ARTICLES
https://doi.org/10.1038/s42255-021-00464-x

nature metabolism

Check for updates

Metabolic remodelling during early mouse embryo development

Jing Zhao[1,11], Ke Yao[2,11], Hua Yu[1,11], Ling Zhang[1,3,11], Yuyan Xu[1,11], Lang Chen[1], Zhen Sun[1], Yuqing Zhu[1], Cheng Zhang[3], Yuli Qian[4], Shuyan Ji[5], Hongru Pan[1], Min Zhang[1], Jie Chen[1], Cristina Correia[6], Taylor Weiskittel[6], Da-Wei Lin[7], Yuzheng Zhao[8], Sriram Chandrasekaran[7], Xudong Fu[1,3], Dan Zhang[4], Heng-Yu Fan[9], Wei Xie[5], Hu Li[6], Zeping Hu[2]✉ and Jin Zhang[1,3,10]✉

During early mammalian embryogenesis, changes in cell growth and proliferation depend on strict genetic and metabolic instructions. However, our understanding of metabolic reprogramming and its influence on epigenetic regulation in early embryo development remains elusive. Here we show a comprehensive metabolomics profiling of key stages in mouse early development and the two-cell and blastocyst embryos, and we reconstructed the metabolic landscape through the transition from totipotency to pluripotency. Our integrated metabolomics and transcriptomics analysis shows that while two-cell embryos favour methionine, polyamine and glutathione metabolism and stay in a more reductive state, blastocyst embryos have higher metabolites related to the mitochondrial tricarboxylic acid cycle, and present a more oxidative state. Moreover, we identify a reciprocal relationship between α-ketoglutarate (α-KG) and the competitive inhibitor of α-KG-dependent dioxygenases, L-2-hydroxyglutarate (L-2-HG), where two-cell embryos inherited from oocytes and one-cell zygotes display higher L-2-HG, whereas blastocysts show higher α-KG. Lastly, increasing 2-HG availability impedes erasure of global histone methylation markers after fertilization. Together, our data demonstrate dynamic and interconnected metabolic, transcriptional and epigenetic network remodelling during early mouse embryo development.

Metabolic reprogramming and its regulation during mammalian early embryo development are hardwired into the complex programme of development[1]. Limitations in obtaining sufficient sample amounts and availability of reliable and scalable methods have hindered a deeper understanding of dynamic changes of metabolites, metabolic pathways and their upstream regulating factors during embryogenesis. Classical analysis on whole-embryo physiology has focused on nutrition uptake, respiratory activity and by-product secretion[1]. For instance, embryos utilize pyruvate earlier and later transition to glucose, and the knowledge led to the development of optimal media conditions currently used during in vitro fertilization clinical practice[2–7]. Moreover, blastocysts exhibit a respiratory burst during development[8,9], but the underlying metabolic remodelling is still elusive. Recent omics analysis using low-input or single-cell RNA-sequencing (RNA-seq) and chromatin analysis has provided insights into molecular regulatory circuitry governing early embryo development[10–15]. But metabolic genes are often labelled as 'housekeeping' genes and metabolic alterations have been broadly neglected. Thus, a comprehensive picture of metabolic gene expression, their epigenetic status and metabolite dynamics in early mouse embryo development has yet to be uncovered.

Our limited knowledge of the metabolism molecular wiring in embryos is primarily built on studies with embryonic stem (ES) cells, and their various pluripotent states representing different embryonic-stage counterparts[16–18], and studies with these cultured cells have provided key insights into their catabolic and anabolic metabolism[19–21], and how metabolic networks are integrated into the well-established pluripotency networks such as *Esrrb*, *Stat3*, *Lin28a* and *Zic3* in energy metabolism[22–24], or *c-Myc* and *mTor* in anabolic metabolism[25,26]. However, there is a need to confirm the physiological importance of these findings from cultured ES cells in the in vivo counterparts of embryo systems.

Epigenetic remodelling, especially histone methylation, is highly dynamic during pre-implantation embryogenesis[13–15,27–30]. How metabolites are involved in this process is largely unknown. Our findings provide insights into the role of reciprocal changes of a pair of competitive metabolites, α-KG and 2-HG, in the dynamic histone methylation erasure during early embryogenesis.

Results

Metabolic reprogramming from two-cell to blastocyst embryos. In order to understand the dynamic metabolic remodelling

[1]Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Zhejiang University, Hangzhou, China. [2]School of Pharmaceutical Sciences, Tsinghua-Peking Center for Life Sciences, Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing, China. [3]Zhejiang Laboratory for Systems & Precision Medicine, Zhejiang University Medical Center, Hangzhou, China. [4]Key Laboratory of Reproductive Genetics (Ministry of Education) and Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China. [5]Center for Stem Cell Biology and Regenerative Medicine, MOE Key Laboratory of Bioinformatics, THU-PKU Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, China. [6]Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, NY, USA. [7]Center of Computational Medicine and Bioinformatics, Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA. [8]Optogenetics & Synthetic Biology Interdisciplinary Research Center, State Key Laboratory of Bioreactor Engineering, Shanghai Collaborative Innovation Center for Biomanufacturing Technology, Research Unit of Chinese Academy of Medical Sciences, East China University of Science and Technology, Shanghai, China. [9]Life Sciences Institute, Zhejiang University, Hangzhou, China. [10]Institute of Hematology, Zhejiang University, Hangzhou, China. [11]These authors contributed equally: Jing Zhao, Ke Yao, Hua Yu, Ling Zhang, Yuyan Xu. ✉e-mail: zeping_hu@tsinghua.edu.cn; zhgene@zju.edu.cn

during pre-implantation embryo development, we used mass spectrometry (MS)-based metabolomics to directly measure metabolite abundance in embryos. Due to difficulties in obtaining a large number of embryos and the limited detection range of current metabolomics technologies, firstly, we optimized a targeted metabolomics approach to detect metabolites using a small number of cells. We titrated a range of mouse ES cells from 2,500 to 80,000 cells (Extended Data Fig. 1a), or a range of zygotes from 15 to 240 embryos (Extended Data Fig. 1b), and we were able to achieve strong correlations between the number of cells/embryos and the MS signal for most of the metabolites even at lower input ranges of the gradients (Extended Data Fig. 1c–e), indicating good quantification of those metabolites with this approach. Then, we collected 100 two-cell-stage embryos and 100 blastocyst-stage (BC) embryos, which represent the totipotent state when zygotic genome activation takes place and the pluripotent state when ES cells can be derived[31], and applied the targeted metabolomics method established above, each with three biological replicates (Fig. 1a). Principal-component analysis (PCA) on targeted metabolite levels showed that two-cell embryos were readily separated from blastocysts (Fig. 1b). Top differential metabolites include citrate, α-KG, succinate, glutamine and malate, which are higher in the blastocyst stage, and 2-HG, *S*-adenosyl-methionine, glutathione (GSH), oxidized glutathione (GSSG) and spermidine, which are higher in the two-cell embryos (Fig. 1c and Supplementary Data 1). Notably, almost all the tricarboxylic acid (TCA) cycle intermediate metabolites were more abundant in blastocyst embryos, including citrate, succinate and α-KG ($P<0.05$), but the abundance of the competitive inhibitor of α-KG-dependent dioxygenases[32], 2-HG, was higher in two-cell embryos ($P<0.001$; Fig. 1d,e). Indeed, metabolite-set enrichment analysis demonstrated that 'TCA' was among the top enriched metabolite sets in blastocyst embryos, whereas 'methionine metabolism', 'spermidine and spermine biosynthesis' and 'nicotinate and nicotinamide metabolism' were among the top enriched metabolite sets in two-cell embryos (Fig. 1f,g). Detailed analysis of the above metabolic pathways including the detected metabolites and the corresponding metabolic genes also revealed the dynamic metabolic signatures of the TCA cycle and purine metabolism pathways in the blastocyst embryos (Extended Data Figs. 2 and 3), and one-carbon metabolism as well as polyamine/GSH/nicotinamide pathways related to the redox state in two-cell embryos (Extended Data Fig. 4). For instance, a higher level of the TCA cycle pathway metabolites in blastocysts is consistent with higher expression of the TCA cycle genes *Aco1*, *Idh2*, *Sucla2*, *Sdha*, *Fh1* and *Mdh2* (Supplementary Fig. 2), and higher levels of GSH, GSSG, GSH/GSSH ratio, spermidine and nicotinamide are associated with higher expression of the redox-related genes such as *G6pd* (Extended Data Fig. 4).

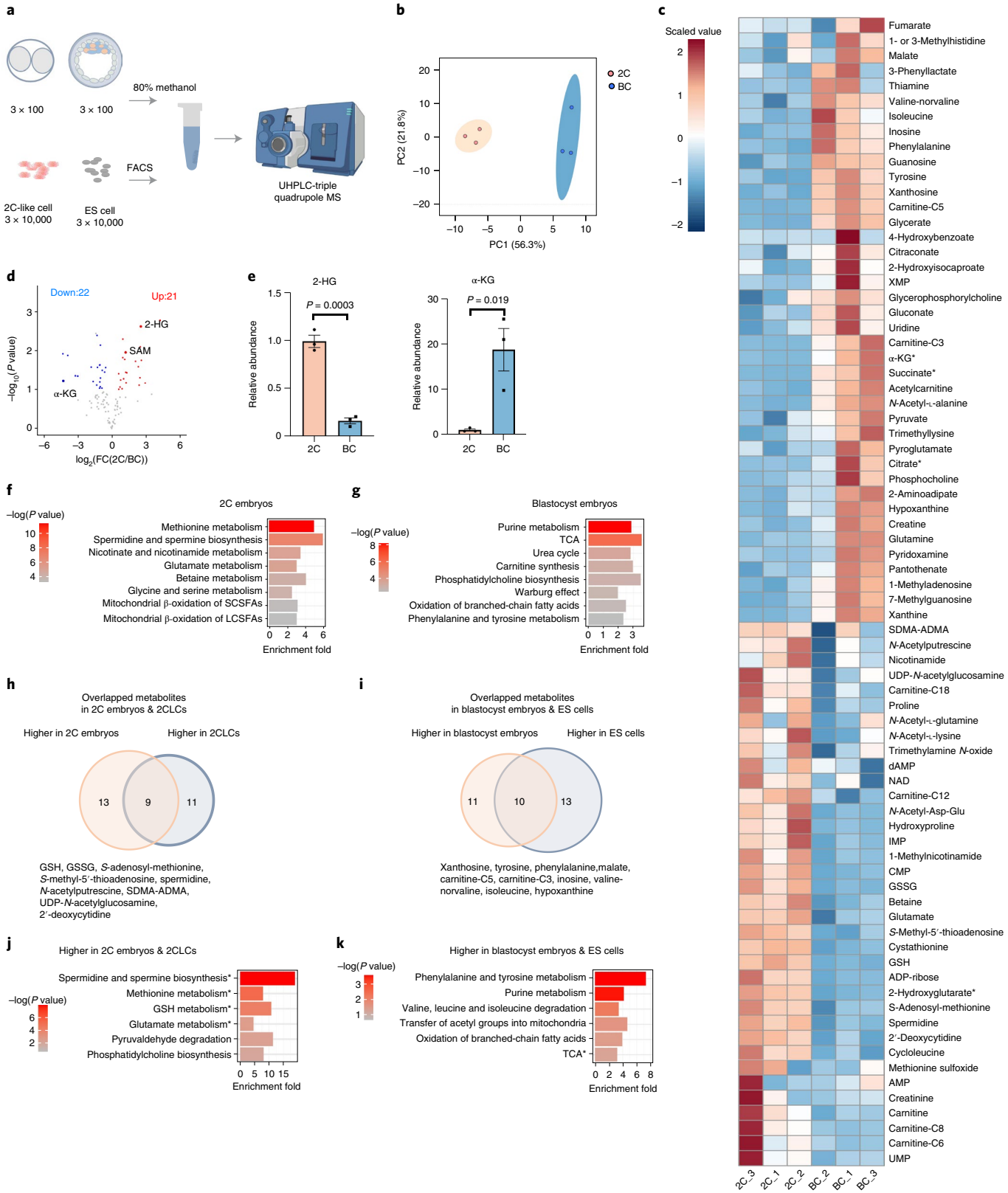
To further analyse the metabolomics data in a systemic way, metabolites were mapped to a metabolic network based on our previously published method in genome-scale metabolic network modelling[21] (Extended Data Fig. 5a), and differential genetic deletion sensitivity was predicted (Extended Data Fig. 5b). Consistently, two-cell embryos were more sensitive to deletion of genes in redox-related 'GSH metabolism', while blastocyst embryos were more sensitive to deletion of genes in 'oxidative phosphorylation' (Supplementary Table 1). In summary, our embryo metabolomics analysis revealed unique metabolic characteristics of two-cell and blastocyst embryos.

It has been reported that the in vitro cultured ES cells have a spontaneously emerged two-cell-like cell (2CLC) population that resembles two-cell embryos in their transcriptomic signature[16]. To examine whether the metabolomes between embryos in vivo and their cultured counterparts in vitro are also similar, we applied the same metabolomics approach to ES cells and 2CLCs. After transduction with a 2C reporter, *2C::tdTomato*, the 2CLCs can be isolated from ES cell culture[16]. The same ES cell culture was also transduced with a *Pou5f1::GFP* reporter to indicate the pluripotent state ES cells (Extended Data Fig. 1f). We thus sorted 10,000 tdTomato-positive 2CLCs or GFP-positive ES cells each with three biological replicates for metabolomics analysis. In general, PCA of common metabolites revealed that cultured cells clustered more closely to each other (Extended Data Fig. 5c), indicating that the in vitro culturing condition has a strong impact on cellular metabolome. To lend further support, we performed metabolomics with another 2CLC system, a 2C gene *Dux* inducible ES cell line[33], to enrich the tdTomato-positive 2CLCs. The ratio of 2CLCs reached 27.2% after *Dux* induction (Extended Data Fig. 5d,e), and the 2C genes were strongly activated (Extended Data Fig. 5f). We observed similar patterns of the PCA as those of sorted 2CLC metabolomics from wild-type cells: globally, two-cell and BC embryos clustered away from 2CLC and ES cells as they represented in vivo and in vitro samples (Extended Data Fig. 5g). Further comparing these 2CLCs and ES cells in the heat map showed GSH, GSSG, GSH/GSSG and spermidine levels were higher in 2CLCs (Extended Data Fig. 5h–j), consistent with the analysis from embryos (Extended Data Figs. 1g and 4d,f,g). For better visualization, we overlapped metabolites higher in two-cell embryos (compared to BC embryos) and higher in 2CLCs (compared to ES cells; Fig. 1h,i), and they were specifically enriched in spermidine, methionine and GSH (Fig. 1j,k). Together, our embryo and cultured cell metabolomics analyses demonstrate that the embryo metabolome is distinguishable from that of their in vitro counterpart cells. Selective metabolic signatures show consistent trends, such as a more reduced metabolism in two-cell embryos and 2CLCs, and a more oxidative state in blastocyst embryos and ES cells.

Integrated transcriptomics and metabolomics analyses. RNA-seq and chromatin immunoprecipitation followed by sequencing (ChIP–seq) analyses of early embryos have shed light on the

Fig. 1 | Metabolomics profiling of two-cell and blastocyst-stage embryos. a, Schematic workflow for sample processing and targeted metabolomics profiling of embryos and ES cells. UHPLC, ultra-high-performance liquid chromatography. **b**, A PCA plot for targeted metabolomics showing clustering of three biological replicates of 2C and BC embryos ($n=113$ metabolites for each sample). **c**, Heat map showing relative abundance of metabolites in 2C and BC embryos. Peak areas for each detected metabolite were normalized against the total ion count of that sample. For each stage, $n=3$ biological replicates. Each biological replicate was from 100 pooled embryos at either the two-cell or the blastocyst stage. TCA metabolites are indicated by an asterisk. **d**, Volcano plot showing $\log_2$ fold change (FC) values between two-cell and blastocyst embryos and *P* values for the comparison of each individual metabolite shown in **c**. Blue dots (21 metabolites, including α-KG) and red dots (22 metabolites, including SAM and 2-HG) indicate significantly increased or decreased metabolites in two-cell embryos in comparison to blastocysts with a *P* value smaller than 0.05. **e**, Relative abundance of 2-HG and α-KG in two-cell and blastocyst embryos. Data in **e** are from $n=3$ biological replicates and are the mean ± s.e.m. Statistical significance was determined by two-tailed unpaired *t*-test. **f**,**g**, Pathway enrichment analysis of metabolites differentially present in two-cell (**f**) and blastocyst embryos (**g**). The analysis was performed using MetaboAnalyst (4.0)[79]. SCSFAs, short-chain saturated fatty acids; LCSFAs, long-chain saturated fatty acids. **h**, Venn diagrams showing the overlap of metabolites higher in two-cell embryos (compared to blastocyst embryos), and higher 2CLCs sorted from the *Dux*-induced cell line (compared to ES cells). **i**, Venn diagrams showing the overlap of metabolites higher in blastocyst embryos (compared to two-cell embryos) and higher in ES cells (compared to the 2CLCs sorted from the *Dux*-induced cell line). **j**,**k**, Pathway enrichment analysis of the overlapping metabolites in **h** and **i** performed using MetaboAnalyst 4.0.

dynamic genetic and epigenetic programmes of these stages, including genes in metabolism. To further examine embryo metabolic features from a genetic regulation perspective, we reanalysed publicly available bulk and single-cell RNA-seq data in mouse embryos of different stages[15,34], and identified 'energy metabolism' and 'translation' as the two most drastically and consistently altered gene categories in early embryo development (Fig. 2a and Extended Data Fig. 6a). We then examined the ~3,000 metabolism-associated genes including metabolic enzymes or transporters[35], and observed contrasting stage-specific gene expression patterns, and these metabolic genes were classified into six stage-specific clusters by *k*-means clustering for downstream analysis (Fig. 2b). Pathway enrichment



analysis of the bulk RNA-seq data revealed that oxidative phosphorylation (OxPhos) and TCA cycle-related genes were enriched in the blastocyst-stage (equivalent to the inner cell mass (ICM) stage) gene cluster, consistent with the embryo metabolomics data, and phosphatidylinositol signalling-related genes were enriched in the metaphase II (MII)-oocyte-stage and two-cell-stage clusters (Fig. 2c). Similarly, single-cell RNA-seq data showed robust enrichment patterns for OxPhos and TCA cycle in the blastocyst stage, and phosphatidylinositol signalling in the zygote and two-cell-stage embryos (Extended Data Fig. 6b,c), together demonstrating that dynamic transcriptome changes with metabolic alterations are associated with early embryo development stages. When we examined the 2CLC and ES cell transcriptomes[16], we also found metabolic genes globally were not clustered with an embryo stage (Fig. 2b). However, gene-set enrichment analysis on selected gene sets such as 'TCA cycle' and 'OxPhos' showed higher expression in ES cells than in 2CLCs (Extended Data Fig. 6d), consistent with the gene expression features in embryos.

