



**HAL**  
open science

# Unraveling the Role of MIXL1 Activation in Endoderm Differentiation of Isogenic Human Induced Pluripotent Stem Cells

Pierre Osteil, Sarah Withey, Nicole Santucci, Nader Aryamanesh, Chi Nam Ignatius Pang, Nazmus Salehin, Jane Sun, Annie Qin, Jiayi Su, Hilary Knowles, et al.

► **To cite this version:**

Pierre Osteil, Sarah Withey, Nicole Santucci, Nader Aryamanesh, Chi Nam Ignatius Pang, et al.. Unraveling the Role of MIXL1 Activation in Endoderm Differentiation of Isogenic Human Induced Pluripotent Stem Cells. 2024. hal-04496012

**HAL Id: hal-04496012**

**<https://hal.science/hal-04496012>**

Preprint submitted on 8 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Unraveling the Role of MIXL1 Activation in Endoderm Differentiation of Isogenic Human Induced Pluripotent Stem Cells

Pierre Osteil,<sup>1\*</sup> Sarah Withey,<sup>3</sup> Nicole Santucci,<sup>1</sup>  
Nader Aryamanesh,<sup>4</sup> Ignatius Pang,<sup>4</sup> Nazmus Salehin,<sup>2</sup> Jane Sun,<sup>1</sup>  
Annie Qin,<sup>1</sup> Jiayi Su,<sup>1</sup> Hilary Knowles,<sup>1</sup> Simon Cai,<sup>4</sup> George Craft,<sup>5</sup> Mark Graham,<sup>5</sup>  
Xiucheng Bella Li,<sup>1</sup> Ernst Wolvetang,<sup>3</sup> and Patrick P. L. Tam,<sup>1,2\*</sup>

<sup>1</sup>Embryology Research Unit, Children's Medical Research Institute, University of Sydney, Australia,

<sup>2</sup>School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Australia,

<sup>3</sup>Australian Institute for Bioengineering and Nanotechnology, University of Queensland, Australia

<sup>4</sup>Bioinformatics Group, Children's Medical Research Institute, University of Sydney, Australia,

<sup>5</sup>, Biomedical Proteomics, Children's Medical Research Institute, University of Sydney, Australia

\* To whom correspondence should be addressed; E-mail: pierre.osteil@uca.fr, ptam@cmri.org.au

<sup>φ</sup> Present address: Université Clermont Auvergne, CNRS, INSERM, GReD Institute, Faculté de Médecine, F-63000 Clermont-Ferrand, France.

**Human induced pluripotent stem cells (hiPSC) possess the ability to differentiate into a multitude of tissue types but display heterogeneity in the propensity of differentiation into specific tissue lineages. An examination of isogenic hiPSC lines revealed variations in the endoderm propensity under directed differentiation conditions. Characterization of the transcriptome and proteome of the hiPSC lines showed that *MIXL1* activity at the early differentiation stage correlated with the efficacy of generating definitive endoderm and further differentiation into endoderm derivatives. Enforced expression of *MIXL1* in the endoderm-incompetent hiPSCs enhanced the propensity of endoderm differentiation, suggesting that modulation of key drivers of lineage differentiation can re-wire hiPSC to the desired lineage propensity for stem cell products.**

## Introduction

Human induced pluripotent stem cells (hiPSCs) are noted for their ability to differentiate into a multitude of cell and tissue types<sup>1-6</sup>. Many protocols have been developed to direct the differentiation of hiPSCs to desirable types of cells or tissues, including the endoderm precursor, the definitive endoderm<sup>7-10</sup> and endoderm derivatives such as intestinal cells<sup>11-13</sup> pancreatic cells<sup>14</sup> and hepatocytes<sup>15-18</sup>. A recent study of a bank of hiPSC lines derived from 125 individuals revealed that hiPSC lines respond differently when directed to differentiate to definitive endoderm<sup>19</sup>, suggesting that there is innate heterogeneity in the propensity for endoderm differentiation among hiPSC lines. In this study the lineage propensity has been linked to specific quantitative trait loci (QTL). Both genetic determinants<sup>20-22</sup>, and somatic or epigenetic memory related to the cell/tissue of origin<sup>23-26</sup>, have been shown to underpin the variable lineage specification and differentiation propensity of hiPSCs. The impact of epigenetic memory on establishment of functional tissue nevertheless remains unresolved, even when examining isogenic cell lines from the same cellular resource and reprogrammed under the same

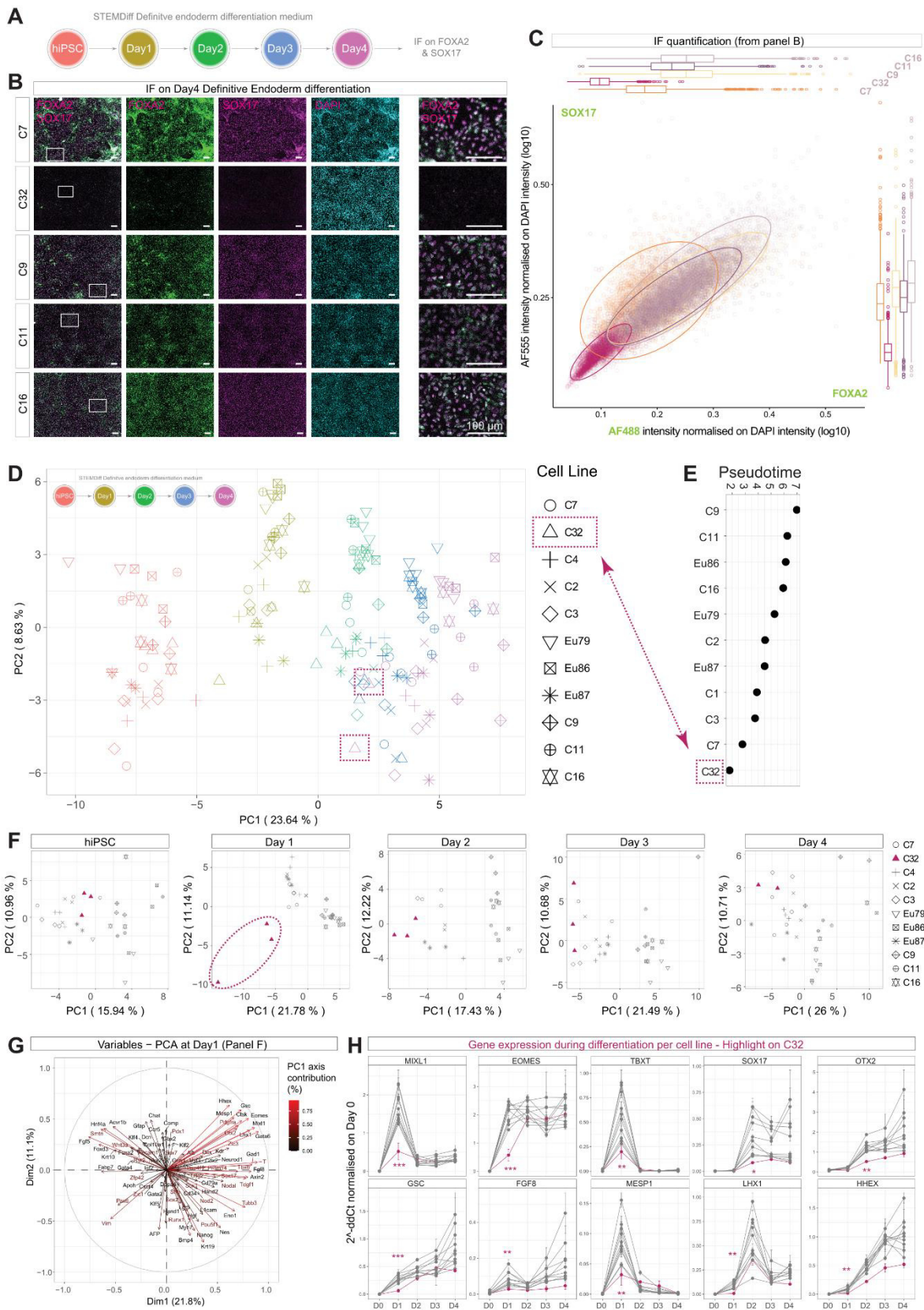
41 conditions<sup>27</sup>. In the present study, we investigated the propensity of endoderm differentiation of  
42 eleven pluripotent lines of four sets of isogenic hiPSCs by tracking the differentiation from  
43 pluripotent cells to definitive endoderm (DE), hepatocytes and intestinal organoids (hIO). We  
44 showed that in isogenic hiPSCs, early activation and a high level of *MIXL1* expression were  
45 associated with enhanced propensity of endoderm differentiation. In the mouse embryo, *Mixl1*  
46 is expressed in the primitive streak and the nascent mesoderm during gastrulation and  
47 expression persists in the primitive streak of the early-somite-stage embryo<sup>28,29</sup>. Loss of *Mixl1*  
48 function is associated with deficiency of DE and under-expansion of the nascent mesoderm<sup>30</sup>.  
49 In the mouse embryonic stem cells, loss of *Mixl1* function leads to inefficient differentiation of  
50 lateral mesoderm tissue and hematopoietic lineages<sup>31</sup>, whereas constitutive *Mixl1* activity  
51 promotes the differentiation of *Foxa2*+/*ECad*+ DE cells<sup>32</sup>. In mouse epiblast stem cells, activation  
52 of *Mixl1* at the early phase of differentiation correlates with improved endoderm differentiation<sup>28</sup>.  
53 Analysis of the molecular attributes of DE differentiation revealed that the activity of *MIXL1*  
54 at the early phase of hiPSC differentiation promotes the differentiation of *SOX17*+ DE cells when  
55 confined to defined size micropattern<sup>33</sup>. We further showed that enhanced expression of *MIXL1*  
56 in hiPSCs augmented endoderm propensity, advancing the understanding of how lineage  
57 propensity can be re-wired to generate fit-for-purpose pluripotent stem cells for translational  
58 application.

## 59 Results

### 60 Early onset of gastrulation is necessary for definitive endoderm formation

61 Eleven hiPSC lines from four isogenic groups (two males and two females)<sup>34,35</sup> were subjected  
62 to DE differentiation by following the manufacturer instruction (STEMDiff definitive endoderm  
63 protocol) (Figure 1A) and assessed for expression of *FOXA2* and *SOX17* on Day 4 of  
64 differentiation (Figure 1B and C, Supplementary Figure S1A and B). Among the hiPSCs, C32  
65 had the lowest expression of both endodermal transcription factors, despite comparable cell  
66 morphology to other cell lines (Figure 1C), suggesting that this cell line is not amenable to  
67 definitive endoderm differentiation. To track the developmental trajectory, cells were collected in  
68 triplicate every day from Day 0 (pluripotency) to Day 4 (DE), and the expression of 96 genes  
69 involved in regulating pluripotency to gastrulation was profiled. Line C32 showed the least  
70 progression across the 4 days of differentiation in the PCA plot (Figure 1D). By taking PC1 as  
71 an endoderm differentiation efficiency proxy<sup>19</sup>, to infer an endoderm specification pseudotime,  
72 the average of triplicate PC eigenvalue along the PC1 axis was calculated to rank the hiPSC  
73 lines (Figure 1E). The results show that C32 ranked last, indicating that low DE differentiation  
74 efficiency could account for the differentiation failure previously suggested by the absence of  
75 *FOXA2* and *SOX17* double positive cells.

76 To unveil the earliest manifestation of such discrepancies, the same PCA as Figure 1D was  
77 plotted for each day, to discern differences among cell lines that may not be visible on the full  
78 timeline of differentiation (Figure 1F). The C32 line displayed disparity on the main axis from the  
79 cohort at Day 1 (Figure 1G). The low differentiation efficiency does not appear to be linked to a  
80 slower down-regulation of pluripotency factors as there are no significant differences among the  
81 11 cell lines for *POU5F1*, *NANOG*, *SOX2*, *PRDM14*, *ZFP42* and *FGF5* mRNA expression



**Figure 1 – Definitive endoderm differentiation heterogeneity among isogenic lines: A)** Differentiation protocol used to generate definitive endoderm (DE) cells; **B)** Immunostaining images of DE cells on FOXA2 (green) and SOX17 (magenta), scale bar = 100μm; **C)** Signal intensity measurement of immunostaining of panel B, n = 3. **D)** PCA obtained from microfluidic RT-qPCR data on DE differentiation time courses. Each day is represented by the same color as Figure 1A (inserted on top left). Purple squares highlight C32 at Day 4 of DE differentiation; **E)** PC1 axis projection of hiPSC at Day 4 of DE differentiation representing efficiency of differentiation as pseudotime. C32 is highlighted in purple. **F)** PCA obtained from microfluidic RT-qPCR data on DE differentiation, identical dataset to figure 1D, but plotted for each day. C32 is highlighted in purple; **G)** Genes' contribution to PC1 and PC2 axis of the Day 1 PCA of Figure 2F. The position of the arrows correlates with the position of the samples on the PCA at Day 1. Colour scale is between 0 and 1. A score of 1 indicates maximum contribution of a particular gene to PC1. **H)** Genes' expression time course during DE differentiation. C32 is highlighted in purple. p.value: \* < 0.05, \*\* < 0.01, \*\*\* < 0.001.

83 down-regulated at Day1 in C32 compared to the cohort (Figure 1H). C32 also has the lowest  
84 levels of expression of *SOX17*, *HHEX*, *OTX2* and *LHX1* at Day 4. *KRT19*, *AXIN2*, *GATA6* and  
85 *GATA4* are expressed at a lower level in the C32 line, although this did not reach significance  
86 (Supplementary Figure S2Aii). Surprisingly, *FOXA2* and *NODAL* are not dysregulated in the C32  
87 line. In addition, mesodermal genes (*BMP4*, *MYH7*, *KLF5*, *KDR*, *PDGFR $\alpha$*  and *CD34*)  
88 (Supplementary Figure S2Aiii) as well as ectodermal genes (*KRT10*, *SOX1*, *NES*, *FOXD3*,  
89 *PAX6* and *DCX*) (Supplementary Figure S2Aiv) showed no significant difference between C32  
90 and other lines.

91 In conclusion, the inability of C32 to activate the molecular program of germ layer differentiation  
92 to a threshold level may be predictive of the low endoderm propensity of this hiPSC line.

93

## 94 **Low endodermal propensity line fails to progress toward functional tissue**

95 Human iPSC lineage propensity was further assessed by the outcome of differentiation into two  
96 endoderm derivatives, hepatocytes (HCm) and intestinal organoids (hIOs) (Figure 2A). C32 was  
97 compared to the higher DE propensity cell line C11<sup>34,35</sup>. Both cell lines were able to differentiate  
98 into hepatocytes (Figure 2C). Microfluidic RT-qPCR analysis of genes specific to hepatocyte  
99 development did not reveal any major differences in the transcriptome between these two cell  
100 lines during hepatocyte differentiation (Figure 2B). The phenotype of AAT- and ALB-expressing  
101 hepatocytes was also similar (Figure 2C, D and Supplementary Figure S1C). However, C32-  
102 derived hepatocytes showed lower Cytochrome P450 3A4 activity across replicates (Figure 2E)  
103 indicating that C32 hepatocytes might have a less efficient metabolism compared to C11.

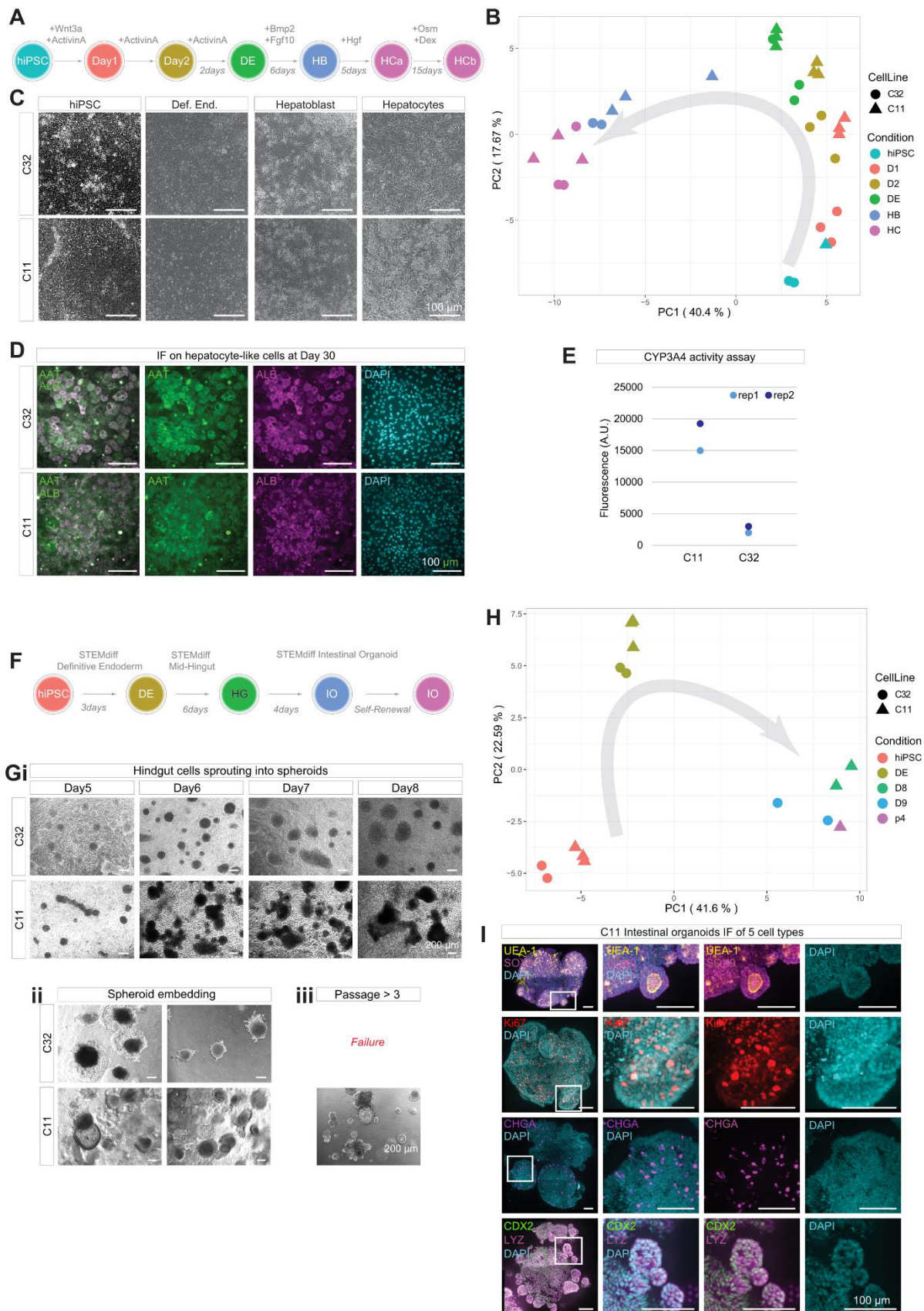
104 In parallel, hIOs were generated from these two cell lines (Figure 2F). At the mid/hindgut budding  
105 spheroid stage, the C32 line generated fewer spheroids than the C11 line (Figure 2Gi). The few  
106 spheroids that were successfully generated were embedded into Matrigel but did not grow as  
107 well as C11 spheroids (Figure 2Gii). Consequently, hIO development was arrested early after  
108 passage 3 (Figure 2Giii). The hIO differentiation of C32 line was repeated with changes in  
109 seeding densities and batches of cells, but the outcome remained unsuccessful (data not  
110 shown). Microfluidic RT-qPCR analysis of genes specific to intestinal organoid differentiation did  
111 not show any discernible signature that could be indicative of inefficient differentiation prior to  
112 the hIO development failure of the C32 cell line (Figure 2H). In contrast, C11 derived hIOs harbor  
113 the representative cell types of the intestinal epithelium: enterocytes (CDX2+), intestinal stem  
114 cells (SOX9), enteroendocrine cells (CHGA), goblet cells (UEA-1) and Paneth cells (LYZ)  
115 (Figure 2I).

116 Together, these results indicate that C32 behaved differently from the C11 line, and displayed a  
117 lower propensity for endoderm differentiation. The C32 cell line therefore was labelled as a  
118 hiPSC line that is refractory to endoderm differentiation in our subsequent analyses.

119

## 120 **Hippo signaling is up regulated in the refractory cell line**

121 Bulk RNA-seq was performed at Day 1 of DE differentiation to discover genes involved in  
122 differential regulation of endodermal propensity at the early phase of germ layer differentiation



**Figure 2 – Low endodermal propensity fails to produce functional tissue:** **A)** Differentiation protocol used to generate hepatocytes from C32 and C11; **B)** PCA obtained from microfluidic RT-qPCR data on hepatocyte differentiation; **C)** Brightfield pictures of hepatocytes differentiation; **D)** Immunostaining images on AAT (green) and ALB (magenta) markers of hepatocytes differentiation; **E)** Results of fluorescent analysis of CYP3A4 activity; **F)** Differentiation protocol used to generate intestinal organoids; **G)** Brightfield pictures of intestinal organoid differentiation at the Hindgut stage (i), spheroid generation stage (ii) and maintenance stage (iii); **H)** PCA obtained from microfluidic RT-qPCR data on intestinal organoid differentiation; **I)** Immunostaining of intestinal organoids of C11 cell lines on different cell types of the gut epithelium: Goblet cells (UEA-1), Intestinal Stem cell (SOX9), enteroendocrine cells (CHGA), epithelium (CDX2) and Paneth cells (LYZ) as well as proliferating cells (Ki67). Nuclei are revealed by DAPI.

