (**Figure 1g-j** and **Figure S1a**), providing support for the reliability of using genomic transcribed barcodes for isogenic sample multiplexing in scRNA-seq experiments.

Generation of a multiplexed atlas of iPSC differentiation *in vitro*

To further evaluate efficacy of these barcoded cell lines for sample multiplexing, we set out to utilise them to capture a dataset that could make full use of the multiplexing potential. Given the diverse signalling and molecular cues shown to inform cell lineage specification during early development, we generated a dataset that could be used to study the role of these molecular determinants during WNT-induced hiPSC differentiation from pluripotency[14,15] (**Figure 2a**). First, we utilised cell hashing to capture a reference time course of differentiation under control conditions, collecting cells at eight time points; every 24 hours from day 2 to day 9 of differentiation. Next, we treated differentiating cells between day 3 (germ layer differentiation) and day 5 (progenitor cell stage) with perturbations including inhibitors and activators of WNT[16], BMP[17], and VEGF[18] signalling pathways known to be essential regulators of developmental lineage differentiation (**Figure 2b** and **Table 1**). In this signalling perturbation dataset, cells were sampled in duplicates prior to perturbation (day 2), following perturbation (day 5), and at the differentiation endpoint of committed cell types (day 9).

In total, 62 independent cell samples were collected across four scRNA-seq libraries with comparable quality control metrics (**Table S2**). Both the cell hashing library for the time course dataset and genomic barcoding libraries for the signalling perturbation dataset had greater sequencing depth, averaging 600 reads per cell in the former and 190 reads per cell in the latter libraries, where majority of cells had reads mapping to all possible HTO or genomic barcode label

(**Table S2** and **Figure S1b**). The cell hashing library showed a greater difference in the percentage of reads mapping to the first and second-highest expressed barcode per cell, however the multiplexing doublet and singlet rates as determined by conventional sample demultiplexing strategies were comparable between the two methods (**Table S2** and **Figure S1b**). Standard quality control and pre-processing of data were performed (**Figure S1c)**, resulting in 13,682 high quality cells from the time course reference and 48,526 cells from the signalling perturbation dataset (**Figure S1d-g**).

The two datasets were integrated using established methods[19] into an atlas of multilineage hiPSC differentiation (**Figure 2c**). Clustering was performed on the integrated dataset to reveal 13 cell clusters, whose marker gene expression and associated enriched GO terms and KEGG pathways were evaluated to reveal capture of mesendoderm (clusters 0 & 1; *MIXL1*), definitive endoderm (clusters 2 & 3; *FOXA2*) anterior (cluster 8; *SOX2*) and posterior (cluster 5; *TTR*) foregut endoderm progenitors, lateral plate (cluster 6; *HAND1*) and paraxial mesoderm (cluster 7; *PRRX1*), cardiomyocytes (cluster 10; *MYH6*), endothelium (cluster 12; *CDH5*), axial mesoderm (cluster 4; *NOG*), ciliated node cells (cluster 11; *FOXJ1*), and neural progenitors (cluster 9; *SOX2*) (**Figure 2d** and **Figure S2a-c**). Further, expression of pluripotency marker *NANOG* in cluster 0 and enrichment of cell division-related GO terms in cluster 2 distinguish them as proliferative mesendoderm and definitive endoderm respectively (**Figure S2a-c**). Label transfer analysis with reference to mouse[2,20] and human[4] embryogenesis scRNA-seq datasets also provided support for these cell type labels (**Figure S2d**). To provide further annotation of the atlas dataset, we used established downstream analysis methods to characterise cell-cell communication between cell types resulting from each treatment condition using CellChat[21] (**Figure S2e, Table S3**, and **Supplementary File 2**), analysis of enriched gene regulons using pySCENIC[22,23] (**Figure S2f** and **Table S4**), and URD-predicted[24] differentiation trajectories relevant to each cluster (**Figure S2g**). Together, this dataset resource presents a deep phenotypic analysis of a time-resolved molecular cell atlas *in vitro* embedded with information of the molecular mechanisms of temporal and signalling control of lineage diversification and cellular heterogeneity during multilineage differentiation of induced pluripotent stem cells.

Determination of cell subtypes during iPSC differentiation

Following characterisation of the broad cell type groups comprising the dataset, we sought to understand changes in specific cell subtypes from each signalling perturbation. A common approach to this task is to increase the resolution, or granularity, of clustering to delineate smaller cell groups. The problem with this purely algorithm-driven approach is a lack of biological justification for a "correct" clustering resolution or cluster size to select. Given the nuanced signalling cues and small temporal increments in our atlas dataset, we anticipated this cell type ambiguity to be particularly relevant such that increasing clustering resolution would result in many small, uninformative clusters, rather than capturing cell subtypes.

To define such a transcriptional signature marking distinct, more biologically robust cell types, we use a package of tools built from the foundational *TRIAGE* method[25]. In brief, *TRIAGE* is a method that uses consortium-level epigenetic patterns to identify genes predicted to have high cell type-specific regulatory functions. In this way, *TRIAGE* provides an independent biological reference point for defining cell states. This method was adapted for application in single cell analysis pipelines in *TRIAGE-Cluster* and *TRIAGE-ParseR* to help identify cell populations and gene regulatory networks underpinning diverse cell types (**Figure 3a** and **Figure S3a**).

First, we used *TRIAGE-Cluster* to identify informative cell subtypes within broader cell type clusters, eliminating the need to evaluate tens or hundreds of small, transcriptionally similar clusters (**Figure 3a**, **Figure S3b** and **Table S5**). Based on the clustering output (**Figure 3a**), we took three approaches to evaluate *in vitro*-derived cell peaks by reference to *in vivo* derived developmental cell types. We first used gene set scoring to analyse cells based on positive gene sets defining seven major cell type domains modelled from 921 RNA-seq datasets[26]. We found that among the high-scoring cells, 5 general cell type domains are represented in our data: mesoderm, surface ectoderm, neural crest, germ, and embryonic domains (**Figure S3c**). We also used label transfer analysis to score *in vitro* derived cells based on their transcriptional similarity to single-cell atlases of mouse organogenesis[2] (**Figure 3b**) and human gastrulation[4] (**Figure 3c**). These reference datasets provide a preliminary indication of the *in vitro*-derived cell type diversity at the mid- to late timepoints (days 5-9), revealing a range of high-confidence matches to cell types including cardiomyocytes, hemogenic endothelium, and gut progenitor cells (**Figure 3b-c**), aligning with the broad cell type annotations from **Figure 2d**. For the earlier timepoints (days 2-4), we assessed transcriptomic similarity of the cell peaks to embryonic mouse spatial transcriptome data[27], revealing early cell types corresponding to gastrulation-stage cell types including those in the epiblast, primitive streak, and mesendoderm (**Figure 3d** and **Figure S3d**).

Evaluation of gene programs underpinning cell subtypes

We next used *TRIAGE* to evaluate cell type-specific gene expression patterns. *TRIAGE* analysis can return a list of the most cell type-defining genes of any given group of cells, akin to differentially expressed or highly variably expressed genes, which can be used as input for common annotation analyses including gene ontology (GO) or KEGG pathway enrichment. (**Table S6**) Unlike differentially expressed genes however, this approach allows each cell group's expression to be represented without comparison to any other cell cluster in the dataset. We show that the top 100 most highly ranked *TRIAGE* genes for each cell type can be used to inform regulatory relationships between cell types by evaluating shared and unique genes for each cell type peak (**Figure 3e**).

Using these 100 top ranked *TRIAGE* genes for each cell type, we implemented *TRIAGE-ParseR*[28], which identifies co-regulated molecular processes driving cell type identity (**Figure 3f**). By separating gene lists into related sub-groups, this analysis can enhance enrichment readout from methods such as Gene Ontology (GO) or STRING network analysis and thus clarify cell type annotation (**Tables S6-8**). As an example, *TRIAGE-ParseR* analysis of cell peak 37 reveals three main gene modules with GO enriched terms for cardiac chamber development and morphogenesis of cardiac septa (**Figure 3g**), suggesting its transcriptional profile aligns with the embryonic ventricular septum.

Annotation of cell types was performed by drawing on all the provided results to assign cell type annotations to each *TRIAGE* peak, taking into consideration expression of marker genes, evaluation against *in vivo* reference points, the top *TRIAGE* identity genes, and enriched *TRIAGE-ParseR* gene programs in each cell type. These results provide an integrated view of iPSC *in vitro* multilineage differentiation, to facilitate interrogation of temporal and signalling-dependent cell subtype differentiation (**Figure 3h** and **Figure S3e**).

The atlas reveals cells representing gastrulation stage cell types at day 2 including epiblast, ectoderm, and mesendoderm. Time course data between days 3-5 reveal the emergence of paraxial, cardiac, and forelimb mesoderm together with definitive endoderm and posterior foregut. A sampling of time course and signalling perturbation data on day 5 reveal the first appearance of endocardial endothelium, cells of the anterior foregut, notochord, and neural tube. As cells progress between days 6 and 9 in the time course, we identify the emergence of first and second heart field

cardiac lineage cells, and cells of posterior foregut lineages, including pancreas and liver. On day 9, integrated analysis of time course and signalling data show differentiation of additional definitive cell types including ventricular cardiomyocytes, and more developed progenitors of the hepatopancreatic, anterior foregut, and neural lineages. Collectively, these wet and dry lab methods enable the elucidation of mechanisms controlling multilineage differentiation from pluripotency.

Analysis of signalling pathway-associated cell differentiation to identify genetic regulators of cell fate

We next assessed differentiation lineage biases arising from each signalling perturbation (**Figure 4a** and **Figure S4a**, see also **Figure 2b**). At day 5, cells are broadly distributed across definitive endoderm, foregut endoderm, axial, and general mesoderm. At day 9, however, marked differences emerge. Cells from the control groups show preferential differentiation of posterior foregut-derived hepatopancreatic progenitors, as well as a small proportion of cardiac progenitor cells. This co-differentiation of foregut and cardiac lineages is aligned with embryonic development, where shared signalling cues induce differentiation of the adjacent progenitor populations[29]. Treatments with a lower dosage of XAV-393 (low XAV) and recombinant VEGF produce similar proportions of cell subpopulations to the control at day 9, suggesting that these two treatments are at insufficient levels to affect differentiation at that post-gastrulation mesendoderm stage.

In contrast, differentiation of anterior foregut endoderm was observed in conditions exposed to treatment with the WNT agonist CHIR-99021 (CHIR) as well as all BMP signalling perturbations (**Figure 4a**). Studies have demonstrated an anteriorising role for BMP inhibition in specifying the foregut endoderm, aligning with the observed effect in the three BMP inhibitory treatments (0.4uM and 4uM Dorsomorphin, and K02288 treatments)[30]. For the anticipated posteriorising WNT and BMP4 signalling cues, it is possible that endogenous feedback mechanisms are activated to correct the effects of the CHIR and BMP4 treatments resulting in similar cell type proportions to the BMP inhibitor treatments at day 9. Lastly, **Figure 4a** shows that higher dosage of WNT inhibitor XAV distinctly promoted differentiation of cardiovascular lineage cell types at the expense of the foregut endoderm lineages, in line with established protocols for cardiac-directed differentiation[31].

To leverage this multilineage atlas to discover genetic factors governing cell differentiation, we first compared differentiated cell types from each signalling perturbation at day 9, evaluating pairwise transcriptional correlation and differential gene expression between treatments (**Figure 4b-c**). These data demonstrate that modulation of the WNT pathway has the most significant impact differentiation trajectory, forming the most distinct cell types by day 9. We therefore used differential gene expression between the two WNT-related treatment conditions affecting on day 9 to identify candidate mediators of WNT-requisite cell fate specification, utilising gene ontology annotations to focus on genes known to influence WNT signalling and its effects (**Figure 4d, y-axis** and **Figure S4b**). This narrowed the list to a set of candidate genes associated with XAV and CHIR modulation of the WNT pathway, primarily mediating cardiac vs. foregut lineage differentiation respectively.

To link these genes to a broader physiological context, we used Summary data-based Mendelian Randomisation (SMR) analysis to evaluate associations between expression of each gene in cardiovascular tissue and effects on genetically-driven anthropometric and cardiovascular traits (**Figure 4d, x-axis**)[32]. We selected the most significantly differentially expressed WNT-related genes in the XAV (*TMEM88* & *NDRG2*) and CHIR (*FZD5* & *TCF7L2*) treatments as candidates for further investigation into specific phenotypes identified from the SMR analysis (**Figure**

4d-f). We found significant associations between eQTL-mediated changes in *TMEM88*, *NDRG2* and *FZD5* expression with myocardial and vascular traits including heart rate and deep venous thrombosis (**Figure 4f**). The bulk of associations for *TCF7L2* stem from variation in its expression in the aorta to affect anthropometric traits including hip circumference and body fat measures, in line with its ties to type 2 diabetes[33]. Given the expression of *FZD5* and *TCF7L2* in the endodermal foregut lineage cell types (**Figure 4e**), we also performed SMR focusing on foregut-related tissue and phenotypes, which showed some enrichment for respiratory phenotypes including asthma (**Figure S4c**).

Of the four candidate genes, *FZD5, TCF7L2,* and *TMEM88* are also significantly upregulated between days 2 and 5 of differentiation, supporting a more direct role during Wnt-driven fate specification at the progenitor stage (**Figure S4d**). *FZD5* and *TCF7L2* have well-established roles in the WNT signalling pathway across foregut, intestinal, and neural development, with the former encoding a receptor in the canonical WNT pathway[34,35] and the latter a transcription factor effector of WNT signalling[36-38]. *NDRG2* has been shown to inhibit β-catenin target genes in heart, brain and liver development, as well as being relevant in heart failure contexts[39-41]. While prior work has implicated *TMEM88* as a WNT inhibitor relevant in cardiac and pharyngeal pouch development[42-45], the SMR analysis indicates its expression in cardiac tissue being most relevant to vascular, rather than myocardial phenotypes in human populations (**Figure 4f**). Furthermore, interrogation of organism-wide gene expression databases shows that *TMEM88* expression localises primarily to vascular endothelium in adult organisms (**Figure S4d-e**), leading us to identify *TMEM88* as a candidate with potential for discovery into a broader role in cardiovascular biology.

Role of *TMEM88* in cardiovascular cell differentiation and physiology

To investigate the role of *TMEM88* on differentiation *in vitro*, we used CRISPRi to generate a doxycycline-inducible *TMEM88* knockdown (KD) hiPSC line (**Figure 5a-b**). We generated a high resolution scRNA-seq time course of mesendoderm differentiation to evaluate the effect of *TMEM88* KD (gain of WNT signalling) compared to control conditions and WNT inhibitory conditions (5μM XAV) (**Figure 5c-d**). We captured 24,841 cells across eight time points between days 2 and 9 of differentiation (**Figure 5d**). UMAP plots show marker genes across the time course dataset and conditions (**Figure 5e**). Using data analysis workflows described in **Figure 2** and **Figure 3**, we identified and annotated 23 cell peaks across the dataset (**Figure 5f-g** and **Figure S5a-e**). Comparing the proportion of cells allocated to each cell type over time for each treatment, we identified notable perturbations in multilineage cell differentiation in *TMEM88* KD cells including persistence of lateral plate and paraxial mesoderm as well as depletion of endothelium and posterior foregut and liver bud progenitor cell types (**Figure 5g**). Genetic validation studies using antibody-based fluorescence activated cell sorting confirmed that, compared to control iPSCs, *TMEM88* KD reduced differentiation of endothelial and cardiac cells *in vitro* (**Figure 5h** and **Figure S5f**), further supporting a role for *TMEM88* in cardiovascular cell identity and differentiation.

