nature cell biology

TECHNICAL REPORT

https://doi.org/10.1038/s41556-022-00884-1

Check for updates

CellComm infers cellular crosstalk that drives haematopoietic stem and progenitor cell development

Edroaldo Lummertz da Rocha[1,13], Caroline Kubaczka[2,3,4,13], Wade W. Sugden[2,3], Mohamad Ali Najia[2,3,4,5,6], Ran Jing[2,3,4], Arianna Markel[2,3,4], Zachary C. LeBlanc[2,3], Rafael dos Santos Peixoto[7], Marcelo Falchetti[8], James J. Collins[5,6,9,10,11,12], Trista E. North[2,3,14] ✉ and George Q. Daley[2,3,4,14] ✉

Intercellular communication orchestrates a multitude of physiologic and pathologic conditions. Algorithms to infer cell–cell communication and predict downstream signalling and regulatory networks are needed to illuminate mechanisms of stem cell differentiation and tissue development. Here, to fill this gap, we developed and applied CellComm to investigate how the aorta–gonad–mesonephros microenvironment dictates haematopoietic stem and progenitor cell emergence. We identified key microenvironmental signals and transcriptional networks that regulate haematopoietic development, including *Stat3*, *Nr0b2*, *Ybx1* and *App*, and confirmed their roles using zebrafish, mouse and human models. Notably, CellComm revealed extensive crosstalk among signalling pathways and convergence on common transcriptional regulators, indicating a resilient developmental programme that ensures dynamic adaptation to changes in the embryonic environment. Our work provides an algorithm and data resource for the scientific community.

Haematopoietic stem and progenitor cell (HSPC) emergence is orchestrated by a highly regulated developmental programme. In the mid-gestation mouse embryo, HSPCs are born in the aorta–gonad–mesonephros (AGM) region around embryonic day (E)10.5. Recent studies have examined haematopoietic development within the AGM using sorted populations to catalogue the transcriptional programme of haemogenic endothelium (HE) specification and its differentiation trajectory to functional haematopoietic stem cells (HSCs)[1–5]. Niche-derived signals are important for HSC specification; however, a comprehensive atlas of the cellular components of the entire AGM niche and their influence on HSC fate has been lacking. In this Technical Report, to close this gap, we applied CellComm to obtain insights into the microenvironmental regulation of HSPC emergence in the AGM region. We performed extensive experimental validation of CellComm's predictions using zebrafish and mouse embryos as well as human induced pluripotent stem cells (iPSCs), then confirmed the roles of several ligand–receptor interactions and downstream transcriptional regulators implicated in haematopoietic development. These findings enhance our understanding of cellular dynamics in the HSC niche and provide further guidance for precise differentiation of iPSCs towards HSPCs. CellComm is a valuable resource for the broader scientific community to investigate critical cell–cell regulatory interactions from single-cell RNA sequencing (scRNA-seq) or spatial transcriptomics data.

Results

The cellular landscape of the AGM microenvironment. CellComm is a systems biology algorithm combining transcriptome data (scRNA-seq or spatial transcriptomics) with protein–protein interaction networks to infer how communication between cells activates downstream signalling pathways and transcriptional programmes dictating cell fates (Fig. 1a and Supplementary Note 1). Briefly, CellComm infers which cell types communicate on the basis of ligand–receptor interactions identified by calculating intracluster mean expression of ligands or receptors among pairwise combinations of cell types. If spatial transcriptomics data are available, CellComm considers co-localization of cell types or niches to predict cell communication. By weighting protein–protein interaction networks using intracluster co-expression measurements, CellComm implements an optimization algorithm to identify paths in the interactome that connect cell surface genes to downstream transcriptional regulators, predicting putative effectors of signalling networks in a fully data-driven manner.

To investigate the process by which HSCs are produced de novo during embryogenesis, we first performed scRNA-seq

[1]Department of Microbiology, Immunology and Parasitology, Federal University of Santa Catarina, Florianópolis, Brazil. [2]Stem Cell Program, Boston Children's Hospital, Boston, MA, USA. [3]Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, MA, USA. [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. [5]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Harvard-MIT Health Sciences & Technology, Massachusetts Institute of Technology, Cambridge, MA, USA. [7]Undergraduate program in Automation and Control Engineering, Federal University of Santa Catarina, Florianópolis, Brazil. [8]Graduate Program of Pharmacology, Center for Biological Sciences, Federal University of Santa Catarina, Florianópolis, Brazil. [9]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [10]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA. [13]These authors contributed equally: Edroaldo Lummertz da Rocha, Caroline Kubaczka. [14]These authors jointly supervised this work: Trista E. North, George Q. Daley. ✉e-mail: trista.north@childrens.harvard.edu; George.Daley@childrens.harvard.edu

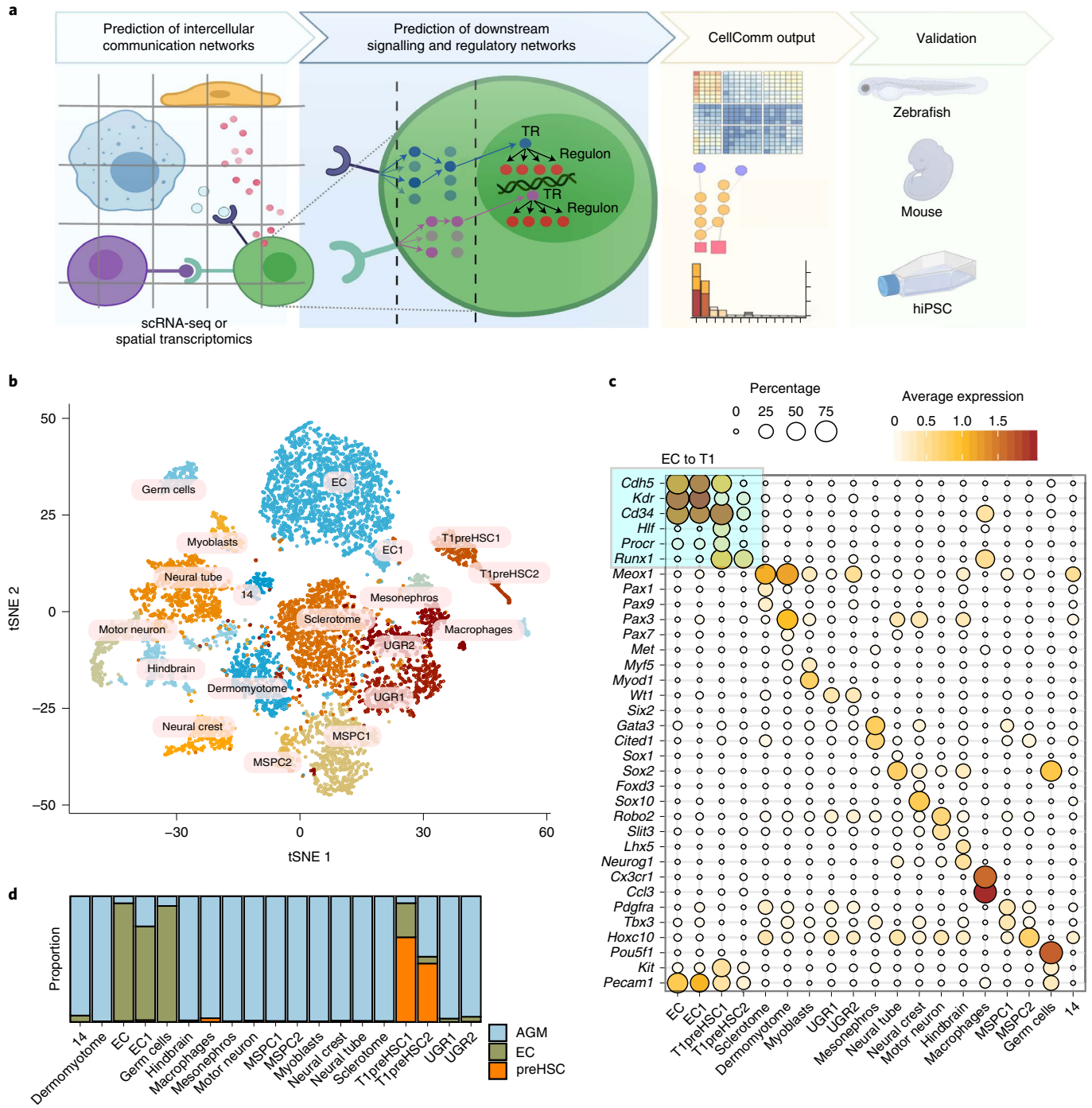


Fig. 1 | Diversity of cell types in the AGM ecosystem. a, CellComm workflow to predict intercellular communication networks and signalling pathways from scRNA-seq or spatial transcriptomics data. (1) CellComm predicts possible cell–cell communication on the basis of co-expression patterns of curated ligand–receptor pairs; when spatial transcriptomics are available, CellComm infers spatially resolved cell–cell interactions. (2) CellComm next implements an optimization algorithm to identify paths between cell surface receptors and transcriptional regulators (TR) in a large-scale weighted protein–protein interaction network. Each interaction in the protein interaction network is weighted by cluster- or cell-type-specific co-expression measurements derived from the scRNA-seq data. The algorithm then finds paths connecting cell surface receptors to transcriptional regulators, predicting a putative signalling pathway. (3) CellComm implements several downstream analytics to extract biological information from the data, such as identification of interacting cell types, ranking of ligand–receptor pairs and identification of signalling pathways as well as gene regulatory networks downstream of cell surface receptors. These analytics aid in elaborating testable hypotheses, which can be validated experimentally in the cell type(s) of interest. Created with BioRender.com. **b**, tSNE analysis colour-coded by transcriptional cluster or cell state; cluster 14 was not assigned to any cell type or state. **c**, Genes preferentially expressed in each cell type (box indicates endothelial and haematopoietic genes). **d**, Proportion of cells from each sample in each cell type or state. $N = 9{,}492$ cells. Source numerical data are available in Source data.

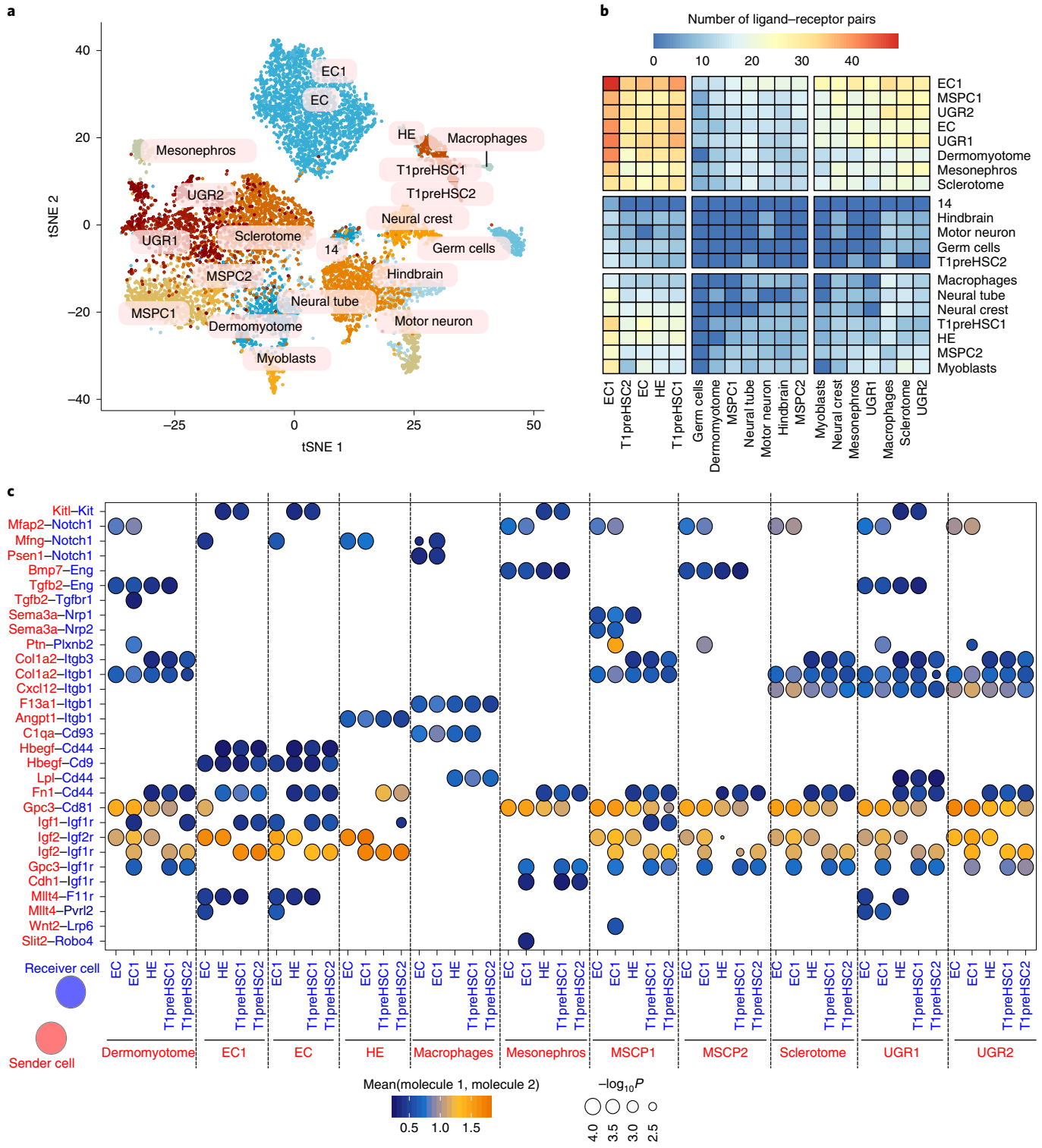


Fig. 2 | Intercellular communication in the AGM ecosystem. a, tSNE analysis of the AGM single-cell data using only ligands and receptors for dimensionality reduction. **b**, Predicted cell–cell communication network in the AGM ecosystem. **c**, Selected, top-ranked ligand–receptor pairs predicted to mediate cell–cell interactions potentially implicated in haematopoietic development. Source numerical data are available in Source data.

on mouse AGM at E10.25, a developmental timepoint during which HSC precursors (T1preHSCs) emerge via the process of endothelial-to-haematopoietic transition (EHT)[6,7]. Following manual dissection, we profiled the entire community of cells within the AGM, as well as fluorescence-activated cell sorting (FACS)-enriched populations of endothelial cells (ECs) and immunophenotypic T1preHSCs, to investigate how cellular crosstalk in the developmental niche regulates HSPC formation (Extended Data Fig. 1a,b).

We used CellRouter to identify cell communities within the AGM ecosystem[8]. Unsupervised clustering of 9,492 cells identified 19 distinct transcriptional clusters, which were annotated on the basis of expression of established marker genes (Fig. 1b,c and

Methods). We applied *t*-stochastic neighbour embedding (tSNE) analysis to visualize clusters and putative cell types or states on the tSNE space (Fig. 1b and Extended Data Fig. 1c–e). Investigating the proportion of cells within the whole AGM and sorted fractions that fell into the identified cell types revealed that the cell state defined as T1preHSC1 was composed of a mixture of cells coming from the EC and T1preHSC fractions, indicative of the presence of a subset of ECs that was already transcriptionally similar to T1preHSCs (Fig. 1d). In-depth analysis of EC and T1preHSC states suggested that this subset has haemogenic potential, and it is thus referred to hereafter as HE (Extended Data Fig. 1f–i, Supplementary Fig. 1 and Supplementary Note 2).

The intercellular communication network of the AGM niche. We next applied CellComm to identify intercellular communication networks potentially active within the AGM. A tSNE analysis using only ligands and receptors showed that cells clustered by cell type, suggesting cell-type-specific expression of signalling molecules (Fig. 2a). Our intercellular communication network of the AGM revealed that cellular states defined as endothelial cell 1 (EC1), HE, mesonephros, urogenital ridges (UGRs) and somites (sclerotome and dermomyotome) as well as the mesenchymal subset mesenchymal stem and progenitor cell 1 (MSPC1) interact with EC1, HE and T1preHSC populations through a higher diversity of ligand–receptor pairs compared with the neural and muscle lineages (Fig. 2b). In addition to known interactions between ECs, HE and T1preHSCs, both somite- and UGR-derived signals were previously implicated in HSC development, consistent with our results[9,10]. Applying Gene Ontology analysis, we found that ligands were enriched in terms such as cell differentiation, migration and angiogenesis, with some biological functions found across several cell types, while others were more cell-type specific (Supplementary Fig. 2a). Analysis of cell surface receptors revealed that signalling-related processes, such as phosphorylation, were broadly enriched across cell types (Supplementary Fig. 2b).