124 (gastrulation). Five cell lines were analyzed including two females (C32 and C7) and 3 males  
125 (C9, C11, C16). Transcriptomic differences showed that C32 differs significantly from the 4 other  
126 cell lines (Supplementary Figure S2B and D), displaying enrichment of transcriptomic signature  
127 of mesoderm (circulatory system and heart: *RUNX1*, *VEGFA*) and ectoderm (neural: *SOX2*,  
128 *POU3F2*) (Supplementary Figure S2C) derivatives. Most of the genes associated with  
129 gastrulation (e.g. *MIXL1*, *EOMES*, *MESP1*, *APLN*, *DKK1*, *GATA6*, etc...) (Supplementary  
130 Figure S2C) were down-regulated in C32. To eliminate the gender bias in the above analysis,  
131 C7, an isogenic clone of C32, was analyzed in parallel. The results of this comparison were  
132 similar to the global analysis. The differentially expressed genes (DEGs) analysis revealed that  
133 C7 expressed genes associated with gastrulation (*MIXL1*, *EOMES*, *TBXT*, *SNAI1*, *CER1*) at a  
134 higher level, while C32 differentially expressed genes associated with cell adhesion (*VIM*, *EZR*,  
135 *FLNA*, *FLNC*, *COL1A1*, *FN1*, *ITGA2/3/6*) and Hippo signaling (*CCN1*, *CCN2*, *AMOT*, *AJUBA*,  
136 *CDH11*) (Supplementary Figure S2E). The transcriptomic analysis thus highlights a possible  
137 ectoderm and mesoderm bias of the C32 cell line during germ layer differentiation that is in  
138 keeping with its lower endodermal propensity. It also indicated a possible role for altered Hippo  
139 signaling and cell adhesion in negatively modulating endoderm differentiation.

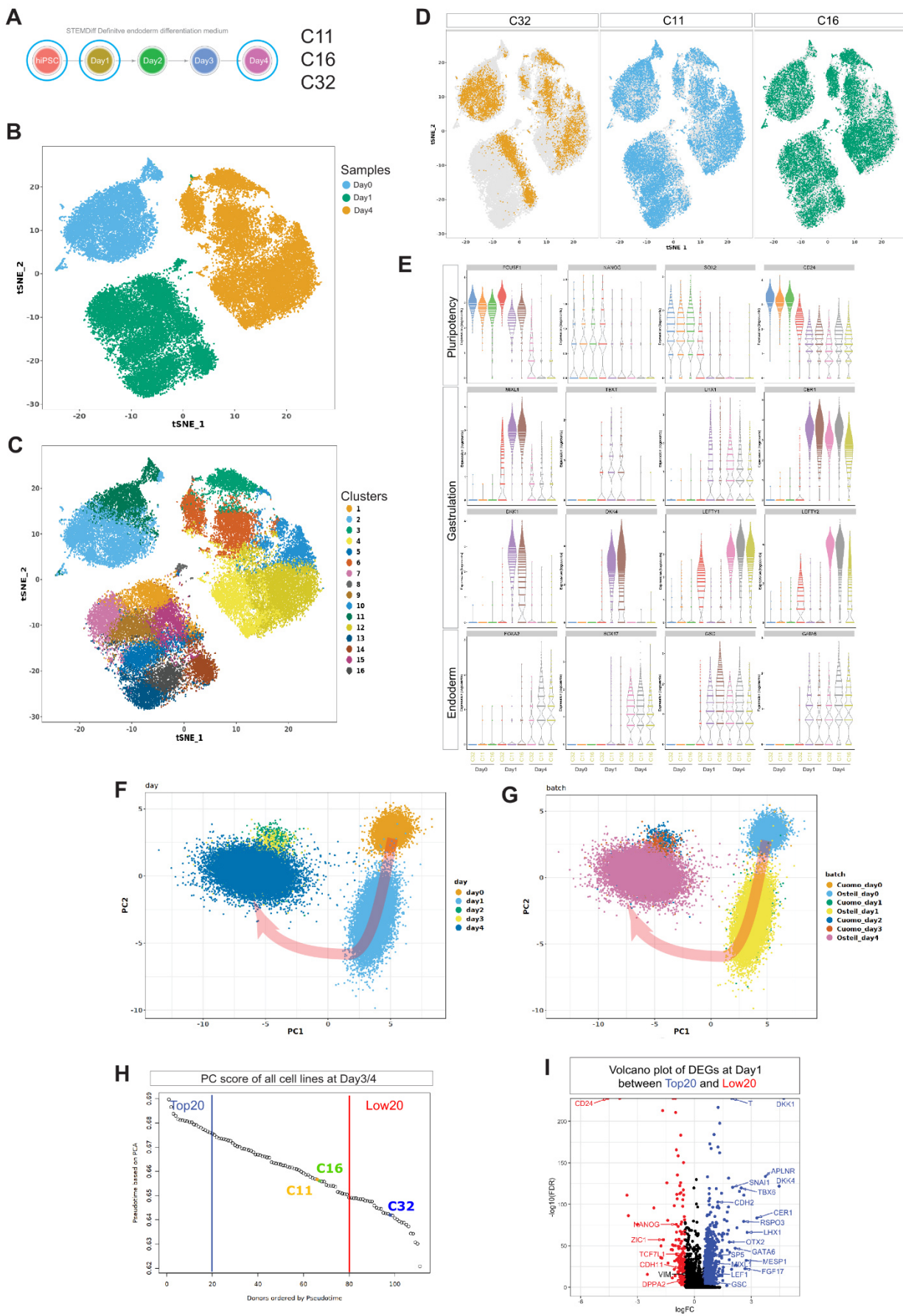
140

## 141 **Refractory cell line harbors a unique molecular signature during** 142 **differentiation of primitive-streak like cells**

143 We further investigated three cell lines (i.e., C32, C11 and C16) by single-cell RNAseq (scRNA-  
144 seq) at 3 time points during differentiation: Day 0 (pluripotency), 1 (peri-gastrulation) and 4  
145 (definitive endoderm) (Figure 3A). After filtering, 41336 cells were retained for analysis. The  
146 tSNE plot (Figure 3B) showed that cells were segregated by time into three major clusters  
147 corresponding to each day. Each cluster were further divided into smaller clusters based on their  
148 transcriptomic differences (Figure 3C). While C32 cells displayed a unique transcriptomic profile  
149 at Day 1, the individual sub-clusters of C32 cells were found within the cell clusters of each of  
150 the other cell lines at Day 4 (Figure 3D), albeit present at different abundance, suggesting that  
151 the molecular signatures were shared by cells of the three hiPSC lines.

152 To better appreciate the discrepancy of the cells at Day 1, the identity of single cells was  
153 annotated based on a human post-implantation embryo dataset<sup>36</sup>, as Epiblast, Primitive Streak,  
154 Nascent Mesoderm, Emergent Mesoderm, Advanced mesoderm and Endoderm  
155 (Supplementary Figure S3A and B). Cells of the C32 line retained an Epiblast signature at Day  
156 1 (55% of cells) and did not display a primitive-streak like cell state like the C11 and C16 cell  
157 lines (Supplementary Figure S3C and D).

158 Finally, individual gene comparison of cell states across the 4-day differentiation showed C32  
159 cells maintained a robust expression of pluripotency-related factors, (*SOX2*, *NANOG*, *POU5F1*).  
160 Interestingly, Nodal targets and antagonists, *LEFTY1* and *LEFTY2*, were up-regulated at Day 1  
161 (Figure 3E). Genes associated with germ layer differentiation, *MIXL1*, *LHX1*, *DKK1*, *DKK4*, and  
162 endoderm related genes (*GSC*, *GATA6*) were down-regulated. While *SOX17* and *FOXA2* were  
163 not down-regulated, the proportion of endoderm cells in the C32 line was reduced relative to  
164 other cell lines (Figure 3E and Supplementary Figure S3D), in agreement with the IHC data.  
165 Collectively, scRNA-seq, bulk RNA-seq and microfluidic RT-qPCR all pointed at a failure or a  
166 delay of the C32 line in differentiating to primitive-streak like cells.



**Figure 3 – scRNA-seq demonstrates that C32 is part of low propensity cell lines: A)** Definitive Endoderm Differentiation protocol used for C32, C16 and C11. Blue circles highlight the samples selected for scRNA-seq. **B)** tSNE plot obtained from scRNA-seq data. **C)** SC3 clustering obtained from scRNA-seq data. **D)** tSNE plot colored for each cell line. **E)** Gene expression for each sample. **F and G)** PCA showing integration of our scRNA-seq data with data from Cuomo et al, **F)** represents cells grouped for each day; **G)** grouped for each sample. The pink arrow indicates time progression during differentiation; **H)** PC1 axis projection of hiPSC at Day 3/4 of DE differentiation representing efficiency of differentiation as pseudotime. **I)** Differentially Expressed Genes (DEGs) between the Top 20 cell lines (better) (blue) versus the Low 20 (less efficient) cell line (red) represented on a Volcano plot.



## 168 **Single-cell transcriptomic analysis as a tool to rank lineage propensity**

169 To test if the information learned from the scRNA-seq analysis from our cell line cohort may be  
170 extrapolated to hiPSC lines in general, the scRNA dataset was combined with that of 125 hiPSC  
171 lines previously profiled for endoderm differentiation *in vitro*<sup>19</sup> (Figure 3F and G). Each cell line  
172 was ranked for endoderm propensity based on their PC1 eigenvalue at Day 4 (Figure 3H) as  
173 described before. The data show that C32 ranks among the lowest 20% propensity cell lines,  
174 while C16 and C11 ranked higher.

175 Using this classification, the Top 20 cell lines (Top20) and the lowest 20 cell lines (Low20), were  
176 analyzed for DEGs. This analysis revealed that genes associated with primitive streak formation  
177 were up-regulated in Top20 cell lines (*TBXT*, *DKK1*, *SNAI1*, *MIXL1*, *LHX1*, etc...) while Low20  
178 cell lines retained a pluripotent signature (*NANOG*, *DPPA2*, *ZIC1*) (Figure 3I). The data infer that  
179 the Low20 lines are less competent for germ layer differentiation and that the C32 characteristics  
180 apply to a number of other hiPSCs of low endoderm differentiation propensity.

181 We conclude that irrespective of the sequencing technology used (SMART-seq2 vs 10X) and  
182 the differentiation protocol (homemade versus commercial), single cell data sets can be used to  
183 rank hiPSC cell lines in terms of endodermal differentiation propensity based on PC1 Eigen  
184 value score. This in turn provides a useful tool that will enable the identification of cell lines with  
185 superior differentiation propensity prior to the use of these cells to productively generate the  
186 endodermal derivatives.

## 187 **Refractory hiPSC showed unique proteomic signature**

188 Although C32 differed from other cell lines at Day 1, the single cell transcriptomic data showed  
189 little difference between C32 and the two other cell lines (C11 and C16) at Day 0, while cells  
190 were pluripotent and undifferentiated. To explore other signatures that may reflect inter-line  
191 differences in endoderm differentiation propensity, the proteome of eleven cell lines was  
192 analyzed at Day 0 (Supplementary Figure S4). This discovery proteome screen revealed that  
193 C32 clustered separately from other groups of hiPSC lines and its isogenic counterpart, C7  
194 clustered with a different group of female cell lines (Supplementary Figure S4A). Compared to  
195 C7, C32 displayed a higher level of expression of pluripotency-related factors, PODXL and  
196 FZD7, and downregulation of FN1 (associated with pharyngeal endoderm)<sup>37</sup> (Supplementary  
197 Figure S4B and C). Of note, PODXL expression is maintained at Day1 of DE differentiation  
198 (Figure S2E) and is also associated with kidney differentiation, a mesoderm derivative<sup>6</sup>,  
199 correlating with our previous observation that C32 might be poised for mesoderm differentiation.  
200 The proteomic signature supports our previous findings that the retention of pluripotency in the  
201 C32 line contributes to the failure of differentiation towards the endoderm lineage.

## 202 **MIXL1 is required for promoting endoderm differentiation**

203 Since our data of the refractory line C32 showed that dysregulation of *MIXL1* expression early  
204 in lineage differentiation may underpin the low endoderm propensity, we next tested the  
205 requirement of *MIXL1* for endoderm differentiation. To this end, frameshift mutations of *MIXL1*  
206 were engineered in C32 and C16 lines by CRISPR editing to generate *MIXL1* loss of function  
207 cell lines: C32-MKO line and C16-MKO line respectively. We validated the introduction of the

208 MIXL1 gene edits by sequencing the targeted regions and confirmed this did not impact  
209 pluripotency of these lines (Supplementary Figure S5).

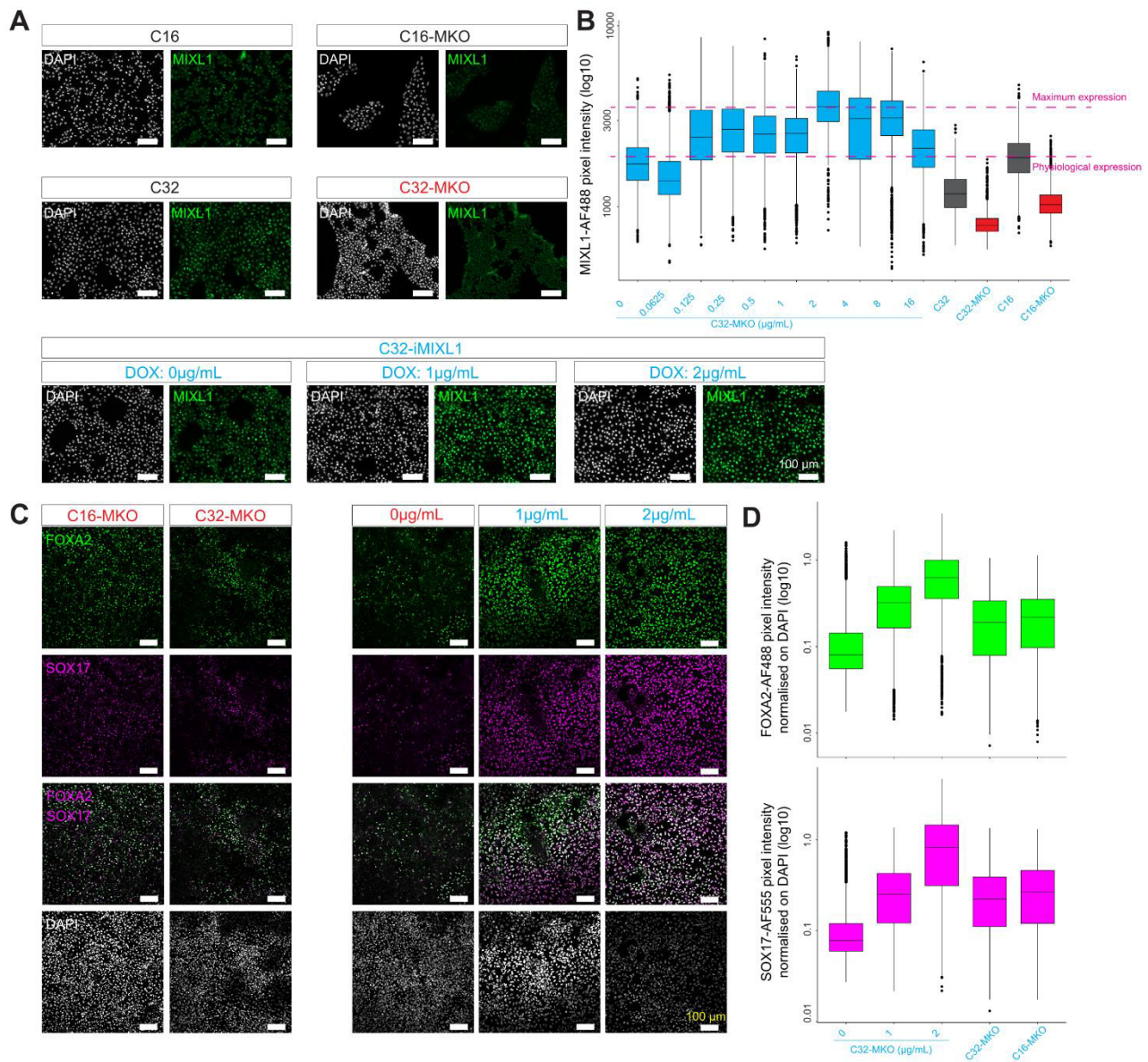
210 Separately, the C32 line was engineered by constitutive expression of a dead Cas9, with no  
211 nuclease activity, linked to VP64, a potent transcriptional activator (dCas9-PVP64)<sup>38,39</sup>. This line  
212 was genetically modified further to express two sgRNAs that targeted the dCas9-VP64 to the  
213 promoter of *MIXL1*, in a doxycycline controllable fashion (C32-Dox)<sup>40</sup>. These two guides  
214 (sgRNA4 and 7 – Figure S6F) displayed the strongest activation of the gene compared to 7 other  
215 guides when tested in HEK cells (Figure S6G).

216 We next quantified *MIXL1* expression in the KO lines and the C32-Dox line with and without  
217 induction at Day 1 of DE differentiation. No *MIXL1* expression can be detected in C32-MKO and  
218 C16-MKO (Figure 4A and B) relative to C32-Dox cells treated with DMSO (C32-Dox0)  
219 (Figure 4A and B). Maximal induction of *MIXL1* expression was achieved, at a saturating  
220 concentration of 2 $\mu$ g/mL Dox. Beyond which, at 16 $\mu$ g/mL Dox, *MIXL1* expression was reduced  
221 possibly due to the toxicity that impacts negatively on mitochondrial gene activity. The levels of  
222 induced expression of *MIXL1* with 1 $\mu$ g/mL of Doxycycline, quantified by immunofluorescence,  
223 were within the physiological range (comparable to C16 at Day 1). Further studies of C32-Dox  
224 cells were therefore performed at 1 $\mu$ g/mL Doxycycline (C32-Dox1) (Figure 4A and B).

225 To assess how outcomes of endoderm differentiation were modulated by different levels of  
226 *MIXL1* expression, FOXA2 and SOX17 expression was quantified after 4 days of differentiation.  
227 Increasing *MIXL1* expression in C32-DOX resulted in higher expression of both markers  
228 (Figure 4C and D), further reinforcing a role of *MIXL1* in promoting endoderm differentiation.  
229 Surprisingly, however, C32-MKO cells displayed FOXA2 and SOX17 expression levels similar  
230 to C16-MKO and C32-Dox0. This finding suggests that *MIXL1* dysregulation may not be the sole  
231 cause of inefficient endoderm specification and differentiation. However, these results indicate  
232 that modulation of *MIXL1* expression can have an effect on DE formation.

### 233 **MIXL1 plays a role in chromatin organization**

234 To elucidate the impact of *MIXL1* on chromatin accessibility, Assay for Transposase-Accessible  
235 Chromatin using sequencing (ATAC-seq) was performed on C16 and C16-MKO cell lines at Day  
236 1 of DE differentiation, when *MIXL1* expression is maximal. Comparing the accessible regions  
237 (or reads pile-up called as peaks) showed that *MIXL1* deletion led to multiple changes in  
238 chromatin accessibility (Supplementary Figure S6A), and in particular, less closing regions  
239 (Supplementary Figure S6B), than opening (Supplementary Figure S6C). This observation  
240 suggests that *MIXL1* may be responsible for opening and closing regions during germ layer  
241 differentiation. To understand the role of these regions, motif discovery was performed on  
242 differentially accessible chromatin regions. This analysis revealed that peaks with less  
243 accessibility in MKO lines contain the motifs TAATNNNATTA (PROP1, PHOXA2), which is the  
244 dual homeobox motif recognized by *MIXL1* (Supplementary Figure S6D). In the absence of  
245 *MIXL1*, more accessible peaks are associated with TEAD1 and FOXH1 motifs (Supplementary  
246 Figure S6E). FOXH1 is known as a cofactor of GSC which negatively regulates *MIXL1* in the  
247 mouse<sup>41</sup>. TEAD1 is the transcription factor bound by YAP/TAZ when Hippo signaling is  
248 inhibited<sup>42</sup>. This is consistent with the observation that the C32 cell line expresses low *MIXL1*



**Figure 4 – MIXL1 functional genomic study reveals its role in endodermal differentiation:**

**A)** Immunostaining images on MIXL1 (green) for C16 and C32 either WT or MIXL1-KO (MKO) and C32 with inducible MIXL1 expression under three different concentrations of doxycycline (0, 1 and 2µg/mL) called C32-iMIXL1. Nuclei are revealed by DAPI (white). **B)** MIXL1 signal intensity normalized on DAPI signal for concentration of doxycycline spanning from 0 to 16µg/mL, WT and MKO conditions. 2 levels are highlighted a physiological level, corresponding to the level of C16 and 1µg/mL of doxycycline and an overexpression level corresponding to 2µg/ml. (n= 3). **C)** Immunostaining images of FOXA2 (green) and SOX17 (magenta) on Day 4 of DE differentiation in MKO and iMIXL1 cell lines. **D)** Quantification of FOXA2 and SOX17 immunostaining.

250 expression and exhibits up-regulation of YAP/TAZ targets (*CCN1*, *CCN2*) (Supplementary  
251 Figure S2E).

252 Collectively the data reveal that *MIXL1* may be a pioneer transcription factor involved in  
253 modulating chromatin accessibility of its targets and possibly influences signaling pathways such  
254 as Hippo and WNT, as inferred from the TCF3 motifs found in open regions in the C16-MKO  
255 lines.