To further validate the signature of 'TCA cycle' and 'OxPhos' (Extended Data Fig. 6e), we tested the ratio of NADH/NAD$^+$ during embryo development with a SoNar biosensor system by injecting in vitro transcribed SoNar mRNA into zygote embryos[36]. The ratio of NADH/NAD$^+$ from two-cell-stage embryos to the blastocyst embryos slightly increased, suggesting an increased mitochondrial TCA cycle and production of NADH (Extended Data Fig. 6f). This is consistent with the metabolomics analysis indicating that the mitochondrial TCA cycle oxidative metabolism was elevated in the blastocyst embryo (Fig. 1g). To validate a strong signature of 'translation' (Fig. 2a and Extended Data Fig. 6a) revealed by an increase of ribosome gene expression (Extended Data Fig. 6g), we determined nascent protein synthesis rate by OP-puromycin staining and found a marked increase from zygotes to morula/blastocyst embryos (Extended Data Fig. 6h), indicating an increased anabolic metabolism programme as the embryos develop, which is critical to support the increased proliferation of embryonic cells. Interestingly, ribosome gene expression and translation activity peaked at the eight-cell to morula stage, and slightly dropped at the blastocyst or ICM stage (Extended Data Fig. 6g), with a concomitant peak in OxPhos gene expression (Extended Data Fig. 6e), suggesting a unique state of slightly braked anabolic programmes, likely for embryos to be primed for the subsequent implantation process.

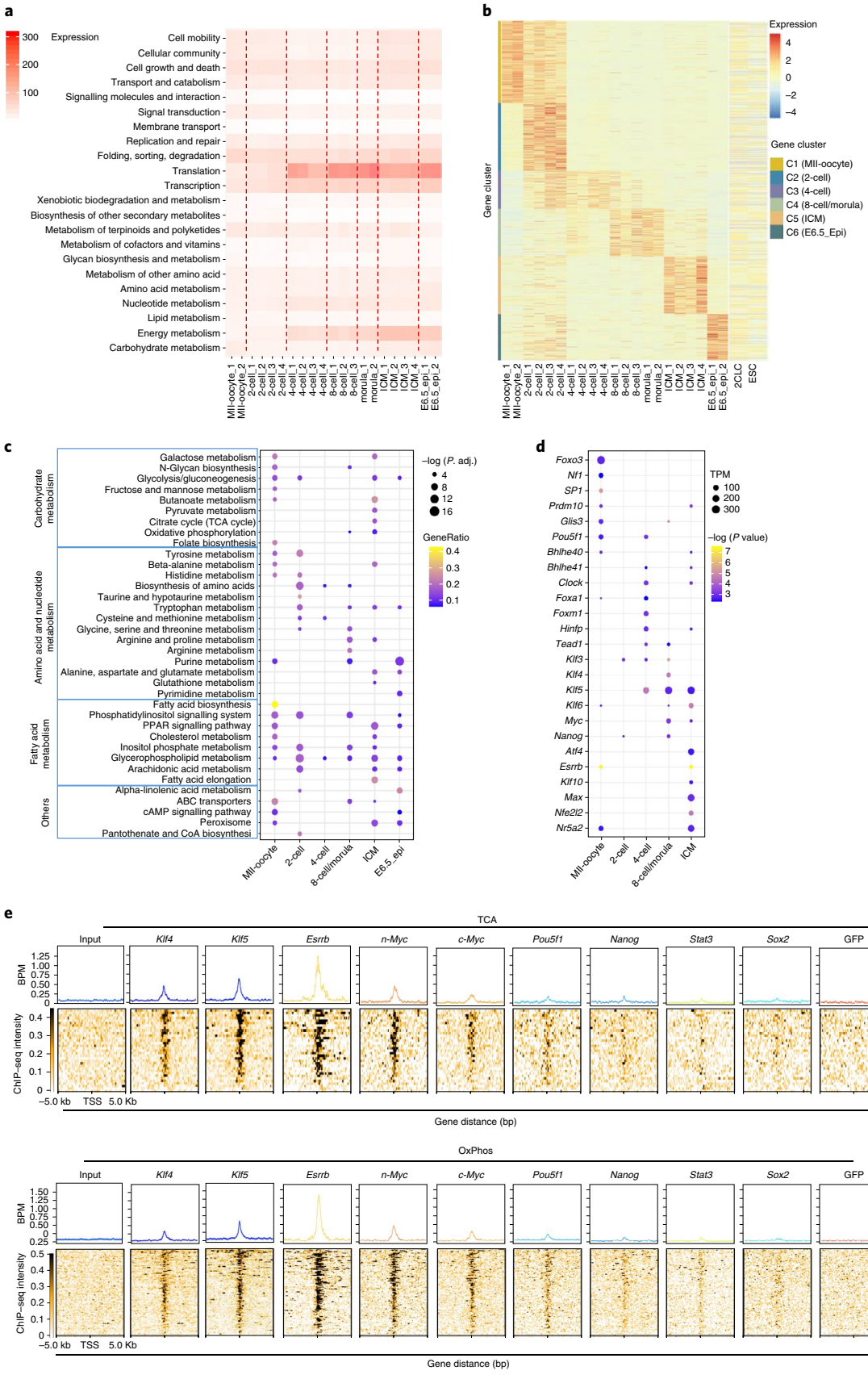
Materials from early embryos can be maternally inherited from oocytes. To distinguish the contribution of maternally inherited versus zygotically transcribed metabolic signatures, we also examined the chromatin status of the ~3,000 metabolism-associated genes using previously published data from the assay for transposase-accessible chromatin using sequencing (ATAC–seq)[11]. ICM stage showed the highest chromatin accessibility of ICM-specific metabolic genes (or cluster 5 genes in Fig. 2b) determined by ATAC–seq peaks at the transcription start sites (TSSs), suggesting these genes are zygotically transcribed at the blastocyst stage (Extended Data Fig. 6i). In contrast, two-cell-stage-specific metabolic genes (or cluster 2 genes in Fig. 2b) did not show notably higher signals at TSSs than the ICM-specific genes (Extended Data Fig. 6i), suggesting that their presence may not reflect active transcription, as most of genes at this stage are likely maternally inherited[31]. Importantly, our metabolomics analysis also identified two-cell-stage-specific metabolic pathways that were not revealed by gene expression analysis such as methionine metabolism (Figs. 1f and 2c). Comparing oocytes with two-cell embryos showed that oocytes are also enriched in metabolites in this pathway, such as methionine (Extended Data Fig. 1h), suggesting that certain metabolites from two-cell-stage embryos may be maternally inherited from oocytes. Together, our transcriptomics analysis showed common and unique signatures compared with the metabolomics analysis, demonstrating the importance of the integrated analysis to obtain a comprehensive understanding of embryo metabolism.

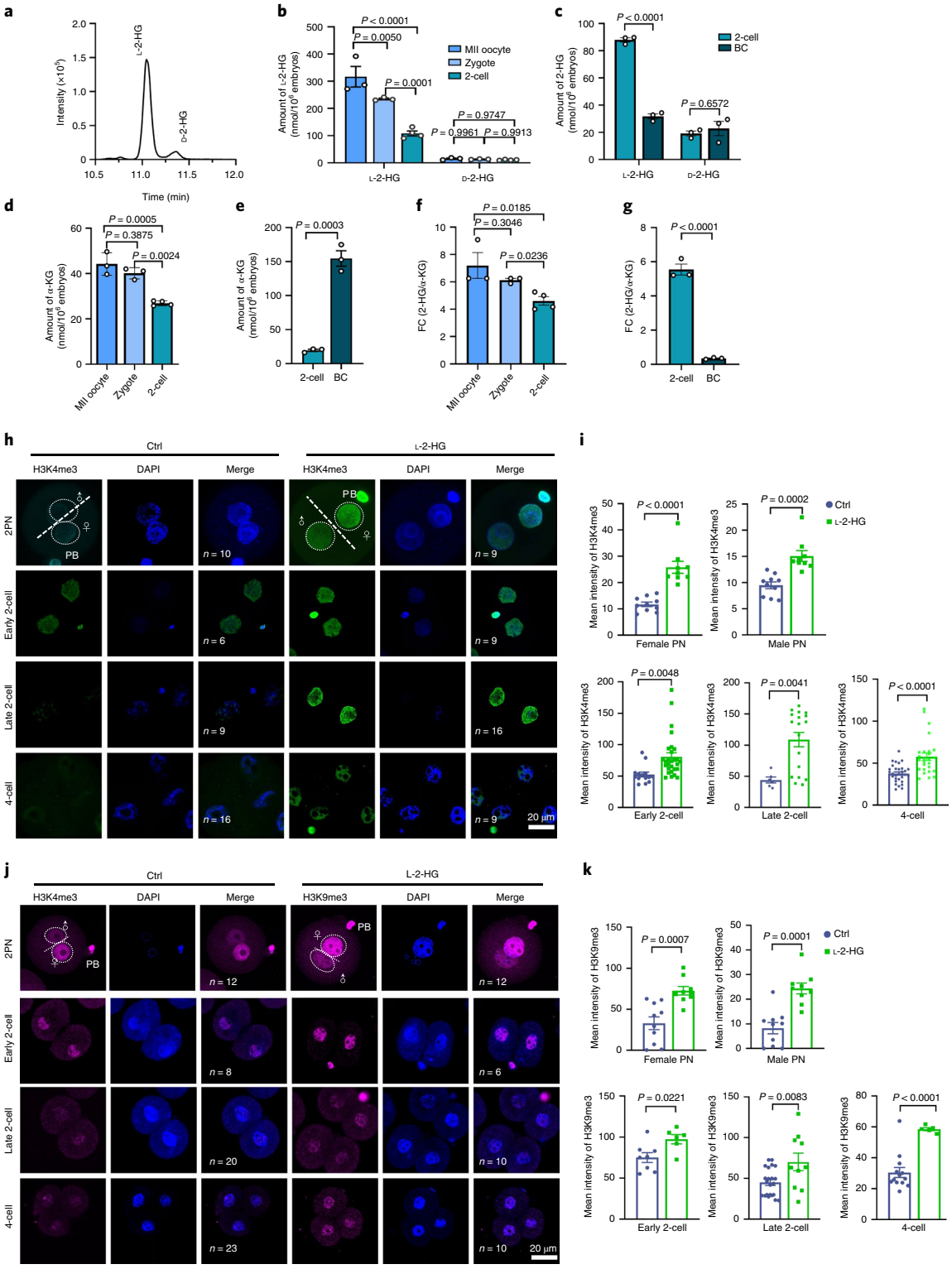
The metabolic network regulated by developmental transcription factor network. The embryo developmental process is tightly governed by transcription regulation from developmental transcription factors (TFs). We performed a TF binding motif enrichment analysis (Extended Data Fig. 6j) across embryo stage-specific clusters of metabolic genes (Fig. 2b). This analysis revealed that TFs like *Esrrb*, *Klf4/5/6*, *Myc* and *Nr5a2* are enriched in the upstream regulators of eight-cell/morula or ICM-specific metabolic genes (Fig. 2d and Extended Data Fig. 6k). Next, co-expression profiles derived from RNA-seq and ChIP–seq binding information were integrated to build a high-confidence TF-metabolic gene regulatory network (Extended Data Fig. 6l) for each embryo stage-specific gene cluster (Fig. 2b). Increasingly complex regulatory networks were established from two-cell/four-cell stages to eight-cell/morula and ICM stages (Extended Data Fig. 6l). Particularly, many mitochondrial TCA cycle and OxPhos-related genes such as *Ndufa13*, *Ndufs2*, *Ndufc1*, *Ndufa1*, *Sdhc*, *Atp4a and Dlat* were among the ICM TF targets. These regulatory networks were consistent with the transcriptome and metabolomics results, implying that the TCA cycle and OxPhos metabolism are highly activated during the blastocyst/ICM stage, but, more importantly, that the metabolic programme is regulated by the classical developmental TFs. We also validated these findings with mouse ES cell ChIP–seq data of key developmental TFs[37], and found *Esrrb*, *Myc and Klf4/5* were largely involved in the regulation of the TCA cycle and OxPhos genes with many shared target genes (Fig. 2e), in contrast to *Sox2, Stat3* and *Nanog*, which appeared to have very little or no influence on the TCA cycle and OxPhos metabolism gene regulation (Fig. 2e). Together, these results demonstrate that metabolic programmes are integrated into a specific set of upstream developmental TF regulatory networks.

Reduction of L-2-HG facilitates histone methylation erasure. Next, to understand whether metabolites are directly involved in gene regulation and embryo development, we chose to examine α-KG and 2-HG, which have established roles in epigenetic regulations and cell fate decision[32,38,39]. α-KG is required for the enzyme activity of many dioxygenases such as histone demethylases, whereas D-2-HG is an oncometabolite produced by mutant IDH1/2, and it acts as an antagonist of α-KG to inhibit histone demethylation[32]. There are two enantiomers of 2-HG, D-2-HG and L-2-HG, and the L-type 2-HG was recently found to be produced under certain physiological conditions[40,41]. The detection of high 2-HG in early embryos when global epigenetic reprogramming takes place made us further explore the subtype and absolute concentration

Fig. 2 | Genomic and metabolic analysis reveals dynamic metabolic network integrated in the developmental TF network in early embryogenesis. **a**, Genome-wide Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of publicly available bulk RNA-seq data with averaged gene expression across indicated pathways showing the dynamics for translation and energy metabolism processes across embryonic development stages. 2C–8C, two-cell- to eight-cell-stage embryos; epi, post-implantation epiblast; E6.5, embryonic day 6.5. **b**, ~3,000 metabolism-associated genes were classified into six stage-specific clusters by *k*-means clustering analysis using bulk RNA-seq data as above. The expression of the same genes in the 2CLCs and ES cells (ESCs) is on the right. **c**, KEGG enrichment terms for the six stage-specific metabolic gene clusters as shown in **b**. **d**, TF binding motifs were identified from the stage-specific metabolism gene clusters as illustrated in **b. e**, ChIP–seq analysis showing the read density distribution in a ±5-kb window around the TSSs of the TCA cycle genes and OxPhos genes bound by key pluripotency TFs. GFP was used as a control. BPMs, bins per million mapped reads.

of this metabolite in MII oocytes, one-cell zygotes and two-cell embryos. Surprisingly, we found that L-2-HG, but not D-2-HG (Fig. 3a and Extended Data Fig. 7a), had the highest abundance in MII oocytes, zygotes and two-cell embryos, and it steadily decreased during the embryo development (Fig. 3b,c). The absolute concentration of L-2-HG that we determined was in the millimolar range,





comparable to some cases of D-2-HG reported in cancer cells harbouring *IDH* mutations or higher than cases of L-2-HG reported in certain physiological conditions (Extended Data Fig. 7b,c)[42–47]. On the contrary, the absolute concentration of α-KG in blastocysts was more than tenfold higher than that in two-cell embryos (Fig. 3e). Even though α-KG was also slightly decreased from MII oocytes to two-cell embryos, the ratio of L-2-HG/α-KG was markedly decreased throughout the early embryo development, from approximately sixfold to less than onefold (Fig. 3f,g). A decreasing L-2-HG concentration and decreasing L-2-HG/α-KG ratio after fertilization suggests that it might allow erasure of various histone methylation marks during this stage[13,27]. To test this hypothesis, we treated embryos with a permeable L-2-HG octyl-L-2-HG[48] during their in vitro development. Global erasure of H3K4me3 and H3K9me3 was indeed impeded or aberrant hyper-methylation was observed in the zygote to four-cell-stage embryos when

