

A genome-wide screen reveals new regulators of the 2-cell-like cell state

Received: 15 April 2022

Accepted: 19 June 2023

Published online: 24 July 2023

 Check for updates

Nikhil Gupta^{1,7,8}✉, Lounis Yakhou^{1,8}, Julien Richard Albert^{1,2},
Anaëlle Azogui¹, Laure Ferry¹, Olivier Kirsh¹, Fumihito Miura³, Sarah Battault¹,
Kosuke Yamaguchi¹, Marthe Laisné¹, Cécilia Domrane¹, Frédéric Bonhomme⁴,
Arpita Sarkar⁵, Marine Delagrangé⁶, Bertrand Ducos⁶, Gael Cristofari^{1,5},
Takashi Ito³, Maxim V. C. Greenberg^{1,2} & Pierre-Antoine Defossez¹✉

In mammals, only the zygote and blastomeres of the early embryo are totipotent. This totipotency is mirrored in vitro by mouse ‘2-cell-like cells’ (2CLCs), which appear at low frequency in cultures of embryonic stem cells (ESCs). Because totipotency is not completely understood, we carried out a genome-wide CRISPR knockout screen in mouse ESCs, searching for mutants that reactivate the expression of *Dazl*, a gene expressed in 2CLCs. Here we report the identification of four mutants that reactivate *Dazl* and a broader 2-cell-like signature: the E3 ubiquitin ligase adaptor SPOP, the Zinc-Finger transcription factor ZBTB14, MCM3AP, a component of the RNA processing complex TREX-2, and the lysine demethylase KDM5C. All four factors function upstream of DPPA2 and DUX, but not via p53. In addition, SPOP binds DPPA2, and KDM5C interacts with ncPRC1.6 and inhibits 2CLC gene expression in a catalytic-independent manner. These results extend our knowledge of totipotency, a key phase of organismal life.

Embryonic development begins with the zygote: a single, totipotent cell that can give rise to all cells of the embryo. In mouse, this totipotency is still present in the 2-cell embryo, as each cell can regenerate a whole embryo and its extra-embryonic tissues, but this capacity quickly wanes with ensuing divisions^{1,2}. Coinciding with the totipotency of the 2-cell blastomeres is the process of zygotic gene activation: that is, the initiation of transcription in the embryo. A key driver of the zygotic gene activation is the transcriptional regulator DUX (*Dux* in mouse, DUX4 in human), which binds to MERV1 repeated elements and activates their transcription^{3–5}. This, in turn, promotes a general opening of chromatin⁴, as well as the induction of the 2-cell embryo transcriptional signature, containing genes such as *Zscan4a-f* (ref. 6). By the 4-cell stage, the totipotency and the

high chromatin accessibility regress, via mechanisms that are only partially understood.

Mouse embryonic stem cells (ESCs) are pluripotent but not totipotent. However, totipotent-like cells do spontaneously arise in ESC culture in serum, where they can constitute roughly 0.1–0.5% of the population⁷. These cells, termed ‘2-cell-like cells’ (2CLCs) due to their shared properties with 2-cell blastomeres, are a tractable system that parallels the 2-cell stage embryo^{8,9}.

The 2CLCs have specific chromatin features: roughly 30% less DNA methylation, more active chromatin marks and increased histone mobility relative to ESCs. Besides DUX, other key regulators of the 2CLC transcriptome are the transcriptional activators DPPA2, DPPA4 (refs. 10,11) and p53 (ref. 12), which all act upstream of DUX.

¹Epigenetics and Cell Fate, Université Paris Cité, CNRS, Paris, France. ²Institut Jacques Monod, Université Paris Cité, CNRS, Paris, France.

³Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka, Fukuoka, Japan. ⁴Epigenetic Chemical Biology, UMR3523, Institut Pasteur, Université Paris Cité, CNRS, Paris, France. ⁵IRCAN, Université Côte d’Azur, Inserm, CNRS, Nice, France. ⁶High Throughput qPCR Facility, Institut de Biologie de l’École Normale Supérieure (IBENS), Laboratoire de Physique de l’ENS CNRS UMR8023, PSL Research University, Paris, France. ⁷Present address: Joint AZ CRUK Functional Genomics Centre, The Milner Therapeutics Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, UK. ⁸These authors contributed equally: Nikhil Gupta, Lounis Yakhou. ✉e-mail: nikhil.gupta@cancer.org.uk; pierre-antoine.defossez@u-paris.fr

Transcriptional hallmarks of the 2CLC state include the derepression of MERVL repeats, and activation of multiple genes also expressed in the 2-cell embryo, including *Zscan4a-f* (ref. 6) and *Zfp352*. Some genes expressed in 2-cell embryos and in 2CLCs become re-expressed in later developmental contexts, such as gametogenesis¹³. For instance, this is the case for *Asz1*, *Spz1* and of *Dazl*, the last of which encodes an RNA-binding protein essential for spermatogenesis¹⁴. This partial overlap between 2-cell-like and germline transcriptomes comes in part from the existence of shared repressive mechanisms that are lifted in both situations, namely DNA methylation and histone modification by the noncanonical Polycomb Repressive Complex 1.6 (ncPRC1.6)^{15–18}. A consequence of this situation is that regulators of *Dazl* could potentially be relevant for germline development, in the 2-cell embryo or both.

Candidate analyses have shown that chromatin opening is a key factor in the reprogramming of ESCs to 2CLCs (ref. 19). In addition, some medium to high-throughput screens have provided important contributions to our understanding of this system^{15,20–22}. However, these screens have often focused on the reactivation of the MERVL repeats while neglecting other markers of totipotency and/or used targeted small-interfering RNA screening, which does not cover the whole genome.

Here, we have undertaken a genetic screen in mouse ESCs, using the reactivation of *Dazl* as a 2CLC marker. This led to the identification of 40 high-confidence hits, including many of the already known actors. Secondary screens based on additional 2CLC markers allowed us to identify the hits that induce a 2CLC state, of which we characterized four in more detail: SPOP, ZBTB14, MCM3AP/GANP and KDM5C. Our functional experiments then establish the epistasis of these factors relative to known regulators. Among our findings, we show that the lysine demethylase KDM5C shares most of its targets with ncPRC1.6 and represses the 2CLC signature in a catalytically independent manner. Together our data bring new concepts and new actors to our comprehension of the 2-cell-like state.

Results

Design and validation of the epigenetic reporter

We first reanalyzed published expression data^{19,23}, confirming that *Dazl* is highly expressed in 2CLCs, but not in serum-grown ESCs (Fig. 1a). Switching ESCs from serum to 2i, which reduces DNA methylation^{24,25}, also results in *Dazl* re-expression (Extended Data Fig. 1a). Mining expression data, we observed that the abundance of *Dazl* messenger RNA in 2i-grown ESCs is comparable to that seen in late 2-cell embryos, whereas the abundance in serum-grown ESCs is similar to that seen in 8-cell-embryos (Fig. 1b).

DNA methylation and ncPRC1.6 are already known to repress *Dazl* transcription in ESCs in serum (Fig. 1c)^{16–18,26,27}. To identify additional regulators, we carried out a CRISPR knockout (KO) screen on serum-grown ESCs, selecting for cells in which *Dazl* becomes reactivated (Fig. 1c). Such factors could be gene-specific, acting only on *Dazl*, or they could be reactivating wider transcriptional programs, of which *Dazl* expression is a component. Specifically, these programs could mirror the 2CLC state, gametogenesis, or both. In addition, some KOs might act by causing the cells to become 2i-like (Fig. 1c).

Before the screen, we verified that in our ESC background, J1, *Dazl* was expressed and its promoter unmethylated in 2i conditions, whereas *Dazl* was repressed, its promoter methylated in serum conditions (Extended Data Fig. 1a,b). We then used CRISPR–Cas9 to knock in two selectable markers into the gene: mScarlet, a bright red fluorescent protein, and the Hygromycin-resistance gene Hygro^R (Fig. 1d and Extended Data Fig. 1c). The markers were inserted within the *Dazl* coding sequence at exon 6 (present in all splicing forms), and were separated by T2A and P2A self-cleaving peptides (Fig. 1d and Extended Data Fig. 1c). Clones were picked and validated by genomic PCR, sequencing and droplet digital PCR (ddPCR) (Extended Data Fig. 1d). To summarize, we generated a heterozygous mouse ESC line in which one allele of *Dazl*

is WT, and the other contains the mScarlet-Hygro^R insertion (Fig. 1f). We refer to this line as the DASH (*Dazl*-Scarlet-Hygro) reporter line.

As expected, DASH cells are mScarlet-positive and Hygromycin resistant in 2i, but mScarlet-negative and Hygromycin-sensitive in serum (Fig. 1e,f and Extended Data Fig. 1e). Also, as expected, the *Dazl* promoter was methylated in serum and unmethylated in 2i (Extended Data Fig. 1f). Two controls support the validity of the DASH reporter to detect 2CLCs. First, the reporter is strongly induced by the addition of sodium acetate, which promotes the 2CLC state²⁸ (Extended Data Fig. 1g). Second, the expression of mScarlet and ZSCAN4 are parallel, as we showed by immunofluorescence experiments. Our positive control was a treatment with sodium acetate that, as expected, induced ZSCAN4-positive cells (Extended Data Fig. 1h). We observed that most ZSCAN4-positive cells (roughly 90%) were found in the mScarlet-positive cells (Extended Data Fig. 1i). In other words, induction of ZSCAN4 is accompanied by induction of the DASH reporter. After these validations, we used the DASH line for a genetic screen.

Genome-wide KO screening yields 40 high-confidence hits

We infected the DASH cells (in serum) with a lentiviral genome-wide CRISPR KO library²⁹. The coverage was roughly 150 times and two independent screens were carried out in parallel. After selecting for infected cells with puromycin, hygromycin was applied in increasing concentrations (Fig. 2a). At the threshold we used, mScarlet-positive cells represented roughly 3% of the starting DASH population; this proportion increased to roughly 25–30% in the Hygromycin-resistant cells (Extended Data Fig. 2a), and the Hygromycin-resistant/mScarlet-positive cells were purified by fluorescence-activated cell sorting (FACS). In this population, quantitative PCR with reverse transcription (RT–qPCR) revealed a number of differences relative to the starting population: upregulation of *Dazl* (as expected), but also increased expression of *Prdm14* and decreased expression of *Dnmt3a*, *Dnmt3b* and *Dnmt3l*, all of which are characteristic of 2i-like cells^{24,25} (Extended Data Fig. 2b). In this selected population, the bulk DNA methylation (Extended Data Fig. 2c) and the *Dazl* promoter methylation (assessed by methylated DNA immunoprecipitation, MeDIP) were lower than in the preselection sample (Extended Data Fig. 2d).

The single-guide RNAs in the Hygro-resistant and/or mScarlet-positive cells were amplified, sequenced and analyzed with MAGeCK³⁰. The top 40 candidates had a *P* value smaller than 5×10^{-4} and were analyzed further (Fig. 2b). Gene ontology terms enriched in this set of candidates include ‘Maintenance of DNA methylation’, along with ‘Glycosaminoglycan synthesis’ and ‘Heparan sulfate synthesis’ (two related terms), and ‘FGFR signaling pathway’ (Fig. 2c). As for UniProt gene ontology terms, ‘Repressor’, ‘Chromatin regulator’ and ‘DNA binding’ were enriched (Extended Data Fig. 2e).

We then grouped the 40 candidates based on known functions and interactions in the STRING database (Fig. 2d). We recovered factors required for DNA methylation maintenance³¹ (UHRF1, DNMT1, USP7), and components of ncPRC1.6 (E2F6, MGA, TFDPI); these hits were expected (Fig. 1c) and validate the screen. We also obtained eight candidates in the ‘TGFβ-Wnt signaling’ and the ‘FGFR signaling’ clusters, genes involved in ‘RNA processing’—including three out of four subunits of the TREX-2 complex (ENY2, SEM1, MCM3AP)—a lysine demethylase, KDM5C (ranked 16), two components of E3 ubiquitin ligase complexes (SPOP³², FBXW7) as well as additional cytoplasmic (PDCL, DPYSL4) and nuclear factors (ZBTB14). New infections were used to generate CRISPR KO populations for 22 selected genes in the top 40, of which 20 (91%) increased the number of mScarlet-positive cells (Fig. 2e), proving the robustness of the screen results.

Refining the list to potential 2CLC regulators

One class of possible irrelevant hits we sought to eliminate were KOs causing the cells to be ‘2i-like’ even in serum condition, with low DNA methylation and high *Dazl* expression. Mutation of genes in the

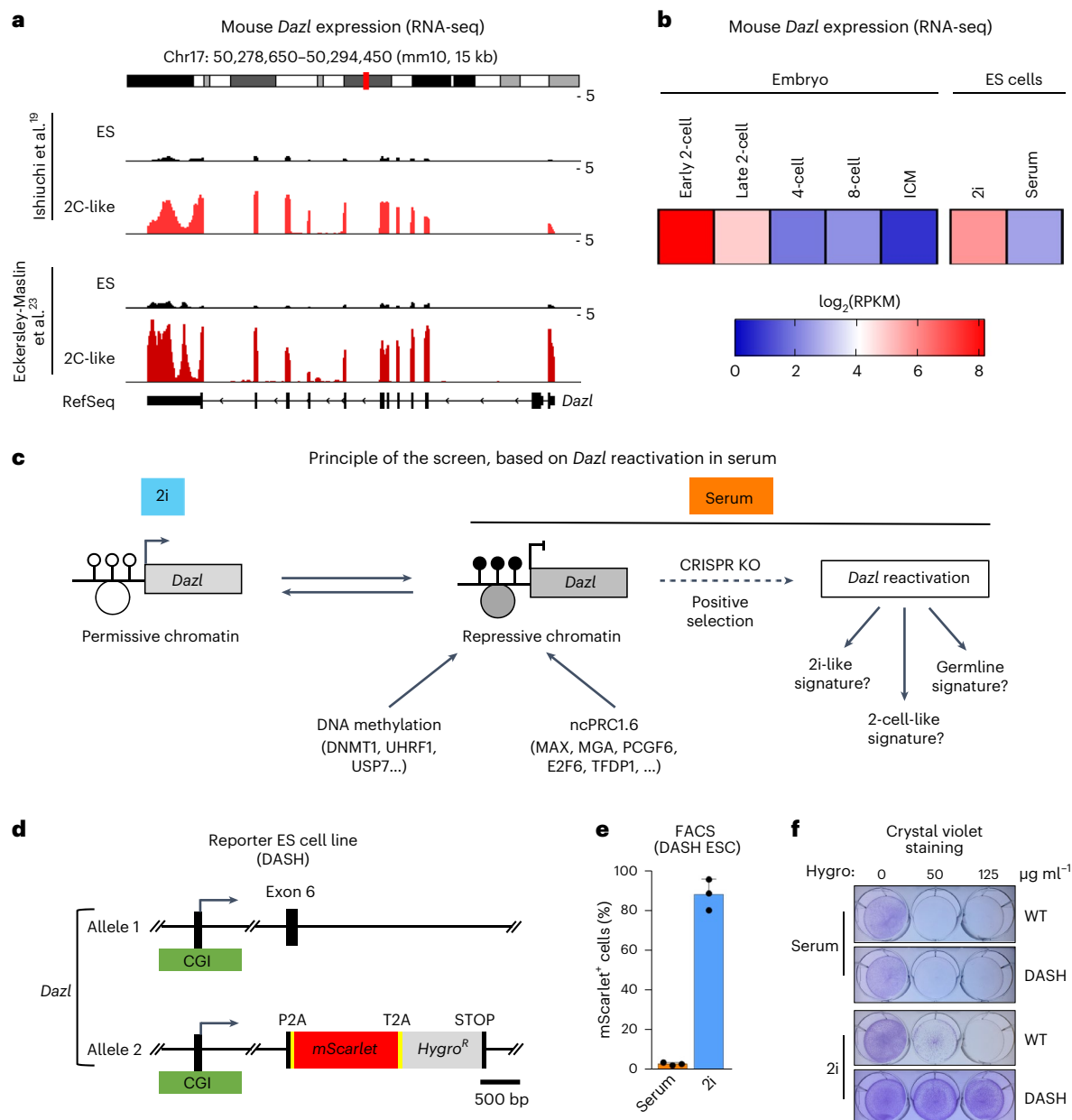


Fig. 1 | *Dazl* as a reporter gene for the 2-cell stage, screen design and generation of the *Dazl*-mScarlet-Hygro^R (DASH) reporter cell line. **a**, *Dazl* is highly expressed in 2CLCs in vitro (analysis of RNA-seq data from the indicated references). RefSeq, Reference Sequence database. **b**, *Dazl* expression in 2i is similar to the late 2-cell embryo, while expression in serum is similar to the 8-cell stage (embryo data from GSE66390 (ref. 65), ESC data, this study). **c**, Principle of the screen. Knocking out factors essential for repression should trigger gene reactivation in serum. Unmethylated CpG, white lollipop; methylated CpG, black

lollipop. Permissive chromatin, white circle; repressive chromatin, gray circle. **d**, The DASH reporter ES cell line. A reporter cassette is inserted into one of the *Dazl* alleles, it encodes two proteins separated by 2 A self-cleaving peptides (P2A, T2A): the red fluorescent mScarlet and the Hygromycin-resistance enzyme (Hygro^R). **e**, DASH cells express mScarlet in 2i, but not in serum. FACS analysis of 50,000 cells per condition from $n = 3$ replicates. Data are presented as mean values \pm s.d. **f**, DASH cells are resistant to Hygromycin in 2i, but not in serum. Surviving cells are stained with crystal violet.

'heparan sulfate-glycosaminoglycan synthesis' pathway decreases FGFR or MAPK signaling, and thus increases the proportion of naïve cells³³; therefore, the hits in this cluster likely affect the differentiation process and were not analyzed further. For the same reason, the 'FGFR' and 'TGF β -Wnt signaling' clusters were also discarded. To identify direct regulators of gene expression, we focused on nuclear factors, excluding proteins already connected to DNA methylation or ncPRC1.6 (refs. 15,20). On the basis of DNA methylation and gene expression analyses, we selected four of the remaining candidates for further experiments: the lysine demethylase KDM5C; the E3 ubiquitin ligase adaptor SPOP; the Zinc-Finger and BTB-containing protein ZBTB14

and MCM3AP (GANP³⁴ in humans), the scaffolding subunit of TREX-2, a complex that couples gene expression and mRNA export^{35–37}.