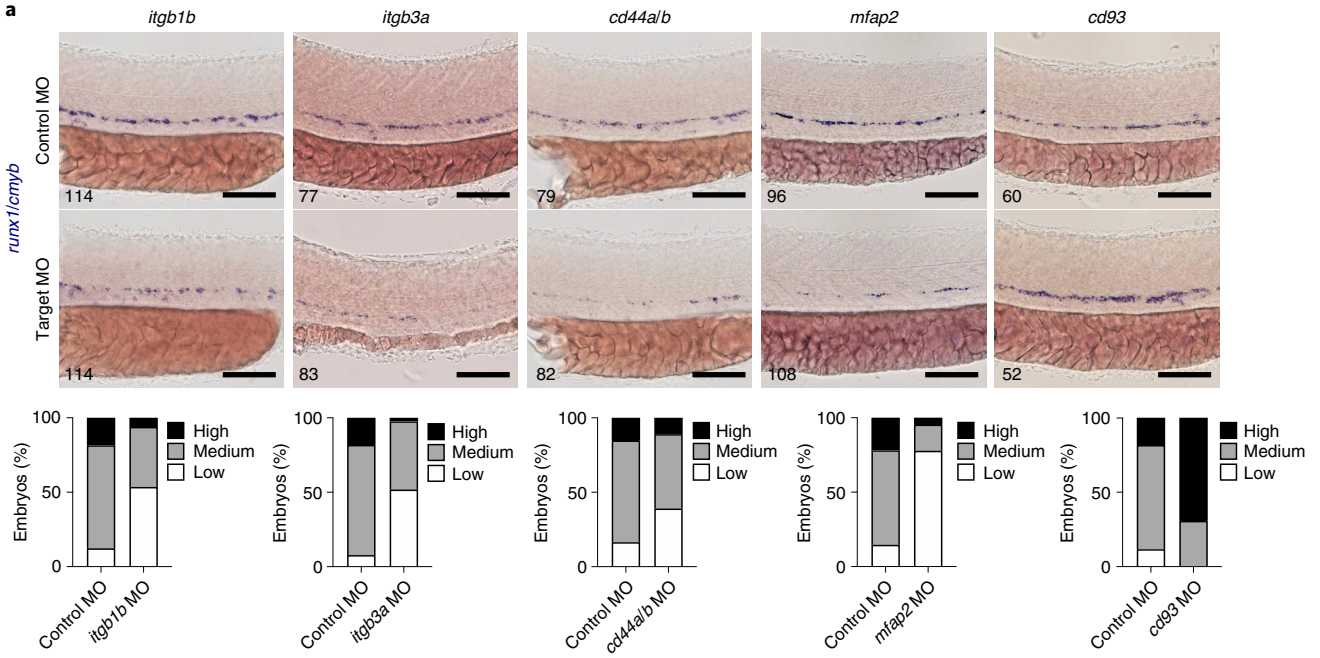
We next ranked ligand–receptor pairs across selected cell–cell interactions on the basis of the magnitude of their co-expression in interacting cell types as reported in Fig. 2b (Supplementary Fig. 2c). Among prioritized pairs, we selected a subset with known function in haematopoiesis (for example, *Cd44* and *Kit*) as well as some candidates with less well-characterized roles (for example, integrins and *Mfap2*; Fig. 2c) for further validation. This analysis implicated several signalling processes in T1preHSC differentiation. For instance, Cd44, expressed by HE and T1preHSCs, was recently identified by scRNA-seq as an HE marker and regulator of EHT[11]. Additionally, we noted several ligand–receptor pairs enriched with integrins, a subset of which were previously described to regulate HSPC development[12].

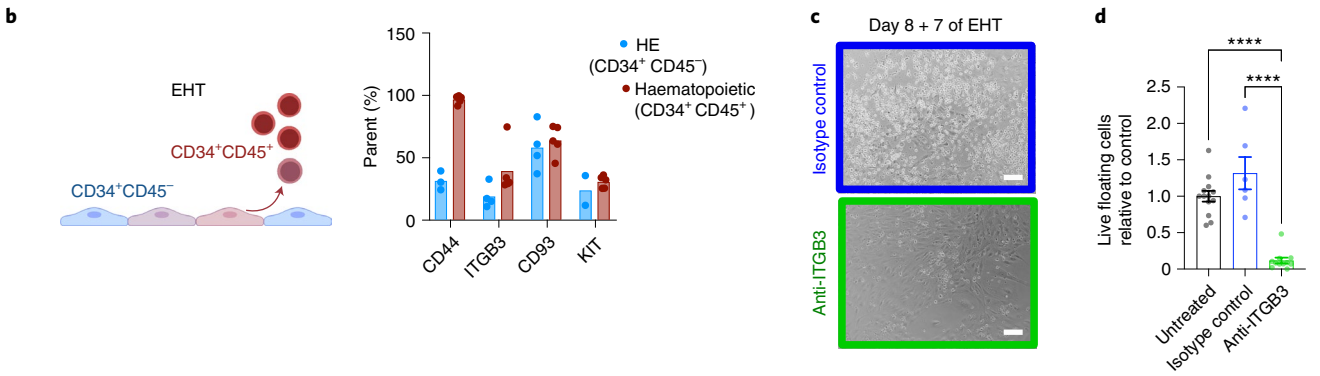
To validate our algorithm and confirm ligand–receptor interactions required for definitive haematopoiesis in vivo, we employed morpholino-based knockdown in zebrafish embryos, an experimentally tractable model that has proven invaluable for identifying genes and pathways impacting HSPC development[13–15]. *Itgb1b* knockdown reduced expression of the conserved HSPC markers, *runx1* and *cmyb*, at 36 h post-fertilization (hpf), as reported[16] (Fig. 3a). Knockdown of *itgb3a*, *mfap2* and *cd44a/b* also reduced *runx1*/*cmyb* expression in the embryo (Fig. 3a), while knockdown of *itgb3b* had no effect (Extended Data Fig. 2a). Interestingly, knockdown of *cd93*, a published marker of emerging AGM HSCs, which fails to enrich for engrafting cells[2], dramatically increased HSPC production as determined by *runx1*/*cmyb* expression, suggestive of negative regulation (Fig. 3a). These data confirm that CellComm accurately predicted receptors influencing HSPC emergence, both positively and negatively.

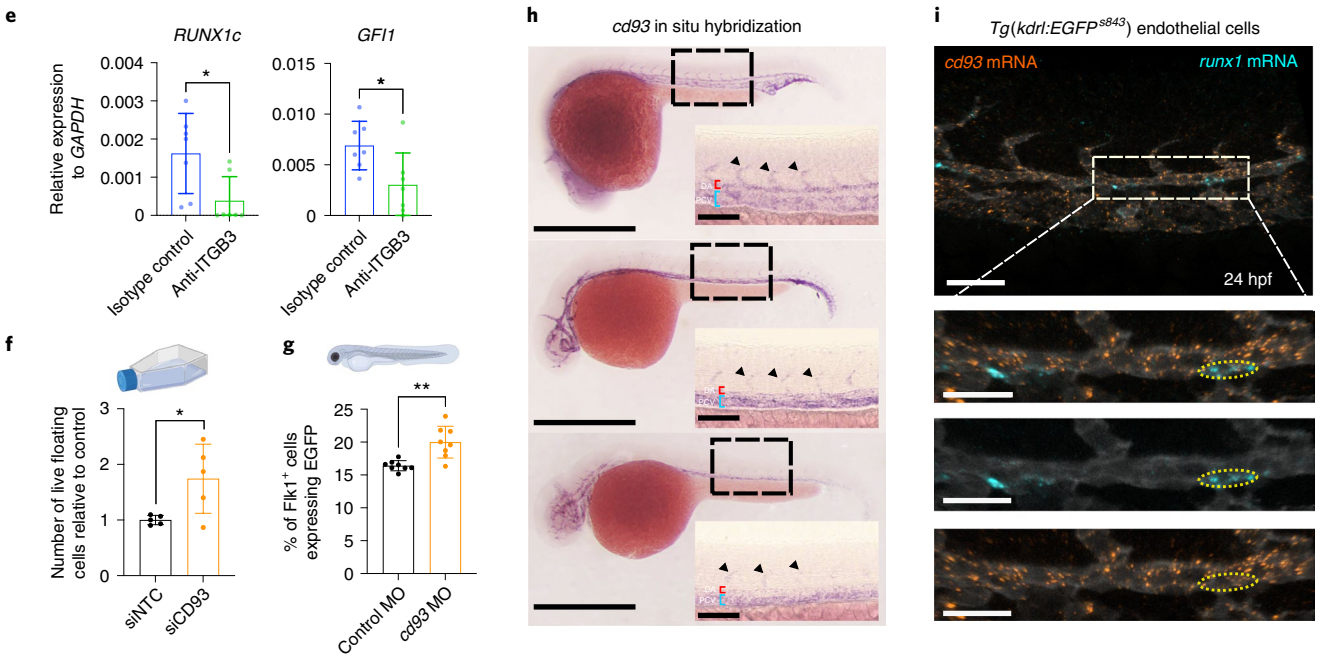
Human iPSCs (hiPSCs) offer a tractable alternative to study the molecular mechanisms of HSPC development from HE. While bona fide HSCs have not been derived from hiPSCs thus far in the absence of transcriptional modulation, in vitro protocols do recapitulate some aspects of HSPC ontogeny. HiPSCs can be differentiated into CD34$^+$ HE that undergoes EHT and acquires haematopoietic surface markers[17]. Using this protocol, we detected cells expressing CD44, ITGB3, CD93 and KIT at the HE (CD34$^+$ CD45$^-$) and HSPC stage (CD34$^+$ CD45$^+$) by flow cytometry (Fig. 3b and Extended Data Fig. 2b). Among the analysed markers, we focused on ITGB3, which was expressed on only a subset of CD34$^+$ cells and increased as cells underwent EHT (Fig. 3b). While ITGB3 expression can be used to enrich long-term repopulating HSCs and label HE in both differentiated hiPSC cultures and mouse embryos, functional assessment of downstream ITGB3 signalling has been lacking[18,19]. We found that inhibition of ITGB3 signalling using a neutralizing antibody severely impaired hiPSC-derived EHT in vitro (Fig. 3c,d). In line with a significant decrease in floating haematopoietic cells, cultures failed to upregulate the haematopoietic transcription factor *RUNX1C* and its downstream target *GFI1*, supporting a requirement for active ITGB3 signalling in EHT (Fig. 3e).

In contrast to that of integrin family members, *cd44*, and *mfap2*, knockdown of *cd93* enhanced HSPC gene expression in zebrafish (Fig. 3a). During iPSC-derived EHT, inhibition of CD93 via short interfering RNA (siRNA)-mediated knockdown significantly increased the number of floating cells without compromising differentiation potential (Fig. 3f and Extended Data Fig. 2c–e). This result is in line with significant increases in HSPCs observed by flow cytometry after morpholino-mediated knockdown of *cd93* in zebrafish (Fig. 3g). Analysis of endogenous expression patterns of *cd93* in zebrafish supports a role as a negative regulator of EHT: *cd93* was present in all trunk vessels across the window of EHT (Fig. 3h).

Fig. 3 | Systematic validation of ligand–receptor interactions. a, WISH for *runx1/cmyb* in standard control and experimental morpholino (MO)-injected embryos at 36 hpf. Scale bars, 100 µm. *N* for each condition is indicated in the lower left corner of the micrographs. **b**, Flow-cytometric quantification of indicated surface markers in hiPSC-derived HE and haematopoietic cells. At least $N=2$ experiments per marker. Each datapoint is from an independent differentiation culture. Bars represent mean. Schematic figure created with BioRender.com. **c**, Brightfield images of hiPSC–EHT cultures with and without neutralizing antibody treatment against ITGB3. Scale bars, 250 µm. **d**, Quantification of live, floating cells after seven days of at least five independent EHT cultures. Bars represent mean ± s.d. One-way ANOVA, *P* values were adjusted for multiple comparisons using Dunnett's multiple comparisons test. $^{****}P<0.0001$. **e**, RT–qPCR for haematopoietic genes *RUNX1c* and *GFI1* at day 7 of EHT after ITGB3 blocking versus control. Bars represent mean ± s.d. Unpaired *t* test, $^{*}P=0.0201$ (*RUNX1c*); $P=0.0234$ (*GFI1*). $N=7$, from four independent differentiations. **f**, Quantification of live, floating cells after 7 days of EHT culture in CD93 knockdown cultures compared with control siRNA-treated wells. $N=5$ independent experiments, unpaired *t* test $P=0.0295$. Bar graphs represent mean ± s.d. **g**, Flow-cytometric quantification of mCherry$^+$/EGFP$^+$ cells from transgenic *flk1:mCherry/cmyb:EGFP* embryos at 48 hpf in control versus *cd93* morpholino-injected embryos. Unpaired *t* test, $^{**}P=0.0013$. Bar graphs represent mean ± s.d. **h**, WISH timecourse for zebrafish *cd93* across the window of EHT. Black arrowheads indicate intersegmental vessels, and red and blue brackets denote the dorsal aorta (DA) and posterior cardinal vein (PCV), respectively. Scale bars, 500 µm and 50 µm (insets). The experiment was performed twice, analysing 20 fish in total. **i**, Maximum intensity projection of a confocal *z* stack of double FISH for zebrafish *cd93* and *runx1* mRNA at 24 hpf. Vasculature is visualized with GFP antibody staining in transgenic embryos. Yellow outline indicates the position of *runx1*$^+$ haemogenic EC. Scale bars, 50 µm (top image) and 25 µm (three lower images, representing enlarged area within the dashed rectangle). Three biological replicates have been analysed. Source numerical data are available in Source data.


a
itgb1b
itgb3a
cd44a/b
mfap2
cd93
runx1/cmyb
Control MO
Target MO
114
77
79
96
60
114
83
82
108
52
Embryos (%)
High
Medium
Low
Control MO
itgb1b MO
itgb3a MO
cd44a/b MO
mfap2 MO
cd93 MO


b
EHT
CD34+CD45+
CD34+CD45−
Parent (%)
HE (CD34+ CD45−)
Haematopoietic (CD34+ CD45+)
CD44
ITGB3
CD93
KIT
c
Day 8 + 7 of EHT
Isotype control
Anti-ITGB3
d
Live floating cells relative to control
Untreated
Isotype control
Anti-ITGB3
****


e
RUNX1c
GFI1
Relative expression to GAPDH
Isotype control
Anti-ITGB3
f
Number of live floating cells relative to control
siNTC
siCD93
g
% of Flk1+ cells expressing EGFP
Control MO
cd93 MO
h
cd93 in situ hybridization
i
Tg(kdrl:EGFPs843) endothelial cells
cd93 mRNA
runx1 mRNA
24 hpf

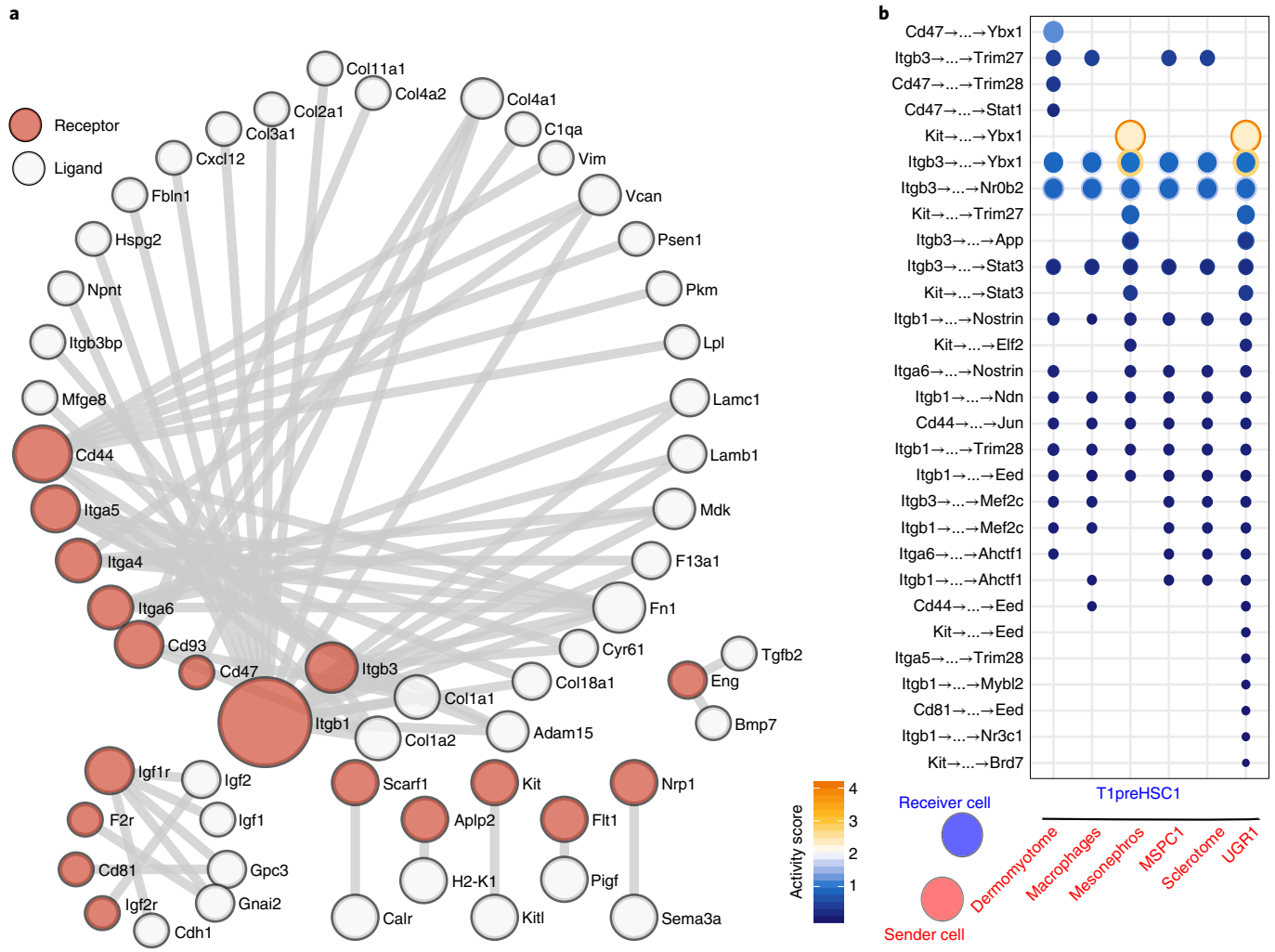


Fig. 4 | Reconstruction of downstream pathways. a, Ligand–receptor network built from ligands and receptors involved in HE and T1preHSC interactions with niche cells. **b**, Predicted downstream signalling pathways from cell surface receptors to transcriptional regulators. Dots between arrows indicate intermediate proteins in the path between cell surface receptors and potential downstream transcriptional regulators indicated. Source numerical data are available in Source data.

However, expression in the dorsal aorta was highest at 24 hpf and decreased over time (red brackets). High-resolution imaging of the aorta revealed that *runx1*$^{+}$ cells had markedly less expression of *cd93* compared with surrounding ECs in the artery (Fig. 3i), consistent with the downregulation of this gene as a corollary of haematopoietic fate acquisition. Finally, in situ hybridization for *mfap2* in zebrafish embryos shows broad expression at 24–36 hpf, consistent with CellComm predictions. Importantly, *mfap2* expression can be found in cells adjacent to *runx1*$^{+}$ HE cells at 24 hpf, indicating a paracrine role in HSPC development (Extended Data Fig. 2f,g). To test whether MFAP2 can act as a soluble Notch ligand, we cultured hiPSC-derived HE in the presence of recombinant human MFAP2. While we observed significant decreases in the number of haematopoietic cells, there was a relative increase in the undifferentiated CD34^{+} CD45^{+} double-positive population (Extended Data Fig. 2h,i). Together, these data suggest that CellComm accurately identified active ligand–receptor pairs that impact vertebrate HSPC development, either positively or negatively regulating the production of human HSPCs from HE. Although CellComm predicted genes that may either promote or buffer HSPC formation, prediction of the directionality of the phenotype for any given gene modulation warrants further refinement of CellComm and experimental confirmation.