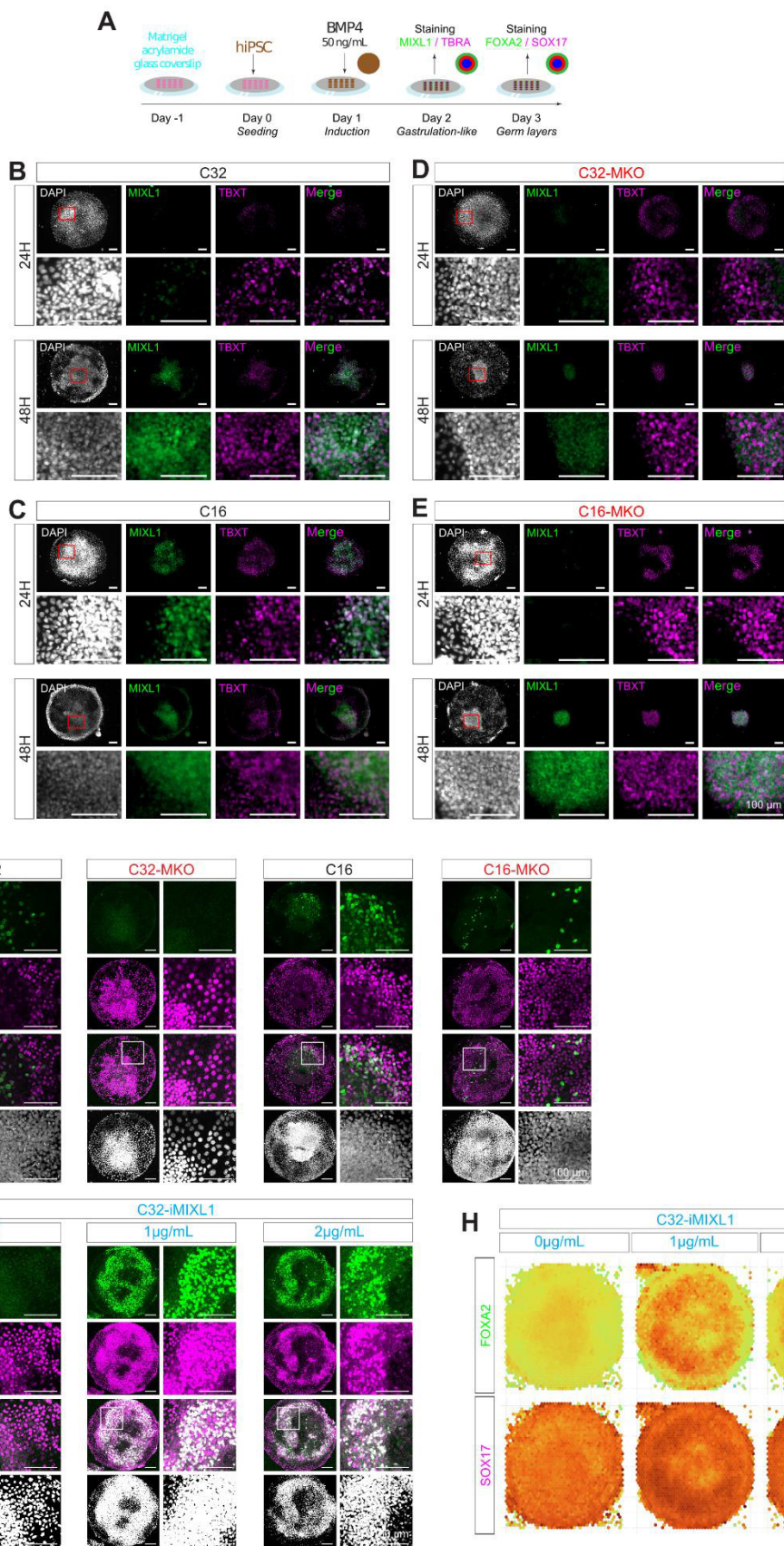
### 256 **Physiological levels of *MIXL1* activity can rescue the endoderm propensity** 257 **of refractory cells in germ layer differentiation.**

258 To assess the function of *MIXL1*, we used the 2D stem cell micropattern model to elucidate its  
259 functional attribute in germ layer differentiation (Figure 5A, Supplementary Figure S7 and S8).  
260 Five cell lines, C32, C32-MKO, C32-Dox, C16 and C16-MKO, were used to generate  
261 micropatterned cultures that recapitulate germ layer formation in response to BMP4<sup>33,43,44</sup>.

262 The emergence of primitive streak-like cells was assessed via the immunostaining of TBXT and  
263 *MIXL1* proteins 24h after BMP4 supplementation. In the C32 cell line, similar expression kinetics  
264 were observed, but was temporally delayed from 48h onwards. At 24h *MIXL1* signal was almost  
265 undetectable (Figure 5B). In the C16 line TBXT and *MIXL1* were expressed at peak level at 24h  
266 followed by decreased expression at 48h. Some cells were co-expressing both proteins (Figure  
267 5C). At 48h, DE differentiation was assessed by FOXA2 and SOX17 expression.  
268 FOXA2+/SOX17+ cells were sparse in the C16 line (Figure 5F). This low propensity for DE cells  
269 in micropattern has already been documented<sup>33,43,45</sup>. However, no double positive cells were  
270 detected in C32 line (Figure 5F). Instead, separate domains of FOXA2+ cells (inner ring) and  
271 SOX17+ cells (outer ring) were identified. These results from the C16 and C32 lines confirm that  
272 a change in culture format could not rescue the phenotype of refractory endoderm differentiation  
273 of the C32 line.

274 MKO lines were analyzed similarly and *MIXL1* could not be detected in both KO lines, and TBXT  
275 did not appear to be affected in its spatiotemporal pattern (Figure 5D and E). Regarding DE  
276 differentiation, the main effect of the loss of *MIXL1* activity was the reduced population of  
277 FOXA2+ cells (Figure 5F).

278 To elucidate the effect of induced *MIXL1* activity on DE differentiation of the incompetent iPSC  
279 line, the C32Dox lines were cultured in micropatterns under BMP4 condition with induction by  
280 DOX at 2 doxycycline concentration: 1µg/mL, corresponding to physiological condition and  
281 2µg/mL, a condition where *MIXL1* is overexpressed. Endoderm differentiation at Day 2 of  
282 differentiation of induced C32Dox line (C32-iMIXL1), compared with C32 line (parental, low  
283 propensity line) and C16 line (high propensity line), was assessed by the presence of SOX17+  
284 and FOXA2+ cells in the micropatterns. C32-induced cells displayed increased number and  
285 density of SOX17+ and FOXA2+ cells (Figure 5G, H) compared to C32 line and C16 line (Figure  
286 5F) and the control condition without doxycycline. Reconstitution of physiological levels of *MIXL1*  
287 activity therefore restored the endoderm propensity of iPSCs that are inherently incompetent for  
288 endoderm differentiation. In addition, increasing the doxycycline concentration increased the  
289 level of double positive cells in the micropatterned cultures indicating a correlation between  
290 *MIXL1* level and DE cells formation.



**Figure 5 – MIXL1 induction rescues DE phenotype in a pseudo-embryo model.** **A)** Protocol to establish stem cell-micropattern model of germ layer differentiation; **B to E)** Immunostaining images of MIXL1 and TBXT at 24h and 48h after BMP4 induction on C32 (**B**), C16 (**C**), C32-MKO (**D**) and C16-MKO (**E**). **F and G)** Immunostaining images of FOXA2 (green) and SOX17 (magenta) at 48h after BMP4 induction on C32, C32-MKO, C16 and C16-MKO (**G**) and C32-iMIXL1 under three different concentrations of doxycycline (0, 1 and 2µg/mL). **H)** Average signaling for FOXA2 and SOX17 measured on micropatterns (n= 10).

## 292 Discussion

293 In this study, isogenic human induced pluripotent stem cell (hiPSC) lines were analyzed and  
294 compared for their propensity in generating definitive endoderm that is capable of progressing  
295 to functional endoderm derivatives, here tested by the formation of intestinal organoids and  
296 hepatocytes (See Figure 2). One cell line of the cohort (C32) was found to be inefficient in  
297 differentiation towards the endoderm lineage. The low propensity for endoderm derivatives is  
298 accompanied by a bias toward the mesoderm lineage. Our results suggest that during germ  
299 layer differentiation, C32 activates the genetic program of vascularization and heart formation  
300 more efficiently than the rest of the cohort (See Supplementary Figure S2C). In addition, the C32  
301 hiPSC line has been used in a previous study to produce kidney organoid<sup>6</sup> supporting the  
302 potential mesoderm bias of the C32 cell line.

303 We sought to determine molecular markers of endoderm propensity of hiPSC lines, at the  
304 pluripotency and early exit stages. Transcriptomic and proteomic analyses did not reveal evident  
305 bias in lineage propensity of the refractory line C32, except for a higher level of expression of  
306 some pluripotency-related factors at the proteomic level (Supplementary Figure S4). The higher  
307 pluripotency level of C32 may underpin the poor performance in endoderm differentiation of this  
308 cell line. To gain a holistic view of the differentiation potential of isogenic hiPSC lines, 4 groups  
309 of isogenic hiPSC lines (2 males and 2 females) were subjected to deep transcriptomic analysis.  
310 This again highlighted the failure of C32 in activating the gastrulation genetic program properly.  
311 In other words, C32 differentiation was inefficient and delayed in initiating germ layer formation  
312 compared to its isogenic clone C7 and the rest of the cohort.

313 The single cell transcriptomic data was analyzed in conjunction with the transcriptomic data of a  
314 previous large scale study<sup>19</sup> surveying 125 hiPSC lines during definitive endoderm  
315 differentiation. Despite the use of different protocols (homemade medium versus commercial kit  
316 for the present study) and sequencing chemistry (SMART-seq2 versus 10X Chromium for the  
317 present study), the data were remarkably comparable. We found that the cell lines with the  
318 lowest differentiation score at the end of the DE differentiation, have a significantly low  
319 expression of genes involved in gastrulation including *MIXL1*. These results further validated our  
320 microfluidic RT-qPCR findings and enabled the identification of a gene panel and novel tool for  
321 ranking hiPSC lines for the propensity of endoderm differentiation and ability to generate mature  
322 endoderm tissues.

323 The transcriptomic survey of the cohort revealed that the gene *MIXL1* is expressed at a low level  
324 in endoderm-incompetent cell lines, and this is corroborated via cross comparison of the  
325 aforementioned study<sup>19</sup>. Our endoderm differentiation data of stem cell-derived micropattern  
326 further points to a causal relationship between the expression of *MIXL1* and efficiency of DE  
327 differentiation. We proposed that *MIXL1* is a useful biomarker for screening human pluripotent  
328 stem cells for competency of endoderm differentiation by quantification of *MIXL1* expression in  
329 cells at 24h of directed differentiation.

330 The ATAC-seq data pointed to a possible role for *MIXL1* in regulating chromatin accessibility. Of  
331 note, the target regions opened in the *MIXL1*-KO cell line are strongly enriched for TEAD1  
332 binding sites. TEAD1 is a transcription factor involved in Hippo signaling<sup>42</sup>. This data, when  
333 combined with our observation that Hippo target genes (e.g., *CCN1*, *CCN2*) are more strongly  
334 activated in the refractory C32 line (Supplementary Figure S2E), suggests an important

335 relationship between chromatin status and the Hippo pathway that is regulated by *MIXL1*.  
336 Interestingly, the closed chromatin region in the MIXL1-KO line mainly contains the dual  
337 homeobox binding sites TAATNNNATTA, recognized by MIXL1. This suggests that MIXL1 may  
338 be involved in regulating accessibility to its own transcriptional targets.

339 Our study has provided a comprehensive survey of endoderm differentiation in an isogenic  
340 cohort and provides a framework for future study of the molecular mechanisms that underpin  
341 endoderm specification during germ layer differentiation of pluripotent stem cells and in  
342 embryonic development.

## 343 **Acknowledgments**

344 We would like to acknowledge the support of Advanced Imaging for all imaging and Vector and  
345 Genome Engineering facilities at CMRI for the help in generating the KO lines. The Advanced  
346 Microscopy Facility at CMRI for the image generation. Prof. Kristopher Kilian and his team for  
347 sharing the micropattern protocols and training. This work was supported by the National Health  
348 and Medical Research Council (NHMRC) (Project Grant ID1127976). PPLT was supported by  
349 Senior Principal Research Fellowship (NHMRC Grant ID110751)

## 350 **Author contributions**

351 PO, EW and PPLT conceived of the study. PO, AW, Ni.S, PPLT designed experimentation. PO  
352 and Ni.S were the primary scientist and completed most of the cell and molecular biology, live  
353 and fixed imaging, and biochemical analysis. AQ, Jy.S, XBL assisted with molecular biology, live  
354 and fixed imaging, and biochemical analysis. HK and Ni.S completed the RNA-seq and scRNA-  
355 seq and NA performed the associated analysis. SC helped with preliminary analysis. Ni.S  
356 completed ATAC-seq and Na.S the associated analysis. Ja.S, IP, GC and MG completed all  
357 mass spectrometry and associated analysis. PO, EW and PPLT secured funding. PO completed  
358 experimental analysis and interpretation and created the figures with assistance from NA and  
359 PPLT. PO, Ni.S and PPLT wrote the manuscript with assistance from all authors.

## 360 **Conflict of interest**

361 We the authors declare no conflict of interest

## 362 **Figures Legend**

### 363 **Figure 1 – Definitive endoderm differentiation heterogeneity among isogenic lines:**

364 **A)** Differentiation protocol used to generate definitive endoderm (DE) cells; **B)** Immunostaining  
365 images of DE cells on FOXA2 (green) and SOX17 (magenta), scale bar = 100µm; **C)** Signal  
366 intensity measurement of immunostaining of panel B, n= 3. **D)** PCA obtained from microfluidic  
367 RT-qPCR data on DE differentiation time courses. Each day is represented by the same color  
368 as Figure 1A (inserted on top left). Purple squares highlight C32 at Day 4 of DE differentiation;

369 **E)** PC1 axis projection of hiPSC at Day 4 of DE differentiation representing efficiency of  
370 differentiation as pseudotime. C32 is highlighted in purple. **F)** PCA obtained from microfluidic  
371 RT-qPCR data on DE differentiation, identical dataset to figure 1D, but plotted for each day. C32  
372 is highlighted in purple; **G)** Genes' contribution to PC1 and PC2 axis of the Day 1 PCA of Figure  
373 2F. The position of the arrows correlates with the position of the samples on the PCA at Day 1.  
374 Colour scale is between 0 and 1. A score of 1 indicates maximum contribution of a particular  
375 gene to PC1. **H)** Genes' expression time course during DE differentiation. C32 is highlighted in  
376 purple. p.value: \* < 0.05, \*\* < 0.01, \*\*\* < 0.001.

### 377 **Figure 2 – Low endodermal propensity fails to produce functional tissue:**

378 **A)** Differentiation protocol used to generate hepatocytes from C32 and C11; **B)** PCA obtained  
379 from microfluidic RT-qPCR data on hepatocyte differentiation; **C)** Brightfield pictures of  
380 hepatocytes differentiation; **D)** Immunostaining images on AAT (green) and ALB (magenta)  
381 markers of hepatocytes differentiation; **E)** Results of fluorescent analysis of CYP3A4 activity; **F)**  
382 Differentiation protocol used to generate intestinal organoids; **G)** Brightfield pictures of intestinal  
383 organoid differentiation at the Hindgut stage (i), spheroid generation stage (ii) and maintenance  
384 stage (iii); **H)** PCA obtained from microfluidic RT-qPCR data on intestinal organoid  
385 differentiation; **I)** Immunostaining of intestinal organoids of C11 cell lines on different cell types  
386 of the gut epithelium: Goblet cells (UEA-1), Intestinal Stem cell (SOX9), enteroendocrine cells  
387 (CHGA), epithelium (CDX2) and Paneth cells (LYZ) as well as proliferating cells (Ki67). Nuclei  
388 are revealed by DAPI.

### 389 **Figure 3 – scRNA-seq demonstrates that C32 is part of low propensity cell lines:**

390 **A) Definitive Endoderm** Differentiation protocol used for C32, C16 and C11. Blue circles  
391 highlight the samples selected for scRNA-seq. **B)** tSNE plot obtained from scRNA-seq data. **C)**  
392 SC3 clustering obtained from scRNA-seq data. **D)** tSNE plot colored for each cell line. **E)** Gene  
393 expression for each sample. **F and G)** PCA showing integration of our scRNA-seq data with data  
394 from Cuomo et al, **F)** represents cells grouped for each day; **G)** grouped for each sample. The  
395 pink arrow indicates time progression during differentiation; **H)** PC1 axis projection of hiPSC at  
396 Day 3/4 of DE differentiation representing efficiency of differentiation as pseudotime. **I)**  
397 Differentially Expressed Genes (DEGs) between the Top 20 cell lines (better) (blue) versus the  
398 Low 20 (less efficient) cell line (red) represented on a Volcano plot.

### 399 **Figure 4 – MIXL1 functional genomic study reveals its role in endodermal differentiation:**

400 **A)** Immunostaining images on MIXL1 (green) for C16 and C32 either WT or MIXL1-KO (MKO)  
401 and C32 with inducible MIXL1 expression under three different concentrations of doxycycline (0,  
402 1 and 2µg/mL) called C32-iMIXL1. Nuclei are revealed by DAPI (white). **B)** MIXL1 signal intensity  
403 normalized on DAPI signal for concentration of doxycycline spanning from 0 to 16µg/mL, WT  
404 and MKO conditions. 2 levels are highlighted a physiological level, corresponding to the level of  
405 C16 and 1ug/mL of doxycycline and an overexpression level corresponding to 2µg/ml. (n= 3).  
406 **C)** Immunostaining images of FOXA2 (green) and SOX17 (magenta) on Day 4 of DE  
407 differentiation in MKO and iMIXL1 cell lines. **D)** Quantification of FOXA2 and SOX17  
408 immunostaining.

### 409 **Figure 5 – MIXL1 induction rescues DE phenotype in a pseudo-embryo model.**



410 **A)** Protocol to establish stem cell-micropattern model of germ layer differentiation; **B to E)**  
411 Immunostaining images of MIXL1 and TBXT at 24h and 48h after BMP4 induction on C32 (**B**),  
412 C16 (**C**), C32-MKO (**D**) and C16-MKO (**E**). **F and G)** Immunostaining images of FOXA2 (green)  
413 and SOX17 (magenta) at 48h after BMP4 induction on C32, C32-MKO, C16 and C16-MKO (**G**)  
414 and C32-iMIXL1 under three different concentrations of doxycycline (0, 1 and 2 $\mu$ g/mL). **H)**  
415 Average signaling for FOXA2 and SOX17 measured on micropatterns (n= 10).

## 416 **Resource availability**

### 417 **Lead contacts**

418 Further information and requests for resources and reagents should be directed to and will be  
419 fulfilled by the lead contacts, Patrick Tam (p.tam@cmri.org.au) and Pierre Osteil  
420 (pierre.osteil@uca.fr).

### 421 **Material availability**

422 The materials used in this study are listed in the key resources table. Materials generated by our  
423 laboratory in this study are available on request, however, there are restrictions to the availability  
424 of human iPSC lines due to a Material Transfer Agreement.

### 425 **Data and Code availability**

- 426 • All raw sequencing data can be found on Gene Expression Omnibus under the accession  
427 number GSE260552, GSE260553, GSE260554 and are available as of the date of  
428 publication.
- 429 • The raw mass spectrometry datasets generated in this study are available via PRIDE:  
430 PXD048788, <http://www.ebi.ac.uk/pride/archive/projects/PXD048788>
- 431 • All microfluidic RT-qPCR data can be found on the GitHub page  
432 (<https://github.com/PierreOsteil/ScriptsForOsteilEtAI2024>) and are available as of the  
433 date of publication.
- 434 • All original codes are available as of the date of publication and can be found on the  
435 following GitHub page: <https://github.com/PierreOsteil/ScriptsForOsteilEtAI2024>.  
436 Bioinformatic source codes and their corresponding DOIs are listed in the key resources  
437 table
- 438 • Any additional information required to reanalyze the data reported in this paper is  
439 available from the lead contact upon request.

## 440 **Experimental Model and Subject Details**

### 441 **Cell Lines**

442 A cohort of 11 human iPSC lines composed of 2 to 3 isogenic cell lines from 4 patients (2 males  
443 and 2 females) was provided by the Australian Institute for Bioengineering and Nanotechnology

444 (AIBN), University of Queensland. Briefly, hiPSC lines derived from fibroblast cells or foreskin  
445 tissue were generated using a non-integrating episomal reprogramming system (oriP/EBNA1-  
446 based pCEP4 episomal vectors pE-P4EO2SCK2MEN2L and pEP4EO2SET2K from Addgene)  
447 carrying *OCT4*, *SOX2*, *KLF4* and *CMYC*. All lines maintained a normal karyotype and were  
448 capable of forming teratomas that contained derivatives of the three germ layers<sup>34,35</sup>. For routine  
449 maintenance, hiPSCs were cultured in mTeSR1 (STEMCELL Technologies) on six-well plates  
450 precoated with hESC-qualified Matrigel (Corning). The culture plates were incubated at 37°C  
451 and 5% CO<sub>2</sub>. The medium was changed daily. The colonies morphology was evaluated under  
452 an inverted microscope. Cells were passaged at 70-80% confluency with ReLeSR(STEMCELL  
453 Technologies) at a split ratio of 1/5 to 1/30 depending on the cell line, into a new well of a 6-well  
454 plate. Experiments were approved by the Sydney Children's Hospitals Network Human  
455 Research Ethics Committee under the reference: HREC/17/SCHN/167.

