

## RESEARCH ARTICLE

# SRRM2 splicing factor modulates cell fate in early development

Silvia Carvalho<sup>1,2,3,4,5,‡</sup>, Luna Zea-Redondo<sup>1,6,\*</sup>, Tsz Ching Chloe Tang<sup>1,\*</sup>, Philipp Stachel-Braun<sup>6,7,8</sup>, Duncan Miller<sup>9,10</sup>, Paulo Caldas<sup>2,3</sup>, Alexander Kukalev<sup>1</sup>, Sebastian Diecke<sup>9,10</sup>, Stefanie Grosswendt<sup>7,8</sup>, Ana Rita Grosso<sup>2,3,‡</sup> and Ana Pombo<sup>1,6,‡</sup>

## ABSTRACT

Embryo development is an orchestrated process that relies on tight regulation of gene expression to guide cell differentiation and fate decisions. The *Srrm2* splicing factor has recently been implicated in developmental disorders and diseases, but its role in early mammalian development remains unexplored. Here, we show that *Srrm2* dosage is critical for maintaining embryonic stem cell pluripotency and cell identity. *Srrm2* heterozygosity promotes loss of stemness, characterised by the coexistence of cells expressing naive and formative pluripotency markers, together with extensive changes in gene expression, including genes regulated by serum-response transcription factor (SRF) and differentiation-related genes. Depletion of *Srrm2* by RNA interference in embryonic stem cells shows that the earliest effects of *Srrm2* heterozygosity are specific alternative splicing events on a small number of genes, followed by expression changes in metabolism and differentiation-related genes. Our findings unveil molecular and cellular roles of *Srrm2* in stemness and lineage commitment, shedding light on the roles of splicing regulators in early embryogenesis, developmental diseases and tumorigenesis.

**KEY WORDS:** *Srrm2*, SRM300, Splicing, Pluripotency, Stemness, Single-cell transcriptomics

## INTRODUCTION

Embryo development is a dynamic complex process that requires refined spatial and temporal gene expression regulation, coordinating cell differentiation and cell fate decisions. A major aspect of gene expression that contributes to gene function diversity is alternative splicing. Changes in the expression levels of specific splicing factors can induce cell differentiation of pluripotent cells and commitment to specific embryonic lineages. Specific splicing events of developmental transcription factors can promote and modulate differentiation (Tsai et al., 2014; Mayshar et al., 2008; Venables et al., 2013; Gabut et al., 2011; Han et al., 2013). Genome-wide programs of alternative splicing are also dynamic during cell differentiation or cell lineage reprogramming (Baralle and Giudice, 2017; Wu et al., 2010; Revil et al., 2010; Cloonan et al., 2008). Although alternative splicing is increasingly understood as a dynamic process that accompanies gene expression rewiring during development, the contributions of splicing regulators to pluripotency and lineage commitment remain ill defined.

Splicing is catalysed by small nuclear ribonucleoproteins (snRNPs) and non-snRNPs splicing factors. The pan-eukaryotic serine/arginine repetitive matrix (SRRM) family of SR-related proteins are splicing factors that lack RNA-binding domains and play important roles in both constitutive and alternative splicing regulation (Blencowe et al., 1998; Torres-Méndez et al., 2019). The SRRM family comprises four elements, SRRM1, SRRM2, SRRM3 and SRRM4. Both SRRM1 and SRRM2 promote early stages of splicing, by mediating interactions between SR proteins bound at exonic enhancers and at splice or branch sites marked by U1/U2 snRNPs, typically promoting exon inclusion (Blencowe et al., 1998, 1999; Eldridge et al., 1999). SRRM1 and SRRM2 proteins are broadly expressed among different tissues, while SRRM3 and SRRM4 are expressed in brain tissues. Misregulation of SRRM2 has been implicated in cancer and neurological disorders (Tomsic et al., 2015; Shehadeh et al., 2010; Cuinat et al., 2022). SRRM3/4 play crucial roles in microexon splicing, with important roles in brain function (Irimia et al., 2014; Quesnel-Vallières et al., 2016), and their misregulation can lead to neurological and neurodevelopmental problems (Nakano et al., 2019; Calarco et al., 2009; Quesnel-Vallières et al., 2016, 2015; Nakano et al., 2012).

SRRM2 has been recently linked to developmental disorders in a meta-analysis of 31,058 human exomes, showing the strongest enrichment for *de novo* mutations (Kaplanis et al., 2020). SRRM2 has a high probability of intolerance to loss-of-function variants in humans (Karczewski et al., 2020; Petrovski et al., 2013), suggesting that it plays a gene-dosage dependent critical role in key biological processes. For example, *Srrm2* expression decreases during the differentiation of human primary macrophages and its depletion upregulates differentiation marks in human myeloid-like cells (Liu et al., 2018; Xu et al., 2022). Data from the International Mouse Phenotyping Consortium also identify *Srrm2*-dosage effects in

<sup>1</sup>Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), Epigenetic Regulation and Chromatin Structure Group, 10115 Berlin, Germany. <sup>2</sup>Associate Laboratory i4HB – Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. <sup>3</sup>UCIBIO – Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. <sup>4</sup>Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, 4050-313 Porto, Portugal. <sup>5</sup>Graduate Program in Areas of Basic and Applied Biology (GABBA), ICBAS, University of Porto, 4050-313 Porto, Portugal. <sup>6</sup>Humboldt-Universität zu Berlin, Institute of Biology, 10115 Berlin, Germany. <sup>7</sup>Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin, Exploratory Diagnostic Sciences (EDS) 10178 Berlin, Germany. <sup>8</sup>Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute for Medical Systems Biology (BIMSB), From Cell State to Function Group, 10115 Berlin, Germany. <sup>9</sup>Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Pluripotent Stem Cells Platform, 13125 Berlin, Germany. <sup>10</sup>DZHK (German Centre for Cardiovascular Research), partner site Berlin, 10785 Berlin, Germany. <sup>‡</sup>Co-authors

<sup>‡</sup>Authors for correspondence (ana.pombo@mdc-berlin.de, ar.grosso@fct.unl.pt, silvia.carvalho@mdc-berlin.de)

© S.C., 0000-0001-6436-205X; L.Z.-R., 0000-0001-9064-9126; T.C.C.T., 0000-0002-4263-1447; P.S.-B., 0009-0006-4622-8838; D.M., 0000-0001-9852-677X; P.C., 0000-0001-6730-4461; A.K., 0000-0001-5847-3773; S.D., 0000-0002-5219-5992; S.G., 0000-0001-7722-2937; A.R.G., 0000-0001-6974-4209; A.P., 0000-0002-7493-6288

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

*Srrm2*-knockout mice, whereby homozygous loss-of-function leads to embryonic lethality at the pre-weaning stage, whereas *Srrm2*-heterozygous mice are viable ([www.mousephenotype.org](http://www.mousephenotype.org), Groza et al., 2022). *Srrm2* dosage effects in embryonic development have been reported for its *C. elegans* ortholog (Fontrodona et al., 2013), but not yet investigated in mammalian development.

In this study, we use constitutive and transient depletion of *Srrm2* in mouse embryonic stem cells (mESC) to investigate the effects of *Srrm2* dosage in early mammalian development. We found that constitutive *Srrm2* heterozygosity leads to a partial loss of stemness characterised by extensive changes in gene expression including the co-expression of naïve and formative pluripotency markers. Transient *Srrm2* depletion identifies alternative splicing of a small number of genes with potential roles in stemness and early differentiation.

## RESULTS

### *Srrm2* reduction impairs the maintenance of mESC pluripotency

To explore the expression of *Srrm1* and *Srrm2* genes in early development, we mined single-cell and bulk gene expression data from public resources (<https://apps.kaessmannlab.org/evodevoapp/>, <https://endoderm-explorer.com/>, Cardoso-Moreira et al., 2019; Nowotschin et al., 2019). We found that *Srrm2* is highly expressed in E3.5-5.5 blastocyst cells, with slightly increased expression from E3.5 inner cell mass (ICM) cells to E5.5 visceral endoderm (VE) cells and especially epiblast cells (Fig. 1A). *Srrm2* is also expressed in all tissue lineages, albeit with different expression levels and with a tendency to decrease, especially in non-neuronal tissues, during embryonic development (Fig. 1B). In contrast, *Srrm1* and *Sf3b1*, a small nuclear ribonucleoprotein with roles in splicing and a spliceosome component, respectively, are more stably expressed between different lineages and during development (Fig. S1A).

To investigate the effects of *Srrm2* heterozygosity in pluripotency and cell lineage commitment, we obtained a gene-trap mouse stem-cell line with a heterozygous truncation in exon 1 of the *Srrm2* gene (*Srrm2*<sup>+/-</sup>) from the Mutant Mouse Resource Center (MMRRC), originally generated by the Sanger Institute Gene Trap Resource (SIGTR) (Fig. 1C). *Srrm2* transcript levels measured by qRT-PCR showed a 60% reduction in *Srrm2*<sup>+/-</sup> cells compared to the parental wild-type ESC line (WT; Fig. 1D). Immunofluorescence experiments showed that SRRM2 protein levels are homogeneously reduced in the *Srrm2*<sup>+/-</sup> cell population, irrespective of whether cells are within or at the periphery of the stem cell colonies (Fig. 1E).

To test whether *Srrm2* depletion affects the ESC pluripotency phenotype, we started by assessing the morphology of cell colonies using brightfield imaging of WT and *Srrm2*<sup>+/-</sup> cells grown in a serum-free medium supplemented with leukemia inhibitory factor (LIF). We found that *Srrm2*<sup>+/-</sup> cells formed fewer well-defined colonies than WT cells, grew in dispersed clusters of cells, and showed an increased number of cells outside colonies (Fig. 1F). We also measured alkaline phosphatase (AP) activity and found that the majority of *Srrm2*<sup>+/-</sup> cells were positive for AP activity but showed lower signal intensity than WT cells (Fig. 1F, insets). The cell clusters and cells outside colonies were typically low or negative in AP activity in *Srrm2*<sup>+/-</sup> cells, suggesting a change in the regulation of the pluripotent state (Štefková et al., 2015).

To further investigate the state of pluripotency of *Srrm2*<sup>+/-</sup> cells, we performed immunofluorescence of classic pluripotency-associated factors POU5F1, SOX2 and NANOG, and found that they were similarly expressed in WT and *Srrm2*<sup>+/-</sup> cells within the

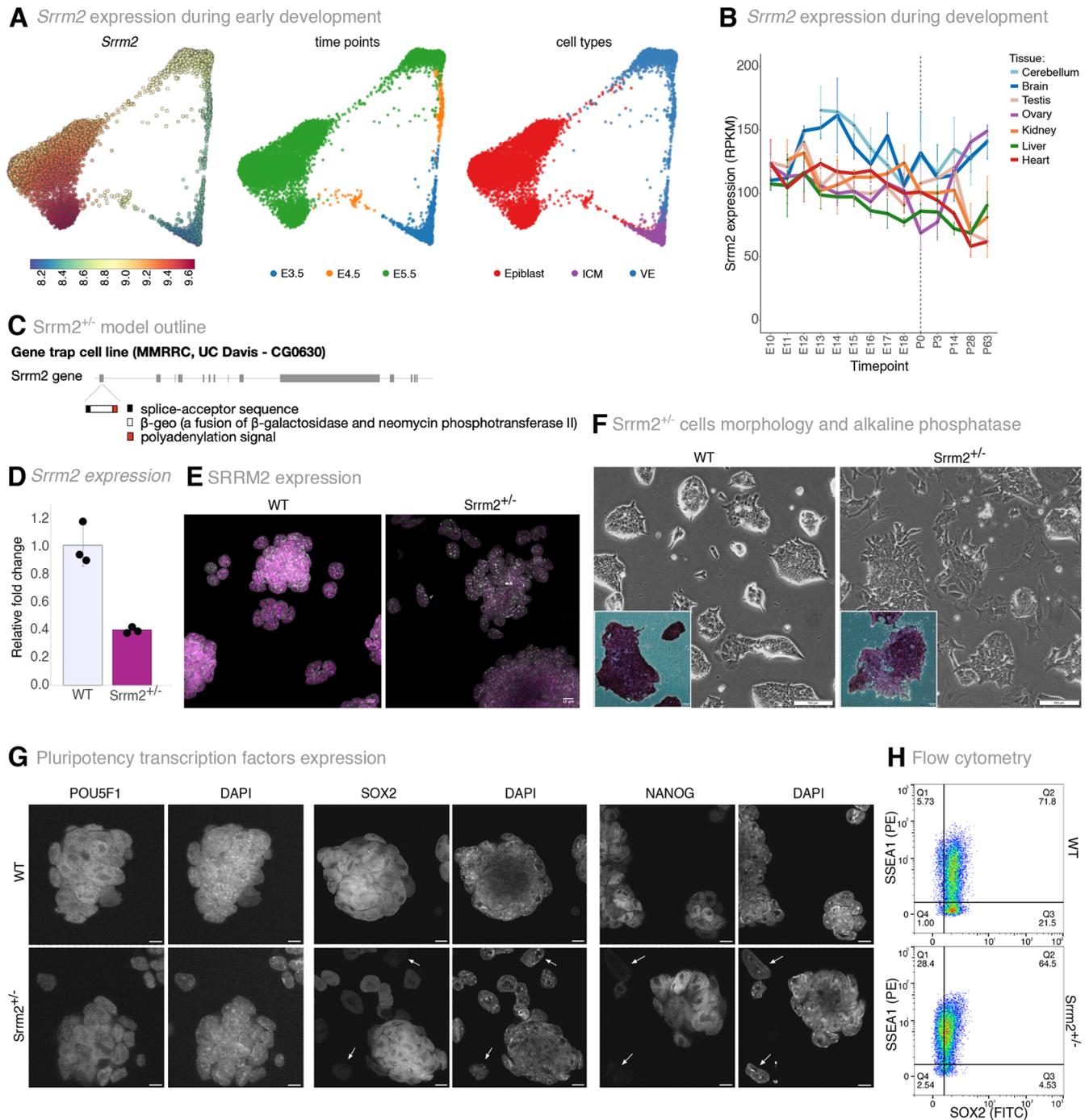
colonies or clusters (Fig. 1G). Consistent with a change in the pluripotent state, the isolated cells outside colonies or cell clusters expressed lower levels of the pluripotency-associated factors, and were observed more frequently in *Srrm2*<sup>+/-</sup> cell cultures (Fig. 1G, arrowheads point to cells with undetectable expression). Western blot analyses confirmed that NANOG is expressed in *Srrm2*<sup>+/-</sup> cells, although in lower bulk levels (Fig. S1B). We found that AP intensity and the expression of pluripotency markers remained constant in the three biological replicates of *Srrm2*<sup>+/-</sup> cell cultures with increased passage. The visual appearance of *Srrm2*<sup>+/-</sup> cell cultures did not noticeably change throughout the study, suggesting that the partial loss of colony integrity and altered heterogenous profile is stable.

To better understand the balance between cell states within the *Srrm2*<sup>+/-</sup> and WT cell cultures, we performed flow cytometry of SOX2 and of the pluripotency-associated cell surface marker SSEA1 (Cui et al., 2004). We found that *Srrm2*<sup>+/-</sup> cells generally express lower levels of SOX2 than WT cells, with an increase in the number of SSEA1-positive cells (Fig. 1H; Fig. S1C). These results show that the loss of colony formation in *Srrm2*<sup>+/-</sup> cells is not trivially related to loss of pluripotency but may represent an intermediate pluripotent stage. Moreover, the lower dosage of *Srrm2* is sufficient to induce a stable culture state characterised by fewer ESCs that express pluripotency markers, which nevertheless retain the ability to form colonies.