Signalling pathways driving T1preHSC differentiation. Somites, UGRs and macrophages have each been previously reported to provide signals to regulate HSC development[9,10,20–23]. Although modelling studies have indicated that secreted ligands can act at distances up to 250 µm from their source[24], quantitative measurements of spatial distances between interacting tissues are often lacking. Regardless, interactions between spatially segregated tissues are important for patterning and specification, which is mediated by gradients of secreted molecules. Therefore, in addition to adjacent cells in the niche microenvironment, molecules secreted by distant tissues, such as somites and UGR, may influence haematopoietic development[22,23]. Using selected cell–cell interactions reported in Fig. 2c, we built a ligand–receptor interaction network predicted to be involved in HE and T1preHSC differentiation. This network contained several integrin, insulin growth factor (Igf1r and Igf2r), Kit and Tgfβ signalling components (Eng, the receptor for Tgfb2 and Bmp7) as well as other cell surface molecules, including Cd44, Cd47 and Cd93 (Fig. 4a). Among these cell surface receptors, CellComm identified a subset for which there were connecting paths to downstream transcriptional regulators, suggesting functional signalling pathways (Fig. 4b). To reconstruct such signalling pathways, CellComm implements an optimization algorithm to find paths in a protein interaction network weighted by gene co-expression measurements

that connect cell surface receptors to transcriptional regulators (Supplementary Note 1). Notably, the mesonephros subset and UGR subset 1 (UGR1) expressed Kit ligand (*Kitl*), which activates Kit signalling and is critical for HSC survival[25] (Fig. 4b). The functional relevance of *Kitl* provided from each cell type to HSPC development warrants further investigation; nonetheless, several transcriptional regulators, including *Ybx1*, *Stat3*, *Nr0b2*, *Trim28* and *Eed*, as well as *App*, whose intracellular domain influences gene expression[26], were identified as potential downstream mediators.

The UGR is a unique component of the embryonic but not the adult haematopoietic niche, which flanks both sides of the dorsal aorta, where nascent HSCs are born. Previous investigations have indicated that the UGR is important for normal haematopoietic development[23]. Thus, we endeavoured to analyse the UGR1–T1preHSC1 interaction in more detail (Fig. 5a). CellComm identified several paths in the weighted protein interaction network leading to putative transcriptional regulators of haematopoietic development (*Ybx1*, *Nr0b2*, *App* and *Stat3)* as well as known regulators, *Mef2c* and *Cebpa*, involved in lymphoid and myeloid development, respectively[27,28]. Interestingly, CellComm reconstructed the Jak–Stat signalling pathway downstream of *Kit*, consistent with previous reports, suggesting that it predicts biologically meaningful signalling pathways from unbiased data analysis[29] (Fig. 5a). The enriched regulons of these transcriptional regulators are involved in hallmark processes of HSC identity, such as a multipotency programme involving myeloid and lymphoid processes, regulated by *Nr0b2* and *Stat3*, and a transcriptional programme regulating ribosome biogenesis, predicted to be regulated by *Ybx1* and *App* (Fig. 5b). Indeed, a gene regulatory network analysis indicated that *Stat3* and *Nr0b2* regulate critical transcription factors for HSC identity, including *Ikzf1*, *Ikzf2*, *Lmo2*, *Myb* and *Hlf (*ref. [2]) (Fig. 5c). On the other hand, *Ybx1* and *App* regulate ribosomal genes, which have been increasingly appreciated to be important for haematopoietic differentiation[30] (Fig. 5c). Interestingly, we independently identified Neurl3, a recently described HE marker[3], as a component of the HSC programme in our T1preHSC gene regulatory network analysis (Fig. 5c). In our dataset, *Neurl3*, expressed in HE and the T1preHSC1 subset, is predicted to be controlled by *Stat3* and *Nr0b2*, suggesting further directions to functionally investigate this regulatory network (Fig. 5c and Extended Data Fig. 1i). To demonstrate the flexibility of CellComm to study cell–cell interactions, we also successfully applied it to identify downstream transcriptional regulators in HE from our dataset, with an independently derived sample, and between ontologically distinct fractions in the AGM region (Fig. 5d,e, Extended Data Fig. 3 and Supplementary Note 3).

CellComm identifies genes involved in developmental haematopoiesis. To better understand the transcriptional dynamics of the identified downstream regulators and obtain further insights about the functional role(s) of the genes selected for experimental validation, we reconstructed the EHT differentiation trajectory using CellRouter[ ] (Extended Data Fig. 4 and Supplementary Note 4). On the basis of these analyses, we selected *Stat3*, *Nr0b2*, *Ybx1* and *App* for experimental validation, as they were highly prioritized by CellComm, dynamically regulated during EHT and less well characterized in developmental haematopoiesis. To more comprehensively test CellComm's predictions, we selected low-scoring genes *Brd7* and *Cebpa* as internal controls to calibrate the lower bound for detection of relevant targets from single-cell transcriptomic data (Figs. 4b and 5a, and Extended Data Fig. 4).

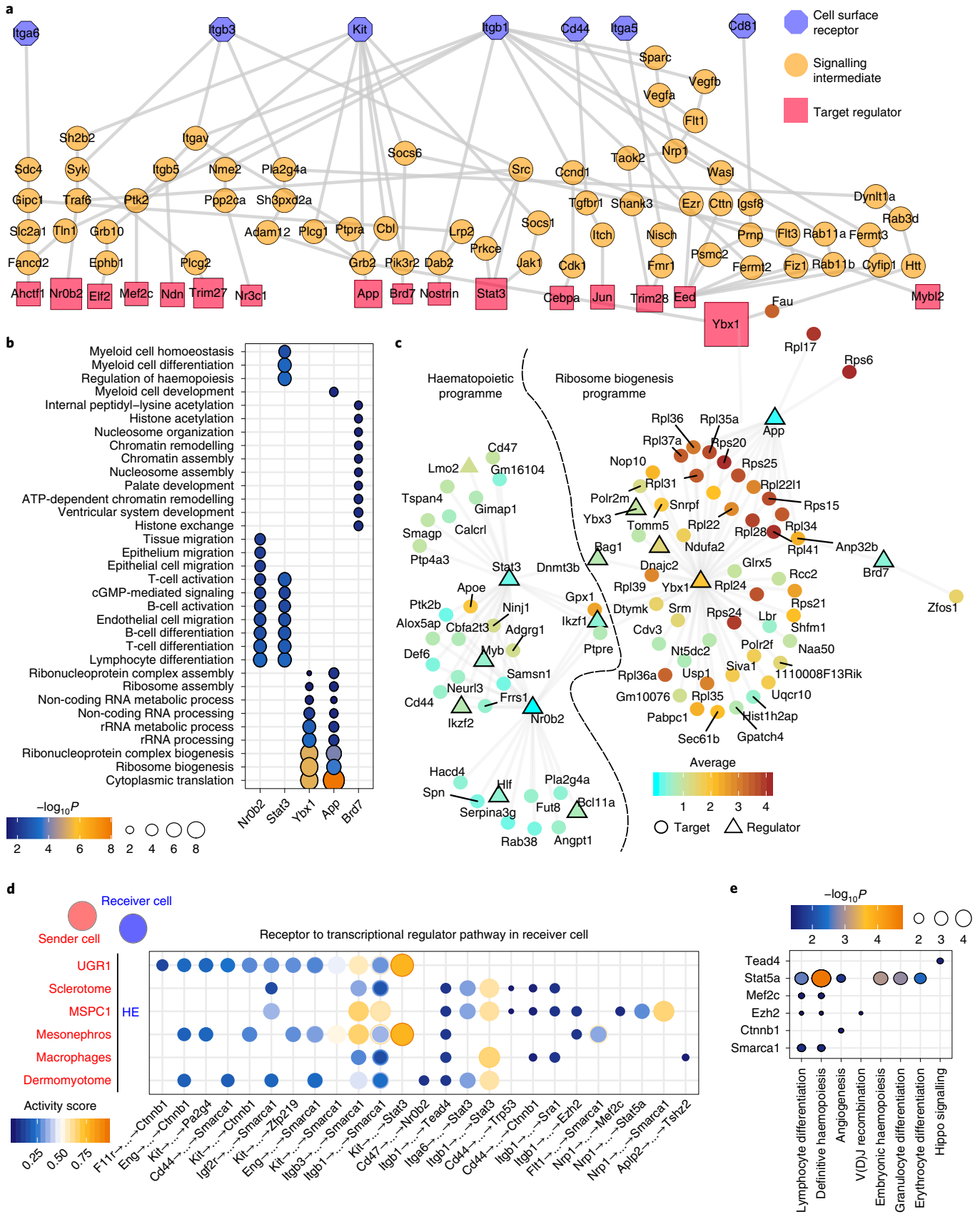
To functionally validate candidate transcriptional regulators during haematopoietic development, we exposed developing zebrafish embryos to C188-9, a small-molecule inhibitor of Stat3, and observed a dose-dependent reduction in *runx1*$^+$/*cmyb*$^+$ HSPC expression (Fig. 6a and Extended Data Fig. 5a). While directed differentiation of HSPCs from hiPSCs in the presence of C188-9 did not significantly reduce CD34$^+$ CD45$^+$ progenitors, it specifically affected their multipotent potential as evidenced by loss of mixed-type colonies (GEMM) in colony-forming unit (CFU) assays and the complete loss of lymphoid differentiation capacity as assayed by CD5$^+$ CD7$^+$ T-cell progenitor differentiation (Fig. 6b,c and Extended Data Fig. 5b–d).

To confirm a role for additional transcriptional regulatory candidates *brd7*, *nr0b2*, *ybx1*, *cebpa/b* and *app* in haematopoietic development, we performed morpholino-knockdown and chemical inhibition in zebrafish embryos. Knockdown of *cebpa* or *cebpb* had no effect and knockdown of *brd7* reduced *runx1*/*cmyb* expression, whereas *nr0b2a* and *appb* increased the appearance of HSPC markers (Fig. 6d and Extended Data Fig. 5e). Utilizing an inducible clustered regularly interspaced short palindromic repeats interference (CRISPRi) approach in hiPSCs, we were able to inhibit *YBX1* coincident with formation of HE. Whereas a 50% reduction in transcript levels of *YBX1* in HE mildly affected formation of phenotypic CD34$^+$CD45$^+$ HSPCs, knockdown impaired overall colony-forming potential as well as lymphoid differentiation potential (Extended Data Fig. 5f–i).

Knockdown-mediated enhancement in *runx1*$^+$/*cmyb*$^+$ staining occurring via loss of the orthologue *appb* was corroborated using 2-PMAP, a small-molecule inhibitor of APP translation and by CRISPR-mediated knockout of *appb* (Fig. 6e). In vivo, *appb* messenger RNA was inversely correlated with *runx1* expression: *appb* transcript was enriched in the dorsal aspect of the aorta, and reduced in HE (Fig. 6f). Moreover, APP inhibition in mouse embryo explant cultures significantly increased the percentage of nascent haematopoietic (CD45$^+$VE-CAD$^+$) cells, which upon more detailed characterization represent phenotypic T2-preHSCs (Fig. 6g and Extended Data Fig. 5j,k). An undescribed role for *App* in HSPC development was further supported by a significant increase in lineage-negative, Sca1$^+$, Kit$^+$ cells in foetal liver samples from E14.5 *App* knockout embryos, compared with littermate controls (Fig. 6h and Extended Data Fig. 5l).

Taken together, our extensive functional validation of cell surface receptors (or ligands) and transcriptional regulators established that out of the 11 gene candidates functionally investigated in this study, CellComm correctly predicted 5 out of 5 cell surface receptors or ligands, and 5 out of 6 downstream transcriptional regulators as either promoting or impairing HSPC development in zebrafish, mouse or hiPSCs. Of note, only *Cebpa* or *Cebpb* did not show an

Fig. 5 | Downstream pathways of UGR1–T1preHSC interaction. a, Signalling network predicted by CellComm identifying transcriptional regulators downstream of cell surface receptors. Ligands are produced by the UGR1 and cell surface receptors are expressed by T1preHSC1. **b**, Gene Ontology processes enriched in regulons (predicted target genes) of the respective transcriptional regulators. The colour and size scales both represent the $-\log_{10}P$ of the pathway enrichment result (hypergeometric test). The *P* values were adjusted for multiple comparisons using the Benjamini–Hochberg method. **c**, Gene regulatory network derived from the transcriptional regulators *Nr0b2, Stat3, Ybx1, App* and *Brd7*. **d**, Predicted downstream signalling pathways from cell surface receptors leading to transcriptional regulators in HE. The *y* axis shows cells producing ligands or receptors (on the left and right hand side of the cell–cell interaction pair, respectively). The *x* axis shows inferred signalling pathways, beginning with a cell surface receptor and ending with downstream transcriptional regulators in receiving cells. **e**, Gene Ontology enrichment analysis of the regulons of each transcriptional regulator. The colour and size scales both represent the $-\log_{10}P$ of the pathway enrichment result. Hypergeometric test. The *P* values were adjusted for multiple comparisons using the Benjamini–Hochberg method. Source numerical data are available in Source data.


a
Cell surface receptor
Signalling intermediate
Target regulator
b
Myeloid cell homoeostasis
Myeloid cell differentiation
Regulation of haemopoiesis
Myeloid cell development
Internal peptidyl–lysine acetylation
Histone acetylation
Nucleosome organization
Chromatin remodelling
Chromatin assembly
Nucleosome assembly
Palate development
ATP-dependent chromatin remodelling
Ventricular system development
Histone exchange
Tissue migration
Epithelium migration
Epithelial cell migration
T-cell activation
cGMP-mediated signaling
B-cell activation
Endothelial cell migration
B-cell differentiation
T-cell differentiation
Lymphocyte differentiation
Ribonucleoprotein complex assembly
Ribosome assembly
Non-coding RNA metabolic process
Non-coding RNA processing
rRNA metabolic process
rRNA processing
Ribonucleoprotein complex biogenesis
Ribosome biogenesis
Cytoplasmic translation
Nr0b2 Stat3 Ybx1 App Brd7
−log10P
c
Haematopoietic programme
Ribosome biogenesis programme
Average
Target
Regulator
d
Receiver cell
Sender cell
Receptor to transcriptional regulator pathway in receiver cell
UGR1
Sclerotome
MSPC1
Mesonephros
Macrophages
Dermomyotome
HE
Activity score
e
−log10P
Tead4
Stat5a
Mef2c
Ezh2
Ctnnb1
Smarca1
Lymphocyte differentiation
Definitive haemopoiesis
Angiogenesis
V(D)J recombination
Embryonic haemopoiesis
Granulocyte differentiation
Erythrocyte differentiation
Hippo signalling

HSPC phenotype in zebrafish MO assays. Supplementary Table 1 contains a list of knockout phenotypes for genes that we have identified. Finally, to demonstrate the broad applicability of CellComm and illustrate its features to infer intercellular communication from spatial transcriptomics data, we applied it to predict spatially resolved intercellular communication networks in the tumour microenvironment, which provided insights into the spatial organization of squamous cell carcinoma, including specialized niches and spatial distribution of regulon activity of prioritized transcriptional regulators, which can be further functionally validated and exploited, as described here for the regulation of HSPC development (Extended Data Figs. 6–8, described in detail in Supplementary Note 5).

Discussion

CellComm predicts cell–cell interactions from single-cell data to illuminate how such communication activates or represses key transcriptional programmes in target cells through reconstruction of signalling pathways and gene regulatory networks. To establish the effectiveness of CellComm, we assembled a comprehensive atlas of the entire AGM ecosystem at a time when HSC precursors emerge through EHT. Our systematic experimental validation of cell-extrinsic and cell-intrinsic factors predicted by CellComm demonstrated its ability to identify biologically meaningful ligand–receptor pairs and downstream transcriptional regulators involved in haematopoiesis.

Ideally, to robustly infer intercellular communication, downstream signalling pathways and regulatory networks within a tissue, an experimental and computational approach is required that quantifies and integrates multimodal parameters, including protein–protein and post-translational modifications, transcriptome, epigenome and potentially other molecular features at single-cell resolution. Technological limitations, however, restrict the scalability to which protein–protein interactions and protein modifications are measured at the single-cell level, imposing substantial challenges to infer cellular communication from orthogonal single-cell omics data. To overcome some of these challenges, CellComm integrates scRNA-seq (or spatial transcriptomics), curated large-scale protein interaction networks and inference of gene regulatory networks to predict how cellular interactions affect downstream transcriptional networks through data-driven reconstruction of pathways in cell types receiving microenvironmental signals. Despite its limitations, CellComm found multiple unknown, highly conserved regulators of HSPC emergence, thereby establishing its utility for hypothesis generation from large-scale single-cell datasets and providing important insights into networks that drive cell fate. CellComm's capabilities to infer cell–cell interactions from spatial transcriptomics data and our extensive comparative analysis to related algorithms demonstrate that CellComm provides an orthogonal, innovative and high-resolution approach to infer intercellular communication from single-cell data (Extended Data Figs. 9 and 10, and Supplementary Notes 5 and 6).

The work described here sheds light on fundamental principles underlying the robustness of haematopoietic development. We found that cells in the AGM microenvironment expressed several ligands that can interact with a defined set of cell surface receptors in the target population, suggestive of both compensatory and complementary mechanisms of action. Moreover, reconstruction of downstream signalling pathways from ligand–receptor interactions ultimately leads to a core gene regulatory circuitry controlling hallmarks of HSPC identity, indicative of functional redundancy that likely confers resilience to dynamic changes in the embryonic environment. Finally, unique signalling mechanisms acting through non-convergent signalling pathways account for combinatorial gene expression regulation and fine-tuning of cell fates during haematopoietic development. Together, these observations highlight the substantial challenges posed by single gene loss of function analysis, and may explain aspects of previously conflicting data in the field of developmental haematopoiesis. Such complexity provides additional opportunities to study developmental biology in vitro and direct the differentiation of hiPSCs into target cell types.