To evaluate the role of *TMEM88* in cardiovascular development *in vivo,* we generated a *Tmem88* conditional knockout mouse model by crossing a floxed *Tmem88* mouse (**Figure 5i**) and a mouse with ubiquitous Cre recombinase expression driven from a CMV promoter. Quantitative RT-PCR analysis of heart tissue showed genotype-dependent loss of *Tmem88* gene expression (**Figure 5j**). All genotypes were born in Mendelian ratios and displayed no observable gross developmental phenotypes (**Figure S5g**). However, micro-computed tomography scans of E15.5 embryos showed significant changes to organ volumes and structures including the vertebrae, pelvic girdle, and ductus arteriosus in *Tmem88* KO embryos (**Figure S5h**).

285 Focusing on cardiovascular physiology, analysis by two-dimensional b-mode and one-dimensional m-mode
286 echocardiography showed no impacts on cardiac structure and function in adult mice (**Figure 5k** and **Figure S5i**),
287 suggesting a minor role for *Tmem88* in cardiac morphogenesis and function. Doppler analysis of blood flow across the
288 mitral valve showed a significant reduction in the E/A and E/E' in *Tmem88* KO mice, suggesting mild diastolic
289 dysfunction (**Figure 5l**). Lastly, based on SMR results indicating significant associations between *TMEM88* and blood
290 pressure phenotypes across diverse tissues (**Figure 5m** and **Table S9**), we evaluated systemic arterial blood pressure
291 in WT and KO littermates using catheter-based conductance micromanometry analysis of pressure traces from the
292 right carotid artery (**Figure 5n**). These data show that, compared to WT controls, *Tmem88* KO mice had significantly
293 increased variability in systolic and diastolic blood pressure, a phenotype associated with diverse clinical pathologies
294 including hypertension and end-organ damage[46]. To our knowledge, this is the first data implicating *TMEM88* in
295 developmental regulation of blood pressure, providing compelling genetic evidence of its tissue specificity and
296 functional significance that justify further work to understand the biological mechanism of action. Collectively, these
297 data demonstrate the power of coupling large-scale multilineage iPSC differentiation *in vitro* with population-level
298 human statistical genetic analysis of complex traits as a basis to discover genetic regulators of developmental
299 physiology.

300

301 **DISCUSSION**

302 This study generated barcoded hPSCs to facilitate multiplexed scRNA-seq analysis of cell differentiation,
303 demonstrated by generation of a single-cell transcriptomic atlas chronicling differentiation of pluripotent stem cells *in*
304 *vitro*. We further implement a two-part computational approach to cell clustering, including mapping to *in vivo* cell
305 types and classification of well-defined cell subtypes across diverse lineages to provide, at single-cell level, the
306 molecular attributes of multilineage cell differentiation. We couple these data with human population-level complex
307 trait data to link candidate genetic regulators of development with insights into putative physiological functions in
308 adult mammals. Using genetic models, we validate a role for *TMEM88* in governing cardiovascular development, with
309 a primary impact on regulating mammalian blood pressure.

310 This study provides diverse new tools and harnesses data to maximise knowledge-gain into mechanisms of
311 development and disease. First, we develop novel cell barcoding strategies. Unlike existing scRNA-seq sample
312 multiplexing strategies[13,47,48], endogenous cellular barcodes broaden the experimental capability for *in vitro* and *in vivo*
313 studies of hPSC differentiation[49]. Barcoding isogenic lines via gene targeting to the *AAVS1* locus allows for
314 constitutively expressed barcodes in all cell types for direct comparison of perturbations, reducing variability related
315 to reprogramming and genotype differences between hPSC[50,51] samples and simplifying sample preparation. The
316 CRISPR-Cas9 integration of stable genomic barcodes in this study notably differs from barcoding methods that utilise
317 viral barcode delivery into cells for isogenic multiplexing and lineage tracing applications[49,52]. Lentiviral barcode
318 delivery is highly flexible and allows for successive rounds of barcode incorporation over the course of an experiment
319 to allow direct capture of ancestor-descendant relationships during differentiation and reprogramming[53]. As the
320 genomic barcodes are introduced prior to the start of an experiment, they are limited to lower throughput clonal
321 tracking experiments, or those where separately-prepared and barcoded progenitor cells are pooled for co-culture. This
322 method however has the benefit of a defined barcode integration location, minimising potential for off-target effects

and more consistent expression across cells and cell types. This proves beneficial in contexts involving high cell type heterogeneity or significant changes in cell states, where lentiviral barcodes are more susceptible to cell type-specific silencing. Overall, as pre-experimental sample multiplexing methods, both barcode delivery methods have complementary use-cases and allow scalable, complex experimental designs for high throughput protocol development, organoid modelling, and *in vivo* cell transplantation studies.

Cell type annotation remains one of the most challenging issues in single-cell studies due to the need to relate cell heterogeneity to known biology. In this study, we implement a computational analysis pipeline outlined in a package of tools (*TRIAGE*[25], *TRIAGE-Cluster*, and *TRIAGE-ParseR*[28]) to define evolving cell types *in vitro* by referencing identity-defining genetic features. By restricting the analysis to biologically grounded cell types, we eliminated noise arising from technical variability in scRNA-seq capture and hPSC culture which complicate data interpretation. We demonstrated that this method facilitates more precise predictions of cell identity relationships. Furthermore, while machine learning approaches for cell type annotation are increasingly popular due to their automatic, unsupervised classification of cell types[54], they rely heavily on prior knowledge of cell type annotations. This is problematic particularly among intermediate cell types such as those derived at early stages of *in vitro* differentiation from pluripotency. This study demonstrates that *TRIAGE* provides an unsupervised pipeline to evaluate cell types in single-cell data atlases spanning developmental and mature cell types.

Transcriptional analysis of *in vitro* multilineage cell diversification provides insights into genetic regulation of cell decisions but has limited capabilities for interpreting tissue-level physiological attributes. We demonstrate the value of complementing *in vitro* atlas data with complex trait genetic data from adult human populations to inform physiological studies of cardiovascular development and disease. In particular we highlight the potential of this approach to facilitate discovery of new mechanisms and roles for minimally-characterised genes such as *TMEM88*. While we identify *TMEM88* as a WNT antagonist that regulates cell fate specification, prior studies have shown that WNT inhibition treatment only partially rescues the effects of its loss-of-function in cardiac development, suggesting possible function through other pathways unrelated to WNT signalling in development. Further, given the mismatch of phenotypes between *in vitro* and *in vivo TMEM88* loss-of-function likely arising from compensatory mechanisms masking more subtle or complex phenotypes *in vivo*[44,55], reference to independent statistical genetics data provided a novel hypothesis for *TMEM88* in an adult context. Our validation of this role *in vivo* reinforces the relevance of studying mechanisms that regulate lineage differentiation in broader developmental and disease contexts[56]. This study and approach therefore provide evidence for a novel drug target to address the ongoing health burden of cardiovascular disease.

Blood pressure variability is a significant risk factor in organ damage and a major determinant for diseases including atherosclerosis, chronic kidney disease, and dementia[46,57,58]. Though it is more common among the 33% of adults experiencing hypertension globally[59], blood pressure variability is independently associated with cardiocerebrovascular events including stroke and myocardial infarction, the two leading causes of death worldwide[60,61]. Analysis of one of the world's largest population-based assessments of middle-aged adults with measured blood pressure found that most treated hypertensives remain uncontrolled[62], and only a subset of antihypertensive drugs have demonstrated capability to target both high and variable blood pressure[63]. As there are

360 also no specific treatments for blood pressure variability[64], identification of novel gene targets governing blood
361 pressure, such as *TMEM88*, provides new opportunities to address this persistent public health concern.

362

363

LEAD CONTACT

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nathan J. Palpant (n.palpant@uq.edu.au)

ACKNOWLEDGEMENTS

We thank Nick Valmas for illustrations; the Australian Genome Research Facility (AGRF) for Sanger sequencing; the Flow Cytometry Facility at the Queensland Brain Institute for facilitating cell sorting; the University of Queensland sequencing facility at the Institute for Molecular Bioscience for performing single-cell capture and library preparation; and the Garvan Sequencing Platform for sequencing. The mutant floxed *Tmem88* mice were produced via CRISPR genome editing by the Monash Genome Modification Platform, Monash University as a node of Phenomics Australia (formerly Australian Phenomics Network). Phenomics Australia is supported by the Australian Government Department of Education through the National Collaborative Research Infrastructure Strategy, the Super Science Initiative, and the Collaborate Research Infrastructure Scheme. We acknowledge the Victor Chang Research Institute Innovation Centre, funded by the New South Wales Government.

Funding support was provided from the National Health and Medical Research Council of Australia (Grants 1143163 (N.J.P., J.E.P., and S.W.L.), 1110751 (P.P.L.T.), 1173469 (P.Y.), GNT2008928 (Q.N.), APP1107599 (J.E.P.), 2007896 (S.L.D)), the Medical Research Future Fund (APP2016033, N.J.P. and M.B.), the Australian Research Council (190102793 (P.P.L.T.)), the National Heart Foundation of Australia (Grants 101889 and 106721, N.J.P.), and AIR@InnoHK administered by Innovation and technology Commission (J.W.K.H).

AUTHOR CONTRIBUTIONS

To generate the barcoded iPS cell lines, S.L. designed the barcodes; S.A. and D.X. designed the plasmids. S.A. cloned the plasmids; T.W. and H.C. carried out the cell culture; D.X. designed the CRISPR targeting strategy. T.W. and D.X. performed the CRISPR gene editing. M.T. and D.X. performed the genotyping and Sanger sequencing. For the pilot barcoding scRNA-seq, T.W. and H.C. carried out the cell culture and collection; S.L. and S.A. processed and sequenced the samples; S.L. and S.S. performed the computational sample demultiplexing and quality control analysis. J.E.P. supervised the barcoding design and assisted with sequencing of barcoded libraries and facilitated data interpretation.

For the time course and *TMEM88* KD scRNA-seq datasets, H.C. generated the CRISPRi hiPSC line; S.S. and X.C. carried out cell culture, collection, and sample preparation; S.S. performed quality control and preliminary computational analysis. For the signalling perturbation scRNA-seq dataset, T.W. and H.C. carried out cell culture, collection, and sample preparation; S.S. and T.W. completed quality control and preliminary computational analysis. For analysis of the combined dataset, S.S. integrated the data and led analysis; E.S., Q.Z., and Y.S. performed additional characterisation analysis of broad cell clusters; Y.S. and S.S. conceptualised and implemented the *TRIAGE-Cluster* approach; Y.S. developed and benchmarked the *TRIAGE-Cluster* method with supervision from W.S.; D.K., P.Y., and P.P.L.T. performed annotation of cell types with reference to spatial gene expression patterns in the developing mouse embryo; D.P. constructed the online dataset visualisation dashboard and website; Z.S. implemented

the visualisation of gene expression to the platform; Q.N. supervised creation of the website; J.W.K.H supervised implementation of the gene expression visualisation; W.S. conceptualised and implemented the *TRIAGE-ParseR* analysis pipeline; M.B. supervised development of the *TRIAGE-ParseR* software.

For analysis of the role of *TMEM88*, D.M. performed statistical genetics analysis of *TMEM88* on human complex trait data; S.N. provided preliminary analysis of GWAS data linking variants in *TMEM88* with vascular physiology; S.K. performed analysis of *TMEM88* KD cardiac and endothelial iPS-derived cell types; E.M.M.A.M. performed micro-computed tomography and data quality control; G.D. performed LAMA analysis micro-computed tomography datasets and interpreted the results; S.L.D. supervised micro-computed tomography and LAMA analysis of the *Tmem88* embryos; S.S. performed echocardiographic imaging and associated analysis with supervision from M.A.R. and J.O.; J.O. performed conductance micromanometry analysis for analysis of blood pressure.

N.J.P. conceptualised the study, supervised the project, and raised funding. S.S. and N.J.P. wrote the manuscript.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

419 **FIGURES AND TABLES**

420 **Table 1: Barcodes and multiplexed experimental design**

Barcode Sequence	Line	Condition		Library 1	Library 2	Library 3
GTGCCGACCAGTATC	BC01	--	Control	Day 5	Day 2	Day 9
ACCACCTGACGCAAA	BC02	1 µM	XAV939	Day 5	Day 2	Day 9
ACGGCCCTATTTAAG	BC03	5 µM	XAV939	Day 5	--	Day 9
AGCCCTGAGTCAGTA	BC04	3 µM	CHIR99021	Day 5	Day 2	Day 9
CAAATTCAAGGCGAT	BC05	0.4 µM	Dorsomorphin	Day 5	Day 2	Day 9
AATCTTGTATAAGTA	BC06	4 µM	Dorsomorphin	Day 5	Day 2	Day 9
CGTCACATTTGAGTC	BC07	0.5 µM	K02288	Day 5	Day 2	Day 9
GGACCTTCTTACGAC	BC08	10 ng/ml	BMP4	Day 5	Day 2	Day 9
ACCCTACGGTGGTTC	BC11	100 ng/ml	VEGF	Day 5	Day 2	Day 9
CGCTAATGTCCGTTT	BC10	--	Control	Day 2	Day 9	Day 5
TACCAATTGTACGCT	BC09	1 µM	XAV939	Day 2	Day 9	Day 5
TGTCCAAGCTGCAAT	BC12	5 µM	XAV939	Day 2	Day 9	Day 5
GTGTATTTAAAGCCG	BC13	3 µM	CHIR99021	Day 2	Day 9	Day 5
ACACCCGTATGTCAC	BC14	0.4 µM	Dorsomorphin	Day 2	Day 9	Day 5
TCTTTCGATGGCGGT	BC15	4 µM	Dorsomorphin	Day 2	Day 9	Day 5
GAGCACCCGCGTATT	BC16	0.5 µM	K02288	Day 2	Day 9	Day 5
TTATTATGTTCTAGC	BC17	10 ng/ml	BMP4	Day 2	Day 9	Day 5
AATCTCTGAAACGAA	BC18	100 ng/ml	VEGF	Day 2	Day 9	Day 5