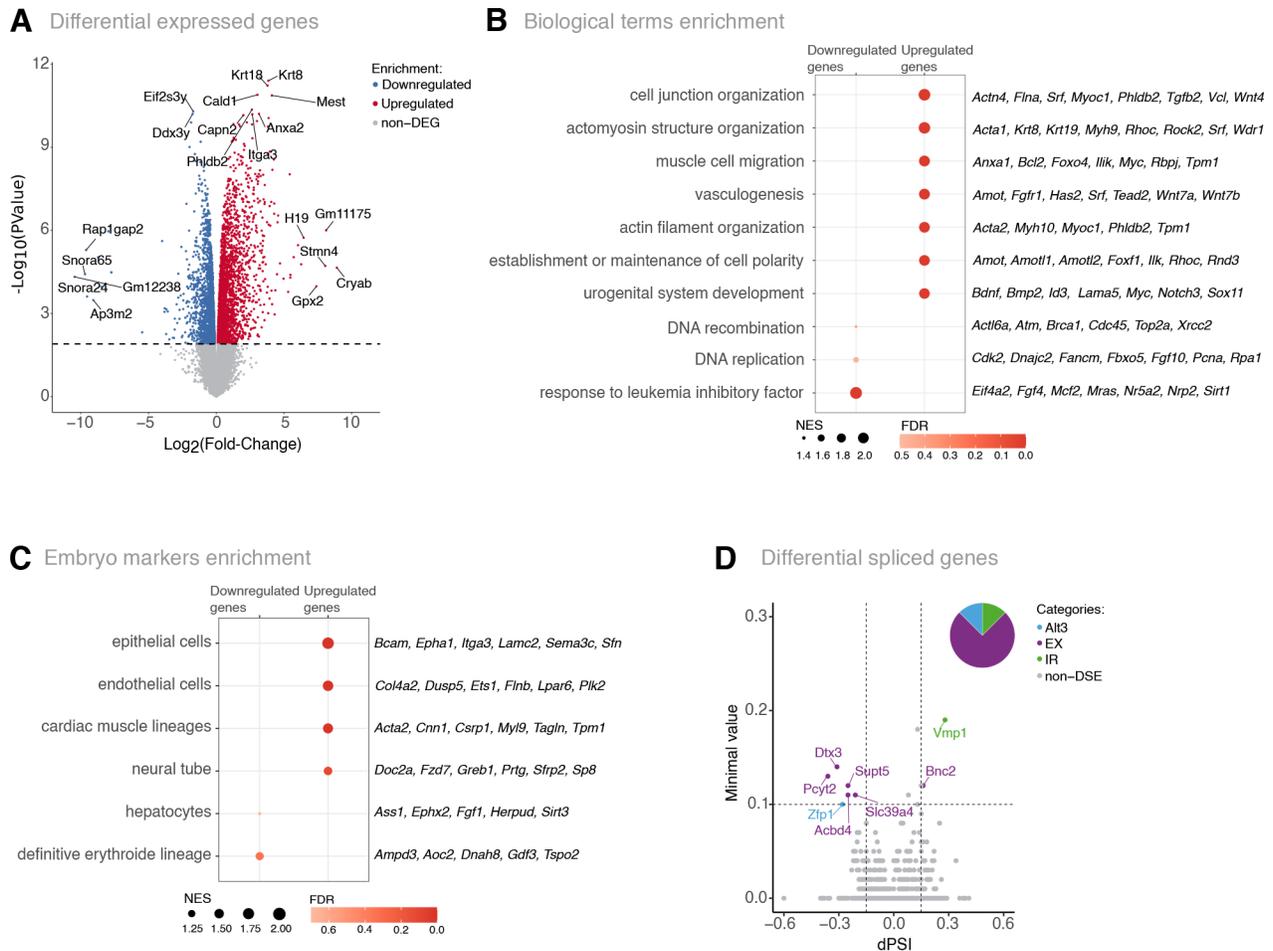
### *Srrm2* heterozygous knockout leads to extensive gene expression changes

To further characterise the *Srrm2*<sup>+/-</sup> cells, we performed total RNA profiling using high-throughput sequencing (RNA-seq) in three biological replicates. We found extensive changes in gene expression between *Srrm2*<sup>+/-</sup> and WT cells, with 2569 upregulated and 2205 downregulated genes (adj. *P*-value<0.05; Fig. 2A; Table S1). We confirmed that *Srrm2* was downregulated in *Srrm2*<sup>+/-</sup> cells (fold change of 1.61, adj. *P*-value 2.36e-05; Fig. S2A, Table S1). The expression of *Nanog* and *Klf4*, the first core pluripotency factors to be downregulated as cells exit pluripotency, were reduced by 1.7- and 1.8-fold, respectively, whereas *Sox2* was reduced by 1.2-fold, and *Pou5f1* was not differentially expressed (Fig. S2B, Table S1). These results are consistent with a partial loss of stemness of the *Srrm2*<sup>+/-</sup> cells, as hinted at by the initial imaging characterisation, with similarities to intermediate states of pluripotency described in mouse and human cells (Kalkan and Smith, 2014; Smith, 2017; Messmer et al., 2019). Amongst the top differential expressed genes, we found many associated with embryo development (e.g. *Mest*, *Itga3* and *H19*), including gametogenesis (e.g. *Ddx3y* and *Eif2s3y*). Most differentially expressed genes were protein coding, with a small tendency for upregulation (Fig. S2C). We also found downregulation of 103 out of 232 snoRNAs (Fig. S2D), including *Gm12238*, *Snora65* and *Snora24*, implicating *Srrm2* in the expression or processing of snoRNAs.

Gene set enrichment analysis (GSEA) focused on biological processes, showed that upregulated genes in *Srrm2*<sup>+/-</sup> cells are associated with cytoskeletal-related terms and differentiation processes, namely mesoderm development, as vasculogenesis (e.g. *Srf*, *Amot*, *Wnt7a*, *Wnt7b*) and urogenital development (e.g. *Sox11*, *Podxl*, *Ca2*, *Plaur*; Fig. 2B; Table S2). In contrast, downregulated genes were significantly associated with response to LIF (e.g. *Nr5a2*, *Nrp2*, *Mras*, *Mcf2*), which plays a central role in promoting a pluripotent ground state. Together with the imaging results, the bulk transcriptomics data show that *Srrm2*<sup>+/-</sup> cells have a



**Fig. 1.** *Srrm2* expression levels influence the pluripotent state of mESC. (A) Single-cell expression of *Srrm2* during E3.5-E5.5 mice embryo development (Nowotschin et al., 2019). Force-directed layouts coloured by *Srrm2* gene expression (left), development time (middle) or cell type labels (right; ICM, inner cell mass; VE, visceral endoderm). (B) *Srrm2* expression levels during pre-natal (E) and post-natal (P) developing mouse tissues (RPKM) (Cardoso-Moreira et al., 2019). The vertical dashed line represents the birth time point. (C) *Srrm2*<sup>+/-</sup> knockout gene-trap embryonic stem cell line description. Gene-trap cassette was inserted in Exon 11 of the *Srrm2* gene. (D) *Srrm2* expression levels. Total RNA levels of exon 11 of *Srrm2* were measured by RT-qPCR. Relative levels are normalised to  $\beta$ -actin. Mean and standard deviation are calculated from three biological replicates. (E) Immunofluorescence of SRRM2 in WT and *Srrm2*<sup>+/-</sup> cell cultures. Depicted by a representative experiment of at least three independent experiments from three biological replicates. Each image corresponds to the merge of the DAPI (grey) and SRRM2 (purple) channels. Scale bars: 10  $\mu$ m. (F) Brightfield images from WT and *Srrm2*<sup>+/-</sup> cell cultures and alkaline phosphatase staining of WT and *Srrm2*<sup>+/-</sup> cells. Brightfield images: representative experiment data of at least three independent experiments from three biological replicates. Scale bars: 100  $\mu$ m. Insets: AP staining of WT and *Srrm2*<sup>+/-</sup> cells. Representative experiment data from three biological replicates in at least three independent experiments. Scale bar: 10  $\mu$ m. (G) Immunofluorescence of POU5F1, SOX2 and NANOG pluripotency transcription factors (left to right) in WT and *Srrm2*<sup>+/-</sup> cell cultures (above to below). Each image corresponds to an individual channel per immunofluorescence (first: pluripotent transcription factor - POU5F1, SOX2 or NANOG; second: DAPI). Representative images from two independent immunofluorescence experiments for SOX2 and from one experiment for POU5F1 and NANOG, performed in one biological replicate. Scale bars: 10  $\mu$ m. Arrows indicate cells with very little or absent expression of the respective pluripotency factor. (H) Flow cytometry data of SSEA1 and SOX2 double labelling in WT and *Srrm2*<sup>+/-</sup> cells. Data shown are representative of one out of three biological replicates.



**Fig. 2. *Srrm2* heterozygous knockout mESC show extensive gene expression alterations.** (A) Volcano plot showing the *Srrm2*<sup>+/-</sup> differentially expressed genes in total RNA from three replicates of each condition. 18,912 genes were detected in the whole dataset and 4774 genes were considered as differentially expressed, using an adjusted *P*-value <0.05 as a cut-off (downregulated and upregulated genes are marked in blue and red, respectively). Gene labels represent the top-10 differentially expressed genes with the lowest adjusted *P*-value, and the top-5 differentially expressed genes with the higher absolute fold change. (B) GSEA for GO biological processes of *Srrm2*<sup>+/-</sup> relative to WT gene expression comparison. Red intensity and the size of the circles represent FDR values and normalised enrichment score (NES), respectively. (C) GSEA for publicly available E9.5-13.5 single-cell embryo tissue markers (Cao et al., 2019) for comparison of *Srrm2*<sup>+/-</sup> versus WT gene expression. Red intensity represents the FDR values, whereas the circle size indicates the NES, respectively. (D) Volcano plot showing the differentially spliced events between *Srrm2*<sup>+/-</sup> and WT cells, where the magnitude of alterations is represented by the difference in Percent Spliced In (dPSI) and the minimal value, computed by Vast-tools. 261,655 splicing events were detected in the whole dataset and eight splicing events were considered as significantly differentially spliced, using a cut-off of  $|\text{abs}(\text{dPSI})| \geq 0.15$  and a minimal value  $\geq 0.1$ . Spliced events with a dPSI <0 correspond to exclusion events in *Srrm2*<sup>+/-</sup> cells, whereas dPSI >0 correspond to inclusion events in *Srrm2*<sup>+/-</sup> cells. Categories of significant differential spliced events are represented by different colours: alternative 3' splice site (Alt3, blue), alternative 5' splice site (Alt5, not detected), exon exclusion (EX, purple), intron retention (IR, green), no alteration (non-DSE, grey). The pie chart represents the proportion of significant differential spliced events.

mixed expression signature with features of pluripotency and differentiation.

### ***Srrm2* heterozygous knockout leads to the acquisition of diverse differentiation signatures**

To further explore whether the differentiation signature of *Srrm2*<sup>+/-</sup> cells has a preferred developmental lineage, we performed a GSEA using publicly available single-cell transcriptome profiles from mouse embryos (Cao et al., 2019). We found that *Srrm2*<sup>+/-</sup> upregulated genes were significantly enriched for markers of epithelial, endothelial and cardiac muscle cells (FDR <0.05; Fig. 2C; Table S3), while downregulated genes included markers of hepatocyte and definitive erythroid lineages (FDR >0.05).

We noticed that many differentiation transcription factors were upregulated in *Srrm2*<sup>+/-</sup> cells, including *Srf*, *Mest* and *Sox11*

(Fig. 2A,B). Over-representation analysis (ORA) using transcription factor lists with the potential to drive transdifferentiation, as predicted by Mogrify (Rackham et al., 2016; Ferrai et al., 2017), showed that *Srrm2*<sup>+/-</sup> upregulated genes are preferential enriched for transcription factors associated with differentiation into cardiomyocytes, fibroblasts and mesenchymal cells (Fig. S2E, Table S4). Transcription factor network analysis showed that many upregulated genes are targets of the Serum Response Factor (SRF) (Fig. S2F,G, Table S5), which is one of the top 15 genes encoding for transcription factors with the lowest *P*-value (Fig. S2H). SRF is involved in cell differentiation, controls cell-to-cell adhesion, expression of cytoskeleton genes during development, and is required for epithelial and cardiac differentiation (Verdoni et al., 2010; Posern and Treisman, 2006; Miano et al., 2007; Deshpande et al., 2022). These results indicate that *Srrm2*<sup>+/-</sup> cells have a

transcriptional profile consistent with partial loss of pluripotency combined with the expression of differentiation markers that represent different cellular states or lineage specialisation.

### **Srrm2 heterozygous knockout shows specific splicing alterations**

Given the known roles of SRRM proteins in splicing regulation, we next asked whether the extensive gene deregulation of *Srrm2*<sup>+/-</sup> cells was associated with splicing defects. We analysed the total RNA-seq data using Vast-tools (Han et al., 2017; Irimia et al., 2014; Braunschweig et al., 2014; Tapial et al., 2017) and found 343,742 differentially splicing events, of which 680 were significant. To identify the most robust alternatively spliced transcripts, we applied stringent thresholds based on read coverage quality in at least two replicates, significance [ $\text{abs(dPSI)} \geq 0.15$  and a minimal value  $\geq 0.1$ ], and PSI distribution of reads over the spliced junctions. We detected eight highly robust splicing events between *Srrm2*<sup>+/-</sup> and WT ESCs, mostly exon exclusion (Fig. 2D; Table S6), as expected in association with *Srrm2* depletion. Six out of the eight genes were also differentially expressed (*Bnc2*, *Dtx3*, *Slc39a4* were downregulated, while *Pcyt2*, *Vmp1* and *Zfp1* were upregulated), and five play important roles in embryo development. VMP1 is important for autophagy and cell adhesion and is involved in lipid metabolism and in embryo implantation (Ropolo et al., 2007; Guo et al., 2012; Morishita et al., 2019; Holzner et al., 2023 preprint). DTX3 is a ubiquitin ligase that regulates NOTCH signaling, with key roles in development (Ding et al., 2020; Wang et al., 2021b). BNC2 and PCYT2 are important for the differentiation and maturation of mesenchyme and muscle, respectively, both with origin in the mesodermal germ layer (Vanhoutteghem et al., 2009; Zhu et al., 2009). *Pcyt2* and *Slc39a4* are essential genes for mouse development, as complete knockout mice die during embryogenesis (Fullerton et al., 2007; Pastor-Arroyo et al., 2021). Our results reveal that the transcriptional profile of *Srrm2*<sup>+/-</sup> cells is consistent with an altered pluripotent state that affects several development-related genes and is compatible with a mixture of different cellular states within the *Srrm2*<sup>+/-</sup> cell cultures.