## 456 **Method details**

### 457 **CRISPR-KO engineering**

#### 458 **gRNA design and cloning**

459 Single *S. pyogenes* Cas9 gRNA (GCGCCGCGTTTCCAGCGTACCGG) targeting *MIXL1* exon 1  
460 was designed using Geneious software, (<http://www.geneious.com><sup>46</sup>) based on the presence of  
461 a canonical NGG PAM (underlined in gRNA sequence) at the target site. Potential off-target sites  
462 were identified using Geneious software, (<http://www.geneious.com><sup>46</sup>). gRNA was cloned in  
463 Addgene plasmid 62988 following adopted protocol from Ran et al<sup>47</sup>. Oligos used for cloning  
464 were:

465 Forward: 5' CACCGCGCCGCGTTTCCAGCGTAC  
466 Reverse: 5' AAACGTACGCTGGAAACGCGGCGC.

#### 467 **Nucleofection, clone selection and sequencing**

468 Cells were transfected using a plasmid expressing Cas9 protein and gRNA targeting MIXL1 exon  
469 1 following Amaxa™ 4D Transfection protocol for 20 µl Nucleocuvette® Strip using P3 Primary  
470 Cell 4D-Nucleofector® X Kit with program CA-137. After transfection cells were plated into 10  
471 cm dish, coated with hESC-qualified Matrigel (BD Biosciences), prefilled with mTESR medium  
472 (Stem Cell Technologies) mixed with 100% CloneR (Stem Cell Technologies). Twenty-four  
473 hours post transfection cells were puromycin (Thermo Fisher Scientific) selected with  
474 concentration of 1µg/ml for the next 48 hours. Following puromycin selection media was  
475 changed every day and the percentage of CloneR (Stem Cell Technologies) in media was  
476 reduced during the next days as single cells were dividing and started forming individual  
477 colonies. Single colonies were picked and transferred individually in a single well of a 96 well  
478 plate where they were grown to be split and frozen for further sequencing analysis. Cells were  
479 detached using ReLeSR (Stem Cell Technologies) and clones were frozen as cell aggregates  
480 in CryoStor® CS10 (Stem Cell Technologies). Clone selection, screening of the CRISPR/Cas9  
481 clones for editing events and validation of allelic deletions of individual clones was done following  
482 protocol from Bauer et al. <sup>48</sup> for genomic deletions in mammalian cells. The PCR was designed  
483 to amplify the sequence flanking the gRNA on exon 1 targeting location with the expected

484 amplicon of 800bp. PCR analysis for the presence of indels were done with primers: Forward:  
485 5'GGAGGGTATAAGTGCGGCC Reverse: 5'CCTCATCTGTGTCTTCTTCCCG

486 All PCR reactions were done in 50 $\mu$ l volume using Q5 high fidelity polymerase (NEB) following  
487 NEB Q5 high fidelity PCR protocol. In short, PCR reaction mix was made by mixing 100ng of  
488 genomic DNA sample from each clone with 10 $\mu$ l of 5xQ5 reaction Buffer, 1 $\mu$ l of 10mM dNTPs,  
489 2.5 $\mu$ l of each (forward and reverse) 10  $\mu$ M primer, 10 $\mu$ l of 5xQ5 High GC Enhancer, 0.5 $\mu$ l of Q5  
490 Polymerase and topped up to 50 $\mu$ l with H<sub>2</sub>O. PCR reaction started with initial denaturation with  
491 temperature of 98°C for 30s followed by 34 cycles of 10s denaturation at 98°C, annealing at  
492 60°C for 20s and extension at 72°C for 20s ending with final extension at 72°C for 5min. PCR  
493 reaction was run on 1.5% agarose gel where expected amplicon of 800bp for each analyze clone  
494 was detected. In total, 43 samples were separately amplified by PCR and analyzed by  
495 sequencing for the presence of indels at the exon 1 targeted site. Next, sequenced clones were  
496 analyzed for genome editing and indel percentages were calculated via TIDE<sup>49</sup> using a control  
497 chromatogram for comparison. Decomposition windows, left boundaries, and indel ranges were  
498 optimized to have the highest alignment possible. After TIDE analysis 11 clones were selected  
499 for validation of biallelic deletion clones for targeted genomic region of exon 1, which was done  
500 following standard protocol from Bauer et al.<sup>48</sup>

## 501 **Differentiation protocols**

### 502 **Definitive endoderm differentiation and characterization**

503 For direct differentiation into Definitive Endoderm, the cells were subject to induction using  
504 STEMdiff Definitive Endoderm kit (STEMCELL Technologies) for 4 days, following  
505 manufacturer's protocol. Briefly, cells were passaged into single cells with StemProAccutaseCell  
506 Dissociation Reagent (Life Technologies) and seeded with mTeSR1 containing Y-27632  
507 dihydrochloride Rock Inhibitor (Tocris, Cat. No. 1245). After 24 hours, the cells are washed with  
508 PBS and then cultured for 4 days in STEMdiff Definitive Endoderm Basal medium with  
509 Supplements A and B for the first day and then Supplement B only for the subsequent 3 days,  
510 with daily medium changes. Samples were harvested daily for RNA and Protein extraction (n=3).

511 To characterise the definitive endoderm, cells were seeded onto glass coverslips coated with  
512 hESC-qualified Matrigel (Corning) before treating with STEMdiff™ Definitive Endoderm kit  
513 (STEMCELL) as described above. Cells were fixed in 4% paraformaldehyde in PBS at RT for  
514 20 min. They were washed with PBS twice and then permeabilized with 0.1% Triton X-100  
515 (Merck) in dPBS (Gibco) (PBST) at RT for 5min. The cells were blocked with 3% bovine serum  
516 albumin (Merck Aldrich) in PBST at room temperature for 1 hour. They were incubated with  
517 primary antibody at 4°C overnight (FOXA2 (Abcam) 1:300, SOX17 (R&D Systems) 1:20). Cells  
518 were washed with dPBS three times, then incubated with corresponding secondary antibodies  
519 at RT for 1 hour. The cell nuclei were stained with DAPI (1 $\mu$ g/ml) (Thermo Fisher Scientific) in  
520 dPBS for 10 min at RT, and then washed three times with dPBS. Cells were mounted with  
521 Fluoromount-G (Thermo Fisher Scientific) and imaged on Zeiss Axio Imager Z2 widefield  
522 microscope.

## 523 **Human Intestinal Organoids differentiation and characterization**

524 hiPSC-derived intestinal organoids were formed using the STEMdiffIntestinal Organoid Kit  
525 (StemCell Technologies), following the manufacturer's protocol. Briefly, cells were passaged as  
526 clumps using ReLeSR (StemCell Technologies). Once 80-90% confluency was reached,  
527 differentiation was initiated with DE Medium (STEMdiffEndoderm Basal Medium plus  
528 STEMdiffDefinitive Endoderm Supplement CJ) for 3 days, with daily medium changes.  
529 Subsequent mid-hindgut differentiation was induced with MH Medium (STEMdiffEndoderm  
530 Basal Medium plus STEMdiffGastrointestinal Supplement PK and STEMdiffGastrointestinal  
531 Supplement UB) for 6 days, with daily medium changes. Free-floating mid-/hindgut spheroids,  
532 collected at 24-hour intervals within the MH Medium treatment, were embedded in Matrigel  
533 (Corning) in wells of NunclonDelta surface 24-well plate (Thermo Fisher Scientific), overlaid with  
534 STEMdiffIntestinal Organoid Growth Medium (STEMdiffIntestinal Organoid Basal Medium plus  
535 STEMdiffIntestinal Organoid Supplement (StemCell Technologies) and GlutaMAX (Gibco)),  
536 performing medium change every 3 - 4 days, incubating at 37°C with 5% CO<sub>2</sub>. After 7 - 10 days  
537 of incubation, cultures were passaged. Briefly, all plasticware were pre-wetted with Anti-  
538 Adherence Rinsing Solution (StemCell Technologies). Matrigel domes containing organoids  
539 were broken manually by pipetting up and down with cold DMEM/F-12 (Gibco), seeding 40-80  
540 organoid fragments per 50µl Matrigel dome.

541 Organoids were removed from Matrigel similar to that described for splitting organoids above  
542 and were fixed in 4% paraformaldehyde in PBS at RT for 30 min. They were washed with PBS  
543 twice and then permeabilised with 0.1% Triton X-100 (Merck) in dPBS (Gibco) (PBST) at RT for  
544 1 hour. The organoids were blocked with CAS-Block (Thermo Fisher Scientific) for 90min and  
545 then incubated with primary antibody (Sox9 (Merck) 1:500, Ki67 (Abcam) 1:250, CHGA (Novus  
546 Biologicals) 1:200, CDX2 (Biogenex) 1:250, Lysozyme (Dako) 1:200) overnight at 4°C.  
547 Organoids were washed with PBST four times, then incubated with corresponding secondary  
548 antibodies and stain (DAPI (1 µg/ml) (Thermo Fisher Scientific) for nuclei in all samples and  
549 UEA-1 (Vector Laboratories) 1:200, for select organoids) in CAS-Block at RT for 3 hours. The  
550 organoids were then washed four times with PBST, followed by clearing in FUNGI solution (50%  
551 (v/v) glycerol, 9.4% (v/v) dH<sub>2</sub>O, 10.6M Tris base, 1.1mM EDTA, 2.5M fructose and 2.5M urea)  
552 for 40min. Organoids were imaged using a µ-slide (Ibidi) on Zeiss Cell Observer Spinning Disc  
553 confocal microscope.

## 554 **Hepatocytes differentiation and characterization**

555 hiPSC lines were differentiated toward hepatocytes following the protocol from Baxter et al.<sup>50</sup>  
556 Briefly, cells were directed into Stage 1a/definitive endoderm-like cells by culturing in RPMI  
557 media containing 1mM L-glutamine, 0.5% FBS, 100ng/mL Activin-A and 25ng/mL Wnt3a for 2  
558 days, followed by Stage 1b/definitive endoderm-like cells by culturing in RPMI media containing  
559 1mM L-Glutamine, 0.5% FBS and 100ng/mL Activin-A for a further 2 days. Stage 2/hepatoblast-  
560 like cells was initiated by incubating for a further 6 days with Hepatocyte culture medium (HCM)  
561 containing 20ng/mL BMP2 and 30 ng/mL Fgf4. Hepatocyte-like cells were made by further  
562 incubating the cultures in Stage 3a media (HCM containing 20ng/mL Hepatocyte Growth Factor  
563 (Peprotech)) for 5 days followed by a further 15 days in Stage 2b media (HCM containing 10  
564 µg/mL Oncostatin M (R&D Systems Cat No. 295-OM) and 10nM dexamethasone).

565 11 and C32, by incubation with P450-Glo™ CYP3A4 assay reagent (Promega). The analysis  
566 was performed according to manufacturer recommendation.

## 567 **Micropatterns preparation**

### 568 **Micropattern chip fabrication**

569 Micropattern chip fabrication was conducted using the protocol of Lee et al.<sup>51</sup>, with specific  
570 modification and optimization. In brief, coverslips were sonicated in 70% ethanol for 15min and  
571 in deionized water for 15min. The clean coverslips were sequentially incubated in 0.5% (3-  
572 aminopropyl)triethoxysilane (APTS) (Merck) for 3min, 0.5% glutaraldehyde (Merck, Cat. No.  
573 G6257) for 30min. After air drying, the coverslips were deposited on a 20μL drop made of 10%  
574 acrylamide (Merck), 0.87% bisacrylamide (Merck), 0.1% ammonium persulfate (Merck, Cat. No.  
575 A3678) and 0.1% N,N,N',N'-Tetramethylethylenediamine (Merck), to make the gel at a stiffness  
576 of 100KPa. After the stiffness droplet was semi-solidified, the whole system was submerged into  
577 70% ethanol, resulting in a smooth polyacrylamide gel forming. Gelled coverslips were  
578 sequentially coated with 64% hydrazine hydrate (Fisher Scientific) for 1h and 2% glacial acetic  
579 acid (Merck) for 1h. To generate polydimethylsiloxane (PDMS) stamp, SYLGARD™ 184 Silicone  
580 Elastomer Curing Agent and Base (Dow) were mixed at a 1:10 ratio before loading to the stamp  
581 mold, provided by the Kilian Lab at the University of New South Wales. Next, the solidified PDMS  
582 stamp was coated with 25μg/mL vitronectin (Life Technologies) and 3.5 mg/mL sodium periodate  
583 (Merck) for 1h. After air-drying the stamp, patterned vitronectin was stamped onto the gelled  
584 coverslip at 0.343N for 1min. Stamped gels were stored overnight in PBS + 1% Penicillin-  
585 Streptomycin at 4°C.

### 586 **Germ layer differentiation on micropatterns and analysis**

587 Differentiation protocol was adapted from Warmflash et al.<sup>44</sup> Since the micropatterned chip  
588 generation required many hands-on manipulations, all culture media were supplemented with  
589 1% Penicillin-Streptomycin. hiPSCs were seeded as single cells to micropattern chip at a density  
590 of  $2.5 \times 10^5$  cells/cm<sup>2</sup> with 10μM Y-27632 ROCK inhibitor in mTeSR Plus supplemented 1%  
591 Penicillin-Streptomycin into a total volume of 1mL per well of a 24-well plate. At about 80%  
592 confluency, germ layer differentiation was induced by adding 50 ng/mL BMP4 (R&D) in mTeSR1.  
593 The cells grown on micropatterns were washed with PBS, fixed in 4% paraformaldehyde in PBS  
594 at RT for 20 min. They were washed with PBS twice and then permeabilised with 0.1% Triton X-  
595 100 (Merck) in dPBS (Gibco) (PBST) at RT for 1 hour. The cells were blocked with 3% bovine  
596 serum albumin (Merck) in PBST at room temperature for 1 hour. They were incubated with  
597 primary antibody at 4°C overnight (MIXL1 (Abcam) 1:50, T/Brachyury (Santa Cruz) 1:50, FOXA2  
598 (Abcam) 1:300, SOX17(R&D Systems) 1:20). Cells were washed with PBST three times, then  
599 incubated with corresponding secondary antibodies at RT for 1 hour. The cell nuclei were stained  
600 with DAPI (1 μg/ml) (Thermo Fisher Scientific) in dPBS for 10 min at RT, and then washed twice  
601 more with PBS. Cells were mounted with Fluoromount-G (Thermo Fisher Scientific).  
602 Micropatterned coverslips were imaged on Zeiss AiryScan LSM880 confocal microscope. All  
603 image analysis was performed using a custom macro. The nuclei from micropatterned images  
604 taken were segmented using the StarDist method<sup>52</sup> via Fiji software<sup>53</sup> using default parameters  
605 (except probability/score threshold = 0.7) and the versatile (fluorescent nuclei).pb model. R

606 software was used with a custom script where target protein immunofluorescence was  
607 normalized to the DAPI intensity of the same cell.

## 608 **Microfluidic RT-qPCR preparation and analysis**

### 609 **RNA extraction**

610 Snap frozen cell pellets had total RNA extracted using ISOLATE II RNA mini kit (Bioline)  
611 following manufacturer's instructions. Briefly, samples were lysed, homogenized and passed  
612 through a spin column containing a silica membrane to which the RNA binds. DNase 1 digestion  
613 eliminated potential genomic DNA contamination and the preparation was washed to remove  
614 impurities such as cellular debris and salts. The purified RNA was eluted with RNase free water  
615 and total RNA concentration was determined using Nanodrop ND-1000 Spectrophotometer  
616 (ThermoFisher Scientific). RNA was used either for Microfluidic RT-qPCR or RNA-sequencing.

### 617 **cDNA synthesis and preparation**

618 Total RNA was adjusted to a concentration of 200ng/ $\mu$ l. A 5 $\mu$ l reaction mix was prepared  
619 composing of 1 $\mu$ l Reverse Transcription Master Mix (Fluidigm), 3 $\mu$ l of RNase free water and 1 $\mu$ l  
620 of RNA and incubated in a thermocycler using the following conditions: 25°C for 5min, 42°C for  
621 30min and 85°C for 5min.

### 622 **cDNA preamplification**

623 3.75 $\mu$ l of preamplification mix (comprising 105.6 $\mu$ l of Preamp MasterMix (Fluidigm), 52 $\mu$ l of 100  
624  $\mu$ M pooled primer and 237.6 $\mu$ l DNase/RNase free water) and 1.25 $\mu$ l of cDNA sample was added  
625 into each well of a 96 well plate and incubated as follows: 95°C for 2min then 10 cycles of 95°C  
626 for 15sec and 60°C for 4min.

### 627 **cDNA clean-up**

628 cDNA clean-up was performed by adding 2 $\mu$ l of the following mix: 168 $\mu$ l DNase free water, 24 $\mu$ l  
629 10x Exo1 reaction buffer and 48 $\mu$ l Exonuclease I (New England Biolabs), into each well and  
630 incubating in a thermocycler for 30min at 37°C followed by 15min at 80°C. Samples were diluted  
631 10x with low EDTA TE buffer.

### 632 **Primer and sample set-up**

633 A sample mix was prepared as follows (per 96-well plate): 495 $\mu$ l of 2X SsoFast EvaGreen  
634 SuperMix with low ROX (Biorad) and 49.5 $\mu$ l 25X DNA Binding Dye (Fluidigm). 4.95 $\mu$ l sample  
635 mix was added with 4.05 $\mu$ l of diluted sample. Primers were prepared in the following mix (per  
636 96-well plate): 450 $\mu$ l 2X Assay Loading Reagent (Fluidigm) and 405 $\mu$ l low EDTA TE buffer. 105 $\mu$ l  
637 of primer mix was added with 0.45 $\mu$ l combined forward and reverse primers. Samples and primer  
638 mixes were loaded onto a 96.96 Dynamic Array IFC plate (Fluidigm) and run on the Biomark  
639 System.

## 640 **ATAC-seq samples preparation and analysis**

### 641 **Cell Preparation**

642 hiPSC were collected at Day1 of the DE diff protocol (STEMDiff) and were processed following  
643 <sup>54</sup>. Briefly,  $5 \times 10^4$  cells were collected at Day 1 of DE differentiation and lysed in cold lysis buffer  
644 (10 mM Tris·Cl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% (v/v) Igepal CA-630). Intact nuclei were  
645 separated by centrifugation at 500xg and immediately digested in transposase mix containing  
646 25 $\mu$ l 2x Tagment DNA buffer, 2.5 $\mu$ l Tagment DNA enzyme I (Illumina) and 22.5 $\mu$ l nuclease-free  
647 water for 30min at 37°C. Digested chromatin fragments were then purified using the MinElute  
648 PCR Purification Kit (Qiagen), according to manufacturer's instructions. The fragments of DNA  
649 were then pre-amplified by adding 10 $\mu$ L purified DNA sample, 10 $\mu$ L RNase-free water, 2.5 $\mu$ L of  
650 each primer (Where each reaction had non-barcoded primer "Ad1\_noMix" and one barcoded  
651 primer 'Ad2.1' - 'Ad2.9' added) and 25 $\mu$ L NEBNext High-Fidelity 2x PCR Master Mix (NEB) and  
652 was run under the following conditions: 72°C for 5min, 98°C for 30sec and then 5 cycles of 98°C  
653 for 10sec, 63°C for 30sec and 72°C for 1min. The number of additional cycles to run was  
654 calculated by running a RT-qPCR side reaction - a reaction mixture containing 5 $\mu$ L of the pre-  
655 amplified PCR product, 3.9 $\mu$ L nuclease-free water, 0.25 $\mu$ L of each primer, 0.6 $\mu$ L 25x SYBR  
656 Green and 5 $\mu$ L NEBNext High-Fidelity 2x PCR Master Mix was run under the following  
657 conditions: 98°C for 30sec and then 20 cycles of 98°C for 10sec, 63°C for 30sec and 72°C for  
658 1min. The linear fluorescence versus cycle number was plotted and the cycle number (N)  
659 required to reach one-third of the maximum relative fluorescence was determined. The final  
660 amplification reaction (the remaining 45 $\mu$ l pre-amplified PCR product) was run under the  
661 following conditions: 98°C for 30sec and then N cycles of 98°C for 10sec, 63°C for 30sec and  
662 72°C for 1min. Amplified samples were then purified using AMPure XP magnetic beads  
663 (Beckman Coulter) to remove small fragments and primer-dimers less than 100 bp long (1.3x  
664 beads) and large fragments (0.5x beads) using a Dynamag-2 magnet (Thermo Fisher Scientific).

665 To determine the integrity, fragment size and concentration, the DNA library was analyzed using  
666 the Agilent HSD5000 ScreenTape System (Agilent). Libraries were then 101 bp paired-end  
667 sequenced on an Illumina HiSeq 4000 (Illumina)

### 668 **Data analysis**

669 ATAC-seq reads were processed using the alignment and filtering functions of the PreNet  
670 pipeline<sup>54</sup>. Paired-end reads were mapped to the hg38 genome using bowtie2<sup>55</sup>, allowing for  
671 local mapping, a maximum insert size of 2000 bp and a maximum of 4 multimapping hits (-local  
672 -X 2000 -k 4). Multimapping reads were allocated using 'assignmultimappers.py' from the  
673 ENCODE ATAC-seq pipeline (<https://github.com/ENCODE-DCC/atac-seq-pipeline/tree/master/src>). Reads with MAPQ < 30 were excluded and only unique, paired reads  
674 that aligned outside blacklisted regions<sup>56</sup> were used for subsequent analyses. Filtering steps  
675 were performed using samtools<sup>57</sup> and sambamba<sup>58</sup>. Qualifying reads were then converted to  
676 pseudo-single end reads and peaks were detected using MACS2<sup>59</sup>, with BED input files and  
677 reads shifted by -100 bp and extended to 200bp to capture Tn5 transposase events: -f BED -  
678 shift -100 -extend 200 -q 0.05. Biological replicates were analyzed individually and then  
679 consensus peak list was created to include only peaks appearing in at least two of the three  
680 replicates. Accessibility within a pooled consensus peak list was estimated by quantifying Tn5  
681 events in each of the biological replicates using featureCounts<sup>60</sup>. The DESeq2<sup>61</sup> package within

683 R was used to identify differentially accessible regions. The regions were filtered for adjusted  
684 P-value < 0.05 and an absolute Log<sub>2</sub>(Fold change) > 1. Coverage tracks were created using  
685 the bamCoverage function of deepTools<sup>62</sup> (-normalization RPKM -bs 10) and visualized within  
686 Integrative Genomics Viewer.<sup>63</sup> findMotifsGenome.pl from HOMER<sup>64</sup> was used to identify over  
687 enriched motifs, between 6 bp and 12 bp in size, within regions of differential accessibility using  
688 a repeat masked version of the hg38 sequence (-mask -len 6,7,8,9,10,11,12). Coverage tracks  
689 summarizing and combining biological replicates were created using WiggleTools<sup>65</sup> to quantify  
690 the mean coverage per 10 bp bin. These tracks were used for heatmap visualizations created  
691 using plotHeatmap from deeptools.

## 692 **Bulk RNA sequencing analysis**

### 693 **Cell preparation and library prep**

694 Illumina RNA Library prep was performed by GENEWIZ  
695 (<https://www.genewiz.com/Public/Services/Next-Generation-Sequencing>). Samples at Day 1 of  
696 DE differentiation (1 - 20 $\mu$ g RNA) were run on HiSeq4000 with a read depth of 20M paired end  
697 reads (2x 150PE).

### 698 **Data pre-processing**

699 Details of the procedure can be found in Aryamanesh and colleagues <sup>66</sup>

### 700 **Statistical Analysis**

701 Details of the procedure can be found in Aryamanesh and colleagues <sup>66</sup>

## 702 **Single-cell RNA sequencing**

### 703 **Cell preparation**

704 hiPSCs were dissociated using Accutase and counted to load 10,000 cells into on channel of a  
705 10X Chromium chip. One channel per sample was used. After emulsion cell lysis and RNA was  
706 extracted followed by library preparation. Libraries were sent to Novogene for sequencing.

707 Single cell suspensions were passed through 40 $\mu$ m cell strainer (Corning) and concentration  
708 was adjusted to 1000 cells/ $\mu$ L. Suspensions were loaded in single-cell-G Chip (10X Genomics)  
709 for target output of 10,000 cells per sample. Single-cell droplet capture was performed on the  
710 Chromium Controller (10X Genomics). cDNA library preparation was performed in accordance  
711 with the Single-Cell 3' v 3.0 or v3.1 protocol. Libraries were evaluated for fragment size and  
712 concentration using Agilent HSD5000 ScreenTape System (Agilent).