421

422

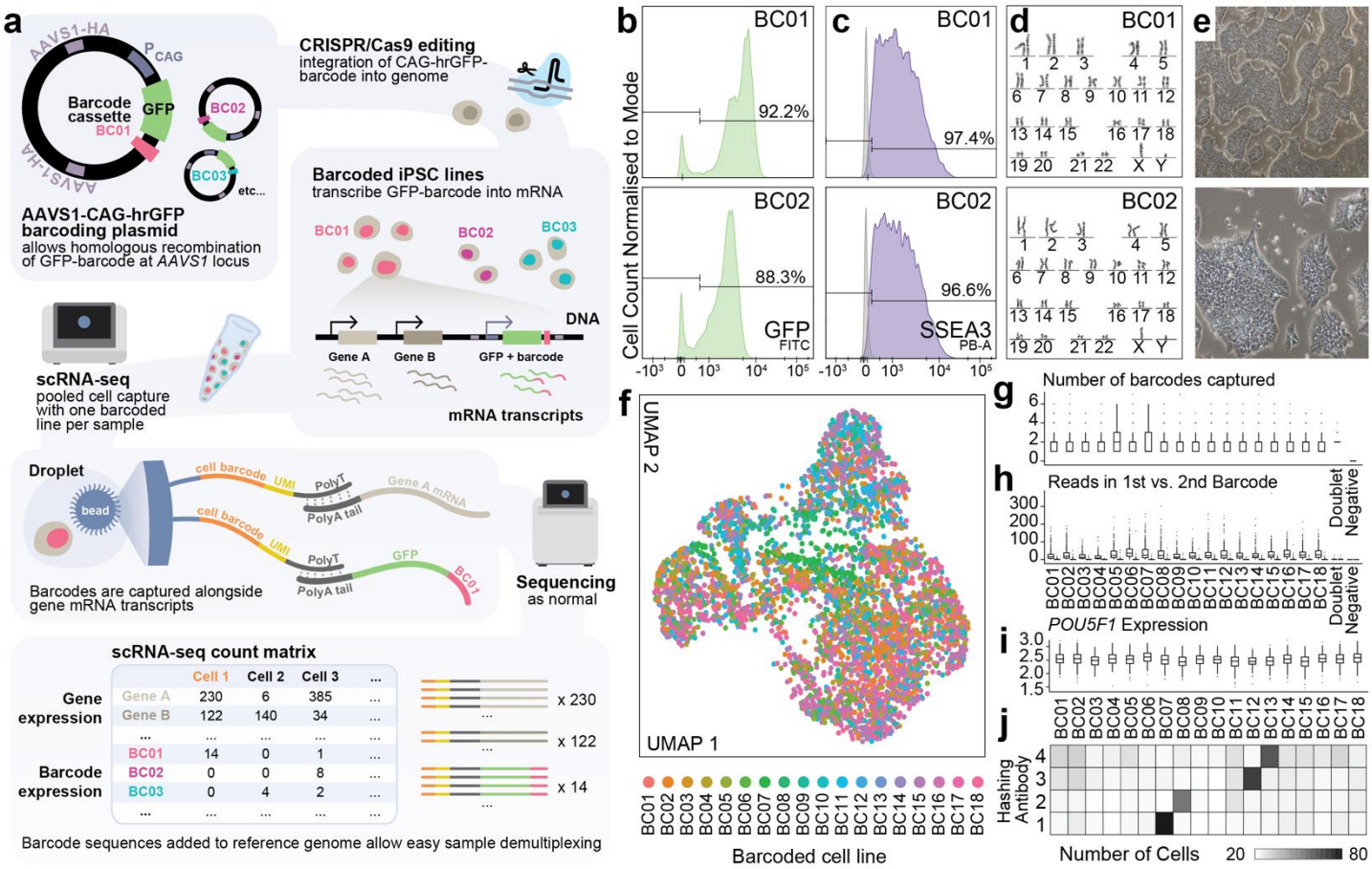

Figure 1: Engineered barcoding of human pluripotent stem cells.

a) Workflow for CRISPR/Cas9 engineering of barcoded hiPS cell lines to facilitate sample multiplexing of isogenic samples in scRNA-seq experiments. AAVS1-HA: *AAVS1* Homology Arm; hrGFP: humanised Renilla Green Fluorescent Protein; P_{CAG}: CAG promoter. See also **Methods** and **Supplementary File 1**.

b-e) Quality control for barcoded cell lines including representative FACS plots for GPF+ (vector integration) **(b)** and SSEA3+ (pluripotent) cells **(c)**, karyotype **(d)**, and images of pluripotent cell morphology **(e)** for barcoded cell lines BC01 and BC02.

f) UMAP of pilot scRNA-seq experiment capturing all 18 barcoded cell lines as undifferentiated iPSCs. Points are coloured by the barcode with the highest number of mapped reads per cell.

g) Number of unique expressed genomic barcodes (out of total 18 possible barcodes) captured per cell in the scRNA-seq experiment grouped by dominant barcode in each cell. See also **Figure S1a**.

h) Number of reads mapped to the top two most highly mapped genomic barcode in each cell, grouped by dominant barcode in each cell. See also **Figure S1a**.

i) Expression level of pluripotency gene *POU5F1* (OCT4) in each barcoded line. See also *SOX2* in **Figure S1a**.

j) Independent validation of barcoding by analysing barcoded cell lines co-labelled with external Cell Hashing antibodies. Antibodies 1-4 were stained to cells in barcoding lines BC07, BC08, BC12, and BC13 respectively.

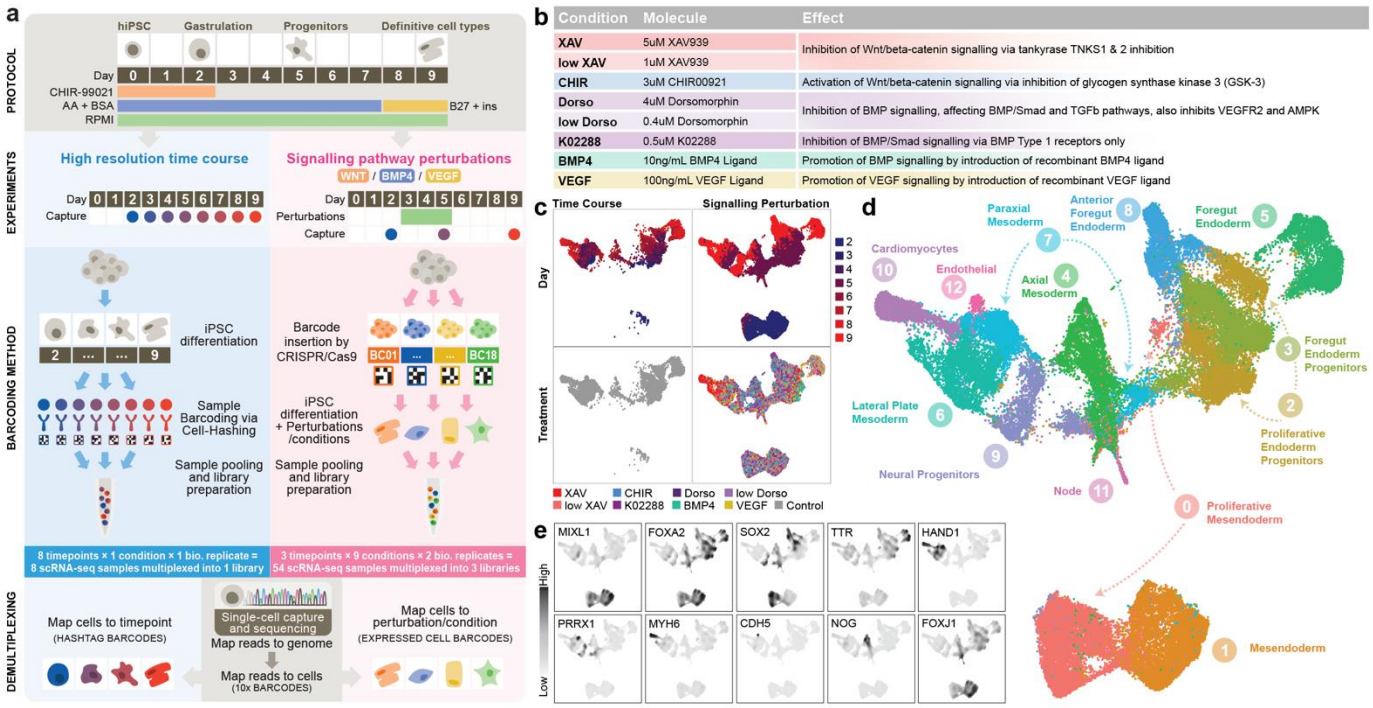

Figure 2: Generating an integrated multilineage atlas of *in vitro* iPSC differentiation.

a) Schematic of multiplexed study design including 1) a high-resolution time course dataset (blue, left) capturing cells every 24 hours between days 2 and 9 of differentiation and 2) a signalling pathway perturbation dataset (pink, right) generated via perturbation of developmental signalling pathways between days 3 and 5 of differentiation with cells captured on days 2, 5, and 9. In total, 62 cell samples were captured and sequenced using 4 single-cell libraries. Samples were demultiplexed by reference to RNA and sample barcodes.

b) Signalling perturbation conditions, concentrations, and mechanism of action.

c) UMAP plots of integrated datasets comprising 62,208 single cells, split by dataset of origin (columns), coloured by time point (top row, 13,682 cells total), and signalling perturbation treatment (bottom row, 48,526 cells total). See also **Table 1**.

d) UMAP showing the 13 cell type clusters identified using low resolution (0.3) Louvain clustering. Cell type annotations are defined based on expression of marker genes.

e) Nebulosa[65] plots showing expression distribution of cell type marker genes. *EOMES* marks mesodermal cells; *SOX2* marks early epiblast, neural progenitors, and foregut progenitors; *FOXA2* marks definitive endoderm; *TTR* marks primitive foregut progenitors; *HAND1* marks precardiac mesoderm; *MYH6* mark cardiomyocytes; *CDH5* marks endothelial cells and *PRRX1* marks limb bud mesoderm. See also **Figure S1f**.

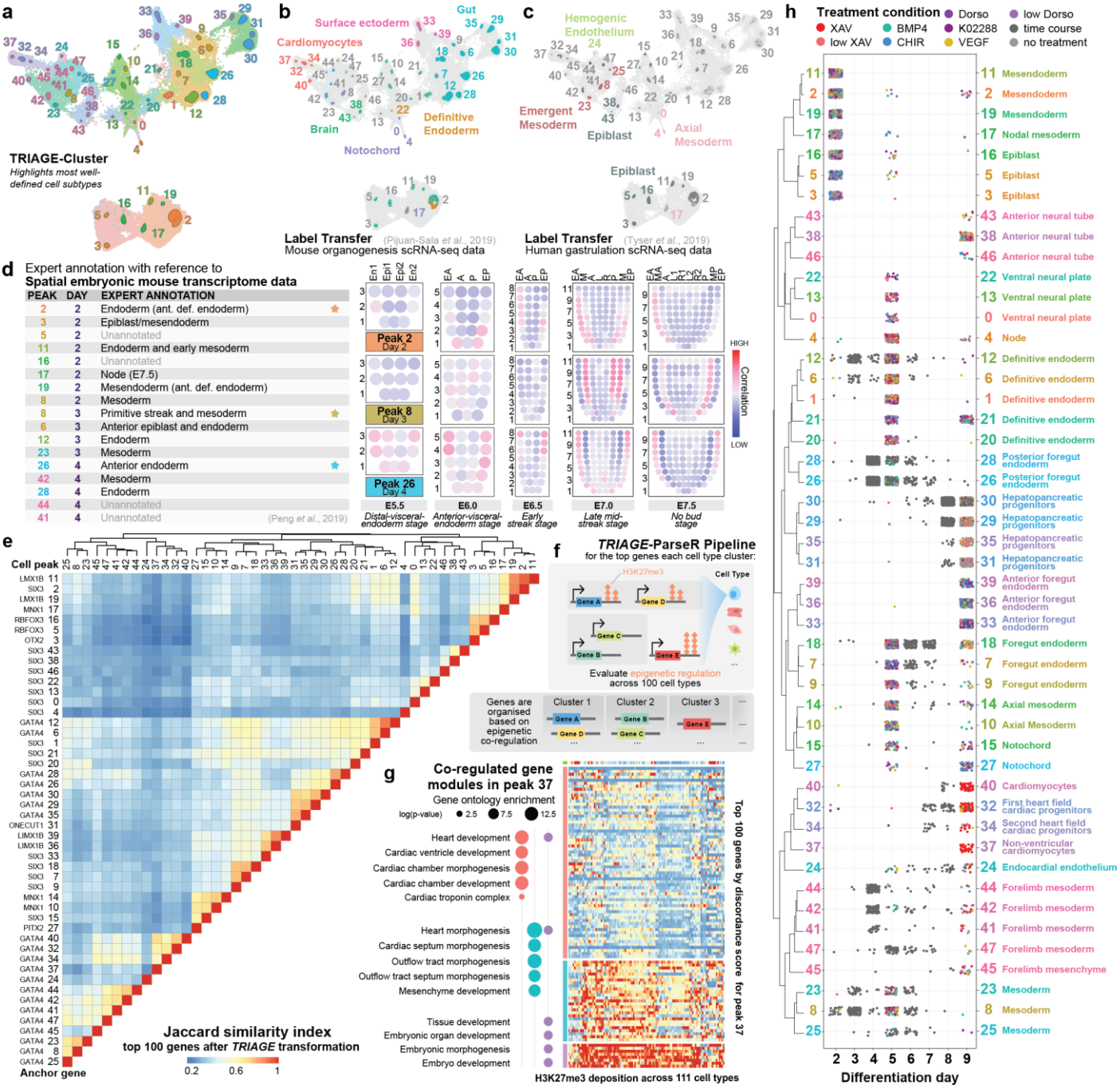

Figure 3: Characterisation and annotation of *in vitro*-derived cell types

a) UMAP showing cell type peaks identified from *TRIAGE-Cluster* compared to broad cell type clusters identified in **Figure 2d**.

b-c) Annotation of *TRIAGE* cell type peaks by reference to three orthogonal *in vivo* single-cell data sets including (**b**) annotated cell types in single-cell datasets of mouse organogenesis[2], and (**c**) the gastrulating human embryo[4]. Cells are coloured on a single-cell basis by their assigned label after thresholding (threshold = 0.5). Dark grey cells do not have a label above the threshold. Cell peak number labels are coloured by the label that makes up the greatest proportion of cells in that peak, where grey labels indicate cell type peaks where most cells do not score above the threshold. Light grey cells do not fall within a cell type peak. See also **Figure S3c**.

d) Annotation of cell peaks by reference to spatial gene expression domains in the developing mouse embryo at five different stages of development between embryonic days 5.5 and 7.5, spanning pre-gastrulation to late primitive streak

468 and amnion formation[27,66]. A Pearson correlation between the expression of selected genes[67] for each *TRIAGE* peak
469 and each embryonic spatial domain is visualised as corn plots. Annotation of the highly correlated domains for each
470 cell peak and timepoint are summarised in the table (left) and the starred rows' corresponding corn plots are displayed
471 (right). For each cell peak and timepoint, the five corn plots capture different stages of development in mouse
472 embryos, where each spot has captured the expression profiles of approximately 5-40 cells. Ant. def. endoderm:
473 Anterior definitive endoderm; En1 and En2: divided endoderm; Epi1 and Epi2: divided epiblast; EA: anterior
474 endoderm; A: anterior; P: posterior; EP: posterior endoderm; M: whole mesoderm; L: left lateral; R: right lateral; MA:
475 anterior mesoderm; L1: anterior left lateral; R1: anterior right lateral; L2: posterior left lateral; R2: posterior right
476 lateral; MP: posterior mesoderm. See also **Figure S2d**.