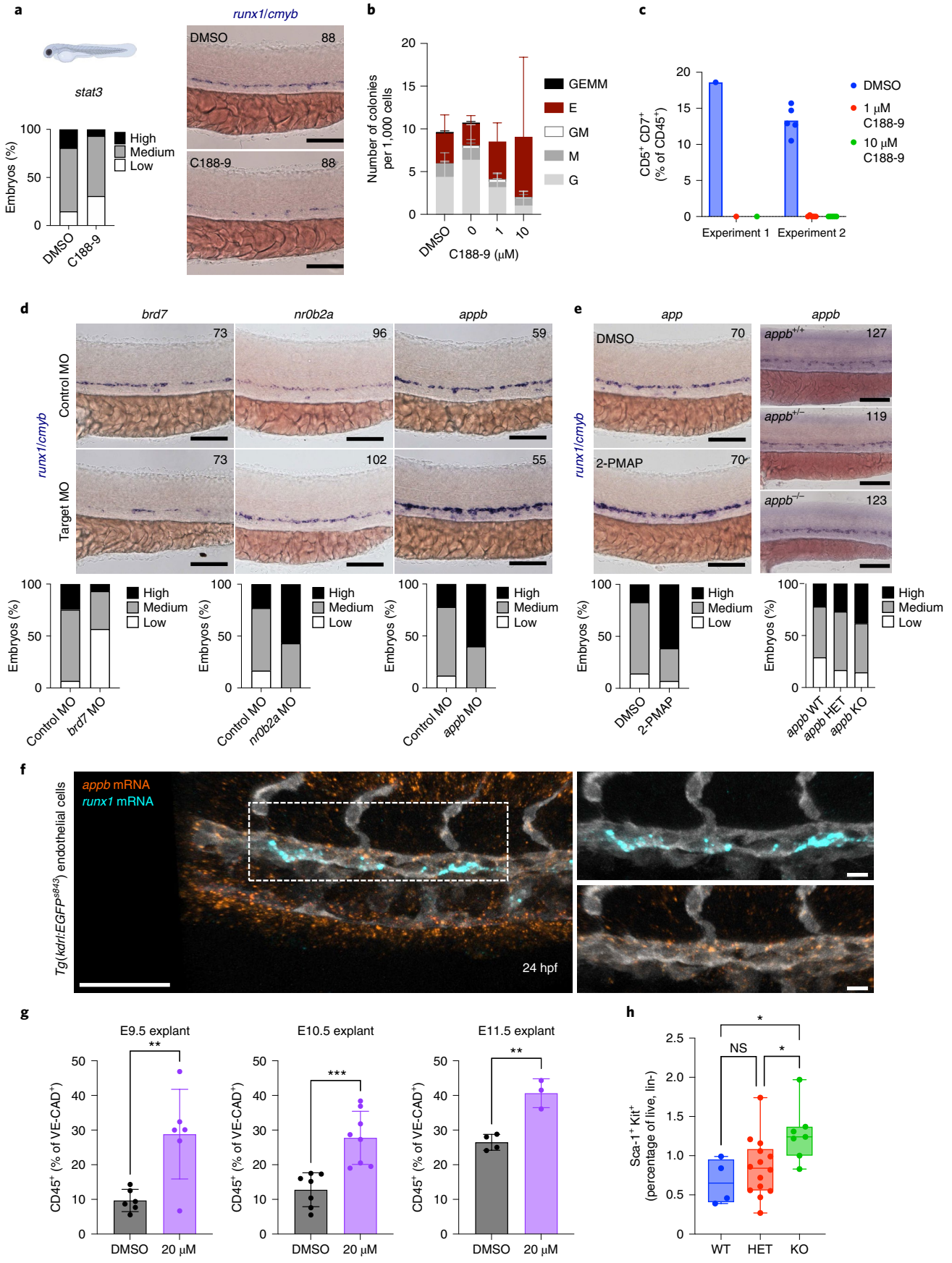
Surprisingly, CellComm predicted a role in EHT for *App*, a gene extensively studied in the context of neurodegenerative disease but never implicated in haematopoietic development. Beyond the clear links with neuronal plaque formation, APP protein is implicated in iron export and homoeostasis. Translation of APP itself is regulated by intracellular iron levels[31] and iron homoeostasis is known to be critical for HSC maintenance and self-renewal[32]. Given our findings that APP inhibition increases HSPC formation in zebrafish and mouse embryos, further exploration of how APP normally restricts haematopoietic fate from ECs may reveal strategies to improve de novo HSPC generation from iPSCs.

The precise recapitulation of developmental pathways to drive differentiation in PSCs remains a major challenge in stem cell engineering. Systems biology approaches have been critical to comprehensively understand and enhance stem cell differentiation, largely by focusing on cell-intrinsic regulatory mechanisms[33–36]. CellComm provides a roadmap for understanding cell communication and differentiation within complex cellular ecosystems by shedding light on cell-extrinsic factors (ligands or receptors) and their coupling to intrinsic signalling pathways and regulatory networks controlling cell fate.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of

Fig. 6 | Systematic validation of transcriptional regulators. a, WISH for *runx1/cmyb* in control (DMSO) and stat3 inhibitor (C188-9)-treated embryos at 36 hpf. Scale bars, 100 µm. $N = 88$ embryos per condition, from three biological replicates (clutches). **b**, CFU potential of cells exposed to Stat3 inhibitor during EHT. G, granulocyte; M, macrophage; E, erythroid; GM, granulocyte and macrophage; GEMM, granulocyte, erythroid, macrophage and megakaryocyte. Bar graphs represent mean ± s.d. $N = 3$ independent differentiation experiments. **c**, Decrease in T-cell progenitor formation in the presence of STAT3 inhibition. $N = 2$ independent experiments, experiment 1 with one culture per condition and experiment 2 with five cultures per condition. Bar graphs represent mean. **d**, WISH for *runx1/cmyb* in control/experimental morpholino-injected embryos at 36 hpf. Scale bars, 100 µm. Number of analysed embryos per condition provided in upper right corner. **e**, WISH for *runx1/cmyb* in DMSO/app inhibitor-treated embryos or *appb* wt/het/ko embryos at 36 hpf. Scale bars, 100 µm. Number of analysed embryos per condition provided in upper right corner. **f**, Maximum intensity projection of a confocal *z* stack of double FISH for zebrafish *appb* and *runx1* mRNA at 24 hpf. Vasculature is visualized with GFP antibody staining in transgenic embryos. Scale bars, 10 µm. **g**, Increase in $CD45^+$ $VE\text{-}Cad^+$ cells in APP inhibitor (2-PMAP)-treated E9.5, E10.5 and E11.5 explant cultures compared with DMSO control. Bars represent mean ± s.d. Unpaired *t* test, E9.5: $N = 6$ explant cultures per condition, $^{**}P = 0.0056$, E10.5: $N = 7$ (DMSO) and $N = 8$ (APP inhibitor) explant cultures, $^{***}P = 0.0007$, E11.5: $N = 4$ (DMSO) and $N = 3$ (APP inhibitor) explant cultures, $^{**}P = 0.0021$ (E11.5). **h**, Increase in lineage negative (lin-), $Sca\text{-}1^+$, Kit^+ cells in E14.5 foetal liver (FL) from *App* KO embryos compared with littermate controls. $N = 4$ (wild type, WT), $N = 14$ (HET) and $N = 7$ (knockout, KO). Embryos are from three litters. Mean, 25th–75th percentiles (box) and minimum and maximum (whiskers) values are indicated. The *P* values from one-way ANOVA are adjusted for multiple comparison using Tukey's multiple comparison test. NS, not significant. $P = 0.034$ (WT versus KO) and $P = 0.0462$ (HET versus KO). Source numerical data are available in Source data.


a
stat3
runx1/cmyb
DMSO
88
C188-9
88
Embryos (%)
High
Medium
Low
DMSO
C188-9
b
Number of colonies per 1,000 cells
GEMM
E
GM
M
G
DMSO
0
1
10
C188-9 (μM)
c
CD5+ CD7+ (% of CD45+)
DMSO
1 μM C188-9
10 μM C188-9
Experiment 1
Experiment 2
d
brd7
nr0b2a
appb
runx1/cmyb
Control MO
Target MO
73
96
59
73
102
55
Control MO
brd7 MO
nr0b2a MO
appb MO
e
app
appb
DMSO
2-PMAP
70
70
appb+/+
appb+/−
appb−/−
127
119
123
appb WT
appb HET
appb KO
f
Tg(kdrl:EGFPs843) endothelial cells
appb mRNA
runx1 mRNA
24 hpf
g
E9.5 explant
E10.5 explant
E11.5 explant
CD45+ (% of VE-CAD+)
DMSO
20 μM
h
Sca-1+ Kit+ (percentage of live, lin-)
NS
WT
HET
KO

author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00884-1.

Received: 7 December 2021; Accepted: 1 March 2022;
Published online: 12 April 2022

References

1. Baron, C. S. et al. Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat. Commun.* **9**, 2517 (2018).
2. Zhou, F. et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487–492 (2016).
3. Hou, S. et al. Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses. *Cell Res.* **30**, 376–392 (2020).
4. Gao, P. et al. Transcriptional regulatory network controlling the ontogeny of hematopoietic stem cells. *Genes Dev.* **34**, 950–964 (2020).
5. Zhu, Q. et al. Developmental trajectory of pre-hematopoietic stem cell formation from endothelium. Blood https://doi.org/10.1182/blood.2020004801 (2020).
6. Taoudi, S. et al. Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin^{+}CD45^{+} pre-definitive HSCs. *Cell Stem Cell* **3**, 99–108 (2008).
7. Rybtsov, S. et al. Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J. Exp. Med.* **208**, 1305–1315 (2011).
8. Lummertz da Rocha, E. et al. Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* **9**, 892 (2018).
9. Nguyen, P. D. et al. Haematopoietic stem cell induction by somite-derived endothelial cells controlled by meox1. *Nature* **512**, 314–318 (2014).
10. McGarvey, A. C. et al. A molecular roadmap of the AGM region reveals BMPER as a novel regulator of HSC maturation. *J. Exp. Med.* **214**, 3731–3751 (2017).
11. Oatley, M. et al. Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition. *Nat. Commun.* **11**, 586 (2020).
12. Khurana, S. et al. Outside-in integrin signalling regulates haematopoietic stem cell function via Periostin–Itgav axis. *Nat. Commun.* **7**, 13500 (2016).
13. Ransom, D. G. et al. Characterization of zebrafish mutants with defects in embryonic hematopoiesis. *Development* **123**, 311–319 (1996).
14. North, T. E. et al. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).
15. Huang, H.-T. et al. A network of epigenetic regulators guides developmental haematopoiesis in vivo. *Nat. Cell Biol.* **15**, 1516–1525 (2013).
16. Rho, S.-S. et al. Rap1b promotes Notch-signal-mediated hematopoietic stem cell development by enhancing integrin-mediated cell adhesion. *Dev. Cell* **49**, 681–696.e6 (2019).
17. Ditadi, A. et al. Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).
18. Umemoto, T. et al. CD61 enriches long-term repopulating hematopoietic stem cells. *Biochem. Biophys. Res. Commun.* **365**, 176–182 (2008).
19. Huang, K. et al. Generation and analysis of GATA2 human ESCs reveal ITGB3/CD61 as a reliable marker for defining hemogenic endothelial cells during hematopoiesis. *Stem Cell Rep.* **7**, 854–868 (2016).
20. Mariani, S. A. et al. Pro-inflammatory aorta-associated macrophages are involved in embryonic development of hematopoietic stem cells. *Immunity* **50**, 1439–1452 (2019).
21. Frame, J. M. et al. Metabolic regulation of inflammasome activity controls embryonic hematopoietic stem and progenitor cell production. Dev. Cell https://doi.org/10.1016/j.devcel.2020.07.015 (2020).
22. Clements, W. K. et al. A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. *Nature* **474**, 220–224 (2011).
23. Souilhol, C. et al. Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells. *Nat. Commun.* **7**, 10784 (2016).
24. Francis, K. & Palsson, B. O. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc. Natl Acad. Sci. USA.* **94**, 12258–12262 (1997).
25. Linnekin, D. Early signaling pathways activated by c-Kit in hematopoietic cells. *Int. J. Biochem. Cell Biol.* **31**, 1053–1074 (1999).
26. Shu, R. et al. APP intracellular domain acts as a transcriptional regulator of miR-663 suppressing neuronal differentiation. *Cell Death Dis.* **6**, e1651 (2015).
27. Stehling-Sun, S., Dade, J., Nutt, S. L., DeKoter, R. P. & Camargo, F. D. Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor Mef2c. *Nat. Immunol.* **10**, 289–296 (2009).
28. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* **7**, 105–117 (2007).
29. Liang, J. et al. The C-kit receptor-mediated signal transduction and tumor-related diseases. *Int. J. Biol. Sci.* **9**, 435–443 (2013).
30. Xu, K. et al. Integrating ChIP-sequencing and digital gene expression profiling to identify BRD7 downstream genes and construct their regulating network. *Mol. Cell. Biochem.* **411**, 57–71 (2016).
31. Lumsden, A. L. et al. Dysregulation of neuronal iron homeostasis as an alternative unifying effect of mutations causing familial Alzheimer's disease. *Front. Neurosci.* **12**, 533 (2018).
32. Muto, Y., Nishiyama, M., Nita, A., Moroishi, T. & Nakayama, K. I. Essential role of FBXL5-mediated cellular iron homeostasis in maintenance of hematopoietic stem cells. *Nat. Commun.* **8**, 16114 (2017).
33. Cahan, P. et al. CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
34. Kinney, M. A. et al. A systems biology pipeline identifies regulatory networks for stem cell engineering. *Nat. Biotechnol.* **37**, 810–818 (2019).
35. Ng, A. H. M. et al. A comprehensive library of human transcription factors for cell fate engineering. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-0742-6 (2020).
36. Morris, S. A. et al. Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell* **158**, 889–902 (2014).

Methods

Mice. Murine housing, maintenance and dissection studies were performed in accordance with guidelines set forth by the Boston Children's Hospital Institutional Animal Care and Use Committee (IACUC study no. 000014940). C57Bl/6 mice (stock no. 000664) and APP knockout mice (B6.129S7-*Apptm1Dbo*/J, stock no. 004133) were purchased from Jackson Laboratory and maintained in a standard barrier facility at housing room conditions of 20–23 °C and 35–70% humidity with a 12 h light–12 h dark cycle and ad libitum food and water. Both male and female mice were used in this study.

Embryo isolation and sequencing. Embryos were collected at E10.25 (66 embryos) and somites were counted, only processing embryos between 28 and 32 somite pairs. AGM regions were dissected, pooled and dissociated following Chen et al.[37]. Cells were filtered (40 µm filter) and stained in PBS/2% FBS on ice using antibodies listed in Supplementary Table 2. Cells were sorted at 15 psi pressure using a 100 µm nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter). Sorted cells were collected into 1.5 ml Eppendorf tubes containing 300 µl PBS/50% FBS. Importantly, both enriched cell populations (ECs/T1preHSCs) and the whole AGM were processed similarly by FACS to deplete red blood cells and ensure that only single, live cells were used for sequencing. Immediately following sorting, each cell population was divided across two lanes on a 10x Genomics scRNA-seq chip as independent technical replicates to encapsulate cells into droplets. scRNA-seq libraries were then prepared per the 10x scRNA-seq v2 protocol. Cells were loaded into 10x lanes at cell concentrations to maintain an estimated doublet rate below 5%. Final 10x libraries were assayed via an Agilent High Sensitivity dsDNA Bioanalyzer, normalized, pooled and shallow sequenced on a MiniSeq, identifying 10,000 high-confidence cell barcodes in total across all conditions. 10x libraries were then renormalized per the distribution of reads/library from the MiniSeq run and deep sequenced on a NovaSeq S4 to a depth of 50,000 reads per cell barcode.

scRNA-seq data processing and analysis. Single-cell sequencing data were aligned and quantified using the Cell Ranger 2.2.0 Single-Cell Software Suite against the mm10 mouse reference genome (version mm10-1.2.0) provided by Cell Ranger. Cells with fewer than five detected genes were removed. Genes expressed in fewer than five cells were also removed. After quality control, our dataset was composed of 9,492 single cells with a median of 2,488 genes detected per cell. Of the 9,492 cells, 268 cells come from the T1preHSC gate, 2,395 cells from the EC gate and 6,829 cells from the random sampling of the AGM niche. We used CellRouter to perform quality control and dimension reduction, identify gene signatures and reconstruct the EHT differentiation trajectory.

Cell type annotation. We used marker genes of ECs (*Cdh5* and *Kdr*), T1preHSCs/HSCs (*Hlf*, *Procr* and *Runx1*), sclerotome (*Meox1*, *Pax1* and *Pax9*), dermomyotome (*Meox1*, *Pax3* and *Pax7*), myoblasts (*Myf5* and *Myod1*), UGRs (*Wt1* and *Six2*), mesonephros (*Gata3* and *Cited1*), neural tube (*Sox1* and *Sox2*), neural crest (*Foxd3* and *Sox10*), motor neurons (*Robo2* and *Slit3*), hindbrain (*Lhx5* and *Neurog1*), macrophages (*Cx3cr1* and *Ccl3*) and mesenchymal stem and progenitor cells (*Pdgfra*, *Tbx3* and *Hoxc10*) to catalogue the diversity of cell types constituting the AGM region. Interestingly, germ cells (*Pou5f1*, *Kit*, *Pecam1* and *Sox2*) were captured in the EC FACS gate, while cluster 14 was an undefined cell type (Fig. 1b,c).

CellComm. CellComm is described in Supplementary Note 1. A step-by-step protocol to apply CellComm to spatial transcriptomics data is available in Nature Protocol Exchange[38].

Trajectory analysis and gene regulatory network inference. Trajectory analysis was performed using CellRouter as previously described. For cell cycle analysis, we calculated signature scores using a curated list of G2M and S genes, and plotted these signature scores along the CellRouter trajectory. We used our previously developed algorithm, the Context of Likelihood of Relatedness, especially the modified version published with CellNet[33,39], to reconstruct gene regulatory networks from the data.

