



Data Article

RNA-Seq dataset of wild type and PRC2 mutant mouse embryonic stem cells undergoing ground state conversion



Cameron M. Chapa^{a,b}, Dustin R. Fetch^c, Amina A. Jumamyradova^a, Alexey A. Soshnev^{a,*}

^a Department of Neuroscience, Developmental and Regenerative Biology, The University of Texas at San Antonio, San Antonio, TX 78249, USA

^b Current affiliation: The David Rockefeller Graduate Program in Bioscience, The Rockefeller University, New York City, NY 10065, USA

^c Graduate Program in Developmental and Regenerative Sciences, The University of Texas at San Antonio, San Antonio, TX 78249, USA

ARTICLE INFO

Article history:

Received 30 June 2025

Revised 21 July 2025

Accepted 29 July 2025

Keywords:

Polycomb

Pluripotency

Transcriptome

mESC

2i

Naïve

Metastable

ABSTRACT

Mouse embryonic stem cells (mESCs) represent an exceptional model for understanding how transcriptional responses are regulated by signalling pathways during development. Treatment with a cocktail of MEK and GSK3 β inhibitors ("2i") induces ground state pluripotency, characterized by increased self-renewal, reduced DNA methylation, and uniformly high expression of pluripotency markers. Polycomb Repressive Complex 2 (PRC2) is a key developmental regulator controlling stem cell self-renewal and differentiation decisions, and altered expression of PRC2 target genes is a signature of 2i-mediated ground state conversion. Here, we generated a comprehensive RNA sequencing dataset from mESCs subjected to 2i conversion time-course across five time points between 0 and 96 h. We analysed two independently derived wild type lines, and two isogenic Cas9-edited lines carrying loss-of-function mutations in core PRC2 subunits Enhancer of Zeste Homolog 2 (Ezh2) or Embryonic Ectoderm Development (Eed). These data provide a resource to understand the temporal patterns of transcriptional re-

* Corresponding author.

E-mail address: alexey.soshnev@utsa.edu (A.A. Soshnev).

sponses to MEK and GSK3 β inhibitors and explore the role of PRC2 function in regulation of pluripotency circuit.

© 2025 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Biology
Specific subject area	Transcriptome characterization of mouse embryonic stem cells
Type of data	Raw (.fastq), Processed (.csv tables)
Data collection	The RNA-Seq data were obtained from 60 samples of mouse embryonic stem cells, representing five individual time points in 2i conversion time-course across four genotypes, performed in biological triplicates. Total RNA from the samples was extracted and enriched for mRNA using polyA+ selection. Libraries were prepared using NEBNext Ultra II kit, pooled and sequenced using Illumina NextSeq 500 platform with a single-end read length of 75 bp. Demultiplexed .fastq files were pre-processed using MultiQC quality assessment and mapped to GRCh39 mouse genome using STAR aligner. Variance-stabilized transformed counts were generated using DESeq2.
Data source location	The University of Texas at San Antonio, San Antonio, TX 78249 USA
Data accessibility	Repository name: Gene Expression Omnibus (GEO) Data identification number: GSE237656 Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE237656
Related research article	n/a

1. Value of the Data

- Mouse embryonic stem (mES) cells are a model of choice for the studies of pluripotency and differentiation, as well as the role of Polycomb Group proteins in development. 2i cocktail is widely used in mES cell media, often without regard for profound transcriptional differences between serum/LIF and 2i conditions. These data present a time-series of gene expression studies in mES cells during 2i “conversion” in two wild type and two PcG mutant backgrounds, providing detailed reference for the effects of 2i and Polycomb in naïve pluripotency.
- These data may provide the basis for exploratory analyses of transcriptional alterations induced by 2i treatment and PcG depletion and may facilitate future studies of how specific genes and/or gene sets are regulated in development.
- Two independently generated PcG mutants may allow for separation of function between subunits of PRC2 core complex, wherein *Ezh2*^{-/-} represents loss of major methyltransferase enzyme while retaining residual methylation due to *Ezh1* paralog activity, and *Eed*^{-/-} represents full loss of PRC2.