Fig. 3 | A decrease of L-2-HG facilitates global erasure of histone methylation during early embryo development. a, A representative chromatogram of L-2-HG and D-2-HG detected in two-cell embryos. **b**, The absolute levels of L-2-HG and D-2-HG in MII oocytes, zygote and two-cell embryos. Data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by two-way analysis of variance (ANOVA) with Tukey's multiple-comparisons post hoc test. **c**, The absolute levels of L-2-HG and D-2-HG in two-cell and blastocyst embryos; data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by two-way ANOVA with Sidak's multiple-comparisons post hoc test. **d**, The absolute level of α-KG in MII oocytes, zygote and two-cell embryos. Data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by one-way ANOVA with Sidak's multiple-comparisons post hoc test. **e**, The absolute level of α-KG in two-cell and blastocyst embryos. Data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by two-tailed unpaired *t*-test. **f**, The ratio of 2-HG/α-KG in MII oocytes, zygote and two-cell embryos. Data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by two-tailed unpaired *t*-test. **g**, The ratio of 2-HG/α-KG in two-cell and blastocyst embryos. Data are presented as the mean ± s.e.m. of three biological replicates. Statistical significance was determined by two-tailed unpaired *t*-test. **h**,**i**, Control and 0.3 mM L-2-HG-treated zygotes, early two-cell, early two-cell and four-cell embryos stained with an H3K4me3 antibody (green). **h**, Representative images of full *z*-series confocal maximum projections of embryos from the indicated number of embryos (*n*) in four independent experiments are shown. ♀, maternal pronuclei; ♂, paternal pronuclei; PB, polar body; 2PN, the 2PN stage of zygote. Scale bar, 20 μm. **i**, Quantification of the mean fluorescence intensity for H3K4me3. Each dot presents the blastomere. Data are presented as the mean ± s.e.m. of an experiment over three biological replicates. Statistical significance was determined by two-tailed unpaired *t*-test. **j**,**k**, Control and 2-HG-treated zygotes, early two-cell, early two-cell and four-cell embryos stained with an H3K9me3 antibody (red). **j**, Representative images of full *z*-series confocal maximum projections of embryos from the indicated number of embryos in four independent experiments are shown. Scale bar, 20 μm. **k**, Quantification of the mean fluorescence intensity for H3K9me3. Each dot presents the blastomere. Data are presented as the mean ± s.e.m. of an independent experiment over three biological replicates. Statistical significance was determined by two-tailed unpaired *t*-test.

supplementing ectopic L-2-HG from the zygote stage (Fig. 3h–k), with a non-toxic L-2-HG concentration that did not affect viability of two-cell or four-cell embryos (Extended Data Fig. 7d,e).

We also found developmental delay and embryo morphological abnormalities (Fig. 4a,b), for example, the formed blastocyst embryos tended to collapse or not be able to hatch normally after L-2-HG treatment (Fig. 4c,d), and the cavity area or averaged cell number per blastocyst decreased (Fig. 4e,f). Moreover, at the blastocyst stage, this effect was more obvious when the treatment was restricted to earlier stages (Fig. 4g,h). RNA-seq analysis with L-2-HG-treated embryos that reached the early blastocyst stage showed a reduction in the TCA cycle and pluripotency genes (Fig. 4i). On the contrary, treatment with a permeable and embryo-tolerable concentration of dimethyl-α-KG[39] showed increased expression of the TCA cycle and pluripotent genes (Fig. 4j). Real-time PCR also validated increased *Nanog* expression following α-KG treatment and decreased *Nanog* expression after 2-HG treatment in blastocyst embryos (Fig. 4k,l). α-KG also rescued the L-2-HG effects on blastocyst formation rate, but could not fully rescue the morphological abnormalities (Fig. 4m,n), suggesting these two metabolites have certain opposing effects, but may also have other non-antagonizing effects[49].

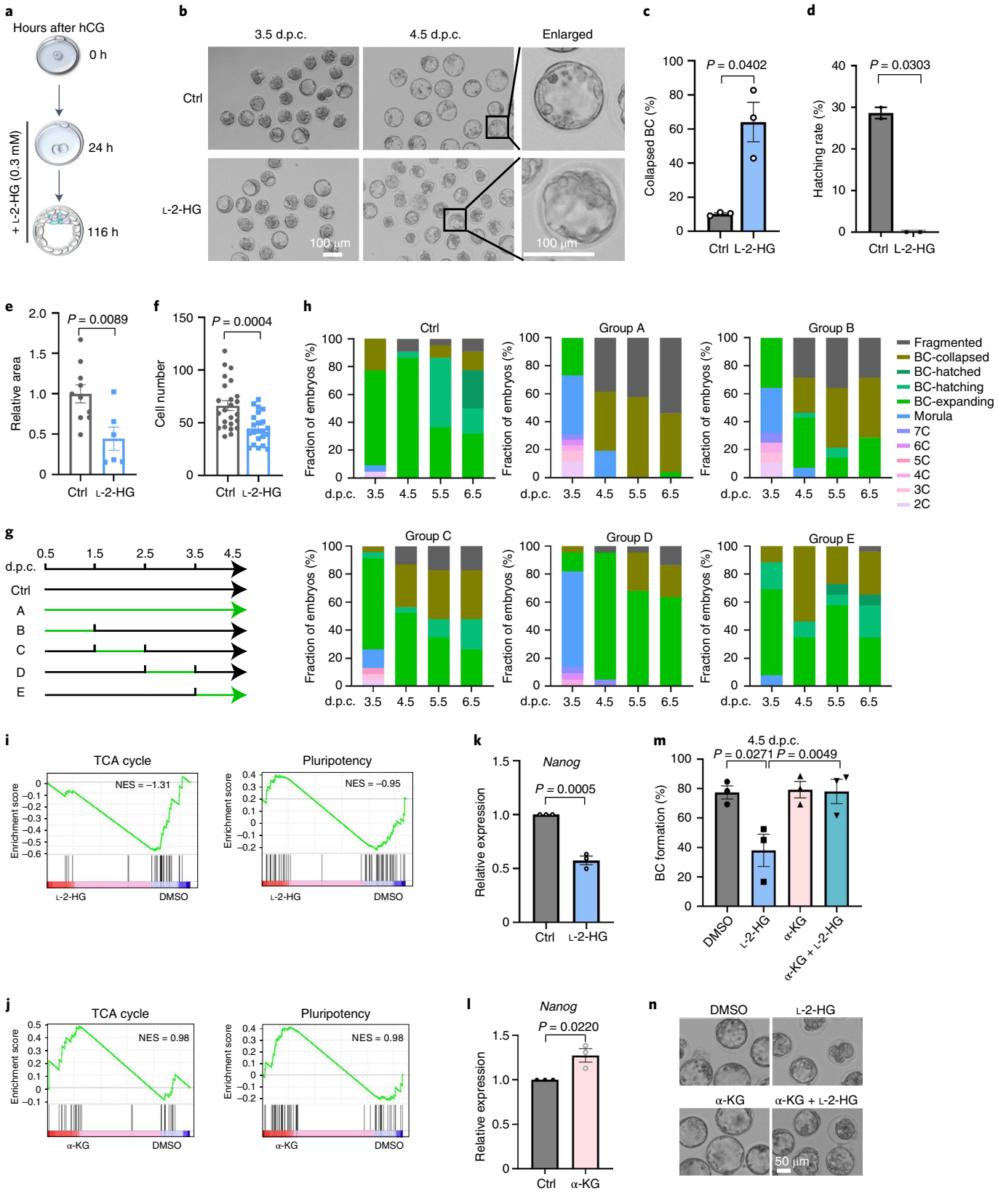
We next explored where the L-2-HG is from and how it is cleared off. It has been reported that malate dehydrogenase (MDH) or lactate dehydrogenase (LDH) can produce L-2-HG under certain physiological conditions[40,41,50], and L-2-hydroxyglutarate dehydrogenase (L-2-HGDH) can consume L-2-HG[51]. We thus determined the expression levels of these enzymes with RNA-seq (Extended Data Fig. 8a), real-time PCR (Fig. 5a and Extended Data Fig. 8b) and western blotting analysis (Extended Data Fig. 8c). For both the mRNA and protein levels, *Ldhb*/LDHB was highly expressed in the oocytes, one-cell and two-cell embryos, with a gradual decrease during embryo development, and showed the lowest expression level at the blastocyst (ICM) stage. The other *Ldh* genes and *Mdh1/2* did not show high expression of their mRNA at or before the two-cell stage of the early embryo development. Simply knocking down the maternal factor *Ldhb* from zygote embryos may not interfere with L-2-HG abundance as it may be produced during oocyte development and inherited maternally from oocytes. Interestingly, mRNA levels of *L2hgdh* increased from two-cell embryos, reaching the highest level at the four-cell stage, before decreasing afterwards (Fig. 5a). Its protein expression was also the highest at the two-cell and four-cell stages and decreased afterwards, suggesting it might have a role in consuming L-2-HG during these stages. When knocking down *L2hgdh* by injecting siRNA in the zygote (Fig. 5b), we found an increase of L-2-HG in the four-cell embryos (Fig. 5c). We also found H3K4me3 hyper-methylation in the four-cell embryos, and in the presence of L-2-HG, this effect was further amplified (Fig. 5d–f), suggesting accumulation of L-2-HG without a consumption enzyme deteriorating the histone methylation erasure process.

Fig. 4 | L-2-HG supplementation impedes early embryo development which is rescuable by α-KG. a, Schematic of the experimental approach. Zygotes were collected 16 h after injection of human chorionic gonadotropin (hCG) and cultured in KSOM medium with 0.3 mM L-2-HG. The embryos were collected at the indicated time for further analysis. **b**, Representative images of BC embryos at 3.5 days post coitum (d.p.c.) and 4.5 d.p.c. after the treatment with L-2-HG. This experiment was repeated three times independently. Scale bar, 100 μm. **c**–**f**, The blastocyst collapsed rate, the hatching rate, the cavity area and the cell number per embryo (counted by DAPI-positive cells) in 0.3 mM L-2-HG-treated (from 16 h after injection of hCG to the blastocyst stage) and untreated embryos. Data in **c** are from $n = 3$ independent experiments; data in **d** are from $n = 2$ independent experiments. Data in **e** are from $n = 10$ and 6 embryos over two independent experiments; data in **f** are from $n = 24$ and 22 embryos over four independent experiments. All data are the mean ± s.e.m. Statistical significance was determined by two-tailed paired *t*-test in **c** and **d** and unpaired *t*-test in **e** and **f**. **g**, Different schemes of the treatment period with 0.3 mM L-2-HG. The 24-h time window for L-2-HG treatment is highlighted in green. **h**, Stacked bar plots showing fraction of embryos at different developmental stages with the different L-2-HG treatment schemes above. The number of embryos from group A to group F were 22, 26, 28, 23, 25 and 26, respectively. **i**,**j**, Gene-set enrichment analysis of the TCA cycle genes and pluripotency genes of L-2-HG-treated (**i**) or α-KG-treated (**j**) BC embryos in comparison to untreated embryos. NES, normalized enrichment score. DMSO, dimethylsulfoxide. **k**,**l**, Quantitative PCR with reverse transcription (RT–qPCR) showing expression of the pluripotency gene *Nanog* in 2-HG-treated (**k**) or α-KG-treated (**l**) BC embryos. Five embryos that reached the blastocyst stage of each condition were pooled for RT–qPCR. Data in **k** and **l** are from $n = 3$ biological experiments and are the mean ± s.e.m. Statistical significance was determined by two-tailed unpaired *t*-test. **m**,**n**, The blastocyst formation rate was determined at 4.5 d.p.c. for embryos treated with 0.3 mM L-2-HG alone, 0.15 mM α-KG alone or α-KG on top of L-2-HG from 24 h after hCG injection to 4.5 d.p.c. Data in **m** are from $n = 3$ biological experiments and are the mean ± s.e.m. Statistical significance was determined by two-tailed paired *t*-test.

Together, these data demonstrate that a reduction of L-2-HG after fertilization and during pre-implantation embryo development is required for the global erasure of histone methylation, and that accumulation of L-2-HG impedes this epigenetic remodelling process and influences embryo development.

Discussion

Mammalian embryogenesis is a complex process with coordinated regulation at multiple molecular levels. Using pooled mouse early embryos and optimized MS-based ultra-sensitive targeted metabolomics for a small number of cells we previously developed[52], here



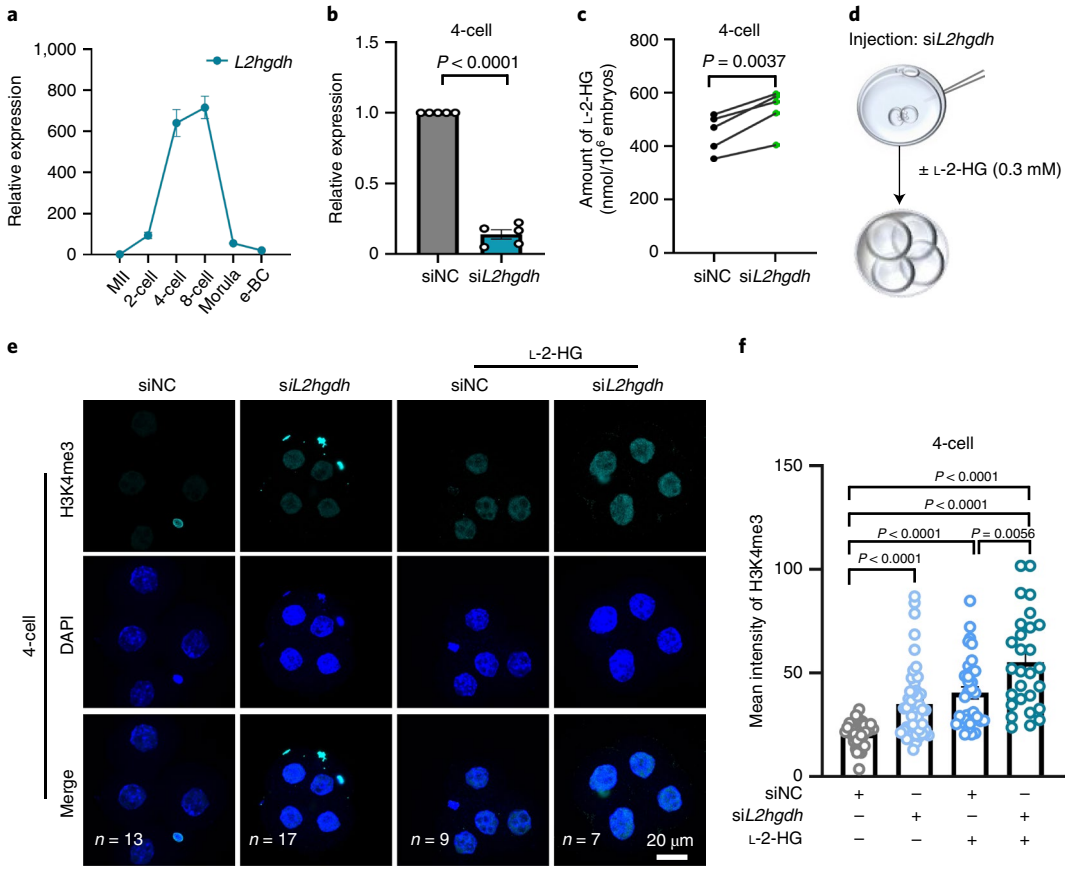


Fig. 5 | *L2hgdh* deficiency impedes H3K4me3 methylation erasure during mouse early embryo development. a, RT–qPCR with RNA from embryos showing the *L2hgdh* mRNA level at different developmental stages. Five embryos were collected and pooled for one biological replicate. Data are presented as the mean ± s.e.m. of three biological replicates from a representative experiment. **b**, RT–qPCR showing the *L2hgdh* mRNA level in four-cell embryos after zygotic injection of siRNAs. Scrambled siRNA duplex without specific targets was used as a negative control (siNC). Five embryos were collected and pooled for one biological replicate. Data in **b** are from $n = 5$ biological replicates and are the mean ± s.e.m. Statistical significance was determined by two-tailed paired *t*-test. **c**, L-2-HG level determined by MS in four-cell embryos after zygotic injection of siRNAs. Fifty embryos were collected and pooled for one sample in five biological replicates. Data are the mean ± s.e.m. Statistical significance was determined by two-tailed paired *t*-test. **d**, Schematic of *L2hgdh* siRNA injection at zygotes and cultured in KSOM supplemented with L-2-HG. Immunofluorescence and RT–qPCR at four-cell embryos were performed. **e**,**f**, Immunostaining of H3K4me3 (cyan) in siNC, si*L2hgdh*, siNC + L-2-HG and si*L2hgdh* + L-2-HG in four-cell embryos. **e**, Representative full *z*-series confocal projection images of embryos from the indicated number of embryos in two independent experiments are shown. Scale bar, 20 μm. **f**, Quantification of the mean fluorescence intensity for H3K4me3. Each dot presents the blastomere. Data in **f** are from $n = 2$ independent experiments and are the mean ± s.e.m. Statistical significance was determined by two-tailed unpaired *t*-test.

we provided a comprehensive and quantitative metabolomics profiling of key stages in mouse early embryo development. By integrating this data with transcriptome, chromatin state and single-cell profiling data on ~3,000 metabolism-associated genes, and by performing metabolic network analysis, we delineated the metabolic landscape of mouse pre-implantation embryo development. Notably, snapshots of metabolomics cannot directly indicate metabolic pathway activity, and isotope tracing analyses on embryos merit further investigations.