### **Srrm2<sup>+/-</sup> cell populations display heterogeneity of pluripotency states**

To further explore the transcriptional and pluripotent state of *Srrm2*<sup>+/-</sup> cells, we generated single-cell RNA-seq (scRNA-seq) data using 10x Genomics technology from two biological replicates of WT and *Srrm2*<sup>+/-</sup> cell cultures. Following quality control processing, we obtained a total of 31,433 cells and a median of 4097 genes detected per cell. The WT and *Srrm2*<sup>+/-</sup> single-cell transcriptome data was integrated using the Harmony algorithm (Korsunsky et al., 2019), revealing four distinct cell clusters, represented through dimensionality reduction via uniform manifold approximation and projection (UMAP; Fig. 3A). Most WT cells (86%) are grouped in cluster 1, whereas 94% of *Srrm2*<sup>+/-</sup> cells populate clusters 2 to 4 (Fig. 3B; Fig. S3A), suggesting that *Srrm2* heterozygosity gives rise to a distinct cell state. In line with the immunofluorescence analyses of SRRM2 protein expression, we found that *Srrm2* was expressed at lower levels in *Srrm2*<sup>+/-</sup> cells independently of the state (Fig. S3B). Cycling cells were also found in all clusters, with clusters 1 and 2 having similar proportions of G1/G0, S and G2 cells, while cluster 4 shows the largest proportion of G1/G0 cells (Fig. S3C). These analyses confirm that *Srrm2*<sup>+/-</sup> cells have a distinct and heterogeneous transcriptional profile which reflects the morphological variability observed in the cell cultures.

We next assessed the expression of pluripotency transcription factors in the different cell states and confirmed that *Pou5f1*, *Sox2*, *Nanog* and *Klf4* are less expressed in *Srrm2*<sup>+/-</sup> than in WT cells (Fig. 3C). *Pou5f1* is expressed in all clusters, although in lower levels in *Srrm2*<sup>+/-</sup> cells from cluster 4 (Fig. S3D). *Sox2* is also expressed in all clusters, but at lower levels in clusters 3 and 4, while *Nanog* and *Klf4* expression is not detected in most cells in clusters 3 and 4 (Fig. S3D). These data show that clusters 1 and 2 resemble the most stem-like states in WT and *Srrm2*<sup>+/-</sup> cultures, whereas clusters 3 and 4, which are more abundant in *Srrm2*<sup>+/-</sup> cell cultures, are characterised by a lower expression of pluripotency transcription factors.

Given the heterogeneous expression profiles of the core pluripotent markers, we asked whether *Srrm2*<sup>+/-</sup> cells depict previously characterised pluripotency states, namely naive (capable of giving rise to all embryo lineages), formative (intermediate state initiating specialisation to a particular lineage), or primed (induction of specialised markers) (Kalkan and Smith, 2014; Smith, 2017; Messmer et al., 2019). We measured the expression of markers of the different pluripotency states (Wang et al., 2021a,b) in the four different clusters and found that naive state markers are expressed in clusters 1 and 2, although at higher levels in cluster 1 (Fig. 3D). Formative markers are most highly expressed in clusters 3 and 4, whereas primed state markers are mostly not detected, except for *Wnt9a* and *Runx1* in cluster 4. Together, these data reveal that *Srrm2*<sup>+/-</sup> cell cultures contain cells in different states of pluripotency, including an intermediate pluripotent state that displays shared features between the naive and formative pluripotency, and other states that are characterised by the absence of pluripotency signatures.

The exit of pluripotent stem cells from the pluripotent naive state is accompanied by alterations in gene expression networks, increased DNA methylation, and metabolism changes, resulting in cell fate determination (Hoogland and Marks, 2021; Kim and Costello, 2017). To further investigate the intermediate pluripotent state of *Srrm2*<sup>+/-</sup> cells, we quantified the transcriptional differences between clusters 1 and 2, and detected 35 downregulated genes and 45 upregulated genes (Fig. 3E; Table S7), from which 50 genes were also differentially expressed in the bulk data, but not differentially spliced. Amongst the downregulated genes, we found *Srrm2*, *Nanog* and *Klf4*, but also the mitochondrial protein *Cox6b2* with previously described roles in the stemness of trophoblast cells (Saha et al., 2022). Within the upregulated genes, we found the transcription factor *Tcf7l1*, which mediates the transition from naive to primed states of pluripotency by repressing the expression of naive and formative genes, and drives cells towards primitive endoderm through the Wnt/ $\beta$ -catenin signalling pathway (Guo et al., 2011; Hoffman et al., 2013; Pereira et al., 2006; Athanasouli et al., 2023). Other upregulated genes included *de novo* DNA methyltransferases, *Dnmt3a* and *Dnmt3b*, which have roles in maintaining the naive pluripotency state (Li et al., 2017; Kinoshita et al., 2021; Betto et al., 2021; Leitch et al., 2013), and ribosomal protein genes, in line with previous reports of global translation increase in early differentiation and cell fate transitions (Li and Wang, 2020).

Finally, to further characterise the different clusters, we identified cluster-specific gene markers (Table S8) and performed Gene Ontology (GO) enrichment analyses (Table S9). The group of gene markers of the WT-linked cluster 1 was enriched for genes associated with response to LIF (e.g. *Klf4*, *Zfp42*) and stem cell pluripotency (e.g. *Nanog*, *Sox2*, *Esrrb1*, *Rest*, *Lefty2*) (Fig. 3F). The most stem-like group of *Srrm2*<sup>+/-</sup> cells (cluster 2) has 23 gene

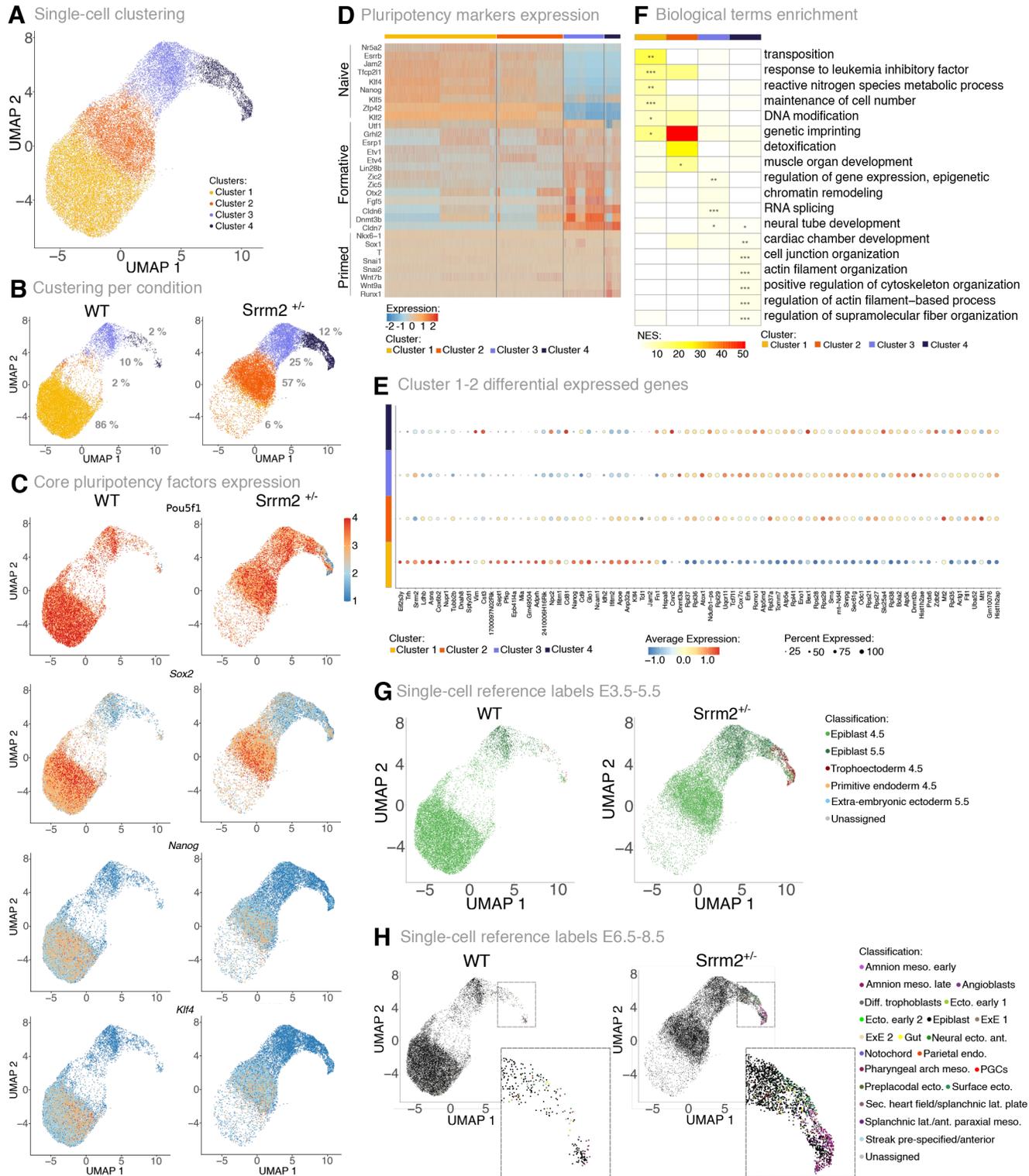


Fig. 3. See next page for legend.

markers, which include *Dppa3*, *Zfp42*, but not *Klf4*, *Nanog* or *Sox2*, as well as ribosomal subunits and genes associated with muscle organ development (e.g. *Rbpj*, *Tcf15*, *Mymx*). Finally, cluster 3 gene markers were enriched for terms such as epigenetic remodelers, neural tube and cardiac chamber development (e.g. *Dnmt3a*, *Dnmt3b*, *Arid1a*, *Sox4* and *Sox11*), whereas cluster 4 was mostly

enriched for cytoskeleton-related terms. Interestingly, cluster 3 and 4 markers are associated with embryonic lethality and abnormal development, including cardiovascular and blood vessel morphology (Fig. S3E, Table S10). Finally, we also observed that the expression of SRF target genes was mostly enriched in cluster 4, despite no detectable changes in *Srf* gene expression (Fig. S3F).

**Fig. 3. Single-cell transcriptome alterations in *Srrm2* heterozygous knockout mESC.** (A-C) Uniform manifold approximation and projection (UMAP) graphs represent the relative similarity between individual cells: coloured by cell cluster (A), shown separately for WT or *Srrm2*<sup>+/-</sup> cells (B), or coloured by the expression levels of core pluripotency factors (C). The displayed numbers in B represent the percentage of cells in each cluster per condition. (D) Expression of a panel of pluripotency markers for naive, formative and primed states along the different cell clusters. Gene markers are labelled on the left side. Each line represents the expression in a single cell. Cells are divided into clusters (from left to right, clusters 1 to 4), illustrated by different colours on the top of the heatmap. (E) Dotplot representing the expression levels of differentially expressed genes between cells of clusters 1 and 2 per cluster. Clusters are represented in different colours on the left of the plot (from down to top, clusters 1 to 4). Circle size corresponds to the percentage of cells expressing the gene of interest. Colour scale represents the average level of expression per cluster. (F) Heatmap representing biological GO terms enrichment from ORAs using gene markers per cluster. The top five significant terms per cluster are displayed. NES of GO terms are represented by a continuous colour scale with yellow representing lowest enrichment and red highest enrichment. Significant NES values are marked with an asterisk (Fisher's exact test \*: adjusted *P*-value<0.05; \*\*: adjusted *P*-value<0.01; \*\*\*: adjusted *P*-value<0.001). (G,H) UMAP projection coloured according to the predicted transferred ID labelling from publicly available *in vivo* scRNA-seq reference atlas of mouse embryogenesis cells from E3.5-5.5 (Nowotchin et al., 2019) (G) and E6.5-8.5 (Chan et al., 2019; Grosswendt et al., 2020; Haggerty et al., 2021) (H). Cells are classified into amnion mesoderm early, amnion mesoderm late, angioblasts, differentiated trophoblasts, ectoderm early 1, ectoderm early 2, epiblast, extraembryonic ectoderm 1 (ExE 1), extraembryonic ectoderm 2 (ExE 2), gut, neural ectoderm anterior, notochord, parietal endoderm, pharyngeal arch mesoderm, primordial germ cells (PGCs), preplacodal ectoderm, surface ectoderm, secondary heart field/splanchnic lateral plate, splanchnic-lateral/anterior-paraxial mesoderm, streak pre-specified/anterior, unassigned (H).

These results suggest that *Srrm2*<sup>+/-</sup> cells have combined features associated with partial loss of pluripotency into more advanced stages of early development and specific embryonic lineages.

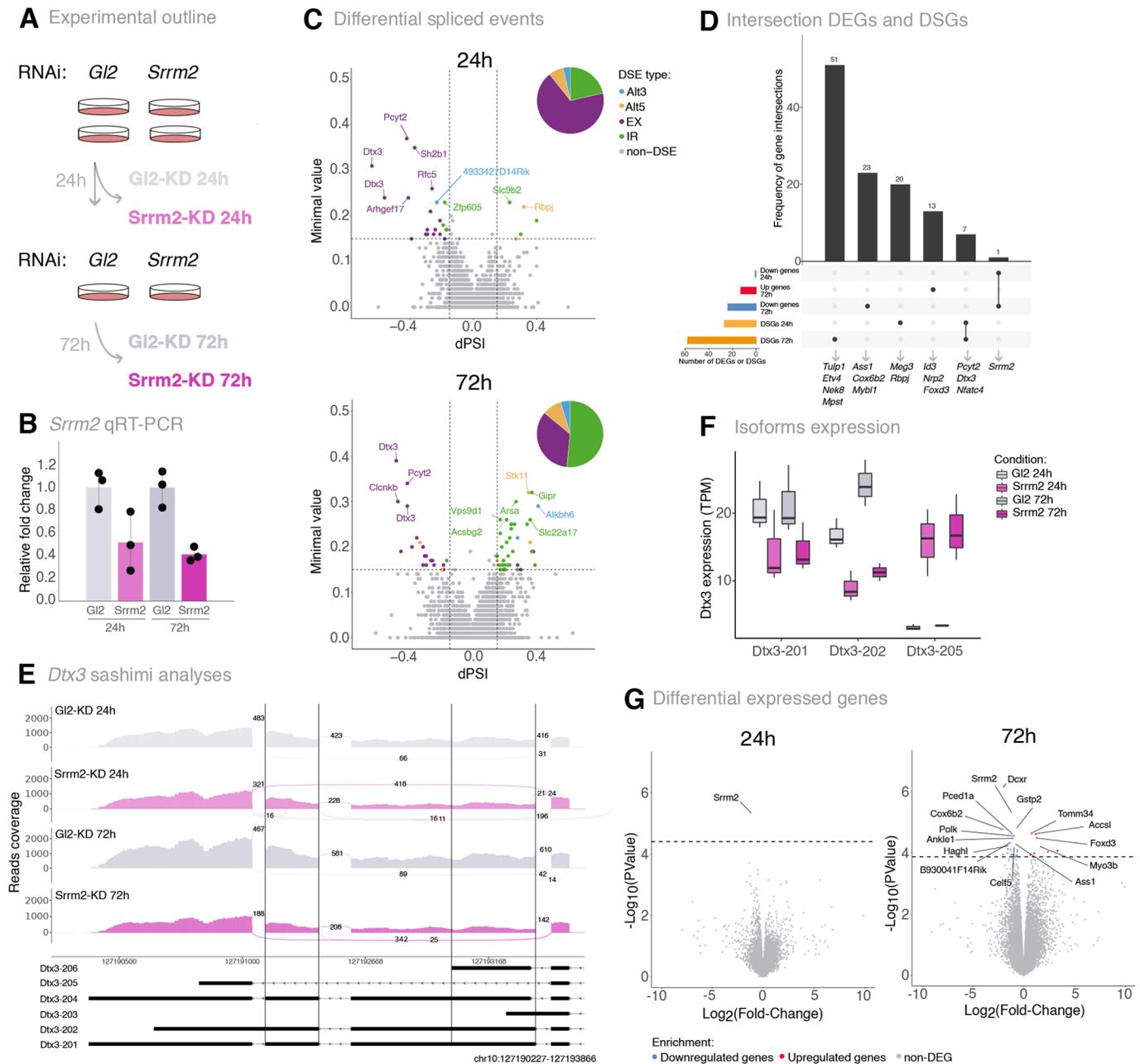
To investigate a potential equivalence between *Srrm2*<sup>+/-</sup> cells and *in vivo* embryonic development, we leveraged published scRNA-seq data of mouse embryogenesis covering either the pre-implantation blastocyst into an implanted egg cylinder stage (E3.5-5.5; Nowotchin et al., 2019), or the gastrulation and early organogenesis stages (E6.5-8.5; Chan et al., 2019; Grosswendt et al., 2020). We transferred cell-type annotations from the two references, E3.5-5.5 and E6.5-8.5, respectively, to the scRNA-seq data collected from *Srrm2*<sup>+/-</sup> and WT cells, using the scmap label transfer method (Kiselev et al., 2018). As expected, when compared to early development cell references, the majority of WT cells are assigned to the earlier epiblast cells and only a small fraction is assigned to the later stage (93% to E4.5 and 6.7% to E5.5 epiblast; Fig. 3G). In contrast, the *Srrm2*<sup>+/-</sup> cells are more frequently assigned to the later-stage epiblast E5.5, although the majority of cells are still assigned to the early epiblast cells of age E4.5 (76% to E4.5 and 20% to E5.5 epiblast). Furthermore, a larger fraction of *Srrm2*<sup>+/-</sup> cells resembled committed cell types, such as trophoblast, extra-embryonic ectoderm and primitive endoderm (3.8% in *Srrm2*<sup>+/-</sup> and 0.3% in WT, Fig. S3G). When compared to the later stage embryos (E6.5-8.5), we observed that most of the WT (98%) and *Srrm2*<sup>+/-</sup> cells (94%) were assigned to the epiblast (Fig. 3H). However, a larger subset of *Srrm2*<sup>+/-</sup> cells (6.1%) compared to WT cells (2.4%) were transcriptionally most similar to differentiated cell types, without an evident cell-type preference (Fig. S3H). Taken together, these results suggest that *Srrm2*<sup>+/-</sup> cells retain transcriptional profiles of early embryonic phenotypes, with a tendency to lose epiblast expression signatures.