2. Background

Mouse embryonic stem cells (mESCs) are among the most versatile models to investigate developmental transitions [1]. When cultured in serum-containing media supplemented with leukaemia inhibitory factor (LIF) [2], mESCs represent a heterogeneous population, with fluctuating levels of pluripotency markers and developmental potential. This metastable state is further characterized by remarkable differences of chromatin modifications observed within the neighbouring cells, often interpreted as indicative of a fluid range of developmental potential [3].

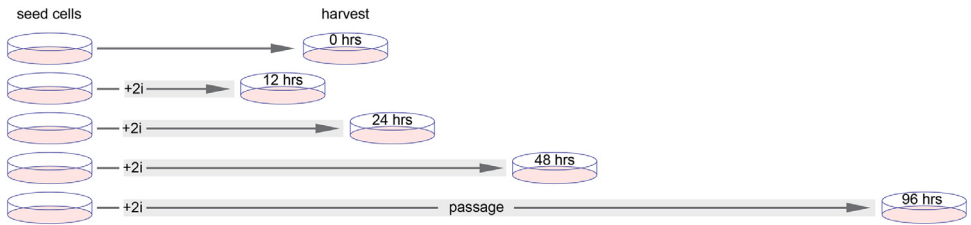


Fig. 1. Outline of the experimental design. Samples were collected at 0, 12, 24, 48, and 96 h after addition of 2i. Cells were harvested at 70–80 % confluence using Trizol reagent. Extracted RNA was DNaseI-treated, selected for polyA+ fraction, reverse-transcribed, barcoded, and pooled for sequencing using Illumina NextSeq 500. Three biological replicates of each genotype (1A and V19, corresponding to wild type, and *Ezh2*^{KO} and *Eed*^{KO}, corresponding to PRC2 mutants) were grown and processed separately.

Chromatin is a complex of nucleic acids and regulatory and structural proteins, of which histones are a major component. Histones, together with the ~147 base pairs of DNA, are assembled into nucleosomes, and are modified by regulatory enzymatic complexes, often referred to as “writers” of post-translational modifications (PTMs), in response to both signalling and metabolic events. Polycomb Repressive Complex 2 is a paradigmatic “writer” which established the methylation of lysine 27 in the N-terminal tail of core histone H3 (H3K27me), a PTM interpreted by several “reader” complexes and associated with developmentally repressed genes [4]. In mammals, PRC2 contains four essential core subunits, including the “reader” of H3K27me Embryonic Ectoderm Development (Eed), and the methyltransferase “writer” Enhancer of zeste homolog 1/2 (*Ezh1/2*), of which *Ezh2* is the primary methyltransferase in mESCs [5,6]. Additional subunits regulate local recruitment and activity of PRC2; of note, many of those were first identified and characterized in mESC model [7,8].

3. Data Description

PRC2 activity is variegated in metastable mESC culture, reflective of cell-to-cell variance in developmental potential [9]. To improve the homogeneity of the mESC culture, a cocktail of two inhibitors, termed 2i, is widely used [10–14]. These inhibitors, targeting mitogen-activated protein kinase (MEK) and glycogen synthase kinase 3 (*GSK3β*), alter mESC signalling to reinforce the pluripotency maintenance network, resulting in naïve, or ground state cell culture with altered transcriptional landscape characterized by low levels of *Myc* and high expression of *Nanog* and *ESRRB* [13]. Here, we extend these studies in two complementary ways. First, we provide the transcriptional profile of two independently derived mESC lines in 2i media, spanning five time points across 96 h of exposure (Fig. 1), corresponding to estimated six consecutive cell divisions. Second, to examine the role of PRC2 in transcriptional response to 2i, we derived two isogenic knock-out mESC lines lacking either core methyltransferase subunit *Ezh2*, or H3K27me-reader *Eed* (Figs. 2 and 3). Both mutations effectively abolish global H3K27 methylation (Fig. 3) yet may reveal non-overlapping functions of individual PRC2 subunits. Together, these data may advance our understanding of stem cell biology by detailing temporal effects of ground state conversion and specifically resolve the role of PRC2 in these developmental processes.