Our comprehensive metabolomics and genomics analysis unravelled a redox state transition during pre-implantation embryogenesis, which is consistent with previous functional analysis on redox or respiration[8,53,54]. Transition from a reduced state to an oxidative state is consistent with the transition from high L-2-HG to high α-KG. We found this decrease of L-2-HG is required for histone methylation erasure during early embryogenesis, establishing a role of metabolic regulation on embryo epigenome remodelling. Interestingly, the phenotypes of H3K4 hyper-methylation and embryo collapse defects revealed by L-2-HG supplementation is reminiscent of embryos with the H3K4me3 demethylase genes *Kdm5a* and *Kdm5b* knocked down, suggesting this histone modification may play a role in the L-2-HG effect[13]. However, the effects of decreased L-2-HG in facilitating demethylation might be broader than one class of histone demethylases. For instance, we also observed increased H3K9me3 and DNA methylation after L-2-HG supplementation, and because the global active demethylation of H3K4me3 at the early embryogenesis period is well characterized[13,14,55], we focused on H3K4me3 here. How the effects of L-2-HG on various methylation marks collectively influences embryo development, and the downstream specificity on different modifications and genes, merit further investigation. Importantly, metabolism in early embryos have also been linked to histone acetylation and zygotic genome activation, implicating the broad range of interaction between metabolic and epigenetic regulation during early embryo development[56].

The two-cell and blastocyst metabolic signatures also indicate that embryos transition from a quiescent state to metabolically active state with high anabolic metabolism and mitochondrial TCA cycle to provide intermediate metabolites[57]. The OxPhos and TCA cycle peak at blastocyst embryo stage, which also produce high level of α-KG that can in turn facilitate pluripotent gene expression and

the developmental process[39]. Along with the activation of the TCA and OxPhos metabolism genes, many pluripotency/developmental TFs are induced. In contrast to *Nanog* and *Sox2*, TF family members such as Esrrb, c-Myc and Klf share many binding targets in the TCA and OxPhos genes, and they might coordinately contribute to activation of these genes. This neglected metabolic regulation adds to the well-established roles of developmental TFs in regulating pluripotency networks[58,59], and thus these TFs can be defined as a new set of developmental regulators with dual roles in establishing both pluripotency and metabolic networks.

Methods

Embryo collection. The C57BL/6J and DBA/2 mice were housed in the animal facility of Zhejiang University. They were maintained on a 12-h light and 12-h dark cycle. The catalogue number of the diet is 1010085 (Xietong).

All experimental procedures were performed in accordance with the Animal Research Committee guidelines of Zhejiang University. To collect pre-implantation embryos, C57BL/6J female mice (4–6 weeks old) were intraperitoneally injected with 7.5 IU each of pregnant mare serum gonadotropin (San-Sheng Pharmaceutical) for 48 h followed by injection of 7.5 IU of hCG (San-Sheng Pharmaceutical). The super-ovulated female mice were mated with DBA2 adult males overnight after hCG administration. Embryos at different stages of pre-implantation development were collected at defined time periods after the administration of hCG: 22–24 h after hCG (zygotes), 40–44 h (mid-two-cell) and 92–94 h (mid-blastocysts). Zygotes were collected from ampullae of oviducts and released with a 100 μg ml^{-1} hyaluronidase/M2 (Sigma-Aldrich) solution for removing cumulus cells. The two-cell embryos were collected from the oviducts and BC embryos were collected from the uterus of the females.

In vitro culture of mouse embryos. For culturing embryos, PN3-stage zygotes were collected based on the size and the conformation of their pronuclei and cultured in a 30-μl drop of KSOM medium (Millipore) in mineral oil (Sigma-Aldrich) at 37 °C, 5% CO_2 and 100% humidity. For metabolites supplementation, pronuclear embryos were cultured in KSOM medium (Millipore) containing α-KG (0.15 mM) or 2-HG (0.3 mM). The cleavage rate at two-cell stage at 44 h and the blastocyst formation rate at 108–112 h were recorded. The two-cell, four-cell, eight-cell, morula and BC embryos were collected for subsequent experiments.

Cell culture. Mouse E14 embryonic stem cells were cultured on mouse embryonic fibroblast cells in LIF/2i medium. LIF/2i medium comprises a 1:1 mix of DMEM/F12 (11320-033, Gibco) and Neurobasal medium (21103-049, Gibco) containing 1× N2 and B27 supplements (17502-048/17504-044, Life Technologies), 100 μM non-essential amino acids (GNM71450, GENOM), 1,000 U ml^{-1} LIF (PeproTech), 1 μM PD03259010, 3 μM CHIR99021 (STEMCELL Technologies), 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin (15140-122, Gibco). E14 ES cells transduced with a *2C::tdTomato* reporter to mark 2CLCs were cultured in LIF/serum medium comprising DMEM medium supplemented with 15% FBS, 2 mM Gluta-MAX, MEM non-essential amino acids, sodium pyruvate, 100 mM β-mercaptoethanol and 50 ng ml^{-1} LIF. The *2C::tdTomato* cells were generated as before[16]. Mouse ES cells or mouse induced pluripotent stem cells were transfected with *2C::tdTomato* using Lipofectamine 2000 and selected with 150 μg ml^{-1} hygromycin 48 h after transfection for 7 d.

Immunofluorescence analysis by confocal microscopy. Embryos were rinsed in PBS with 3 mg ml^{-1} polyvinylpyrrolidone (PVP-PBS, Sigma, PVP-360) and fixed in 4% paraformaldehyde (PFA) in PBS for 30 min, and then rinsed in PVP-PBS followed by permeabilization in 0.25% Triton X-100 in PVP-PBS for 1 h. After permeabilization, embryos were rinsed three times in PVP-PBS and blocked in blocking buffer (0.2% BSA, 0.25% Triton X-100, 0.01% Tween 20, 2% donkey serum and 3 mg ml^{-1} PVP in PBS) for 1 h. Then embryos were incubated with H3K4me3 (Cell Signaling, 9751; 1:100 dilution) and H3K9me3 (Cell Signaling, 13969; 1:100 dilution) prepared in blocking reagent at 4 °C overnight. Subsequently, embryos were washed four times for 10 min each time and incubated with a secondary antibody labelled with Alexa Fluor 488 (Abcam, ab150073; 1:200 dilution) or Alexa Fluor 555 (Abcam, ab150062; 1:200 dilution) for 1 h at room temperature. After washing four times with blocking buffer, nuclei were stained with DAPI for 10 min. Embryos were observed using an Olympus FV3000 fluorescence microscope at ×60 magnification with an oil immersion objective. *Z*-sections were taken every 0.5 μm. For all experiments, acquisition parameters were set to make sure the detector was not saturated for obtaining fluorescence intensity signals in a linear range of hybrid detectors. Hence, the fluorescence signal accurately reflects the level of antigen present in the system. Total nuclear fluorescence intensity was calculated as the mean intensity multiplied by the average volume for each embryo stage using Fiji software.

Western blotting. The embryo cells were lysed with 1× SDS Loading buffer at 95 °C for 5 min. The protein was separated by 10% SDS–PAGE and transferred to PVDF membranes. Blocking was performed for 1 h in 5% non-fat milk/PBS-T buffer followed by incubation overnight with primary antibodies against L2HGDH (Proteintech, 15707-1-AP; 1:1,000 dilution) and LDHB (Proteintech, 14828-1-AP; 1:1,000 dilution) at 4 °C. The next day, membranes were incubated with the appropriate secondary antibodies (EARTHOX, E030120-01; 1:3,000 dilution) conjugated to horseradish peroxidase for 2 h at room temperature, and the bands were detected by ECL reagent and autoradiography.

Evaluation of apoptosis by TUNEL. The TUNEL assay was performed on different stage embryos using the One Step TUNEL Apoptosis Assay Kit (Bytone, C1086) according to the manufacturer's instructions. The embryos were washed three times with PVP-PBS and fixed in 4% PFA in PBS for 30 min, then permeabilized in 0.25% Triton X-100 in PVP-PBS for 1 h. The embryos were stained in TdT-FITC (fluorescein) solution at 37 °C for 60 min, washed with PVP-PBS and stained with DAPI for 10 min. Embryos mounted on glass slides were examined by scanning laser confocal microscopy (Olympus, FV3000). Positive labelling for nuclear accumulation of FITC indicated dead cells. The numbers of total cells and labelled dead cells were analysed using Fiji.

Collection of embryos for metabolomics. For the embryo metabolomics gradient experiment, different numbers (15, 30, 59, 83, 120 and 240) of zygotes were collected and rinsed in 0.9% NaCl three times and lysed with cold 80% methanol on dry ice. Eventually 100 2C-stage or BC-stage embryos were collected with three replicates for each group. All collected embryos were kept at −80°C.

Collection of embryonic stem cells and 2C-like cells for metabolomics. Cell isolation methods for metabolomics analysis were optimized to minimize metabolic changes during isolation as previously described (that is, processing speed and temperature)[52]. Specifically, ES cells with a GFP knock-in at the *Pou5f1* locus and transfected with the *2C::tdTomato* reporter were digested with trypsin for 5 min, the reaction was stopped with serum-containing medium, and then left in the incubator at 37 °C for 30 min to allow for dissociation and separation of mouse embryonic fibroblast cells. Floating ES cells in the supernatant fraction were collected and filtered through a 45-μm nylon mesh into 5-ml tubes. ES cells were then sorted in a FACS system (Beckman) running with a sheath fluid of 0.1% BSA/0.5× PBS to eliminate ion suppression of MS signals from salts in the sheath fluid. The pilot experiment contained a gradient of unsorted ES cells of 2,500, 5,000, 10,000, 20,000, 40,000 and 80,000 cells and three replicates in each group. Sorted 10,000 2CLCs (tdTomato positive) and ES cells (GFP positive) were collected for metabolomics. The sorting for ES cells took 1 min and about 8 min for 2CLCs. All cells were maintained on ice before sorting and directly sorted into 500 μl 80% methanol pre-chilled on dry ice. All samples were stored at −80 °C. The inducible *Dux* ES cell line was made as before[33]. Flow cytometry analysis showed the induced tdTomato$^+$ 2CLCs after addition of 1 μg ml^{-1} doxycycline for 24 h.

Metabolite extraction and targeted metabolomics analysis. Mouse ES cells and embryos in cold 80% methanol were stored at −80 °C overnight. Lysed cells were centrifuged at 14,000*g* for 15 min at 4 °C, and the supernatant was transferred to a new tube (pre-chilled on dry ice) and evaporated with a speed vacuum. Dried metabolites were reconstituted in 30 μl of 0.03% formic acid in water, vortexed, centrifuged at 14,000*g* for 15 min at 4 °C and the supernatant was analysed using liquid chromatography–tandem mass spectrometry (LC–MS/MS). A UHPLC system (Nexera X2 LC-30A, Shimadzu) was used for liquid chromatography, with an ACQUITY UPLC HSS-T3 UPLC column (150 × 2.1 mm, 1.8 μm, Waters) and the following gradient: 0–3 min 1% mobile phase B; 3–15 min 1–99% B; 15–17 min 99% B; 17–17.1 min 99–1% B; and 17.1–20 min 1% B. Mobile phase A was 0.03% formic acid in water and mobile phase B was 0.03% formic acid in acetonitrile (ACN). The flow rate was 0.25 ml min^{-1}, the column was kept at 35 °C, and the samples in the autosampler were maintained at 4 °C. The injection volume was 20 μl. MS was performed with a triple-quadrupole mass spectrometer (QTRAP 6500+, SCIEX) in multiple-reaction monitoring (MRM) mode. A total of 262 metabolites were monitored with 165 ion transitions in positive mode and 97 ion transitions in negative mode. For metabolite identification, the parent/product ions of MRM transitions of each metabolite were optimized by injecting analytical standards to the MS spectrometer. And then the retention time of each metabolite was determined by injecting analytical standards to the LC–MS system. Finally, 256 metabolites of common and important metabolic pathways (for example, energy metabolism, carbohydrate metabolism, amino acid metabolism and nucleotide metabolism) were included in this method. This metabolomics method has been widely used in metabolism studies in various fields, including cancer[60–63], infectious diseases[64], stem cells[65] and cardiovascular diseases[66].

Chromatogram review and peak area integration were performed using MultiQuant software v.3.0 (SCIEX). The peak area for each detected metabolite was normalized against the total ion count of that sample, and thus is a fraction of the total detected metabolite content of that sample. This normalization method was also described in other studies[67]. Because the size and the total biomass of the two-cell embryos and blastocyst embryos were similar, we chose to normalize to the total metabolite content, instead of normalizing to the contrasting cell numbers for these two stages. Normalized peak areas were used as variables for multivariate and univariate statistical data analyses.

Absolute quantification of L-2-HG and D-2-HG. For the absolute quantification of L-2-HG and D-2-HG, 20 µl internal standard ([$^{13}C_5$]DL-2-hydroxyglutaric acid, 100 nM) was added to the scraped cells in 80% methanol. After vortex mixing and centrifugation at 14,000*g* for 15 min at 4 °C, the supernatant was evaporated to dryness using a SpeedVac concentrator. Measurement of L-2-HG and D-2-HG was performed according to the previously described method[68,69] with some modification. Next, 50 µl TSPC (12.5 mM in ACN with 5% pyridine) was added to the dried metabolite extracts for chiral derivatization. After reacting at 25 °C for 20 min, the mixture was dried with SpeedVac and then redissolved in 50% aqueous ACN for LC–MS analysis. The quantification of TSPC-labelled L-2-HG and D-2-HG was performed on the LC–MS system with a QTRAP 6500+ Triple–Quadrupole Mass Spectrometer (SCIEX) with electrospray ionization source and a Nexera X2 LC-30A UHPLC system (Shimadzu). The HPLC separation was performed on an ACQUITY UPLC HSS-T3 Column (2.1 × 150 mm, 1.8 µm) at 45 °C. The mobile phase A was water with 15 mM ammonium acetate and 0.1 % formic acid (vol/vol), and mobile phase B was ACN. The gradient was optimized as follows: 0–4 min 10–31% B, 4–15.5 min 31–32% B, 15.5–16 min 32–95% B, 18–18.5 min 95–10% B and 18.5–22 min 10% B. The MS detection was performed under negative-ion mode using MRM. MRM transitions were monitored at 448.1 > 155.1 for both L-2-HG and D-2-HG and 453.1 > 155.1 for [$^{13}C_5$]DL-2-hydroxyglutaric acid.

Absolute quantification of α-ketoglutarate. [$^{13}C_5$]DL-2-hydroxyglutaric acid (100 nM) was used as an internal standard for the quantification of α-KG. The extraction and measurement methods were same with the targeted metabolomics method. The MRM transitions for α-KG and [$^{13}C_5$]DL-2-hydroxyglutaric acid were 145 > 101 and 152 > 134, respectively.

RNA extraction and RT–qPCR analysis. Total RNA was isolated from ES cells using miRNeasy kit (217004, QIAGEN) according to the manufacturer's protocol, and 1 µg RNA was reverse transcribed to cDNA with HiScript II Q RT Super Mix (R223-01, Vazyme). Gene expression was determined with SYBR-Green qPCR Master mix (Vazyme) on a Bio-Rad PCR machine (CFX-96 Touch). *GAPDH* was used as an endogenous control. Primers are listed in Supplementary Table 2.

OP-puromycin labelling. To measure nascent proteins, different stages of mouse embryos were cultured for 1 h in the EmbryoMax KSOM media, and labelling experiments were performed using Click-iT Plus OPP Protein Synthesis Assay Kit (Life Technologies, C10456). After labelling, the samples were fixed for 15 min at room temperature in PBS supplemented with 4% PFA and then permeabilized in PBS supplemented with 0.25% Triton X-100 for 15 min at room temperature. Nuclei were stained with DAPI for 2 min. Embryo imaging was carried out using a Zeiss LSM 880 fluorescence microscope at ×63 magnification with an oil immersion objective.

Embryo siRNA or mRNA microinjection. In total, 50 µM of pooled siRNA solution was injected into the cytoplasm of the zygotes using an Eppendorf FemtoJet microinjector and Narishige micromanipulators. The microinjected zygotes were further cultured at 37 °C under 5% CO_2 up to the morula or blastocyst stage. The cleavage rate and blastocyst formation rate were recorded. The siRNA sequences are listed in Supplementary Table 3.

For injection of SoNar mRNA[36], it was in vitro transcribed from a pcDNA3.1-SoNar under a T7 promoter, and 0.4 µg µl^{-1} mRNA was injected into the zygotes. Developing embryos were observed with 405/488 nm of excitation under a live embryo imaging confocal system using an Olympus FV3000.