713 Samples were sequenced on an Illumina HiSeq4000 instrument according to manufacturer's  
714 instructions (Illumina). Sequencing was carried out using 2x150 paired-end (PE) configuration  
715 with a sequencing depth of 20,000 reads per cell. Sequencing was performed by GENEWIZ.

### 716 **Data pre-processing**

717 Details of the procedure can be found in Aryamanesh and colleagues <sup>66</sup>



718 **Statistical Analysis**

719 Details of the procedure can be found in Aryamanesh and colleagues <sup>66</sup>

720

721 **Proteomics**

722 Sample were prepared and data were obtained by the Proteomic facility at CMRI.

723 **Proteomics sample preparation and Mass spectrometry**

724

725 Eleven hiPSC lines from four isogenic groups were prepared for proteomics analyses using the  
726 Accelerated Barocycler lysis and extraction digestion sample preparation method<sup>67</sup>. The tryptic  
727 peptides were desalted using Waters Oasis C18 HLB 30mg SPE cartridges. The amount of  
728 peptide in each sample was measured using the absorption of 280 nm light (Implen  
729 Nanophotometer, Labgear, Australia).

730

731 A reference sample was prepared by pooling equal amount of peptide from 25 of the 45 hiPSC  
732 samples. Aliquots containing 10 µg of peptide from each sample were labeled with tandem mass  
733 tag (TMT) 16-plex reagents (TMTpro, Thermo Fisher Scientific), according to the manufacturer's  
734 instructions. Three separate TMT16plex sets were prepared with eleven samples per set and  
735 the reference sample included in each set.

736

#	hiPSC line_Name	cell	TMT_ Set-1	TMT_ Set-2	TMT_ Set-3	TMTpro 16-plex label
1	Eu79		+	+	+	126
2	Eu86		+	+	+	127N
3	Eu87		+	+	+	127C
4	C9		+			128N
5	C11				+	128C
6	C16		+			129N
7	C7		+			129C
8	C32		+			130C
9	C4		+	+	+	131N
10	C2		+	+	+	131C
11	C3			+	+	132N
12	Universal reference sample	/				134

737

+, samples were combined to produce the reference sample used as the common sample in the three TMT 16plex experiments.

738

739

740 High pH fractionation was performed using the Pierce High pH Reversed-Phase Peptide  
741 Fractionation Kit according to the manufacturer's instructions. A total of 80 µg of peptide for each  
742 of the TMT16-plex sets were loaded, washed with water then 5% Acetonitrile/0.1% Triethylamine  
743 solution. A total of fifteen High-pH step elution's were collected from 8% to 50% Acetonitrile /  
744 0.1% Triethylamine. The high-pH elution's were dried to completeness and resuspended in 0.1%  
745 Formic acid and the peptide concentration was determined using the absorption of 280 nm light.  
746

747 The peptides from each High pH elution were resolved by reversed phase chromatography on  
748 a 300 x 0.075 mm column packed with ReproSil Pur C18 AQ 1.9 mm resin (Dr Maisch, Germany)  
749 using an Ultimate 3000 RSLC nano system (Thermo Fisher Scientific). The column was heated  
750 to 50 °C using a column oven (PRSO-V1, Sonation lab solutions).  
751

752 The chromatography buffer A was 0.1% formic acid in water and buffer B was 0.1% formic acid,  
753 90% acetonitrile and 9.99% water and the flow rate was 250 nL/min. For each high pH fraction,  
754 1 to 2 µg of peptide was directly loaded onto the column in 99% buffer A and 1% buffer B for 30  
755 min. The gradient started from 1% to 7% buffer B in 6 min, then to 30% buffer B in 51 min, then  
756 to 35% buffer B in 10 min and to 99% buffer B in 3 min and held at 99% buffer B for 8 min. MS  
757 acquisition was performed for the entire 120 min. The fifteen high pH elution steps collected for  
758 each TMT16-plex set were individually analysed using a data-dependent acquisition LC-MS/MS  
759 method. Between each TMT16plex set, one blank was run.  
760

761 Peptides were detected by tandem mass spectrometry using a Q Exactive Plus hybrid  
762 quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific). The Nanospray Flex ion  
763 source (Thermo Fisher Scientific) spray operated at 2.3 kV. The capillary temperature was  
764 250°C and the S lens radio frequency level was 50. The MS scan was from m/z 375 to 1500 at  
765 a resolution of 70,000 full width at half maximum with an automatic gain control target of  $3 \times 10^6$   
766 counts for a maximum ion time of 100 ms. For each MS scan, up to 12 of the most intense ions  
767 above a threshold of  $5.2 \times 10^4$  counts were selected for an MS/MS scan. MS/MS scans were at  
768 a resolution of 35,000 full width at half maximum for a maximum ion time of 115 ms and  
769 automatic gain control target of  $2 \times 10^5$  counts. The isolation window was 1.1 units of the m/z  
770 scale, the fixed first mass was set at m/z 120 and the normalized collision energy was 30.  
771 Peptides with charge state  $<2+$  or  $>8+$  or with unassigned charge were excluded. Dynamic  
772 exclusion of previously scanned peptides was for 35 s.  
773

774 The raw LC-MS/MS data were processed with MaxQuant v1.6.7.0<sup>68</sup> using the following settings:  
775 The fasta file was the Human reference proteome downloaded from UniProtKB on January 12,  
776 2022 and containing 101,017 entries including protein isoforms and Retention time standards.  
777 Protease specificity was Trypsin/P with up to 3 missed cleavages. Carbamidomethyl (C) was a  
778 fixed modification and the TMTpro 16plex reagents were designated isobaric labels.  
779 Deamidation (N and Q), oxidation (M) and acetylation (protein N-terminus) were variable  
780 modifications. A maximum of 5 modifications per peptide was allowed. The minimum score for  
781 modified peptides was 40. The minimum peptide length was 6 and maximum peptide mass was  
782 6,000 Da. The peptide spectrum match, protein, and modification site false discovery rate was  
783 1%. A dependent peptide search was performed with a 1% false discovery rate. Modified  
784 peptides and their counterpart non-modified peptides were excluded from protein quantification.  
785 A second peptide search was enabled. The tolerance for MS and MS/MS spectra was 4.5 ppm

786 and 20 ppm, respectively. All other settings were left as the default within MaxQuant v1.6.7.0.  
787 The three TMTpro 16plex sets were searched together (Set1, Set2, Set3); fractions 1 to 15 for  
788 each TMT set representing the High pH elution fractions.

789

## 790 **Data cleaning, normalization and hypothesis testing**

791 The data cleaning, normalization and hypothesis testing were performed using the  
792 ProteomeRiver pipeline<sup>69</sup> and the implementation is briefly described here. To extract the protein  
793 abundance data from the MaxQuant output, the 'proteinGroups.txt' output file from MaxQuant<sup>70</sup>  
794 were processed. Each protein group must have at least one unique peptide. Any proteins were  
795 removed from further analysis if they match any entries in the contaminants or the reverse  
796 sequence decoy databases, in which the protein accession starts with CON\_ or REV\_ prefixes  
797 respectively. The 'reporter intensity corrected' column were used for further analysis. Proteins  
798 with one or more missing values in any samples were removed from further analysis.

799 The following rules from Engholm-Keller et al.<sup>71</sup> were used to identify a representative UniProt  
800 accession for each protein group. 1. For proteins that mapped to multiple UniProtKB protein  
801 accessions, the accession with the highest 'protein existence (PE)' value was kept as the best  
802 evidence. Where the protein accession was an isoform (therefore lacking PE information), the  
803 PE value was taken from the parent protein. 2. When the PE value was equal, a Swiss-Prot (sp)  
804 entry was taken over a TrEMBL (tr) entry. 3. If both entries were Swiss-Prot, the non-isoform  
805 was selected. 4. If both entries were isoforms, the longest isoform was selected.

806 To perform data normalization, samples were log (base 2) transformed and between sample  
807 normalization were performed using scaled normalization from the 'limma' R package. The  
808 remove unwanted variation 'ruv' R package<sup>72</sup> was used to remove batch effects. The method  
809 relies on having a set of endogenous negative control proteins, which are proteins with little  
810 changes in protein abundances between different cell types or experimental treatments. For this  
811 study, a set of 500 empirical negative control proteins with high q-values indicating little or no  
812 change in protein expression across sample were identified from an initial ANOVA test. The  
813 RUVIII method<sup>73</sup> was used to remove the unwanted variations across the samples and six  
814 unwanted components (k = 7) were removed by the tool. The RUVIII method requires the  
815 experiment design matrix, a matrix describing the replicates for each treatment condition, and  
816 the list of negative control proteins.

817 Differential abundance analysis of proteins was performed using the adjusted abundance matrix.  
818 Differential abundance analysis of proteins involved pairwise hypothesis testing of samples from  
819 FA3 cell line with samples of another type of cell line and all possible pairs were analyzed. Linear  
820 model for comparing each pair of time points was fitted using the 'lmFit' function and the p-values  
821 calculated using the empirical Bayes method 'eBayes' function. Trended and robust analysis  
822 were enabled. The false discovery rate correction was applied to the moderated p-values by  
823 calculating the q-values<sup>74</sup>. Significant differentially expressed proteins were defined as those  
824 with q-values less than 0.05.

## 825 Clustering Analysis

826 For a protein to be included in the clustering, it must be statistically down-regulated (q-value  
827 <0.05) in six or more FA3 versus another cell line, or statistically up-regulated (q-value <0.05) in  
828 six or more FA3 versus another cell line. The protein must also be statistically significant (q-  
829 value <0.05) in at least one of the comparisons of FA3 with any of the KO or DOX cell lines. The  
830 z-standardized log<sub>2</sub> abundance of the sample, excluding samples of the KO or DOX cell lines,  
831 were used in the clustering analysis (e.g. 10 values were used in clustering). Consensus  
832 clustering were performed 'diceR' R library<sup>75</sup>. All statistically significant differentially abundant  
833 proteins. After assessment with the consensus clustering tools, using multiple clustering  
834 algorithms, including 'pam', 'km', 'som', 'hc', and 'diana', and different distance metrics  
835 'euclidean', 'canberra', 'minkowski', and 'spearman', self-organizing maps (som) with seven  
836 clusters (k=7) was identified to be a reliable method to use. The diceR tool automatically  
837 identifies robust consensus clusters by merging the results from 100 runs of self-organizing  
838 maps. The clusters identified were used for functional enrichment analysis (see below for  
839 details).

## 840 Functional Enrichment

841 Functional enrichments were performed using the Fisher's exact test implemented in the  
842 'clusterProfiler' R library<sup>76</sup>. The background list of proteins included all the proteins in the dataset  
843 after the data cleaning step. The query list of proteins includes the following: 1) significantly  
844 differentially abundant proteins with positive log fold-change, 2) significantly differentially  
845 abundant proteins with negative log fold-change, 3) the list of proteins from each cluster from  
846 self-organizing maps. Gene ontology annotations from UniProt<sup>77</sup> and the KEGG<sup>78</sup> and  
847 Reactome<sup>79</sup> pathway databases were used for the enrichment analyses.

## 848 Quantification and Statistical Analysis

849 Statistical analysis was performed using R software. The type of statistical test performed, the  
850 meaning of dispersion and precision measurements as well as the significance of each  
851 experiment is indicated in the corresponding figure, figure legends and/or in the method details.  
852 Outliers have been omitted to facilitate visualization. For micropattern quantification, images  
853 were taken from 3 to 10 micropatterns within a coverslip. For microfluidic PCR, three biological  
854 replicates were collected for each cell line and timepoint.

## 855 Supplementary Figures

856 **Figure S1: A)** Brightfield pictures of definitive endoderm differentiation for the 5 cell lines. Day4  
857 images were zoomed in (inlet) to show the cell morphology of the endodermal cells at the end  
858 of the differentiation protocol. Scale bar are indicated on the lowest right image of each panel;  
859 **B)** Immunostaining quantification of FOXA2 (green) and SOX17 (magenta); related to Figure 1;  
860 **C)** Immunostaining images on AAT (green) and ALB (magenta) markers of hepatocytes  
861 differentiation;

862 **Figure S2: A)** Genes' expression time course during DE differentiation related to **i)** Pluripotency,  
863 **ii)** Endoderm, **iii)** Mesoderm and **iv)** Ectoderm. C32 is highlighted in purple. **B)** Heatmap of  
864 differentially expressed genes (DEGs) of C32 cell lines versus the rest of the cell line used in the  
865 study. **C)** Venn diagram of up and down DEGs compared with relevant ontologies. **D)** PCA  
866 representing data of bulk RNAseq of Day1 samples. **E)** DEGs between C32 and C7. Dots that  
867 are colored have a p.value < 0.05.

868 **Figure S3: A)** Schematic of the annotation transfer performed in our study; **B)** Result of the  
869 annotation transfer applied on our scRNA-seq dataset; **C)** The same results faceted for each  
870 cell line; **D)** Stacked barplot representing the proportion of each cell type for each cell line.

871 **Figure S4: A)** PCA representing the proteomic signature of hiPSC; **B)** Volcano plot representing  
872 peptides significantly differentially expressed between C32 and C7; **C)** Allocation of peptides to  
873 proteins and respective LogFC

874 **Figure S5:** Staining for SOX2 (green) and OCT4 (magenta) on three colonies of **A)** C16; **B)**  
875 C16-MKO; **C)** C32; **D)** C32-MKO. Scale bar = 100µm

876 **Figure S6: A)** Volcano plot representing Differentially Accessible Chromatins (DACs) region  
877 between C16 and C16-MKO. **B and C)** Heatmap showing ATAC-seq signal from C16 and C16-  
878 KO for DACs (**B)** more accessible in C16 and (**C)** in C16-MKO. **D and E)** Motifs enriched in  
879 DACs more accessible in (**D)** C16 and (**E)** C16-MKO. **F)** Genomic region of 350bp upstream of  
880 *Mixl1* TSS. 7 guides RNA predicted by Benchling are highlighted. **G)** RT-qPCR results on *Mixl1*  
881 expression in HEK cells transfected with single guides or tandems.

882 **Figure S7:** Panels of 10 micropatterns at 48h post BMP4 treatment, stained with FOXA2, SOX17  
883 and DAPI for **A)** C32, **B)** C32-MKO, **C)** C16 and **D)** C16-MKO. **E)** Indicates the average cell  
884 density for all micropattern analyzed.

885 **Figure S8: A)** Panels of 9 micropatterns at 48h post BMP4 treatment, stained with FOXA2,  
886 SOX17 and DAPI for C32-iMXL1 at 0, 1 and 2µg/mL of Doxycycline. **B)** Indicates the average  
887 cell density for all micropattern analyzed.

888

889 **References**

- 890 1. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined  
891 Factors. *Cell* **131**, 861–872 (2007).
- 892 2. Dimos, J. T. *et al.* Induced Pluripotent Stem Cells Generated from Patients with ALS Can Be  
893 Differentiated into Motor Neurons. *Science* **321**, 1218–1221 (2008).
- 894 3. Friedman, C. E. *et al.* Single-Cell Transcriptomic Analysis of Cardiac Differentiation from Human  
895 PSCs Reveals HOPX-Dependent Cardiomyocyte Maturation. *Cell Stem Cell* **23**, 586-598.e8  
896 (2018).
- 897 4. Nayler, S. P. *et al.* Human iPSC-Derived Cerebellar Neurons from a Patient with Ataxia-  
898 Telangiectasia Reveal Disrupted Gene Regulatory Networks. *Front. Cell. Neurosci.* **11**, 321  
899 (2017).
- 900 5. Park, I.-H. *et al.* Disease-Specific Induced Pluripotent Stem Cells. *Cell* **134**, 877–886 (2008).
- 901 6. Takasato, M. *et al.* Kidney organoids from human iPS cells contain multiple lineages and model  
902 human nephrogenesis. *Nature* **526**, 564–568 (2015).
- 903 7. Cheng, X. *et al.* Self-Renewing Endodermal Progenitor Lines Generated from Human Pluripotent  
904 Stem Cells. *Cell Stem Cell* **10**, 371–384 (2012).
- 905 8. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive  
906 endoderm. *Nat. Biotechnol.* **23**, 1534–1541 (2005).
- 907 9. Loh, K. M. *et al.* Efficient Endoderm Induction from Human Pluripotent Stem Cells by Logically  
908 Directing Signals Controlling Lineage Bifurcations. *Cell Stem Cell* **14**, 237–252 (2014).
- 909 10. Green, M. D. *et al.* Generation of anterior foregut endoderm from human embryonic and induced  
910 pluripotent stem cells. *Nat. Biotechnol.* **29**, 267–272 (2011).

- 911 11. McCracken, K. W., Howell, J. C., Wells, J. M. & Spence, J. R. Generating human intestinal tissue  
912 from pluripotent stem cells in vitro. *Nat. Protoc.* **6**, 1920–1928 (2011).
- 913 12. Miura, S. & Suzuki, A. Brief summary of the current protocols for generating intestinal organoids.  
914 *Dev. Growth Differ.* **60**, 387–392 (2018).
- 915 13. Spence, J. R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue  
916 in vitro. *Nature* **470**, 105–109 (2011).
- 917 14. D'Amour, K. A. *et al.* Production of pancreatic hormone–expressing endocrine cells from human  
918 embryonic stem cells. *Nat. Biotechnol.* **24**, 1392–1401 (2006).
- 919 15. Hannan, N. R. F., Segeritz, C.-P., Touboul, T. & Vallier, L. Production of hepatocyte-like cells from  
920 human pluripotent stem cells. *Nat. Protoc.* **8**, 430–437 (2013).
- 921 16. Varghese, D. S., Alawathugoda, T. T. & Ansari, S. A. Fine Tuning of Hepatocyte Differentiation  
922 from Human Embryonic Stem Cells: Growth Factor vs. Small Molecule-Based Approaches. *Stem*  
923 *Cells Int.* **2019**, 1–18 (2019).
- 924 17. Basma, H. *et al.* Differentiation and Transplantation of Human Embryonic Stem Cell–Derived  
925 Hepatocytes. *Gastroenterology* **136**, 990-999.e4 (2009).
- 926 18. Hay, D. C. *et al.* Highly efficient differentiation of hESCs to functional hepatic endoderm requires  
927 ActivinA and Wnt3a signaling. *Proc. Natl. Acad. Sci.* **105**, 12301–12306 (2008).
- 928 19. Cuomo, A. S. E. *et al.* Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic  
929 genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
- 930 20. Burrows, C. K. *et al.* Genetic Variation, Not Cell Type of Origin, Underlies the Majority of  
931 Identifiable Regulatory Differences in iPSCs. *PLOS Genet.* **12**, e1005793 (2016).
- 932 21. Kajiwara, M. *et al.* Donor-dependent variations in hepatic differentiation from human-induced  
933 pluripotent stem cells. *PNAS* **109**, 12538–12543 (2012).

- 934 22. Kyttälä, A. *et al.* Genetic Variability Overrides the Impact of Parental Cell Type and Determines  
935 iPSC Differentiation Potential. *Stem Cell Rep.* **6**, 200–212 (2016).
- 936 23. Nishizawa, M. *et al.* Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is  
937 an Indicator of Differentiation Capacity. *Cell Stem Cell* **19**, 341–354 (2016).
- 938 24. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in  
939 human iPS cells. *Nat. Cell Biol.* **13**, 541–549 (2011).
- 940 25. Phetfong, J. *et al.* Cell type of origin influences iPSC generation and differentiation to cells of the  
941 hematoendothelial lineage. *Cell Tissue Res.* **365**, 101–112 (2016).
- 942 26. Shao, K. *et al.* Induced Pluripotent Mesenchymal Stromal Cell Clones Retain Donor-derived  
943 Differences in DNA Methylation Profiles. *Mol. Ther.* **21**, 240–250 (2013).
- 944 27. Sanchez-Freire, V. *et al.* Effect of Human Donor Cell Source on Differentiation and Function of  
945 Cardiac Induced Pluripotent Stem Cells. *J. Am. Coll. Cardiol.* **64**, 436–448 (2014).
- 946 28. Pearce, J. J. H. & Evans, M. J. Mml, a mouse Mix-like gene expressed in the primitive streak. *Mech.*  
947 *Dev.* **87**, 189–192 (1999).
- 948 29. Wolfe, A. D. & Downs, K. M. Mixl1 localizes to putative axial stem cell reservoirs and their  
949 posterior descendants in the mouse embryo. *Gene Expr. Patterns* **15**, 8–20 (2014).
- 950 30. Hart, A. H. *et al.* Mixl1 is required for axial mesendoderm morphogenesis and patterning in the  
951 murine embryo. *Development* **129**, 3597–3608 (2002).
- 952 31. Ng, E. S. *et al.* The primitive streak gene *Mixl1* is required for efficient haematopoiesis and BMP4-  
953 induced ventral mesoderm patterning in differentiating ES cells. *Development* **132**, 873–884  
954 (2005).
- 955 32. Lim, S. M. *et al.* Enforced Expression of *Mixl1* During Mouse ES Cell Differentiation Suppresses  
956 Hematopoietic Mesoderm and Promotes Endoderm Formation. *Stem Cells* **27**, 363–374 (2009).



- 957 33. Tewary, M. *et al.* *High-Throughput Micro-Patterning Platform Reveals Nodal-Dependent*  
958 *Dissection of Peri-Gastrulation-Associated versus Pre-Neurulation Associated Fate Patterning.*  
959 <http://biorxiv.org/lookup/doi/10.1101/465039> (2018) doi:10.1101/465039.
- 960 34. Becherel, O. J. *et al.* A new model to study neurodegeneration in ataxia oculomotor apraxia type  
961 *2. Hum. Mol. Genet.* **24**, 5759–5774 (2015).
- 962 35. Briggs, J. A. *et al.* Integration-Free Induced Pluripotent Stem Cells Model Genetic and Neural  
963 Developmental Features of Down Syndrome Etiology. *Stem Cells* **31**, 467–478 (2013).
- 964 36. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo.  
965 *Nature* **600**, 285–289 (2021).
- 966 37. Wang, X. *et al.* Fn1 Regulates the Third Pharyngeal Pouch Patterning and Morphogenesis. *J. Dent.*  
967 *Res.* **101**, 1082–1091 (2022).
- 968 38. Omachi, K. & Miner, J. H. Comparative analysis of dCas9-VP64 variants and multiplexed guide  
969 RNAs mediating CRISPR activation. *PLoS ONE* **17**, e0270008 (2022).
- 970 39. Jinek, M. *et al.* A programmable dual RNA-guided DNA endonuclease in adaptive bacterial  
971 immunity. *Science* **337**, 816–821 (2012).
- 972 40. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors.  
973 *Nat. Methods* **10**, 973–976 (2013).
- 974 41. Izzi, L. *et al.* Foxh1 recruits Gsc to negatively regulate Mixl1 expression during early mouse  
975 development. *EMBO J.* **26**, 3132–3143 (2007).
- 976 42. Ota, M. & Sasaki, H. Mammalian Tead proteins regulate cell proliferation and contact inhibition  
977 as transcriptional mediators of Hippo signaling. *Development* **135**, 4059–4069 (2008).