4. Experimental Design, Materials and Methods

Cell culture. Cells were grown in vacuum-plasma treated plastic flasks and dishes (Falcon) preincubated with 0.1 % gelatin (Embryomax Millipore, ES-006-B) in mES cell media containing Knock-Out DMEM (Gibco 10829018) supplemented with 15 % FBS (Gemini 100–525), 2 mM L-Glutamine (Gibco 25030081), 100 U/ml Penicillin/Streptomycin (Gibco 15140122), 0.1 mM β-

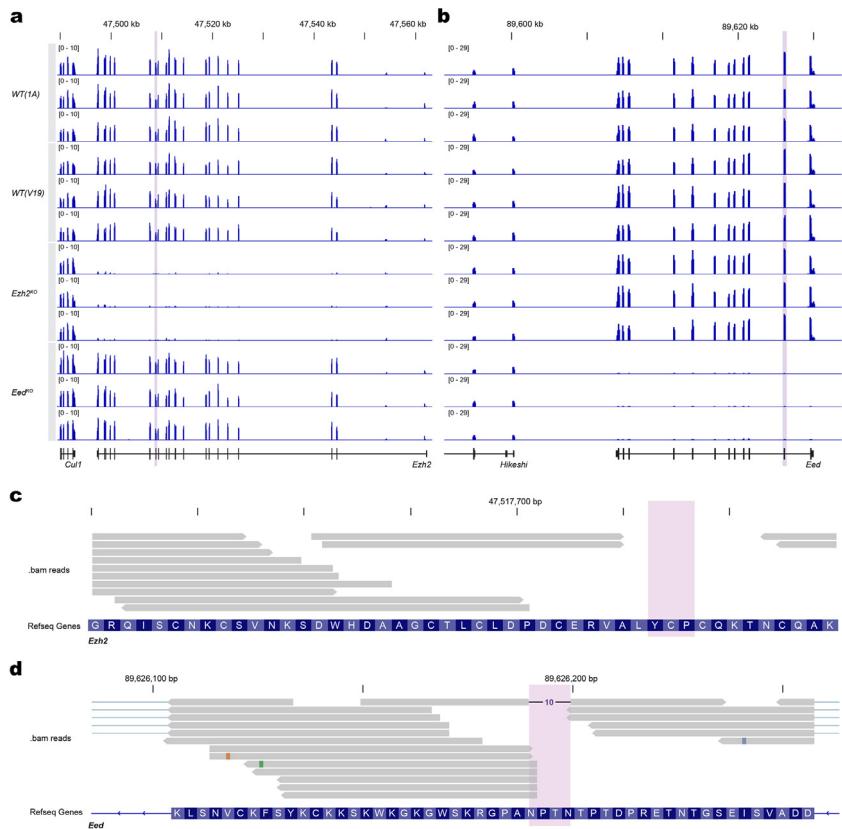


Fig. 2. Validation of PRC2 knock-out in RNA-Seq data. IGV browser view of *Ezh2* (a) and *Eed* (b) genomic regions with all replicates at 0 hr of treatment shown individually. Regions encompassing deletions are highlighted by light purple bars and expanded below. (c,d) individual reads plotted from .bam files corresponding to *Ezh2*^{KO} (0 hr, replicate 1, c) and *Eed*^{KO} (0 hr, replicate 1, d) to illustrate extent of genomic deletions (highlighted in purple).

mercaptoethanol (Fisher O3446I-100) and 1:1000 LIF (Gemini 400–495). Cells were seeded at 200,000 per 5 ml in T25 flasks with media exchange every 24 h and passaged every 48 h. For 2i treatment, CHIR99021 and PD0325901 (Stemgent) were added at 3 μ M and 1 μ M, respectively, and exchanged daily. Two wild type mouse Embryonic Stem Cell (mESC) cell lines, termed 1A (ESC-1) and V19 (ESC-V19) were independently derived and extensively characterized previously [12].

Derivation and validation of *Ezh2* and *Eed* knock-out lines. Guide RNA sequence targeting conserved second exon of *Eed* (TGTCGATTGGCGTATTG) was designed using CRISPOR [15]. Guide RNA sequence targeting catalytic domain of *Ezh2* (TGATGGTTAATGGTGACCAC) was reported previously [16]. Annealed and phosphorylated oligonucleotides were cloned into pSpCas9(BB)-2A-GFP (PX458) plasmid (Addgene #48138) linearized with BbsI (New England Biolabs R3539). Inserts were validated by Sanger sequencing (Azenta) and transfected into early passage (p10) culture of 1A cells using Xfect mESC Transfection Reagent (Takara 631320). Live GFP+ singlets were sorted into gelatin-coated 96-well plates (Rockefeller University Flow Cytometry Resource Center) and expanded over 6–8 passages for subsequent validation.