### Transient *Srrm2* knockdown reveals alterations in specific splicing events

Gene expression alterations observed in the constitutive *Srrm2*<sup>+/-</sup> heterozygous knockout ESC can comprise both direct effects of the SRRM2 protein, but also cellular responses due to genetic compensation response to the permanent depletion. To identify genes associated with the earliest cascade of deregulation driven by *Srrm2* dosage decrease, we used RNA interference (RNAi) to knockdown *Srrm2* in WT E14tg2a.4 cells for 24 h and 72 h, using as control small interfering RNA (siRNA) that target the firefly luciferase gene *Gl2* (*Srrm2*-KD and *Gl2*-KD, respectively; Fig. 4A). qRT-PCR analyses confirmed depletion of *Srrm2* transcript to 50 and 40% at 24 h and 72 h, respectively (Fig. 4B). RNAi-induced reduction in the dosage of *Srrm2* transcripts did not induce noticeable changes in the morphology of ESC colonies or an increased number of colony escapees at 24 h, whereas at 72 h the *Srrm2*-siRNA treated colonies are no longer fully round and the cell morphology becomes more typical of cells that begin to differentiate (Fig. S4A), suggesting that the conditional *Srrm2* depletion eventually interferes with pluripotency.

To determine genome-wide changes in splicing events, we collected poly-adenylated RNA at 24 h and 72 h upon RNAi treatment in three independent biological replicates. Using Vast-tools, we found 382,645 and 387,215 splicing events, of which 758 and 1398 were differential, respectively at 24 and 72 h of siRNA treatment (Han et al., 2017; Irimia et al., 2014; Braunschweig et al., 2014; Tapial et al., 2017). To identify the most robust differential splicing events, we applied stringent thresholds based on read coverage quality in at least two replicates, a significance  $\text{abs}(\text{dPSI}) \geq 0.15$  and a minimal value  $\geq 0.15$ , and a curation of PSI distribution of reads over the spliced junctions. We found significant differentially splicing events (DSEs) in 78 differentially spliced genes (DSGs), with 28 DSEs detected in 27 genes at 24 h, and 64 DSEs detected in 58 genes at 72 h (Fig. 4C; Tables S11,12). The splicing alterations corresponded predominantly to exon-skipping events at 24 h, similar to the type of splicing alterations identified in constitutive *Srrm2*<sup>+/-</sup> cells (Fig. 2D) and were followed by an increase in intron retention events in the longer 72 h exposure. All exon skipping events in the *Srrm2*-KD were characterised by exon exclusion without alterations in microexon splicing, as previously reported in other systems (Xu et al., 2022; Torres-Méndez et al., 2019; Cui et al., 2023).

GO analyses revealed no statistically significant enrichment in specific biological functions of the 78 DSGs, but a closer inspection revealed 22 genes associated with development, including *Pcyt2* (phosphate cytidylyltransferase 2), *Dtx3* (Deltex E3 Ubiquitin Ligase 3), transcription factors *Rbpj*, *Nfatc4*, *Etv4* and *Tulp1*, and long non-coding RNA *Meg3*. Interestingly, *Rbpj* is also a cluster 2 marker in the scRNA-seq analyses and is associated with exit from naive pluripotency to the formative state (Kalkan et al., 2019; Lackner et al., 2021). NFATC4 is a transcription factor required for cardiac development (Bushdid et al., 2003; Graef et al., 2001), which undergoes an exon skipping event that alters the relative abundance of protein-coding isoforms (Table S13). ETV4 expression can impair ectodermal commitment (Akagi et al., 2015), and also undergoes an exon-skipping event in *SRRM2*-depleted differentiated human cells (Cui et al., 2023). TULP1 is important for neuron and photoreceptor cell development (Jia et al., 2022), and *Meg3* transcripts sponge the microRNA *miR-423-5p* that in turn downregulates *Srf* expression (Cheng et al., 2020), which is upregulated in *Srrm2*<sup>+/-</sup> cells.



**Fig. 4. Transcriptome alterations upon *Srrm2* RNAi knockdown.** (A) Overview of the experimental outline. mESC were treated with RNAi targeting *Gl2* luciferase gene (control) or *Srrm2*. Cells were harvested at 24 h or submitted to another round of RNAi and harvested at 72 h. RNA was extracted simultaneously. (B) *Srrm2* expression levels. Total RNA levels of exon 11 of *Srrm2* were measured by RT-qPCR. Relative levels are normalised to  $\beta$ -actin. Mean and standard deviation are calculated from three replicates. (C) Volcano plot showing the differentially spliced events upon *Srrm2* knockdown at 24 h (top) and 72 h (bottom). The magnitude of alterations is represented by the dPSI and the minimal value. 324,859 and 303,695 splicing events were detected at 24 h and 72 h, respectively. 28 and 64 splicing events were considered as significantly differentially spliced, at 24 h and 72 h, respectively, using a cut-off of  $abs(dPSI) \geq 0.15$  and a minimal value  $\geq 0.15$ . Spliced events with a dPSI < 0 correspond to inclusion events in *Srrm2* RNAi-treated cells, whereas dPSI > 0 correspond to exclusion events in *Srrm2* RNAi-treated cells. Categories of significant differential spliced events are represented by different colours: alternative 3' splice site (Alt3, blue), alternative 5' splice site (Alt5, orange), exon exclusion (EX, purple), intron retention (IR, green), no alteration (non-DSE, grey). The pie chart represents the proportion of significant differential spliced events. Gene labels represent the top ten genes undergoing the differentially spliced events with the highest minimal value. (D) Upset plot representing the comparison of differentially altered genes upon *Srrm2* knockdown at the expression and splicing levels for 24 h and 72 h. (E) Sashimi plots displaying the differential splicing events for the *Dtx3* gene region (chr10:127190227-127193866) in control samples and *Srrm2* knockdown samples at 24 h and 72 h. Splicing events are delimited with vertical black lines (*Dtx3*: chr10:127192609-127193357 and chr10:127191336-127191553). Numbers in the plot represent the average number of junction reads from three replicates. Isoforms that cover the represented genomic region are shown and identified as following: ENSMUST00000144322 (*Dtx3*-206), ENSMUST00000137151 (*Dtx3*-205), ENSMUST00000130855 (*Dtx3*-204), ENSMUST00000125254 (*Dtx3*-203), ENSMUST00000116229 (*Dtx3*-202), ENSMUST0000038217 (*Dtx3*-201). (F) Expression levels (TPM) of *Dtx3* isoforms upon *Gl2* (control) and *Srrm2* knockdown at 24 h and 72 h. Isoforms with expression levels above 5 TPM and encoding for proteins are displayed. (G) Volcano plots showing the differentially expressed genes upon *Srrm2* knockdown at 24 h (left) and 72 h (right). 19,461 genes were detected in the whole dataset, at 24 h and 72 h. 1 and 37 genes were considered as differentially expressed, at 24 h and 72 h, respectively, using an adjusted  $P$ -value < 0.05 as a cut-off (downregulated and upregulated genes are marked in blue and red, respectively).

Seven genes had common differential splicing events between 24 h and 72 h (Fig. 4D), with eight common differential splicing events, all of which are exon cassette skipping (Fig. S4B). Among these genes, *Dtx3* and *Pcyt2* were also differentially spliced in *Srrm2*<sup>+/-</sup> cells, with two and one exon skipping events, respectively (Fig. 2C), showing that their splicing deregulation occurs upon acute *Srrm2* depletion and prevails in an adapted heterozygous knockout ESC. DTX3 can act both as an activator or repressor of NOTCH signalling, a key pathway that regulates cell fate determination (Ding et al., 2020; Wang et al., 2021b). *Pcyt2* is an embryonic-lethal gene that encodes a protein involved in phospholipid synthesis (Fullerton et al., 2007). Visual inspection of the density of RNA-seq reads along splice junctions, using sashimi plots, confirmed that the skipped exons have a lower proportion of coverage of sequencing reads at both genes, in *Srrm2*-KD cells at both 24 h and 72 h (Fig. 4F; Fig. S4C).

To further explore the impact of the exon skipping in *Dtx3* and *Pcyt2*, we compared the expression levels of their annotated protein-coding isoforms. At both 24 and 72 h, we found a decrease in the two highest expressed *Dtx3* protein-coding isoforms (Dtx3-201 and Dtx3-202) accompanied by a strong upregulation of another protein-coding isoform (Dtx3-205) (Fig. 4F). Notably, Dtx3-205 encodes for a DTX3 protein that lacks a RING domain present in Dtx3-201 and Dtx3-202, which is responsible for ubiquitination transfer. For *Pcyt2*, the highest expressed isoform, which encodes the canonical PCYT2 (PCYT2 $\alpha$ ) protein isoform, did not show notable expression differences, whereas a *Pcyt2* protein-coding isoform that lacks exon 7 (Pcyt2 $\beta$ ) was more highly expressed at both 24 h and 72 h (Fig. S4D). The corresponding protein isoform has the same functional domains as *Pcyt2* $\alpha$  but differs in catalytic and kinetic activities and is expressed in a tissue-specific manner (Tie and Bakovic, 2007). *Pcyt2* $\alpha$  is more prominently expressed than *Pcyt2* $\beta$  in most mouse tissues, except in muscle and heart tissues where their expression is similar. The same splicing alteration in PCYT2 was reported after SRRM2 depletion in the human hepatoma cell line HepG2 (Cui et al., 2023).

### Transient *Srrm2* depletion in ESC leads to mild changes in genome-wide gene expression

Finally, we investigated gene expression changes upon *Srrm2* RNAi treatment. At 24 h, only *Srrm2* was significantly downregulated, while 24 genes were downregulated at 72 h, including *Srrm2*, and 13 were upregulated (Fig. 4G; Tables S14,15). Among the 78 differentially spliced genes in *Srrm2*-KD, none is differentially expressed at 24 h or 72 h (Fig. 4D), suggesting that the splicing alterations at specific genes related with *Srrm2* dosage decrease precede the more global gene expression changes detected in constitutive *Srrm2* heterozygosity.

Within the 37 differential expressed genes detected in *Srrm2*-KD cells, we found embryonic development-related and transcription factor genes, such as *Id3*, *Mybl1*, *Nrp2*, *Fgf11* and *Foxd3* (Fig. 4G), with *Id3*, *Mybl1* and also *Cox6b2*, being differentially expressed in the same direction in *Srrm2*<sup>+/-</sup> cells (Fig. S4E). Out of the 37 genes found differentially expressed upon *Srrm2* knockdown, only 16 were differentially expressed in the constitutive *Srrm2*<sup>+/-</sup> cells, and only 9 were differentially expressed in the same direction (Fig. S4E). These results suggest that the imbalance in cell identity found upon constitutive *Srrm2* dosage reduction in *Srrm2*<sup>+/-</sup> cells, likely results from a few direct effects of SRRM2 likely through splicing deregulation, which lead to a cascade of indirect effects on stemness and cell lineage regulatory networks.

To zoom in further on the genes that are deregulated in both constitutive and RNAi depletion of *Srrm2*, we compared the expression of the *Srrm2*-KD DEGs with the earliest gene expression alterations of *Srrm2*<sup>+/-</sup> cells (Fig. S4F). Although the correlation of expression was very low ( $R=0.3$ ,  $P\text{-value}=0.12$ ), we found that the genes that are differentially expressed after RNAi also tend to change expression in the same direction in the most pluripotent states of *Srrm2*<sup>+/-</sup> cells detected by scRNA-seq, clusters 1 and 2. Amongst these, we find *Cox6b2*, a gene marker of cluster 1, and significantly downregulated in cluster 2, and at 72 h upon *Srrm2* RNAi treatment (Fig. 4G; Fig. S4F,G). *Ass1*, which encodes for a protein that catalyses the penultimate step of the arginine biosynthetic pathway, is also a marker gene of cluster 1, is found significantly downregulated at 72 h in *Srrm2*-KD cells, and its expression decreases from cluster 1 to 2, although not significantly. In contrast, the myogenesis-related transcription factor *Id3* is upregulated 72 h after *Srrm2* depletion and it is highly expressed in cluster-4 state cells, corresponding to the most differentiated group of ESC within the cultures.

Taken together, our results show that transient depletion of *Srrm2* leads to exon cassette exclusion and intron retention splicing alterations in specific genes, including development-related genes. These specific splicing alterations precede gene expression alterations, suggesting that a small number of specific SRRM2-dependent splicing defects drive the gene expression alterations that ultimately lead to an impairment in the stemness of ESC upon *Srrm2* depletion.