We carried out two more experiments to rule out a 'pseudo-2i' state, in which cells would fail to respond to serum and differentiation cues. First, we measured the expression of genes expressed in 2i (*Tcf2l1*; *Prdm14*; *Kit*), or in fetal bovine serum (FBS) (*Id1*; *Dnmt3l*, *Myc*). We observed that all four mutants resembled FBS-grown embryonic stem (ES) cells, and not 2i-grown cells (Extended Data Fig. 2f). Second, we let the serum-grown cells undergo spontaneous differentiation by removing leukemia inhibitory factor (LIF). While *Tcf7l1* mutant cells, as expected, failed to differentiate, the colonies of these four

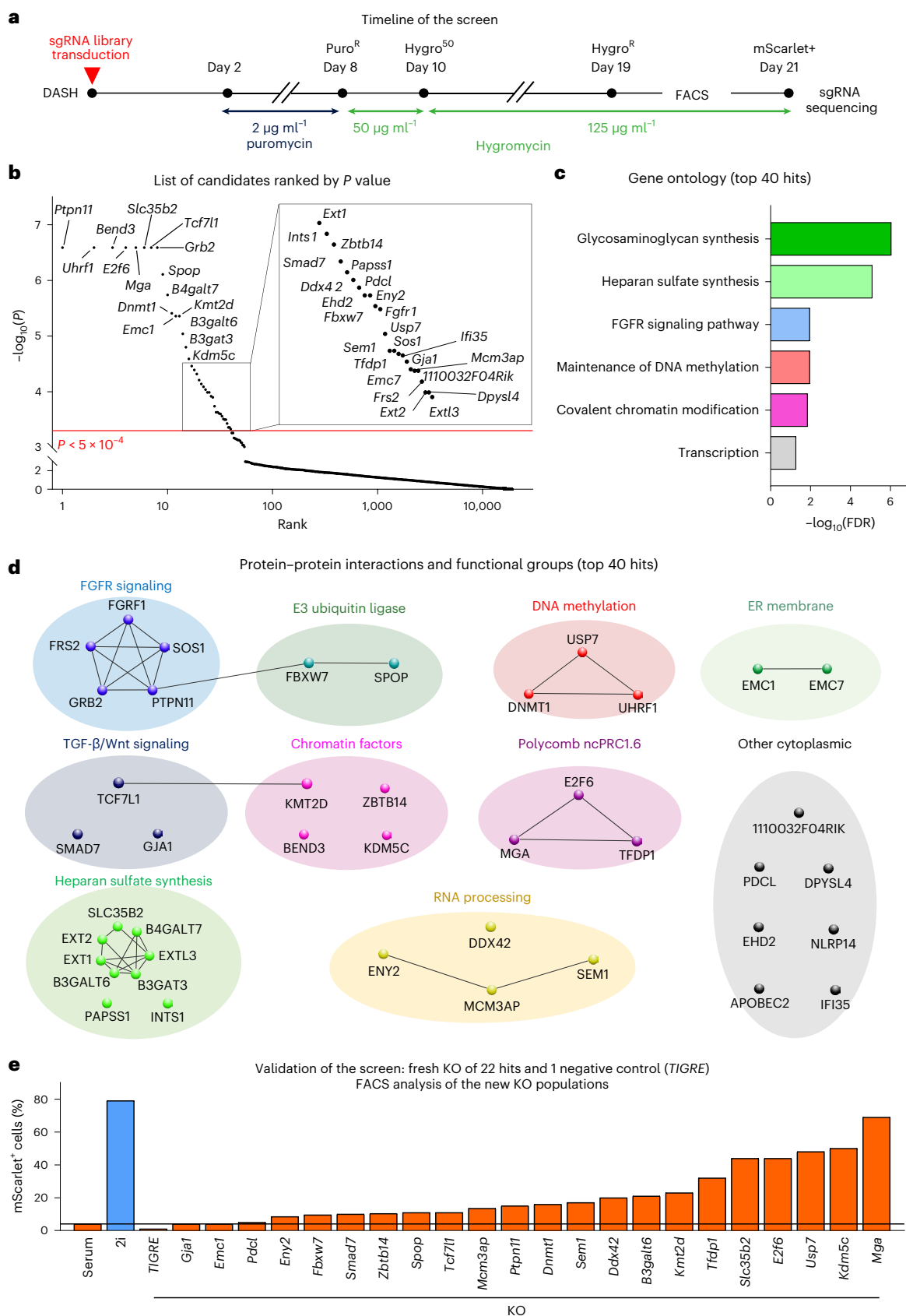


Fig. 2 | A genome-wide CRISPR KO screen for *Dazl* gene reactivation in the DASH reporter line. a, Timeline of the screen. **b**, Ranking of the hits by one-sided P value (MAGECK, robust rank aggregation). The top 40 hits ($P < 0.0005$) are above the red line. **c**, Gene ontology (GO) terms (Biological Process) significantly enriched among the top 40 hits. **d**, Protein-protein interaction network

clustering of the top 40 hits, generated by STRING analysis. **e**, FACS analysis (50,000 cells per condition) of mScarlet expression after KO of the indicated candidate genes. The parental DASH cells (cultured in serum or 2i), and cells expressing a sgRNA targeting the inert *TIGRE* locus served as controls.

candidates (*Mcm3ap*, *Spop*, *Kdm5c* and *Zbtb14*) were rounded in serum and on LIF removal became irregularly shaped, a mark of differentiation (Extended Data Fig. 2g). Besides, during spontaneous differentiation, the KO cells repressed pluripotency genes (*Prdm14*, *Pou5f1*) and induced the differentiation marker *Fgf5* to a similar extent to their wild-type (WT) counterpart (Extended Data Fig. 2h). These data rule out the possibility that the KO cells are locked in a 2i-like state. We therefore continued our characterization of the four candidates: *Mcm3ap*, *Spop*, *Kdm5c* and *Zbtb14*.

Generation and rescue of individual KO clones

Next, we isolated three individual KO clones for each of the four genes of interest. All had loss-of-function mutations (Fig. 3a and Extended Data Fig. 3a), and reactivated mScarlet, to various extents (Fig. 3b). The clones also expressed the untagged allele of *Dazl* at the RNA (Extended Data Fig. 3b) and protein level (Fig. 3c).

Genetic rescue was tested using piggyBac transposons expressing V5-tagged constructs³⁸. In all cases, expression of the functional gene in the corresponding KO decreased the number of mScarlet-positive cells (Fig. 3d), rendered the cells sensitive to Hygromycin (Fig. 3e) and silenced the expression of the *Dazl* RNA (Extended Data Fig. 3b) and protein (Fig. 3f). These experiments show that loss-of-function mutations in *Kdm5c*, *Mcm3ap*, *Spop* and *Zbtb14* caused the reactivation of *Dazl*.

The clones were further characterized by whole-genome bisulfite sequencing (WGBS) at ten times median coverage (Extended Data Fig. 3c). A PCA analysis showed that all four mutants clustered together and away from the serum-grown or 2i-grown WT cells (Extended Data Fig. 3d). The level of CpG methylation was reduced from 60% in WT to roughly 55% in the mutants (Fig. 3g and Extended Data Fig. 3e,f), the difference being statistically significant for the *Mcm3ap* and *Spop* KOs. Nevertheless, the amount of DNA methylation lost in the KO cells is modest, unlike that seen in 2i cells (Fig. 3g and Extended Data Fig. 3e,f), or the 30% decrease reported in 2CLCs (ref. 23). We conclude that the four KOs we identified do not cause a widespread loss of DNA methylation. We then turned to a transcriptomic approach.

The KOs induce a 2-cell-like transcriptional signature

mRNA-seq was performed on each of the KOs (three distinct clones each); they showed roughly 1,000 to 1,200 differentially expressed genes (Fig. 4a), split roughly equally between up- and downregulated genes (Extended Data Fig. 4a). As expected, *Dazl* was upregulated in all mutants (Fig. 4a and Extended Data Fig. 4b). Among 92 validated 2CLC markers¹⁵, between 21 and 32 were differentially expressed in the different KOs, including *Zscan4* and *Taf7l* (Fig. 4a and Extended Data Fig. 4b). The master regulator of the 2CLC state, *Dux*, was also induced in all the KOs (Fig. 4a). The results were validated by RT-qPCR, and we verified that the genes induced in the KOs return to basal level on rescue (Extended Data Fig. 4c).

The four KOs have 201 commonly upregulated genes (Fig. 4b), and 2CLC markers are enriched in this set (Fig. 4c and Extended Data Fig. 4d). Correspondingly, the 201 shared upregulated genes are more expressed at the 2-cell stage than at other early developmental stages (Fig. 4d). This provides evidence that a 2CLC signature is present in each of the four KOs.

An additional feature of 2CLCs is a reduction of oxygen consumption, glycolytic activity and ROS accumulation²⁸. In agreement, gene set enrichment analysis (GSEA) with Hallmark pathways showed that ‘Oxidative phosphorylation’, ‘Glycolysis’ and ‘Reactive Oxygen Species’ were among the top significantly depleted pathways in the combined ‘4KO’ profile (Extended Data Fig. 4e).

Two more sets of data support the possibility that the four KO lines are 2-cell-like. First, in the absence of any selection, all KO populations showed more ZSCAN4-positive cells than the WT DASH cells (Extended Data Fig. 4f). Second, the marker that is most commonly used to identify 2CLCs is the transcriptional activation of MERVL repeats^{8,9}. Therefore we tested whether the four genes we have identified also repress a MERVL-LTR reporter. We used the cell line tbG4 (ref. 19) and, as a positive control of MERVL induction, we treated the cells with sodium acetate. Each of the four small-interfering RNA (siRNA) knockdowns (against *Kdm5c*, *Mcm3ap*, *Spop* or *Zbtb14*) also induced the MERVL-LTR-GFP reporter (Extended Data Fig. 4g). Whereas the percentages of positive cells are small, they are in line with previous reports¹⁹. Altogether, these data show that the four KOs revealed by the screen reactivate a gene expression signature that is typical of 2CLCs.

Activation of repeat elements and chimeric transcripts

Next, we explored the expression of repeat elements (short interspersed nuclear elements, long interspersed nuclear elements and endogenous retroviruses). Our data showed an induction of MERVL repeats—and of their long terminal repeats (LTRs) MT2_{mm}—but of few other repeats in *Kdm5c* KO, *Mcm3ap* KO, *Spop* KO and *Zbtb14* KO clones (Fig. 4e).

The individual repeats activated in our four KOs are similar to the repeats reported to be upregulated in 2CLCs (refs. 7,19,23), and to those induced in the 2-cell embryo in vivo³⁹ (Fig. 4f). Many of these upregulated MERVL-LTR copies led to the expression of chimeric transcripts, known to positively regulate the 2CLC state¹⁰. Finally, we validated our findings with RT-qPCR (Extended Data Fig. 4c), and verified that ZSCAN4 and MERVL-Gag were induced at the protein level in the four KOs and silenced again when the KOs were rescued (Fig. 4g).

Epistasis with Dux, Dppa2 and Trp53

We next sought to connect our findings to known 2CLC regulators. First, we knocked down *Dux*³ (Fig. 5a); this caused the number of mScarlet-positive (Fig. 5b) and Hygromycin-resistant cells (Fig. 5c) to decrease to WT levels in the *Mcm3ap*, *Spop* and *Zbtb14* KOs. By contrast, the *Kdm5c* KO/sh*Dux* population maintained a large fraction of mScarlet-positive and Hygromycin-resistant cells (Fig. 5b,c). The knockdown also caused a major decrease of 2CLC markers (*Dazl*, *Zscan4*, *Usp17le*, MERVL) in all KOs, except for the *Kdm5c* KO, in which *Dazl* expression stayed significantly higher than in WT cells (Fig. 5d).

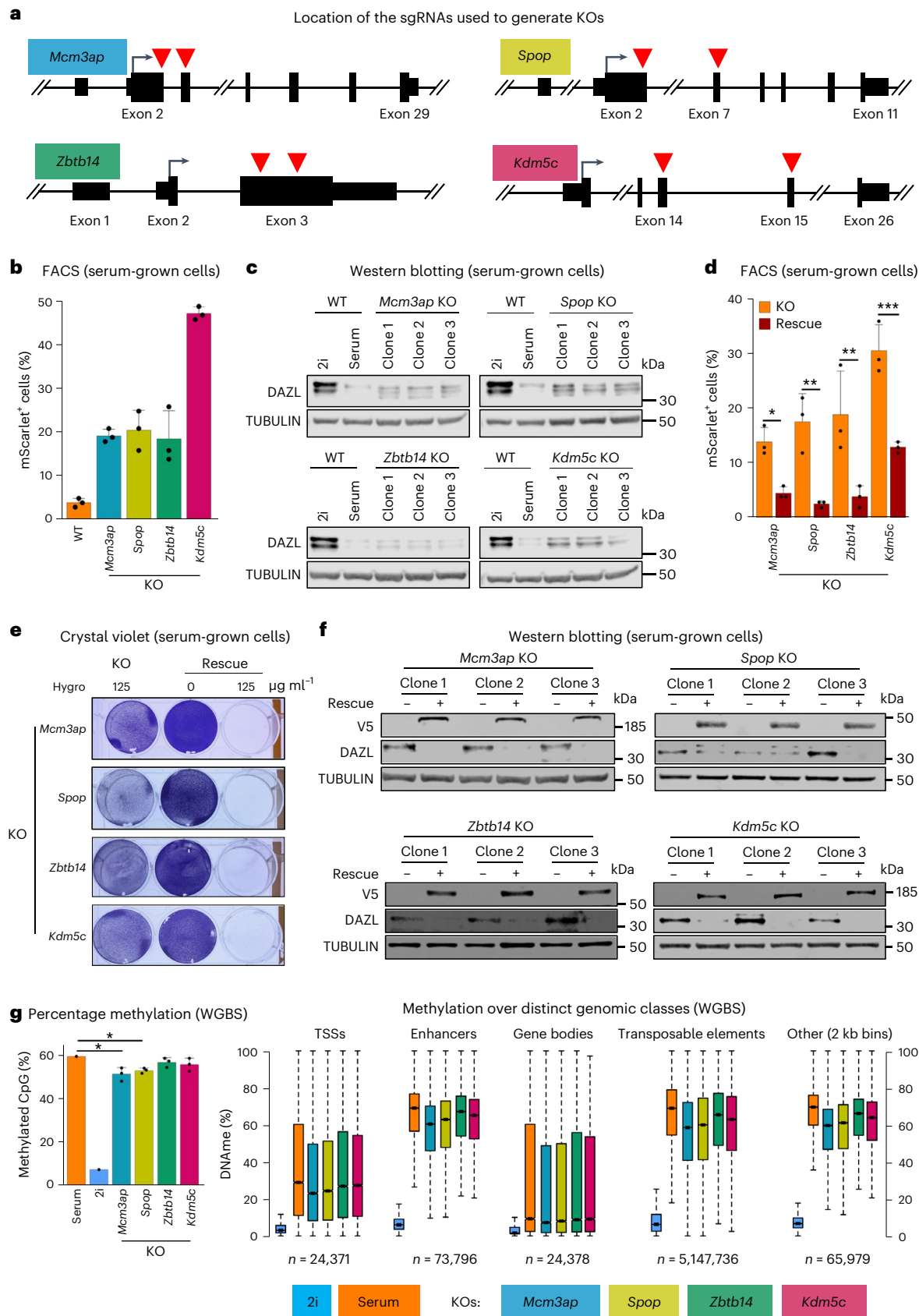
Similarly, DPPA2 was necessary for 2CLC gene induction in all KOs, the exception being *Kdm5c* KO, in which *Dazl* induction persisted even after DPPA2 removal (Fig. 5e–h). Last, we investigated p53 (ref. 12) (Extended Data Fig. 5a). Its knockdown (Extended Data Fig. 5b,c) did not affect the expression of *Dazl*, *Dux*, *Usp17le* or MERVL in any of the four KOs (Extended Data Fig. 5d), showing that they act independently of p53.

Fig. 3 | Four loss-of-function mutants that induce *Dazl* expression: *Mcm3ap* KO, *Spop* KO, *Zbtb14* KO and *Kdm5c* KO. a, Depiction of the gene structures and position of the sgRNAs used to generate mutations (red arrowheads). **b**, FACS analysis (50,000 cells per condition) of mScarlet expression after KO of the indicated genes. **c**, Western blot analysis of *Dazl* expression in the indicated cells. **d**, FACS analysis (50,000 cells per condition): genetic rescue of each KO suppresses mScarlet expression. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-sided Holm–Sidak post hoc test following analysis of variance (ANOVA)). **e**, Genetic rescue of each KO suppresses Hygromycin resistance. **f**, Western blot

analysis: genetic rescue of each KO suppresses *DAZL* protein induction. Each rescue construct contained a V5 tag. **g**, WGBS detects varying decreases of DNA methylation in the different KO populations. The left panel shows the global levels of CpG methylation. * $P < 0.05$ (two-tailed t -test). Right panel shows analysis at the indicated genomic regions. In the boxplots, the thick line indicates the median, the box limits indicate the upper and lower quartiles and the whiskers extend to minimum and maximum values. The color code is indicated below the figure. Data from $n = 3$ independent KO clones are presented as mean values \pm s.d. for **b**, **d** and **g**.

To summarize, removing MCM3AP, SPOOP or ZBTB14 induces 2CLC gene expression in a manner that is DPPA2 and DUX-dependent, showing that they act upstream of these transcriptional activators. Removal of KDM5C has a more complex, dual effect: it induces *Zscan4*, *Usp17*

and MERVL in a DUX- and DPPA2-dependent manner, again arguing for KDM5C being an upstream regulator. However, removing KDM5C induces *Dazl* even when DUX or DPPA2 are absent, suggesting that KDM5C is repressing the *Dazl* promoter directly.



Last, we reanalyzed single-cell RNA-sequencing (RNA-seq) data of ES cells sorted according to ZCAN4 and MERVL expression²³ (Extended Data Fig. 5e). We observe that the double-positive ZSCAN4⁺/MERVL⁺ cells express higher levels of *Dazl* than single-positive or double-negative cells, and also that they express lower levels of *Kdm5c*, *Spop*, *Zbtb14* and *Mcm3ap*, showing that a transcriptional down-regulation of the four hits occurs in 2CLCs (Extended Data Fig. 5e), reinforcing the idea that they are negative regulators of this cellular state.