477 **e)** Heatmap and dendrogram showing pairwise Jaccard similarity score between each *TRIAGE-Cluster* cell peak using
478 the top 100 genes by *TRIAGE*-transformation. Anchor genes for each cell peak are listed.

479 **f)** Schematic describing *TRIAGE-ParseR* analysis pipeline[28] to identify gene clusters from each cell peak to reveal
480 functional and characteristic gene programs of cell subtype.

481 **g)** Example analysis of cardiomyocytes (peak 37) using *TRIAGE-ParseR.* Heatmap shows H3K27me3 deposition
482 across 111 cell and tissue types in the NIH Epigenome Roadmap data across the 100 top ranked genes for cell type
483 peak 37 after *TRIAGE*-transformation. The genes are grouped using the gene co-modulation analysis and GO term
484 enrichment (-log(p value)) of select terms for each of the main gene clusters are shown (left). See **Tables S5-7** for full
485 list of parsed genes and enriched ontologies for all cell type peaks.

486 **h)** Jitter plot showing every cell in the dataset based on the *TRIAGE-Cluster* peak, time point, and signalling condition
487 under-which each cell was derived. Final cell peak annotation is provided (right). Dendrogram (left) reflects pairwise
488 Jaccard similarity of the top 100 genes after *TRIAGE*-transformation[25] for each cell peak. See also **Figure S3e**.

489

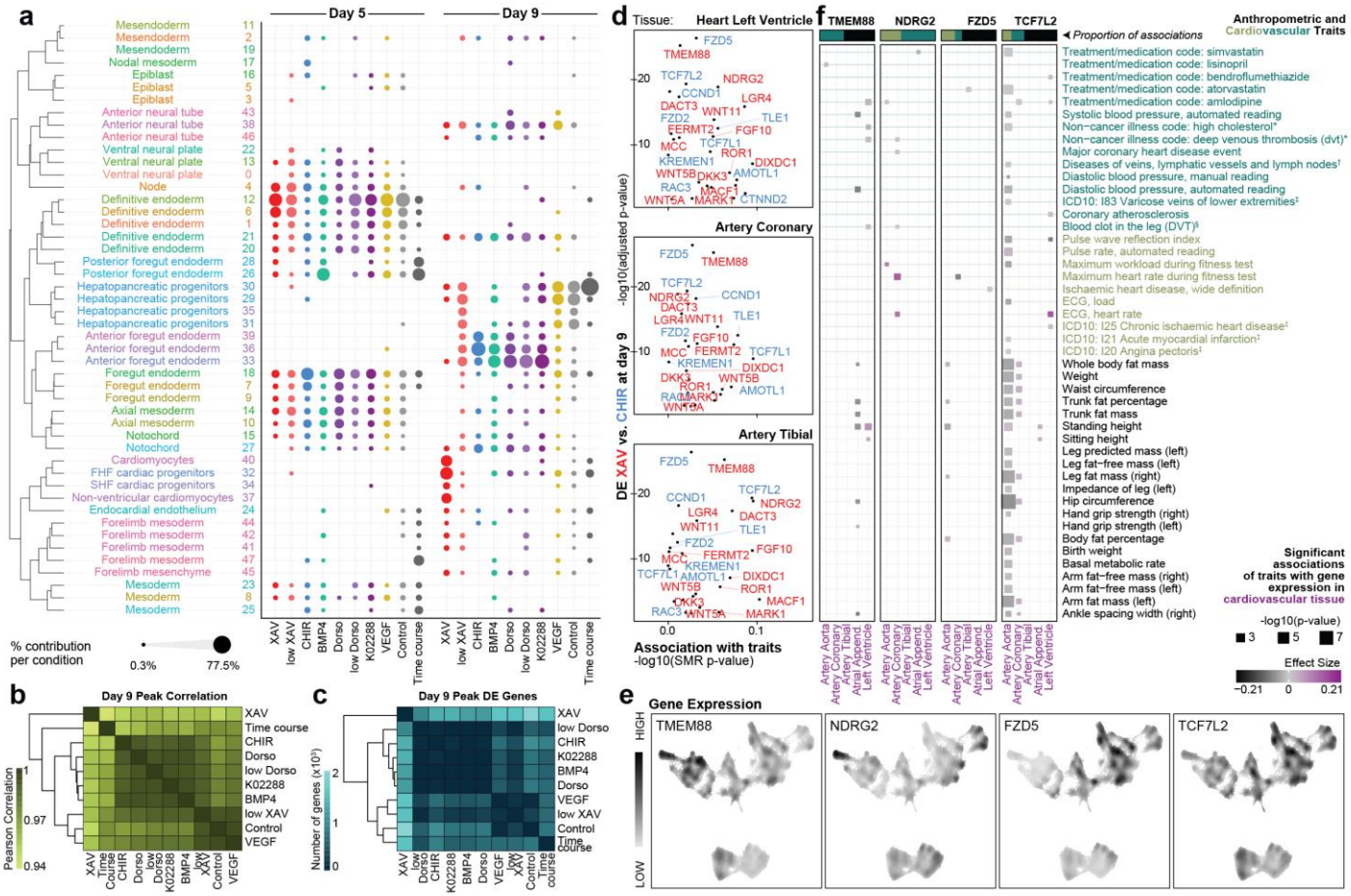

Figure 4: Signalling analysis of cell differentiation

(a) Proportion of cells from each treatment condition contributing to each cell type peak at days 5 and 9 of differentiation. Each column represents a treatment condition and sums to 100% across all cells produced by that treatment for the corresponding day of differentiation. FHF: First heart field; SHF: Second heart field.

(b-c) Heatmaps summarising the pairwise correlation **(b)** and number of differentially expressed genes **(c)** between cells from each treatment condition at day 9.

(d) WNT-related genes that are differentially expressed between XAV (red) and CHIR (blue) treated *TRIAGE-Cluster* peak cells at day 9 (y-axis) and the significance of the cardiovascular trait most strongly associated with their expression in three cardiovascular tissues, as determined by Summary data-based Mendelian Randomisation (SMR) analysis (x-axis).

(e) Nebulosa[65] plots showing expression pattern of four candidate WNT-related genes in the iPSC atlas of mesendoderm differentiation.

(f) SMR analysis evaluating associations of candidate WNT-related genes' expression in cardiovascular tissues with cardiovascular and anthropometric traits in the UK Biobank. Some trait and tissue names have been shortened from the following: *non-cancer illness code, self-reported: …; †Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified; ‡Diagnoses – main ICD10: …; §Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor: Blood clot in the leg (DVT); Atrial Append. : Atrial Appendage.

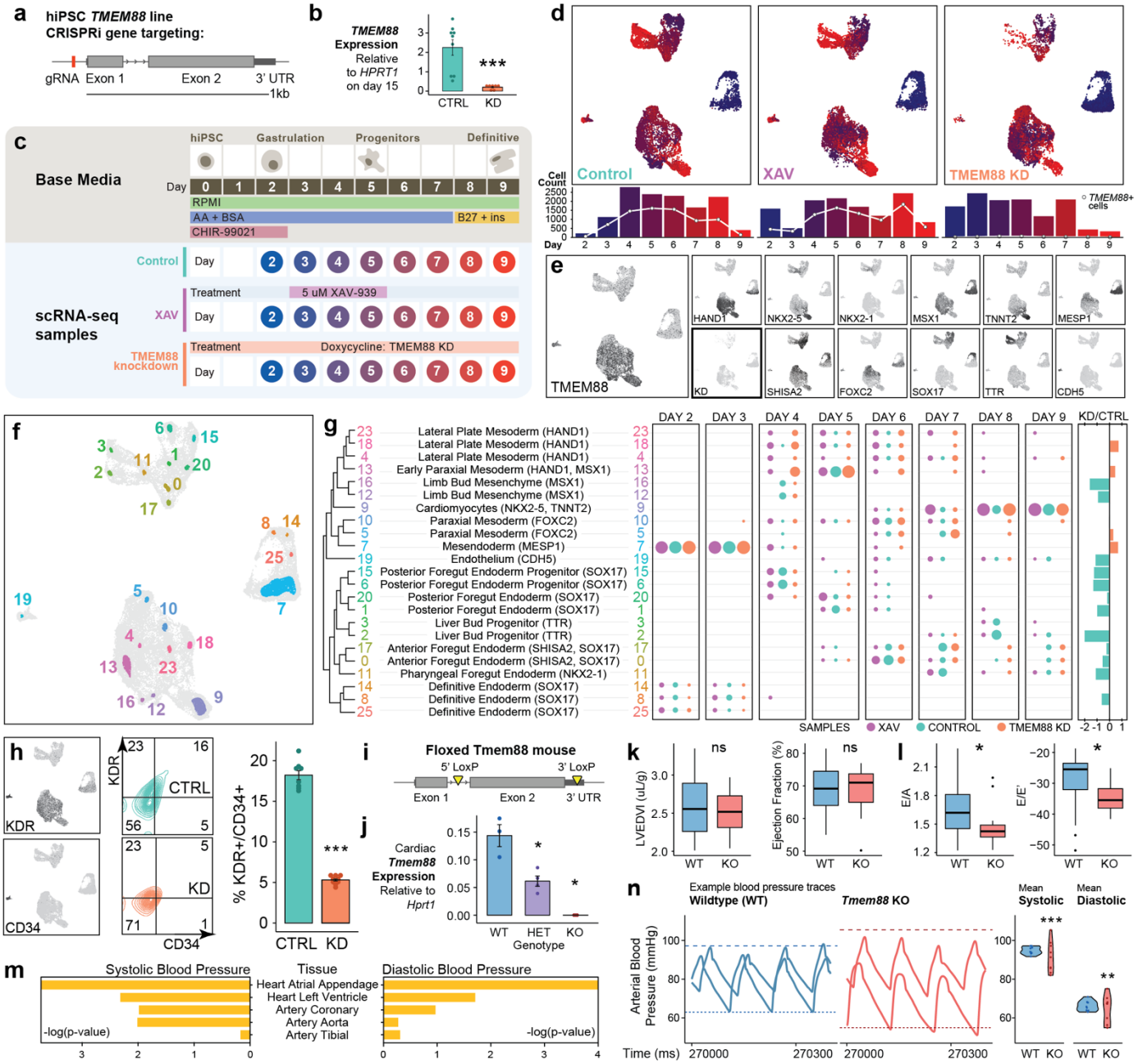

Figure 5: WNT-inhibitor *TMEM88* as a regulator of cardiovascular differentiation and physiology.

(a-b) CRISPRi targeting of *TMEM88* to generate doxycycline-inducible *TMEM88* KD hiPS cell line **(a)** and qPCR confirmation of *TMEM88* knockdown efficiency **(b)**, n=7-9, data are represented as mean ± SEM, ***p<0.001 by *t*-test.

(c) Experimental design of scRNA-seq experiment comparing *TMEM88* KD vs control and XAV-treated cells with cells captured every 24 hrs across day 2-9 of mesendoderm differentiation.

(d) Distribution of cells at each time point and treatment. Points in the UMAP are coloured by day of collection, and bar plots show number of cells collected at each time point. White points and lines graph overlaid on the bar plot show number of cells expressing *TMEM88* at each time point.

(e) Expression of marker genes. *TMEM88* (left) in the integrated data set and confirmed loss of *TMEM88* expression in the knockdown sample (KD, bottom row in bold). *HAND1* marks lateral plate mesoderm; *NKX2-5* marks cardiac mesoderm and pharyngeal foregut endoderm; *NKX2-1* marks pharyngeal foregut endoderm; *MSX1* marks paraxial

limb bud mesoderm; *TNNT2* marks cardiomyocytes; *MESP1* marks mesendoderm; *SHISA2* marks anterior foregut endoderm; *FOXC2* marks paraxial mesoderm; *SOX17* marks definitive endoderm and its primitive derivatives; *TTR* marks posterior foregut endoderm and *CDH5* marks endothelium. See also **Figure S3c**.

(f) UMAP showing the 23 cell type peaks identified by *TRIAGE-Cluster* after filtering out those with 20 cells or less.

(g) Final cell type annotations for cell types identified by *TRIAGE-Cluster* (left) and proportion of cells from each group allocated to each cell type over time (right). Dendrogram (left) reflects pairwise Jaccard similarity of the top 100 genes after *TRIAGE* transformation[25] for each cell peak. Far right panel quantifies gain or loss of cells in for each cell type peak for *TMEM88* KD vs control. Relevant marker genes are included in brackets next to the cell type annotation.

(h) Validation of depletion of endothelium from *TMEM88* knockdown *in vitro* by flow cytometry. Expression of endothelial markers KDR and CD34 in the scRNA-seq dataset (left) and bivariate contour plots showing representative flow cytometry analysis (middle) and quantification (right) of control vs *TMEM88* KD for KDR/CD34-labelled cells at day 5 of differentiation, represented as mean ± SEM. n = 8-9 per group. ***$p < 0.001$ by *t*-test.

(i-j) Design for CRISPR floxed knockout of *Tmem88* in mice (**i**) and validation of genotype-dependent expression of *Tmem88* in cardiac tissue, represented as mean ± SEM (**j**), n = 3-4 per group. *$p < 0.05$ by *t*-test.

(k-l) Echocardiographic analysis of heart function evaluating cardiac parameters by b-mode (**k, left**), m-mode (**k, right**) and pulsed wave doppler (**l**) in 8 week old *Tmem88* KO vs wildtype (WT) mice. Left ventricular end diastolic volume index (LVEDVI) normalised to body weight. n = 13-22, * $p < 0.05$ by *t*-test. In box plots, centre line shows median, box limits show upper and lower quartiles, whiskers show 1.5x interquartile range, and points show outliers.

(m) Summary-data based Mendelian Randomisation analysis showing significance of *TMEM88*-associated eQTLs from cardiovascular tissues on systolic and diastolic pressure traits.

(n) Analysis of blood pressure in the left carotid artery by conductance micromanometry. Raw pressure traces of two example animals per group are provided (left) and mean systolic and diastolic pressures are summarised for *Tmem88* KO versus wildtype (WT) mice. Dashed line indicates the highest mean systolic blood pressure and dotted line the lowest mean diastolic blood pressure per group. n = 9, ** $p<0.01$ and *** $p<0.001$ by Bartlett test for homogeneity of variances after Shapiro normality test.

547 **STAR METHODS**

548 **hiPSC culture and standard monolayer mesendoderm differentiation.** All human pluripotent stem cell studies
549 were carried out in accordance with consent from the University of Queensland's Institutional Human Research Ethics
550 approval (HREC#: 2015001434). Undifferentiated hiPSCs were cultured on Vitronectin XF (STEMCELL
551 Technologies #07180) coated plates in mTeSR1 media (STEMCELL Technologies #05850) with supplement at 37°C
552 with 5% CO_2. On the day prior to differentiation (day -1), cells were dissociated using 0.5mM EDTA solution and
553 seeded onto separate coated plates in pluripotency media with ROCK Inhibitor (STEMCELL Technologies #72308)
554 and cultured overnight. Once forming an ~80% confluent monolayer, differentiation was induced (day 0) by changing
555 the culture media to RPMI (ThermoFisher, #11845119) containing 3μM CHIR99021 (STEMCELL Technologies,
556 #72054), 500mg/mL BSA (Sigma #A9418), and 213mg/mL ascorbic acid (Sigma #A8960). On days 3 and 5, the
557 media was replaced with the same media cocktail excluding the CHIR99021 supplemental cytokine. On day 7 and
558 every subsequent other day, the cultures were fed with RPMI containing 1xB27 (ThermoFisher #17504001)
559 supplement plus insulin.