Pathway enrichment analysis. We used the package clusterProfiler with standard parameters. Briefly, as input we used either ligands or receptors identified by CellComm or the regulons of each predicted transcriptional regulator. When performing Gene Ontology analysis with clusterProfiler, the function enrichGO will use all genes in the database as a background. The number of genes used as input varies depending on the number of ligand–receptor pairs identified, or the size of the associated regulons of each transcriptional regulator. We consider pathways with adjusted P values <0.05 statistically significant, and those are shown in figures.

hiPSC culture and differentiation. HE differentiation from hiPSCs was induced after one passage on murine embryonic fibroblasts before embryoid body (EB) formation. Colonies were detached using collagenase IV and seeded into low-attachment dishes (Corning) in aggregation medium[40]. Twenty-four hours later, basic fibroblast growth factor (bFGF; 5 ng ml^{-1}) was added. On day 2, medium was changed to aggregation medium supplemented with bFGF (5 ng ml^{-1}), BMP4 (10 ng ml^{-1}), CHIR99021 (3 µM) and SB431542 (6 µM). On day 3, medium was changed to EB medium (Stempro-34; Thermo Fisher Scientific) supplemented with ascorbic acid (1 mM), holo-transferrin (150 µg ml^{-1}) and α-monothioglycerol (0.4 mM)). On day 6, medium was changed to EB medium supplemented with bFGF (5 ng ml^{-1}), vascular endothelial growth factor (15 ng ml^{-1}), interleukin (IL)-6 (10 ng ml^{-1}), IL-11 (5 ng ml^{-1}), insulin-like growth factor 1 (25 ng ml^{-1}), stem cell factor (50 ng ml^{-1}) and erythropoietin (2 U ml^{-1}). EBs were collected on day 8 and dissociated first with trypsin/EDTA, then with collagenase IV. HE was enriched by magnetic-activated cell sorting for CD34$^+$ cells (Miltenyi). HE was plated onto Matrigel (Thermo Fisher Scientific) coated dishes in EHT medium (EB medium + BMP4 (10 ng ml^{-1}), bFGF (5 ng ml^{-1}), IL-3 (30 ng ml^{-1}), IL-6 (10 ng ml^{-1}), IL-11 (5 ng ml^{-1}), insulin-like growth factor 1 (25 ng ml^{-1}), vascular endothelial growth factor (5 ng ml^{-1}), stem cell factor (100 ng ml^{-1}), erythropoietin (2 U ml^{-1}), thrombopoietin (30 ng ml^{-1}), FMS-like tyrosine kinase 3 ligand (10 ng ml^{-1}), Sonic hedgehog (20 ng ml^{-1}), angiotensin II (10 µg ml^{-1}) and losartan (100 µM). Throughout EB and EHT differentiation, cells were kept at 5% CO_2/5%O_2/90% N_2 environment. YBX1 CRISPRi lines have been generated following the protocol provided in ref. [41].

T-cell differentiation. Magnetic-activated cell sorting purified, hiPSC derived-CD34$^+$ haemogenic ECs (1–1.5 × 10^4) were seeded on individual wells of a 96-well plate and induced to differentiate into pro-T cells using StemSpan T-cell generation kit (#09940, StemCell Technologies) following the manufacturer's instructions. Cells were treated with DMSO or stat3 inhibitor (1 µM and 10 µM) on day 1 of differentiation, and CD5$^+$ CD7$^+$ pro-T cells were detected after 2 weeks of differentiation.

CFU assay. At d8 + 7 of hiPSC-derived EHT culture, floating cells were collected and between 1,000 and 5,000 cells were plated into serum-free methylcellulose-based medium with recombinant cytokines (SF H4636) on 3.5 cm dishes in a humidified chamber inside the incubator. Colonies were scored blindly after 12–14 days.

Mouse embryo explant culture. Caudal parts of E9.5–E11.5 embryos from wild-type C57Bl/6 mice (The Jackson Laboratory, stock no. 000664) harbouring the para-aortic splanchnopleura or AGM region were dissected and cultured for 72 h on Durapore membrane filter paper at the air–liquid interface. Filter paper was floating on top of MyeloCult M5300 medium supplemented with 10 µM hydrocortisone succinate and penicillin–streptomycin.

Foetal liver flow analysis. Homozygous APP knockout mice (The Jackson Laboratory, stock no. 004133) were mated with wild-type C57Bl/6 mice (The Jackson Laboratory, stock no. 000664) to obtain heterozygous animals. For foetal liver dissection, timed matings from both heterozygous female and male mice were performed and pregnant females were sacrificed on E14.5. Foetal livers were collected and dissociated into a single-cell suspension by passing them through a 23 G needle. Red blood cell lysis was performed for 20 min on ice followed by a wash and antibody staining (for antibodies, see Supplementary Table 2).

Zebrafish husbandry. Zebrafish were maintained in accordance with Beth Israel Deaconess Medical Center and Boston Children's Hospital Institutional Animal Care and Use Committee protocols (protocol number 00001259). Translation-blocking morpholinos were obtained from Gene Tools, and injected into one-cell-stage zebrafish oocytes (for morpholino sequences and injection amounts, see Supplementary Table 3). Chemical treatments were performed in six-well plates, from 14 hpf to 36 hpf using ≤30 embryos per well, with the indicated amount of compound administered in 5 ml E3 fish medium. Whole-mount in situ hybridization (WISH) analysis of *runx1* and *cmyb* expression was performed using published probes and protocols. For fluorescence in situ hybridization (FISH), endothelial enhanced green fluorescent protein (EGFP) in Tg(kdrl:EGFP)s843 embryos was detected with rabbit anti-green fluorescent protein (GFP) antibody and goat anti-rabbit AF647 secondary (Supplementary Table 2). FACS analysis of HSPC numbers was performed at 48 hpf in embryos carrying the *Tg(kdrl:mCherry)s916* and *Tg(cmyb:EGFP)zf169* transgenes. In brief, double-transgenic embryos (five embryos per sample, with a minimum eight biological replicates from at least two experiments) were dissociated in PBS containing 0.5 mg ml^{-1} Liberase (Roche) at 34 °C for 2 h with gentle agitation and manual pipetting. Filtered cells were resuspended in 1× PBS, labelled with 5 nm SYTOX Red (Life Technologies) as a live/dead stain, and the number of live GFP$^+$/mCherry$^+$ cells quantified using a BD LSRFortessa cell analyser and FlowJo 10.4.2 software (Tree Star).

Statistics and reproducibility. Flow cytometry, quantitative PCR with reverse transcription (RT–qPCR) and cell culture experiments were analysed with Prism9 (version 9.3.1) software (GraphPad). Figure legends indicate the statistical test used in each experiment. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. Where indicated, investigators were blinded to allocation during experiments and outcome assessment. For *runx1*/*cmyb* in situ hybridization experiments in

zebrafish embryos, at least 20 embryos from a minimum of two independent clutches were analysed per treatment condition and/or genotype, with scoring independently confirmed. Stained embryos were compared with stage-matched sibling or timed-mating controls, across multiple replicate clutches. Expression intensity was qualitatively categorized as high, medium or low, relative to the median staining pattern of the control group for a given clutch. At 36 hpf, for *runx1*/*cmyb* expression, a score of 'high' reflects continuous aortic signal with a consistent thickened pattern, reflective of numerous cell clusters or budding cells; 'medium' denotes a pattern of distinct cell budding/clusters, with continuous aortic signal; and 'low' indicates a spotty/ discontinuous aortic staining pattern. Distribution graphs represent combined totals for each expression category from all embryos scored, across clutches; images were selected from matched embryos within a single experimental replicate to be representative of the most abundant expression pattern for each treatment or condition.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The processed and raw scRNA-seq data supporting the findings of this study have been deposited in the Gene Expression Omnibus under accession code GSE160526. We re-analysed publicly available datasets from the following accession numbers: GSE137117 and GSE144240. Reference genome data for scRNA-seq analyses were downloaded from https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/2.2. All other data supporting the findings of this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

Code availability

CellComm and CellRouter are available through the Framework for Unified Single-Cell Analysis (FUSCA) R package available at https://github.com/edroaldo/fusca. We also provide a web resource to run CellComm and other algorithms at http://hematopoieticniches.com.

References

37. Chen, M. J. et al. Transcriptome dynamics of hematopoietic stem cell formation revealed using a combinatorial Runx1 and Ly6a reporter system. *Stem Cell Rep.* **14**(5), 956–971 (2020).
38. Lummertz da Rocha, E. et al. Prediction of intercellular communication networks using CellComm. Protoc. Exch. https://doi.org/10.21203/rs.3.pex-1843/v1 (2022).
39. Faith, J. J. et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
40. Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 554–561 (2014).
41. Mandegar, M. A. et al. CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* **18**, 541–553 (2016).

Acknowledgements

This work was supported by the following grants from the National Institutes of Health: RC2DK120535 (G.Q.D. and T.E.N.), U01HL134812 (G.Q.D. and T.E.N.), R24DK092760 (G.Q.D. and T.E.N.), R01DK098241 (T.E.N.), F32DK122715 and K01DK129409 (W.W.S.). E.L.d.R. thanks Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil. M.F. is supported by a PhD fellowship from the Brazilian National Council for Scientific and Technological Development (CNPq). C.K. was supported by a postdoctoral fellowship from the German Research Foundation (DFG; grant KU 3580/1-1).

Author contributions

E.L.d.R., C.K. and G.Q.D. conceived and designed the project. E.L.d.R. designed and developed CellComm. E.L.d.R. and C.K. designed experiments to sequence the AGM niche. C.K. and M.A.N. generated the scRNA-seq dataset. E.L.d.R. processed the raw scRNA-seq and analysed all data described. C.K. designed and performed mouse and iPSC experiments, with A.M. W.W.S. designed and performed all zebrafish experiments, with Z.C.L. R.J. performed some T-cell differentiation experiments. R.d.S.P. created the R package with CellComm and the updated version of CellRouter. M.F. wrote the protocol for Protocol Exchange. E.L.d.R., C.K. and W.W.S. wrote the paper. J.J.C. provided critical feedback on the manuscript. G.Q.D. and T.E.N. oversaw the project, critiqued experiments and revised the paper.

Competing interests

G.Q.D. and Boston Children's Hospital have filed provisional patent applications (US11261430B2, US20220049221A1) covering derivation of blood cells from pluripotent stem cell sources. J.J.C. is an academic co-founder and board member of Cellarity. The other authors declare no competing interests.

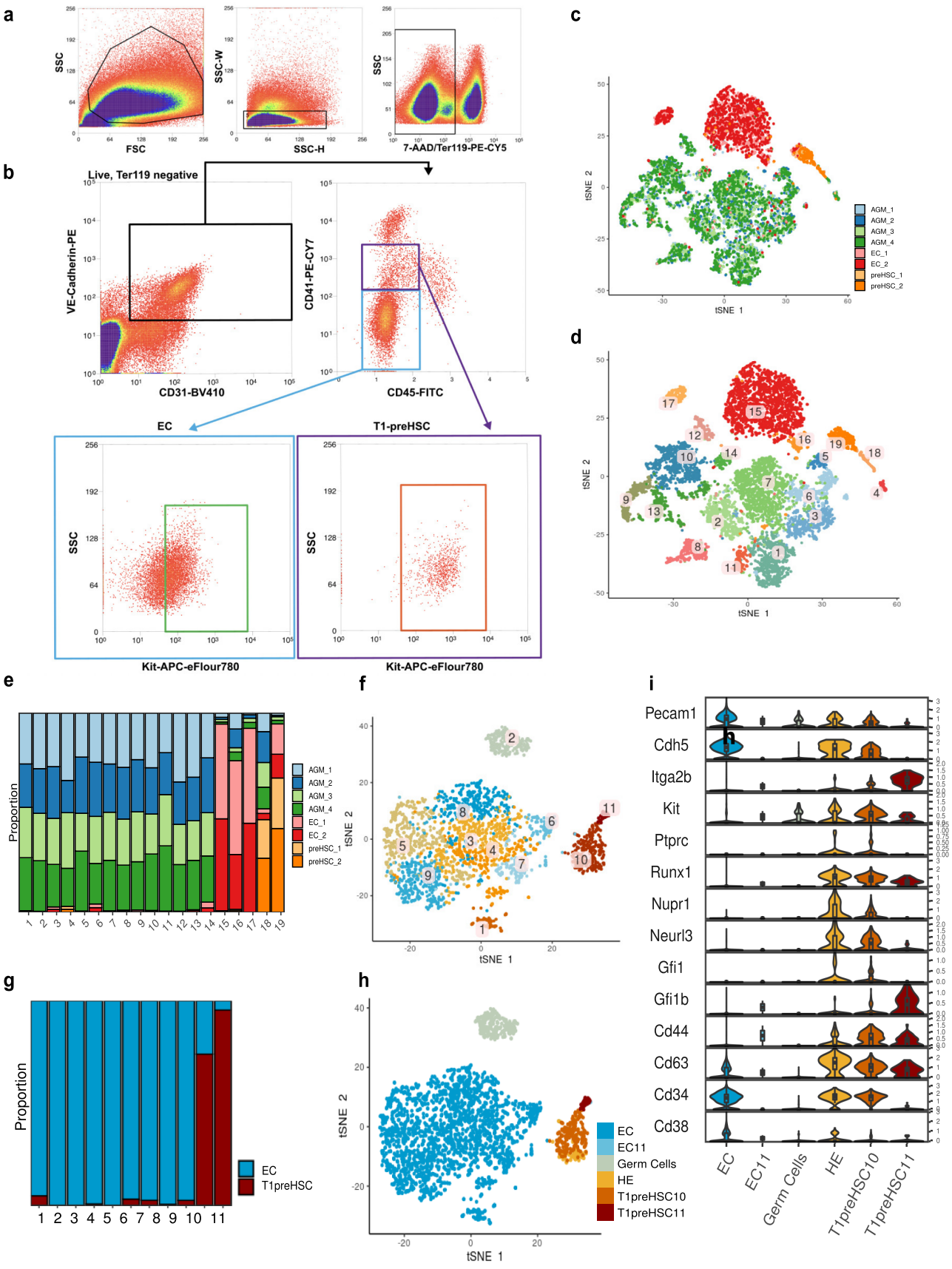
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00884-1.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00884-1.

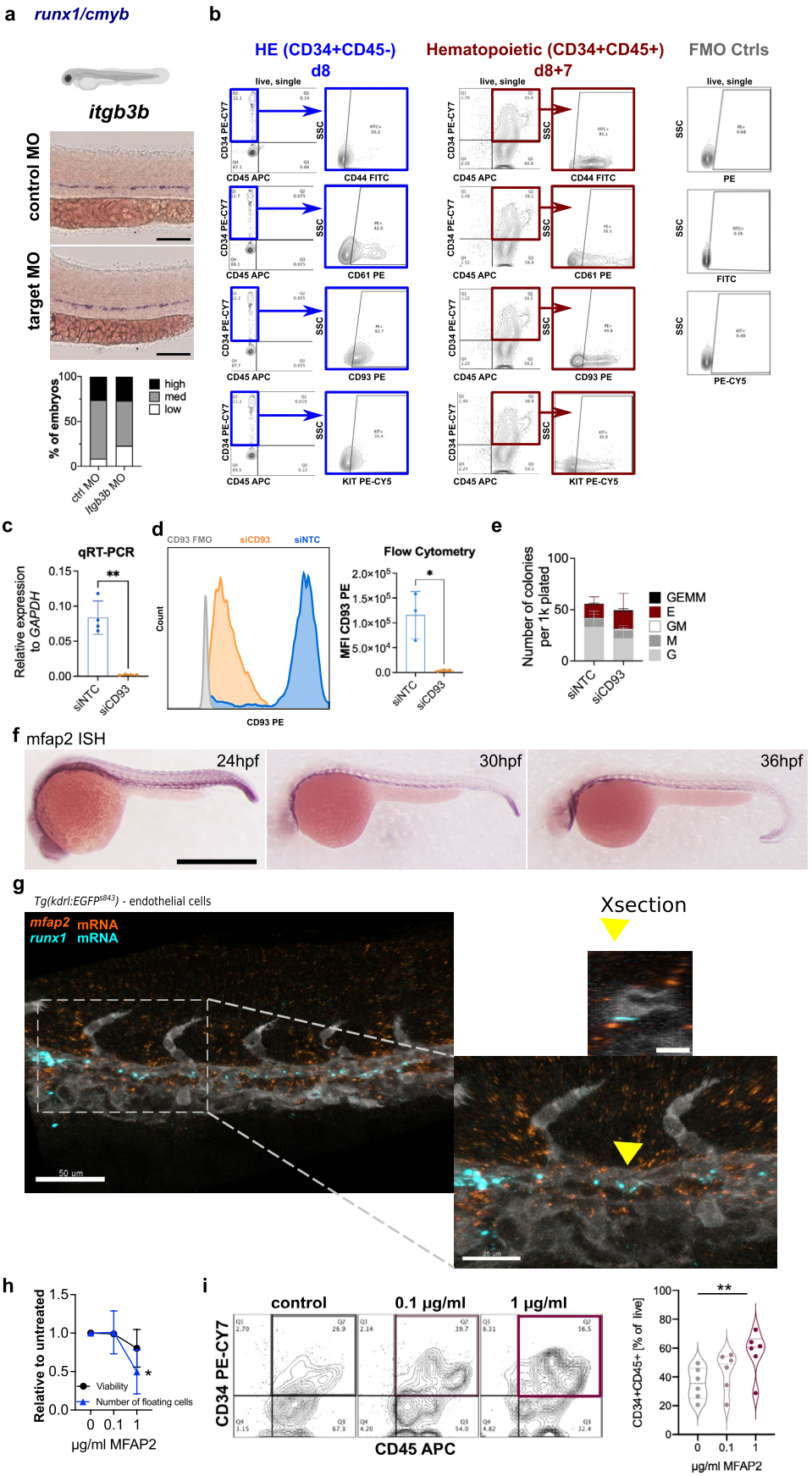
Correspondence and requests for materials should be addressed to Trista E. North or George Q. Daley.