RNA-seq library preparation and sequencing. Embryos were collected (five embryos per sample) in a 0.2-ml PCR tube with a micro-capillary pipette and each sample was added with 4 µl lysis buffer (0.2% Triton X-100, RNase inhibitor, dNTPs, oligo-dT primers) and immediately incubated at 72 °C for 3 min, and processed into cDNA with Superscript II reverse transcriptase[70]. The cDNA was amplified with KAPA Hifi HotStart using ~12 cycles. Sequencing libraries were constructed from 1 ng of pre-amplified cDNA using DNA library preparation kit (TruePrep DNA Library Prep Kit V2 for Illumina, Vazyme). Libraries were sequenced on a HiSeq PE150, with paired-end reads of 150-bp long each.

Next-generation sequencing data analysis. All bulk transcriptome sequencing reads, including publicly downloaded and experimentally generated RNA-seq samples, were first trimmed using Trimmomatic 0.36 software with the parameters 'ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36' (ref. [71]; and then filtered using fastq_quality_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/) with the follow settings: minimum quality score 20 and minimum percentage of 80% bases that have a quality score larger than this cut-off value. The filtering-passed reads were mapped to the mm10 genome using HISAT2 (v2.1.0) with the --dta paramenter[72]. PCR duplicate reads were removed using Picard tools and only uniquely mapped reads were kept for further analysis. The expression levels of genes were calculated by StringTie v1.3.4d[73] with --e --B --G parameters using release M18 gene annotations downloaded from GENCODE data portal. For each metabolic gene, reads mapped to mm10 were counted as transcripts per million reads based on their genome locations. For publicly available single-cell RNA-seq data of different mouse embryo developmental stages, we directly downloaded the gene expression data and used the fragments per kilobase of transcript per million mapped read values as expression abundance metrics. *K*-means clustering analysis of metabolic genes was performed using pheatmap (v1.0.10) in the R package. KEGG pathway enrichment analysis of stage-specific metabolic gene clusters was carried out by clusterProfiler (v3.16.0) in R[74].

To tailor and filter ATAC–seq and ChIP–seq reads, we used the same procedure as that for RNA-seq reads processing. For each sample, the retained ATAC–seq and ChIP–seq reads were first aligned to mm10 genomes using Bowtie2 (v2.3.4.1). The ATAC–seq reads were aligned with the parameters: --t --q --N 1 --L 25 --X 2000 no-mixed no-discordant. The ChIP–seq reads were aligned to mm10 with the options: --t --q --N 1 --L 25. All unmapped reads and PCR duplicates were removed. The bamCoverage and bamCompare commands contained in deepTools (v2.5.3)[75], a suite of Python tools for exploring deep-sequencing data, were adopted for downstream analysis. Using the bamCoverage command with the parameters: --normalizeUsing BPM --of bigwig --binSize 100, we normalized the raw reads signal to BPM signal and converted the alignment bam files to bigwig signal files. The 'computeMatrix' and 'plotProfile' commands of deepTools were used to produce the reads density distribution plot for ATAC–seq and ChIP–seq signals in a given genomic region.

Regulatory network construction for stage-specific transcription-factor metabolism-associated genes. To reconstruct stage-specific transcription-factor metabolism-associated gene (TF-MG) regulatory networks, we used a three-step 'enrichment-searching-selecting' procedure. First, we customized the HOMER (v4.11) software to find TF binding motif for each identified cluster. Next, we selected TFs with an enrichment *P* value < 0.1 for subsequent analysis. Then, for each individual cluster, we searched for known links between TFs and MGs in MouseNet (v2)[76], STRING (v11; https://string-db.org/cgi/input.pl) and a gene co-expression network constructed by NetMiner (v1.0.0)[77] that used bulk RNA-seq gene expression data across different developmental stages of mouse embryos as inputs[77]. Finally, we calculated Pearson correlation coefficients between TFs and MGs for each cluster and selected the links contained in MouseNet v2, STRING v11 or NetMiner network and Pearson correlation coefficients > 0.1 to obtain stage-specific TF-MG regulatory networks.

Metabolic network and differential sensitivity analysis. The metabolic network analysis was performed using the Recon1 metabolic model containing the relationship between 3,744 reactions, 2,766 metabolites, 1,496 metabolic genes and 2,004 metabolic enzymes[21]. After constructing the metabolic network based on the metabolomics data from two-cell and blastocyst embryos, simulation of metabolic gene knockout in these two conditions was performed. Gene knockout growth rates were determined using an optimization approach that determines the metabolic flux state that satisfies constraints from metabolomics, nutrient conditions, thermodynamics and stoichiometry[78]. The differences in growth rate of gene knockouts between the two conditions were then *z*-transformed for visualization.

Statistical analysis. All statistical analyses were conducted with R or GraphPad. Details of individual tests are outlined within each figure legend, including the number and type of replication performed (*n*) and the reported error as the s.e.m. All statistical significances are indicated as $*P < 0.033$, $**P < 0.02$ and $***P < 0.001$, and calculated by two-tailed Wilcoxon signed-rank test (for paired samples) and two-tailed Mann–Whitney *U* test (for independent samples) unless stated otherwise.

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

Data availability

RNA-seq data have been deposited in the NCBI Gene Expression Omnibus under accession code GSE181648. Previously published RNA-seq data that were reanalysed here are available under accessions GSE45719, GSE98150 and GSE33923. Published ChIP–seq data for TFs are available under accession GSE11431. Published ATAC–seq data were downloaded from GSE66390. Source data are provided with this paper. The remaining data that support the findings of this study and uncropped versions of blots are available from the corresponding authors upon reasonable request.

Code availability

All the analysis in this study was made based on custom Perl (v5.30.0), Python (v2.7.16) and R (v4.0.2) codes and is available upon reasonable request.

Received: 2 March 2020; Accepted: 31 August 2021;
Published online: 14 October 2021

References

1. Zhang, J. et al. Metabolism in pluripotent stem cells and early mammalian development. *Cell Metab.* **27**, 332–338 (2018).
2. Chronopoulou, E. & Harper, J. C. IVF culture media: past, present and future. *Hum. Reprod. Update* **21**, 39–55 (2015).
3. Conaghan, J., Handyside, A. H., Winston, R. M. & Leese, H. J. Effects of pyruvate and glucose on the development of human preimplantation embryos in vitro. *J. Reprod. Fertil.* **99**, 87–95 (1993).
4. Brinster, R. L. Studies on the development of mouse embryos in vitro. The effect of energy source. *J. Exp. Zool.* **158**, 59–68 (1965).
5. Brown, J. J. & Whittingham, D. G. The roles of pyruvate, lactate and glucose during preimplantation development of embryos from F1 hybrid mice in vitro. *Development* **112**, 99–105 (1991).
6. Gardner, D. K. Changes in requirements and utilization of nutrients during mammalian preimplantation embryo development and their significance in embryo culture. *Theriogenology* **49**, 83–102 (1998).
7. Gardner, D. K. & Lane, M. Culture and selection of viable blastocysts: a feasible proposition for human IVF? *Hum. Reprod. Update* **3**, 367–382 (1997).
8. Houghton, F. D., Thompson, J. G., Kennedy, C. J. & Leese, H. J. Oxygen consumption and energy metabolism of the early mouse embryo. *Mol. Reprod. Dev.* **44**, 476–485 (1996).
9. Leese, H. J. Metabolism of the preimplantation embryo: 40 years on. *Reproduction* **143**, 417–427 (2012).
10. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
11. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
12. Lu, F. et al. Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell* **165**, 1375–1388 (2016).
13. Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
14. Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
15. Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20**, 620–631 (2018).
16. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
17. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
18. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).
19. Teslaa, T. & Teitell, M. A. Pluripotent stem cell energy metabolism: an update. *EMBO J.* **34**, 138–153 (2015).
20. Zhou, W. et al. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* **31**, 2103–2116 (2012).
21. Chandrasekaran, S. et al. Comprehensive mapping of pluripotent stem cell metabolism using dynamic genome-scale network modeling. *Cell Rep.* **21**, 2965–2977 (2017).
22. Sone, M. et al. Hybrid cellular metabolism coordinated by Zic3 and Esrrb synergistically enhances induction of naive pluripotency. *Cell Metab.* **25**, 1103–1117 (2017).
23. Carbognin, E., Betto, R. M., Soriano, M. E., Smith, A. G. & Martello, G. Stat3 promotes mitochondrial transcription and oxidative respiration during maintenance and induction of naive pluripotency. *EMBO J.* **35**, 618–634 (2016).
24. Zhang, J. et al. LIN28 regulates stem cell metabolism and conversion to primed pluripotency cell stem cell article LIN28 regulates stem cell metabolism and conversion to primed pluripotency. *Cell Stem Cell* **19**, 66–80 (2016).
25. Scognamiglio, R. et al. Myc depletion induces a pluripotent dormant state mimicking diapause. *Cell* **164**, 668–680 (2016).
26. Bulut-Karslioglu, A. et al. Inhibition of mTOR induces a paused pluripotent state. *Nature* **540**, 119–123 (2016).
27. Xu, R., Li, C., Liu, X. & Gao, S. Insights into epigenetic patterns in mammalian early embryos. *Protein Cell* https://doi.org/10.1007/s13238-020-00757-z (2020).
28. Xia, W. & Xie, W. Rebooting the epigenomes during mammalian early embryogenesis. *Stem Cell Reports* https://doi.org/10.1016/j.stemcr.2020.09.005 (2020).
29. Eckersley-Maslin, M. A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 (2018).
30. Burton, A. et al. Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat. Cell Biol.* **22**, 767–778 (2020).
31. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic genome activation in vertebrates. *Dev. Cell* **42**, 316–332 (2017).
32. Xu, W. et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
33. Fu, X., Djekidel, M. N. & Zhang, Y. A transcriptional roadmap for 2C-like-to-pluripotent state transition. *Sci. Adv.* **6**, eaay5181 (2020).
34. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
35. Birsoy, K. et al. An essential role of the mitochondrial electron transport chain in cell proliferation is to enable aspartate synthesis. *Cell* **162**, 540–551 (2015).
36. Zhao, Y. et al. In vivo monitoring of cellular energy metabolism using SoNar, a highly responsive sensor for NAD+/NADH redox state. *Nat. Protoc.* **11**, 1345–1359 (2016).
37. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
38. Lu, C. et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
39. Carey, B. W., Finley, L. W., Cross, J. R., Allis, C. D. & Thompson, C. B. Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413–416 (2015).
40. Intlekofer, A. M. et al. Hypoxia induces production of L-2-hydroxyglutarate. *Cell Metab.* **22**, 304–311 (2015).
41. Intlekofer, A. M. et al. L-2-hydroxyglutarate production arises from noncanonical enzyme function at acidic pH. *Nat. Chem. Biol.* **13**, 494–500 (2017).
42. Steinert, E. M., Vasan, K. & Chandel, N. S. Mitochondrial metabolism regulation of T cell-mediated Immunity. *Annu. Rev. Immunol.* **39**, 395–416 (2021).
43. Shim, E. H. et al. L-2-hydroxyglutarate: an epigenetic modifier and putative oncometabolite in renal cancer. *Cancer Discov.* **4**, 1290–1298 (2014).
44. Amankulor, N. M. et al. Mutant IDH1 regulates the tumor-associated immune system in gliomas. *Genes Dev.* **31**, 774–786 (2017).
45. Terunuma, A. et al. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J. Clin. Invest.* **124**, 398–412 (2014).
46. Gross, S. et al. Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J. Exp. Med.* **207**, 339–344 (2010).
47. Bunse, L. et al. Suppression of antitumor T cell immunity by the oncometabolite (*R*)-2-hydroxyglutarate. *Nat. Med.* **24**, 1192–1203 (2018).
48. Tyrakis, P. A. et al. *S*-2-hydroxyglutarate regulates CD8+ T lymphocyte fate. *Nature* **540**, 236–241 (2016).
49. Fu, X. et al. 2-hydroxyglutarate inhibits ATP synthase and mTOR signaling. *Cell Metab.* **22**, 508–515 (2015).
50. Oldham, W. M., Clish, C. B., Yang, Y. & Loscalzo, J. Hypoxia-mediated increases in L-2-hydroxyglutarate coordinate the metabolic response to reductive stress. *Cell Metab.* **22**, 291–303 (2015).
51. Ye, D., Guan, K. L. & Xiong, Y. Metabolism, activity, and targeting of D- and L-2-hydroxyglutarates. *Trends Cancer* **4**, 151–165 (2018).
52. Agathocleous, M. et al. Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature* **549**, 476–481 (2017).
53. Nasr-Esfahani, M. H. & Johnson, M. H. Quantitative analysis of cellular glutathione in early preimplantation mouse embryos developing in vivo and in vitro. *Hum. Reprod.* **7**, 1281–1290 (1992).
54. Dumollard, R., Ward, Z., Carroll, J. & Duchen, M. R. Regulation of redox metabolism in the mouse oocyte and embryo. *Development* **134**, 455–465 (2007).
55. Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
56. Nagaraj, R. et al. Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. *Cell* **168**, 210–223 (2017).
57. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
58. Orkin, S. H. et al. The transcriptional network controlling pluripotency in ES cells. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 195–202 (2008).
59. Ng, H. H. & Surani, M. A. The transcriptional and signalling networks of pluripotency. *Nat. Cell Biol.* **13**, 490–496 (2011).
60. DeNicola, G. M. et al. NRF2 regulates serine biosynthesis in non-small cell lung cancer. *Nat. Genet.* **47**, 1475–1481 (2015).
61. Huang, F. et al. Inosine monophosphate dehydrogenase dependence in a subset of small cell lung cancers. *Cell Metab.* **28**, 369–382 (2018).
62. Kim, J. et al. CPS1 maintains pyrimidine pools and DNA synthesis in KRAS/LKB1-mutant lung cancer cells. *Nature* **546**, 168–172 (2017).
63. Piskounova, E. et al. Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–191 (2015).
64. Li, X. K. et al. Arginine deficiency is involved in thrombocytopenia and immunosuppression in severe fever with thrombocytopenia syndrome. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.aat4162 (2018).
65. Liu, X. et al. Regulation of mitochondrial biogenesis in erythropoiesis by mTORC1-mediated protein translation. *Nat. Cell Biol.* **19**, 626–638 (2017).

66. Nakada, Y. et al. Hypoxia induces heart regeneration in adult mice. *Nature* **541**, 222–227 (2017).
67. Park, J. S. et al. Mechanical regulation of glycolysis via cytoskeleton architecture. *Nature* **578**, 621–626 (2020).
68. Zheng, J. Y. et al. A readily ^{16}O-/^{18}O isotopically paired chiral derivatization approach for the quantification of 2-HG metabolic panel by liquid chromatography–tandem mass spectrometry. *Anal. Chim. Acta* **1077**, 174–182 (2019).
69. Cheng, Q. Y. et al. Sensitive determination of onco-metabolites of D- and L-2-hydroxyglutarate enantiomers by chiral derivatization combined with liquid chromatography/mass spectrometry analysis. *Sci. Rep.* **5**, 15217 (2015).
70. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
71. Bolger, A. M., Marc, L. & Bjoern, U. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
72. Daehwan, K., Ben, L. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
73. Mihaela, P. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
74. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
75. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
76. Kim, E. et al. MouseNet v2: a database of gene networks for studying the laboratory mouse and eight other model vertebrates. *Nucleic Acids Res.* **44**, D848–D854 (2016).
77. Yu, H. et al. NetMiner—an ensemble pipeline for building genome-wide and high-quality gene co-expression network using massive-scale RNA-seq samples. *PLoS ONE* **13**, e0192613 (2018).
78. Shen, F., Cheek, C. & Chandrasekaran, S. Dynamic network modeling of stem cell metabolism. *Methods Mol. Biol.* **1975**, 305–320 (2019).
79. Chong, J. et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–w494 (2018).

Acknowledgements

We thank J. Sheng, L. Shen, D. Ye, Y. Ni, X. Huang, M. Guan, Y. Yang, C. Navdeep and A. Intlekofer for helpful discussion and sharing of facilities. We thank G. Daley and M. Teitell for their long-term support. J. Zhang is supported by the National Key Research and Development Program of China (2018YFA0107100). Z.H. is supported by grants from National Natural Science Foundation of China (92057209). J. Zhang is also supported by the National Key Research and Development Program of China (2018YFA0107103 and 2018YFC1005002), the National Natural Science Foundation projects of China (31871453 and 91857116), Zhejiang Innovation Team grant (2019R01004) and the Zhejiang Natural Science Foundation projects of China (LR19C120001). Z.H. is also supported by grants from National Key R&D Program of China (2019YFA0802102), Tsinghua-Peking Center for Life Sciences (100084) and Beijing Frontier Research Center for Biological Structure.

Author contributions

J. Zhang and Z.H. designed and supervised the study. J. Zhao, K.Y., L.Z., Y.X., L.C., Z.S., Y.Z., Y.Q., S.J., H.P., M.Z. and J.C. performed the experiments. J. Zhao and L.Z. performed sample preparation and metabolomics results data analysis. K.Y. and Z.H. developed the metabolomics method, performed the metabolomics experiments and data analysis. H.-Y.F., J. Zhang, Z.H., C.Z., C.C., W.T. and D.-W.L. performed the bioinformatics analysis. J. Zhang, Z.H., H.L., W.X., H.-Y.F., D.Z., X.F., S.C. and Y.Z. contributed to writing and discussing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s42255-021-00464-x.