560 **Barcode cassette design.** The barcodes (**Table 1**) were introduced into the cells as a part of a barcode cassette, which
561 also incorporated the reverse complement of a partial Chromium Read2 adaptor sequence (truncated slightly to allow
562 oligo length of <= 60bp). This Read2 sequence is used as a PCR handle to generate barcode-containing amplicons
563 compatible with Chromium scRNA library preparation. Additionally, restriction enzyme recognition sequences were
564 added or regenerated to enable easy transfer of the cassette between different vectors. The exact structure of the
565 cassette is as follows: EcoRV site 3' 3bp – MluI site – partial 10x Read2 adaptor reverse complement – 15bp barcode
566 – MluI complementary sequence. Barcoding cassettes were ordered as complementary single-stranded oligos, which
567 could be annealed and ligated into a digested plasmid backbone.

568 **Vector design.** AAVS1-CAG-hrGFP (Addgene #52344) was used as the plasmid backbone. It contains hrGFP under
569 the control of the CAG promoter, and AAVS1 homology arms to allow precise integration by homologous
570 recombination of the donor sequence into the *AAVS1* locus of the genome when paired with the CRISPR/Cas9 system
571 using a well-described guide RNA[68]. The barcode cassette was introduced between the EcoRV and MluI sites of the
572 plasmid between hrGFP and the polyadenylation site, to enable expression as part of the hrGFP transcriptional unit
573 (**Supplementary File 1**).

574 **Generation of barcoding plasmids.** AAVS1-CAG-hrGFP (Addgene #52344) was digested by incubation with
575 EcoRV-HF (New England BioLabs; NEB #R3195S) followed by addition of MluI-HF (NEB #R3198S) and further
576 incubation. Successful digestion was confirmed by running a small amount on an agarose gel, and remainder was
577 purified using QIAquick PCR Purification Kit (QIAGEN #28104). Top and bottom strands of barcode oligos were
578 annealed by mixing 1uL each of 100 μM oligo in 1x T4 DNA ligase buffer in a volume of 10 μL, heating to 94°C for
579 2 min, then allowing to cool to 25°C at a rate of 1°C/s. Annealed oligos were further diluted 1 in 10 with nuclease-free
580 water. Ligation of barcode cassettes to vector was performed by combining 100ng digested plasmid, 4 μL diluted
581 annealed oligos, and 1 μL T4 DNA ligase (NEB #M0202S; 400U/μL) in 10μL total volume with 1x T4 DNA ligase
582 buffer. Reaction was incubated at 16°C for 16 hours. Separate reactions were performed for each barcode oligo. 3μL
583 of ligation reactions were added to 20 μL Stellar competent cells (Takara #636763) and heat shock transformation
584 performed at 42°C for 1 min in 1.5mL tubes. 350 μL SOC media was added for recovery, with shaking at 37°C for 1

585 hour. 100 µL was spread onto selective ampicillin-containing agar plates, which were incubated overnight at 37°C. 3
586 colonies were picked from each plate for screening with colony PCR for expected insert in 10 µL reactions using
587 MangoTaq (Bioline #BIO-21083). Colonies showing successful amplification of insert were grown overnight in 5 mL
588 LB broth containing ampicillin, and plasmid purified using QIAGEN Plasmid Miniprep kit. 100% sequence identity
589 was confirmed across the barcode insert by Sanger sequencing using a universal sequencing primer (5'-
590 ttttggcagagggaaaaaga-3'), performed by the Australian Genome Research Facility (AGRF). 50 mL cultures were
591 grown from glycerol stocks of plasmids with confirmed barcode sequence insert, and plasmid was purified using
592 Nucleobond Xtra Midi Kit (Macherey-Nagel #740410.50) to give endotoxin-free, high concentration stock for
593 transfection.

594 **Stable generation of WTC BC01-BC18 barcoded iPS cell lines.** For gene editing, WTC WT-11 hiPSCs (Gladstone
595 Institute of Cardiovascular Research, UCSF; Karyotype: 46, XY; RRID: CVCL_Y803, generated as previously
596 described[69,70]) cells were grown to about 50-80% confluency, dissociated using 1XTrypLE and 100-200K cells were
597 used for each 10 µl reaction of the Neon Transfection System. The transfection mixture included 0.5 µg of barcoding
598 plasmid DNA, 20 pmol AAVS1-taregting sgRNA (protospacer sequence: GGGGCCACTAGGGACAGGAT,
599 chemically synthesised by Agilent technology) and 20 pmol spCas9 protein (IDT). After electroporation with 1 pulse
600 of 1300 V for 30 ms, cells were seeded in mTeSR1 with ROCK Inhibitor (Y-27632; STEMCELL Technologies
601 #72308) and CloneR (STEMCELL Technologies #05888). Selection was performed with 1 µg/ml puromycin, and
602 purified cell lines were frozen down in CryoStor CS10 Cell Freezing Medium (STEMCELL Technologies #07930)
603 and stored in liquid nitrogen.

604 **Quality control of cell lines.** All cell lines underwent quality testing for correct genetic insertion, selection efficiency,
605 pluripotency, chromosomal abnormalities, and mycoplasma contamination. Genomic DNA from all cell lines was
606 extracted using QuickExtract DNA Extraction Solution (Epicentre #QE09050). Correct targeting of donor construct at
607 the *AAVS1* locus was confirmed by junction PCR using the following primer pair: AAVS1 F1: 5'-
608 ggttcggcttctggcgtgtgacc-3', AAVS1 R1: 5'-tcaagagtcacccagagacagtgac-3'. The purified PCR product was then Sanger
609 sequenced using a universal sequencing primer (5'-cccatatgtccttccgagtg-3') to validate correct barcode insertion in
610 each cell line. Flow cytometry was performed on live cells for endogenous GFP expression and after labelling for the
611 pluripotency marker SSEA3 (BD Biosciences #562706) and a corresponding isotype control. Karyotyping was carried
612 out as a professional service by Sullivan Nicolaides Pathology. iPSCs were grown in a 25 cm^2 flask to about 70-80%
613 confluency and sent for analysis, 15 cells were examined per culture.

614 **Single-cell RNA-sequencing of undifferentiated barcoded iPSCs.** All barcoded iPS cell lines were cultured in
615 parallel as described above, dissociated using 0.5mM EDTA and 600K cells from each cell line were combined. Prior
616 to this, four cell lines were additionally labelling with different TotalSeq-A cell hashing antibodies[13] according to the
617 manufacturer's protocol. The combined sample was transferred into 2% BSA (Sigma #A9418) in PBS, stained with
618 Propidium Iodide and 500K viable cells were sorted using a BD Influx™ Cell Sorter (BD Biosciences) with
619 FACSDiva software (BD Biosciences).

620 Single-cell RNA-seq libraries were generated using the 10X Genomics Chromium Single Cell 3' v2 protocol, with
621 minor modifications to the workflow, outlined by Stoeckius and Smibert to capture the fraction of droplets containing
622 the Hashtag Oligonucleotide (HTO)-derived cDNA (<180bp). HTO additive primers and Illumina TruSeq DNA

623 D7xx_s primer (containing i7 index) were ordered from IDT, and used according to the cell hashing protocol. Hashtag
624 libraries were quantified using the Agilent Bioanalyzer. Sequencing was performed using the Illumina Nextseq
625 instrument, with gene expression and HTO libraries pooled on a single flow cell using a ratio of 90:10. Raw
626 sequencing data were processed using the the 10X Genomics Cell Ranger pipeline to derive gene expression count
627 matrices.

628 HTO-tagged cells were identified and extracted from the fastq files using the CITE-seq-Count[71] program with default
629 parameters, to generate a count matrix of cells and their respective HTO expression values. This allowed the pooled
630 hashing cells to be independently deconvoluted. From the 3' gene expression data, barcoded cells were identified by
631 the expression of a barcode from the whitelist, which was included in the transcriptome reference with unique
632 identifiers (e.g., 'BC01').

633 Transcriptome-based quality control filtering was performed, keeping cells with library sizes between 7500 and
634 60,000 reads, between 2000 and 8000 distinct features, mitochondrial reads making up no greater than 20% of all
635 reads, and ribosomal reads mapped to between 22 and 45% of total reads. The *scds* R package[72] was also used to
636 remove doublets predicted based on transcriptomic features. For this pilot study, a simplistic demultiplexing method
637 for both the genomic barcodes and the cell hashing multiplexing was used where the highest barcode was taken to be
638 the sample identity for each cell, cells with no counts mapped to any barcode were assigned as 'Negative' and cells
639 with two or more barcodes expressing the same maximum count were determined to be 'Doublets' or 'Multiplets'.
640 The agreement between genomic and hashing barcode assignments used is the sum of the double positives and double
641 negatives divided by the total number of cells. Standard normalisation, scaling, and dimensionality reduction of this
642 dataset was done using the *Seurat* R package to generate the plots.

643 **Cell hashed high resolution time course of mesendoderm-directed differentiation.** The cell line used to generate
644 the time course and *TMEM88* scRNA-seq datasets were WTC CRISPRi TMEM88-g2.3 GCaMP hiPSCs (Karyotype:
645 46, XY; RRID: CVCL_VM38; generously provided by M. Mandegar and B. Conklin, Gladstone Institute, UCSF),
646 generated as previously described[73]. In brief, guide RNA targeting the transcriptional start site of *TMEM88* in the
647 human genome was cloned into the pQM-u6g-CNKD doxycycline-inducible construct as previously described[73]
648 (Forward oligo: 5'-TTG GAG AGC CGC ATT CCA GGA TTA-3'; Reverse oligo: 5'-AAA CTT ATC CTG GAA
649 TGC GGC TCT-3'). The construct was then transfected into WTC CRISPRi GCaMP hiPSCs using the GeneJuice
650 (Novagen), followed by selection using rounds of replating with 10ug/mL blasticidin treatment. Populations were
651 tested for knockdown efficiency of *TMEM88* (Forward primer 5'-CCT ACT GGT CAC CGG ATT CCT-3', Reverse
652 primer 5'- GAC GCC GAT AAA GGG CTC G-3'), normalised to housekeeping gene *HPRT1* by qPCR (see also
653 **Quantitative RT-PCR**) following continuous 1ug/mL doxycycline treatment from day 0 of differentiation.

654 For the time course experiment, cells were not exposed to doxycycline and treated as transcriptionally wildtype.
655 Mesendoderm-directed differentiation as described above was induced in separate plates for each collection timepoint.
656 Cells were dissociated using 0.5mM EDTA in 2.5% Trypsin (ThermoFisher, #15400054) and neutralised with 50%
657 foetal bovine serum (GE Healthcare Life Sciences, #SH30084.03) in DMEM/F12 media (Sigma #11320033). 1 x 106
658 cells from each sample were labelled with a different TotalSeq-A cell hashing antibody (BioLegend antibodies
659 A2051-8) as per the recommended protocol [13] and sorted for viability on a BD Influx Cell Sorter (BD Biosciences)
660 using propidium iodide. 5 x 105 live cells per time point were collected and pooled for Chromium Single Cell 3' V3

661 (10x Genomics) reactions following the manufacturer's protocol, targeting $2x10^4$ cells. Gene expression libraries were
662 sequenced on an Illumina NovaSeq 6000 and the cell hashing libraries were sequenced on an Illumina NextSeq
663 550Dx.

664 **Multiplexing of signalling perturbations during mesendoderm differentiation.** 18 barcoded iPS cell lines (WTC
665 BC01-BC18) were cultured in parallel and multiple staggered set-ups of mesendoderm-directed differentiation. Cell
666 lines were divided into 2 batches (BC01-BC09 and BC10-BC18) to allow for the capture of biological duplicates for
667 all 9 conditions. Cells were collected for scRNA-seq on day 2 as a reference prior to any perturbations, as well as on
668 days 5 and 9 of differentiation. In total, 3 sequencing libraries were generated, with each library consisting of all 18
669 cell lines. A combination of 2 different timepoints from the two batches was pooled in each library to allow for easy
670 detection and removal of potential batch effects during downstream analysis (**Table 1**).

671 **Signalling pathway perturbation experiment single-cell library preparation and sequencing.** Sample pools were
672 assessed for quality using a hemocytometer with Trypan Blue exclusion. Cell viability ranged from 82-88%; cell
673 concentration was between 1.3E+06 and 2.1.9E+06 cells/mL. Chromium Single Cell 3' v3 (10x Genomics) reactions
674 were performed for each sample according to the manufacturer's protocol, targeting 20,000 cells per reaction. 11
675 cycles of cDNA amplification were performed in a C1000 Touch thermocycler with Deep Well Reaction Module
676 (Bio-Rad). After clean-up of full-length amplified cDNA, 10µL was used for the construction of the gene expression
677 library according to the manufacturer's protocol, with 11 indexing PCR cycles. Additionally, 5µL of full-length
678 amplified cDNA was used to generate a barcoding library for each sample pool. Briefly, a first round of PCR was
679 performed to specifically amplify cDNA regions containing the barcode cassette and append partial P5 and P7
680 sequencing adaptors. Each reaction contained 1x KAPA HiFi HotStart ReadyMix and 300nM each barcode_amp_F
681 (5'-ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3') and barcode_amp_R (5'-CTA CAC GAC GCT CTT
682 CCG ATC T-3') primers in a final volume of 50µL. A 2-step PCR protocol was performed with annealing/extension at
683 71°C for 30s, for six cycles. After a 1.2X SPRI clean-up to remove primers, a second round of PCR was performed
684 with the entire volume of purified product from PCR1. Each reaction contained 1x KAPA HiFi HotStart ReadyMix,
685 500nM SI-PCR primer (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT
686 C-3'), and 5µL of a unique i7 indexed R primer from Chromium i7 Multiplex Kit (10x Genomics), in a total volume
687 of 50µL. PCR was performed as for the SI-PCR protocol in the Chromium gene expression library construction
688 workflow. Eight indexing PCR cycles were performed, for a total of 14 cycles over two rounds of PCR. Final
689 barcoding libraries were purified using 1X SPRI beads, and fragment size and library concentration verified along
690 with gene expression libraries using a BioAnalyzer DNA High Sensitivity Kit (Agilent). Final gene expression
691 libraries were 62-82nM, with average size 457-494bp. Barcoding libraries were 30-38nM, with average size 358-
692 362bp.

693 A single pool was prepared from the three gene expression and three barcoding libraries for sequencing. The samples
694 were pooled equimolar within each library type and combined so that the gene expression libraries together made up
695 90% of the pool, and the barcoding libraries 10%. Samples were sequenced on an Illumina NovaSeq 6000 (NovaSeq
696 control software v1.6.0/Real Time Analysis v3.4.4) using a NovaSeq 6000 S4 Kit V1 (200 cycles) (Illumina,
697 20027466) in standalone mode as follows: 28 bp (Read 1), 8 bp (I7 Index), and 91 bp (Read 2). Gene expression count
698 matrices were derived using the standard 10X cell ranger pipeline.