Peer review information *Nature Cell Biology* thanks Fuchou Tang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



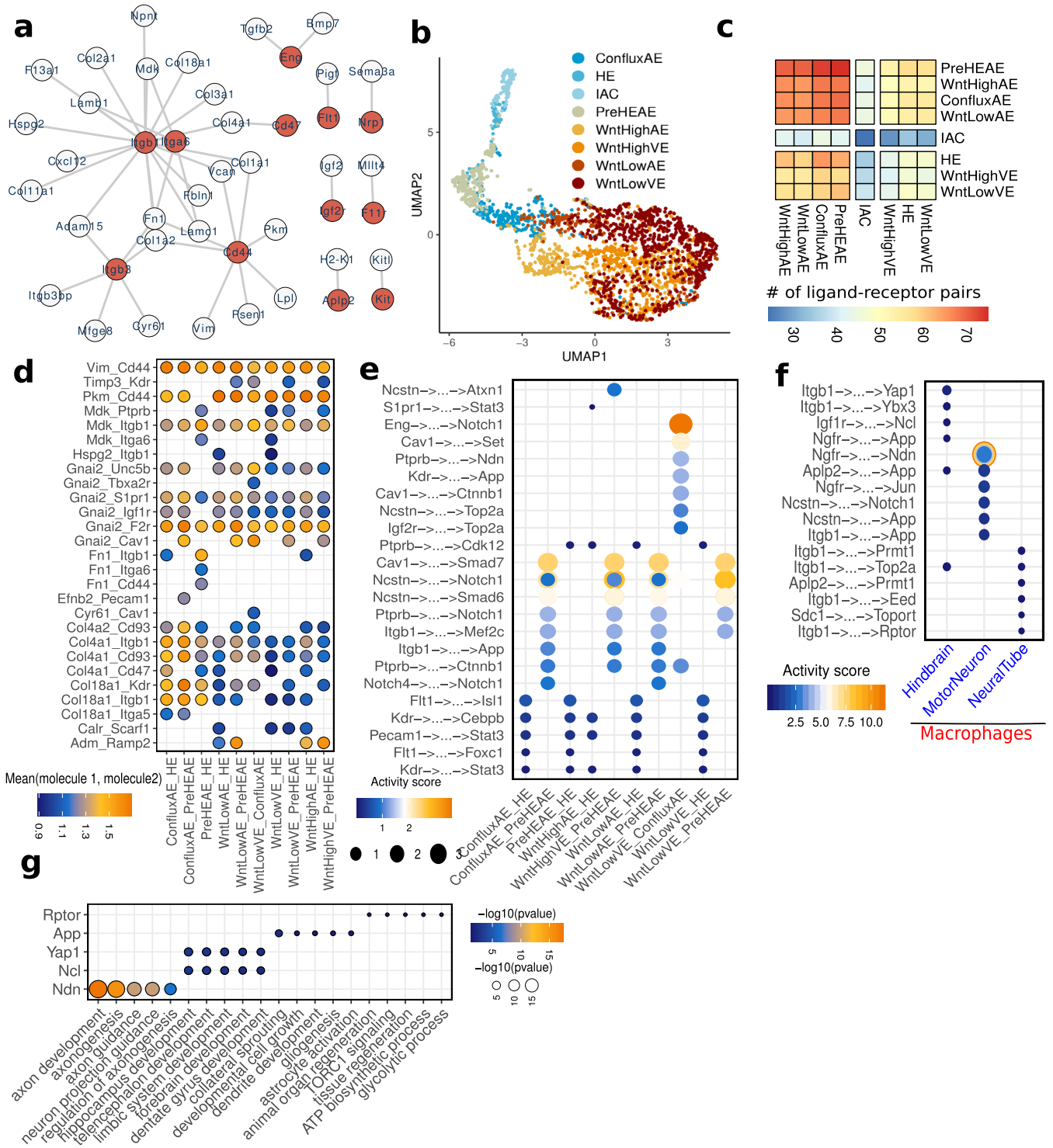
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Gating strategy and Identification of the hemogenic endothelium population. a, Single, live, non-red blood cells (7-AAD^{negative}TER119^{negative}) were sampled. **b**, Gating strategy for isolation of endothelial cells (EC) and type 1 preHSCs (T1preHSCs). Cells gated in A were further gated on VE-Cadherin^{+}CD31^{+}CD45^{negative}CD41^{negative}KIT$^{low/+}$ for EC and VE-Cadherin^{+}CD31^{+}CD45^{negative} CD41^{low}KIT$^{low/+}$ for T1-preHSC. **c**, tSNE analysis of the entire AGM ecosystem data color-coded by sample replicates. **d**, tSNE analysis of the entire AGM ecosystem data color-coded by transcriptional clusters. **e**, Proportion of replicates in each transcriptional cluster. **f**, tSNE analysis of sorted endothelial cells (ECs) and immunophenotypic T1preHSCs colored-coded by transcriptional clusters. **g**, Proportion of sorted cells (ECs, T1reHSCs) across transcriptional clusters. **h**, Manual cluster annotation based on transcriptional clusters and the sample of origin. **i**, Gene expression distribution of selected EC, HE (hemogenic endothelium), and T1preHSC genes. The middle line in the box plot indicates the median. The lower and upper hinges correspond to the 25th and 75th percentiles. Whiskers show min to max. Data beyond the whiskers are 'outlying' points and are plotted individually. Source numerical data are available in source data.

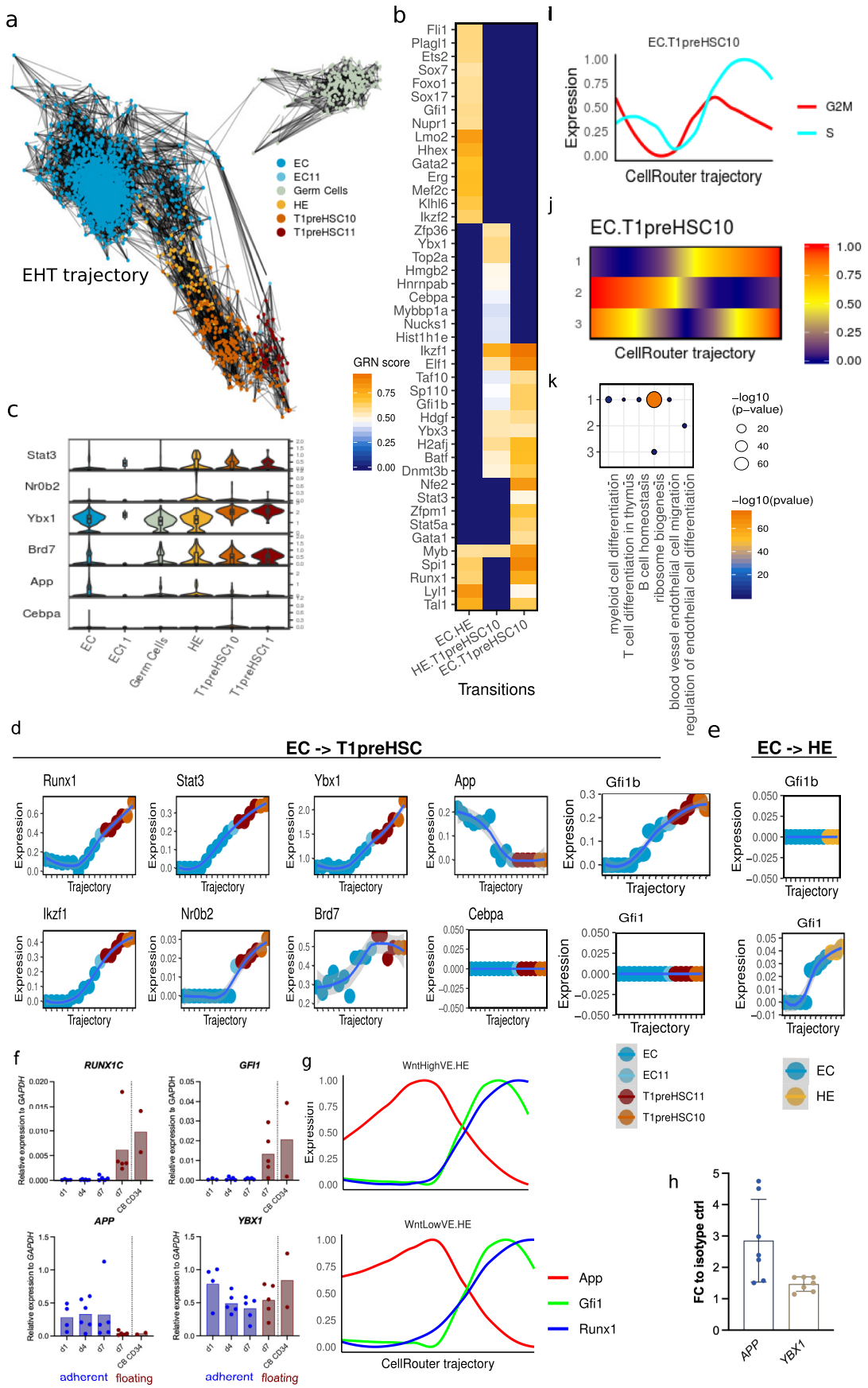


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Experimental validation of ligands and receptors. a, Whole-mount *in situ* hybridization (WISH) for *runx1/cmyb* in control and *itgb3b* morpholino-injected embryos at 36 hpf. Scale bar, 100 µm, N = 23 embryos (control) and N = 26 embryos (*itgb3b* morpholino). **b**, Representative flow plots for tested surface receptors on hemogenic endothelial cells from hiPSCs (HE, $CD34^+CD45^-$, day 8) and Hematopoietic cells ($CD34^+CD45^+$, day 8 + 7), pre gated on live, single cells. **c**, qPCR confirming significant reduction of *CD93* transcript 48 h after siRNA transfection. Bars represent mean +/- SD. N = 4 independent EHT cultures. Paired t-test, **p = 0.0065. **d**, Flow cytometric analysis of CD93 surface expression 48 h after siRNA transduction. Bars represent mean +/- SD. N = 3 independent differentiation experiments. Paired t-test, one-tailed, *p = 0.0261. **e**, CFU assay from d7 floating cells after hiPSC-derived EHT with and without CD93 siRNA KD. Bars represent mean +/- SD of N = 6 assays. **f**, WISH timecourse for zebrafish *mfap2* across the window of EHT. Scale bar 500 µm. **g**, Maximum intensity projection of a confocal z-stack of double fluorescent *in situs* for zebrafish *mfap2* and *runx1* mRNA at 24 hpf. Vasculature is visualized with GFP antibody staining in transgenic embryos. Yellow arrowhead indicates position of cross section. Scale bars 50 µm, 25 µm and 10 µm. **h**, Viability and quantification of floating, hematopoietic cells at d8 + 7 of endothelial to hemogenic transition at increasing concentrations of recombinant human MFAP2. Mean +/- SD is plotted, N = 6 independent differentiation experiments. **i**, Representative flow plots (left) and violin plots (median and quartiles) for CD34 and CD45 staining of floating hematopoietic cells after hiPSC derived EHT culture in the presence of increasing recombinant human MFAP2 concentrations (N = 6 independent experiments). RM One-way ANOVA, p-values have been adjusted for multiple comparison using Dunnett's test, **p = 0.0029. Source numerical data are available in source data.

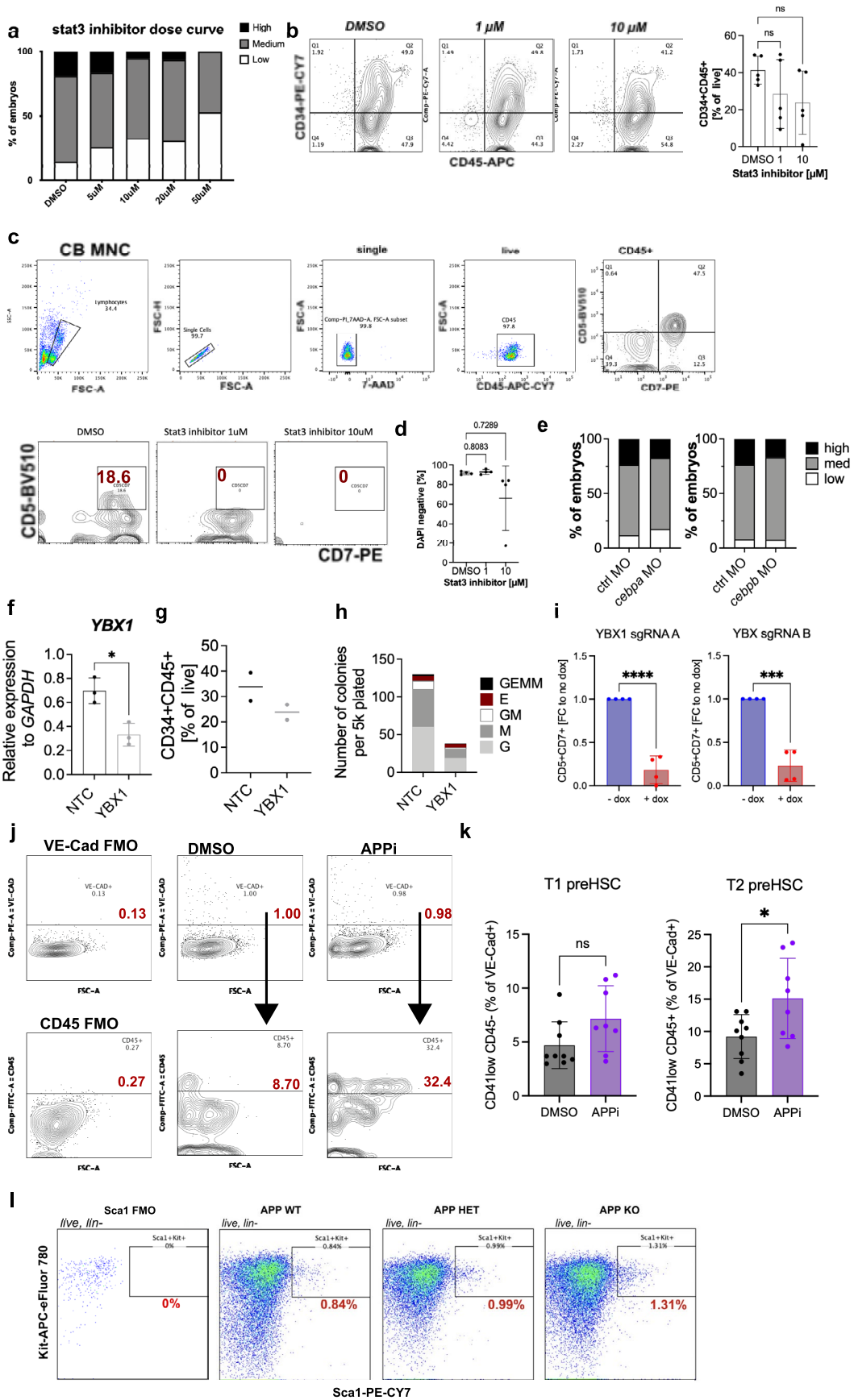


Extended Data Fig. 3 | Application of CellComm to other cell-cell interactions. a, UMAP analysis of scRNA-seq of E + HE + IAC sorted populations from E10.5 mouse embryos. **b**, Inferred intercellular communication network. **c**, Top ligand-receptor pairs across selected cell types prioritized by CellComm. **d**, Reconstructed signaling pathways connecting cell surface receptors to TRs across reported cell-cell interactions. **e**, Predicted downstream signaling pathways from cell surface receptors leading to transcriptional regulators in neuron subtypes upon interaction with macrophages. **f**, Gene ontology enrichment analysis of the regulons of selected transcriptional regulators identified in **e**. The color and size scales both represent the -log10 (P value) of the pathway enrichment result. AE: arterial endothelial cell. VE: ventral endothelial cell. Source numerical data are available in source data.