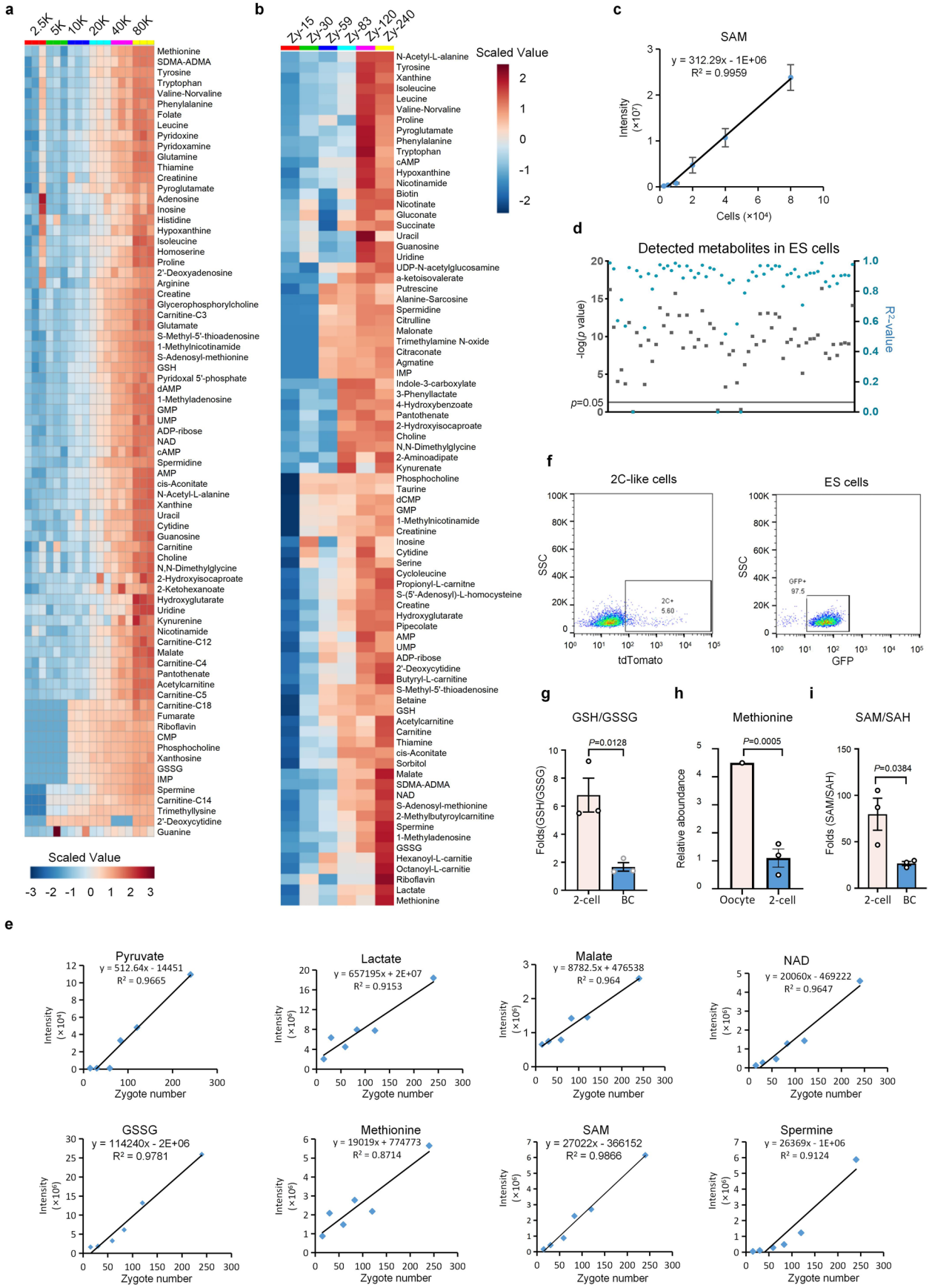
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42255-021-00464-x.

Correspondence and requests for materials should be addressed to Zeping Hu or Jin Zhang.

Peer review information *Nature Metabolism* thanks Navdeep Chandel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. **Primary Handling Editors**: Ashley Castellanos-Jankiewicz; Pooja Jha.

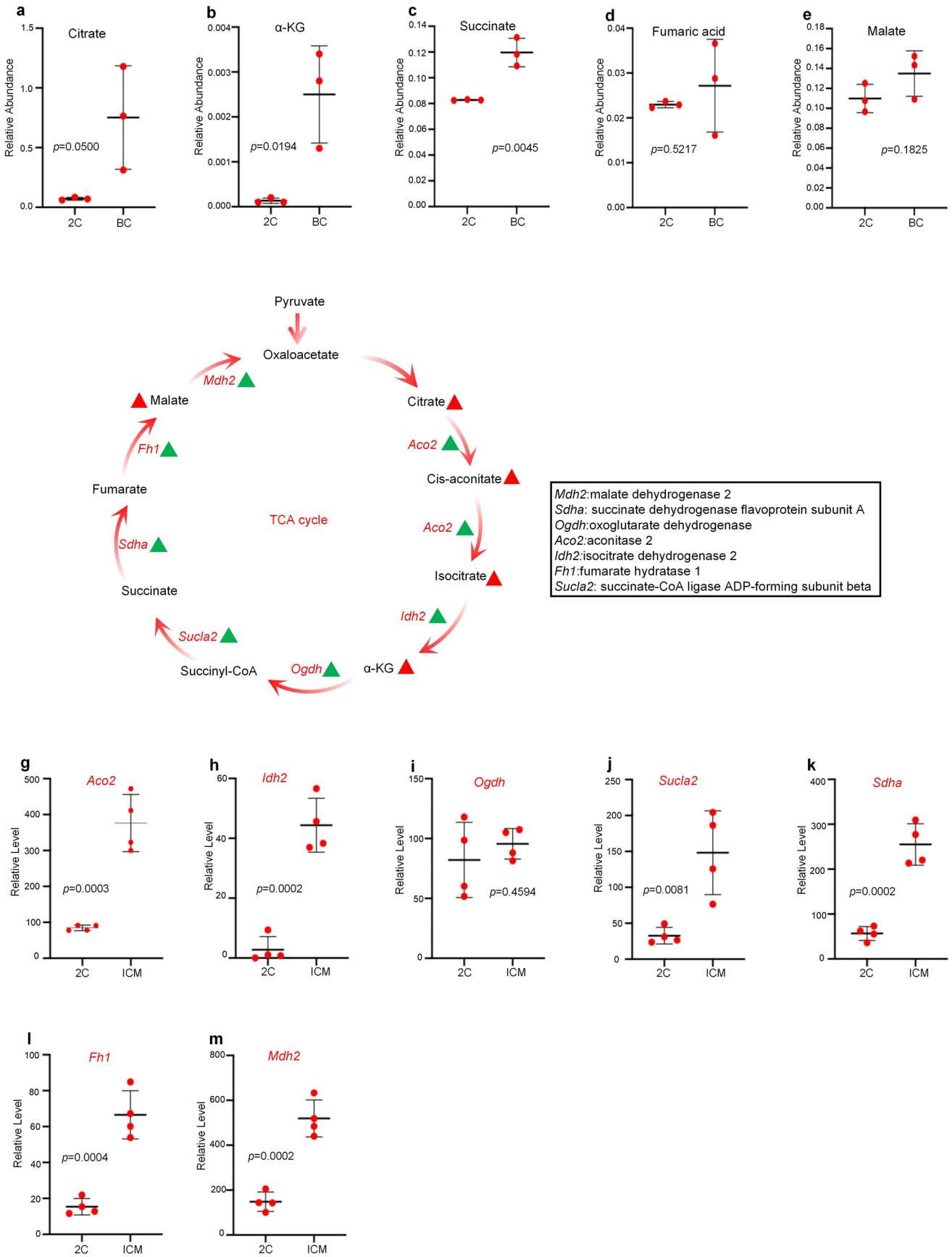
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



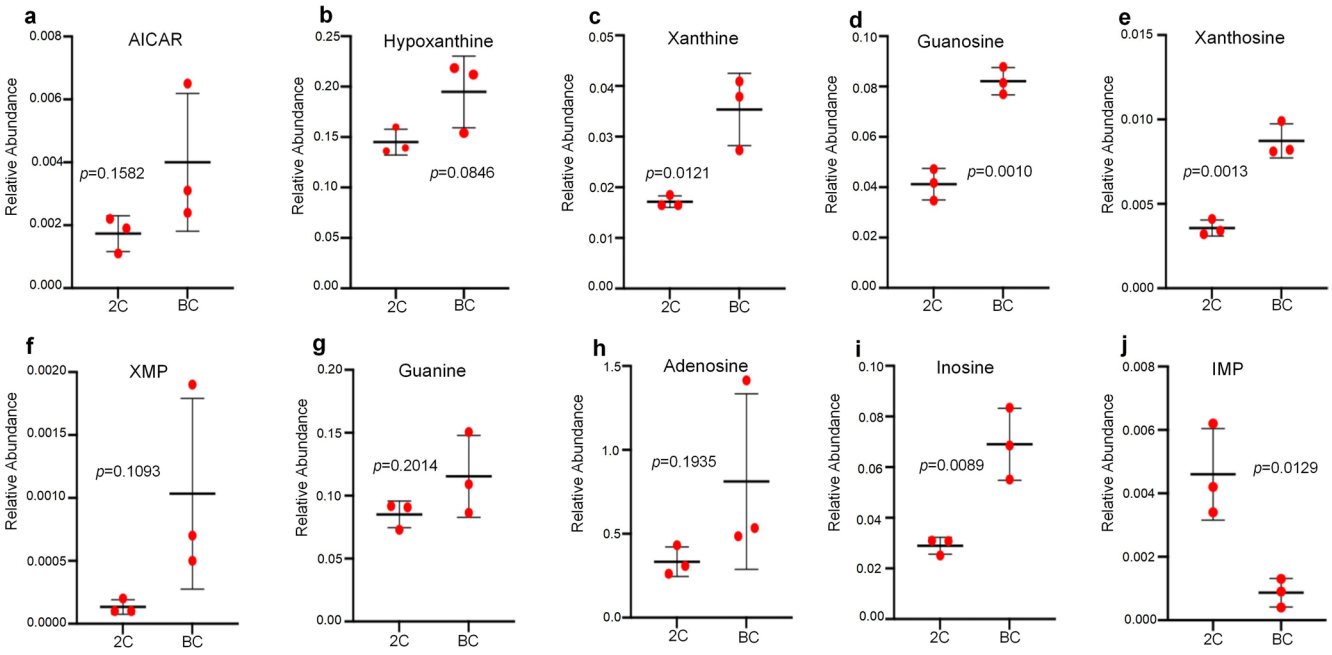
Extended Data Fig. 1 | See next page for caption.

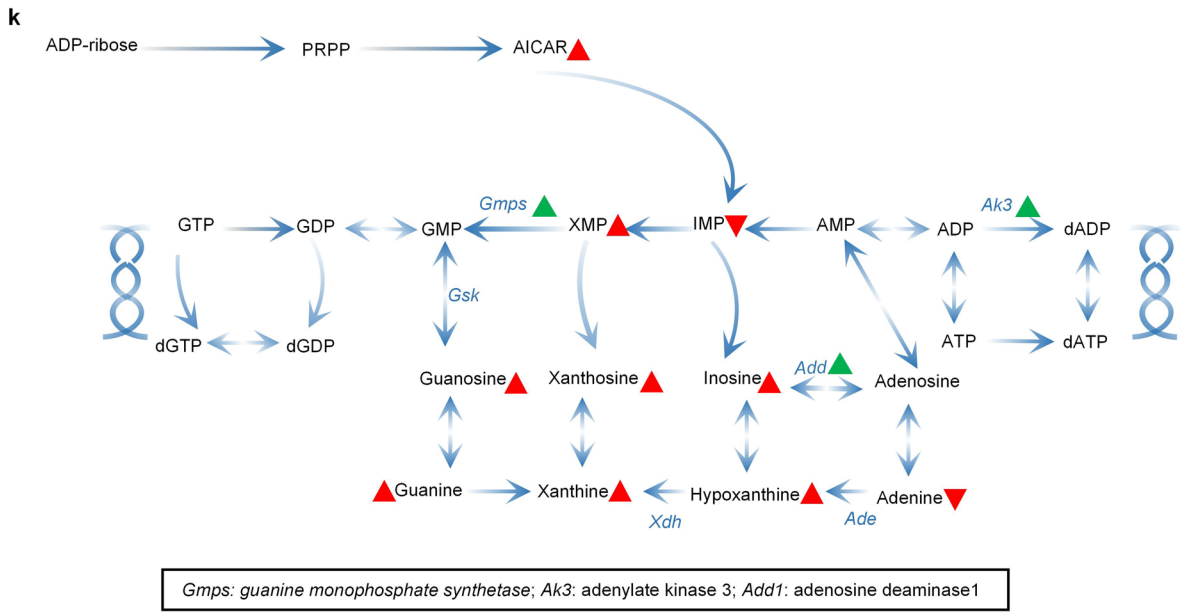
Extended Data Fig. 1 | Metabolomics sample gradient titration, and metabolomics profiling for in vivo derived embryos and in vitro cultured embryo stem cells. (a, b) Metabolites from 2.5 K to 80 K mouse embryonic stem cells or 15 to 240 zygotes were extracted and titrated for targeted metabolomics. Heatmap showing relative metabolite abundance normalized with MetaboAnalyst 4.0. Zy: zygote. (c) Representative metabolite SAM was shown for the correlation between the ES cell number used and relative mass spectrometry intensity obtained. R is the Pearson correlation coefficient. Data in c are from n = 3 biological replicates. Data are mean ± SAM. SAM:S-Adenosyl-methionine. (d) A plot showing R^2 and -log (*p*-value) of each detected metabolite in the ES cell titration experiment. The significance level or *p*-value of Pearson correlation coefficients was obtained using cor.test function in R. (e) Representative metabolites were shown for correlation between the number of embryos and relative mass spectrometry intensity obtained. R is the correlation coefficient. (f) Flow cytometry showing tdTomato positive 2C-like cells (2CLC) and GFP positive ES cells (ESC). (g) The GSH/GSSG ratio in 2-cell and BC embryos. Data in g are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. BC: blastocyst. (h) Relative methionine levels (signal peak areas normalized to total metabolites) in oocytes and 2 C embryos. Data in h are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. (i) The SAM/SAH ratio in 2-cell and BC embryos. Data in i are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed paired t-test.



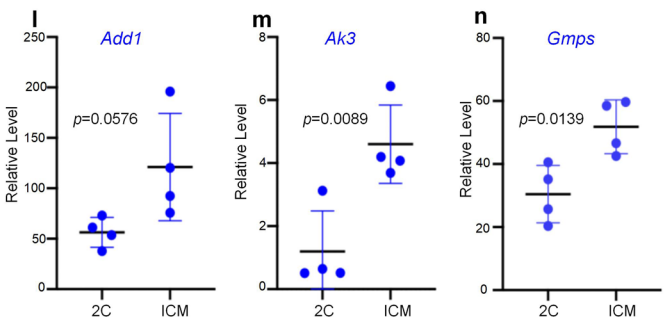
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Analysis of metabolites and the corresponding metabolic enzyme genes in the TCA cycle in 2-cell and blastocyst embryos. (a-e) Abundance of metabolites in the TCA cycle. Data in **a-e** are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. **(f)** The TCA cycle pathway with differential metabolites and genes indicated by the arrows. **(g-m)** Expression of the TCA cycle metabolic genes in 2-cell and ICM of blastocyst. Data in **g-m** are from n = 4 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. ICM: Inner Cell Mass.