Demultiplexing and quality control. Barcoding reads used to assign sample barcodes to each cell were from the separately amplified and sequenced barcode libraries, and cell barcodes without both transcriptome and sample barcoding reads were removed. For each barcode sequencing library, the 'HTODemux' function in the *Seurat* R package[74] with the default 0.99 quantile cutoff was used to determine the dominant sample barcode for each cell and annotate negative and doublet cells based on their sample barcode reads alone. Three transcriptome-based doublet detection methods in the *scds* R package were used to further assign doublet annotations to each cell, and cells labelled as doublets by at least three methods were removed. Transcriptome-based cell filtering as part of the *Seurat* pipeline removed cells with fewer than 2000 and greater than 7500 detected genes; fewer than 5000 and greater than 50,000 total read counts; or mitochondrial reads accounting for greater than 25% of total reads. Following filtering, sample barcodes were assigned to the remaining cells based on the barcode with the highest expression in each cell.

Visualisation of individual scRNA-seq datasets. Normalisation and UMAP dimensionality reduction of the data were done following the standard pipeline in the *Seurat* pipeline. To visualise the expression of marker gene expression in the UMAP plots, we used the R package *Nebulosa*[65] which represents gene expression using kernel density estimation to account for overplotting and noise from expression drop-out.

Integration of single-cell datasets. Integration of the two scRNA-seq datasets was done using the RPCI method as part of the *RISC* R package[19], following the suggested pipeline to pre-process raw counts. We used only genes expressed in both datasets to generate 15 gene eigenvectors and using the signalling pathway perturbation dataset as the reference dataset, performed integration to return 50 principal components for further analysis. UMAP dimensionality reduction of the integrated data was also done using the *RISC* package, using 15 components of the 'PLS' embeddings as recommended for integrated values. Gene expression UMAP plots were generated using the *Nebulosa* R package. Broad cell type clustering was performed using the *Seurat* package with the resolution parameter set to 0.3.

Cell-cell interaction analysis. The *CellChat*[21] R package (v2.1.2) was used to infer, analyse, and visualise cell-cell communication networks. We utilized *CellChatDB* v2, which contains approximately 3,300 validated molecular interactions. In total, we have a combination of 10 treatment conditions ("XAV", "lowXAV", "CHIR", "BMP4", "Dorso", "lowDorso", "K02288", "VEGF", "no_treatment", and "no_treatment_tc") and three time points (days 2, 5, and 9), resulted in a total of 30 datasets. We applied the CellChat pipeline to each dataset. Briefly, we created the CellChat object based on the corresponding expression matrix and metadata for each dataset with the 'createCellChat' function. Communication probability was computed with the 'computeCommunProb' function, using the default 'triMean' parameter to approximate a 25% truncated mean. This ensures that the average gene expression is set to zero if less than 25% of cells in a group express the gene. The 'filterCommunication' function was then applied to filter cell-cell communications using the default setting, requiring a minimum of 10 cells in each group to infer communication. The inferred cellular communication networks were extracted at both the ligand-receptor and signaling pathway levels using the 'subsetCommunication' function. Aggregated cell-cell communication networks for each dataset were generated using the 'aggregateNet' function. We used CellChat plot functions to visualize three key signaling pathways - WNT, BMP, and VEGF - if they passed statistical significance, along with the top three pathways with the highest probability of cell-cell communication. Network centrality measures from weighted-directed networks were employed to identify the dominant senders, receivers, mediators, and influencers of

intercellular communication, using the 'netAnalysis_computeCentrality' and 'netAnalysis_signalingRole_network' functions. In addition, CellPhoneDB (v5.0.0) along with cellphonedb-data (v5.0), comprising a total of ~3,000 interactions, was used to infer cell-cell interactions for each of the 30 datasets (**Supplementary File 2**). We generated counts and metadata files from the Seurat object to serve as input for CellphoneDB. Following the guidelines provided in the CellPhoneDB Jupyter notebooks (https://github.com/ventolab/CellphoneDB/tree/master/notebooks), we employed the 'simple analysis' mode (method 1) to obtain the mean interaction values for each pair of cell clusters, and employed the 'statistical_analysis' mode (method 2) to evaluate the significance of the interactions present in the dataset. Circle plots, including those shown in **Figure S2e**, were generated using the 'netVisual_aggregate' function in CellChat with the parameter 'layout=circle'.

Gene regulatory network analysis using pySCENIC. To assess regulon activities within individual cells, we used the Python-based package pySCENIC (version 0.12.1)[22,23]. Unlike traditional methods that focus on individual gene expression, pySCENIC determines rank-based regulon activities by inferring co-expression between transcription factor-target modules and identifying enriched cis-regulatory motifs. The analysis began by identifying co-expression modules comprising transcription factors and their potential target genes. This step was performed using the GRNBoost2 algorithm and applied to the gene expression matrix with reference to the hg38 database within pySCENIC. Following the identification of co-expression modules, we performed cis-regulatory motif enrichment analysis to generate candidate transcription factor regulons using cisTarget with the motif collection database version 10 in pySCENIC. Finally, the activity levels of these candidate regulons were quantified at the single-cell level using AUCell. This approach evaluates whether the regulon gene set is significantly enriched at the top of the gene ranking for each cell, providing a measure of regulon activity across different cellular contexts.

Lineage inference analysis using URD. Differentiation trajectory reconstruction was using the URD method (version 1.1.1)[24] following the instructions outlined on the authors' GitHub (https://github.com/farrellja/URD) and utilising default parameters for pre-processing and normalisation of the raw count matrix. For calculation of the diffusion map to determine transition probabilities between cells, 116 nearest neighbours (knn) and a sigma of 20 were used. Next, pseudotime values were assigned to each cell, ordering them along the developmental process. For this, all cells collected from days 2 and 3 of differentiation were used as root cells and 50 flood simulations determined the visitation structure of the lineage specification graph. The URD tree was constructed using all day 9 cells, annotated by their whole-dataset cluster assignment (**Figure 2d**), as tip cells (terminal cell populations), with a p-value threshold of 0.001, 8 bins per pseudotime window, 25 cells per pseudotime bin, and "preference" as the divergence method.

***TRIAGE-Cluster* analysis.** To assign each cell a repressive tendency score (RTS), we identified the expressed gene with the highest *TRIAGE* RTS in each cell. The RPCI[19]-generated UMAP coordinates were input into the python *scipy.stats gaussian_kde* function, using the RTS vector as the "weights" parameter to estimate a probability density function of the repressive tendency across the data in UMAP space. Using a bandwidth of 0.25, a contour representation of the data was generated, splitting the density function into 10 levels. For each level, the spatial clustering algorithm *DBSCAN* from the python module *sklearn* was used to identify all spatially separated groups of cells. *TRIAGE* "peak" regions were found by identifying the *DBSCAN* clusters with no smaller clusters within them at higher contour levels. Code is available at https://github.com/palpant-comp/TRIAGE-Cluster.

774 **Cell type annotation with reference to *in vivo* datasets.** For gene set scoring of cell type domain genes from 921
775 mouse tissue RNA-seq samples[26], R package *biomaRt getLDS* function was used to map the mouse Ensembl gene IDs
776 to human Ensembl gene IDs. R package *topGO* was used to determine enriched GO terms for each gene set. For each
777 gene set, the *AddModuleScore* function from the *Seurat* package was used to assign a gene set score for each cell in
778 each cell type peak. After thresholding scores to be greater than 0, the gene set with the highest score was assigned as
779 the cell type domain annotation for that cell. For label transfer, both our query cell type peaks and the reference *in vivo*
780 datasets[2,4] were *TRIAGE*[25] transformed as it has been observed to result in higher prediction confidence in label
781 transfer analysis. The label transfer prediction scores were thresholded only to consider cells with scores over 0.5, and
782 the highest scoring cell type was used to annotate each cell. For both gene set scoring and label transfer annotations,
783 each cell peak's overall annotation is defined as the most common label after thresholding and is assigned as NA if
784 most cells do not have a score greater than the respective thresholds.

785 **Cell type annotation with reference to spatial domains of embryonic cell types in gastrulating mouse embryo.**
786 To annotate early *in vitro* cell types from days 2-4 in our dataset with reference to early *in vivo* cell types, we
787 compared we compared the expression profiles of each *TRIAGE* cell type peak to spatial mouse embryonic
788 transcriptome data that spans embryonic days 5.5 to 7.5[27]. For days 2-4 in our *in vitro* data, the top 50 cell type marker
789 genes were determined for each time point using Cepo, which identifies differentially stable genes between cell types
790 using stability metrics including proportion of zeros and coefficient of variation of a log-transformed normalised
791 expression matrix[67]. Cell type peaks with fewer than 20 cells and genes expressed in less than 5% of any peak were
792 excluded from further analysis. Next, we took the union of all cell type marker genes identified by Cepo and
793 calculated the Pearson correlation of their expression profiles with the spatial mouse transcriptome data. This is used
794 as a proxy to assess how well the expression profile of a cluster recapitulates specific cell types in the gastrulating
795 mouse embryo.

796 ***TRIAGE-ParseR* analysis.** We calculated the average gene expression value for each *TRIAGE-Cluster* cell type peak
797 and converted these values to the *TRIAGE* discordance score (DS) by multiplication of each gene to its repressive
798 tendency score[25]. We used the top 100 genes ranked by DS from each cell type peak to calculate the pairwise Jaccard
799 similarity index between all peaks. Cell peaks were clustered based on Euclidean distance with complete linkage using
800 the *pheatmap* R package. The top 100 DS genes for each cell peak were then parsed into functionally related modules
801 using gene co-modulation analysis[28]. In brief, principal component (PC) analysis was performed on H3K27me3
802 deposition breadth data across all genes and cell types in the NIH Roadmap Epigenomics dataset[75], where each PC
803 represents a biologically meaningful H3K27me3 deposition pattern. The top 67 PCs, which capture over 96% of
804 variance in the data, are used to inform probabilistic clustering of the input genes from each cell peak into co-
805 modulated sets using a Gaussian Mixture Model. The optimal number of clusters is determined by Bayesian
806 Information Criterion, and only clusters with significant ($p<0.001$, Fisher's Exact Test) protein-protein interactions as
807 determined by STRING analysis[76] were kept for further analysis. Code is available at https://github.com/palpant-
808 comp/TRIAGE-ParseR.

809 **Differential gene expression, KEGG and GO enrichment analysis.** For differential gene expression analysis, we
810 utilised the 'FindMarkerGenes' function from the *Seurat* R package. Heatmap visualisation displaying number of
811 differentially expressed genes between treatments was generated using the *pheatmap* R package. WNT-related GO

term associations were identified using the *biomaRt* R package, and DE genes were annotated by their most specific directional GO term. GO term and KEGG pathway enrichment analysis for the differentially expressed genes compared between the XAV and CHIR treatments at day 9 was performed using the 'enrichGO' and 'enrichKEGG' functions in the *clusterProfiler* R package. WNT -related genes in Figure 4d are defined as those associated with the following GO terms: Wnt signalling pathway (GO:0016055), Positive regulation of Wnt signalling pathway (GO:0030177), Negative regulation of Wnt signalling pathway (GO:0030178), Positive regulation of canonical Wnt signalling pathway (GO:0090263), Negative regulation of canonical Wnt signalling pathway (GO:0090090), Positive regulation of non-canonical Wnt signalling pathway (GO:2000052), and Negative regulation of non-canonical Wnt signalling pathway (GO:2000051).

Generation of *TMEM88* KD scRNA-seq dataset. The WTC CRISPRi TMEM88-g2.3 cells in each sample were cultured in 8 temporally staggered setups of mesendoderm-directed differentiation. The 'XAV' group was treated with 5uM XAV-393 for 48 hours, from day 3 to day 5 of differentiation, and the 'TMEM88 KD' group was treated with 1ug/mL doxycycline for the entire duration of differentiation starting from day 0. Cells were dissociated using 0.5% EDTA in 2.5% Trypsin (ThermoFisher, #15400054) and neutralised with 50% foetal bovine serum (GE Healthcare Life Sciences, #SH30084.03) in DMEM/F12 media (Sigma #11320033). 1 x 10^6 cells from each sample were labelled with a different TotalSeq-A cell hashing antibody (BioLegend antibodies A2051-8) as per the recommended protocol [13] and sorted for viability on a BD Influx Cell Sorter (BD Biosciences) using propidium iodide. 5 x 10^5 live cells per time point were collected and pooled for Chromium Single Cell 3' V3 (10x Genomics) reactions following the manufacturer's protocol, targeting 2 x 10^4 cells. Gene expression libraries were sequenced on an Illumina NovaSeq 6000 and the cell hashing libraries were sequenced on an Illumina NextSeq 550Dx.

Quality control processing of *TMEM88* scRNA-seq dataset. This dataset was pre-processed in the same manner as the time course dataset; the cellranger pipeline was used to process raw reads, cell hashing barcodes (time points) were demultiplexed with the 'HTODemux' function from the *Seurat* package, and doublet detection was performed with the *scds* package, removing cell barcodes described as doublets by at least 3 methods. Transcriptome-based cell filtering as part of the *Seurat* pipeline removed cells with fewer than 1000 and greater than 10,000 total read counts; fewer than 100 and greater than 300 detected genes; mitochondrial reads comprising greater than 15% of the library; or ribosomal reads making up greater than 45% of the library. Following filtering, time points were assigned to the remaining cells based on their cell hashing barcode with the highest expression in each cell.

Quantitative RT-PCR. Cells were lysed using β-mercaptoethanol in RLT buffer (QIAGEN) and snap frozen on dry ice. Total RNA was collected with the RNeasy Mini Kit (QIAGEN) according to manufacturer's instructions. RNA was reverse transcribed into cDNA using Superscript III Reverse Transcriptase (Thermo Fisher Scientific). qRT-PCR was performed using SYBR Green PCR Master Mix reagents on a ViiA 7 Real-Time PCR System (Thermo Fisher Scientific). Gene expression was analysed in GraphPad Prism or R, where relative gene expression levels were calculated for each gene by raising two to the power of the measured gene (averaged over two technical replicates), divided by the mean control housekeeping gene expression. For human samples, *HPRT1* was used as the control (Forward primer 5'-TGA CAC TGG CAA AAC AAT GCA-3', Reverse primer 5'-GGT CCT TTT CAC CAG CAA GCT-3'), and *Hprt1* was used for mouse samples (Forward primer 5'-TCA GTC AAC GGG GGA CAT AAA-3', Reverse primer 5'-GGG GCT GTA CTG CTT AAC CAG-3'). Data were analysed in R, where outlier values (those

850 exceeding mean ± 2 standard deviations) in each group were removed. The statistical test used to compare the
851 normalised expression means from separate experiments was a two-tailed *t*-test with a 95% confidence interval using
852 R.