### DISCUSSION

In this work, we investigated the roles of *Srrm2* dosage in stemness through the modulation of splicing and gene expression networks, using *Srrm2* heterozygous knockout or transient knockdown in mouse ESC. The RNAi knockdown revealed that the most acute effects of *Srrm2* half dosage are alternative splicing defects in a small number of genes with roles in gene and metabolism regulation, without immediate changes in gene expression levels. In contrast, single-cell transcriptomic analyses showed that the constitutive half dosage of *Srrm2* in ESC gives rise to heterogeneous cultures reliably composed of cells in different transcriptional and stemness states. Embryonic stem cells have the unique capacity to differentiate into a wide range of specialised cell types (pluripotency) and to perpetually propagate in culture while retaining their undifferentiated state through self-renewal (Martin, 1981; Evans and Kaufman, 1981; Thomson et al., 1998; Huang et al., 2015). Our study reveals that a persistent reduction in *Srrm2* transcripts in stem cells reduces the competency to maintain stemness and results in the emergence of a mixed cell population with distinct phenotypes, which include defective colony formation and altered expression of pluripotency markers. *Srrm2*<sup>+/-</sup> cells significantly overexpress genes related to development and cytoskeleton, whereas the group of downregulated genes are enriched in genes associated with the LIF pathway, crucial for ESC self-renewal (Smith et al., 1988; Williams et al., 1988; Huang et al., 2015). Notably, LIF withdrawal or the direct inhibition of downstream effectors leads to differentiation of mouse ESC into morphologically mixed cell populations, similar to what we observe in *Srrm2*<sup>+/-</sup> population (Niwa et al., 1998). We identify *Srrm2* dosage as a key player in balancing self-renewal and differentiation processes, to ensure appropriate proportions of stem cell niches.

Single-cell RNA-seq analysis revealed distinct cell states among *Srrm2*<sup>+/-</sup> cells. While the majority of *Srrm2*<sup>+/-</sup> cells retain pluripotency expression signatures, namely *Pou5f1*, *Sox2*, *Nanog*, *Klf4*, their pluripotency state is markedly different from the most abundant pluripotent state in WT cells. Mouse ESC cultured in the presence of differentiation-restricting inhibitors, such as LIF, are in

*in vitro* counterparts of the naive preimplantation epiblast E4.5. EpiLCs, derived from mESC, resemble the early post-implantation epiblast (E5.5-6.5) and exhibit features of the formative pluripotent state which precedes the induction for multi-lineage specialisation (Hayashi et al., 2011; Rossant and Tam, 2017; Kurimoto et al., 2015; Yang et al., 2019). Compared to the WT condition, the population of *Srrm2*<sup>+/-</sup> ESC contains a higher proportion of cells resembling E5.5 stage cells, and we found that most *Srrm2*<sup>+/-</sup> ESC exist in an intermediate state between the naive and formative pluripotency. This intermediate state is characterised by a reduction in the expression of *Nanog* and *Klf4*, and overexpression of the DNA methyltransferases, *Dnmt3a* and *Dnmt3b*, and several ribosomal protein genes, such as *Rps21*, *Rps29*, *Rpl29*, *Rpl36*. Moreover, transient *Srrm2* depletion by RNAi has a rapid effect on the expression of metabolic-related genes. In particular, we found misregulation of *Cox6b2*, involved in mitochondria oxidative phosphorylation, and *Ass1*, a component of the arginine pathway important in protein synthesis. Metabolic cues are tightly regulated during pluripotency and differentiation, and proper regulation of ribosomal gene expression is crucial for the requirements of different cell states. This highlights a possible role of *Srrm2*, and more specifically its dosage, as an early modulator of gene expression networks, but also in controlling metabolic pathways, both key to balancing a coordinated and dynamic progression of pluripotent states during embryo development.

The modulation of transcription factor expression cascades, and the putative alteration of the epigenetic and metabolic landscape in *Srrm2*<sup>+/-</sup> cells, suggests that *Srrm2* depletion leads to a partial dismantling of stem regulatory networks, promoting the onset of cell lineage commitment, into a state where the majority of cells are beyond their naive pluripotency features but are still not committed to a particular lineage. Alternatively, it is possible that the intermediate state of *Srrm2*<sup>+/-</sup> cells between naive and formative pluripotency does not exist *in vivo* in embryos, or in standard cultures of ESC. Regardless, our study suggests that *Srrm2* levels play important roles in coordinating the transitions or the temporal dynamics between naive and formative pluripotent stem cells, and the robustness of their stemness. Further studies may help elucidate possible roles of *Srrm2*-mediated mechanisms in promoting fluctuations of pluripotent states that provide windows of opportunities for cell commitment.

*Srrm2*<sup>+/-</sup> cells express cell-type-specific markers and transcription factors important for mesoderm or cardiac tissues. Importantly, upon *Srrm2* RNAi, we also observed misregulation of genes associated with mesoderm and angiogenesis, such as *Id3*, *Myo3b* and *Nrp2*, before detectable changes in pluripotency regulation genes. A stable and continuous depletion of *Srrm2* in *Srrm2*<sup>+/-</sup> mESC leads to a significant enrichment of the *Srf* transcription factor, a master regulator of cardiac development, and its target genes, especially in the most differentiated cells. SRF activation is driven by cell mechanic cues such as extracellular physical or chemical signals that are transduced by actin cytoskeleton remodelling signalling cascades coordinated with activation of myocardin-related transcription factors (Olson and Nordheim, 2010; Posern and Treisman, 2006; Xiong et al., 2022). Cytoskeletal and mechanical cues play key roles in regulating pluripotency by modulating the transcriptome of mouse ESC in response to external signalling cues (David et al., 2019; Heo et al., 2017; Putra et al., 2023). The misregulation of cytoskeletal-related genes in *Srrm2*<sup>+/-</sup> cells, such as *Krt8* and *Krt18*, could potentially identify important players in controlling cell mechanical cues, which may be interconnected with alterations in SRF-related pathways. Transient depletion of *Srrm2* led to splicing alterations

in *Meg3* lncRNA, which indirectly regulates *Srf* transcript levels through its sponging functions. Alternative splicing of *Meg3* could lead to alterations in its three-dimensional folding, impacting its sponging abilities, and potentially explaining the increase in SRF targets. Further studies are required to better understand whether the alterations in *Meg3* isoforms, cytoskeleton and SRF-pathway drive or result from the loss of stemness in *Srrm2*<sup>+/-</sup> ESC.

While *Srrm2*<sup>+/-</sup> cells show extensive gene expression alterations with few splicing alterations, transient *Srrm2* depletion leads to minor expression changes and reveals that splicing misregulation, mostly involving exon skipping and intron retention, as previously shown in differentiated cells (Cui et al., 2023; Xu et al., 2022), occurs before gene expression effects. Notably, *Dtx3* and *Pcyt2* exhibit differential splicing events in protein-coding isoforms both in acute and constitutive *Srrm2* dosage reduction, indicating that they may play pivotal roles in triggering the cascade of *Srrm2*-associated phenotypes. DTX3 is a ubiquitin ligase with a key role in cell fate determination and is a regulator of Notch signalling, which governs numerous processes during embryo development. PCYT2 is an enzyme involved in the Kennedy pathway of phospholipid synthesis, and its roles in development are still not known. These findings suggest that *Srrm2*-dependent splicing of specific genes, including *Dtx3* and *Pcyt2*, as a likely mechanism driving subsequent changes in gene expression and ultimately impairing pluripotency.

Further research is needed to provide insights into the phenotypical and molecular mechanisms underlying the effects of *Srrm2* dosage reduction in pluripotency maintenance and early differentiation. The observation of different outcomes from the *Srrm2* depletion by constitutive heterozygous knockout or acute RNA interference recapitulates studies in *C. elegans*, in which depletion of the *Srrm2* ortholog (*rsr-2*) by RNA interference and knockout leads to different phenotypes, from viable worms able to reach the adult stage, to embryonic lethality and larvae arrest (Fontrodona et al., 2013). The complex effects of *Srrm2* depletion during development can be reconciled with roles in the regulation of cell fate and stemness that are likely to be sensitive to *Srrm2* levels, and an accumulation of deregulated pathways, and additional factors that require investigation through complementary approaches. Future studies should involve different genetic manipulation technologies, additional clones, developmental models, such as gastruloids, combined with molecular approaches that directly identify target RNA isoforms dependent on *Srrm2* abundance. SRRM2 has also been shown to be important for the biogenesis and composition of splicing speckles, which are nuclear domains enriched for splicing factors (Ilik et al., 2020; Xu et al., 2022; Hu et al., 2019). For a better understanding of SRRM2-mediated mechanisms in early development, it will also be important to disentangle its longer-term effects that result in altered cell identity seen in mouse and *C. elegans*, with its direct molecular roles in alternative splicing, splicing speckle formation and genome organisation in the context of pluripotency and development.

Taken together, our study sheds light on the roles of splicing factors in the intricate interplay between alternative splicing and gene expression networks that govern early developmental processes. *Srrm2* emerges as a key regulator of pluripotency and early development potentially through its dosage-dependent regulation of specific splicing events and its ability to balance unscheduled gene expression alterations. Furthermore, the identification of early and long-lasting *Srrm2*-mediated splicing and expression defects at specific genes prompts further investigation of their roles in biological and medical contexts, opening opportunities for exploring

new therapeutic strategies targeting *Srrm2* and its downstream effectors in developmental and cancer diseases.

## MATERIALS AND METHODS

### *Srrm2* expression analysis during *in vivo* mouse development

*Srrm2* expression levels were assessed during E3.5–5.5 mice embryo development using publicly available data (Nowotwschin et al., 2019), and the force-directed layouts plots in Fig. 1A were produced using the interactive data website [www.endoderm-explorer.com](http://www.endoderm-explorer.com). Gene expression colour per cell represents the post-imputation with the MAGIC algorithm (van Dijk et al., 2018) following normalisation and log transformation. *Srrm2* expression levels during E10–P63 mice embryo development were assessed using publicly available data (Cardoso-Moreira et al., 2019). The data used to produce Fig. 1B was obtained from the interactive app <https://apps.kaessmannlab.org/evodevapp/>.

### mESC lines

*Srrm2*<sup>+/-</sup> cells (CG0630) were generated by a gene trap-derived truncation in the *Srrm2* gene by the Sanger Institute Gene Trap Resource (SIGTR), Cambridge, UK, using mouse ESC lines clone E14tg2a.4, as a recipient. *Srrm2*<sup>+/-</sup> and parental WT E14tg2a.4 ESC were obtained from the Mutant Mouse Resource Center (MMRRC) at the University of California at Davis, a National Institutes of Health (NIH) funded strain repository. MMRRC screening in *Srrm2*<sup>+/-</sup> cells reported *Mycoplasma spp.* and mouse Parvo viruses free, and a chromosome count of 40 chromosomes. The specificity of the gene-trap cassette integration in *Srrm2*<sup>+/-</sup> cells was confirmed by MMRRC through 5' RACE genotyping which identified a unique sequence mapping to *Srrm2* Exon 1 (NNNNNNNNNNNNNNNTNNGCNCGCT-TCTGGTGCCGGAACCAGGCAAAGCGCCATTCGCCATTCAGGCTG-CGCAACTGTTGGGAAGGGCGATCGGTGCGGGCTCTTCGCTAT). E14tg2a.4 WT cells were expanded and tested negative for *Mycoplasma spp.*, upon screening according to the manufacturer's instructions (AppliChem, A3744). E14tg2a.4 WT and *Srrm2*<sup>+/-</sup> cells were expanded and cryopreserved at three different passage numbers (WT: passage 19–21, *Srrm2*<sup>+/-</sup>: passage 4–7) in three independent batches, that were further thawed and treated as biological replicates. Bulk sequencing datasets produced from E14tg2a.4 WT and *Srrm2*<sup>+/-</sup> were mapped and tested negative for *Mycoplasma spp.*

### mESC culture

mESC were grown as previously described (Beagrie et al., 2017). Briefly, cells were grown at 37°C, 5% (v/v) CO<sub>2</sub>, on gelatine-coated (0.1% v/v, Sigma-Aldrich, cat# G1393) Nunc T25 flasks in Gibco Glasgow's MEM (Invitrogen, 21710-082), supplemented with 10% (v/v) heat-inactivated FCS (Invitrogen, 10270-106, batch number 41F8126K), 2000 units ml<sup>-1</sup> LIF (Millipore, ESG1107), 0.1 mM β-mercaptoethanol (Invitrogen, 31350-010), 2 mM L-glutamine (Invitrogen, 25030-024), 1 mM sodium pyruvate (Invitrogen, 11360-070), 1% penicillin–streptomycin (Invitrogen, 15140-122) and 1% non-essential amino acids (Invitrogen, 11140-035). The cell culture medium was changed every day and cells were split every second day. Before sample collection, cells were grown for 24 h, 48 h or 72 h in serum-free ESGRO Complete Clonal Grade medium (Merck, SF001-500P), supplemented with 1000 units ml<sup>-1</sup> LIF, on 0.1% gelatine-coated (Sigma-Aldrich, G1393-100ml, 0.1% v/v) Nunc dishes, with a daily medium change. Cultures were visually inspected and brightfield images were acquired in an Olympus IX53 microscope.

### Alkaline phosphatase assay

WT and *Srrm2*<sup>+/-</sup> E14tg2a.4 cells were seeded at a clonal density of 8.5 × 10<sup>3</sup> cells/cm<sup>2</sup> and grown for 72 h under normal conditions. The Alkaline Phosphatase assay was carried out according to the manufacturer's instructions (Sigma-Aldrich, 86R-1KT). Alkaline phosphatase (AP) activity is highest in stem cells, and it reduces as cells differentiate into other cell types, thus AP activity is commonly used as a means of determining the ability of cells to differentiate. Cells with higher AP activity show higher Hematoxylin staining. AP staining was visually inspected and brightfield images were acquired in an Olympus IX53 microscope.

### *Srrm2* RNAi knockdown in mESC

WT E14tg2a.4 ESC were transfected using OptiMEM (Invitrogen, 31985062), Lipofectamine RNAiMAX (Invitrogen, 13778075) and 50 nM synthetic siRNA duplexes (Eurogentec) targeting either *Srrm2* or firefly luciferase (Gl2), which was used as a control. RNA interference was done according to the manufacturer's instructions. Briefly, cells were seeded and reverse-transfected at a density of 3.1 × 10<sup>4</sup> cells/cm<sup>2</sup>. 24 h after the first transfection, cells were harvested or re-transfected with the same siRNA duplexes and transfection reagents and harvested 72 h after seeding. The siRNA sequences are reported from 5' to 3': Gl2 sense CGUACGCG-GAAUACUUCGA, Gl2 antisense UCGAAGUAUUCGCGUACG, *Srrm2* E8/9 sense GAAGAAGCACAGGUCAGAA, *Srrm2* E8/9 antisense UUCUGACCUGUGCUUCUUC, *Srrm2* E13 sense CGUGGAGAUCU-CACUCUA, *Srrm2* E13 antisense UAGAGUGAGAUCUUCACG.