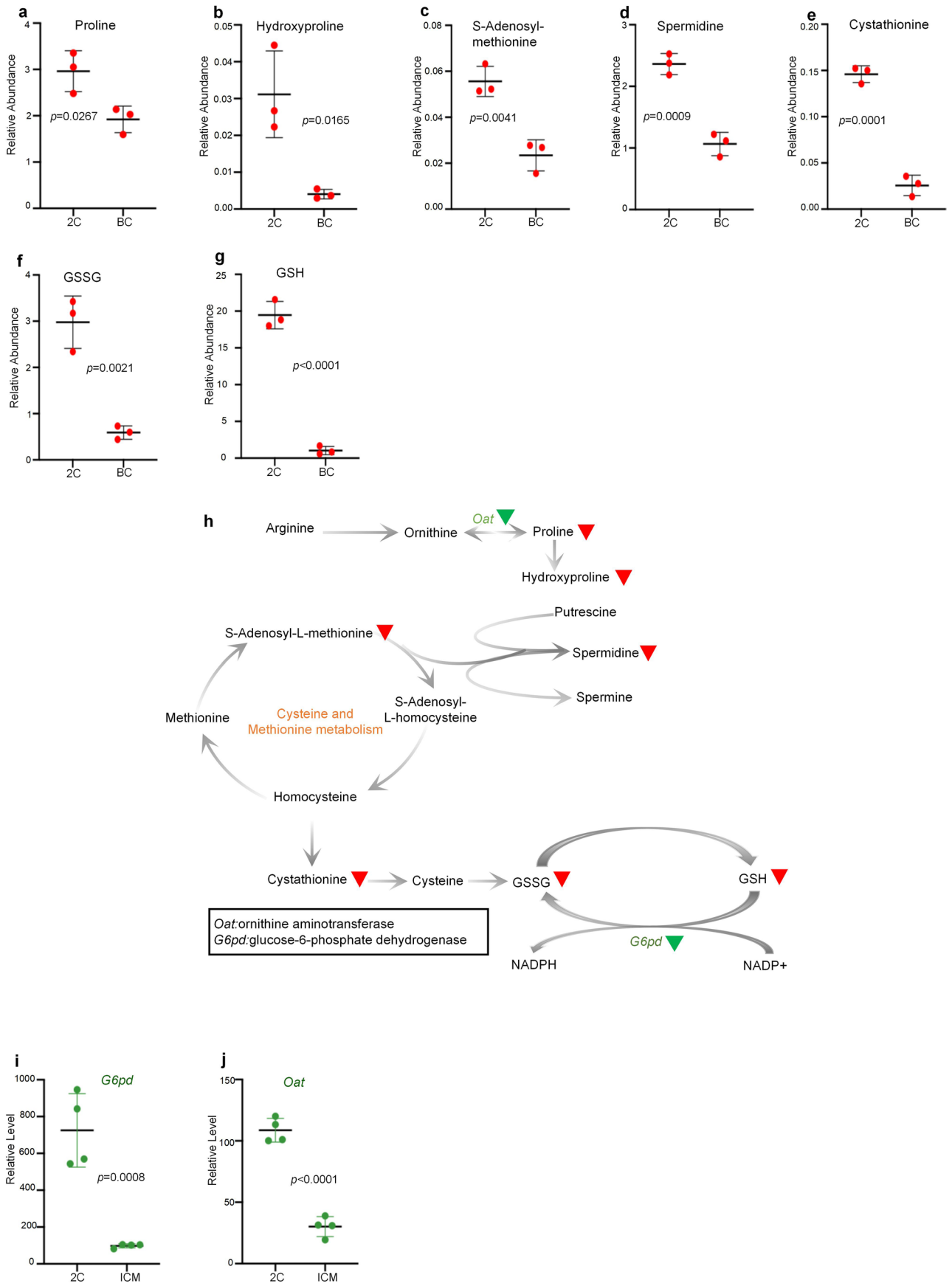


Gmps: guanine monophosphate synthetase; *Ak3*: adenylate kinase 3; *Add1*: adenosine deaminase1



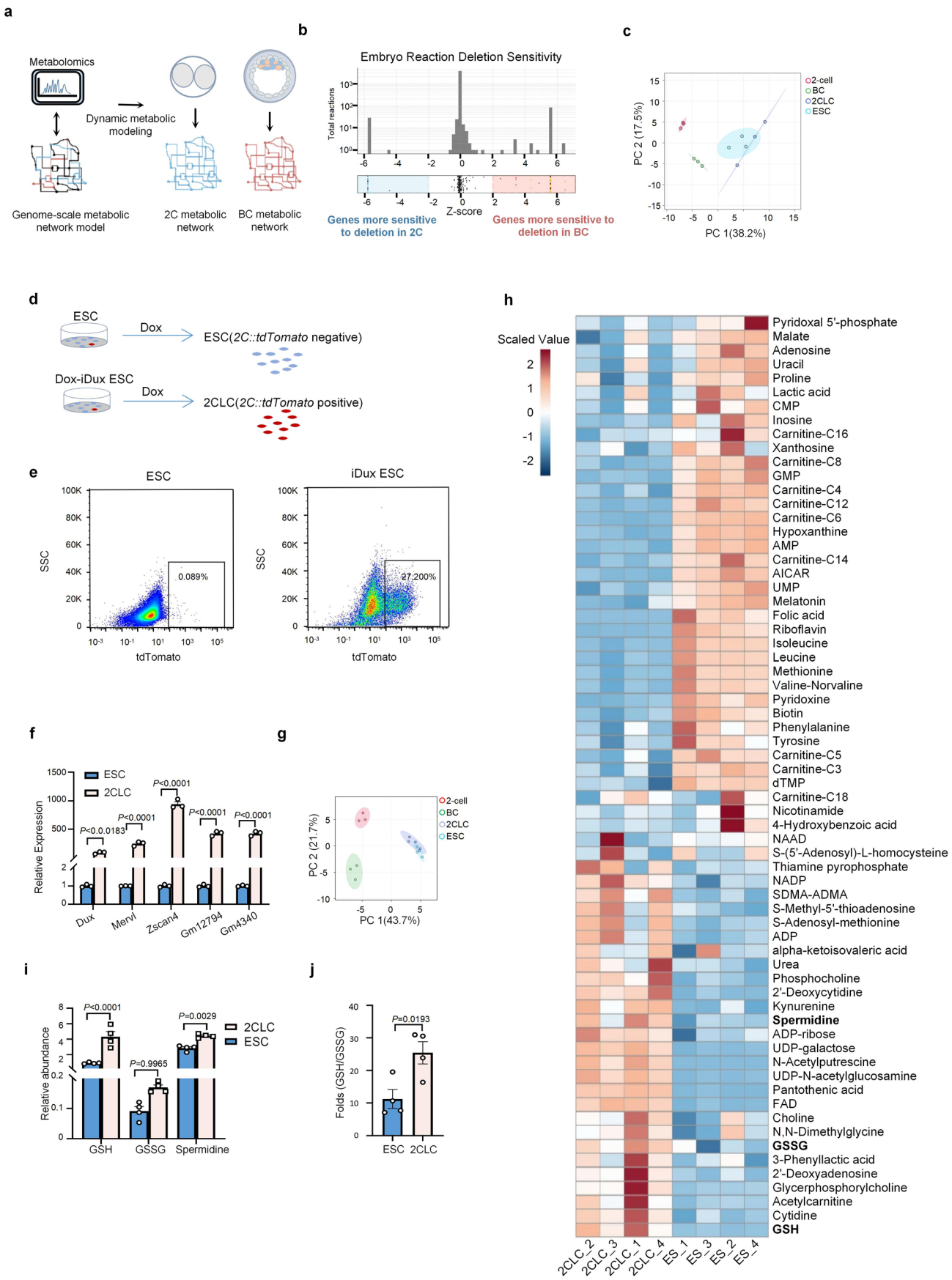
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Abundance of metabolites and expression the corresponding metabolic enzyme genes in the purine metabolism pathway in 2-cell and blastocyst embryos. (a-j) Abundance of metabolites in the purine metabolism pathway. Data in **a-j** are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. **(k)** The purine metabolism pathways with differential metabolites and genes indicated by the arrows. **(l-n)** Expression of the purine metabolism genes in the 2-cell embryos and blastocyst ICM. Data in **l-n** are from n = 4 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test.



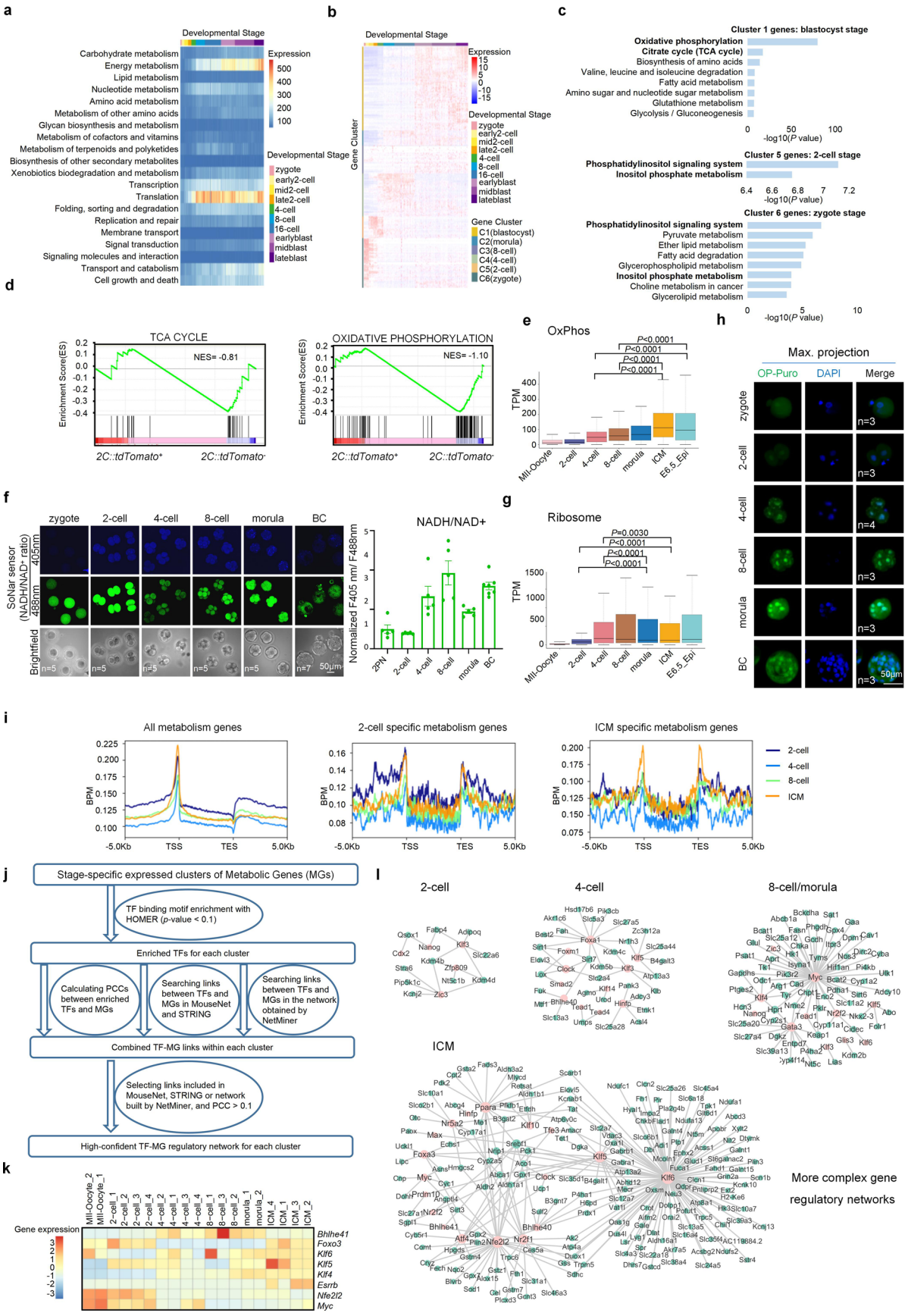
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Metabolite abundance and metabolic gene expression in one carbon metabolism and redox metabolism-related pathways in 2-cell and blastocyst embryos. (a-g) Abundance of metabolites in one carbon metabolism, glutathione metabolism and polyamine metabolism pathways. Data in **a-g** are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test. **(h)** The above metabolism pathways with differential metabolites and genes indicated by the arrows. **(i-j)** Expression of the above pathway metabolism genes in the 2-cell embryos and ICM of blastocyst. Data in **i-j** are from n = 4 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test.



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

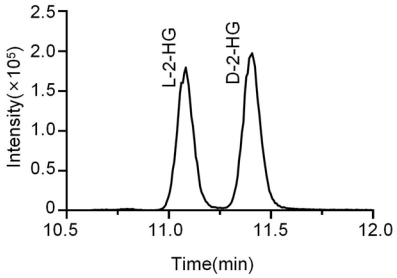
Extended Data Fig. 5 | Metabolic network analysis and metabolomics analysis with 2CLCs. (a, b) Metabolic network of 2 C and BC embryos constructed from metabolomics data as the method before[21]. Deletions of metabolic enzyme genes were simulated, and the differential sensitivity analysis was presented in **b**. **(c)** The PCA analysis of metabolomics from 2-cell, blastocyst (BC) embryos, 2-cell like cells (2CLC) and ES cells. **(d)** Schematics for obtaining the tdTomato$^+$ 2CLCs from the 2 C transcription factor *Dux* inducible ES cell line (Dox-iDux ESC) and the *tdTomato*$^-$ ESCs. **(e)** Flow cytometry analysis showing the induced tdTomato$^+$ 2CLCs upon addition of 1 μg/ml doxycycline(dox) for 24 hours. **(f)** qRT-PCR showing higher expression of 2 C genes in 2CLCs obtained above compared with ESCs. Data in **f** are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-way(ANOVA) with Sidak's multiple comparisons post-test. **(g)** The PCA analysis of metabolomics from 2-cell embryos, blastocyst embryos (BC), 2CLCs from the inducible *Dux* line or 2CLC (iDux), and ESCs. **(h)** Heatmap of the metabolomics analysis of the 2CLC and ESC samples. **(i)** Relative abundance of GSH, GSSG, and spermidine from the 2CLC and ESC samples described above. Data in **i** are from n = 3 biological replicates; Data are mean ± SEM. Statistical significance was determined by two-way ANOVA with Sidak's multiple comparisons post-test. **(j)** The ratio of GSH/GSSG from the 2CLC and ESC samples described above. Data in **j** are from n = 3 biological experiments; Data are mean ± SEM. Statistical significance was determined by two-tailed unpaired t-test.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Single cell RNA-seq analysis, epigenetics analysis of metabolic genes, and anabolic metabolism analysis during pre-implantation embryo development. (a) Genome-wide KEGG pathway analysis showing dynamic gene expression in different developmental stages using publicly available single cell RNA-seq data[34]. Blast: blastocyst. KEGG secondary category pathways are shown. **(b)** All metabolic genes are clustered into 6 clusters (C1-C6) by k-means analysis using single-cell RNA-seq data as **in a**. **(c)** Enriched KEGG terms for Cluster 1 (blastocyst stage), Cluster 5 (2-cell stage) and Cluster 6 (zygote stage) as defined in **b**. **(d)** GSEA analysis of the TCA cycle and Oxidative Phosphorylation between 2CLCs (tomato positive) and ESCs (tomato negative). **(e)** Boxplots showing expression levels of all OxPhos genes in different mouse embryo development stages analyzed from the same bulk RNA-seq data as in **a**. Data in **e** are from n = 4 biological replicates; In **e**, the center line is the median, the bottom of the box is the 25th percentile boundary, the top of the box the 75th, and the whiskers define the bounds of the data that are not considered outliers, with outliers defined as greater/lesser than ± 1.5 x IQR, where IQR = inter quartile range.v. Statistical significance was determined by Wilcox signed rank test. **(f) Left:** Representative images showing the NADH/NAD + ratio in the developing embryos. Scale bar: 50 μm. **Right:** Normalized F405 nm/F488 nm of different stages of embryos. Data are mean ± SEM. **(g)** Boxplots showing expression levels of all Ribosomal genes in different mouse embryo development stages analyzed from the same bulk RNA-seq data as above. Data in **e** are from n = 4 biological replicates; In **g**, the center line is the median, the bottom of the box is the 25th percentile boundary, the top of the box the 75th, and the whiskers define the bounds of the data that are not considered outliers, with outliers defined as greater/lesser than ± 1.5 x IQR, where IQR = inter quartile range.v. Statistical significance was determined by Wilcox signed rank test. **(h)** Translation activity of different stage embryos stained with **OP-Puromycin**. Representative images of full z-series confocal max projections of embryos are shown. Data in **h** are from n = 2 biological experiments with similar results; Scale bar: 50 μm. **(i)** Analysis of publicly available ATAC-seq data for the developmental stage-specific metabolic gene clusters defined in Fig. 2b. **(j)** Schematic diagram for constructing transcription factor-metabolic genes (TF-MG) regulatory network in each developmental stage. **(k)** Heatmap showing gene expression patterns of a few hub TFs identified in early embryo development as in Fig. 2d**. (l)** Gene regulatory networks between TFs or epigenetic factors and metabolic genes for 2-cell, four-cell, 8-cell/morula embryos and ICM in blastocyst embryos.

a



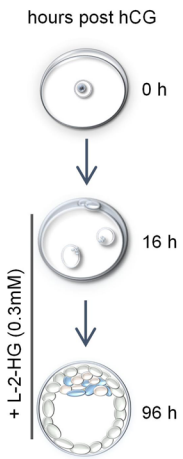
b

2-HG enantiomer	Concentration (Unit)	MII	Zygote	2-cell	BC
L-2-HG	nmol/10^6 embryos (dm=70μm)	316.69	234.96	111.21	31.79
	mM	1.76	1.31	0.62	0.18
	nmol/μg			1.61	0.41
D-2-HG	nmol/10^6 embryos (dm=70μm)	15.36	13.57	11.33	22.88
	mM	0.09	0.08	0.06	0.13
	nmol/μg			0.36	0.30