SPOP binds DPPA2

The results above show that SPOP acts upstream of DPPA2. An interaction between SPOP and DPPA2 has been reported in yeast two-hybrid⁴⁰, so we tested their interaction and indeed the co-immunoprecipitation between V5-tagged SPOP and FLAG-tagged DPPA2 was positive (Fig. 6a). As SPOP is an E3 ubiquitin ligase adaptor, we speculated that it might target DPPA2 for degradation; however, the removal of SPOP by siRNA did not increase the level of DPPA2 (Fig. 6b). These data argue that SPOP binds DPPA2, but does not target a large fraction of the protein for degradation.

ZBTB14 binds nonmethylated CpG islands

We next focused on ZBTB14, by performing CUT&RUN in the *Zbtb14* KO line expressing V5-ZBTB14 (Fig. 6e,f). Of the 2832 ZBTB14 peaks, 75% were in promoters (Fig. 6c). RNA-seq showed similar numbers of genes upregulated or downregulated in the *Zbtb14* KO cells (639 versus 617, Extended Data Fig. 6a). 2CLC genes (*Zscan4a/b/c/f*, *Usp17a/b/e...*) are among the induced genes. However, the overlap between promoters bound by ZBTB14, and genes differentially expressed in the *Zbtb14* KO was limited (Extended Data Fig. 6b). We carried out de novo motif discovery on the ZBTB14-bound peaks and recovered the motif GCGCGCGCGC, which is nearly identical to the motif identified in vitro with recombinant protein⁴¹ (Fig. 6d). As this motif is CG-rich, we tested whether the promoters bound by ZBTB14 are CpG islands, and indeed they are overwhelmingly so (Fig. 6e). Certain CpG-binding zinc-finger proteins such as CFP1 (ref. 42) or BANF⁴³ are repelled by methylation, and in vitro experiments suggested that it might be the case for ZBTB14 (ref. 31). We therefore asked whether methylated CpG islands are bound by ZBTB14, and found that the answer was negative (Fig. 6f,g). One of the promoters most bound by ZBTB14 is the *Zbtb14* promoter itself; it contains an unmethylated CpG island with eight instances of the ZBTB14 binding motif (Fig. 6h). By contrast, the *Dazl* CpG island, contains a single ZBTB14 binding motif in the methylated portion of the CpG island and it is not bound by ZBTB14 (Fig. 6i).

These data indicate that ZBTB14 most likely regulates *Dazl* expression and the 2C state indirectly. We used the CladeOScope tool⁴⁴ to determine which genes have most closely coevolved with *Zbtb14*, and found that *Zscan4* was in the top five (Extended Data Fig. 6c). Conversely, *Zbtb14* is one of the top five genes that have most closely coevolved with *Zscan4*. This shared evolutionary history is suggestive of functional interplay between ZBTB14 and the key 2CLC regulator ZSCAN4.

KDM5C binds ncPRC1.6 and has a noncatalytic role

We next investigated the role of KDM5C. Chromatin immunoprecipitation with sequencing (ChIP-seq) on endogenous KDM5C in DASH cells

revealed a peak over the *Dazl* promoter (Fig. 7a). More generally, we observed roughly 800 gene promoters bound by KDM5C, of which many belonged to germline genes such as *Taf7l* and *Ddx4* (Extended Data Fig. 7a) and, more specifically, to germline genes that have a *Dazl*-like pattern of regulation¹⁸ (Extended Data Fig. 7b). This represents a description of endogenous KDM5C distribution in mouse ESCs, refining data obtained with overexpression⁴⁵. Genes such as *Dazl* or *Taf7l*, that are both bound and repressed by KDM5C, represent only a minority of the targets (Fig. 7b) and they are highly enriched in germline genes (Extended Data Fig. 7c).

A motif discovery analysis on the KDM5C peaks yielded as the top result the sequence CACGTG (Fig. 7c). This is an E-box that can be bound by MAX-MGA, a DNA-binding module within ncPRC1.6, therefore suggesting that many sites directly bound by KDM5C are also bound by ncPRC1.6. The examination of MAX ChIP-seq data validated this hypothesis, as we found that 77% of KDM5C targets were also MAX targets (Fig. 7d). We next tested the possibility of a physical interaction between KDM5C and the ncPRC1.6 complex, and co-immunoprecipitation experiments showed that KDM5C does indeed interact (directly or indirectly) with the ncPRC1.6 component PCGF6 (Fig. 7e).

These data show that most genes bound by KDM5C are also bound by ncPRC1.6, but does this also apply to genes that are repressed by KDM5C even though they are not directly bound? We assessed this by comparing the transcriptomes of cells lacking *Kdm5c* or *Pcgf6* (a component of ncPRC1.6), and indeed found that there was extensive overlap: roughly 40% of genes repressed by KDM5C are also repressed by ncPRC1.6 (Fig. 7f), including *Dux* and other 2CLC markers (*Usp17l* cluster, *Zscan4* cluster, *Zfp352*,...). Our ChIP-seq data did not support direct binding of KDM5C to the *Dux* locus, so we hypothesize that an indirect recruitment occurs, maybe via ncPRC1.6 (Fig. 7g).

Next, we examined the requirement for the catalytic activity of KDM5C in the transcriptional repression process. *Kdm5c* KO cells rescued either with the WT version of the enzyme or with the catalytically inactive mutant H514A (ref. 46). As observed previously (Fig. 3), rescue with the WT form strongly reduced the number of mScarlet-positive cells emerging in the *Kdm5c* KO population. The catalytically inactive mutant had the same effect (Fig. 7h). Therefore, repression of *Dazl* expression does not require lysine demethylation by KDM5C.

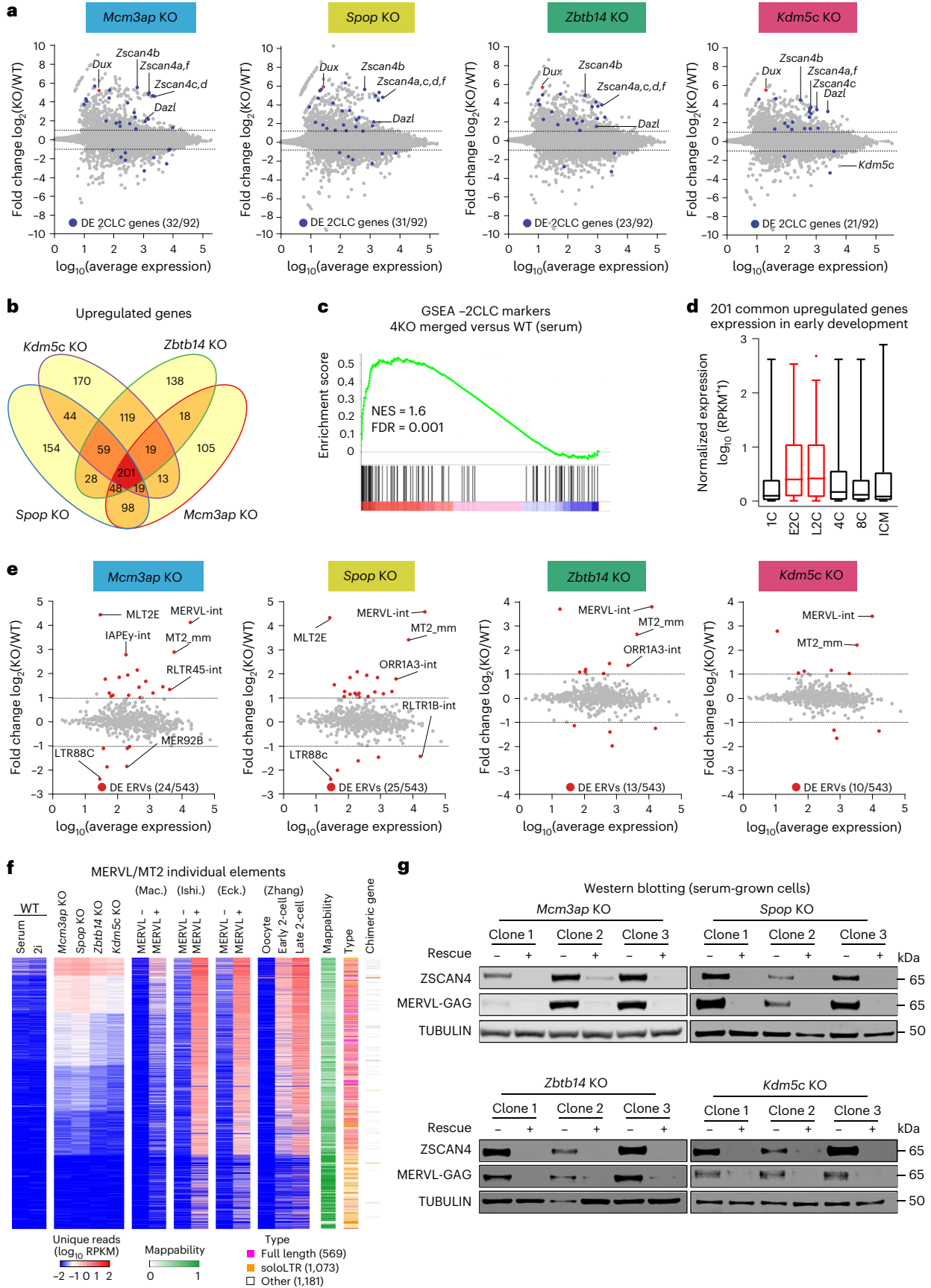
In addition to repressing *Dazl*, KDM5C also inhibits the 2CLC state more generally, and does so in a DUX-dependent manner (Fig. 5). The WT form of KDM5C efficiently repressed *Dux* in *Kdm5c* KO cells, as it repressed MERVL, *Zscan4* and *Usp17le* (Fig. 7i). Again, catalytically inactive KDM5C repressed just as efficiently as WT (Fig. 7i), an observation we confirmed by western blotting (Fig. 7j). A parallel finding was that KDM5C, whether catalytically active or not, had no discernible effect on the global levels of H3K4me1/2/3 (Extended Data Fig. 7d). To conclude, our data indicate a surprising function of KDM5C in our experimental context: catalysis-independent repression of *Dazl* and *Dux*, mediated in part by direct binding to chromatin, and in part by events that do not involve direct binding but may rely on KDM5C interaction with ncPRC1.6.

Discussion

We have used a CRISPR KO screen in mouse ESCs to identify negative regulators of a 2CLC marker, *Dazl*. The screen yielded many of the

Fig. 4 | Induction of a 2-cell-like signature in the *Mcm3ap*, *Spop*, *Zbtb14* and *Kdm5c* KOs. a, RNA-seq analysis of the indicated KOs. The MA plots are shown, with *Dux* depicted in red, and the differentially expressed (DE) genes of the 2CLC signature¹⁵ depicted in blue. **b**, Venn diagram of upregulated genes (RNA-seq) after KO of the indicated genes. **c**, The 2CLC signature¹⁵ is significantly enriched in the transcriptome of the four KOs. **d**, Boxplots of normalized expression (RPKM) for the 201 shared upregulated genes in early developmental stages. In the boxplots, the line indicates the median, the box limits indicate the upper and lower quartiles, and the whiskers extend to 1.5 IQR from the quartiles. Transcriptome data was obtained from GSE66390 (ref. 65). E2C, early 2-cell; L2C,

late 2-cell. **e**, MA plots for repeated elements in our RNA-seq data. **f**, Heatmap depicting the normalized expression (RPKM) of MERVL/MT2 single elements (full length, solo LTR or others) in four candidates KO and parental DASH ESCs. Mappability score and the presence of chimeric genes are indicated. Published data of MERVL-positive cells and 2C embryos were included for comparison and were obtained from GSE33923 (ref. 7, Macfarlan (Mac.)); E-MTAB-2684 (ref. 19, Ishiuchi (Ishi.)); GSE75751 (ref. 23) (Eckersley-Maslin (Eck.)) and GSE71434 (ref. 66, Zhang). **g**, Induction of the ZSCAN4 and MERVL-GAG proteins in the KOs, and inhibition in the rescue clones. Data from $n = 3$ independent KO clones are presented for **a**, **e** and **f**.



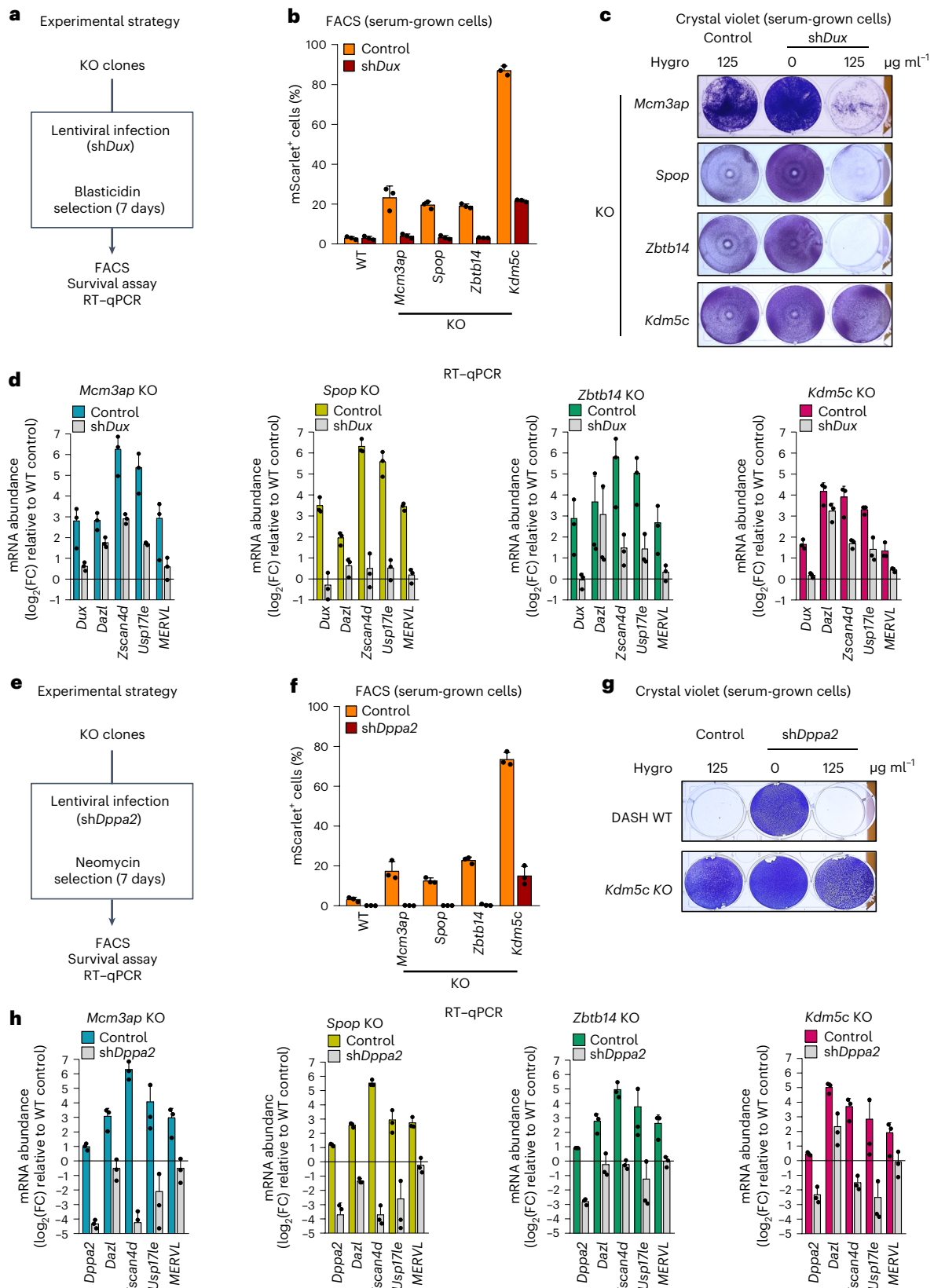


Fig. 5 | Epistasis analyses show that KDM5C removal activates *Dazl* independently of *Dux* and *Dppa2*. **a**, Experimental scheme for *Dux* depletion. **b**, FACS analysis (50,000 cells per condition) after sh*Dux* shows that depleting *Dux* brings back the expression of mScarlet to WT levels in all KOs, except for *Kdm5c* KO. **c**, Depleting *Dux* cancels Hygromycin resistance in all KOs, except for *Kdm5c* KO. **d**, RT-qPCR analysis: the induction of 2CLC markers is *Dux* dependent, except for *Dazl* induction in the *Kdm5c* KO. **e**, Experimental scheme for *Dppa2*

depletion. **f**, FACS analysis (50,000 cells per condition) after sh*Dppa2* shows that depleting *Dppa2* brings back the expression of mScarlet to WT levels in all KOs, except for *Kdm5c* KO. **g**, Depleting *Dppa2* cancels Hygromycin resistance in all KOs, except for *Kdm5c* KO. **h**, RT-qPCR analysis: the induction of 2CLC markers is *Dppa2* dependent, except for *Dazl* induction in the *Kdm5c* KO. Data from $n = 3$ independent KO clones are presented as mean values \pm s.d. for **b**, **d**, **f** and **h**. FC, fold change.