853 **Flow cytometry.** Cells were dissociated with 0.5mM % EDTA + 0.25% Trypsin (1:10) then neutralised with foetal
854 bovine serum (FBS) in DMEM/F12 or RPMI media (1:1). Cells were fixed using 4% paraformaldehyde,
855 permeabilised in 0.75% saponin with 5% FBS in PBS and labelled for flow cytometry using either cardiac troponin T
856 (Creative Biolabs, #M40283), KDR (R&D Systems, #FAB357P), or CD34 (BD Biosciences, #340430) with the
857 corresponding immunoglobin G isotype control. Flow cytometry data was captured using a BD FACSCANTOII with
858 the FACDiva software. Analysis of this data was performed using *FlowJo* (v10.07). Summaries of the cell populations
859 were analysed and visualised in R, where outlier values (exceeding mean ± 2 standard deviations) in each group were
860 removed before performing a two-tailed *t*-test for the comparison of means with a 95% confidence interval.

861 **Summary-data based Mendelian Randomisation (SMR) analysis.** We used 61 anthropometric and cardiovascular
862 phenotypes [32] from the UK Biobank[77], whose heritability estimates were given as "medium" or "high" by Neale's lab,
863 to explore the broad effect of select genes on human traits. Publicly available summary statistics generated by Neale's
864 lab were used to perform Summary-data Mendelian Randomisation (SMR) analysis using eQTL (expression
865 quantitative trail loci) information derived in the V8 GTEX release[78]. Individuals with European ancestry from the
866 1000 genomes project were used as a LD (linkage disequilibrium) reference. SMR analysis was performed with a p-
867 value threshold for eQTL inclusion of 0.05 to ensure all eQTLs affecting gene expression were captured. SMR
868 analysis across all phenotypes was performed on the vascular (coronary artery, tibial artery, aortic artery), cardiac (left
869 ventricle, atrial appendage), respiratory (lung), and gut tube derivative (gastroesophageal junction, oesophageal
870 mucosa, pancreas, stomach, and thyroid) tissues. Further analysis for *TMEM88* was performed on systolic blood
871 pressure (4080) and diastolic blood pressure (4079) phenotypes using all 49 GTEX tissues.

872 **Generation of *Tmem88* conditional knockout mouse line.** The floxed *Tmem88* mice were generated by the Monash
873 Genome Modification Platform, Monash University (see acknowledgements). In brief, CRISPR RNAs were ordered
874 from Integrated DNA Technologies and annealed with trans-activating CRISPR RNA (Alt-R CRISPR-Cas9 tracrRNA
875 from IDT #107319 to form single guide RNA (gRNA) targeting Exon 2 of Tmem88 (gRNA1: 5'-AAT GGG CCA
876 CTC TGC CCA GG-3', gRNA2: 5'-TTT GCC CAT TTG TAG TCT TG-3'). 10ng/uL Cas9 nuclease (IDT Alt-R®
877 S.p. Cas9 nuclease) was incubated with 10ng/uL gRNAs to form a ribonucleoprotein (RNP) complex. A 1247bp
878 ssDNA repair template containing LoxP sites flanking the second exon of *Tmem88* (see also **Figure 5i**) was generated
879 using Invitrogen Dynabeads™ MyOne™ Streptavidin C1 according to manufacturer's instructions. The RNP and
880 ssDNA repair template (10ng/uL) were microinjected into the pronucleus of C57BL/6J zygotes at the pronuclei stage.
881 Microinjected zygotes were transferred into the uterus of pseudo pregnant F1 females.

882 All mouse breeding was conducted in accordance with consent from the University of Queensland's Molecular
883 Biosciences Animal Ethics committee approval (Molecular Biosciences AEC-MBS approval #: 2018/AE000177 &
884 2021/AE000421). Floxed *Tmem88* animals were crossed with C57BL/6 mice expressing Cre recombinase driven by a
885 CMV promoter, to generate animals with ubiquitous loss of the *Tmem88* gene. Both hetero- and homozygous mutants
886 were viable and showed no obvious phenotype. For genotyping the animals, DNA was extracted from ear or toe
887 samples using 50uL of QuickExtract (Biosearch Technologies, #QE09050), incubated at 65°C for 15 minutes, then

888 100°C for 5 mins and diluted to 50ng/uL for genotyping. Each mouse was tested for genomic presence of Cre-
889 Recombinase (Forward primer 5'-CTG ACC GTA CAC CAA AAT TTG CCT G-3', Reverse primer 5'-GAT AAT
890 CGC GAA CAT CTT CAG GTT C-3'; Cre = 200bp), Floxed *Tmem88* (Forward primer 5'-CGG CTC AGA ATC
891 TGC TGG TG-3', Reverse Primer 5'-TTT CTA GCG GGA AGC GGT GT-3'; Wt = 1300bp, Flox = 1350bp, KO =
892 600bp), and both LoxP sites (5'LoxP Forward primer 5'-CGG CTC AGA ATC TGC TGG TG -3', 5'LoxP Reverse
893 primer 5'-GCA GAT GCG AGG TGC AAA GG-3', 3'LoxP Forward primer 5'-TCC AAT CGC TCC TGC ACT GG-
894 3', 3'LoxP Reverse primer 5'-TTT CTA GCG GGA AGC GGT GT-3'; Wt = 300bp, LoxP = 350bp). Quantitative
895 RT-PCR was also used to confirm the loss of Tmem88 expression in the mutants (Forward primer 5'-CTG GTG GCT
896 GTT TTC AAT CTC C-3', Reverse primer 5'-GTG CCT GAG AGC GCA GAA A-3'). For RNA extraction, 30mg
897 heart tissue was homogenised in liquid nitrogen using a mortar and pestle before proceeding with the RNeasy Mini Kit
898 (QIAGEN #74106) following the manufacturer's instructions. cDNA synthesis and quantitative RT-PCR were
899 performed as described above.

900 **Mouse echocardiography.** All mouse experiments were carried out in accordance with consent from the University
901 of Queensland's Molecular Biosciences Animal Ethics committee approval (Molecular Biosciences AEC-MBS
902 approval #: 2018/AE000171 & 2021/AE000999). Blinded echocardiography was performed on 8 week old mice using
903 a Vevo 3100 Preclinial Imaging System (VisualSonics, Toronto, Canada) with a 25-55 MHz linear transducer
904 (MX550D). Mice were anaesthetised using 4% isofluorane, and general anaesthesia was maintained with 1%
905 isofluorane during the procedure. Mice were weighed and placed in supine position on a heating pad during
906 echocardiography with body temperature, heart rate, and electrocardiography recorded throughout. Warmed echo
907 transmission gel was applied to the hairless chest of the mouse before collecting images. Two-dimensional B-mode
908 images recorded in the parasternal long axis view were used to determine left ventricular end-diastolic volume
909 (LVEDV). One-dimensional M-mode images in the parasternal short-axis view were also used to measure ejection
910 fraction (EF) and thicknesses of the posterior walls of the left ventricle in diastole (LVPWd). Pulsed-wave doppler
911 recordings in apical four-chamber view were used to measure mitral inflow velocity of blood during passive filling of
912 the ventricle (E-wave) and active filling during atrial systole (A-wave). Pulsed-wave tissue doppler in apical four-
913 chamber view was used to capture mitral annular velocity during early (E') and late (A') diastole. Analysis of
914 echocardiography images was performed in VevoLab (version 3.1.1) software (VisualSonics, Toronto, Canada).
915 LVEDV, and LVPWd were normalised to body weight to achieve left ventricle end diastolic index (LVEDI/LVESI),
916 and LVPWd index, respectively. For each animal, all parameters were measured on average 4 times (minimum twice)
917 and averages are presented. Animals with abnormal body weights and heart rates were excluded (4 animals removed),
918 as well as all measurements from images with consistent outlier measurements indicating incorrect image capture (12
919 images removed). Statistical analysis was carried out using R with the *ggpubr* package to perform a two-sided *t*-test
920 with a 95% confidence interval.

921 **Mouse blood pressure measurements.** Adult mice between 8 and 28 weeks of age were anaesthetised using 4%
922 isofluorane before endotracheal intubation and mechanical ventilation. Isofluorane was reduced to 1.5% to maintain
923 general anaesthesia for the remainder of the procedure. A Millar Pressure-Volume (PV) catheter (Millar #SPR-839,
924 1.4F Catheter) was used to cannulate the right carotid artery, to allow a stable pressure trace as measured by the
925 MVPS Ultra Pressure-Volume Unit (Millar, #880-0168). The continuous baseline blood pressure was measured for 5-

926 10 minutes and recorded using the LabChart software (ADInstruments, Dunedin, New Zealand). The average
927 maximum (systolic) and minimum (diastolic) blood pressure across the pressure cycles in the measured time frame are
928 presented.

929 **Micro-computed tomography and automated image analysis (LAMA).** Embryos were dissected at embryonic day
930 (E)15.5 in PBS, rinsed in cold water and stained in 0.7% phosphotungstic acid diluted in 70% ethanol at RT for 14
931 days[79]. Stained embryos were mounted in agarose and scanned on a Skyscan 1272 micro-CT scanner (Bruker) using
932 an Al 0.5 + Cu 0.038 filter and voltage of 87 kV. Images were captured at 1640 × 2452 resolution (6 μm^3 voxel size)
933 through 360° at 0.4° angle increments and reconstructed with smoothing, ring artefact reduction and beam hardening
934 options in NRecon software (Bruker). 3D datasets were downsampled to 14 μm^3 and each registered using the LAMA
935 phenotyping pipeline[80] to an E15.5 population average so that its 43-organ atlas (in preparation) maps onto each
936 embryo dataset. The resulting organ and whole embryo volumes (WEV) were fitted to a linear model organ
937 volume/WEV~genotype+WEV with heterozygous and homozygous *Tmem88* embryos statistically compared to
938 wildtype littermate controls. P-values were corrected for multiple comparisons using the LAMA permutation statistics
939 script.

940

941 **MATERIALS AVAILABILITY**

942 Details for all resources presented in this study are available at our web accessible dashboard
943 (http://cellfateexplorer.d24h.hk/) and will be made available upon request.

944

945 **DATA AND CODE AVAILABILITY**

946 All three single-cell RNA sequencing raw datasets generated in this study are available upon request. Additionally, all
947 figures, data, and an interactive version of the combined, annotated dataset are available at
948 http://cellfateexplorer.d24h.hk/ , where additional co-modulation analysis outputs, metadata, and gene expression for
949 individual cell type peaks can be queried. *TRIAGE-Cluster* and *TRIAGE-ParseR* scripts are available at
950 https://github.com/palpant-comp/TRIAGE-Cluster and https://github.com/palpant-comp/TRIAGE-ParseR. Raw and
951 processed sequencing data will be made available on the Gene Expression Omnibus platform.

952

REFERENCES

1. Wu, Z., Shen, S., Sun, Y., Werner, T., Bradford, S.T., and Palpant, N.J. (2022). Analysing Genetic Programs of Cell Differentiation to Study Cardiac Cell Diversification. In Advanced Technologies in Cardiovascular Bioengineering, J. Zhang, and V. Serpooshan, eds. (Springer International Publishing), pp. 125-167. 10.1007/978-3-030-86140-7_8.
2. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490-495. 10.1038/s41586-019-0933-9.
3. Tabula Sapiens Consortium, Jones Robert, C., Karkanias, J., Krasnow Mark, A., Pisco Angela, O., Quake Stephen, R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. 10.1126/science.abl4896.
4. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285-289. 10.1038/s41586-021-04158-y.
5. Volpato, V., Smith, J., Sandor, C., Ried, J.S., Baud, A., Handel, A., Newey, S.E., Wessely, F., Attar, M., Whiteley, E., et al. (2018). Reproducibility of Molecular Phenotypes after Long-Term Differentiation to Human iPSC-Derived Neurons: A Multi-Site Omics Study. Stem Cell Reports *11*, 897-911. 10.1016/j.stemcr.2018.08.013.
6. Jiang, C.L., Goyal, Y., Jain, N., Wang, Q., Truitt, R.E., Coté, A.J., Emert, B., Mellis, I.A., Kiani, K., Yang, W., et al. (2022). Cell type determination for cardiac differentiation occurs soon after seeding of human-induced pluripotent stem cells. Genome Biology *23*, 90. 10.1186/s13059-022-02654-6.
7. Mellough, C.B., Collin, J., Queen, R., Hilgen, G., Dorgau, B., Zerti, D., Felemban, M., White, K., Sernagor, E., and Lako, M. (2019). Systematic Comparison of Retinal Organoid Differentiation from Human Pluripotent Stem Cells Reveals Stage Specific, Cell Line, and Methodological Differences. Stem Cells Transl Med *8*, 694-706. 10.1002/sctm.18-0267.
8. Gomez-Garcia, M.J., Quesnel, E., Al-attar, R., Laskary, A.R., and Laflamme, M.A. (2021). Maturation of human pluripotent stem cell derived cardiomyocytes in vitro and in vivo. Seminars in Cell & Developmental Biology *118*, 163-171. https://doi.org/10.1016/j.semcdb.2021.05.022.
9. Subramanian, A., Sidhom, E.-H., Emani, M., Vernon, K., Sahakian, N., Zhou, Y., Kost-Alimova, M., Slyper, M., Waldman, J., Dionne, D., et al. (2019). Single cell census of human kidney organoids shows reproducibility and diminished off-target cells after transplantation. Nature Communications *10*, 5462. 10.1038/s41467-019-13382-0.
10. Friedman, C.E., Nguyen, Q., Lukowski, S.W., Helfer, A., Chiu, H.S., Miklas, J., Levy, S., Suo, S., Han, J.-D.J., Osteil, P., et al. (2018). Single-Cell Transcriptomic Analysis of Cardiac Differentiation from Human PSCs Reveals HOPX-Dependent Cardiomyocyte Maturation. Cell Stem Cell *23*, 586-598.e588. https://doi.org/10.1016/j.stem.2018.09.009.
11. Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B., et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nature Biotechnology *27*, 851-857. 10.1038/nbt.1562.
12. Oceguera-Yanez, F., Kim, S.-I., Matsumoto, T., Tan, G.W., Xiang, L., Hatani, T., Kondo, T., Ikeya, M., Yoshida, Y., Inoue, H., and Woltjen, K. (2016). Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives. Methods *101*, 43-55. https://doi.org/10.1016/j.ymeth.2015.12.012.
13. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biology *19*, 224. 10.1186/s13059-018-1603-1.
14. Burridge, P.W., Keller, G., Gold, J.D., and Wu, J.C. (2012). Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. Cell Stem Cell *10*, 16-28. 10.1016/j.stem.2011.12.013.
15. Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L.B., Azarin, S.M., Raval, K.K., Zhang, J., Kamp, T.J., and Palecek, S.P. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. Proceedings of the National Academy of Sciences *109*, E1848. 10.1073/pnas.1200250109.
16. McMahon, A.P., and Moon, R.T. (1989). Ectopic expression of the proto-oncogene int-1 in Xenopus embryos leads to duplication of the embryonic axis. Cell *58*, 1075-1084. 10.1016/0092-8674(89)90506-0.
17. Xu, P.-F., Houssin, N., Ferri-Lagneau, K.F., Thisse, B., and Thisse, C. (2014). Construction of a Vertebrate Embryo from Two Opposing Morphogen Gradients. Science *344*, 87-89. 10.1126/science.1248252.

18. Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K.S., Powell-Braxton, L., Hillan, K.J., and Moore, M.W. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. Nature *380*, 439-442. 10.1038/380439a0.
19. Liu, Y., Wang, T., Zhou, B., and Zheng, D. (2021). Robust integration of multiple single-cell RNA sequencing datasets using a single reference space. Nature Biotechnology. 10.1038/s41587-021-00859-x.
20. Qiu, C., Martin, B.K., Welsh, I.C., Daza, R.M., Le, T.-M., Huang, X., Nichols, E.K., Taylor, M.L., Fulton, O., O'Day, D.R., et al. (2024). A single-cell time-lapse of mouse prenatal development from gastrula to birth. Nature *626*, 1084-1093. 10.1038/s41586-024-07069-w.
21. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nature Communications *12*, 1088. 10.1038/s41467-021-21246-9.
22. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nature Methods *14*, 1083-1086. 10.1038/nmeth.4463.
23. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nature Protocols *15*, 2247-2276. 10.1038/s41596-020-0336-2.
24. Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*. 10.1126/science.aar3131.
25. Shim, W.J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., et al. (2020). Conserved Epigenetic Regulatory Logic Infers Genes Governing Cell Identity. Cell Systems. 10.1016/j.cels.2020.11.001.
26. Hutchins, A.P., Yang, Z., Li, Y., He, F., Fu, X., Wang, X., Li, D., Liu, K., He, J., Wang, Y., et al. (2017). Models of global gene expression define major domains of cell type and tissue identity. Nucleic Acids Research *45*, 2354-2367. 10.1093/nar/gkx054.
27. Peng, G., Suo, S., Cui, G., Yu, F., Wang, R., Chen, J., Chen, S., Liu, Z., Chen, G., Qian, Y., et al. (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. Nature *572*, 528-532. 10.1038/s41586-019-1469-8.
28. Sun, Y., Shim, W.J., Shen, S., Sinniah, E., Pham, D., Su, Z., Mizikovsky, D., White, M.D., Ho, Joshua W.K., Nguyen, Q., et al. (2023). Inferring cell diversity in single cell data using consortium-scale epigenetic data as a biological anchor for cell identity. Nucleic Acids Research, gkad307. 10.1093/nar/gkad307.
29. Ng, W.H., Varghese, B., Jia, H., and Ren, X. (2023). Alliance of Heart and Endoderm: Multilineage Organoids to Model Co-development. Circulation Research *132*, 511-518. 10.1161/CIRCRESAHA.122.321769.
30. Loh, Kyle M., Ang, Lay T., Zhang, J., Kumar, V., Ang, J., Auyeong, Jun Q., Lee, Kian L., Choo, Siew H., Lim, Christina Y.Y., Nichane, M., et al. (2014). Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically Directing Signals Controlling Lineage Bifurcations. Cell Stem Cell *14*, 237-252. 10.1016/j.stem.2013.12.007.
31. Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., and Murry, C.E. (2007). Biphasic role for Wnt/β-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. Proceedings of the National Academy of Sciences *104*, 9685. 10.1073/pnas.0702859104.
32. Mizikovsky, D., Naval Sanchez, M., Nefzger, C.M., Cuellar Partida, G., and Palpant, N.J. (2022). Organization of gene programs revealed by unsupervised analysis of diverse gene-trait associations. Nucleic Acids Res *50*, e87. 10.1093/nar/gkac413.
33. del Bosque-Plata, L., Martínez-Martínez, E., Espinoza-Camacho, M.Á., and Gragnoli, C. (2021). The Role of TCF7L2 in Type 2 Diabetes. Diabetes *70*, 1220-1228. 10.2337/db20-0573.
34. Li, T., Chan, R.W.S., Lee, C.L., Chiu, P.C.N., Li, R.H.W., Ng, E.H.Y., and Yeung, W.S.B. (2022). WNT5A Interacts With FZD5 and LRP5 to Regulate Proliferation and Self-Renewal of Endometrial Mesenchymal Stem-Like Cells. Front Cell Dev Biol *10*, 837827. 10.3389/fcell.2022.837827.
35. Poulain, M., and Ober, E.A. (2011). Interplay between Wnt2 and Wnt2bb controls multiple steps of early foregut-derived organ development. Development *138*, 3557-3568. 10.1242/dev.055921.
36. Chodelkova, O., Masek, J., Korinek, V., Kozmik, Z., and Machon, O. (2018). Tcf7L2 is essential for neurogenesis in the developing mouse neocortex. Neural Development *13*, 8. 10.1186/s13064-018-0107-8.
37. Miao, D., Ren, J., Jia, Y., Jia, Y., Li, Y., Huang, H., and Gao, R. (2024). PAX1 represses canonical Wnt signaling pathway and plays dual roles during endoderm differentiation. Cell Commun Signal *22*, 242. 10.1186/s12964-024-01629-3.

38. Zhang, X., Chen, Y., Ye, Y., Wang, J., Wang, H., Yuan, G., Lin, Z., Wu, Y., Zhang, Y., and Lin, X. (2017). Wnt signaling promotes hindgut fate commitment through regulating multi-lineage genes during hESC differentiation. Cell Signal *29*, 12-22. 10.1016/j.cellsig.2016.09.009.
39. Kim, Y.J., Yoon, S.Y., Kim, J.T., Song, E.Y., Lee, H.G., Son, H.J., Kim, S.Y., Cho, D., Choi, I., Kim, J.H., and Kim, J.W. (2009). NDRG2 expression decreases with tumor stages and regulates TCF/beta-catenin signaling in human colon carcinoma. Carcinogenesis *30*, 598-605. 10.1093/carcin/bgp047.
40. Zhao, J., Zhao, Q., and Mao, S. (2021). N-myc downstream regulated gene 2 ameliorates myocardial remodeling and cardiac function in heart failure rats. Human & Experimental Toxicology *40*, 1296-1307. 10.1177/0960327121993208.
41. Zhu, H., Zhao, J., Zhou, W., Li, H., Zhou, R., Zhang, L., Zhao, H., Cao, J., Zhu, X., Hu, H., et al. (2012). Ndrg2 regulates vertebral specification in differentiating somites. Developmental Biology *369*, 308-318. https://doi.org/10.1016/j.ydbio.2012.07.001.
42. Ge, Y.X., Wang, C.H., Hu, F.Y., Pan, L.X., Min, J., Niu, K.Y., Zhang, L., Li, J., and Xu, T. (2018). New advances of TMEM88 in cancer initiation and progression, with special emphasis on Wnt signaling pathway. J Cell Physiol *233*, 79-87. 10.1002/jcp.25853.
43. Liu, J., Yang, L., Lu, Z., and Wang, Q. (2023). Tmem88 plays an essential role in pharyngeal pouch progenitor specification by inhibiting Wnt/β-catenin signaling. Life Medicine *2*, lnad044. 10.1093/lifemedi/lnad044.
44. Novikov, N., and Evans, T. (2013). Tmem88a mediates GATA-dependent specification of cardiomyocyte progenitors by restricting WNT signaling. Development *140*, 3787-3798. 10.1242/dev.093567.
45. Palpant, N.J., Pabon, L., Rabinowitz, J.S., Hadland, B.K., Stoick-Cooper, C.L., Paige, S.L., Bernstein, I.D., Moon, R.T., and Murry, C.E. (2013). Transmembrane protein 88: a Wnt regulatory protein that specifies cardiomyocyte development. Development *140*, 3799-3808. 10.1242/dev.094789.
46. Kubozono, T., Akasaki, Y., Kawasoe, S., Ojima, S., Kawabata, T., Makizako, H., Kuwahata, S., Takenaka, T., Maeda, M., Fujiwara, S., et al. (2022). The relationship between day-to-day variability in home blood pressure measurement and multiple organ function. Hypertension Research *45*, 474-482. 10.1038/s41440-021-00818-8.
47. McGinnis, C.S., Patterson, D.M., Winkler, J., Conrad, D.N., Hein, M.Y., Srivastava, V., Hu, J.L., Murrow, L.M., Weissman, J.S., Werb, Z., et al. (2019). MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. Nature Methods *16*, 619-626. 10.1038/s41592-019-0433-8.
48. Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176. 10.1126/science.aam8999.
49. Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2019). CellTag Indexing: genetic barcode-based sample multiplexing for single-cell genomics. Genome Biology *20*, 90. 10.1186/s13059-019-1699-y.
50. Volpato, V., and Webber, C. (2020). Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. Dis Model Mech *13*, dmm042317. 10.1242/dmm.042317.
51. Brooks, I.R., Garrone, C.M., Kerins, C., Kiar, C.S., Syntaka, S., Xu, J.Z., Spagnoli, F.M., and Watt, F.M. (2022). Functional genomics and the future of iPSCs in disease modeling. Stem Cell Reports. https://doi.org/10.1016/j.stemcr.2022.03.019.
52. Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219-224. 10.1038/s41586-018-0744-4.
53. VanHorn, S., and Morris, S.A. (2020). Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development. Developmental Cell. 10.1016/j.devcel.2020.10.021.
54. Huang, Y., and Zhang, P. (2021). Evaluation of machine learning approaches for cell-type identification from single-cell transcriptomics data. Brief Bioinform *22*. 10.1093/bib/bbab035.
55. Zhang, M., Liu, J., Mao, A., Ning, G., Cao, Y., Zhang, W., and Wang, Q. (2023). Tmem88 confines ectodermal Wnt2bb signaling in pharyngeal arch artery progenitors for balancing cell cycle progression and cell fate decision. Nature Cardiovascular Research *2*, 234-250. 10.1038/s44161-023-00215-z.
56. Kazenwadel, J., Venugopal, P., Oszmiana, A., Toubia, J., Arriola-Martinez, L., Panara, V., Piltz, S.G., Brown, C., Ma, W., Schreiber, A.W., et al. (2023). A Prox1 enhancer represses haematopoiesis in the lymphatic vasculature. Nature *614*, 343-348. 10.1038/s41586-022-05650-9.
57. Ma, Y., Tully, P.J., Hofman, A., and Tzourio, C. (2020). Blood Pressure Variability and Dementia: A State-of-the-Art Review. Am J Hypertens *33*, 1059-1066. 10.1093/ajh/hpaa119.
58. Sasaki, T., Sakata, S., Oishi, E., Furuta, Y., Honda, T., Hata, J., Tsuboi, N., Kitazono, T., Yokoo, T., and Ninomiya, T. (2022). Day-to-Day Blood Pressure Variability and Risk of Incident Chronic Kidney Disease in a General Japanese Population. J Am Heart Assoc *11*, e027173. 10.1161/jaha.122.027173.

59. Zhou, B., Carrillo-Larco, R.M., Danaei, G., Riley, L.M., Paciorek, C.J., Stevens, G.A., Gregg, E.W., Bennett, J.E., Solomon, B., Singleton, R.K., et al. (2021). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. The Lancet *398*, 957-980. https://doi.org/10.1016/S0140-6736(21)01330-1.
60. Rakugi, H., Yu, H., Kamitani, A., Nakamura, Y., Ohishi, M., Kamide, K., Nakata, Y., Takami, S., Higaki, J., and Ogihara, T. (1996). Links between hypertension and myocardial infarction. American Heart Journal *132*, 213-221. https://doi.org/10.1016/S0002-8703(96)90556-6.
61. Stevens, S.L., Wood, S., Koshiaris, C., Law, K., Glasziou, P., Stevens, R.J., and McManus, R.J. (2016). Blood pressure variability and cardiovascular disease: systematic review and meta-analysis. Bmj *354*, i4098. 10.1136/bmj.i4098.
62. Tapela, N., Collister, J., Clifton, L., Turnbull, I., Rahimi, K., and Hunter, D.J. (2021). Prevalence and determinants of hypertension control among almost 100 000 treated adults in the UK. Open Heart *8*, e001461. 10.1136/openhrt-2020-001461.
63. Omboni, S., Kario, K., Bakris, G., and Parati, G. (2018). Effect of antihypertensive treatment on 24-h blood pressure variability: pooled individual data analysis of ambulatory blood pressure monitoring studies based on olmesartan mono or combination treatment. J Hypertens *36*, 720-733. 10.1097/hjh.0000000000001608.
64. Parati, G., Bilo, G., Kollias, A., Pengo, M., Ochoa, J.E., Castiglioni, P., Stergiou, G.S., Mancia, G., Asayama, K., Asmar, R., et al. (2023). Blood pressure variability: methodological aspects, clinical relevance and practical indications for management - a European Society of Hypertension position paper∗. Journal of Hypertension *41*.
65. Alquicira-Hernandez, J., and Powell, J.E. (2021). Nebulosa recovers single-cell gene expression signals by kernel density estimation. Bioinformatics *37*, 2485-2487. 10.1093/bioinformatics/btab003.
66. Rivera-Pérez, J.A., Jones, V., and Tam, P.P.L. (2010). Chapter 11 - Culture of Whole Mouse Embryos at Early Postimplantation to Organogenesis Stages: Developmental Staging and Methods. In Methods in Enzymology, P.M. Wassarman, and P.M. Soriano, eds. (Academic Press), pp. 185-203. https://doi.org/10.1016/S0076-6879(10)76011-0.
67. Kim, H.J., Wang, K., Chen, C., Lin, Y., Tam, P.P.L., Lin, D.M., Yang, J.Y.H., and Yang, P. (2021). Uncovering cell identity through differential stability with Cepo. Nature Computational Science *1*, 784-790. 10.1038/s43588-021-00172-2.
68. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826. 10.1126/science.1232033.
69. Kreitzer, F.R., Salomonis, N., Sheehan, A., Huang, M., Park, J.S., Spindler, M.J., Lizarraga, P., Weiss, W.A., So, P.L., and Conklin, B.R. (2013). A robust method to derive functional neural crest cells from human pluripotent stem cells. Am J Stem Cells *2*, 119-131.
70. Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat Methods *11*, 291-293. 10.1038/nmeth.2840.
71. Roelli, P., bbimber, Flynn, B., santiagorevale, and Gui, G. (2019). Hoohm/CITE-seq-Count: 1.4.2. Zenodo. https://doi.org/10.5281/zenodo.2590196.
72. Bais, A.S., and Kostka, D. (2020). scds: computational annotation of doublets in single-cell RNA sequencing data. Bioinformatics *36*, 1150-1158. 10.1093/bioinformatics/btz698.
73. Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell Stem Cell *18*, 541-553. 10.1016/j.stem.2016.01.022.
74. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e1821. 10.1016/j.cell.2019.05.031.
75. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330. 10.1038/nature14248.
76. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research *47*, D607-D613. 10.1093/nar/gky1131.
77. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine *12*, e1001779. 10.1371/journal.pmed.1001779.

78. GTEx (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318. 10.1126/science.aaz1776.
79. Dullin, C., Ufartes, R., Larsson, E., Martin, S., Lazzarini, M., Tromba, G., Missbach-Guentner, J., Pinkert-Leetsch, D., Katschinski, D.M., and Alves, F. (2017). µCT of ex-vivo stained mouse hearts and embryos enables a precise match between 3D virtual histology, classical histology and immunochemistry. PLoS One *12*, e0170597. 10.1371/journal.pone.0170597.
80. Horner, N.R., Venkataraman, S., Armit, C., Casero, R., Brown, J.M., Wong, M.D., van Eede, M.C., Henkelman, R.M., Johnson, S., Teboul, L., et al. (2021). LAMA: automated image analysis for the developmental phenotyping of mouse embryos. Development *148*. 10.1242/dev.192955.