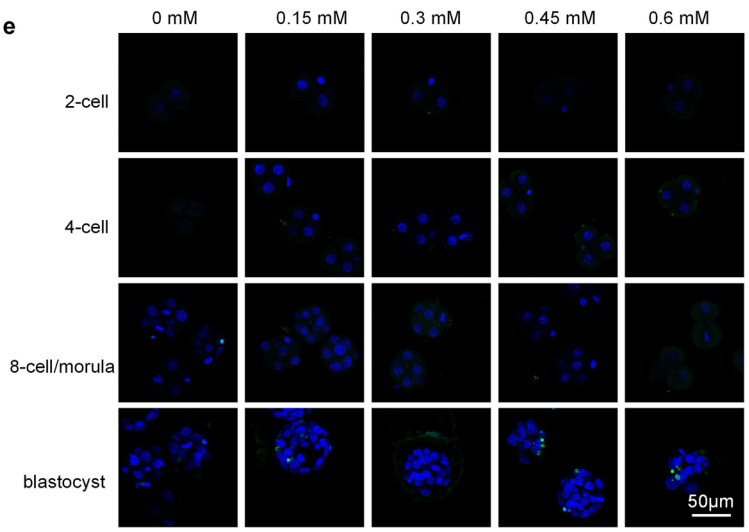
c

2-HG enantiomer	Cell type/tissue	IDH Genotype	Concentration	References
L-2-HG	Drosophila	LDH	2 pmol/μg	Hongde Li et al,2016
D-2-HG	Gliomas	IDH1 mutant	10-100nmol/μg protein	Nduka M. Amankulor et al,2017 Lukas Bunse et al,2018
R-2-HG	Sacoma cell A2DR1	IDH1 mutant	2 pmol/μg protein	Lukas Bunse et al,2018
R-2-HG	LN229	IDH1 mutant	8 pmol/μg protein	Lukas Bunse et al,2018
L-2-HG	T-cell		0.05-0.10nmol/10^6 cells	Tyrakis et al,2016
2-HG	AML	IDH1 mutant	5.36-35.37 nmol/10^6 cells	Stefan Gross et al,2010
2-HG	AML	IDH2 mutant	65.02 nmol/10^6 cells	Stefan Gross et al,2010
2-HG	ER-negative breast tumors		1-20 mmol/kg	Terunuma et al,2013
2-HG	Breast cancer cell lines: MDA-MB-231;MDA-MB-468;SUM159T;HCC-1419;MDA_MB-361		2-5 mM	Atsushi Terunuma et al,2014
L-2-HG	Renal cancer		1-3 pmol/μg protein	Fun-See Shim et al, 2014

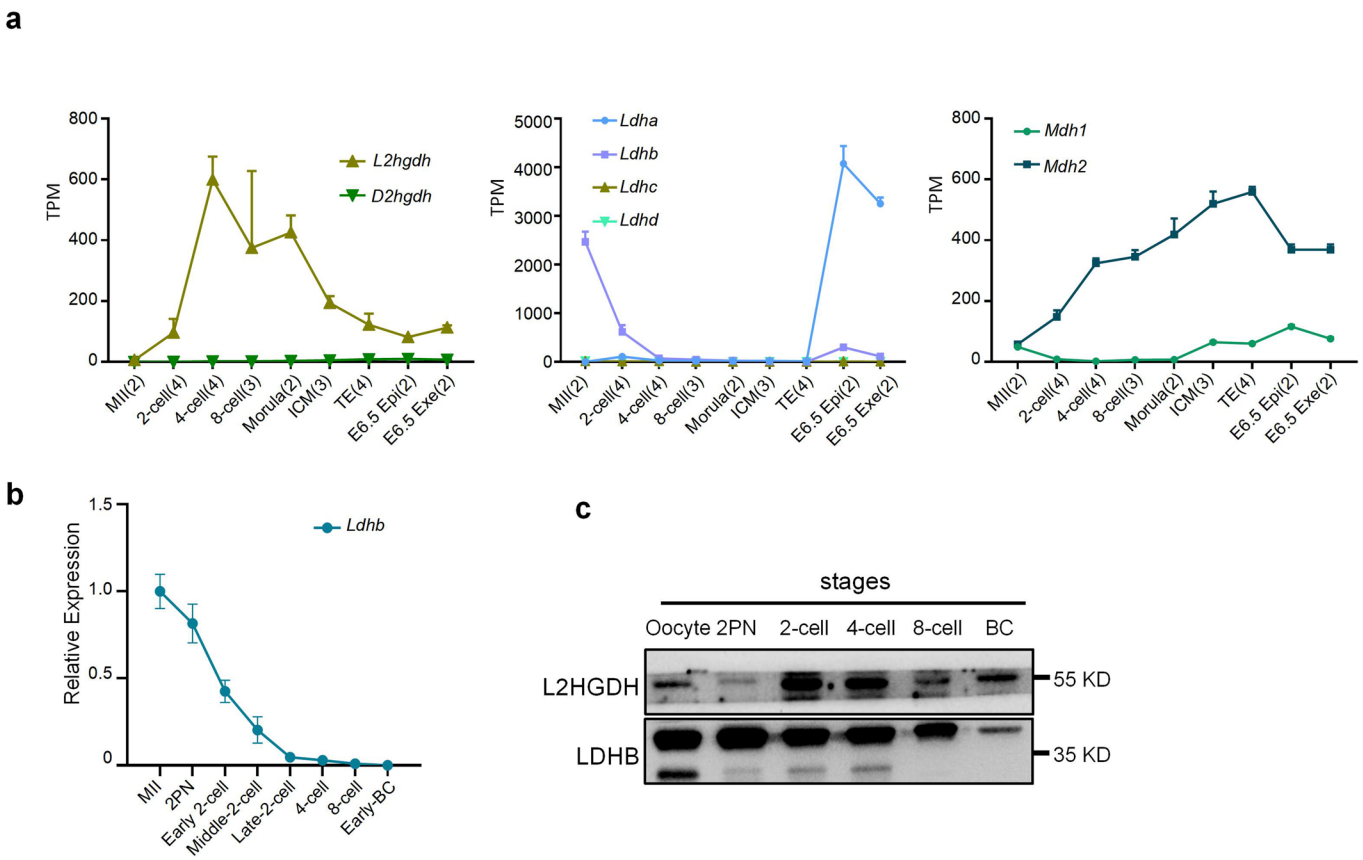
d



e



Extended Data Fig. 7 | Absolute concentration of L-2-HG, and concentration titration for treating embryos. (a) The extracted-ion chromatogram of L-2-HG and D-2-HG standards (concentration 100 nM). It shows clear separation between these two enantiomers. **(b)** The absolute concentration of L-2-HG and D-2-HG in MII, zygote, 2-cell and blastocyst embryos. **(c)** Reported absolute concentration of 2-HG in various cell lines, tumor and tissues. **(d)** Schematic illustration of Tunel experimental approach. Zygotes were collected 16 hours after injection of hCG and cultured in KSOM with different concentration L-2-HG. **(e)** Tunel assay was performed at different stage embryos with different concentration. Embryos were cultured in KSOM with L-2-HG at 0 mM, 0.15 mM, 0.3 mM, 0.45 mM and 0.6 mM. Representative images of full z-series confocal max projections of embryos in three independent experiments are shown. The scale bar is 50 μm.



Extended Data Fig. 8 | Expression of Ldh, Mdh and L2hgdh/D2hgdh mRNA and their protein during early embryo development. (a) mRNA expression of *Ldh*, *Mdh* and *L2hgdh/D2hgdh* genes from RNA-seq results of early embryos. Data are from n = 4 biological replicates; Data are mean ± SEM. **(b)** qRT-PCR of *Ldhb* expression in early embryos. Data are from n = 2 biological experiments with similar results; Data are mean ± SEM. **(c)** Western blotting of LDHB and L2HGHD in early embryos. Equal microgram of protein from protein lysate extracted from equal number of embryos were used for loading.

nature research

Corresponding author(s): JIN ZHANG & ZEPING HU

Last updated by author(s): Aug 18, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No computer code was used to collect the data.
Data analysis	All NGS reads were first processed using Trimmomatic (v0.36) software with the parameters "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36" and fastq_quality_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/) with the follow settings: 1) minimum quality score 20 and 2) minimum percent of 80% bases that has a quality score larger than this cutoff value to trim adaptor and low-quality reads. ChIP-seq reads were aligned to the mouse genome build mm10 using the bowtie2 (v2.3.4.1) command. Signal tracks for each sample were generated using the deepTools' BamCoverage command (v2.5.3) and were normalized to Bins Per Million mapped reads (BPM) for visualization. The "computeMatrix" and "plotProfile" commands of deepTools were used to produce the reads density distribution curve for ATAC-seq and ChIP-seq signals in a given genomic region. RNA-seq reads were mapped to the mm10 reference genome HISAT2 (v2.1.0) with -dta paramenter. Expression levels for all metabolic genes were quantified to TPM using StringTie (v1.3.4d) with -e -B -G parameters and TPM values of replicates were averaged (except for single-cell RNA-seq for which we directly downloaded the gene expression data and used the FPKM as expression abundance metrics). K-means clustering analysis of metabolic genes was performed using pheatmap (v1.0.10) R package. KEGG pathway enrichment analysis of stage-specific metabolic gene clusters were carried out by clusterProfiler (v3.16.0) R package. HOMER (v4.11) software was used to find TF binding motif for each identified metabolic gene cluster. All custom codes were generated using Perl (v5.30.0), Python (v2.7.16) or R(v4.0.2), and can be available upon request.

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

Data

Policy information about availability of data

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

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

RNA-seq data have been deposited in the NCBI Gene Expression Omnibus database under accession code GSE181648. Previously published RNA-Seq data that were re-analyzed here are available under accession GSE45719; GSE98150 and GSE33923. Published ChIP-Seq data for TFs are available under accession code GSE11431. Published ATAC-seq data was downloaded from GSE66390. The remaining data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Usually, sample size was chosen in order to make sure it will be sufficient for statistic analysis. The number of stem cells used to titration was from 2.5k to 80k; the number of zygotes used to titration was from 15 to 240. For metabolomics studies, 3x100 embryos or 3x10000 mouse embryonic stem cells were used. The number of embryos for RT-qPCR and RNA-seq was 5-10 embryos. Twenty-five embryos were used for 2-HG and α-KG absolute quantification. No statistical method were used to predetermine the sample size.
Data exclusions	No data were excluded from the analyses.
Replication	All the experiments were performed 1-5 times with the number of replications shown in the figure legends. The number of replicates is sufficient to determine the corresponding results. All attempts at replication were successful.
Randomization	All the oocytes and embryos were collected from the oviducts of female mice. embryos were randomly allocated for Metabolics, RNA-seq and ctulture.
Blinding	During the siRNA injection, investigators injected the siRNA to the embryos without knowing the types of the siRNA.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

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

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions

Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location

State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access and import/export

Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance

Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

IF:Histone H3K4me3 antibody(Cell signaling, #9751, 1:100), H3K9me3 (Cell signaling, #13969, 1:100); WB:L2HGDH (Proteintech, 15707-1-AP, 1:1000), LDHB (Proteintech, 14824-1-AP, 1:1000). Secondary antibodies: Donkey Anti-Rabbit IgG H&L (Alexa Fluor® 488) (ab150073), 1:200; Donkey Anti-Rabbit IgG H&L (Alexa Fluor® 555) (ab150062),1:200.

Validation

Histone H3K4me3 antibody (Cell signaling Technology, #9751) This antibody can be used for Western Blotting , Immunohistochemistry, Immunofluorescence (Immunocytochemistry) , Flow CytometryChIP, immunoprecipitation, Chromatin IP ,
Chromatin IP-seq and CUT&RUN in human ,mouse, rat and monkey cells according to the manufacture's description. Western Blotting and Immunofluorescence in human ,mouse, rat and monkey cells according to the manufacture's description. There are validation data for IF with Hela cells on the manufactures's website.https://www.cellsignal.cn/products/primary-antibodies/tri-methyl-histone-h3-lys4-c42d8-rabbit-mab/9751?_=1629269716028&Ntt=9751&tahead=true
Histone H3K9me3 antibody (Cell signaling Technology, #13969) This antibody can be used for Western Blotting , Immunohistochemistry, Immunofluorescence (Immunocytochemistry) and Chromatin IP in human ,mouse, rat and monkey cells to the manufacture's description. There are validation data for IF with Hela cells on the manufactures's website. https://www.cellsignal.cn/products/primary-antibodies/tri-methyl-histone-h3-lys9-d4w1u-rabbit-mab/13969.
L2HGDH antibody(Proteintech, 15707-1-AP):This antibody can be used for WB, IHC and ELISA in Human, Mouse and Rat. There are validation data for WB with A549 cells, rat brain tissue, mouse small intestine tissue on the manufactures's website. https://www.ptglab.com/products/L2HGDH-Antibody-15707-1-AP.htm
LDHB(Proteintech, 14824-1-AP): This antibody can be used for WB, IP, IHC, IF,FC and ELISA in Human, Mouse and Rat. There are validation data for WB in DU 145 cells, HeLa cells, mouse brain tissue, PC-3 cells, rat brain tissue, rat liver tissue, mouse heart tissue, rat heart tissue on the manufactures's website. https://www.ptglab.com/products/LDHB-Antibody-14824-1-AP.htm

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The E14 ESCs were gifts from George Q. Daley's lab at Harvard Medical School.

Authentication

For ES cell lines, we performed immunofluorescence staining of OCT4 and NANOG, and RT-qPCR of marker gene Pou5f1 and Nanog, and they are highly expressed indicating these are authentic ES cells.

Mycoplasma contamination

We regularly perform mycoplasma tests and cells used in this study are negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

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

Laboratory animals

C57BL/6 female mice (4–6 weeks old) and DBA2 male mice (10-16 weeks old) were used in our study. They were maintained in temperature and humidity controlled rooms on 12 hours light and 12 hours dark cycle.

Wild animals

No wild animals inolved.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

All animal experiments and study protocols were approved by the Animal Research Committee guidelines of Zhejiang University, and this is noted in the manuscript.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	ftp://47.94.193.106/pub/metaproj/
Files in database submission	Ctrl K9_BKDL192550793-1a_NHi1.bw a-KG K9_BKDL192550794-1a_NHi1.bw 2-HG K9_BKDL192550795-1a_NHi1.bw gene_TPM_NC1_vs_aKG_and NC2_vs_2HG.xlsx gene_TPM_siKlf5_vs_siNC.xlsx
Genome browser session (e.g. UCSC)	track type=bigWig name="CTRLH3K9me3" description="CTRLH3K9me3" visibility=full color=0,100,200 bigDataUrl=ftp://47.94.193.106/pub/metaproj/CrtlK9_BKDL192550793-1a_NHi1.bw track type=bigWig name="aKGH3K9me3" description="aKGH3K9me3" visibility=full color=100,100,100 bigDataUrl=ftp://47.94.193.106/pub/metaproj/a-KGK9_BKDL192550794-1a_NHi1.bw track type=bigWig name="2HGH3K9me3" description="2HGH3K9me3" visibility=full color=200,100,0 bigDataUrl=ftp://47.94.193.106/pub/metaproj/2-HGK9_BKDL192550795-1a_NHi1.bw

Methodology

Replicates	ChIP-seq was performed with one biological replicate for each condition
Sequencing depth	The samples were sequenced using Illumina HiSeq 2500 with 150bp read length and paired-end sequenced. Detailed mapping summary could be referenced in Supplementary Table S1.
Antibodies	Histone H3K9me3 antibody (abcam,, ab8898). This antibody can be used for ChIP,Flow Cyt, Western Blotting, ChIP-seq and Immunofluorescence with wide range reactivity(such as Mouse, Rat, Chicken, Human, Saccharomyces cerevisiae) according to the manufacture's description.
Peak calling parameters	All ChIP-seq and control files were mapped to mouse genome mm10 using bowtie2 (v2.3.4.1) command with the paremeter "-t -q -N 1 -L 25", the mapped sam files for replicates were merged. We did not perform peak calling. Signal tracks for each sample were generated using the deepTools' BamCoverage command (v2.5.3) and were normalized to Bins Per Million mapped reads (BPM) for visualization.
Data quality	Signal tracks for each sample were generated using the deepTools' BamCoverage command (v2.5.3) and were normalized to Bins Per Million mapped reads (BPM) for visualization.

Software	The "computeMatrix" and "plotProfile" commands of deepTools were used to produce the reads density distribution curve for ATAC-seq and ChIP-seq signals in a given genomic region.

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	ES cells transduced with reporter constructs such as 2C::tdTomato, OCT4::GFP or iDux-2C::tdTomato ESC were digested with trypsin for 5 min, stopped with serum containing media, and depleted with MEF at least 30min, and then mESCs with reporter constructs were applied to fluorescence activated cell sorting system to sort tdTomato positive cells or GFP positive cells into 80% methanol.
Instrument	Beckman moflo Astrios EQ
Software	Summit V6.0
Cell population abundance	We sorted 10000 tdTomato or GFP positive cells each sample and applied 3 replicates. The sorted cells are pure 2C-like cells or ES cells determined by their marker gene expression with qPCR and single cell RNA-sequencing.
Gating strategy	Stringent gating strategies were always applied, leaving a significant gap in between negative and positive populations. Dead cells were excluded in FSC/SSC gating.

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.