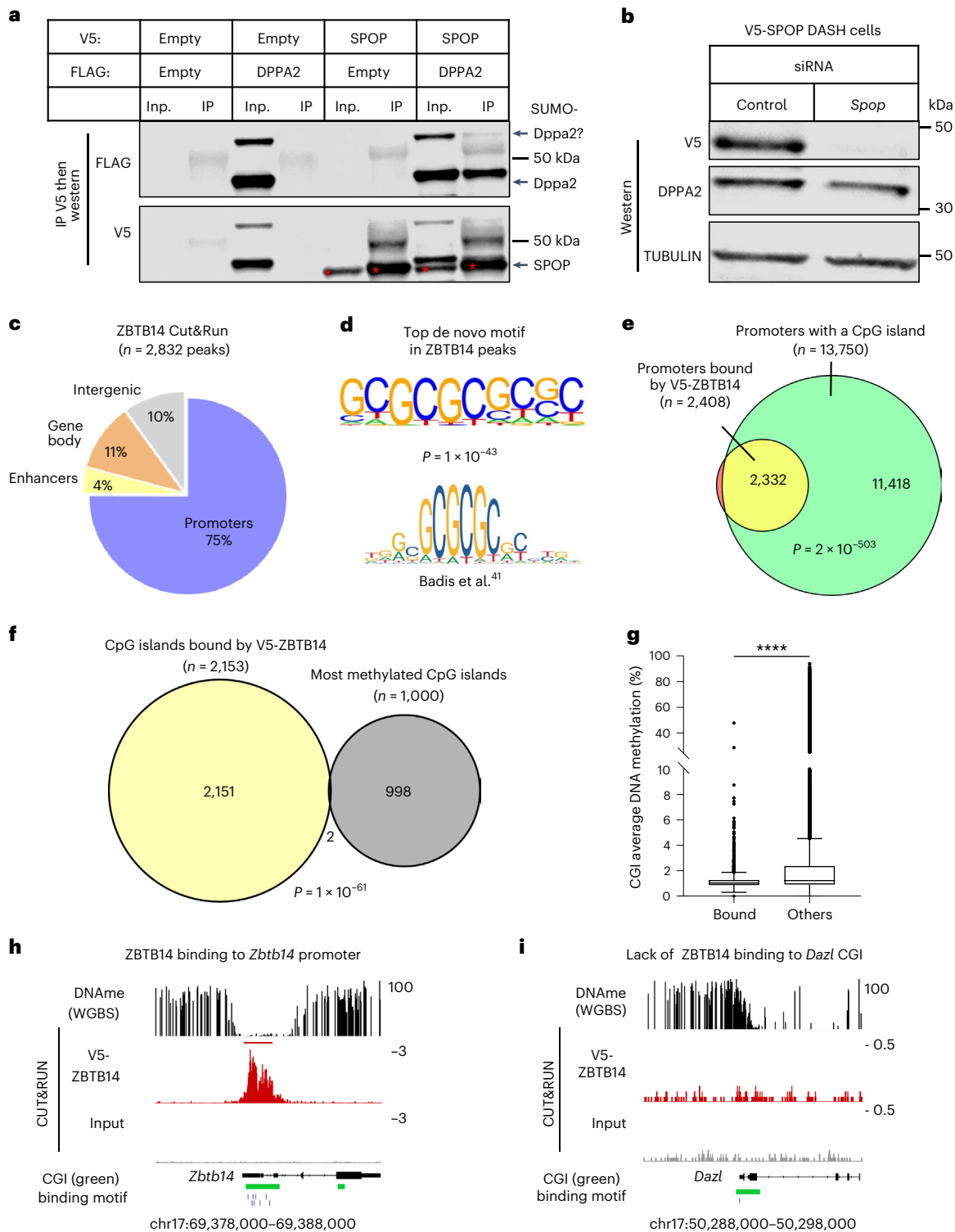


Fig. 6 | Mechanistic studies on SPOP and ZBTB14. **a**, SPOP and DPPA2 interact. Result of a co-immunoprecipitation experiment after transfection of the indicated plasmids in DASH cells. The SPOP band is shown with a red asterisk. The other bands correspond to DPPA2, which is also detected because the anti-DPPA2 and anti-SPOP antibodies are from the same species and DPPA2 was probed first. Inp., input; IP, immunoprecipitation. **b**, SPOP does not destabilize DPPA2. Western blotting on the endogenous DPPA2 protein in a V5-SPOP-expressing DASH line, after the indicated siRNA. **c**, Most of the ZBTB14 binding peaks occur in promoters. **d**, De novo motif discovery in ZBTB14 peaks reveals a CG-rich motif highly similar to the one reported in vitro by Badis et al.⁴¹. Significance is the *P* value computed by homer2. **e**, The promoters bound by

ZBTB14 are overwhelmingly CpG islands. **f**, The most highly methylated CpG islands are not bound by ZBTB14. **g**, Quantitative analysis of CpG methylation at CGIs (CpG islands) that are bound (*n* = 2,153), or not bound (*n* = 13,715), by ZBTB14. In the boxplots, the line indicates the median, the box limits indicate the upper and lower quartiles and the whiskers extend to 1.5 IQR from the quartiles. *****P* < 0.0001 (two-tailed Mann–Whitney test). **h**, Illustration of the CUT&RUN results for V5-ZBTB14: the protein binds the CGI of its own gene, which is unmethylated and contains multiple ZBTB14 consensus sites (vertical tick marks). **i**, ZBTB14 binding is not detected at the *Dazl* CpG island, which is partly methylated. Significance: hypergeometric test for **e–f**.

expected hits (DNMT1, UHRF1, E2F6...), as well as a number of other hits. A prioritization strategy led to the identification of four factors in which loss-of-function mutations reactivate not only *Dazl*, but a broader 2CLC signature, suggesting that they are regulators of the 2CLC status. Epigenomics, transcriptomics and functional studies allow us to place these hits relative to known actors, and to extend the conceptual framework of 2CLC identity, with likely relevance to physiological totipotency.

A genetic screen to identify regulators of the 2CLC state

Our knowledge of the 2CLC state has been greatly advanced by genetic screens in which MERVL and/or *Zscan4* reporter genes were used as proxy for the activation of the 2CLC signature. We opted to use a distinct reporter, *Dazl*. We note that *Dazl* is not as specific of the 2CLC state as MERVL or *Zscan4* are, and indeed we found many hits that were unrelated to the 2CLC state but instead promoted the re-expression of *Dazl* because they placed the cell in a 2i-like status (Fig. 2d).

Nevertheless, our procedure coupled to adequate secondary screens allowed us to recover regulators of the 2CLC state. We identified previously uncharacterized regulators of the 2CLC signature, but there was also overlap between the hits we obtained and those found in previous approaches, including a recent CRISPR screen based on a *Dux* induction system²⁰. Some known inhibitors of the 2CLC state failed to reach the top 40, but still scored highly in our candidate list: the SUMO E3-ligase PIAS4 (ref. 47) ranked no. 41; the chromatin remodeler p400 (ref. 15) ranked no. 52; the histone demethylase LSD1¹⁷ ranked no. 111 and the ncPRC1 subunits¹⁵ RING1B and RYBP ranked nos. 100 and 245, respectively. This suggests that relaxing our statistical cutoff may still yield important regulators of totipotency. In summary, our experimental scheme allowed us to fruitfully recover some known, but also some new, regulators of the 2CLC state.

One potential shortcoming of our KO approach (as opposed to siRNA or CRISPRi) is that loss of genes that are essential or near-essential is counter-selected. This likely explains why some published inhibitors of the 2CLC state, such as the histone chaperone CAF-1 (ref. 19) or the transcription factor TRIM28 (ref. 3) did not rank high in our screen. Another constraint of the screening approach is that it uses a single reporter gene. We selected *Dazl* because it is expressed at the 2-cell-like stage^{15,20}, and its transcription intensity is compatible with antibiotic selection⁴⁸. This reporter yielded hits that had not been reported using other systems such as MERVL-based reporters^{15,20}, showing its complementarity to these approaches. However, *Dazl* is also expressed in the germline, a property shared with other 2CLC markers such as *Spz1* or *Taf7l*. For this reason, we had to find ways to focus our attention on those hits that regulate the 2CLC state generally, and not just *Dazl*, or not just germline genes. A positive corollary, however, is that our list of candidates may reveal interesting regulators of germline gene expression that could be explored in the future. Keeping these caveats in mind, our reporter strain can be readily used in the future for additional screens. For instance, given the known roles of lncRNAs

in the 2CLC (refs. 49,50), a genome-wide CRISPRi or CRISPRoff screen on these elements may be warranted.

Possible mechanisms of action of the four factors

KDM5C demethylates H3K4me2/3 and is therefore a repressive demethylase^{46,51,52}. We find that removing KDM5C from ESCs eases their conversion into 2CLCs. The protein seems to have a two-pronged function. First, KDM5C directly binds and represses the promoter of certain 2CLC markers, including *Dazl* and *Taf7l*. This first observation fits with data obtained in neurons⁵³. Second, KDM5C represses the *Dux* locus, in a catalytic-independent manner. It has been observed before that lysine demethylases can have catalysis-independent functions⁵⁴, but we are not aware of any such report for KDM5C. As KDM5C interacts with E2F6 in HeLa cells⁵¹, and with PCGF6 in mouse dendritic cells⁵⁵, we hypothesize that KDM5C might be recruited by ncPRC1.6 at some of the many targets they share, including *Dux*.

Three subunits of the TREX-2 complex were recovered in our screen: MCM3AP-GANP, ENY2 and SEM1. TREX-2 has pleiotropic roles: it couples gene expression to mRNA export³⁵⁻³⁷ and it also regulates the outcome of DNA repair⁵⁶. Future experiments, for instance with separation-of-function mutants, will be required to determine which of these functions contribute to repressing the 2CLC state in ESCs.

SPOP is the adaptor module for an E3 ubiquitin ligase complex that promotes protein degradation³², and we show that SPOP interacts with DPPA2, yet this does not seem to strongly limit the amount of DPPA2 in ES cells. It is possible that SPOP degrades a small but critical subpopulation of DPPA2 molecules; alternatively SPOP might inhibit DPPA2 by sequestering it.

ATR-dependent replication stress triggers *Dux* expression⁵⁷, and ZBTB14 has been recently shown to stabilize the RPA-ATR-ATRIP complex at stalled replication forks⁵⁸. However, the loss of ZBTB14 leads to decreased ATR signaling, therefore it seems unlikely that the induction of *Dux* and of the 2CLC signature in the *Zbtb14* KO involves ATR signaling. Alternatively, given the recent discovery that replication speed controls the emergence of 2CLCs (ref. 59), ZBTB14 might be involved in this process.

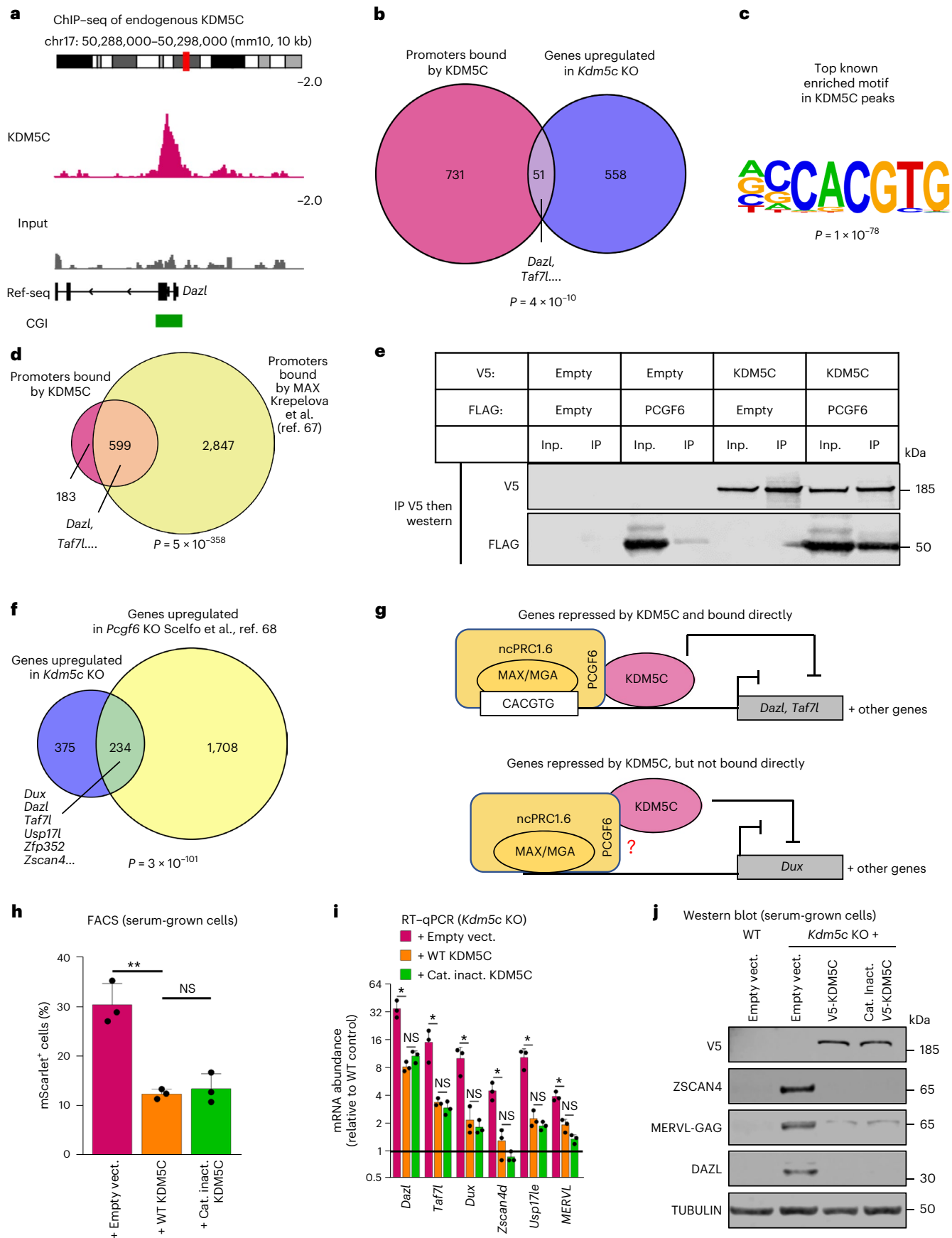
Relevance to totipotency, disease and therapy

Our work identifies new regulators of the 2CLC state *in vitro*. Many of the discoveries obtained using 2CLCs have proved to also apply to the 2-cell embryo^{1,2,8,9}, therefore it is possible that MCM3AP, KDM5C, SPOP and ZBTB14, also regulate totipotency *in vivo*. The formal demonstration of this assumption will require future experiments on early mouse embryos.

Our findings have interesting consequences for human pathologies. *Dux* expression is normally restricted to a specific developmental window, and inappropriate expression of *Dux* (and its downstream targets) in skeletal muscle leads to the human disease facioscapulo-humeral dystrophy⁵. Some mutations that induce *Dux* expression in ESCs seem to also induce it in muscle cells: this is at least the case for

Fig. 7 | Catalysis-independent role of KDM5C. **a**, KDM5C binds the *Dazl* promoter. ChIP-seq of endogenous KDM5C in DASH cells. **b**, KDM5C is necessary for the repression of some of its promoter targets, including *Dazl*. **c**, The top known binding motif called by HOMER for KDM5C peaks is the E-box, which is the target of the MAX-MGA dimer. Significance: *P* value computed by homer2. **d**, 77% (599 of 782) of all promoters bound by KDM5C are also bound by MAX, a DNA-binding subunit of ncPRC1.6 (data from GSE48175, ref. 67). **e**, Co-immunoprecipitation of KDM5C and PCGF6, a component of the ncPRC1.6 complex. DASH cells were transfected with the indicated plasmids, then immunoprecipitation and western blotting were performed as indicated. **f**, 38% (234/609) of genes upregulated in *Kdm5c* KO are also upregulated in *Pcgf6* KO, including 2CLC markers (data from GSE122715, ref. 68). **g**, Among the 609 genes repressed by KDM5C, some are directly bound by KDM5C and they are frequently cobound by ncPRC1.6 (top part). Some other genes depend on KDM5C for

repression, but are not directly bound by KDM5C (bottom part). Many of these genes also depend on PCGF6 for repression, and we speculate that KDM5C may be recruited by ncPRC1.6. **h**, The WT version of KDM5C efficiently represses mScarlet expression in the *Kdm5c* KO, and so does the catalytically inactive (H514A) mutant (50,000 cells per condition). NS, *P* > 0.05, ***P* < 0.01 (two-sided Holm-Sidak post hoc test following ANOVA). **i**, RT-qPCR analysis: the WT version of KDM5C efficiently represses 2CLC gene expression in the *Kdm5c* KO, and so does the catalytically inactive (Cat. inact.) mutant. Empty vect., empty vector. NS, *P* > 0.05; **P* < 0.05 (two-sided Holm-Sidak corrected two-tailed *t*-tests). **j**, Western blot analysis: the WT version of KDM5C efficiently represses 2CLC protein expression in the *Kdm5c* KO, and so does the catalytically inactive mutant. Data from *n* = 3 independent KO clones are presented as mean values ± s.d. for **h**–**i**. Significance: hypergeometric test for **b**, **d** and **f**.



DNMT3B (ref. 60), as well as TRIM28 and CAF-1 (ref. 61). Our results indicate that mutations altering the function of SPOP, KDM5C, MCM3AP or ZBTB14 could potentially contribute to facioscapulohumeral dystrophy. Conversely, mutations of KDM5C cause mental retardation⁶² and mutations of MCM3AP cause Charcot–Marie–Tooth disease⁶³; it is possible that misexpression of 2CLC genes contributes to these pathologies.

Finally, in humans, the 8-cell-embryo is similar to the mouse 2-cell embryo in that it is the site of zygotic gene activation, and its cells are totipotent. Very recent results have shown that human ‘8-cell-like’ cells can be obtained from human ESCs in vitro⁶⁴, paving the way for further research into how these cells emerge. It will be of interest to determine whether the regulators described in our work play a conserved role in humans. This effort and those of others will help harness the potential of totipotent cells for basic research and for medicine.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-023-01038-z>.