- 978 43. Morgani, S. M., Metzger, J. J., Nichols, J., Siggia, E. D. & Hadjantonakis, A.-K. Micropattern  
979 differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate  
980 patterning. *eLife* **7**, e32839 (2018).
- 981 44. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to recapitulate early  
982 embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847–854 (2014).
- 983 45. Plouhinec, J.-L., Simon, G., Vieira, M., Collignon, J. & Sorre, B. Dissecting signaling hierarchies in  
984 the patterning of the mouse primitive streak using micropatterned EpiLC colonies. *Stem Cell Rep.*  
985 S2213-6711(22)00262-4 (2022) doi:10.1016/j.stemcr.2022.05.009.
- 986 46. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the  
987 organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
- 988 47. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308  
989 (2013).
- 990 48. Bauer, D. E., Canver, M. C. & Orkin, S. H. Generation of Genomic Deletions in Mammalian Cell  
991 Lines via CRISPR/Cas9. *J. Vis. Exp.* 52118 (2014) doi:10.3791/52118.
- 992 49. Etard, C., Joshi, S., Stegmaier, J., Mikut, R. & Strähle, U. Tracking of Indels by DEcomposition is a  
993 Simple and Effective Method to Assess Efficiency of Guide RNAs in Zebrafish. *Zebrafish* **14**, 586–  
994 588 (2017).
- 995 50. Baxter, M. *et al.* Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells  
996 better mimic fetal rather than adult hepatocytes. *J. Hepatol.* **62**, 581–589 (2015).
- 997 51. Lee, L. H. *et al.* Micropatterning of human embryonic stem cells dissects the mesoderm and  
998 endoderm lineages. *Stem Cell Res.* **2**, 155–162 (2009).
- 999 52. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex Polyhedra for 3D Object  
1000 Detection and Segmentation in Microscopy. in *2020 IEEE Winter Conference on Applications of*

1001 *Computer Vision (WACV)* 3655–3662 (IEEE, Snowmass Village, CO, USA, 2020).  
1002 doi:10.1109/WACV45572.2020.9093435.

1003 53. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**,  
1004 676–682 (2012).

1005 54. Salehin, N., Tam, P. P. L. & Osteil, P. Prenet: Predictive network from ATAC-SEQ data. *J. Bioinform.*  
1006 *Comput. Biol.* **18**, 2040003 (2020).

1007 55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–  
1008 359 (2012).

1009 56. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic  
1010 Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

1011 57. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079  
1012 (2009).

1013 58. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS  
1014 alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).

1015 59. Zhang, Y. *et al.* Model-based Analysis of CHIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

1016 60. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning  
1017 sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

1018 61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-  
1019 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

1020 62. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.  
1021 *Nucleic Acids Res.* **44**, W160–W165 (2016).

1022 63. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-  
1023 performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

1024 64. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-  
1025 regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589  
1026 (2010).

1027 65. Zerbino, D. R., Johnson, N., Juettemann, T., Wilder, S. P. & Flicek, P. WiggleTools: parallel  
1028 processing of large collections of genome-wide datasets for visualization and statistical analysis.  
1029 *Bioinformatics* **30**, 1008–1009 (2014).

1030 66. Aryamanesh, N. A Reproducible and Dynamic Workflow for Analysis and Annotation of scRNA-  
1031 Seq Data. *Methods Mol. Biol.* **2490**, 101–140 (2022).

1032 67. Lucas, N. *et al.* Accelerated Barocycler Lysis and Extraction Sample Preparation for Clinical  
1033 Proteomics by Mass Spectrometry. *J. Proteome Res.* **18**, 399–405 (2019).

1034 68. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-  
1035 based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

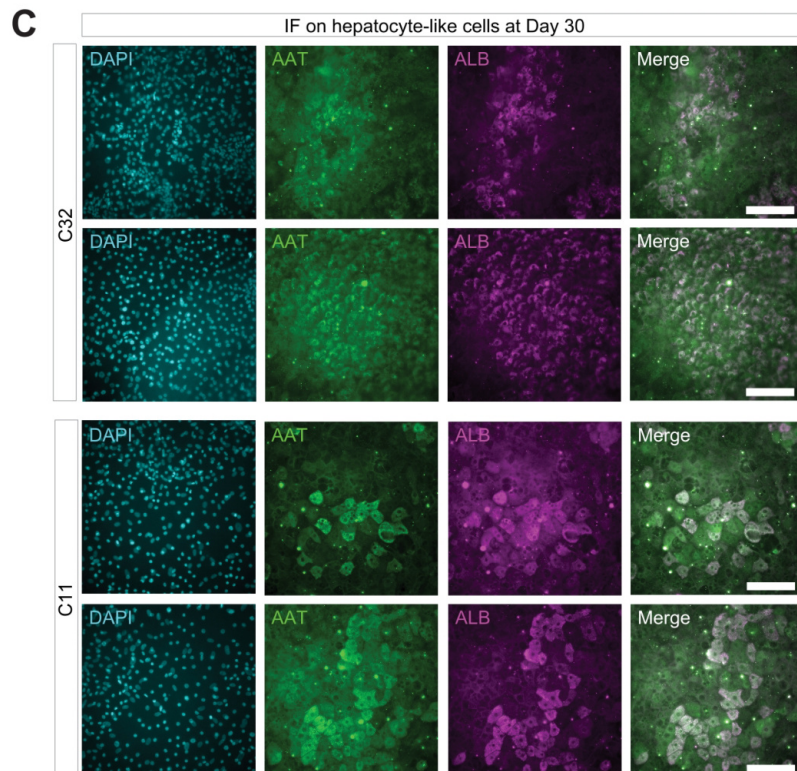
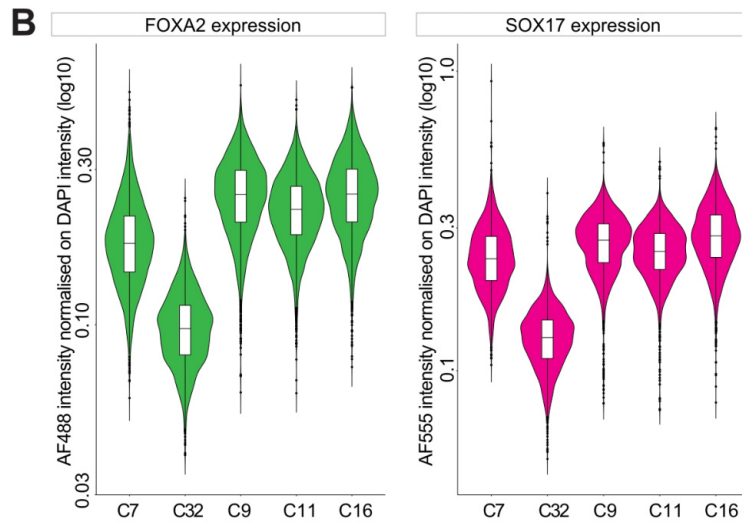
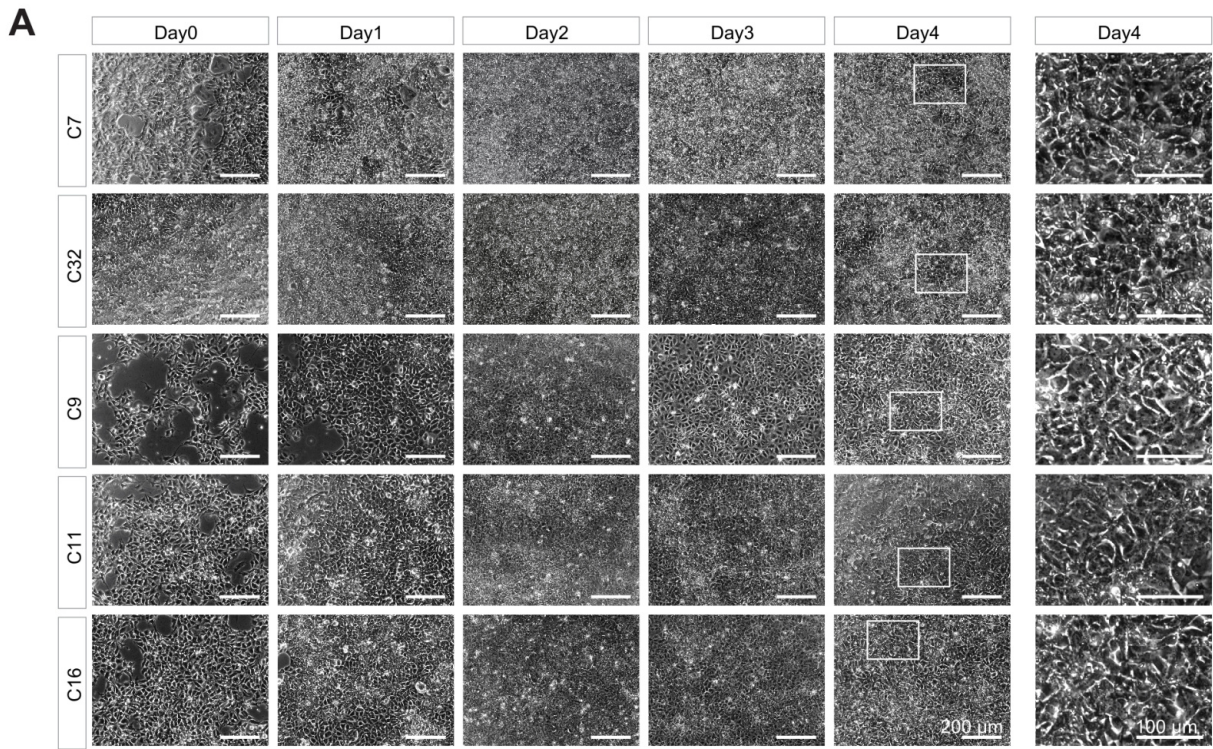
1036 69. Pang, Ignatius C, Waardenberg, Ashley J, Aryamanesh, Nader, & Graham, Mark E. ProteomeRiver  
1037 pipeline. (2023).

1038 70. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-  
1039 range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–  
1040 1372 (2008).

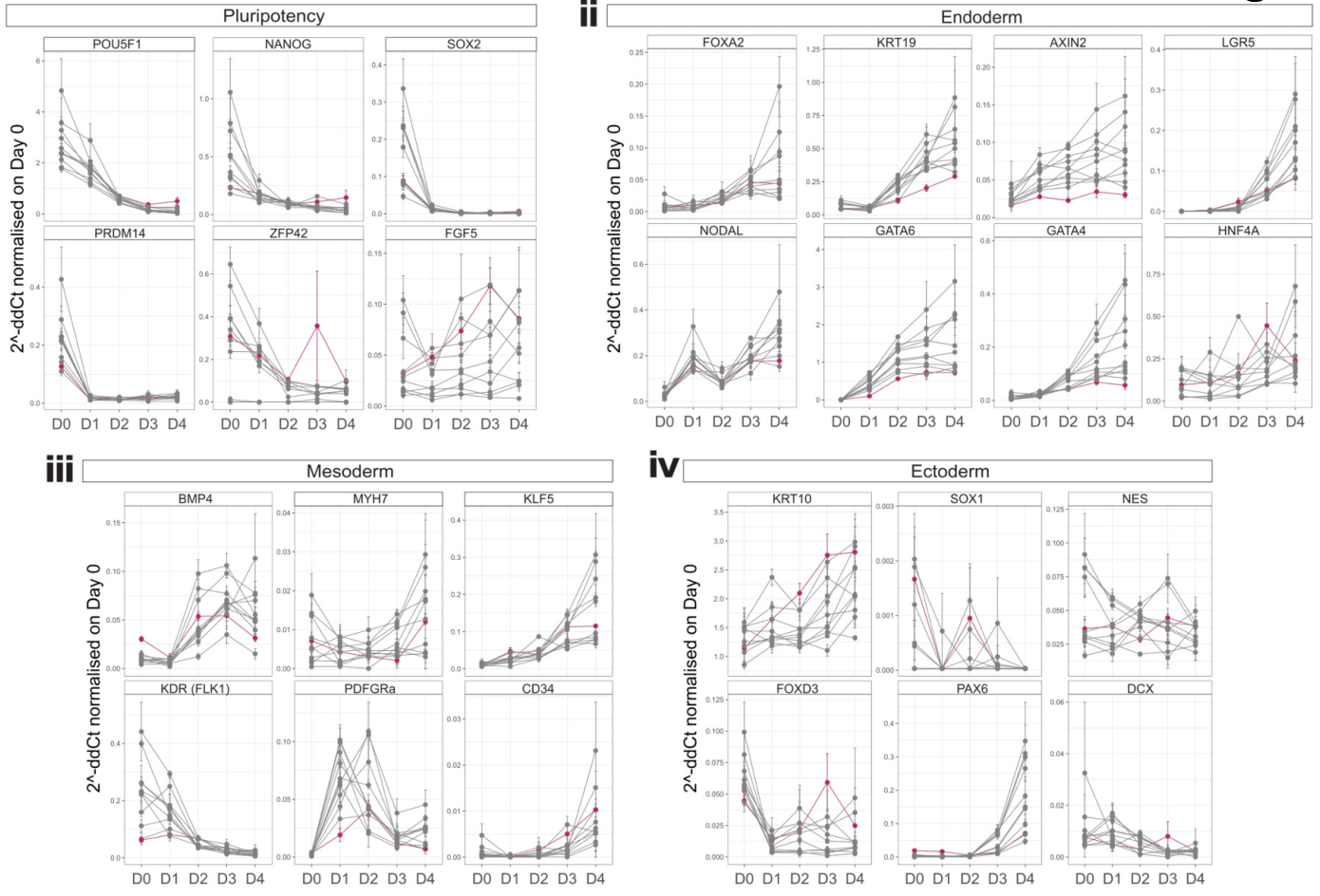
1041 71. Engholm-Keller, K. *et al.* The temporal profile of activity-dependent presynaptic phospho-  
1042 signalling reveals long-lasting patterns of poststimulus regulation. *PLoS Biol.* **17**, e3000170  
1043 (2019).

1044 72. Gagnon-Bartsch, J. A. & Speed, T. P. Using control genes to correct for unwanted variation in  
1045 microarray data. *Biostatistics* **13**, 539–552 (2012).

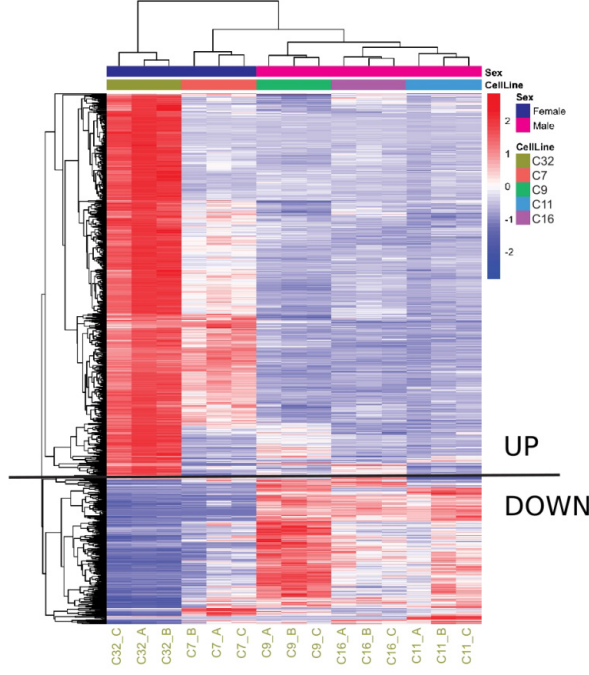
- 1046 73. Molania, R., Gagnon-Bartsch, J. A., Dobrovic, A. & Speed, T. P. A new normalization for Nanostring  
1047 nCounter gene expression data. *Nucleic Acids Res.* **47**, 6073–6083 (2019).
- 1048 74. Storey, J. D. A Direct Approach to False Discovery Rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**,  
1049 479–498 (2002).
- 1050 75. Chiu, D. S. & Talhouk, A. diceR: an R package for class discovery using an ensemble driven  
1051 approach. *BMC Bioinformatics* **19**, 11 (2018).
- 1052 76. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The*  
1053 *Innovation* **2**, 100141 (2021).
- 1054 77. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids*  
1055 *Res.* **51**, D523–D531 (2023).
- 1056 78. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for  
1057 taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
- 1058 79. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–  
1059 D692 (2022).
- 1060



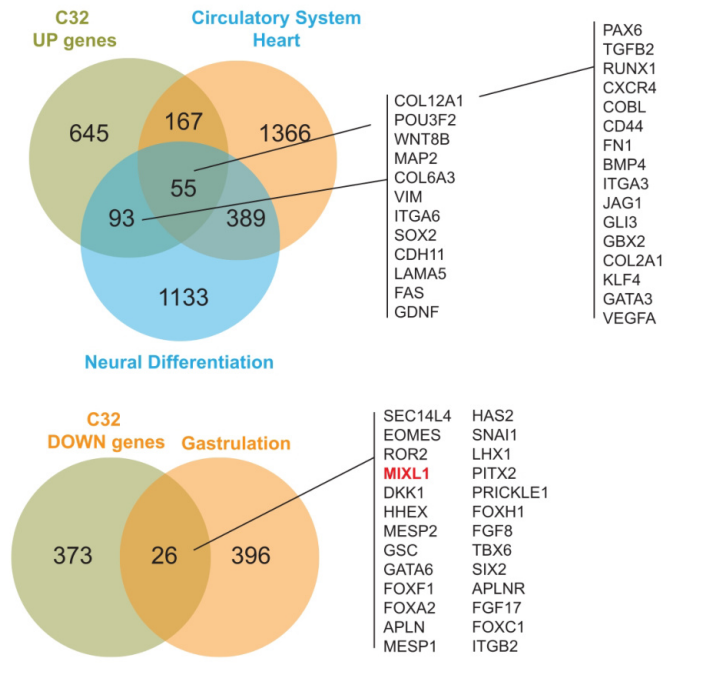
**A** Gene expression during differentiation per cell line - Highlight on C32



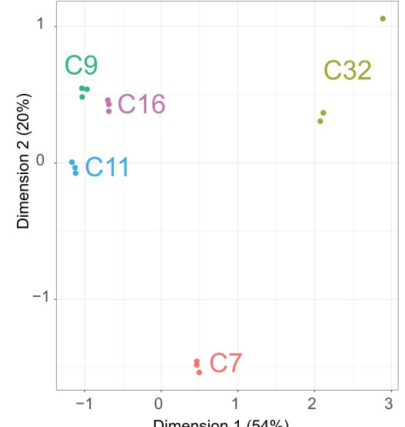
**B** Heatmap on DEGs in C32 versus the cohort



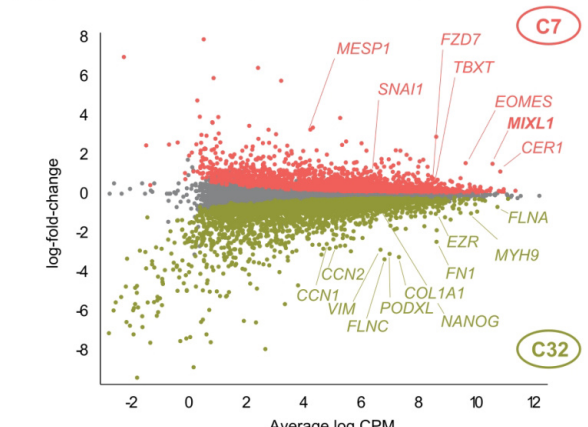
**C** Number of genes in Gene Ontologies