### RNA isolation

For total RNA-seq and qRT-PCR, *Srrm2*<sup>+/-</sup> and WT ESC were seeded (5.2 × 10<sup>4</sup> cells/cm<sup>2</sup>) and grown for 48 h to 70–80% confluency. Total RNA was isolated with TRIzol (Invitrogen, 15596-026) following the manufacturer's instructions. Lysates were either processed immediately for qRT-PCR, or snap-frozen in liquid nitrogen and stored at –80°C until further processing. DNA was degraded with TURBO DNase (Invitrogen, AM1907), according to the manufacturer's protocol. Purified RNA was either stored at –80°C or processed immediately.

### cDNA synthesis and real-time quantitative PCR

Purified RNA (1 μg) was reverse transcribed into cDNA using 10 U/μl of Superscript II Reverse Transcriptase (Invitrogen, 18064-022), 50 ng of random primers (Promega, C118A), 0.5 mM dNTPs (NEB, N0447L), 5 mM RNaseOUT (Invitrogen, 10777-019) in a 20 μl solution with 10 mM DTT and Superscript buffer, according with the manufacturer's instructions. RNA-DNA hybrids were removed with 2 U of RNase H (NEB, M0297S). The synthesised DNA was diluted 1:4 and 1 μl was used for qRT-PCR, using SensiMix™ SYBR® No-ROX (Meridian Bioscience, QT650-05), 250 nM of each primer in a final volume of 20 μl. The primers sequence are reported from 5' to 3': *Srrm2* Exon 11 FW AGTCTCTCGTAGA-AGCCGGT, *Srrm2* Exon 11 RV CTTCTGCGTCTTGGTGGAGT, *Actb* FW TCTTTGCGACTCCTTCGTTG and *Actb* RV ACGATGGAGGG-GAATACAGC. *Srrm2* expression was measured from three biological replicates per condition and were assessed according to the 2–ΔΔCT method (Livak and Schmittgen, 2001), and normalised to *Actb*. The relative RNA expression for each condition was estimated by calculating the ΔCT average per replicate. The mean and standard deviation (s.d.) were then determined from the replicates.

### Immunofluorescence

*Srrm2*<sup>+/-</sup> and E14tg2a.4 WT cells were seeded (1 × 10<sup>5</sup> cells/cm<sup>2</sup>) on ethanol-washed and autoclaved glass 0.1% gelatin-coated coverslips and grown under regular conditions for 1 day. Cells were rinsed and fixed in 4% EM-grade PFA (Alfa Aesar, 43368) in 125 mM HEPES-NaOH pH 7.6, for 30 min at room temperature. Aldehydes were quenched (10 min) with 20 mM glycine in PBS. Cells were permeabilised (10 min) in 0.5% Triton X-100 in PBS (w/v). Cells were washed and incubated (three times over 1 h) in blocking solution (1% BSA, 0.2% fish skin gelatin, 0.1% casein, in PBS, pH 7.8). Primary antibodies were diluted in blocking solution and incubated overnight at 4°C in a humid chamber. Cells were washed (three times over 1 h) in blocking solution, before incubation with secondary antibody diluted in blocking solution for 1 h at RT. After incubation, cells were washed in PBS three times for 1 h. DNA was stained with 0.5 μg/ml DAPI (Sigma-Aldrich, D9542) in PBS for 2 min. Cells were washed three times in PBS and mounted in Vectashield (Vector Laboratories, H-1800) before imaging. The antibodies used are listed in Table S16.

### Confocal imaging

Immunofluorescence images were acquired with a Leica TCS SP8-STED confocal microscope (Leica Microsystems DMI6000B-CS) using Leica Application Suite X v.3.5.5.19976 and an HC PL APO CS2 63x/1.40 oil

objective, with a pinhole equivalent to 1 Airy disk. Images were acquired using 405 nm, 488 nm, 555 nm, 633 nm and 680 nm excitation using a long-pass filter at 1024×1024 pixel resolution (pixel size 180×180 nm). For intensity signal comparison, representative images from the same experiment set were collected without intensity signal saturation, in the same session with the same settings. Images were post-processed using Fiji software (version 2.9.0/1.53t; Schindelin et al., 2012; Linkert et al., 2010), and comparative images were contrast-stretched using the same parameters.

### Western blotting

E14tg2a.4 WT and *Srrm2*<sup>+/-</sup> cells were seeded (5.2×10<sup>4</sup> cells/cm<sup>2</sup>) and grown for 48 h to 70–80% confluency. The organic phase was extracted following TRIzol protocol (Invitrogen, 15596-026), according to the manufacturer's instructions. Proteins were precipitated with isopropanol and washed with 0.3 M of guanidine hydrochloride in 95% ethanol for 20 min. Protein pellets were resuspended in 1% SDS and protein concentration was measured using the Bradford assay kit (Thermo Fisher Scientific, 1863028), according to the manufacturer's instructions. 10 µg of protein extracts were denatured in 5x Laemmli-buffer (125 mM Tris-HCl pH 6.8, 10% SDS, 50% glycerol, 0.5% bromophenol blue, 0.2% α-mercaptoethanol) for 5 min at 95°C. Protein lysates were loaded on 10% polyacrylamide gels and resolved at 120 V for 2 h at RT. Proteins were transferred to a PVDF membrane using a Mini Trans-Blot<sup>®</sup> Cell (Bio-Rad Laboratories) at 100 V for 2 h, at 4°C. The membrane was blocked in 5% skimmed milk for at least 1 h. Primary antibodies incubation was performed overnight at 4°C and secondary antibodies were incubated for 2 h. Chemiluminescent substrate was added according to the manufacturer's instructions (Amersham ECL Prime Western Blotting Detection Reagent, Cytiva #RPN2232) and was detected with an Amersham Imager 600 (GE Healthcare, 10600023). The antibodies used in this study are listed in Table S16.

### Flow cytometry

*Srrm2*<sup>+/-</sup> and E14tg2a.4 WT cells were grown under regular conditions for 48 h. Triplicate wells (technical replicates) were each harvested at three different passages using TrypLE (Gibco). Cells were fixed and permeabilised using the permeabilising buffers from the FoxP3 staining buffer kit (Miltenyi; 130-093-142) and stained with conjugated antibodies (Table S16). Protein expression was analysed using a MACSQuant VYB flow cytometer (Miltenyi). Gating and plots were generated using FlowJo 10.

### Bulk RNA-seq library preparation and sequencing

Purified RNA was produced in three biological replicates from WT and *Srrm2*<sup>+/-</sup> E14tg2a.4 cells, or from RNAi-treated E14tg2a.4 WT cells and tested for RNA quality using the 2100 Bioanalyzer Agilent system with the Agilent RNA 6000 Nano assay (Agilent Technologies, 5067-1511). RNA spike-ins (Invitrogen, 4456740) were added to each purified RNA sample following the manufacturer's recommendations. Total RNA libraries were produced from 600 ng of RNA using the TruSeq Stranded total RNA library preparation kit (Illumina, 20020596) according to the manufacturer's protocol. PolyA+ RNA-seq libraries were produced from 1 µg of RNA using the TruSeq<sup>®</sup> Stranded mRNA Library Prep (Illumina, 20020594) kit, according to the manufacturer's instructions. Libraries were analysed on the 2100 Bioanalyzer Agilent system using the Agilent High Sensitivity DNA kit (Agilent Technologies, 5067-4626). DNA concentrations were measured by Qubit Quant IT kit (Invitrogen, Q10212) according to the manufacturer's protocol. Total RNA libraries were paired-end sequenced using the NextSeq 500/550 system at 75 bp length, according to the manufacturer's protocol. Sequencing depth ranged from 78–99 million reads in WT and *Srrm2*<sup>+/-</sup> replicates. PolyA+ RNA libraries were paired-end sequenced using the NovaSeq 6000 system at 100-bp length, according to the manufacturer's protocol. Sequencing depth ranged from 91–251 million reads per replicate.

### Total and polyA+ RNA-seq data analyses

#### Gene expression

Total and polyA expression levels from RNA-seq profiles were obtained using Kallisto v0.46.1 (Bray et al., 2016), where reads were pseudo-aligned

to the mouse gencode transcriptome release M24. Differential expression analysis was conducted using edgeR (version 3.38.4). Briefly, sample comparison was performed by filtering for low expressed genes, applying a quasi-likelihood method (glmQLFTest function), and selecting significantly differentially expressed genes with FDR adjusted *P*-values lower than 0.05, as implemented in the edgeR package. For polyA+ RNA-seq, an additional step of batch effect correction for the processing date of the samples was performed indicating the RNA processing day as a covariate in the design matrix. To assess the accuracy of the transcriptome alteration, we analysed the correlation between the observed and expected fold changes of the spike-ins (total RNA: *R*=0.97, *P*-value<2.2×10<sup>-6</sup>; 24 h and 72 h polyA+ RNA: *R*=0.98, *P*-value<2.2×10<sup>-6</sup>).

### GSEA and over-representation analyses

GSEA and ORA were performed with WEB-based GENE SeT AnaLysis Toolkit (WebGestalt 2019), using the default parameters and weighted set to reduce redundancy of the gene sets in the enrichment result (Zhang et al., 2005; Wang et al., 2013, 2017; Liao et al., 2019). For GSEA all the detected genes were used as an input and ranked according to their *P*-value and the sign of the fold change, as previously described (Reimand et al., 2019). GSEA was performed for the default functional databases: biological processes (GO) and predicted transcription factor targets of networks (Gene Transcription Regulation Database, GTRD) (Kolmykov et al., 2020). Furthermore, external databases were used as functional databases in WebGestalt to perform GSEA and ORA. To assess commitment to a specific cell lineage using signatures previously identified by single-cell transcriptome profiles of mouse embryos (Cao et al., 2019) and retrieved from MSigDB's hallmark collection (<https://www.gsea-msigdb.org/>; Liberzon et al., 2015). To identify the transcription factors necessary to induce cell-type reprogramming, we used the publicly available Mogrify (Rackham et al., 2016) gene lists (Ferrai et al., 2017) as a functional database to perform an ORA, providing *Srrm2*<sup>+/-</sup> upregulated genes as input and all detected genes as background. Finally, the list of SRF targets predicted by the GTRD was downloaded from the Human Molecular Signatures Database (MSigDB), C3 subcollection TFT: Transcription factor targets and converted into *Mus musculus*, using the Bioconductor BiomaRt R package and the *M. musculus* transcription factors list was obtained from the AnimalTFDB 3.0 database (Hu et al., 2018).

### Splicing alterations

For AS event-level analysis, raw reads were mapped to previously annotated AS events for the mouse reference assembly (mm10 vastdb.mm2.23.06.20) from VastDB (VastDB v2 released). Abundance of AS events in PSI values was estimated using Vast-tools (version 2.5.1) align and combine commands (Irimia et al., 2014; Tapial et al., 2017). Vast-tools quantifies exon and microexon skipping, intron retention, and alternative donor and acceptor site choices. For further analyses, we filtered splicing events that had at least two replicates per condition with a minimum read coverage of LOW. To calculate differential AS between samples, we used Vast-tools *diff* function. Vast *diff* uses Bayesian inference followed by differential analysis to calculate percent spliced inclusion variations (dPSI) and minimum value difference (MV) between two samples for each AS event. We selected, as significantly differentially spliced, events with an absolute dPSI≥0.15 and a MV≥0.1 or 0.15 (total and polyA+ RNA, respectively) at a 95% confidence interval. Significant events were individually assessed for the quality of the PSI distributions in both of the compared conditions. For visualisation of splicing events, read alignments from three replicates were merged per condition and visualised in the form of sashimi plots, using ggsashimi tool v1.1.5 (Garrido-Martín et al., 2018). We applied a threshold of ten minimum reads to support a junction in sashimi plots. Expression levels for each individual isoform were estimated and normalised in Transcripts per Million (TPM) with Kallisto, as described above.

### Single-cell mRNA library preparation

Two independent replicates of *Srrm2*<sup>+/-</sup> and E14tg2a.4 WT cells were grown for 48 h under regular conditions. Cells were resuspended in 0.04% BSA in PBS and the suspension was loaded on a 40 µm Flowmi<sup>™</sup> Tip

Strainer to remove cell debris and clumps. The cell concentration and viability of the single-cell suspension were determined using a Countess® II Automated Cell Counter (Thermo Fisher Scientific, AMQAX1000), following the manufacturer's instructions. Cell viability ranged between 86-96%. Cells were resuspended into  $10^6$  cells/ml in 0.04% BSA in PBS. Single-cell mRNA-seq libraries were prepared with the Chromium Single Cell 3' v3 chemistry Kit (10x Genomics), according to the manufacturer's instructions. Briefly, a droplet emulsion was generated in a microfluidic chip followed by barcoded cDNA generation inside these droplets. Purified and amplified cDNA was then subjected to library preparation and sequenced on a NovaSeq 6000 instrument (Illumina).

### Single-cell RNA-seq data analyses

#### Mapping, filtering, pre-processing, data integration, clustering and cell cycle assignment

The four 10x Genomics scRNA-seq libraries were simultaneously mapped to the mouse genome (refdata-gex-mm10-2020-A version provided by cellranger) using the *cellranger count* function from 10x Genomics Cellranger software (version cellranger-6.1.2) with default parameters, except `-expect-cells 10,000` (total number of cells targeted per experiment). The four 10x Genomics scRNA-seq libraries were aggregated on one unique count matrix using *cellranger aggr* function from 10x Genomics Cellranger software (version cellranger-6.1.2) with default parameters. The analysis of the single-cell data was performed using Seurat version 4.1.0. To remove low-quality cells, we applied sample-specific thresholds for both number of UMIs and number of features, by subtracting one standard deviation from the mean value across all cells of UMIs or features, respectively. Cells that did not pass both of these two thresholds were removed from the analyses. Additionally, cells having higher than 10% of mitochondrial reads were also excluded.

Single-cell datasets were normalised using the *LogNormalize* method, with a scale factor of 10,000, and the top 2000 variable genes were calculated using the *vst* method. Lastly, the data were scaled using the *ScaleData* function from Seurat, using all genes as input.

PCA analysis was performed by running the *RunPCA* function from Seurat with default parameters, and it was followed by the integration of all samples using Harmony (version 1.0; Korsunsky et al., 2019). The batch-corrected data was subjected to uniform manifold approximation and projection (UMAP) using the first ten dimensions. To explore distinct biologically meaningful groupings, we proceed with clustering by running *FindNeighbors* (first 20 dimensions) and *FindClusters* (0.2 resolution) from the Seurat package. To estimate the cell cycle phase of each cell in our study, we used the *project\_cycle\_space* and *estimate\_cycle\_position* from the tricycle package (version 1.2.1). For simplicity, numerical values were converted to categorical labels following the developer's recommendations: cells with a tricycle position between 0.5pi and pi were classified as 'S' phase, those between pi and 1.75pi as 'G2M' phase, those greater than or equal to 1.75pi or less than 0.25pi as 'G1/G0' phase, and the rest as 'NA'.