References

- Ladstätter, S. & Tachibana, K. Genomic insights into chromatin reprogramming to totipotency in embryos. *J. Cell Biol.* **218**, 70–82 (2019).
- Riveiro, A. R. & Brickman, J. M. From pluripotency to totipotency: an experimentalist’s guide to cellular potency. *Development* **147**, dev189845 (2020).
- De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
- Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
- Whiddon, J. L., Langford, A. T., Wong, C.-J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
- Falco, G. et al. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* **307**, 539–550 (2007).
- Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Genet, M. & Torres-Padilla, M.-E. The molecular and cellular features of 2-cell-like cells: a reference guide. *Development* **147**, dev189688 (2020).
- Iturbide, A. & Torres-Padilla, M.-E. A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. *Curr. Opin. Genet. Dev.* **64**, 26–30 (2020).
- Eckersley-Maslin, M. et al. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).
- De Iaco, A., Coudray, A., Duc, J. & Trono, D. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep.* **20**, e47382 (2019).
- Grow, E. J. et al. p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat. Genet.* **53**, 1207–1220 (2021).
- Yang, M. et al. Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* <https://doi.org/10.1016/j.stem.2022.01.010> (2022).
- Morgan, M., Kumar, L., Li, Y. & Baptissart, M. Post-transcriptional regulation in spermatogenesis: all RNA pathways lead to healthy sperm. *Cell. Mol. Life Sci.* **78**, 8049–8071 (2021).
- Rodriguez-Terrones, D. et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).
- Endoh, M. et al. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. *eLife* **6**, e21064 (2017).
- Dahlet, T. et al. E2F6 initiates stable epigenetic silencing of germline genes during embryonic development. *Nat. Commun.* **12**, 3582 (2021).
- Mochizuki, K. et al. Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat. Commun.* **12**, 7020 (2021).
- Ishiyachi, T. et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
- Fu, X., Wu, X., Djekidel, M. N. & Zhang, Y. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.* **21**, 835–844 (2019).
- Alda-Catalinas, C. et al. A single-cell transcriptomics CRISPR-activation screen identifies epigenetic regulators of the zygotic genome activation program. *Cell Syst.* **11**, 25–41.e9 (2020).
- Gretarsson, K. H. & Hackett, J. A. Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-020-0445-1> (2020).
- Eckersley-Maslin, M. A. et al. MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
- Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Leitch, H. G. et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
- Velasco, G. et al. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl Acad. Sci. USA* **107**, 9281–9286 (2010).
- Laisné, M., Gupta, N., Kirsh, O., Pradhan, S. & Defossez, P.-A. Mechanisms of DNA methyltransferase recruitment in mammals. *Genes* **9**, 617 (2018).
- Rodriguez-Terrones, D. et al. A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep.* **21**, e48354 (2020).
- Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
- Li, W. et al. MAGECK enables robust identification of essential genes from genome-scale CRISPR-Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
- Petryk, N., Bultmann, S., Bartke, T. & Defossez, P.-A. Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkaa1154> (2020).
- Clark, A. & Bursleson, M. SPOP and cancer: a systematic review. *Am J Cancer Res* **10**, 704–726 (2020).
- Li, M. et al. Genome-wide CRISPR-KO screen uncovers mTORC1-Mediated Gsk3 regulation in naive pluripotency maintenance and dissolution. *Cell Rep.* **24**, 489–502 (2018).
- Singh, S. K. et al. GANP regulates recruitment of AID to immunoglobulin variable regions by modulating transcription and nucleosome occupancy. *Nat. Commun.* **4**, 1830 (2013).
- Aksenova, V. et al. Nucleoporin TPR is an integral component of the TREX-2 mRNA export pathway. *Nat. Commun.* **11**, 4577 (2020).
- Umlauf, D. et al. The human TREX-2 complex is stably associated with the nuclear pore basket. *J. Cell Sci.* **126**, 2656–2667 (2013).

37. Schneider, M. et al. The nuclear pore-associated TREX-2 complex employs mediator to regulate gene expression. *Cell* **162**, 1016–1028 (2015).
38. Fukuda, K., Okuda, A., Yusa, K. & Shinkai, Y. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.* **28**, 846–858 (2018).
39. Zhang, W. et al. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz594> (2019).
40. Rual, J.-F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
41. Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
42. Thomson, J. P. et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
43. Grand, R. S. et al. BANP opens chromatin and activates CpG-island-regulated genes. *Nature* **596**, 133–137 (2021).
44. Tsaban, T. et al. CladeOScope: functional interactions through the prism of clade-wise co-evolution. *NAR Genom. Bioinform.* **3**, lqab024 (2021).
45. Outchkourov, N. S. et al. Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. *Cell Reports* **3**, 1071–1079 (2013).
46. Iwase, S. et al. The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**, 1077–1088 (2007).
47. Yan, Y.-L. et al. DPPA2/4 and SUMO E3 ligase PIAS4 oppositely regulate zygotic transcriptional program. *PLoS Biol.* **17**, e3000324 (2019).
48. Nakatake, Y. et al. Kinetics of drug selection systems in mouse embryonic stem cells. *BMC Biotechnol.* **13**, 64 (2013).
49. Percharde, M. et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405.e19 (2018).
50. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
51. Tahiliani, M. et al. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447**, 601–605 (2007).
52. Christensen, J. et al. RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128**, 1063–1076 (2007).
53. Scandaglia, M. et al. Loss of Kdm5c causes spurious transcription and prevents the fine-tuning of activity-regulated enhancers in neurons. *Cell Rep.* **21**, 47–59 (2017).
54. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* **8**, e1002964 (2012).
55. Boukhald, G. M. et al. The transcriptional repressor polycomb group factor 6, PCGF6, negatively regulates dendritic cell activation and promotes quiescence. *Cell Rep.* **16**, 1829–1837 (2016).
56. Evangelista, F. M. et al. Transcription and mRNA export machineries SAGA and TREX-2 maintain monoubiquitinated H2B balance required for DNA repair. *J. Cell Biol.* **217**, 3382–3397 (2018).
57. Atashpaz, S. et al. ATR expands embryonic stem cell fate potential in response to replication stress. *eLife* **9**, e54756 (2020).
58. Kim, W. et al. ZFP161 regulates replication fork stability and maintenance of genomic stability by recruiting the ATR/ATRIP complex. *Nat. Commun.* **10**, 5304 (2019).
59. Nakatani, T. et al. DNA replication fork speed underlies cell fate changes and promotes reprogramming. *Nat. Genet.* **54**, 318–327 (2022).
60. van den Boogaard, M. L. et al. Mutations in DNMT3B modify epigenetic repression of the D4Z4 repeat and the penetrance of facioscapulohumeral dystrophy. *A. J. Hum. Genet.* **98**, 1020–1029 (2016).
61. Campbell, A. E. et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *eLife* **7**, e31023 (2018).
62. Bonefas, K. M. & Iwase, S. Soma-to-germline transformation in chromatin-linked neurodevelopmental disorders? *FEBS J.* <https://doi.org/10.1111/febs.16196> (2021).
63. Ylikallio, E. et al. MCM3AP in recessive Charcot-Marie-Tooth neuropathy and mild intellectual disability. *Brain* **140**, 2093–2103 (2017).
64. Taubenschmid-Stowers, J. et al. 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell* **29**, 449–459.e6 (2022).
65. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
66. Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
67. Krepelova, A., Neri, F., Maldotti, M., Rapelli, S. & Oliviero, S. Myc and max genome-wide binding sites analysis links the Myc regulatory network with the polycomb and the core pluripotency networks in mouse embryonic stem cells. *PLoS ONE* **9**, e88933 (2014).
68. Scelfo, A. et al. Functional landscape of PCGF proteins reveals both RING1A/B-dependent and RING1A/B-independent specific activities. *Mol. Cell* **74**, 1037–1052.e7 (2019).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Cell culture

J1 mouse ESCs (129S4/SvJae, XY) were cultured at 37 °C under 5% CO₂ on gelatin-coated dishes in serum/LIF medium containing DMEM/GlutaMAX supplemented with 15% FBS, nonessential amino acids, penicillin–streptomycin and 1,000 U ml⁻¹ LIF. When necessary, ESCs were adapted to 2i, vitamin C and LIF medium containing serum-free DMEM-F12 and neurobasal media supplemented with 1% N2, 2% B27, 100 µg ml⁻¹ ascorbic acid, 1 µM PD0325901 and 3 µM CHIR99021. For spontaneous differentiation, cells were seeded at clonal density in a serum medium without LIF.

Cloning of sgRNA, transfection and transduction in ESCs

The single-guide RNAs (sgRNAs) were designed using Benchling/CRISPOR or were obtained from the Brie library. The sgRNAs were cloned in the PX459 vector (Addgene no. 62988) or in lentiCRISPRv2 (Addgene no. 52961). For ESC transfection, we used an Amaxa 4D-Nucleofector (Lonza). Production of lentiviral particles was performed by calcium-phosphate transfection of HEK293T with psPAX2 and pMD2.G plasmids.

Generation of the DASH reporter cell line

The reporter cassette P2A-mScarlet-T2A-Hygro^R (GenScript) was inserted within exon 6 of the mouse *Dazl* gene. The cassette was flanked by *Dazl* homology arms corresponding to endogenous intron 5-exon 6 and intron 6 sequences, respectively, and cloned into pUC57. The two sgRNAs targeting *Dazl* exon 6 were cloned into the pSpCas9(BB)-2A-GFP backbone (Addgene no. 48138). The homologous integration of the reporter cassette in one of the alleles of *Dazl* was confirmed by PCR and sequencing.

CRISPR KO screen: amplification of sgRNA library, lentiviral transduction and sample collection

We performed the screen in two biological replicates, with a multiplicity of infection (MOI) of roughly 0.1, using the lentiviral Brie sgRNA library (Addgene no. 73632)²⁹. After 48 h, transduced cells were selected with 2 µg ml⁻¹ puromycin for 5 days. Coverage was 150× (150 transduced cells per sgRNA) for each biological replicate. Following this, cells were initially selected with 50 µg ml⁻¹ Hygromycin for 3 days followed by additional selection for 11 days at 125 µg ml⁻¹ Hygromycin. Three weeks postinfection, Hygromycin-resistant cells were sorted by FACS for mScarlet expression.

CRISPR KO screen: sequencing and analysis

PCR was performed with Platinum Taq polymerase (Thermo), using a pool of P5 primers and a unique P7 barcode primer. The PCR conditions were: 94 °C for 4 min; then 28 cycles of 94 °C for 30 s, 53 °C for 30 s and 72 °C for 30 s per kb; final extension at 72 °C for 10 min. The PCR products were retrieved using QIAquick PCR Purification Kit (Qiagen). The DNA was further purified using AMPure XP (Beckman Coulter). Libraries were sequenced on an Illumina HiSeq 1500 in single-end 100 basepair (bp) output mode. The sgRNA distribution and enrichment at different time points were analyzed with the MAGeCK workflow³⁰ (Supplementary Table 1). A statistical threshold of $P < 0.0005$ resulted in a list of 40 candidates whose KO led to the expression of mScarlet and Hygro^R in DASH ESCs.

Generation of individual gene KO and clonal cell lines

Individual KOs in DASH ESCs were generated using the top two most efficient sgRNAs (as determined by MAGeCK analysis). During lentiviral production, both sgRNA plasmids targeting the same gene were mixed to increase KO efficiency. Transduced cells were selected with 2 µg ml⁻¹ puromycin for 3 days; followed by Hygromycin selection (50 µg ml⁻¹ for 3 days, and 125 µg ml⁻¹ for the next 7 days). For *Kdm5c*, *Mcm3ap*, *Spop* and *Zbtb14*, three independent clonal KO lines were established.

Rescue experiments, piggyBac system

For rescue experiments, the coding sequence of candidates was synthesized (*Spop* and *Zbtb14*) (GenScript), amplified from complementary DNA (*Kdm5c*), or obtained from colleagues (*Mcm3ap*, from N. Saka-guchi³⁴). In all cases, silent mutations were incorporated within the PAM and/or sgRNA sequence. These coding sequences were cloned into a piggyBac vector and cotransfected with phosphate buffered transposase for stable insertion³⁸. Transfected cells were selected with 5 µg ml⁻¹ blasticidin for 5 days.

Knockdown experiments, siRNA/shRNA transfection

The pLKO.1-blasticidin short hairpin RNA vector for *Dux* was kindly shared by D. Trono³, and the pLKO.1-neomycin shRNA vector for *Dppa2* was ordered (Sigma, TRCN0000174599). Transduced cells were selected with 5 µg ml⁻¹ blasticidin for 5 days or 400 µg ml⁻¹ geneticin for 7 days. The siRNA for *Tp53* was kindly shared by G. Velasco²⁶, and other siRNA pools were ordered (see Supplementary Table 2 for references). Cells were reverse transfected with 50 nM siRNA and 3 µl ml⁻¹ Lipofectamine RNAiMAX (Thermo Fisher Scientific) diluted in Opti-MEM. Total RNA was extracted 2 days after transfection for RT–qPCR.

Flow cytometry

mScarlet expression was determined by flow cytometry using a yellow laser (561 nm) at the ImagoSeine core facility (Institut Jacques Monod). Data were analyzed with FlowJo software.

Crystal violet staining

Cells were seeded at the same density in all wells and grown with or without Hygromycin for 7 days. Surviving cells were fixed with absolute ethanol for 15 min, stained with 1% crystal violet dye (Sigma) for 30 min and washed extensively with water to remove the unbound stain.

Immunofluorescence

Cells grown on sterile round glass coverslips were fixed directly with 4% paraformaldehyde for 10 min at room temperature, and permeabilized with 0.5% Triton X-100 in PBS for 5 min at room temperature. Samples were blocked in 2% BSA in PBS supplemented with 0.1% Tween 20 for 1 h at room temperature before incubation with primary antibodies for 1 h and secondary antibodies conjugated to Alexa Fluor 488 or 594 (Invitrogen) for 30 min at room temperature. Coverslips were mounted in Vectashield medium with 4,6-diamidino-2-phenylindole (Vector Laboratories), and images were acquired with a Leica DMI6000 epifluorescence microscope. The intensity of Zscan4-positive cells was obtained by scoring at least 100 cells in each experiment, applying the same treatment of fluorescence levels to all images with the ImageJ software (v.1.53).

Digital droplet PCR

The PCR reaction mixture composed of 2× EvaGreen ddPCR Supermix (Bio-Rad), primers at a final concentration of 100 nM and 10 ng of template DNA were partitioned into up to 20,000 droplets by water–oil emulsion. After droplet generation, a regular PCR was performed with the following conditions: 95 °C for 5 min; 95 °C for 30 s and 60 °C for 1 min (40 cycles); 4 °C for 5 min, 90 °C for 5 min and 4 °C hold. For all steps, a ramp rate of 2 °C s⁻¹ was used. Cycled droplets were read individually (Bio-Rad QX-200 droplet reader). Each run included technical duplicates and no-template controls.

RT–qPCR

Total RNA was extracted from cells with RNeasy Plus Mini kit (Qiagen) according to the manufacturer's instructions and quantified using Qubit RNA BR Assay kit on Qubit v.2.0 Fluorometer (Thermo Fisher Scientific). One microgram of total RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) and Oligo dT primers (Promega). RT–qPCR was performed using Power

SYBR Green (Applied Biosystems) on a ViiA 7 Real-Time PCR System (Thermo Fisher Scientific). *Actinb*, *Ppia* and *RplpO* were used for normalization. Primer sequences are available in Supplementary Table 2.

Western blotting

Cells were gathered and lysed in RIPA buffer (Sigma) with protease inhibitor cocktail (Thermo Fisher Scientific), sonicated with a series of 30 s ON and 30 s OFF for 5 min on a Bioruptor (Diagenode), and centrifuged at 16,000g for 5 min at 4 °C. The supernatant was collected and quantified by BCA assay (Thermo Fisher Scientific). Thirty micrograms of protein extract per sample was mixed with NuPage 4× LDS Sample Buffer and 10× Sample Reducing Agent (Thermo Fisher Scientific) and denatured at 95 °C for 5 min. Samples were resolved on a precast SDS-PAGE 4–12% gradient gel (Thermo Fisher Scientific) with 120 V electrophoresis for 90 min and blotted onto a nitrocellulose membrane (Millipore). The membrane was blocked with 5% fat-free milk and PBS at room temperature for 1 h, then incubated overnight at 4 °C with appropriate primary antibodies. After three washes with PBS and 0.1% Tween20, the membranes were incubated with the cognate fluorescent secondary antibodies and revealed in the LI-COR Odyssey-Fc imaging system. The list of antibodies used in this study is available in the Nature Research Reporting Summary.

ChIP

Here, 1×10^7 cells were cross-linked with 1% formaldehyde for 5 min at room temperature. Cross-linking was stopped by adding Glycine (125 mM final), and the cells were washed with PBS. Cells were resuspended in Swelling Buffer (0.5% NP-40, 0.85 mM KCl, 1 mM PMSF, 5 mM PIPES pH 8.0) supplemented with protease inhibitor cocktail (Roche) and incubated for 20 min on ice. After centrifugation, cell nuclei were resuspended in immunoprecipitation buffer (0.1% SDS, 1% Triton X-100, 3 mM EDTA, 1 mM PMSF, 150 mM NaCl, 25 mM Tris-HCl pH 8.0, 1× protease inhibitor cocktail) and sonicated for 5 min (series of 30 s ON/30 s OFF) on a Bioruptor Pico (Diagenode) to generate 200 to 500 bp fragments. Fragmented chromatin (50 µg) was immunoprecipitated in immunoprecipitation buffer with 1 µg of antibody (α -KDM5C, Bethyl Laboratories no. A301-034A) overnight at 4 °C. Subsequent steps (including incubation with magnetic beads, multiple washes, elution) were performed with the Pierce Magnetic ChIP Kit (Thermo Fisher Scientific), following the manufacturer's instructions. Reverse cross-linked DNA was purified by ChIP DNA Clean&Concentrator kit (Zymo Research). Libraries were prepared with KAPA HyperPrep kit (Roche), following the manufacturer's instructions.

ChIP-seq analysis

FASTQ reads were trimmed using Trimmomatic (v.0.39) and parameters: ILLUMINACLIP:illumina_adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:36. Trimmed reads were aligned using Bowtie2 (v.2.4.1) in local mode. Following alignment, Picard (v.2.23.4) CleanSam, SamFormatConverter, SortSam and MarkDuplicates were used to generate a duplicate-marked bam file. The resulting bam files were converted to bigwig using Deeptools (v.3.3.0) Bamcoverage and options `-ignoreDuplicates -normalizeUsing CPM -minMappingQuality 10 -ignoreForNormalization chrX chrY chrM`. Peaks were called using MACS2 with default parameters, and motif enrichment analysis was performed with homer2 findMotifs (v.4.11). For datasets already published, SRA files were downloaded from National Center for Biotechnology Information (NCBI) GEO using E-utilities `esearch` and `efetch` (v.15.9) and converted to FASTQ using SRAToolkit (v.2.8.0).

Isolation of genomic DNA

Genomic DNA was isolated from cells using overnight 200 µg ml⁻¹ proteinase K treatment at 55 °C followed by 20 µg ml⁻¹ RNase A treatment at 37 °C for 1 h and extracted by a standard phenol, chloroform and alcohol method. Alternatively, genomic DNA was isolated from cells

using QIAamp DNA Mini kit (Qiagen), following the manufacturer's instructions. Genomic DNA was resuspended in water and quantified with Qubit dsDNA BR Assay kit on Qubit v.2.0 Fluorometer (Thermo Fisher Scientific). DNA integrity was assessed with Genomic DNA ScreenTape on TapeStation system (Agilent) and samples with a DNA integrity number greater than nine were used for subsequent analysis.