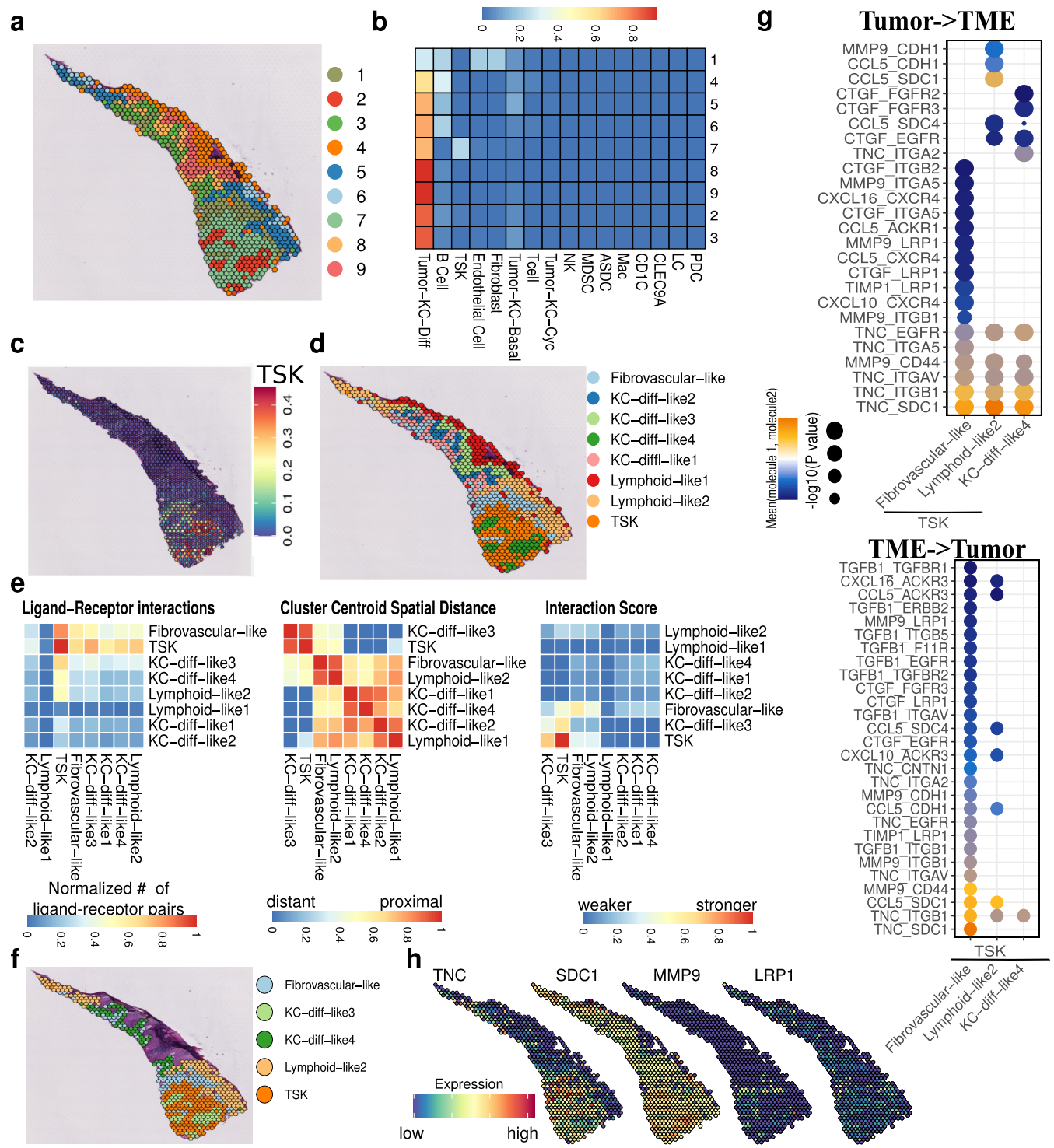
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Further prioritization of transcriptional regulators. a, The endothelial-to-hematopoietic transition trajectory reconstructed by CellRouter using sorted ECs and T1preHSCs. **b**, Transcriptional regulators predicted to be important for the cell state transitions reported in the x-axis, for example EC.HE means the differentiation trajectory from EC to HE. **c**, Gene expression distribution of genes selected for experimental validation. The middle line in the box plot indicates the median. The lower and upper hinges correspond to the 25th and 75th percentiles. Whiskers show min to max. Data beyond the whiskers are 'outlying' points and are plotted individually. **d**, Kinetic profiles of selected genes along the CellRouter trajectory from endothelial cells (ECs) to the T1preHSC10 state. **e**, Kinetic profiles of *Gfi1b* and *Gfi1* along the EC to HE differentiation trajectory. **f**, Expression dynamics by qPCR of hematopoietic transcripts *RUNX1C* and *GFI1* as well as CellComm predicted candidate regulators *APP* and *YBX1*. N = 4 (d1), N = 6 (d4), N = 5 (d7) independent EHT cultures, N = 2 independent CB CD34 + donor samples. Bar graphs represent mean. **g**, Kinetic profiles of selected genes along the CellRouter predicted EHT trajectory using the dataset generated by[1]. **h**, Increase in *APP* and *YBX1* expression upon ITGB3 inhibition in hiPSC-derived HSPCs. N = 6, from 3 independent cultures. Bar graphs represent mean +/- SD. **i**, Cell cycle signature scores along the CellRouter trajectory from EC to the T1preHSC10 state. **j**, Clustered kinetic profiles, which group genes with similar gene expression trends along the reported CellRouter trajectory. **k**, Gene ontology enrichment of gene sets in each kinetic cluster identified in **I**. The color and size scales both represent the -log10 (P value) of the pathway enrichment result. Hypergeometric test. The p-values were adjusted for multiple comparisons using the Benjamini-Hochberg method. Source numerical data are available in source data.

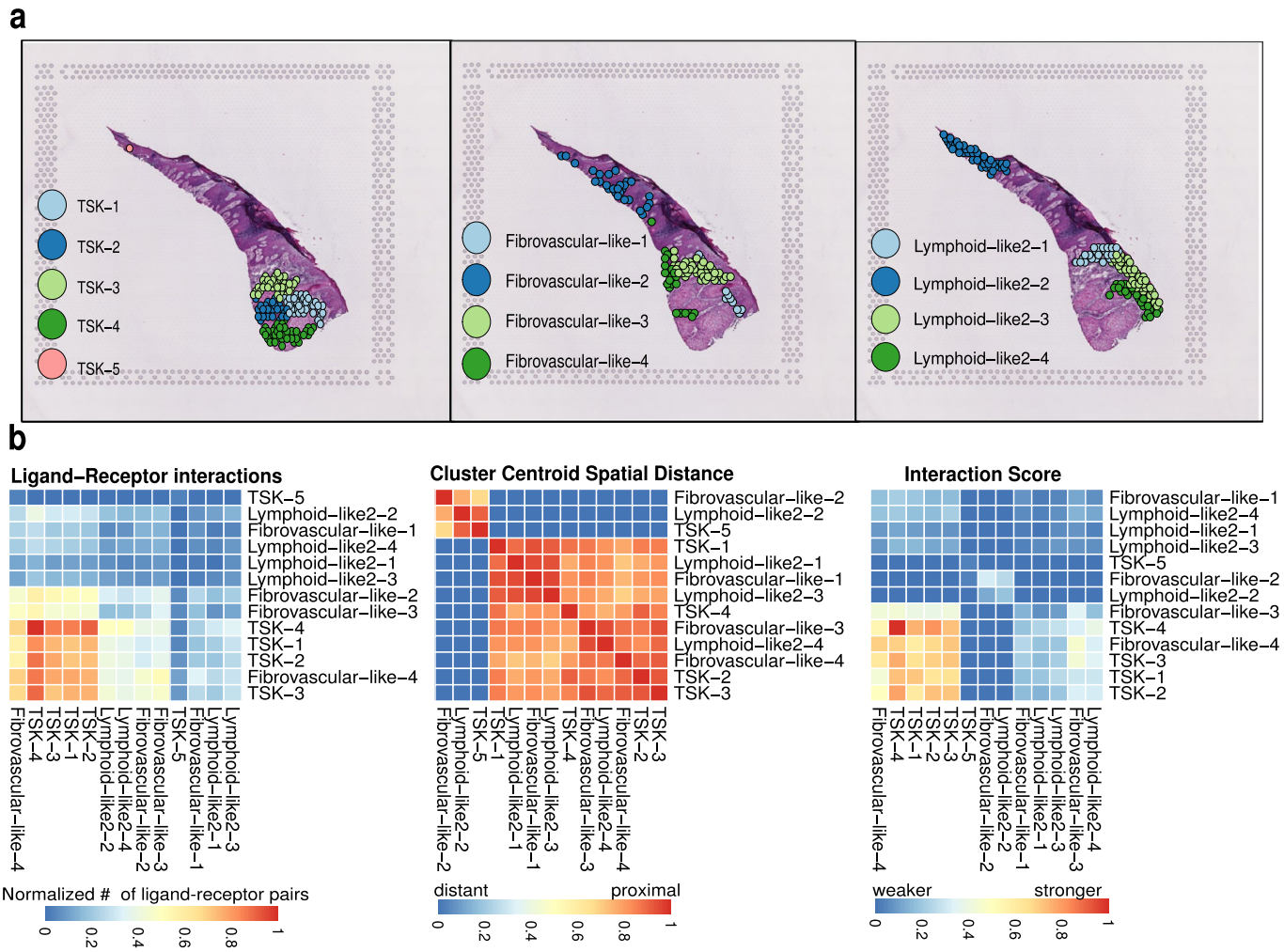


Extended Data Fig. 5 | See next page for caption.

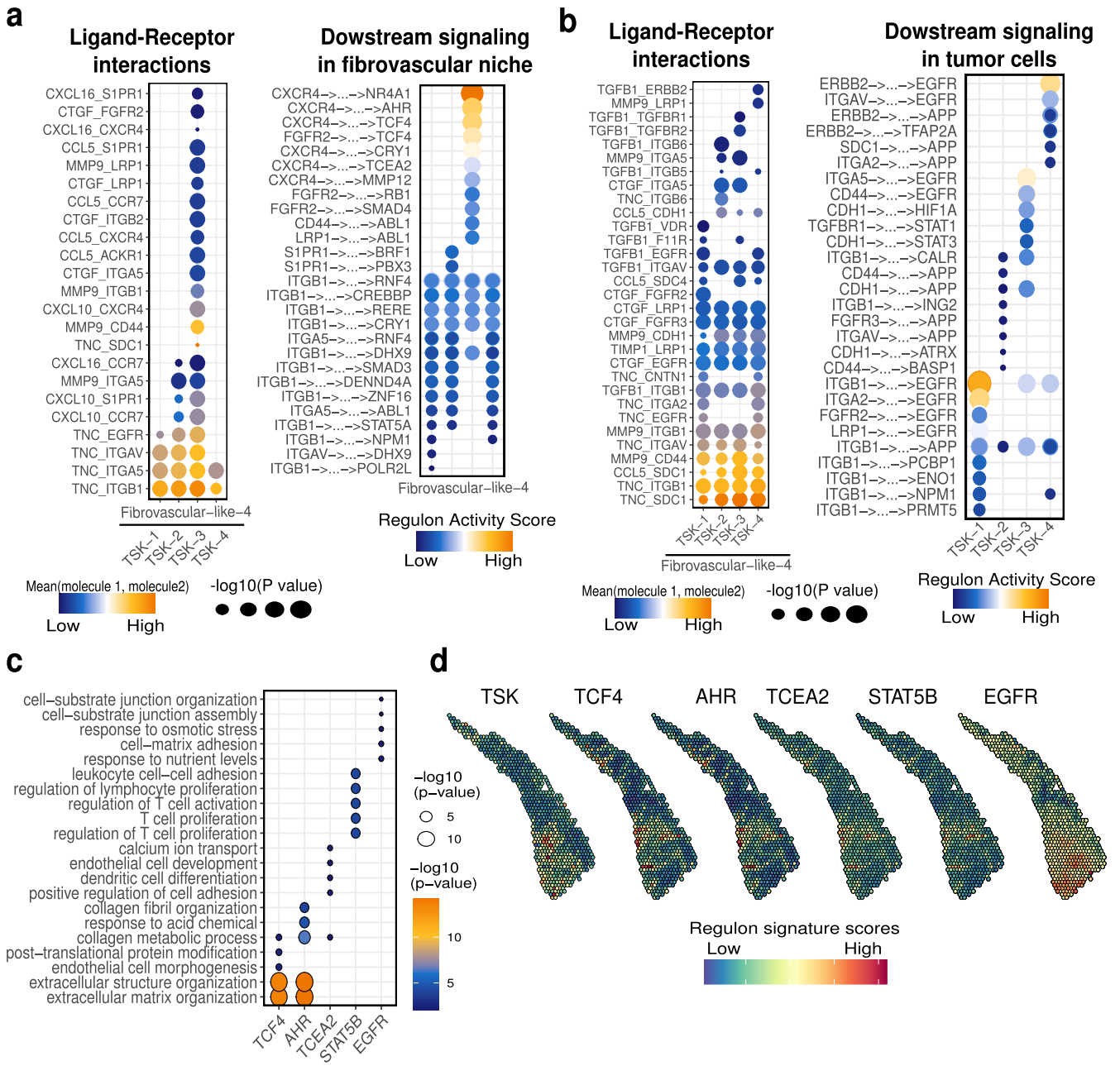
Extended Data Fig. 5 | Experimental validation of transcriptional regulators. a, Stat3 inhibitor dose-response curve on *runx1/cmyb* expression in zebrafish embryos treated from 14-36hpf. **b**, Representative flow plots for CD34 and CD45 staining of floating hematopoietic cells after hiPSC derived EHT culture in the presence of increasing STAT3 inhibitor concentrations. Bar graphs represent mean +/- SD. Ordinary One-way ANOVA, Dunnett's multiple comparisons test, N = 5, from 4 independent cultures. **c**, Gating strategy on positive control samples (CB MNC) and representative flow plots of anti-CD5 and anti-CD7 stained cells, 14 days in T cell differentiation culture in the presence of indicated STAT3 inhibitor concentrations. **d**, No decrease in viability (DAPI negativity) in hiPSC-derived EHT cultures treated with a STAT3 inhibitor. Mean +/- SD is depicted. N = 4, from 3 independent differentiation experiments. **e**, Phenotypic distribution plots of *runx1/cmyb* + staining in 36hpf zebrafish embryos after control and experimental morpholino injection. N = 42 and 46 (ctrl MO and *cebpa* MO) and N = 51 and 53 (ctrl MO and *cebpb* MO) from two independent experiments each. Of note, ybx1 morphants displayed severe embryonic toxicity at doses as low as 1 ng, which prevented further analysis in zebrafish. **f**, qRT-PCR verifying *YBX1* knockdown after 72 h in dox inducible hiPSC derived HE carrying a dCAS9-KRAB fusion protein and sgRNAs against *YBX1*. Bars represent mean +/- SD. N = 3 independent cultures, unpaired t-test, p = 0.0115. **g**, CRISPRi of YBX1 does not affect the percentage of CD34 + CD45 + cells. N = 2 independent sgRNAs. Data and mean is plotted. **h**, Reduction of CFU potential upon CRISPRi-mediated reduction of YBX1 expression in hiPSC derived HSPCs. Bar graphs represent mean. N = 3 sgRNAs for NTC, N = 2 sgRNAs for YBX1. **i**, Reduced lymphoid differentiation potential (N = 4 independent cultures) upon CRISPRi-mediated reduction of YBX1 expression in hiPSC derived HSPCs. ****p < 0.0001, ***p = 0.0001. Bar graphs represent mean +/- SD. **j**, Representative flow cytometric plots of nascent preHSCs (VE-CAD + CD45 +) in DMSO or APPi-treated E9.5 explant cultures. **k**, Quantification of phenotypic T1- and T2-preHSCs in control E10.5 AGM explant cultures (N = 9) or cultures treated with APP inhibitor (N = 8). Bars represent mean +/- SD. Unpaired t-test, two-tailed, T1-preHSCs p = 0.0722, T2-preHSCs p = 0.0258. **l**, Representative flow cytometric plots of E14.5 fetal liver samples from APP wild type (wt), heterozygous (het) and knockout (ko) embryos pregated on single, live, lineage negative cells. Source numerical data are available in source data.



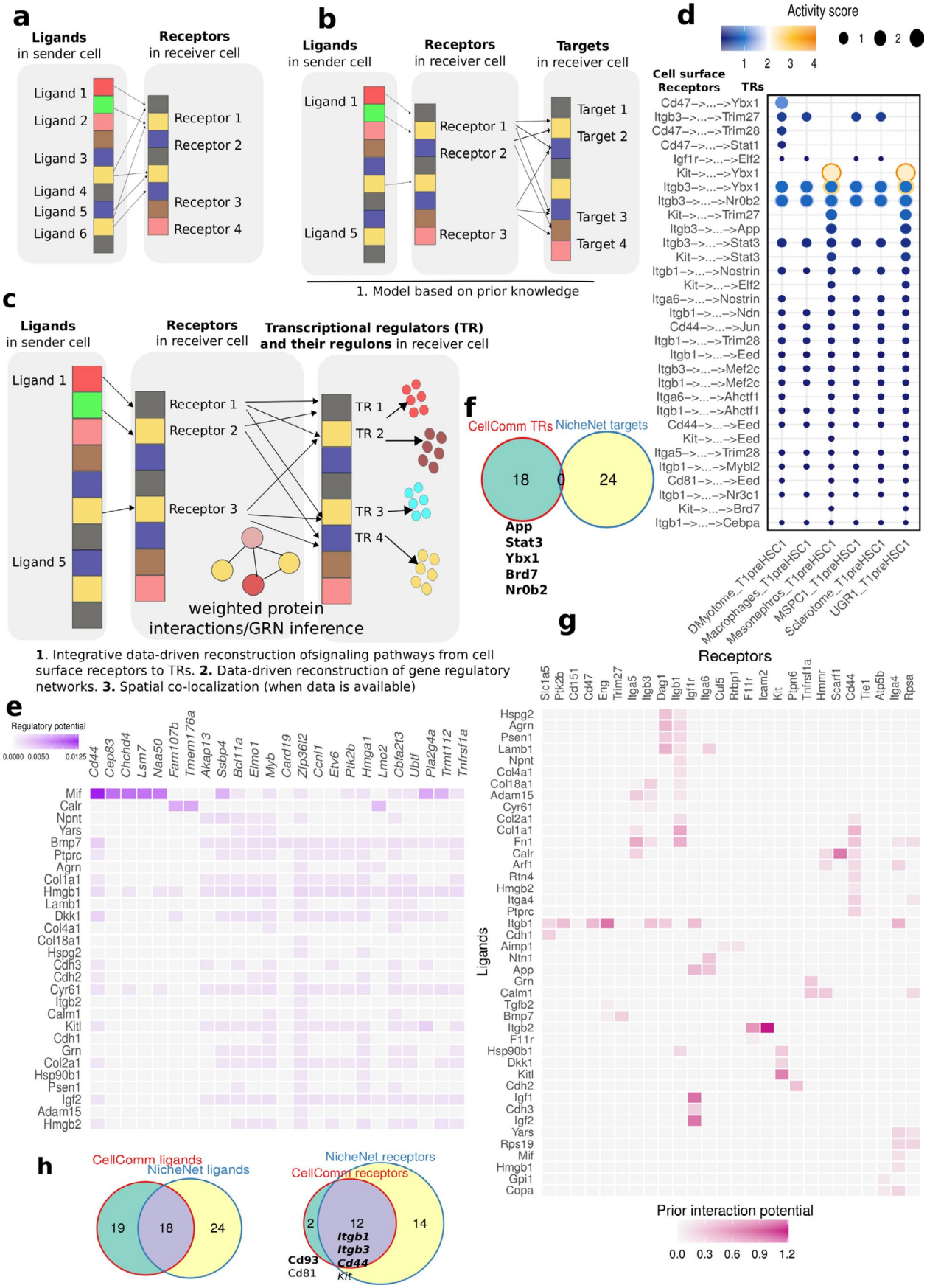
Extended Data Fig. 6 | CellComm infers cell communication from spatial transcriptomics data. a, Transcriptional clusters overlaid on the tissue image. Only spots are shown. **c**, Probabilistic transfer of annotations from a reference cSCC cell atlas to the spatial transcriptomics data, **c**, Probabilistic scores for a Tumor Specific Keratinocyte (TSK) population overlaid on the tissue image. **d**, Spots color-coded by niche identity overlaid on the tissue. Only spots are shown. **e**, left panel: number of ligand-receptor interactions co-expressed between reported cell types; middle panel: spatial proximity between niche centroids in the tissue image; right panel: interaction score taking into account the number of ligand-receptor interactions and the co-localization of niche centroids in the tissue image. **f**, Selected niches with higher interaction scores overlaid on the tissue image. **g**, Ligand-receptor pairs mediating crosstalk between the tumor microenvironment (TME) and the tumor cells. **h**, Spatially resolved expression of selected ligand-receptor pairs identified by CellComm. Source numerical data are available in source data.