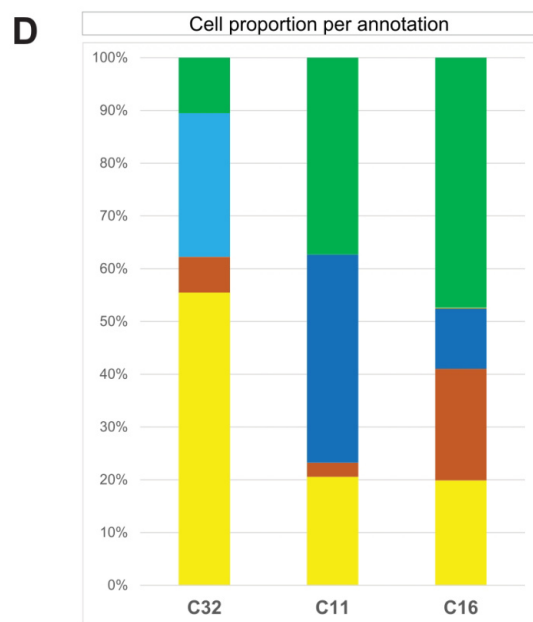
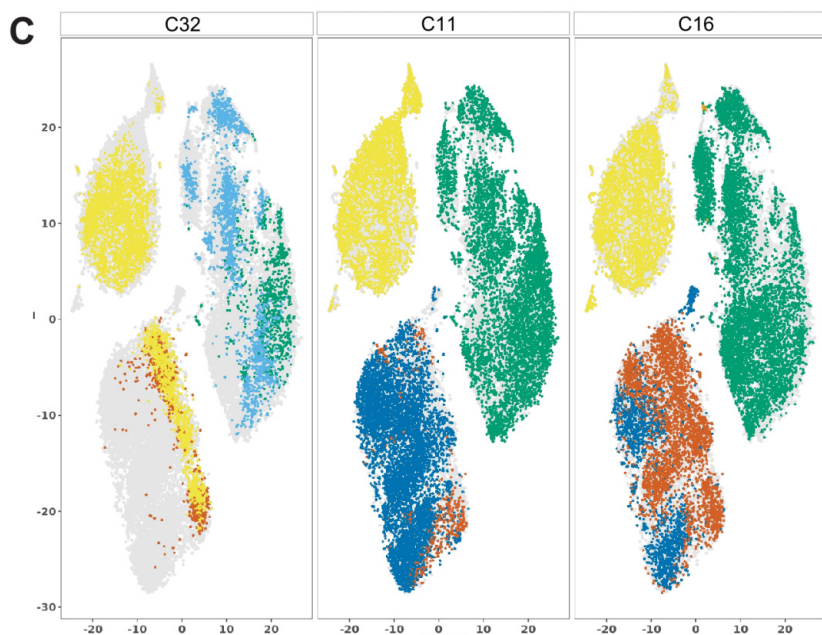
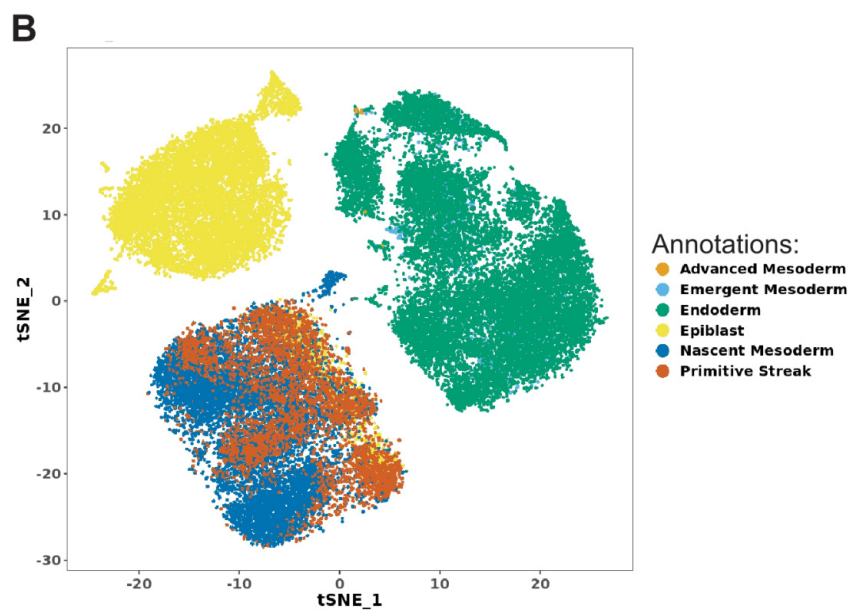
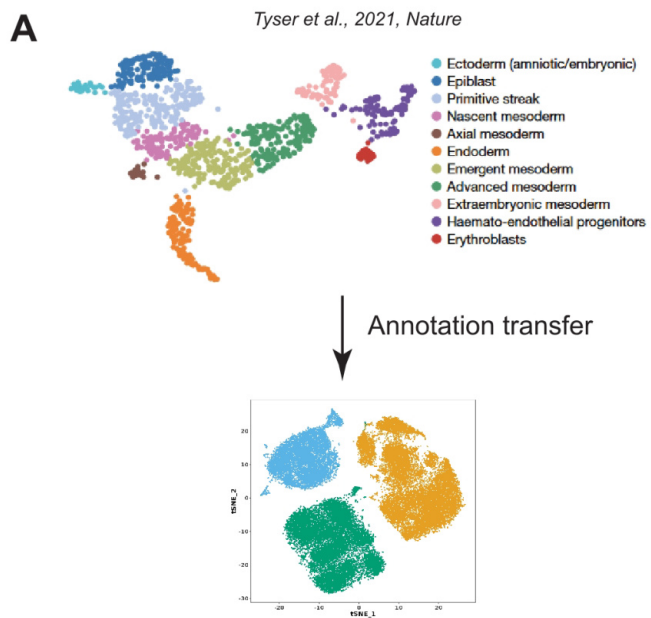


**D** PCA RNA-seq at Day1



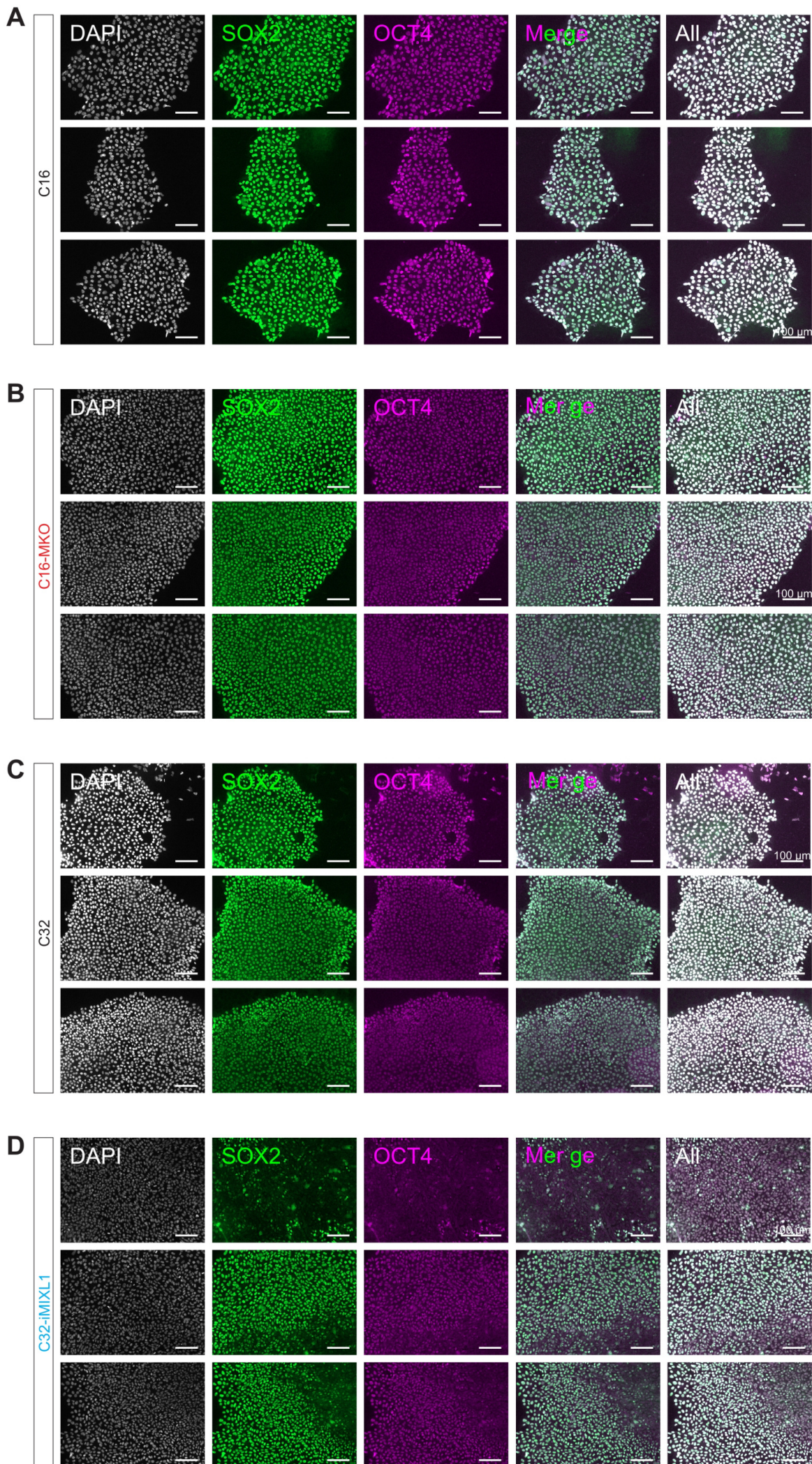
**E** MD-plot RNA-seq at Day1



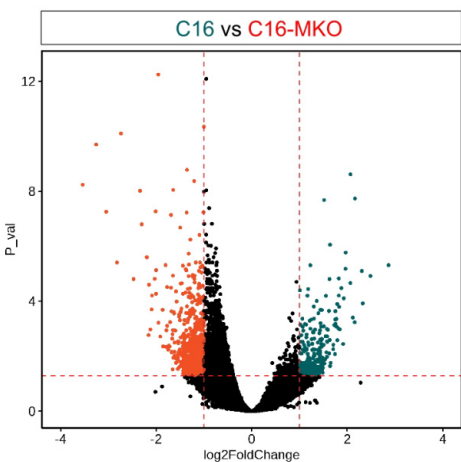




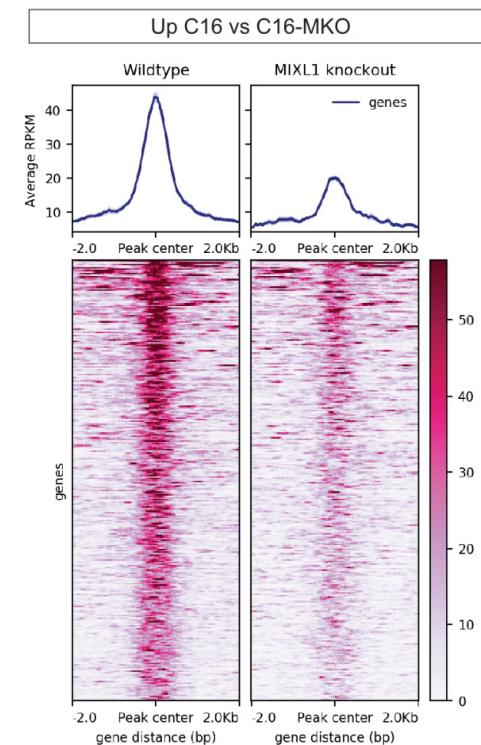




**A**



**B**



**D**

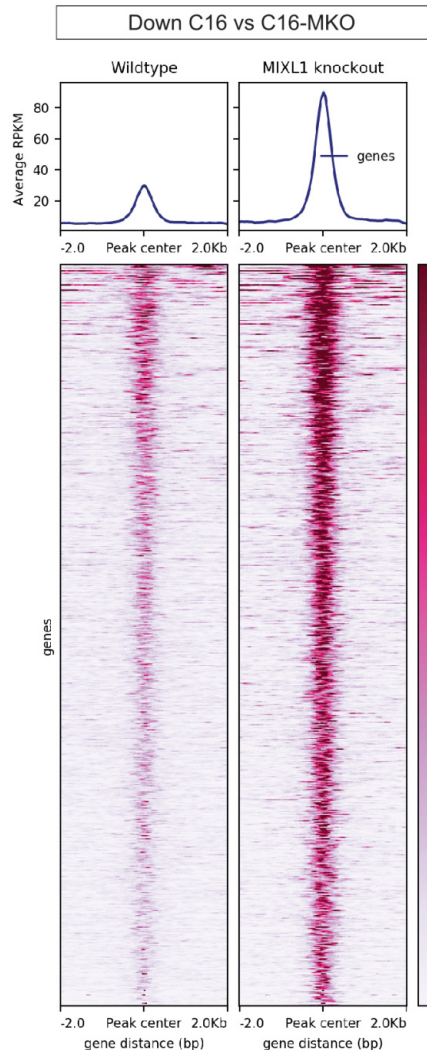
Known factor motifs

Motif	Factor	P-value
	PROP1	10 <sup>-13</sup>
	PHOX2A	10 <sup>-13</sup>
	PAX7	10 <sup>-8</sup>
	LHX9	10 <sup>-7</sup>
	LHX2	10 <sup>-7</sup>

De novo motifs

Motif	Best match	P-value
	PHOX2B	10 <sup>-21</sup>
	SOX15	10 <sup>-10</sup>
	E2A	10 <sup>-10</sup>
	SPDEF	10 <sup>-9</sup>
	RBPJ	10 <sup>-8</sup>

**C**



**E**

Known factor motifs

Motif	Factor	P-value
	TEAD1	10 <sup>-9</sup>
	FOXH1	10 <sup>-8</sup>
	JUN-AP1	10 <sup>-8</sup>
	BATF	10 <sup>-7</sup>
	CTCF	10 <sup>-7</sup>

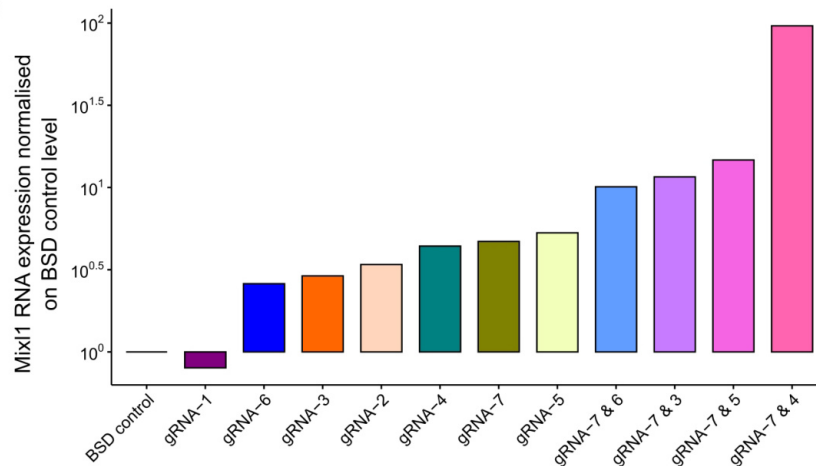
De novo motifs

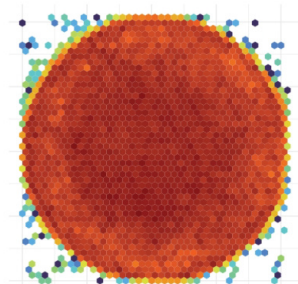
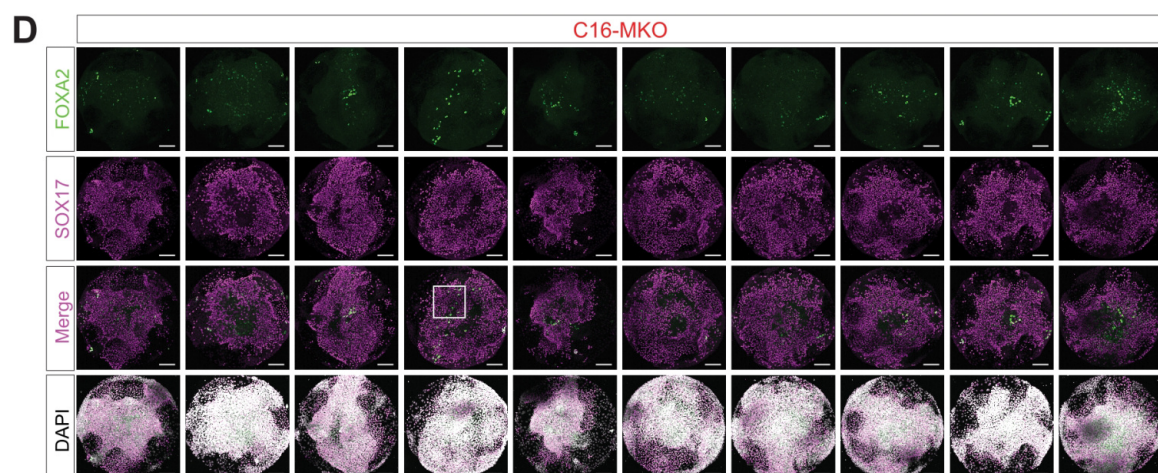
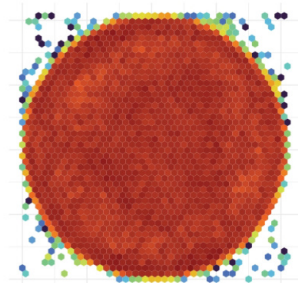
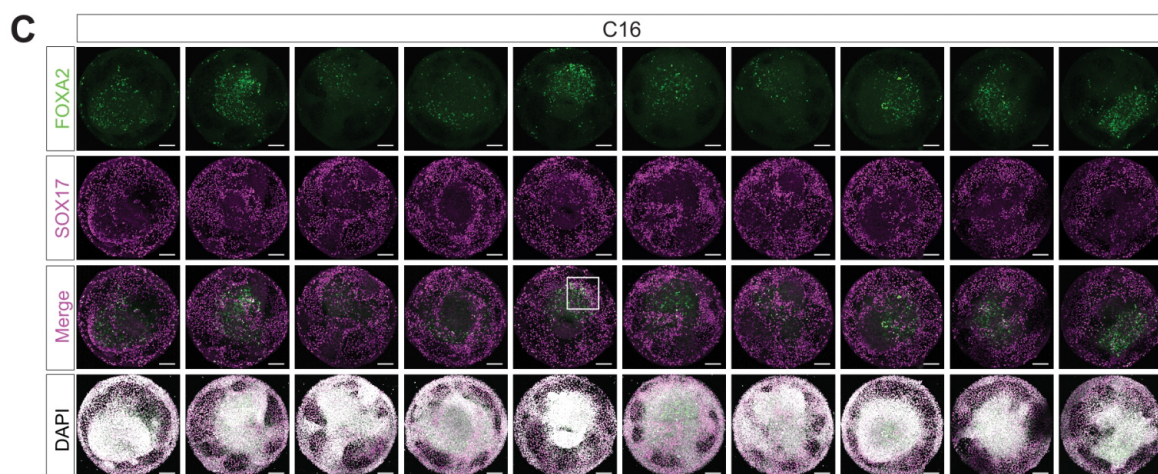
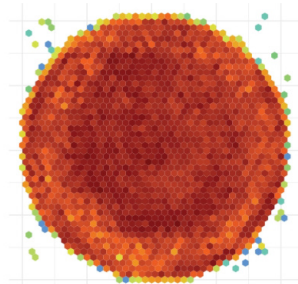
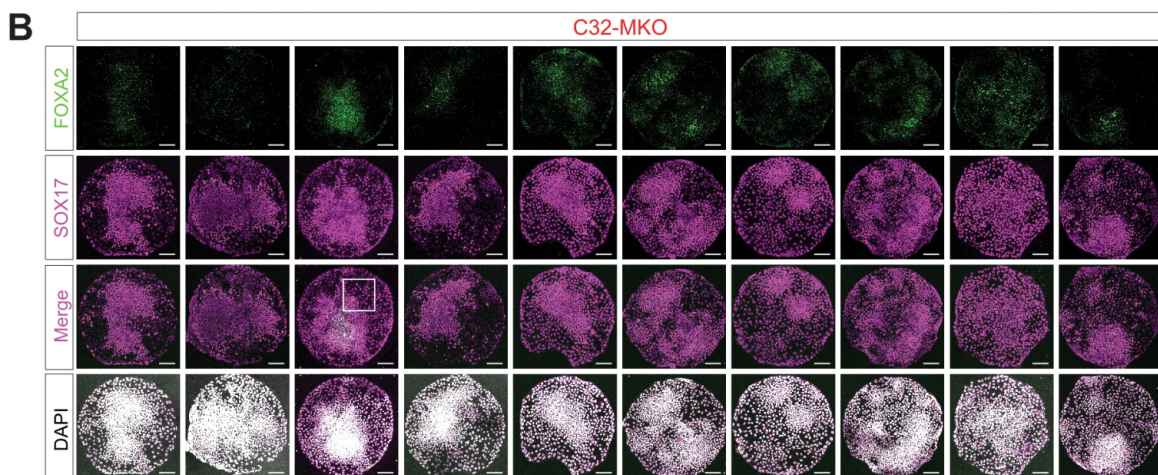
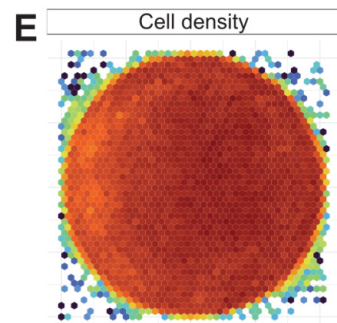
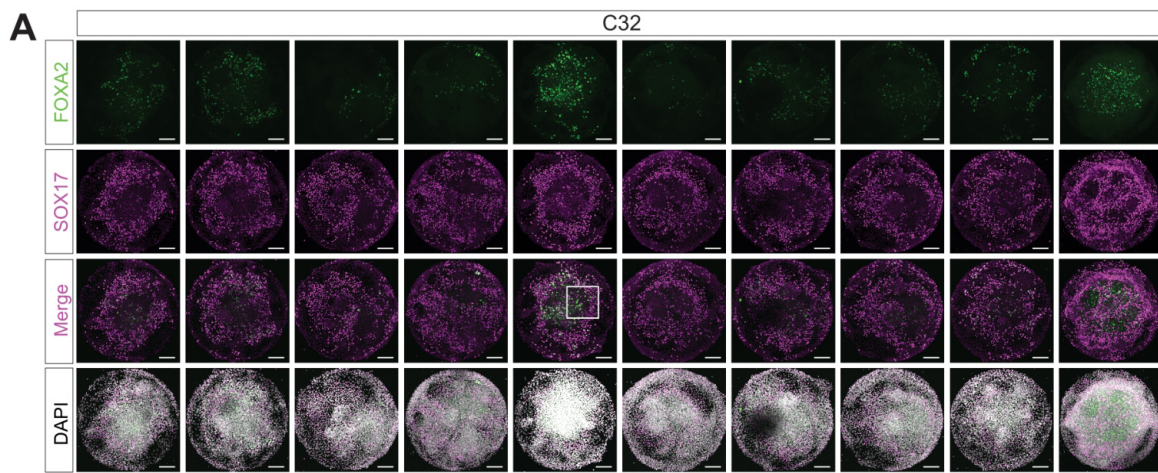
Motif	Best match	P-value
	FOXH1	10 <sup>-18</sup>
	CTCF	10 <sup>-15</sup>
	FOSL2	10 <sup>-13</sup>
	SPIB	10 <sup>-13</sup>
	TCF3	10 <sup>-13</sup>