DNA methylation analysis: MeDIP

MeDIP was performed using the Auto MeDIP Kit on an automated platform SX-8G IP-Star Compact (Diagenode). Briefly, 2.5 µg of DNA was sheared using a Bioruptor Pico to approximately 500 bp fragments, as assessed with D5000 ScreenTape (Agilent). Cycle conditions were as follows: 15 s ON and 90 s OFF, repeated six times. A portion of sheared DNA (10%) was kept as input and the rest of the sheared DNA was immunoprecipitated with α -5-methylcytosine antibody (Diagenode no. C15200081, clone 33D3, 1 µg per immunoprecipitation), bound to magnetic beads and was isolated. qPCR for selected genomic loci was performed and efficiency was calculated as a percentage (me-DNA-IP/total input). Primer sequences are listed in Supplementary Table 2.

DNA methylation analysis: luminometric methylation assay

To assess global CpG methylation, 500 ng of genomic DNA was digested with MspI+EcoRI and HpaII+EcoRI (NEB) in parallel reactions, EcoRI was included as an internal reference. CpG methylation percentage is defined as the HpaII/MspI ratio. Samples were analyzed using a PyroMark Q24 Advanced pyrosequencer.

DNA methylation analysis: LC-MS/MS

The genomic DNA was extracted as described above with an additional step of digestion with RNase A. One microgram of DNA was treated with 10 U of DNA Degradase Plus (ZymoResearch) at 37 °C for 4 h. After enzyme inactivation at 70 °C for 20 min, the solution was filtered with Amicon Ultra-0.5 ml 10 K centrifugal filters (Merck Millipore). The reaction mix retained in the centrifugal filter was processed for liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis. Analysis of global levels of 5-mdC were performed on a Q exactive mass spectrometer (Thermo Fisher Scientific). It was equipped with an electrospray ionization source (H-ESI II Probe) coupled with an Ultimate 3000 RS HPLC (Thermo Fisher Scientific). Digested DNA was injected onto a ThermoFisher Hypersil Gold aQ chromatography column (100 × 2.1 mm, 1.9 µm particle size) heated at 30 °C. The flow rate was set at 0.3 ml min⁻¹ and run with an isocratic eluent of 1% acetonitrile in water with 0.1% formic acid for 10 min. Parent ions were fragmented in positive ion mode with 10% normalized collision energy in parallel-reaction monitoring mode. MS2 resolution was 17,500 with an AGC target of 2×10^5 , a maximum injection time of 50 ms and an isolation window of 1.0 *m/z*. The inclusion list contained the following masses: dC (228.1) and 5-mdC (242.1). Extracted ion chromatograms of base fragments (± 5 ppm) were used for detection and quantification (112.0506 Da for dC; 126.0662 Da for 5-mdC). Calibration curves were previously generated using synthetic standards in the ranges of 0.2 to 10 pmol injected for dC and 0.02 to 10 pmol for 5 mdC. Results are expressed as a percentage of total dC.

Whole-genome bisulfite sequencing

Genomic DNA was extracted as described for LC-MS/MS. The library preparation for WGBS was performed with the tPBAT protocol described previously^{69,70}. One hundred nanograms of genomic DNA spiked with 1% (w/w) of unmethylated lambda DNA (Promega) was used for the library preparation. The sequencing was performed by MacroGen Japan Inc. using the HiSeq X Ten system. We assigned eight lanes for the analysis of 20 samples. The sequenced reads were mapped with BMap and summarized with an in-house pipeline as described previously⁷⁰, with custom scripts archived using GitHub (Code availability statement). The basic metrics of the methylome data are provided

in Supplementary Table 3. DNA methylation levels over CpGs covered by at least five sequencing reads were averaged over the following regions of interest: genome-wide 2 kb bins, enhancer elements (Ensembl Regulatory features release 81, $n = 73,796$), promoters (NCBI RefSeq, TSS ± 1 kb, $n = 24,371$), gene bodies and transposable elements (RepeatMaster, $n = 5,147,736$). PCA plot was generated using DeepTools multiBamSummary with default parameters.

RNA-seq: library preparation for transcriptome sequencing

One microgram total RNA per sample was used as input and RNA samples were spiked with ERCC RNA Spike-In Mix (Thermo). Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB). Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under an elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5 \times). First-strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were replaced by dUTP. The remaining overhangs were converted into blunt ends via exonuclease and/or polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially 250–300 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter). Then 3 μ l USER Enzyme (NEB) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. Finally, products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

RNA-seq: read alignment

FASTQ reads were trimmed using Trimmomatic (v.0.39) and parameters: ILLUMINACLIP:adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:36. Read pairs that survived trimming were aligned to the mouse reference genome (build mm10) using STAR (v.2.7.5c) and default single-pass parameters. PCR duplicate read alignments were flagged using Picard-tools (2019) MarkDuplicates (v.2.23.4). Uniquely aligned, non-PCR-duplicate reads were kept for downstream analysis using Samtools view (v.1.10) and parameters: -q 255 -F 1540. Gene expression values were calculated over the mm10 NCBI RefSeq Genes annotation using VisRseq (v.0.9.12) and normalized per million aligned reads per transcript length in kilobases (RPKM). Bigwig files were generated using DeepTools bamCoverage (v.3.3.0) using counts per million normalization and visualized in IGV (v.2.11).

RNA-seq: differential expression and gene set enrichment analysis

DESeq2 (v.1.30.0) was used using apeglm LFC shrinkage to calculate differential expression. Genes or transposable elements were categorized as significantly differentially expressed if they showed an absolute expression fold-change of two or more, and associated adjusted $P < 0.01$. Differentially expressed genes are listed in Supplementary Table 4. Gene set enrichment analysis was performed using GSEA (v.4.1.0) and default parameters (1,000 permutations, permutation type = gene_set. Selected significant terms from Hallmark gene sets ($n = 50$ 'h.all.v7.4.symbols.gmt') were displayed.

RNA-seq: transposable element quantification

RepeatMasker (last updated 6 February 2012) was downloaded from the UCSC Table Browser. To measure the expression of transposable element families, PCR duplicates were removed and all reads, including uniquely mapped and multi-mapped reads, were enumerated using VisRseq. Multi-mapped reads were counted once, and all individual

elements were aggregated to calculate family-wide expression in read count for differential expression analysis. Heatmaps were generated using Morpheus (<https://software.broadinstitute.org/morpheus>).

RNA-seq analysis: MERVL/MT2_Mm analysis

MERVL internal sequences and their MT2_Mm LTR promoters were extracted from the RepeatMasker annotation (last updated 6 February 2012). Internal sequences and their LTRs located within 88 bp were merged into a single element using bedtools merge (v.2.27.0) to account for an 87 bp insertion of a related ORR1A3 element. Elements were categorized as full-length MERVL elements if they contained both LTR elements and internal sequences and spanned more than 6,000 bp. MT2_Mm elements under 500 bp in length were defined as 'Solo-LTRs'. All other elements, such as those composed of MERVL internal sequences and only one LTR, were categorized as 'other'. Genome-wide mappability scores were calculated using iGEM (v.1.315) and parameters: K_MER_SIZE = 300 MAX_MISMATCHES = 0.04 and the mappability of each MERVL element was calculated using VisRseq. A list of MERVL elements that generate chimeric transcripts was downloaded⁷ and mapped onto the mm10 genome using UCSC LiftOver. To measure individual transposable element expression, only uniquely aligned, non-PCR duplicate reads were counted. Elements were grouped and sorted by K-medoid clustering on log₁₀-transformed RPKM values using the R package 'cluster' and VisRseq.

Statistics and reproducibility

Details on samples sizes and statistical tests performed are presented in the corresponding figure legends. When representative results are shown, each experiment was independently repeated 2–4 times with similar results. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

Data availability

All sequencing data generated in this study are available in the Gene Expression Omnibus under accession number [GSE221710](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221710). For Figs. 1 and 4 and Extended Data Fig. 5, published RNA-seq were reanalyzed from [GSE33923](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33923) (Macfarlan)⁷; [E-MTAB-2684](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=E-MTAB-2684) (Ishiuuchi)¹⁹; [GSE75751](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75751) (Eckersley-Maslin)²³; [GSE71434](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71434) (Zhang)⁶⁶ and [GSE66390](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66390) (Wu)⁶⁵. Other online databases used in the study were STRING (<https://string-db.org/>) in Fig. 2 and GSEA (<https://www.gsea-msigdb.org/gsea/index.jsp>) in Fig. 4 and Extended Data Fig. 4. Source data are provided with this paper.

Code availability

Custom scripts used for WGBS mapping are available at: <https://github.com/FumihitoMiura/Project-2>.

References

- Miura, F., Enomoto, Y., Dairiki, R. & Ito, T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **40**, e136 (2012).
- Miura, F. et al. Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **47**, e85 (2019).

Acknowledgements

We are grateful to the following colleagues for useful advice:

A. Bardin, B. Rodgers, C. Francastel, C. Rougeulle, S. Polo, P. Navarro, R. Margueron, G. Velasco, G. Fillion, M. Casanova, S. Donakonda,

M. Weber, L. Tora, Y. Shinkai, S. Khochbin and T. Bartke. We thank the following colleagues for useful reagents: D. Bourc'his (Institut Curie, Paris) for J1 mESCs, M.-E. Torres Padilla (Helmholtz Zentrum Munich) for tbg4 reporter mESCs, M. Timmers (DKFZ Heidelberg) for KDM5C tagged lines, N. Sakaguchi (Kumamoto University) for a mouse GANP cDNA, H. Niwa and Y. Shinkai (RIKEN Saitama) for piggyBac constructs, J. Sharif and H. Koseki (RIKEN Yokohama) for a FLAG-PCGF6 expression plasmid, J. Hackett (EMBL Rome) for a FLAG-DPPA2 expression plasmid. We thank the Vectorology platform, Epigenetics platform, Microscopy platform and Bioinformatics/Biostatistics Core Facility at the CNRS Epigenetics and Cell Fate Unit (Université Paris Cité), for providing access and technical advice. We thank E. Jeannot at Institut Curie for help with ddPCR. We thank S. Bultmann (LMU, Munich) for help with sgRNA sequencing and MAGECK analysis. We acknowledge the ImagoSeine core facility of the Institut Jacques Monod, member of the France Biolmaging (grant no. ANR-10-INBS-04) and the support of La Ligue contre le Cancer (grant no. R03/75-79). Microfluidic RT-qPCR (Fluidigm) analysis was carried out on the qPCR-HD-Genomic Paris Centre Core Facility and was supported by grants from Région Ile-de-France, grant no. DIMBIO-RVT-INSERM-ADR-P11 21016711. P.-A.D. is supported by Agence Nationale de la Recherche (PRCI INTEGER grant no. ANR-19-CE12-0030-01), LabEx 'Who Am I?' (grant no. ANR-11-LABX-0071), Université de Paris IdEx (grant no. ANR-18-IDEX-0001) funded by the French Government through its 'Investments for the Future' program, Fondation pour la Recherche Médicale, Fondation ARC (Programme Labellisé grant no. PGA1/RF20180206807). P.-A.D. and M.C.V.G. are supported by Agence Nationale de la Recherche (grant no. PRC REMEDY ANR-21-CE12-0015-03). P.-A.D., A.S. and G.C. were supported by grant RETROMET, no. ANR-16-CE12-0020, from Agence Nationale de la Recherche. J.R.A. and M.V.C.G. were supported by Laboratoire d'excellence Who Am I? (grant no. Labex 11-LABX-0071) Emerging Teams Grant and by the European Research Council (grant no. ERC-StG-2019 DyNAMEcs). This research was supported by Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under grant no. JP20am0101103 (support no. 2652). K.Y. was the recipient of a postdoctoral fellowship from Fondation Association pour la

Recherche sur le Cancer, and of a subsequent postdoc fellowship from Labex WhoAmI. M.L. thanks the Ligue contre le Cancer for a fourth year PhD fellowship.

Author contributions

N.G. and P.-A.D. conceived the project. N.G., L.Y. and P.-A.D. planned the experiments. N.G., L.Y., L.F., F.B., S.B., K.Y. and A.A. performed experiments and analyzed the data. F.M. performed WGBS. C.D. performed MeDIP. F.B. performed mass spectrometry. M.D. and B.D. performed Fluidigm experiments. O.K. performed WGBS analysis. J.R.A., O.K., L.Y., M.L., A.S., K.Y. and G.C. performed other bioinformatic analyses. N.G., L.Y. and P.-A.D. wrote the manuscript. P.-A.D., T.I. and N.G. supervised the project. M.V.C.G., G.C., T.I. and P.-A.D. acquired funding. All authors reviewed the manuscript. J.R.A., A.A., L.F. and O.K. contributed equally.

Competing interests

The authors declare no competing interests.

Additional information

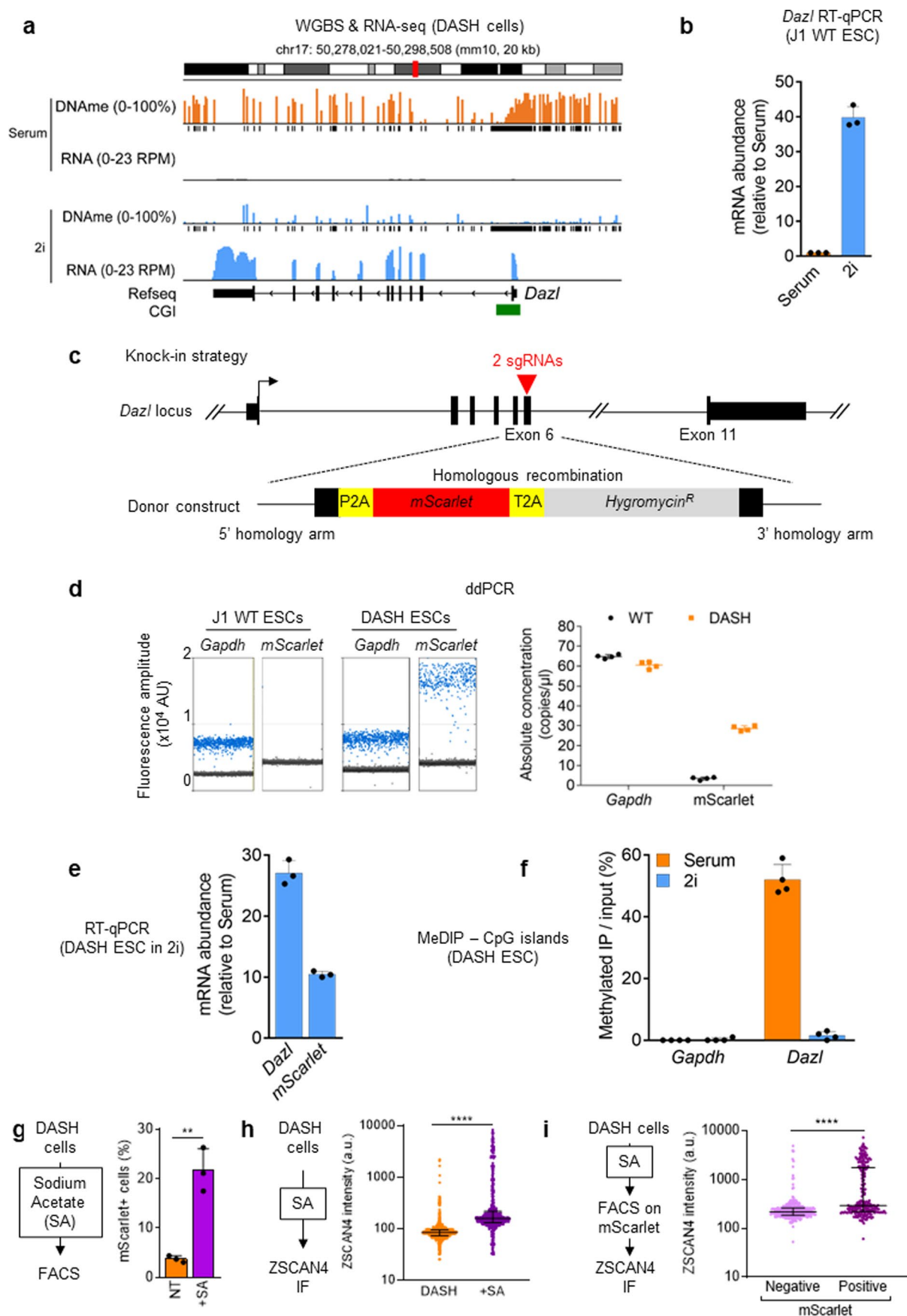
Extended data is available for this paper at <https://doi.org/10.1038/s41594-023-01038-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-023-01038-z>.

Correspondence and requests for materials should be addressed to Nikhil Gupta or Pierre-Antoine Defossez.

Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Carolina Perdigoto and Dimitris Typas were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

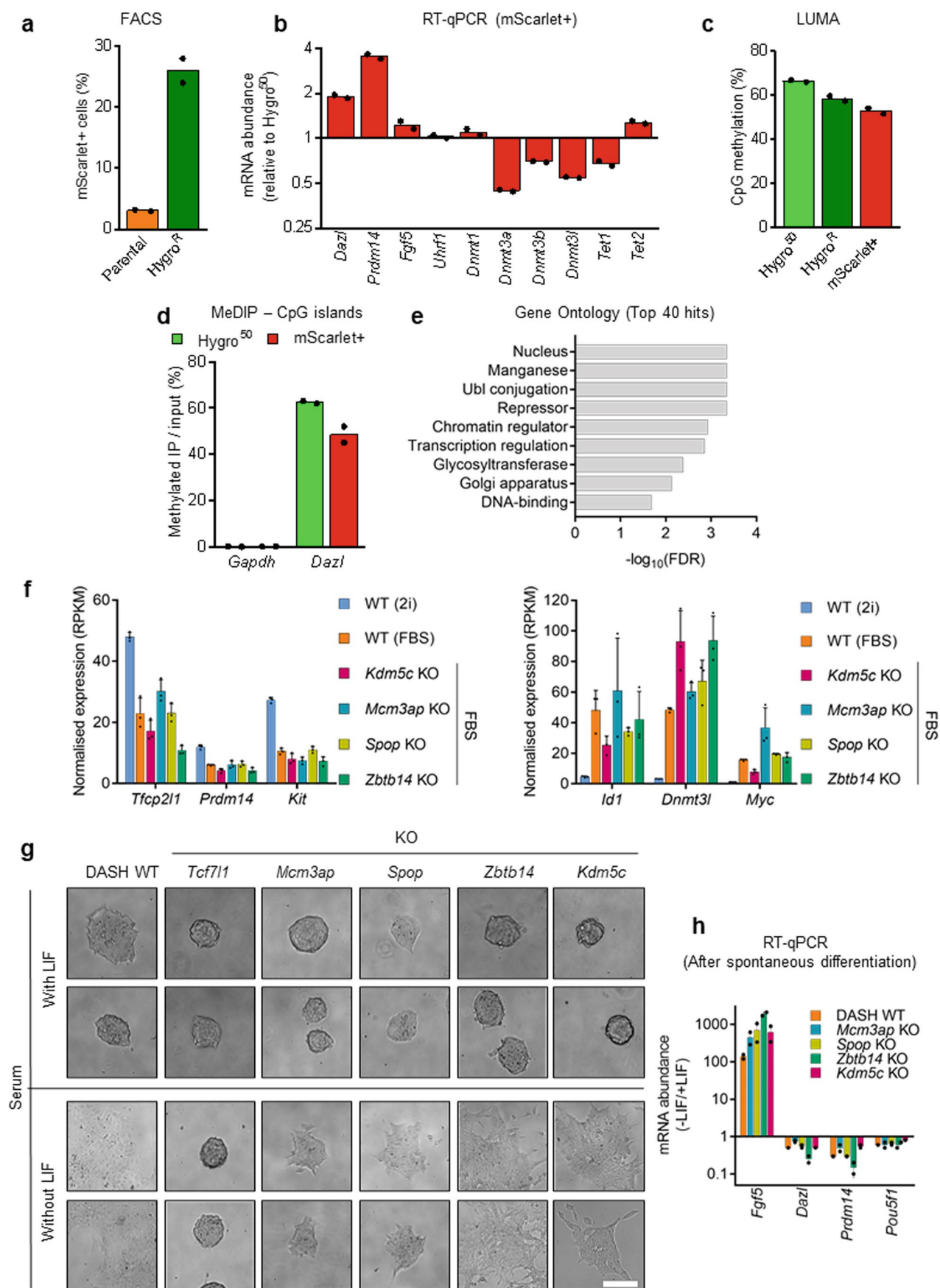
Reprints and permissions information is available at www.nature.com/reprints.



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

Extended Data Fig. 1 | Generation and validation of the *Dazl*-mScarlet-Hygromycin (DASH) reporter cell line. (a) WGBS and RNA-seq in our cellular background (DASH cells) confirm that *Dazl* promoter is highly methylated and repressed in serum, but hypo-methylated and expressed in 2i. CGI: CpG island. (b) RT-qPCR confirms the upregulation of *Dazl* in ESCs cultured in 2i (relative to serum condition). (c) *Dazl* exon 6 was targeted by 2 independent sgRNAs (red arrowhead) to insert the reporter cassette by homologous recombination. The donor construct contains *Dazl* homology arms flanking genes for the red fluorescent protein mScarlet, and the Hygromycin resistance enzyme (Hygro^R) separated by 2A self-cleaving peptides (P2A, T2A). (d) DASH ESCs have a single insertion at one of the *Dazl* alleles, as determined by ddPCR. Left panel: blue droplets are positive for the corresponding amplification; black droplets are negative. About 18,000 droplets were analyzed for each amplification. Right panel: quantitative analysis confirming single insertion of the donor construct.

Gapdh served as a control present at 2 copies/cell. (e) RT-qPCR showing the up-regulation of *Dazl* and *mScarlet* in DASH ESCs cultured in 2i (relative to serum condition). (f) MeDIP assay showing the relative levels of 5mC at *Gapdh* and *Dazl* promoters in DASH ESCs grown in serum or 2i conditions. (g) Treatment of DASH cells with Sodium Acetate, a chemical inductor of 2CLCs, activates mScarlet expression. NT: not treated. SA: Sodium Acetate (40 mM, 48 h). ** $P < 0.01$ (two-tailed *t*-test). (h) Detection of ZSCAN4-positive cells in the SA-treated population ($n = 633$ cells) versus DASH ($n = 812$ cells). **** $P < 0.0001$ (two-tailed Mann-Whitney test). (i) Most ZSCAN4-positive cells are also mScarlet-positive. **** $P < 0.0001$ (two-tailed Mann-Whitney test). Data from $n = 3$ replicates are presented as mean values \pm SD for panels b, e and g. Data from $n = 4$ replicates are presented as mean values \pm SD for panels d and f. Data from $n = 278$ cells (Negative) and $n = 228$ cells (Positive) are presented for panel i.



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

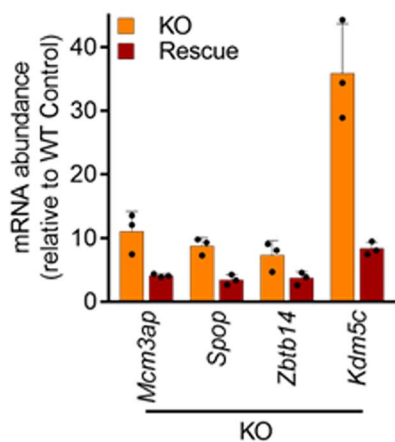
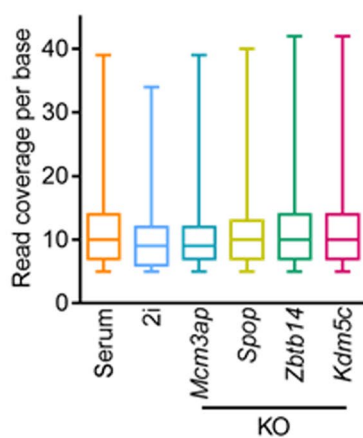
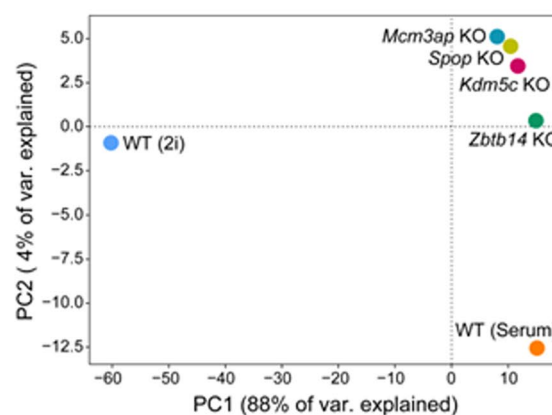
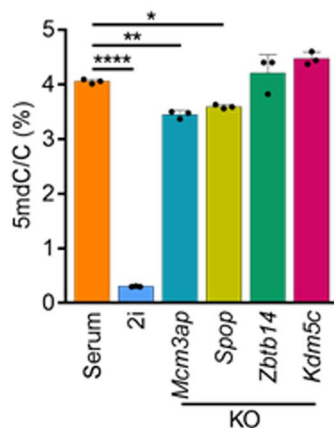
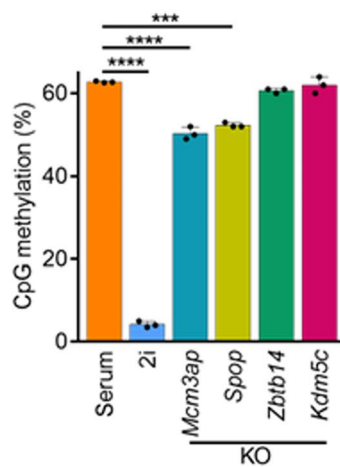
Extended Data Fig. 2 | Screen quality controls and validations of selected hits. (a) FACS analysis (50,000 cells per condition) of mScarlet expression after Hygromycin selection. $**P < 0.01$ (two-tailed *t*-test). (b) RT-qPCR: comparison of mScarlet-expressing cells to Hygro⁴⁹ cells. $*P < 0.05$, $***P < 0.005$ (two-sided Holm-Sidak post-hoc test following ANOVA). (c) Decrease of global DNA methylation in mScarlet+ cells in comparison to Hygro⁴⁹ cells from the screen, as measured by LUMA. $*P < 0.05$, $**P < 0.01$ (two-sided Holm-Sidak post-hoc test following ANOVA). (d) MeDIP assay showing the relative levels of 5mC at *Gapdh* and *Dazl* promoters in Hygro⁴⁹ and mScarlet+ screen samples. (e) Gene ontology (GO) terms (Uniprot keywords) significantly enriched among the top 40 hits. (f)

Expression of 'naïve' markers (left panel) and 'FBS' markers (right panel) in the KOs, FBS-grown and 2i-grown ESC. The KOs behave like FBS-grown ESC, not like 2i-grown cells. (g) Spontaneous differentiation induced by LIF removal is not impeded in the *Mcm3ap*, *Spop*, *Zbtb14*, and *Kdm5c* KOs. The *Tcf7l1* KO is known to be unable to differentiate and is used as a control. Scale bar: 200 μm . (h) RT-qPCR on pluripotency and differentiation markers after LIF removal in the indicated KOs. Data from $n = 2$ independent screen replicates are presented as mean values \pm SD for panels a-d. Data from $n = 3$ independent KO clones are presented as mean values \pm SD for panels f and h.

a

Identification of the mutations in the CRISPR KO clones

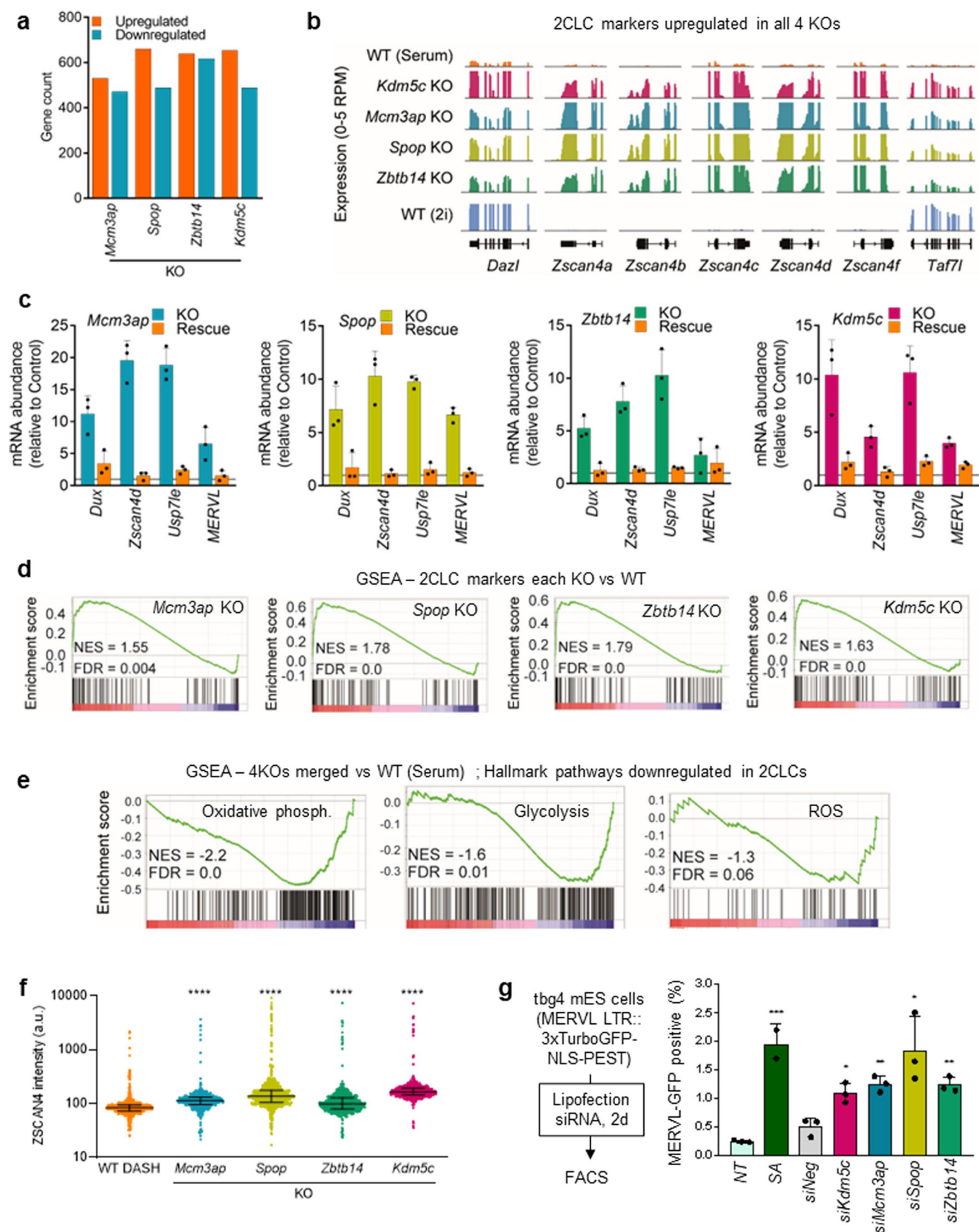
Gene	Protein ID	Clone number	Mutation	
			Allele 1	Allele 2
<i>Mcm3ap</i>	NP_062307	1	p.R297Kfs*9	p.R298Tfs*10
		2	p.R298Sfs*4	p.R298Tfs*10
		3	p.P296Efs*8	p.R298Tfs*10
<i>Spop</i>	NP_079563	1	p.C23Hfs*5	
		2	p.C23*	
		3	p.C23Sfs*3	
<i>Zbtb14</i>	NP_033573	1	p.E77Dfs*2	p.R82*
		2	p.S83_S136del	
		3	p.L81Sfs*34	p.N81Tfs*9
<i>Kdm5c</i>	NP_038696	1	p.R635_R637del / p.R744Ffs*17	
		2	p.R634Efs*24	
		3	p.S743Ifs*15	

bRT-qPCR *Dazl*
(Serum-grown cells)**c**WGBS
CpG coverage**d**WGBS – PCA
(3 replicates of each clone merged)**e**% CpG methylation
(LC-MS/MS)**f**% CpG methylation
(LUMA)

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

Extended Data Fig. 3 | Genetic analysis of the KO clones, rescue, and WGBS confirmation. (a) Identification of the mutations found in each of the KO clones. (b) RT-qPCR analysis: genetic rescue of each KO suppresses *Dazl* mRNA induction. (c) WGBS coverage statistics (n = 3 independent KO clones). In the boxplots, the thick line indicates the median, the box limits indicate the upper and lower quartiles, and the whiskers extend to min and max values (d) Principal Component Analysis (PCA) on the WGBS results. The 4 KOs cluster together, away from serum cells and from 2i cells. (e) Liquid chromatography

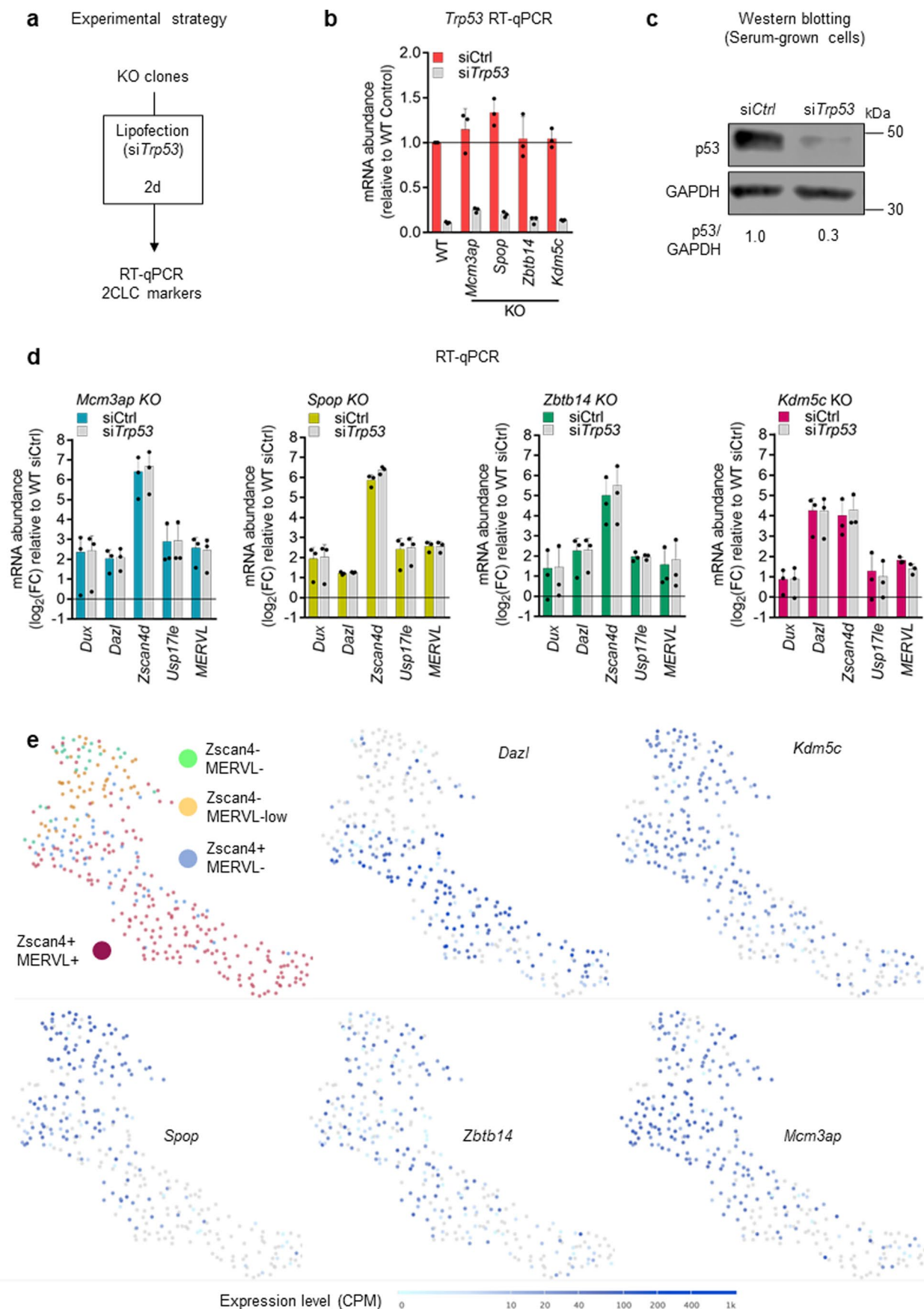
followed by tandem Mass Spectrometry (LC-MS/MS) confirms the decrease of DNA methylation in *Mcm3ap* and *Spop* KOs. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ (two-sided Holm-Sidak post-hoc test following ANOVA). (f) A restriction-enzyme based technique (LUMA) confirms the decrease of DNA methylation in *Mcm3ap* and *Spop* KOs. *** $P < 0.001$, **** $P < 0.0001$ (two-sided Holm-Sidak post-hoc test following ANOVA). Data from n = 3 independent KO clones are presented as mean values \pm SD for panels b and e-f.