Extended Data Fig. 7 | CellComm analysis at spatial subcluster resolution. a, Subclustering of niche identities based on spatial information, showing spatial subclusters of tumor-specific keratinocytes (TSK), the fibrovascular niche identified, as well as spots classified as lymphoid cells. **b**, Number of ligand-receptor pairs co-expressed between spatial subclusters (left panel); spatial proximity between subcluster centroids of reported niches (middle) and interaction score taking into account the number of ligand-receptor pairs and spatial distance of subclusters. Source numerical data are available in source data.

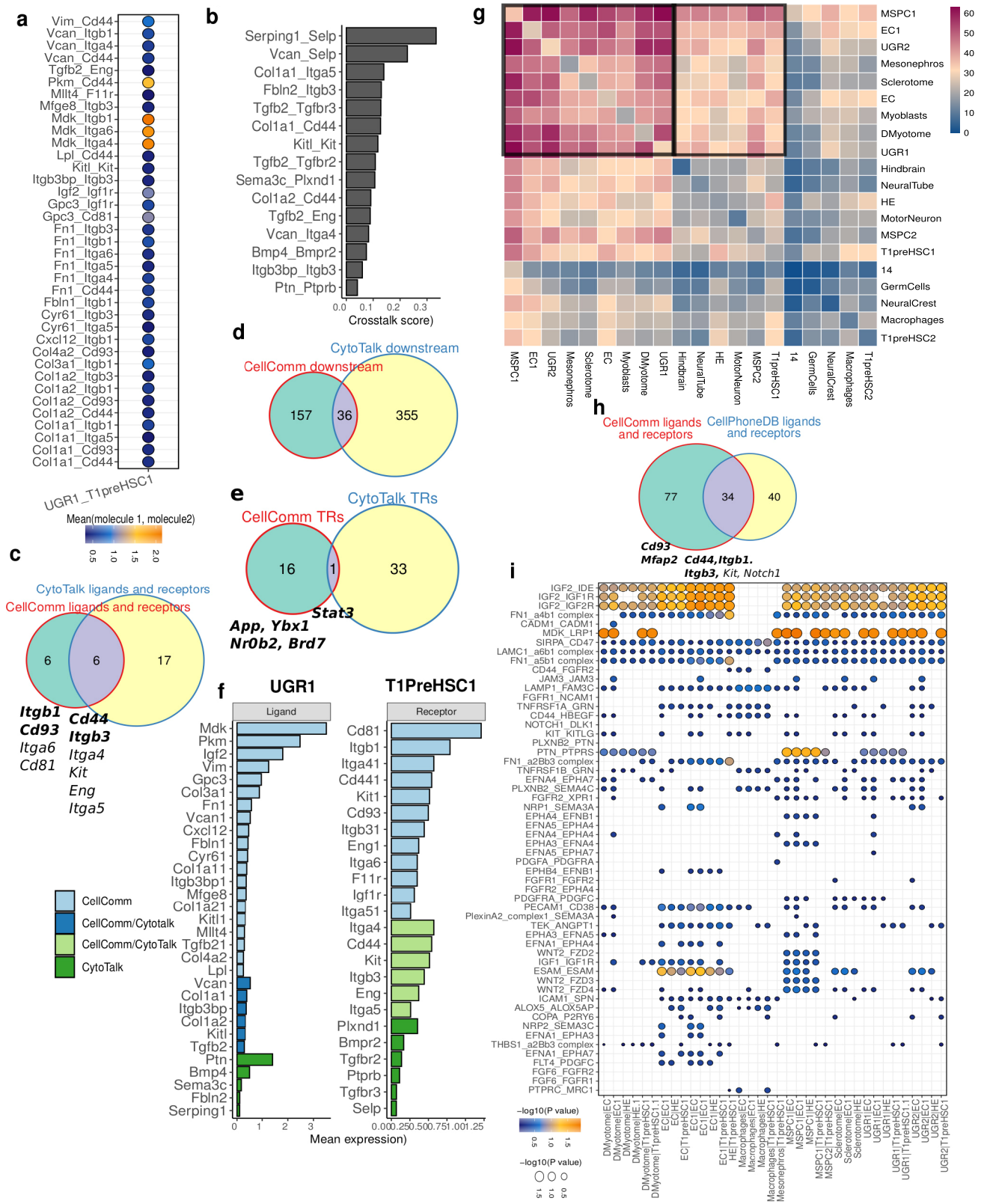


Extended Data Fig. 8 | Signaling processes in the tumor microenvironment. a, Ligand-receptor pairs mediating interactions between TSK spatial subclusters and the fibrovascular niche (left) as well as the downstream transcription factors in the fibrovascular niche identified by CellComm (right). **b**, Ligand-receptor pairs mediating interactions between the fibrovascular niche and the TSK spatial subclusters (left) as well as the downstream transcriptional regulators in tumor cells identified by CellComm. **c**, Gene ontology analysis of the regulons predicted to be controlled by the reported transcriptional regulators. Hypergeometric test. The p-values were adjusted for multiple comparisons using the Benjamini-Hochberg method. **d**, Signature scores for TSKs, demonstrating spatial localization of this tumor-specific population, as well as signature scores calculated from the regulons of the reported transcriptional regulators. Source numerical data are available in source data.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Conceptual and practical comparison of CellComm to related algorithms. Conceptual design of CellPhoneDB, NicheNet and CellComm. **a**, CellPhoneDB infers intercellular communication based on ligand-receptor pairs. **b**, NicheNet uses prior knowledge of signaling pathways and regulatory interactions to build a model to predict which ligands regulate expression of target genes, such as differentially expressed genes or any other gene set of interest. **c**, CellComm infers intercellular communication by exploring co-expression patterns of ligand-receptor pairs across cell types. Then, CellComm weights a large-scale protein interaction network based on cell type-specific co-expression measurements derived from the scRNA-seq data. CellComm implements an optimization algorithm to search for paths connecting cell surface receptors to downstream transcriptional regulators and prioritizes signaling pathways and transcriptional regulators based on the statistical enrichment of the regulons of each transcriptional regulator on cell type-specific signatures. The predicted regulons are identified by gene regulatory network reconstruction from the scRNA-seq data. TR = Transcriptional Regulator. **d**, CellComm analysis showing predicted signaling pathways connecting cell surface receptors in the T1preHSC1 subset to downstream transcriptional regulators. **e**, NicheNet analysis showing which ligands are predicted to regulate the expression of T1preHSC1 signature genes (predicted target genes). **f**, The Venn diagram of the downstream transcriptional regulators identified by CellComm and target genes identified by NicheNet shows no overlap, demonstrating the orthogonal approach to cell-cell communication implemented in CellComm. **g**, Ligands and receptors identified by NicheNet. **h**, Overlap between ligands and receptors identified by CellComm and NicheNet. Genes in bold were experimentally validated in our study (hypergeometric test). Genes not in bold are known genes involved in hematopoiesis. Source numerical data are available in source data.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Comparison of CellComm to CytoTalk and CellPhoneDB. a, Ligand-receptor interactions predicted by CellComm for the UGR1-T1preHSC1 interaction. **b**, Ligand-receptor interactions predicted by CytoTalk for the UGR1-T1preHSC1 interaction. **c**, Overlap of ligands/receptors identified by CellComm and CytoTalk. Genes in bold were experimentally validated in our study. **d**, Overlap of downstream genes predicted by CellComm and CytoTalk. **e**, Overlap of transcriptional regulators predicted by CellComm and transcriptional regulators in the downstream CytoTalk subnetwork (hypergeometric test). **f**, Mean expression of ligands and receptors identified by CellComm and CytoTalk. **g**, Cell-cell interaction network inferred by CellPhoneDB using the AGM scRNA-seq data generated in this study. **h**, Overlap of ligands/receptors identified by CellComm and CellPhoneDB (hypergeometric test). Genes in bold were experimentally validated in our study. Genes not in bold are known genes involved in hematopoiesis. **i**, Ligand-receptor pairs identified by CellPhoneDB in the AGM scRNA-seq data. Source numerical data are available in source data.

nature research

Corresponding author(s): George Q. Daley/Trista E. North

Last updated by author(s): Feb 17, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Embryos were harvested at E10.5 and somites counted; only embryos between 28 and 32 somite pairs were processed further. AGM regions were dissected, pooled and dissociated following Chen et al., 2020. Cells were filtered (40 µm filter) and stained in PBS/2%FBS on ice using antibodies provided in Supplementary table 1. Cells were sorted at 15 psi pressure using a 100 µm nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter). Sorted cells were collected into 1.5 ml Eppendorf tubes containing 300 µl PBS/50%FBS. Immediately following sorting, each cell population was divided across two lanes on a 10X Genomics scRNA-Seq chip as independent technical replicates to encapsulate cells into droplets. scRNA-Seq libraries were then prepared per the 10X scRNA-Seq v2 protocol. Cells were loaded into 10X lanes at cell concentrations to maintain an estimated doublet rate below 5%. Final 10X libraries were assayed via an Agilent High Sensitivity dsDNA Bioanalyzer, normalized, pooled and shallow sequenced on a MiniSeq, identifying 10,000 high confidence cell barcodes in total across all conditions. 10X libraries were then renormalized per the distribution of reads/library from the MiniSeq run and deep sequenced on a NovaSeq S4 to a depth of 50,000 reads per cell barcode.
Data analysis	Single-cell sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite. Cells with fewer than five detected genes were removed. Genes expressed in fewer than 5 cells were also removed. After quality control, our dataset was composed of 9,492 single cells with a median of 2,488 genes detected per cell. We used CellRouter to perform quality control, dimension reduction, identify gene signatures and reconstruct the EHT differentiation trajectory. CellComm was used to infer cell communication networks and downstream signaling pathways and regulatory networks. CellRouter and CellComm are available through the Framework for Unified Single-Cell Analysis R package (https://github.com/edroaldo/fusca). Versions of all software packages used in this study are provided below: R version 4.0.1 (2020-06-06), Seurat_3.2.3, clusterProfiler_3.18.0, fusca_1.3.1, Cell Ranger 2.2.0, FlowJo v_10.4.2, Prism9 (9.3.1)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Our dataset is freely available under the Gene Expression Omnibus (GEO) accession number: GSE160526.

We re-analyzed publicly available datasets from the following accession numbers:
GSE137117: Zhu, Q. et al. Developmental trajectory of pre-hematopoietic stem cell formation from endothelium. Blood (2020) doi:10.1182/blood.2020004801.
GSE144240: Ji, A. L. et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. Cell 182, 1661–1662 (2020).

We used the mouse genome version mm10-1.2.0. Reference genome data for scRNA-seq analyses were downloaded from https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/2.2.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods to predetermine sample size were used. Animal experiments were performed in biological replicates according to the standards of the field. For zebrafish morpholino experiments, more than 20 embryos per experimental condition were scored, for mouse embryos at least 4 embryos per genotype were analyzed.
Data exclusions	No data were excluded
Replication	Number of replicates per experiment and statistical analyses are described in each figure legend.
Randomization	Mice and zebrafish embryos were assigned randomly to groups and not blinded
Blinding	Blinding was performed where subjective bias would influence data collection, for example in CFU assays. For APP mouse mutant experiments, genotypes were not know at the time of analysis. Morpholino experiments in fish were not performed blindly, however, independent scoring validation was performed to verify robustness of data. For other in vitro experiments, blinding was not performed due to the complexity of the experimental setup.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Rat anti mouse Ter119 PE-CY5 TER-119 Thermo Fisher Scientific Cat# 15-5921-82 1/100 Rat anti mouse CD31 BV421 MEC 13.3 BD Bioscience Cat# 562939 1/100

Rat anti mouse CD41 PE-CY7 eBioMWReg30 Thermo Fisher Scientific Cat# 25-0411-82 1/100
Rat anti mouse CD144/VE-Cadherin PE eBioBV13 Thermo Fisher Scientific Cat# 12-1441-82 1/100
Rat anti mouse CD45 FITC 30-F11 BioLegend Cat# 103108 1/100
Rat anti mouse CD117 (cKit) APC-eFluor 780 2B8 Thermo Fisher Scientific Cat# 47-1171-82 1/100
Rat anti mouse Sca-I (PE-CY7) D7 Thermo Fisher Scientific Cat# 25-5981-82 1/100
Rat anti mouse CD3 (eFluor 450) 17A2 Thermo Fisher Scientific Cat# 48-0032-82 1/100
Rat anti mouse CD45R (eFluor 450) RA3-6B2 Thermo Fisher Scientific Cat# 48-0452-82 1/100
Rat anti mouse CD11b (eFluor 450) M1/70 Thermo Fisher Scientific Cat# 48-0112-82 1/100
Rat anti mouse Ter-119 (Pacific blue) TER-119 BioLegend Cat# 116231 1/100
Rat anti mouse Gr-1/Ly-6G/Ly-6C (eFluor 450) RB6-8C5 Thermo Fisher Scientific Cat# 48-5931-82 1/100
Mouse anti human CD5 BV510 UCHT2 Thermo Fisher Scientific Cat# BDB563381 1/100
Mouse anti human CD7 PE M-T701 Thermo Fisher Scientific Cat# BDB555361 1/100
Mouse anti human CD45 APC-CY7 2D1 Thermo Fisher Scientific Cat# BDB557833 1/100
Mouse anti human CD34 PE-CY7 581 Thermo Fisher Scientific Cat# BDB560710 1/100
Mouse anti human CD45 APC J33 Beckman Coulter Cat# IM2473U 1/100
Mouse anti human CD61 PE VI-PL2 Thermo Fisher Scientific Cat# BDB561912 1/100
Mouse anti human CD93 PE VIMD2 BioLegend Cat# 336108 1/100
Mouse anti human CD44 FITC BJ18 BioLegend Cat# 338804 1/100
Mouse anti human CD117 (KIT) PE-CY5 104D2 BioLegend Cat# 313210 1/100
Mouse anti human CD61 neutralizing ab PM6/13 Novus Biologicals Cat# NBP1-28398 1 µg/ml
rabbit anti-GFP antibody Origene Cat#TP401 1/100
goat anti-rabbit AF647 Thermo Fisher Scientific Cat# A-21245 1/1000

Validation

Antibodies against human targets were previously validated using CB mononuclear cells or peripheral blood T-cells , while antibodies against mouse targets were tested on mouse embryos and mouse bone marrow by the researcher during pilot studies.
All antibodies used were also validated by the manufacturer (BD bioscience, Thermo Fisher Scientific and Biolegend):
The TER-119 antibodies were tested by flow cytometric analysis of mouse splenocytes and bone marrow cells by the manufacturer.
The CD31 antibody was validated on mouse bone marrow cells by the manufacturer.
The CD41 antibody was tested by flow cytometric analysis of mouse platelets by the manufacturer.
The CD144 antibody was tested by flow cytometric analysis of bEnd.3 cells by the manufacturer.
The CD45 antibody was quality control tested by immunofluorescent staining of mouse splenocytes with flow cytometric analysis by the manufacturer.
The CD117 antibody was tested by flow cytometric analysis of mouse bone marrow cells by the manufacturer.
The CD45R antibody was tested by flow cytometric analysis of mouse splenocytes by the manufacturer.
Both CD11b antibodies were tested by flow cytometric analysis of mouse bone marrow cells by the manufacturer.
For goat anti-Rabbit IgG: To ensure specificity, the secondary antibody has been cross adsorbed against bovine IgG, goat IgG, mouse IgG, rat IgG and human IgG by the manufacturer

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

iPSCs, generated as previously described (Sugimura et al., Nature, 2017), were obtained from the Boston Children's Hospital Human Embryonic Stem Cell Core (hESC) and verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping. Primary cells (CD34+ cord blood) were obtained from AllCells.

Authentication

iPSCs were verified by immunohistochemistry for pluripotency markers, teratoma formation and karyotyping

Mycoplasma contamination

All lines routinely tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used in this study

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

All mouse embryos used in this study were derived from timed matings at E9.5-E14.5 as indicated in the figure legends. The genotypes were either C57BL/6J (The Jackson Laboratory, Stock number 000664) or APP KO (B6.129S7-Apptm1Dbo/J, Stock No: 004133). The sex of the embryos was not determined at the time of analysis.

Morpholino injections were performed with: AB (WT)
Transgenic lines used were: Tg(kdrl:mCherry) s916, Tg(cmyb:EGFP)zf169

Wild animals

no wild animals were used in this study

Field-collected samples

no field-collected samples were used in this study

Ethics oversight

All mice were housed in pathogen-free animal facilities, and all experiments were performed with the approval of the Boston Children's Hospital Institutional Animal Care and Use Committee (IACUC Study#000014940).

Zebrafish were maintained in accordance with approved Beth Israel Deaconess Medical Center (Protocol #: 045-2021) and Boston Children's Hospital (Protocol #: 00001259) Institutional Animal Care and Use Committee protocols.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Embryos were harvested at E10.5 and somites counted; only embryos between 28 and 32 somite pairs were processed further. AGM regions were dissected, pooled and dissociated following Chen et al., 2020. Cells were filtered (40 µm filter) and stained in PBS/2%FBS on ice using antibodies provided in Supplementary table 1. Cells were sorted at 15 psi pressure using a 100 µm nozzle on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter). Sorted cells were collected into 1.5 ml Eppendorf tubes containing 300 µl PBS/50%FBS.
Instrument	Acquisition was performed on a BD Fortessa cytometer and all sorting was perforemd on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter).
Software	All flow cytometry data was analyzed using FlowJo v_10.4.2
Cell population abundance	No post sort acquisition was performed due to limited number of obtained cells.
Gating strategy	All cells were first gated on FSC/SSC according to cell size and granularity. Doublet exclusion was performed using FSC-A vs FSC-H and SSC-W vs SSC-H. For sorting of AGM cells, dead cells and red blood cells were excluded by gating on 7-AAD negative/Ter119-PE-CY5 negative cells. For hiPSC related experiments, dead cells were excluded by gating on DAPI negative cells. Positive gates were set according to FMO controls.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.