**F**

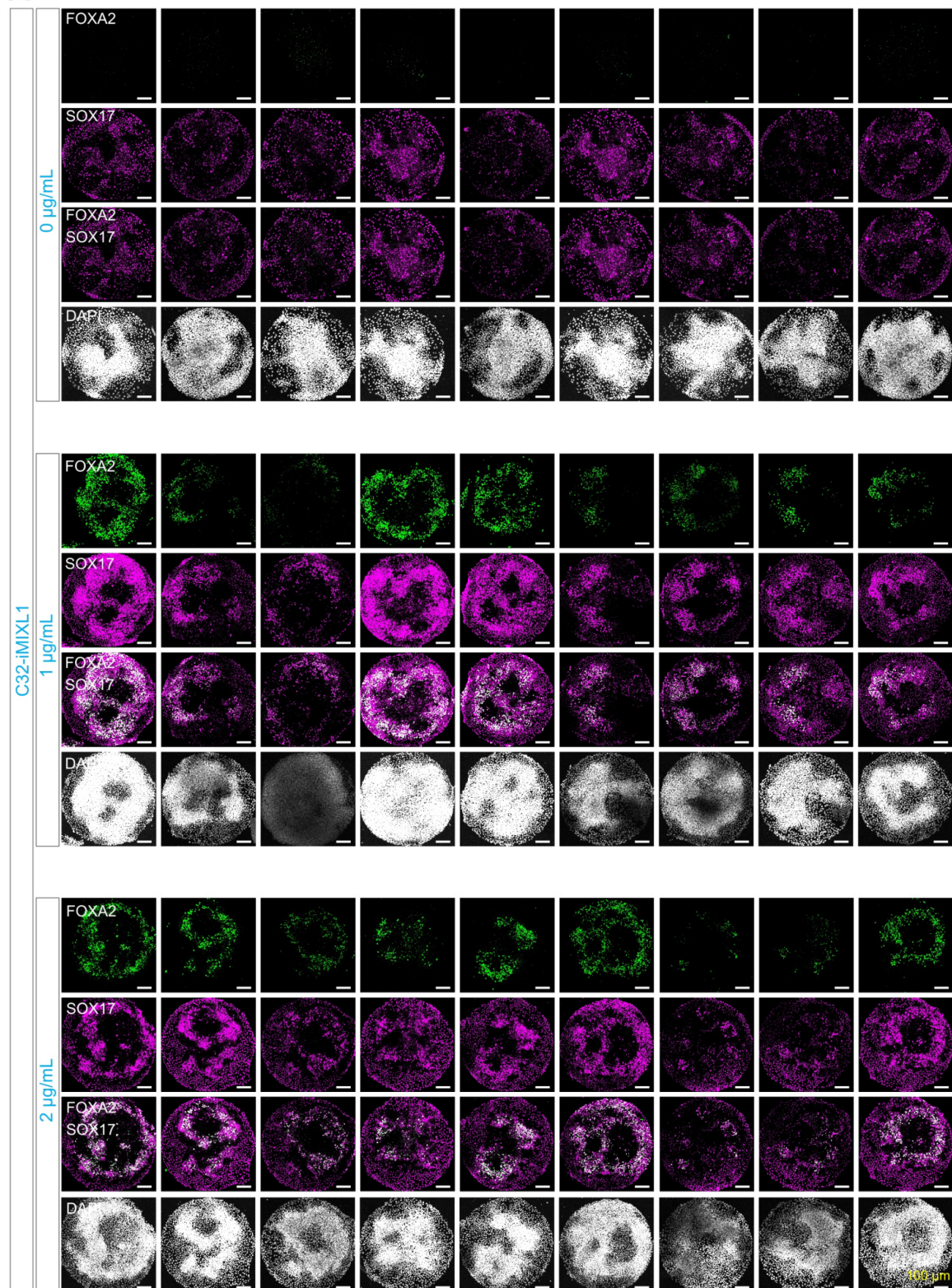
1 AGCTTTGATG AGGACAGACG GGACTAGGGC AAGGCTCAAG gRNA-1  
 41 GCCCAGGCCA TGTAAGGCAC AGGCCAGGCA AGGTGCTTCC gRNA-2  
 81 GCCGCTCTTA ACGGGGATTT GACCCGGAGA AGAGAGTTCT gRNA-3  
 121 GTCGCCCTGCG GCCCGCGCAC CCGGGCCCTT CCGGGCTGCC gRNA-4  
 161 CCTCCCAGGC AAATAGTCCT TCGGGGATGT GGATTGCGCC gRNA-5  
 181 GTCCGGGCGG GTGGGCGGGG AAGAACACGG AGGGGGCCGG gRNA-6  
 201 GACCCAAGTT CACGCTCCC TCCTGCATCC CGCAGCCGGC gRNA-7  
 241 GGAAGCGATT ATTCCCCGGC GTCTGGCGGG GCCGGGGGCG gRNA-7  
 281 GGCTCGCACC CGGAGGAGAC ACGGGTCTCC CCGAGCCAC gRNA-7  
 321 CTTTGATCGC ACCGCCCGC CTCCAGCC CTTGCCCGG gRNA-7  
 361 GAGGTATAA GTGCGGCCCG CGCC Mixl1 TSS

**G**

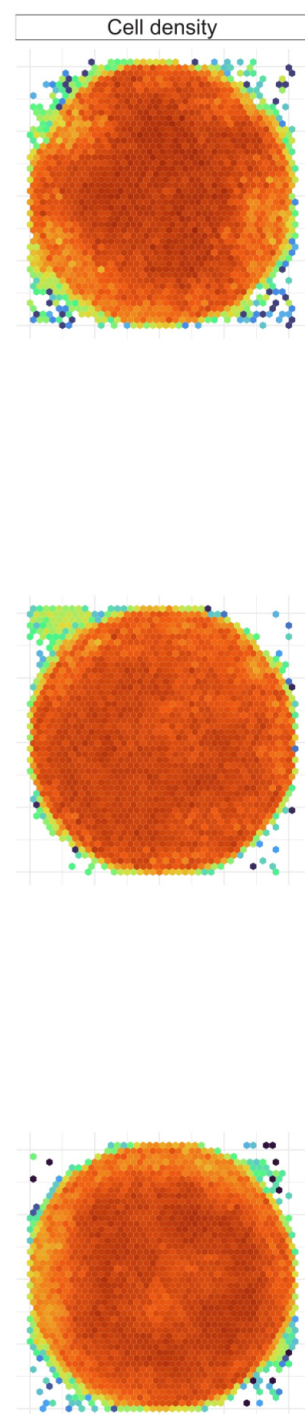




A



B



**Key resources table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Polyclonal anti-MixL1	Abcam	Cat# ab28411; RRID: AB_881536
Goat Polyclonal anti-T/Brachyury	Santa Cruz Biotechnology	Cat# sc-17745; RRID: AB_2200243
Rabbit Monoclonal anti-FoxA2	Abcam	Cat# ab108422; Clone: EPR4466; RRID: AB_11157157
Goat Polyclonal anti-Sox17	R&D Systems	Cat# AF1924; RRID: AB_355060
Rabbit Polyclonal anti-Sox9	Millipore	Cat# AB5535; RRID: AB_2239761
Rabbit Monoclonal anti-Ki67	Abcam	Cat# AB16667; Clone: SP6; RRID: AB_302459
Mouse Monoclonal anti-Chromogranin A	Novus Biologicals	Cat# NBP2-29428; Clone: CGA/414
Mouse Monoclonal anti-CDX2	BioGenex	Cat# MU392A-UC; Clone: CDX2-88; RRID: AB_2923402
Rabbit Polyclonal anti-Lysozyme	Agilent	Cat# A0099; RRID: AB_2341230
Mouse Monoclonal anti-Oct3/4	Santa Cruz Biotechnology	Cat# sc-5279; Clone: C10; RRID: AB_628051
Rabbit Monoclonal anti-Sox2	Cell Signaling Technologies	Cat# 3579; Clone: D6D9; RRID: AB_2195767
Donkey anti-goat IgG (H+L), Alexa Fluor 488 Conjugated	Invitrogen	Cat# A11055; RRID: AB_2534102
Donkey anti-rabbit IgG (H+L), Alexa Fluor 594 Conjugated	Invitrogen	Cat# R37119; RRID: AB_2556547
Donkey anti-mouse IgG (H+L), Alexa Fluor 488 Conjugated	Invitrogen	Cat# A21202; RRID: AB_141607
Donkey anti-rabbit IgG (H+L), Alexa Fluor 647 Conjugated	Invitrogen	Cat# A31573; RRID: AB_2536183
Chemicals, peptides, and recombinant proteins		
hESC-qualified Matrigel	Corning	Cat# 85850
mTeSR1	StemCell Technologies	Cat# 85850
CloneR	StemCell Technologies	Cat# 05888
Puromycin dihydrochloride	Thermo Fisher Scientific	Cat# A1113802
ReLeSR	StemCell Technologies	Cat# 05872
CryoStor CS10	StemCell Technologies	Cat# 07930
dNTP Mix (10mM each)	ThermoFisher Scientific	Cat# R0192
StemPro Accutase Cell Dissociation Reagent	Life Technologies	Cat# A1110501
Y-27632 dihydrochloride Rock Inhibitor	Tocris	Cat# 1245
Basement Membrane Matrigel	Corning	Cat# 354234
STEMdiff Intestinal Organoid Growth Medium	StemCell Technologies	Cat# 05145

GlutaMAX Supplement, 100x	Gibco	Cat# 35050061
Anti-Adherence Rinsing Solution	StemCell Technologies	Cat# 07010
(3-aminopropyl)triethoxysilane	Sigma-Aldrich	Cat# A3648
Gluteraldehyde solution, 25%	Sigma-Aldrich	Cat# G6257
Acrylamide solution, 40%	Sigma-Aldrich	Cat# A4058
N-N'-Methylenebisacrylamide solution, 2%	Sigma-Aldrich	Cat# M1533
Ammonium persulfate	Sigma-Aldrich	Cat# A3678
N,N,N',N'-Tetramethylethylenediamine	Sigma-Aldrich	Cat# T7024
Hydrazine hydrate, 100% (hydrazine, 64%)	Fisher Scientific	Cat# AC196711000
Acetic acid, glacial	Sigma-Aldrich	Cat# PHR1748
SYLGARD 184 L silicone Elastomer kit	Dow	Cat# 1317318
Vitronectin (VTN-N) Recombinant Human Protein, Truncated	Life Technologies	Cat# A14700
Sodium periodate	Sigma-Aldrich	Cat# 311448
Penicillin-Streptomycin	Sigma-Aldrich	Cat# P4333
Recombinant Human BMP-4 Protein	R&D Systems	Cat# 314-BP
Paraformaldehyde	Scharlau	Cat# PA00950500
Triton X-100	Sigma-Aldrich	Cat# X100
DPBS (1x)	Gibco	Cat# 14190-144
Bovine serum albumin	Sigma-Aldrich	Cat# A3311
DAPI	Thermo Fisher Scientific	Cat# 62247
Fluoromount-G	Thermo Fisher Scientific	Cat# 00-4958-02
CAS-Block Histochemical Reagent	Thermo Fisher Scientific	Cat# 008120
Ulex Europaeus Agglutinin I (UEA I), Rhodamine	Vector Laboratories	Cat# RL-1062-2
Glycerol	Chem-Supply	Cat# GA010
Tris Base	Sigma-Aldrich	Cat# T1503
EDTA	Sigma-Aldrich	Cat# 798681
Fructose	Sigma	Cat# F0127
Urea	Chem-Supply	Cat# UA001
Reverse Transcription Master Mix	Fluidigm	Cat# 100-6298
PreAmp Master Mix	Fluidigm	Cat# 100-5580
Exonuclease I	New England Biolabs	Cat# M0293S
SsoFast EvaGreen SuperMix with low ROX	BIO-RAD	Cat# 1725211
20x DNA Binding Dye	Fluidigm	Cat# 100-7609
2x Assay Loading Reagent	Fluidigm	Cat# 100-7611
Sodium chloride	Sigma-Aldrich	Cat# 71376
Magnesium chloride	Sigma-Aldrich	Cat# 208337
Igepal CA-630	Sigma-Aldrich	Cat# I8896
NEBNext High Fidelity 2x PCR Mix	New England Biolabs	Cat# M0541S
AmPure XP magnetic beads	Beckman Coulter	Cat#A63880
Fluoromount-G Mounting Medium	Thermo Fisher Scientific	Cat#00-4958-02
mTeSR Plus	StemCell Technologies	Cat#100-0276
<b>Critical commercial assays</b>		
P3 Primary Cell 4D-Nucleofector X Kit	Lonza	Cat# V4LP-3002
Q5 High Fidelity DNA Polymerase kit	New England Biolabs	Cat# M0491
STEMDiff Definitive Endoderm Kit	StemCell Technologies	Cat# 05110

STEMDiff Intestinal Organoid Kit	StemCell Technologies	Cat# 05140
P450-Glo CYP3A4 Assay with Luciferin-IPA	Promega	Cat# V9001
Isolate II RNA Mini kit	Bioline	Cat# BIO-52073
Tagment DNA TDE1 Enzyme and Buffer kit	Illumina	Cat# 20034197
MinElute PCR Purification Kit	Qiagen	Cat# 28004
Deposited data		
Micropattern data	This paper	<a href="https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024">https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024</a>
ATAC-seq data	This paper	GSE260552
RNA-seq data	This paper	GSE260553
scRNA-seq data	This paper	GSE260554
Proteomic data	This paper	PXD048788
Codes	This paper	<a href="https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024">https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024</a>
Experimental models: Cell lines		
Human: CRL1502 C7 hiPSC line	AIBN, UQ, Australia	N/A
Human: CRL1502 C32 hiPSC line	AIBN, UQ, Australia	N/A
Human: C32-MKO hiPSC line	VGEF, CMRI, Australia	N/A
Human: C32-Dox	AIBN, UQ, Australia	N/A
Human: FB C2 hiPSC line	AIBN, UQ, Australia	N/A
Human: FB C3 hiPSC line	AIBN, UQ, Australia	N/A
Human: FB C4 hiPSC line	AIBN, UQ, Australia	N/A
Human: CCL54 Eu79 hiPSC line	AIBN, UQ, Australia	N/A
Human: CCL54 Eu86 hiPSC line	AIBN, UQ, Australia	N/A
Human: CCL54 Eu87 hiPSC line	AIBN, UQ, Australia	N/A
Human: CRL2429 C9 hiPSC line	AIBN, UQ, Australia	N/A
Human: CRL2429 C11 hiPSC line	AIBN, UQ, Australia	N/A
Human: CRL2429 C11 MKO hiPSC line	VGEF, CMRI, Australia	N/A
Human: CRL2429 C16 hiPSC line	AIBN, UQ, Australia	N/A
Human: C16-MKO hiPSC line	VGEF, CMRI, Australia	N/A
Oligonucleotides		
<i>S. pyogens</i> Cas9 <i>MIXL1</i> gRNA sequence forward 5'-GCGCCGCGTTTTCCAGCGTACCGG-3'	This paper	N/A
<i>MIXL1</i> gRNA sequence Forward: 5' CACCGCGCCGCGTTTTCCAGCGTAC-3'	This paper	N/A
<i>MIXL1</i> gRNA sequence Reverse: 5' AAACGTACGCTGGAAACGCGGCGC-3'	This paper	N/A
<i>MIXL1</i> gRNA exon 1 indel sequence Forward: 5'GGAGGGTATAAGTGGCGCC-3'	This paper	N/A
<i>MIXL1</i> gRNA exon 1 indel sequence Reverse: 5'CCTCATCTGTGTCTTCTTCCCG-3'	This paper	N/A
Microfluidic qRT-PCR oligonucleotides	This paper	<a href="https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024">https://github.com/PierreOsteil/ScriptsForOsteilEtAl2024</a>
ATAC-seq oligonucleotides	T Buenrostro J.D, et. al. (2013)	NA



Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459) V2.0	Addgene	Cat# 62988
Software and algorithms		
Geneious	Kearse et al. (2012); DOI: <a href="https://doi.org/10.1093/bioinformatics/bts199">10.1093/bioinformatics/bts199</a>	<a href="http://www.geneious.com/">http://www.geneious.com/</a> ; RRID: SCR_010519
TIDE	Brinkman et al, (2014); DOI: <a href="https://doi.org/10.1093/nar/gku936">10.1093/nar/gku936</a>	<a href="https://tide.nki.nl/">https://tide.nki.nl/</a> ; RRID: SCR_023704
Fiji	Schindelin <i>et al.</i> , (2012); DOI: <a href="https://doi.org/10.1038/nmeth.2019">10.1038/nmeth.2019</a>	<a href="http://fiji.sc/">http://fiji.sc/</a> ; RRID: SCR_002285
Fluidigm Real Time PCR Analysis software	Fluidigm	<a href="https://www.fluidigm.com/software/">https://www.fluidigm.com/software/</a> ; RRID: SCR_015686
R Project for Statistical Computing	R Core Team (2021)	<a href="https://www.r-project.org/">https://www.r-project.org/</a> ; RRID: SCR_001905
PreNet Pipeline	Salehin et al. (2020); DOI: <a href="https://doi.org/10.1142/S02197202040003X">10.1142/S02197202040003X</a>	<a href="https://doi.org/10.1142/S02197202040003X">https://doi.org/10.1142/S02197202040003X</a>
Bowtie2	Langmead et al. (2012); DOI: <a href="https://doi.org/10.1038/nmeth.1923">10.1038/nmeth.1923</a>	<a href="https://bio.sourceforge.net/bowtie2/index.shtml">https://bio.sourceforge.net/bowtie2/index.shtml</a> ; RRID: SCR_016368
ENCODE ATAC-seq pipeline	Kim DS (2023); DOI: <a href="https://doi.org/10.1007/978-1-0716-2899-7_17">10.1007/978-1-0716-2899-7_17</a>	<a href="https://github.com/ENCODE-DCC/atac-seq-pipeline/tree/master/src">https://github.com/ENCODE-DCC/atac-seq-pipeline/tree/master/src</a> ; RRID: SCR_023100
SAMTOOLS	Heng et al. (2009); DOI: <a href="https://doi.org/10.1093/bioinformatics/btp352">10.1093/bioinformatics/btp352</a>	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a> ; RRID: SCR_002105
Sambamba	Tarasov et al. (2015); DOI: <a href="https://doi.org/10.1093/bioinformatics/btv098">10.1093/bioinformatics/btv098</a>	<a href="http://www.open-bio.org/wiki/Sambamba.Sambamba">http://www.open-bio.org/wiki/Sambamba.Sambamba</a>
MACS, Version 2	Zhang et. al (2008); DOI: <a href="https://doi.org/10.1186/gb-2008-9-9-r137">10.1186/gb-2008-9-9-r137</a>	<a href="https://github.com/macs3-project/MACS">https://github.com/macs3-project/MACS</a> ; RRID: SCR_013291
featureCounts	Liao et al. (2014); DOI: <a href="https://doi.org/10.1093/bioinformatics/btt656">10.1093/bioinformatics/btt656</a>	<a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a> ; RRID: SCR_012919
DESeq2	Love et al. (2014); DOI: <a href="https://doi.org/10.1186/s13059-014-0550-8">10.1186/s13059-014-0550-8</a>	<a href="http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a> ; RRID: SCR_015687
deeptools	Ramirez et al. (2016); DOI: <a href="https://doi.org/10.1093/nar/gkw257">10.1093/nar/gkw257</a>	<a href="https://github.com/deeptools/deepTools">https://github.com/deeptools/deepTools</a> ; RRID: SCR_016366
Integrative Genomics Viewer	Thorvaldsdóttir et al. (2013); DOI: <a href="https://doi.org/10.1093/bib/bbs017">10.1093/bib/bbs017</a>	<a href="http://www.broadinstitute.org/igv/">http://www.broadinstitute.org/igv/</a> ; RRID: SCR_011793

HOMER	Heinz et al. (2010); DOI: 10.1016/j.molcel.2010.05.004	<a href="https://github.com/sipcapture/homer">https://github.com/sipcapture/homer</a> ; RRID: SCR_010881
WiggleTools	Zerbino et al. (2014); DOI: 10.1093/bioinformatics/btt737	<a href="http://www.github.com/Ensembl/Wiggletools">www.github.com/Ensembl/Wiggletools</a> ; RRID: SCR_001170
ProteomeRiver pipeline	Pang I, Waardenberg A, Aryamanesh N, Graham M	<a href="https://bitbucket.org/cmri-bioinformatics/proteomeriver/src/main/">https://bitbucket.org/cmri-bioinformatics/proteomeriver/src/main/</a>
MaxQuant	Cox et al. (2008); DOI: 10.1038/nbt.1511	<a href="http://www.biochem.mpg.de/5111795/maxquant">http://www.biochem.mpg.de/5111795/maxquant</a> ; RRID: SCR_014485
RUV-2	Gagnon-Bartsch et al. (2012); DOI: 10.1093/biostatistics/kxr034	<a href="https://github.com/cran/ruv/blob/master/R/RUV2.R">https://github.com/cran/ruv/blob/master/R/RUV2.R</a>
RUV-III	Molania et al. (2019); DOI: 10.1093/nar/gkz433	<a href="https://github.com/cran/ruv/blob/master/R/RUVIII.R">https://github.com/cran/ruv/blob/master/R/RUVIII.R</a>
DiceR	Chiu et al. (2018); DOI: 10.1186/s12859-017-1996-y	<a href="https://CRAN.R-project.org/package=diceR">https://CRAN.R-project.org/package=diceR</a>
clusterProfiler	Wu et al. (2021); DOI: 10.1016/j.xinn.2021.100141	<a href="http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html">http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html</a> ; RRID: SCR_016884
UniProt	UniProt Consortium (2023); DOI: 10.1093/nar/gkac1052	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a> ; RRID: SCR_002380
KEGG	Kanehisa et al. (2023); DOI: 10.1093/nar/gkac963	<a href="https://www.kegg.jp">https://www.kegg.jp</a> ; RRID: SCR_012773
Reactome	Gillespie et al. (2022); DOI: 10.1093/nar/gkab1028	<a href="http://www.reactome.org">http://www.reactome.org</a> ; RRID: SCR_003485
Other		
Nunclon Delta surface 24-well plate	Thermo Fisher Scientific	Cat# NUN142475
Dynamag-2 magnet	Thermo Fisher Scientific	Cat# 12321D
HSD5000 ScreenTape	Agilent	Cat# 5067-5592
μ-slide 18-well - flat	Ibidi	Cat# 81826
40μm cell strainer	Corning	Cat# CLS431750
Single-Cell-G Chip	10x Genomics	Cat# 1000127