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

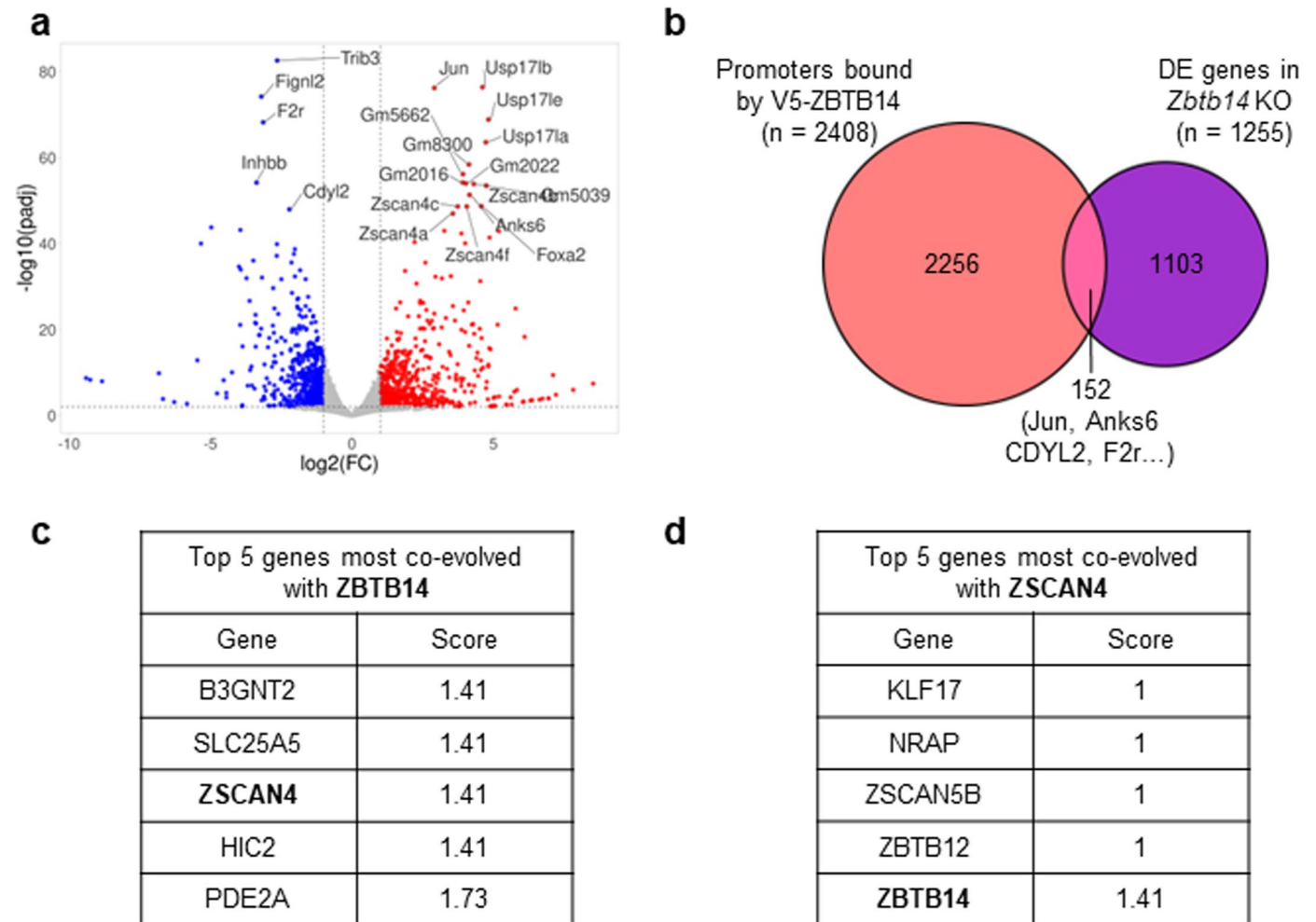
Extended Data Fig. 4 | Additional characterizations of the transcriptional 2-cell-like signature in the *Mcm3ap*, *Spop*, *Zbtb14*, and *Kdm5c* Kos. (a) RNA-seq statistics: differentially expressed genes ($|FC| > 2$; $FDR < 1\%$) in each KO condition. (b) Genome browser tracks depicting RNA-seq profiles of 2CLC markers reactivated in the KOs. (c) RT-qPCR analysis: genetic rescue of each KO suppresses 2CLC marker induction. (d) Gene Set Enrichment Analysis (GSEA): the 2CLC signature is enriched in each individual KO. (e) GSEA: metabolic pathways downregulated in 2CLCs³⁸ are also downregulated in the KOs. (f)

Increased ZSCAN4-positive staining in the indicated KO populations. Data from $n = 812$ cells; $n = 501$ cells; $n = 757$ cells; $n = 576$ cells and $n = 547$ cells are presented for, DASH, *Mcm3ap* KO, *Spop* KO, *Zbtb14* KO and *Kdm5c* KO, respectively. **** $P < 0.0001$ (Dunn's post-hoc test following Kruskal-Wallis test) (g) Reactivation of an LTR-GFP reporter after siRNA of the indicated factors. Data from $n = 3$ independent replicates are presented as mean values \pm SD. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-sided Holm-Sidak post-hoc test following ANOVA).



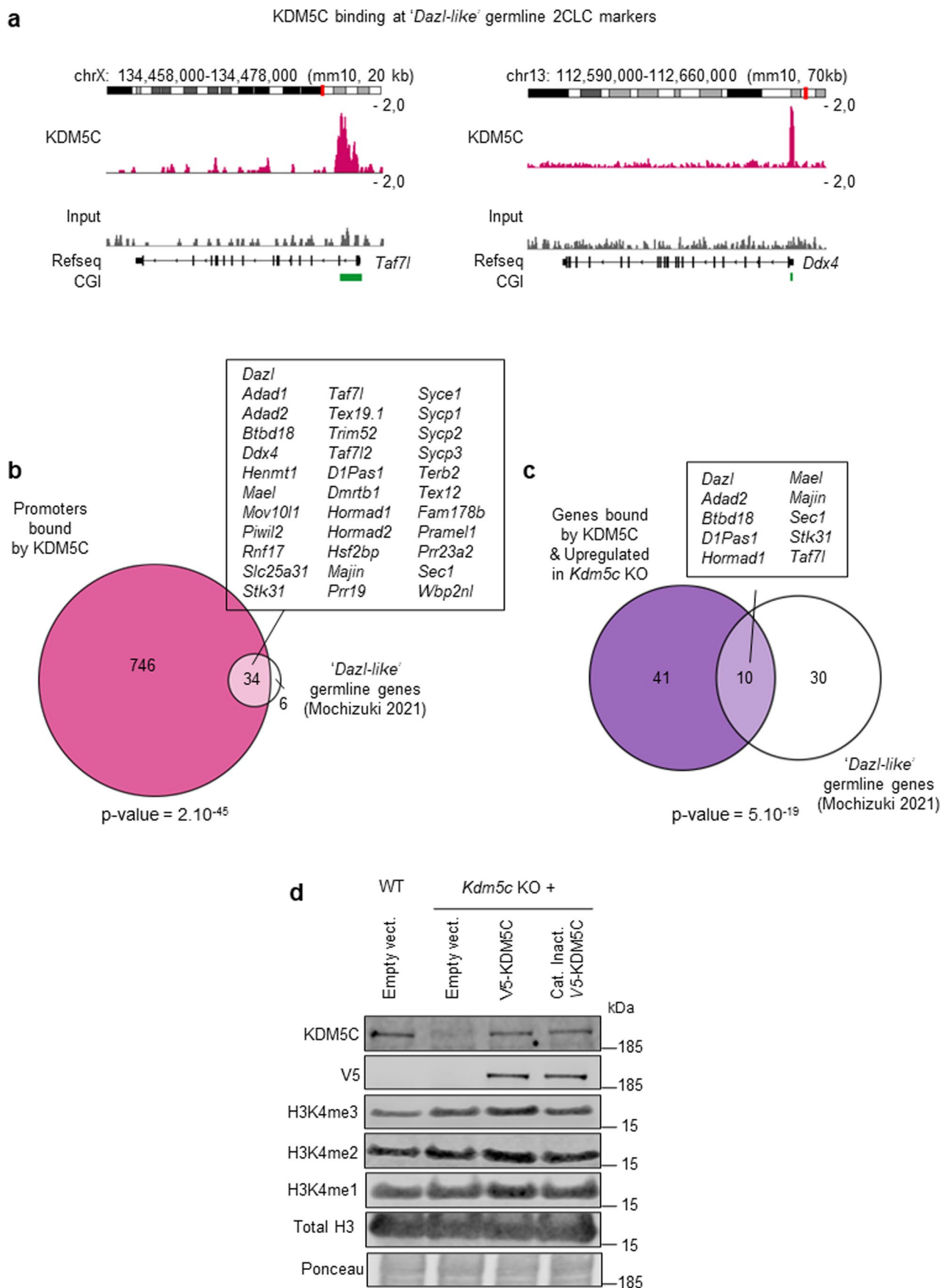
Extended Data Fig. 5 | *Trp53* is not required for the activation of 2CLC markers in the *Mcm3ap*, *Spop*, *Zbtb14*, and *Kdm5c* Kos. (a) Experimental scheme for *Trp53* depletion. **(b)** RT-qPCR analysis: *Trp53* mRNA is efficiently depleted in all KOs. **(c)** The p53 protein is efficiently depleted, example of western blot on WT

cells. **(d)** RT-qPCR analysis: the induction of 2CLC markers is *Trp53*-independent. **(e)** Expression of the indicated genes in single ES cells sorted according to ZSCAN4 and MERVL expression. Data from $n = 3$ independent KO clones are presented as mean values \pm SD for panels b and d.



Extended Data Fig. 6 | Limited overlap between ZBTB14 binding and transcriptional response to ZBTB14 loss. (a) MA-plot on the RNA-seq data from *Zbtb14* KO cells relative to WT ES cells. This is the same data as in Fig. 4a. (b) Only a minority of promoters bound by ZBTB14 respond transcriptionally to *Zbtb14*

KO, and vice versa. (c) and (d) ZBTB14 and ZSCAN4 have co-evolved closely. The scores are from CladeOScope, <http://cladeoscope.cs.huji.ac.il>. A smaller score means tighter co-evolution.

**Extended Data Fig. 7 | KDM5C binds additional germline/2CLC genes in ESCs.**

(a) Genome browser tracks illustrating the binding of KDM5C to the *Taf7l* and *Ddx4* promoters. (b) 85% of the germline genes that are regulated similarly to *Dazl* are bound by KDM5C in ESCs. (c) 25% of the germline genes that are

regulated similarly to *Dazl*^{fls} are bound and silenced by KDM5C in ESCs.

(d) KDM5C expression has no effect on the global level of H3K4me1/2/3. Western blotting was performed in the indicated cellular backgrounds. Significance: hypergeometric test for panels b-c.

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

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

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | 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 custom software was used in this study. FACS data were collected with BD FACSAria Fusion and FlowJo software v10.8.1. Sequencing data was collected using Illumina HiSeq 1500 or NovaSeq 6000 or HiSeq X Ten platforms.
Data analysis	All softwares used for data analysis are publicly available. Data analysis was performed using R software v4.0, ImageJ v1.53, DESeq2 v1.30.0, MAGeCK v0.5.9, DAVID v6.8, Cytoscape v3.7.2, GSEA v4.1.0, IGV v2.11, Trimmomatic v0.39, STAR v2.7.5c, Picard-tools v2.23.4, Samtools v1.10, VisRseq v0.9.12, deeptools v3.3.0, iGEM v1.315, Bowtie2 v2.4.1, homer2 findMotifs v4.11, E-utilites esearch and efetch v15.9, SRAToolkit v2.8.0, STAR v2.7.5c, MarkDuplicates v2.23.4, Samtools view v1.10, bamCoverage v3.3.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data generated in this study are available in the Gene Expression Omnibus under accession number GSE221710. For Fig. 1, Fig. 4 and Extended Data Fig. 5, published RNA-seq were reanalyzed from GSE33923 (Macfarlan 2012); E-MTAB-2684 (Ishuchi 2015); GSE75751 (Eckersley-Maslin 2016); GSE71434 (Zhang

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](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	No statistical method was used to pre-determine sample size. For CRISPR screening, duplicate independent screens at 100x coverage were performed, in line with literature recommendations. For subsequent experiments, we used 3 independent replicates whenever possible to better represent biological variations.
Data exclusions	No data were excluded from the analysis.
Replication	Most of the screen hits were validated through generation of independent pooled knockout lines and functionally assayed. Reproducibility between independent RNA-seq and WGBS samples was assessed using clustering approaches (PCA, unsupervised hierarchical clustering) with satisfying results. Replicated confirmed other experimental findings.
Randomization	Randomization was not used in this cell-culture based study, as all generated cell lines were generated from the same parental mouse embryonic stem cell line.
Blinding	Blinding was not relevant to this study as all measurements were made with objective quantitative methods, mostly sequencing-based.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

For Western blotting: α -DAZL (Abcam #ab34139; 1:500), α -KDM5C (Bethyl Laboratories #A301-034A; 1:1000), α -DPPA2 (Merck #MAB4356; 1:1000), α -V5 (Abcam #ab206566; 1:1000), α -MuERVL-Gag (HuaBio #ER50102; 1:1000), α -ZSCAN4 (Merck #AB4340; 1:5000), α -p53 (CST #2524; 1:1000), α -TUBULIN (Abcam #7291; 1:10000), α -GAPDH (Abcam #ab9485; 1:10000), IRDye 800CW Donkey α -Rabbit (Licor #926-32213, 1:15000), IRDye 680RD Donkey α -Mouse (Licor #926-68072, 1:15000).

For MeDIP: α -5-methylcytosine (Diagenode #C15200081, clone 33D3, 1 μ g/IP).
For ChIP-seq: α -KDM5C (Bethyl Laboratories #A301-034A, 1 μ g/50 μ g chromatin).
For CUT&RUN-seq: α -V5 (Abcam #ab206566, 1 μ g/IP).

Validation

α -DAZL (Abcam #ab34139) - WB positive control: human and mouse testis tissue lysate.
 α -DPPA2 (Merck #MAB4356) - validated for WB.
 α -V5 (Abcam #ab206566) - validated for IF, IP and WB.
 α -MuERVL-Gag (HuaBio #ER50102) - IF positive control : mouse embryo.
 α -ZSCAN4 (Merck #AB4340) - WB positive control : ES cell lysate.
 α -p53 (CST #2524) - validated for IF, ChIP and WB. Cited in 1200+ publications.
 α -TUBULIN (Abcam #7291) - validated for IF and WB. Cited in 800+ publications.
 α -GAPDH (Abcam #ab9485) - validated for IF and WB. Cited in 2200+ publications.
 α -5-methylcytosine (Diagenode #C15200081) - validated for MeDIP-seq.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The ES-J1 cell line (129S4/SvJae, XY) was kindly provided by the laboratory of Deborah Bourc'his. HEK293T cell line were from the ATCC.
Authentication	Every clonal knockout cell lines used has been precisely genotyped by NGS. HEK293T cell line was not authenticated.
Mycoplasma contamination	All cell lines were routinely tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

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 <i>May remain private before publication.</i>	All sequencing data generated in this study are available in the Gene Expression Omnibus under accession number GSE221710.
Files in database submission	ChIPseq_KDM5C_ES_WT_rep1.fastq ChIPseq_input_ES_WT_rep1.fastq ChIPseq_KDM5C_ES_WT_rep1.bw ChIPseq_input_ES_WT_rep1.bw ChIPseq_KDM5C_ES_WT_rep2.fastq ChIPseq_input_ES_WT_rep2.fastq ChIPseq_KDM5C_ES_WT_rep2.bw ChIPseq_input_ES_WT_rep2.bw
Genome browser session (e.g. UCSC)	bigwig files are provided in the GEO submission for visualisation in any genomic browser.

Methodology

Replicates	We sequenced two independent replicates of ChIP and input samples from WT ES cells.
Sequencing depth	Sequencing was performed in paired-end 150bp. ChIPseq_KDM5C_ES_WT_rep1 : 17 190 599 total reads, 9 390 289 uniquely mapped reads. ChIPseq_input_ES_WT_rep1 : 11 857 536 total reads, 7 032 288 uniquely mapped reads. ChIPseq_KDM5C_ES_WT_rep2 : 15 544 549 total reads, 13 194 743 uniquely mapped reads. ChIPseq_input_ES_WT_rep2 : 16 578 897 total reads, 9 533 174 uniquely mapped reads.
Antibodies	α -KDM5C, Bethyl Laboratories #A301-034A
Peak calling parameters	Peaks were called using MACS2 with default parameters, and motif enrichment analysis was performed with HOMER.
Data quality	MACS2 analysis called 2173 peaks (at FDR < 5%) for KDM5C ChIP-seq, including 1547 peaks (71%) above 4-fold enrichment.
Software	FASTQ reads were trimmed using Trimmomatic (v0.39) and parameters: ILLUMINACLIP:illumina_adapters.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:36. Trimmed reads were aligned using Bowtie2 (v2.4.1) in --local mode. Following alignment, Picard (v2.23.4) CleanSam, SamFormatConverter, SortSam and MarkDuplicates were used to generate a duplicate-marked bam file. The resulting bam files were converted to bigwig using deeptools (v3.3.0) Bamcoverage and options --ignoreDuplicates --normalizeUsing CPM --minMappingQuality 10 --ignoreForNormalization chrX chrY chrM.