#### Differential expression analyses and over-representation analyses

To identify cluster-specific marker genes that may characterise all the different cellular populations, we performed differential gene expression analysis using the *FindAllMarkers* function in Seurat. This analysis was conducted using the Wilcoxon rank-sum test and considering only positively regulated genes detected in at least 10% of cells within each cluster with a minimum fold change of 0.25. We excluded clusters that were composed of less than 2% of the total number of cells in the dataset, specifically clusters 5 and 6. Analyses of the expression of pluripotency markers showed a progressive reduction in its levels along the identified clusters (Fig. 3; Fig. S3). This suggests that the UMAP projection captures a progressive state that underlies a loss of stemness from cluster 1 to cluster 4 and that a trajectory analysis would not provide additional insight into the cellular dynamics of our system. Pairwise differential expression analysis between cluster 2 (mostly composed of *Srrm2*<sup>+/-</sup> cells, making up 95% of the cluster) and cluster 1 (mostly containing WT cells, making up 94% of the cluster) was performed using *FindMarkers* function (Seurat). Genes with  $\log_2$  Fold change >0 represent upregulated genes in cluster 2, whereas genes

with  $\log_2$  Fold change <0 are upregulated in cluster 1. Differentially expressed genes were expressed in at least 50% of the cells in at least one of the two clusters being compared (min.pct=0.5).

To investigate the pluripotency state of the single cells, we assessed the expression of gene markers, as previously described (Wang et al., 2021a). Over-representation analysis for biological terms (<http://geneontology.org/>) and phenotype terms ([https://www.informatics.jax.org/vocab/mp\\_ontology/](https://www.informatics.jax.org/vocab/mp_ontology/)) were performed with WEB-based GENE SeT AnaLysis Toolkit (WebGestalt 2019), using the default parameters and weighted set to reduce redundancy of the gene sets in the enrichment result (Zhang et al., 2005; Wang et al., 2013, 2017; Liao et al., 2019). Gene markers were provided as input against a gene universe of all genes expressed in at least ten of the cells that were included in the analysis. The top five terms with a higher normalised enrichment score (NES) for each cluster were selected to produce the heatmaps shown in Fig. 3F and Fig. S3E.

#### Single-cell labelling transfer

Mice *in vivo* wild-type scRNA-seq data and cell type annotations from E3.5-5.5 and E6.5-8.5 development stages were acquired from GSE123046 (Nowotchin et al., 2019) and GSE137337 (Chan et al., 2019; Grosswendt et al., 2020; Haggerty et al., 2021) repositories, respectively. Data normalisation was conducted using the Seurat package's *NormalizeData* function (v. 4.3.0.1). For the E3.5-5.5 reference dataset, ICM cells were identified and annotated *de novo* based on increased and specific expression of key pluripotency genes (*Pou5f1*, *Sox2*, *Nanog*, *Klf2*, *Klf5*). The most similar *in vivo* cell types were identified for wild-type and *Srrm2*<sup>+/-</sup> cells using the *scmapCluster* function (Scmap v. 1.22.3 using 2000 highly variable features and an assignment threshold of 0.5) of the Scmap cell-cluster label transfer method (Kiselev et al., 2018).

#### Manuscript preparation

R version 4.2.1 (2022-06-23) and Inkscape version 1.2.1 (<https://inkscape.org/>) were used for data visualisation and figure preparation, respectively.

#### Acknowledgements

The authors thank Vedran Franke (Bioinformatics and Omics Data Science Platform, Berlin Institute for Medical Systems Biology, MDC) for advice on single-cell analyses, Manuel Irimia for advice on alternative splicing analysis, the System Biology Imaging facilities from the Berlin Institute for Medical Systems Biology, and the Genomics Technology Platform of Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) and Berlin Institute of Health at Charité–Universitätsmedizin Berlin, for their important contributions to this work, especially to Andrew Woehler, Thomas Conrad and Sarah Nathalie Vitcetz. S.C. thanks colleagues from the Pombo lab and GABBA PhD program for the productive discussions and support.

#### Competing interests

The authors declare no competing or financial interests.

#### Author contributions

Conceptualization: S.C., A.R.G., A.P.; Methodology: S.C., L.Z.-R., T.C.C.T., P.S.-B., D.M., P.C., A.K.; Validation: S.C.; Formal analysis: S.C., L.Z.-R., T.C.C.T., P.S.-B., D.M.; Investigation: S.C., T.C.C.T., P.S.-B., D.M., A.K., S.D., S.G., A.R.G., A.P.; Data curation: L.Z.-R., P.S.-B., P.C., A.K.; Writing - original draft: S.C., P.S.-B., S.G., A.R.G., A.P.; Writing - review & editing: S.C., L.Z.-R., T.C.C.T., P.S.-B., D.M., P.C., A.K., S.D., S.G., A.R.G., A.P.; Visualization: S.C., P.S.-B., D.M.; Supervision: S.C., S.D., S.G., A.R.G., A.P.; Project administration: S.C., A.R.G., A.P.; Funding acquisition: S.C., P.C., S.D., S.G., A.R.G., A.P.

#### Funding

This work was supported by the Helmholtz Association (to A.P., S.G. and S.D.), by the Fundação para a Ciência e Tecnologia (PD/BD/135453/2017 and COVID/BD/152489/2022 to S.C., UIDP/04378/2020, UIDB/04378/2020, LA/P/0140/2020 to A.R.G.), by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) (International Research Training Group IRTG2403 to A.P. and L.Z.-R.), by the DFG under Germany's Excellence Strategy (EXC-2049–390688087 to A.P.), by Deutsches Zentrum für Herz-Kreislaufforschung (DZHK), partner site Berlin (Standortprojekt 81Z0100101 to D.M. and to S.D.), by HORIZON EUROPE Marie Skłodowska-Curie Actions (FOX-MTN-HORIZON-MSCA – 2021-PF-01-01 to P.C.). Open Access funding provided by Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft. Deposited in PMC for immediate release.

## Data availability

Datasets produced in this study can be found in the GEO repository under the accession number [GSE243433](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243433), which includes raw fastq files and raw count matrices for the RNA-seq bulk datasets, and fastq files and matrix processed files for single-cell datasets. This study used the genome assembly mm10 release M24 (GRCh38.p6). The publicly available datasets used in this study are listed in the Materials and Methods section. Supporting data can be found in the [supplementary information](#).

## References

- Akagi, T., Kuure, S., Uranishi, K., Koide, H., Costantini, F. and Yokota, T. (2015). ETS-related transcription factors ETV4 and ETV5 are involved in proliferation and induction of differentiation-associated genes in embryonic stem (ES) cells. *J. Biol. Chem.* **290**, 22460-22473. doi:10.1074/jbc.M115.675595
- Athanasouli, P., Balli, M., De Jaime-Soguero, A., Boel, A., Papanikolaou, S., van der Veer, B. K., Janiszewski, A., Vanhessche, T., Francis, A., El Laithy, Y. et al. (2023). The Wnt/TCF7L1 transcriptional repressor axis drives primitive endoderm formation by antagonizing naive and formative pluripotency. *Nat. Commun.* **14**, 1210. doi:10.1038/s41467-023-36914-1
- Baralle, F. E. and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437-451. doi:10.1038/nrm.2017.27
- Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., Barbieri, M., de Santiago, I., Lavitas, L.-M., Branco, M. R. et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524. doi:10.1038/nature21411
- Betto, R. M., Diamante, L., Perrera, V., Audano, M., Rapelli, S., Lauria, A., Incarnato, D., Arboit, M., Pedretti, S., Rigoni, G. et al. (2021). Metabolic control of DNA methylation in naive pluripotent cells. *Nat. Genet.* **53**, 215-229. doi:10.1038/s41588-020-00770-2
- Blencowe, B. J., Issner, R., Nickerson, J. A. and Sharp, P. A. (1998). A coactivator of pre-mRNA splicing. *Genes Dev.* **12**, 996-1009. doi:10.1101/gad.12.7.996
- Blencowe, B. J., Bowman, J. A., McCracken, S. and Rosonina, E. (1999). SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.* **77**, 277-291. doi:10.1139/o99-048
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M. and Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcripts. *Genome Res.* **24**, 1774-1786. doi:10.1101/gr.177790.114
- Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527. doi:10.1038/nbt.3519
- Bushdid, P. B., Osinska, H., Waclaw, R. R., Molckentin, J. D. and Yutzey, K. E. (2003). NFATc3 and NFATc4 are required for cardiac development and mitochondrial function. *Circ. Res.* **92**, 1305-1313. doi:10.1161/01.RES.0000077045.84609.9F
- Calarco, J. A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinias, D., van der Kooy, D. et al. (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* **138**, 898-910. doi:10.1016/j.cell.2009.06.012
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J. et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502. doi:10.1038/s41586-019-0969-x
- Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascensão, K., Rummel, C., Ovchinnikova, S. et al. (2019). Gene expression across mammalian organ development. *Nature* **571**, 505-509. doi:10.1038/s41586-019-1338-5
- Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., Jost, M., Quinn, J. J., Yang, D., Jones, M. G. et al. (2019). Molecular recording of mammalian embryogenesis. *Nature* **570**, 77-82. doi:10.1038/s41586-019-1184-5
- Cheng, X., Li, L., Shi, G., Chen, L., Fang, C., Li, M. and Li, C. (2020). MEG3 promotes differentiation of porcine satellite cells by sponging miR-423-5p to relieve inhibiting effect on SRF. *Cells* **9**, 449. doi:10.3390/cells9020449
- Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G. et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613-619. doi:10.1038/nmeth.1223
- Cui, L., Johkura, K., Yue, F., Ogiwara, N., Okouchi, Y., Asanuma, K. and Sasaki, K. (2004). Spatial distribution and initial changes of SSEA-1 and other cell adhesion-related molecules on mouse embryonic stem cells before and during differentiation. *J. Histochem. Cytochem.* **52**, 1447-1457. doi:10.1369/jhc.3A6241.2004
- Cui, H., Diedrich, J. K., Wu, D. C., Lim, J. J., Nottingham, R. M., Moresco, J. J., Yates, J. R., Blencowe, I. I., Lambowitz, B. J. and Schimmel, A. M. (2023). Arg-tRNA synthetase links inflammatory metabolism to RNA splicing and nuclear trafficking via SRRM2. *Nat. Cell Biol.* **25**, 592-603. doi:10.1038/s41556-023-01118-8
- Cuinat, S., Nizon, M., Isidor, B., Stegmann, A., van Jaarsveld, R. H., van Gassen, K. L., van der Smagt, J. J., Volker-Touw, C. M. L., Holwerda, S. J. B., Terhal, P. A. et al. (2022). Loss-of-function variants in SRRM2 cause a neurodevelopmental disorder. *Genet. Med.* **24**, 1774-1780. doi:10.1016/j.gim.2022.04.011
- David, B. G., Fujita, H., Yasuda, K., Okamoto, K., Panina, Y., Ichinose, J., Sato, O., Horie, M., Ichimura, T., Okada, Y. et al. (2019). Linking substrate and nucleus via actin cytoskeleton in pluripotency maintenance of mouse embryonic stem cells. *Stem. Cell Res.* **41**, 101614. doi:10.1016/j.scr.2019.101614
- Deshpande, A., Shetty, P. M. V., Frey, N. and Rangrez, A. Y. (2022). SRF: a seriously responsive factor in cardiac development and disease. *J. Biomed. Sci.* **29**, 38. doi:10.1186/s12929-022-00820-3
- Ding, X., Hu, H., Huang, K., Wei, R., Min, J., Qi, C., Tang, H. and Qin, X. (2020). Ubiquitination of NOTCH2 by DTX3 suppresses the proliferation and migration of human esophageal carcinoma. *Cancer Sci.* **111**, 489-501. doi:10.1111/cas.14288
- Eldridge, A. G., Li, Y., Sharp, P. A. and Blencowe, B. J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc. Natl Acad. Sci. USA* **96**, 6125-6130. doi:10.1073/pnas.96.11.6125
- Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156. doi:10.1038/292154a0
- Ferrai, C., Torlai Triglia, E., Risner-Janiczek, J. R., Rito, T., Rackham, O. J., de Santiago, I., Kukalev, A., Nicodemi, M., Akalin, A., Li, M. et al. (2017). RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation. *Mol. Syst. Biol.* **13**, 946. doi:10.15252/msb.20177754
- Fontrodona, L., Porta-de-la-Riva, M., Morán, T., Niu, W., Díaz, M., Aristizábal-Corrales, D., Villanueva, A., Schwartz, S., Reinke, V. and Cerón, J. (2013). RSR-2, the caenorhabditis elegans ortholog of human spliceosomal component SRm300/SRRM2, regulates development by influencing the transcriptional machinery. *PLoS Genet.* **9**, e1003543. doi:10.1371/journal.pgen.1003543
- Fullerton, M. D., Hakimuddin, F. and Bakovic, M. (2007). Developmental and metabolic effects of disruption of the mouse CTP:Phosphoethanolamine Cytidyltransferase Gene (Pcvt2). *Mol. Cell. Biol.* **27**, 3327-3336. doi:10.1128/MCB.01527-06
- Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E. O. et al. (2011). An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming. *Cell* **147**, 132-146. doi:10.1016/j.cell.2011.08.023
- Garrido-Martín, D., Palumbo, E., Guigó, R. and Breschi, A. (2018). ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLoS Comput. Biol.* **14**, e1006360. doi:10.1371/journal.pcbi.1006360
- Graef, I. A., Chen, F., Chen, L., Kuo, A. and Crabtree, G. R. (2001). Signals transduced by Ca2+/calineurin and NFATc3/c4 pattern the developing vasculature. *Cell* **105**, 863-875. doi:10.1016/S0092-8674(01)00396-8
- Grosswendt, S., Kretzmer, H., Smith, Z. D., Kumar, A. S., Hetzel, S., Witterl, L., Klages, S., Timmermann, B., Mukherji, S. and Meissner, A. (2020). Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102-108. doi:10.1038/s41586-020-2552-x
- Groza, T., Gomez, F. L., Mashhadi, H. H., Muñoz-Fuentes, V., Gunes, O., Wilson, R., Cacheiro, P., Frost, A., Keski-Valko, P., Vardal, B. et al. (2022). The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038-D1045. doi:10.1093/nar/gkac972
- Guo, G., Huang, Y., Humphreys, P., Wang, X. and Smith, A. (2011). A piggyback-based recessive screening method to identify pluripotency regulators. *PLoS ONE* **6**, e18189. doi:10.1371/journal.pone.0018189
- Guo, L., Yang, L., Fan, C., Chen, G. and Wu, F. (2012). Novel roles of Vmp1: Inhibition metastasis and proliferation of hepatocellular carcinoma. *Cancer Sci.* **103**, 2110-2119. doi:10.1111/cas.12025
- Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A. S., Mattei, A. L., Bailly, N., Gottfreund, J., Giesselmann, P., Weigert, R., Brändl, B. et al. (2021). Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* **28**, 594-603. doi:10.1038/s41594-021-00603-8
- Han, H., Irimia, M., Ross, P. J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I. P., Nachman, E. N. et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241-245. doi:10.1038/nature12270
- Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R. J., Hirsch, C. L., Ha, K. C. H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J. et al. (2017). Multilayered control of alternative splicing regulatory networks by transcription factors. *Mol. Cell* **65**, 539-553.e7. doi:10.1016/j.molcel.2017.01.011
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519-532. doi:10.1016/j.cell.2011.06.052

- Heo, S.-J., Cosgrove, B. D., Dai, E. N. and Mauck, R. L. (2017). Mechano-adaptation of the stem cell nucleus. *Nucleus* **9**, 9-19. doi:10.1080/19491034.2017.1371398
- Hoffman, J. A., Wu, C.-I. and Merrill, B. J. (2013). Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development* **140**, 1665-1675. doi:10.1242/dev.087387
- Holzner, M., Sonicki, T., Hunn, H., Gade, V. R., Weis, K., Wutz, A. and Minin, G. D. (2023). The scramblases VMP1 and TMEM41b are required for primitive endoderm specification by targeting WNT signaling. *bioRxiv*. doi:10.1101/2023.10.31.564914
- Hoogland, S. H. A. and Marks, H. (2021). Developments in pluripotency: a new formative state. *Cell Res.* **31**, 493-494. doi:10.1038/s41422-021-00494-w
- Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q. and Guo, A.-Y. (2018). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33-D38. doi:10.1093/nar/gky822
- Hu, S., Lv, P., Yan, Z. and Wen, B. (2019). Disruption of nuclear speckles reduces chromatin interactions in active compartments. *Epigenetics Chromatin* **12**, 43. doi:10.1186/s13072-019-0289-2
- Huang, G., Ye, S., Zhou, X., Liu, D. and Ying, Q.-L. (2015). Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network. *Cell. Mol. Life Sci.* **72**, 1741-1757. doi:10.1007/s00018-015-1833-2
- Ilik, I. A., Malszycki, M., Lübke, A. K., Schade, C., Meierhofer, D. and Aktas, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. *eLife* **9**, e60579. doi:10.7554/eLife.60579
- Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D. et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511-1523. doi:10.1016/j.cell.2014.11.035
- Jia, D., Gao, P., Lv, Y., Huang, Y., Reilly, J., Sun, K., Han, Y., Hu, H., Chen, X., Zhang, Z. et al. (2022). Tulp1 deficiency causes early-onset retinal degeneration through affecting ciliogenesis and activating ferroptosis in zebrafish. *Cell Death and Disease* **13**, 962. doi:10.1038/s41419-022-05372-w
- Kalkan, T. and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130540. doi:10.1098/rstb.2013.0540
- Kalkan, T., Bornelöv, S., Mulas, C., Diamanti, E., Lohoff, T., Ralsler, M., Middelkamp, S., Lombard, P., Nichols, J. and Smith, A. (2019). Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. *Cell Stem Cell* **24**, 785-801.e7. doi:10.1016/j.stem.2019.03.017
- Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F. et al. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757-762. doi:10.1038/s41586-020-2832-5
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P. et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443. doi:10.1038/s41586-020-2308-7
- Kim, M. and Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. *Exp. Mol. Med.* **49**, e322-e322. doi:10.1038/emm.2017.10
- Kinoshita, M., Li, M. A., Barber, M., Mansfield, W., Dietmann, S. and Smith, A. (2021). Disabling de novo DNA methylation in embryonic stem cells allows an illegitimate fate trajectory. *Proc. Natl Acad. Sci. USA* **118**, e2109475118. doi:10.1073/pnas.2109475118
- Kiselev, V. Y., Yiu, A. and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* **15**, 359-362. doi:10.1038/nmeth.4644
- Kolmykov, S., Yevshin, I., Kulyashov, M., Sharipov, R., Kondrakhin, Y., Makeev, V. J., Kulakovskiy, I. V., Kel, A. and Kolpakov, F. (2020). GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* **49**, D104-D111. doi:10.1093/nar/gkaa1057
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289-1296. doi:10.1038/s41592-019-0619-0
- Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T. et al. (2015). Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* **16**, 517-532. doi:10.1016/j.stem.2015.03.002
- Lackner, A., Sehlke, R., Garmhausen, M., Giuseppe Stirparo, G., Huth, M., Titz-Teixeira, F., van der Lelij, P., Ramesmayer, J., Thomas, H. F., Ralsler, M. et al. (2021). Cooperative genetic networks drive embryonic stem cell transition from naive to formative pluripotency. *EMBO J.* **40**, e105776. doi:10.15252/embj.202105776
- Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A. et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311-316. doi:10.1038/nsmb.2510
- Li, D. and Wang, J. (2020). Ribosome heterogeneity in stem cells and development. *J. Cell Biol.* **219**, e202001108. doi:10.1083/jcb.202001108
- Li, M. A., Amaral, P. P., Cheung, P., Bergmann, J. H., Kinoshita, M., Kalkan, T., Ralsler, M., Robson, S., von Meyenn, F., Paramor, M. et al. (2017). A lncRNA fine tunes the dynamics of a cell state transition involving Lin28, let-7 and de novo DNA methylation. *eLife* **6**, e23468. doi:10.7554/eLife.23468
- Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199-W205. doi:10.1093/nar/gkz401
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P. and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell Systems* **1**, 417-425. doi:10.1016/j.cels.2015.12.004
- Linker, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D. et al. (2010). Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777-782. doi:10.1083/jcb.201004104
- Liu, H., Lorenzini, P. A., Zhang, F., Xu, S., Wong, M. S. M., Zheng, J. and Roca, X. (2018). Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator. *Nucleic Acids Res.* **46**, 6069-6086. doi:10.1093/nar/gky401
- Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔCT</sup> method. *Methods* **25**, 402-408. doi:10.1006/meth.2001.1262
- Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634-7638. doi:10.1073/pnas.78.12.7634
- Mayshar, Y., Rom, E., Chumakov, I., Kronman, A., Yayon, A. and Benvenisty, N. (2008). Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. *Stem Cells* **26**, 767-774. doi:10.1634/stemcells.2007-1037
- Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A. T. L., Marioni, J. C. and Reik, W. (2019). Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. *Cell Reports* **26**, 815-824.e4. doi:10.1016/j.celrep.2018.12.099
- Miano, J. M., Long, X. and Fujiwara, K. (2007). Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. *Am. J. Physiol. Cell Physiol.* **292**, C70-C81. doi:10.1152/ajpcell.00386.2006
- Morishita, H., Zhao, Y. G., Tamura, N., Nishimura, T., Kanda, Y., Sakamaki, Y., Okazaki, M., Li, D. and Mizushima, N. (2019). A critical role of VMP1 in lipoprotein secretion. *eLife* **8**, e48834. doi:10.7554/eLife.48834
- Nakano, Y., Wiechert, S. and Bánfi, B. (2019). Overlapping activities of two neuronal splicing factors switch the GABA effect from excitatory to inhibitory by regulating REST. *Cell Rep.* **27**, 860-871.e8. doi:10.1016/j.celrep.2019.03.072
- Nakano, Y., Jahan, I., Bonde, G., Sun, X., Hildebrand, M. S., Engelhardt, J. F., Smith, R. J. H., Cornell, R. A., Fritzsche, B. and Bánfi, B. (2012). A mutation in the *srnm4* gene causes alternative splicing defects and deafness in the bronx waltzer mouse. *PLoS Genet.* **8**, e1002966. doi:10.1371/journal.pgen.1002966
- Niwa, H., Burdon, T., Chambers, I. and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12**, 2048-2060. doi:10.1101/gad.12.13.2048
- Nowotzsch, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C. et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367. doi:10.1038/s41586-019-1127-1
- Olson, E. N. and Nordheim, A. (2010). Linking actin dynamics and gene transcription to drive cellular motile functions. *Nat. Rev. Mol. Cell Biol.* **11**, 353-365. doi:10.1038/nrm2890
- Pastor-Arroyo, E. M., Rodriguez, J. M. M., Pellegrini, G., Bettoni, C., Levi, M., Hernandez, N. and Wagner, C. A. (2021). Constitutive depletion of Slc34a2/NaPi-11b in rats causes perinatal mortality. *Sci. Rep.* **11**, 7943. doi:10.1038/s41598-021-86874-z
- Pereira, L., Yi, F. and Merrill, B. J. (2006). Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol. Cell Biol.* **26**, 7479-7491. doi:10.1128/MCB.00368-06
- Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. and Goldstein, D. B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709. doi:10.1371/journal.pgen.1003709
- Posern, G. and Treisman, R. (2006). Actin' together: serum response factor, its cofactors and the link to signal transduction. *Trends Cell Biol.* **16**, 588-596. doi:10.1016/j.tcb.2006.09.008
- Putra, V. D. L., Kilian, K. A. and Knothe Tate, M. L. (2023). Biomechanical, biophysical and biochemical modulators of cytoskeletal remodelling and emergent stem cell lineage commitment. *Commun. Biol.* **6**, 75. doi:10.1038/s42003-022-04320-w
- Quesnel-Vallières, M., Irimia, M., Cordes, S. P. and Blencowe, B. J. (2015). Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* **29**, 746-759. doi:10.1101/gad.256115.114
- Quesnel-Vallières, M., Dargaoui, Z., Irimia, M., Gonatopoulos-Pournatzis, T., Ip, J. Y., Wu, M., Sterne-Weiler, T., Nakagawa, S., Woodin, M. A., Blencowe, B. J. et al. (2016). Misregulation of an activity-dependent splicing network as a

- common mechanism underlying autism spectrum disorders. *Mol. Cell* **64**, 1023-1034. doi:10.1016/j.molcel.2016.11.033
- Rackham, O. J. L., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W. et al. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* **48**, 331-335. doi:10.1038/ng.3487
- Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C. et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482-517. doi:10.1038/s41596-018-0103-9
- Revil, T., Gaffney, D., Dias, C., Majewski, J. and Jerome-Majewska, L. A. (2010). Alternative splicing is frequent during early embryonic development in mouse. *BMC Genomics* **11**, 399. doi:10.1186/1471-2164-11-399
- Ropolo, A., Grasso, D., Pardo, R., Sacchetti, M. L., Archange, C., Re, A. L., Seux, M., Nowak, J., Gonzalez, C. D., Iovanna, J. L. et al. (2007). The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *J. Biol. Chem.* **282**, 37124-37133. doi:10.1074/jbc.M706956200
- Rossant, J. and Tam, P. P. L. (2017). New insights into early human development: lessons for stem cell derivation and differentiation. *Cell Stem Cell* **20**, 18-28. doi:10.1016/j.stem.2016.12.004
- Saha, S., Bose, R., Chakraborty, S. and Ain, R. (2022). Tipping the balance toward stemness in trophoblast: Metabolic programming by Cox6B2. *FASEB J.* **36**, e22600. doi:10.1096/fj.202200703RR
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682. doi:10.1038/nmeth.2019
- Shehadeh, L. A., Yu, K., Wang, L., Guevara, A., Singer, C., Vance, J. and Papapetropoulos, S. (2010). SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PLoS ONE* **5**, e9104. doi:10.1371/journal.pone.0009104
- Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365-373. doi:10.1242/dev.142679
- Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690. doi:10.1038/336688a0
- Štefková, K., Procházková, J. and Pacherník, J. (2015). Alkaline phosphatase in stem cells. *Stem Cells Int.* **2015**, 628368. doi:10.1155/2015/628368
- Tapial, J., Ha, K. C. H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y. et al. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* **27**, 1759-1768. doi:10.1101/gr.220962.117
- Tie, A. and Bakovic, M. (2007). Alternative splicing of CTP:phosphoethanolamine cytidyltransferase produces two isoforms that differ in catalytic properties. *J. Lipid Res.* **48**, 2172-2181. doi:10.1194/jlr.M600536-JLR200
- Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147. doi:10.1126/science.282.5391.1145
- Tomsic, J., He, H., Akagi, K., Liyanarachchi, S., Pan, Q., Bertani, B., Nagy, R., Symer, D. E., Blencowe, B. J. and de la Chapelle, A. (2015). A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Sci. Rep.* **5**, 10566. doi:10.1038/srep10566
- Torres-Méndez, A., Bonnal, S., Marquez, Y., Roth, J., Iglesias, M., Permanyer, J., Almudí, I., O'Hanlon, D., Guitart, T., Soller, M. et al. (2019). A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat. Ecol. Evol.* **3**, 691-701. doi:10.1038/s41559-019-0813-6
- Tsai, S. C., Chang, D. F., Hong, C.-M., Xia, P., Senadheera, D., Trump, L., Mishra, S. and Lutzko, C. (2014). Induced overexpression of OCT4A in human embryonic stem cells increases cloning efficiency. *Am. J. Physiol. Cell Physiol.* **306**, C1108-C1118. doi:10.1152/ajpcell.00205.2013
- van Dijk, D., Sharma, R., Nainys, J., Yin, K., Kathail, P., Carr, A. J., Burdzyak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D. et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716-729.e27. doi:10.1016/j.cell.2018.05.061
- Vanhoutteghem, A., Maciejewski-Duval, A., Bouche, C., Delhomme, B., Hervé, F., Daubigney, F., Soubigou, G., Araki, M., Araki, K., Yamamura, K. et al. (2009). Basonuclin 2 has a function in the multiplication of embryonic craniofacial mesenchymal cells and is orthologous to disco proteins. *Proc. Natl Acad. Sci. USA* **106**, 14432-14437. doi:10.1073/pnas.0905840106
- Venables, J. P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B. et al. (2013). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat. Commun.* **4**, 2480. doi:10.1038/ncomms3480
- Verdoni, A. M., Ikeda, S. and Ikeda, A. (2010). Serum response factor is essential for the proper development of skin epithelium. *Mamm. Genome* **21**, 64-76. doi:10.1007/s00355-009-9245-y
- Wang, J., Duncan, D., Shi, Z. and Zhang, B. (2013). WEB-based GENE SeT analysis toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77-W83. doi:10.1093/nar/gkt439
- Wang, J., Vasaike, S., Shi, Z., Greer, M. and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130-W137. doi:10.1093/nar/gkx356
- Wang, X., Xiang, Y., Yu, Y., Wang, R., Zhang, Y., Xu, Q., Sun, H., Zhao, Z.-A., Jiang, X., Wang, X. et al. (2021a). Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell Res.* **31**, 52. doi:10.1038/s41422-020-00410-8
- Wang, L., Sun, X., He, J. and Liu, Z. (2021b). Functions and molecular mechanisms of deltex family ubiquitin E3 ligases in development and disease. *Front. Cell Dev. Biol.* **9**, 6-541.
- Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687. doi:10.1038/336684a0
- Wu, J. Q., Habegger, L., Noisa, P., Szekely, A., Qiu, C., Hutchison, S., Raha, D., Egholm, M., Lin, H., Weissman, S. et al. (2010). Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc. Natl Acad. Sci. USA* **107**, 5254-5259. doi:10.1073/pnas.0914114107
- Xiong, X., Li, W., Nam, J., Qu, M., Kay, S. A. and Ma, K. (2022). The actin cytoskeleton-MRTF/SRF cascade transduces cellular physical niche cues to entrain the circadian clock. *J. Cell Sci.* **135**, jcs260094. doi:10.1242/jcs.260094
- Xu, S., Lai, S.-K., Sim, D. Y., Ang, W. S. L., Li, H. Y. and Roca, X. (2022). SRRM2 organizes splicing condensates to regulate alternative splicing. *Nucleic Acids Res.* **50**, 8599-8614. doi:10.1093/nar/gkac669
- Yang, P., Humphrey, S. J., Cinghu, S., Pathania, R., Oldfield, A. J., Kumar, D., Perera, D., Yang, J. Y. H., James, D. E., Mann, M. et al. (2019). Multi-omic profiling reveals dynamics of the phased progression of pluripotency. *Cell Syst.* **8**, 427-445.e10. doi:10.1016/j.cels.2019.03.012
- Zhang, B., Kirov, S. and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741-W748. doi:10.1093/nar/gki475
- Zhu, L., Michel, V. and Bakovic, M. (2009). Regulation of the mouse CTP: Phosphoethanolamine cytidyltransferase gene Pcyt2 during myogenesis. *Gene* **447**, 51-59. doi:10.1016/j.gene.2009.07.014