RNA isolation and cleanup. At specified time points of the time course, media was aspirated from the wells of 12-well plates, followed by gentle rinse with warm PBS (Corning 21–040-CV) and immediate lysis in 1 ml of TRIzol reagent (Thermo 15596026) added directly into

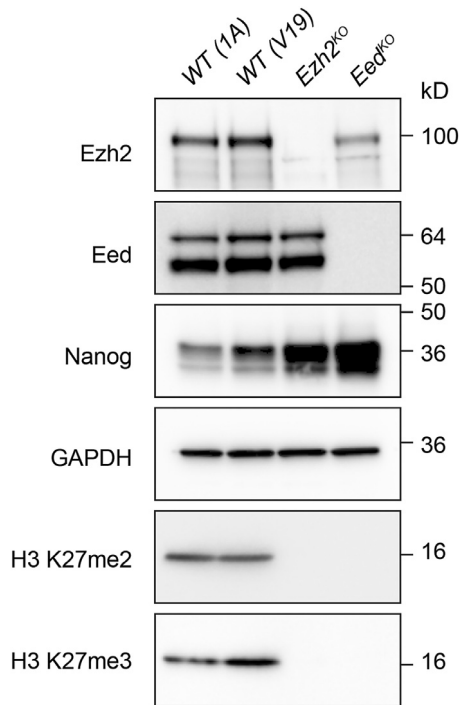


Fig. 3. Western blot validation of PRC2 knock-out lines. Whole cell lysates of wild type and PRC2 mutant mESC lines grown in serum/LIF media were resolved on SDS-PAGE, transferred to PVDF membrane, and probed with antibodies to indicated epitopes. GAPDH serves as loading control. Note loss of H3 K27 methylation in both EZH2 and EED knock-out backgrounds. Increased Nanog accumulation in PRC2 mutants is corroborated by RNA-Seq in Fig. 4e. Matching uncropped, unedited membranes are shown in Supplemental Fig. 1.

the well. Following chloroform addition, aqueous phase was separated using 5Prime PhaseLock Heavy tubes (QuantaBio 2302830), precipitated with isopropanol, and washed twice with 75 % ethanol. Air-dried RNA was resuspended in RNase-free water (Corning 46-000-CM), and 5 µg total RNA per sample were treated with DNaseI using Qiagen RNeasy on-column treatment and cleanup protocol (Qiagen 79254 and Qiagen 74104).

Library preparation and RNA Sequencing. 1 µg of DNaseI-treated total RNA was subjected to polyA enrichment using magnetic isolation (NEB E7490), and mRNA-enriched fraction was converted to barcoded library using NEBNext Ultra II kit (NEB E7770) according to manufacturer's instructions, using Ampure XP SPRI beads (Beckman Coulter A63881) for size selection, and 96 index primers for barcoding (NEB E6609). Molarity of each library was estimated from electrophoresis profile (Agilent Tapestation 4200) and equimolar amounts were combined in one pool. Samples were sequenced in two Illumina NextSeq 500 High Output 75 SR flowcells, demultiplexed, and merged to combine two runs into single .fastq files. Quality control and initial processing of the resulting data are outlined below.

Cell line identity. mES cells were derived previously [12] and authenticated using CellCheck STR service (IDEXX BioAnalytics, Supplemental Table 1 and 2). Early passage (p10) cells were used for gene editing, and matched passage was used for final experiment. All cells were routinely tested for mycoplasma (e-Myco PCR Detection kit, Boca Scientific 25235). Gene edited clones were validated using conventional PCR followed by Sanger sequencing (Azenta) using the following primer pairs: Ezh2-ex19-fw: GCCAGGTGGCTCTATGAATAA, Ezh2-ex19-rev: ACTTGTGGGCTGAGTAGTTTC; Eed-ex2-fw: TTAAGATGTGTTGGCATTGCT, Eed-ex2-rev: GTAGCTGTACAGGTGTTTGT. *Ezh2*^{KO} clone contains a homozygous deletion of seven base

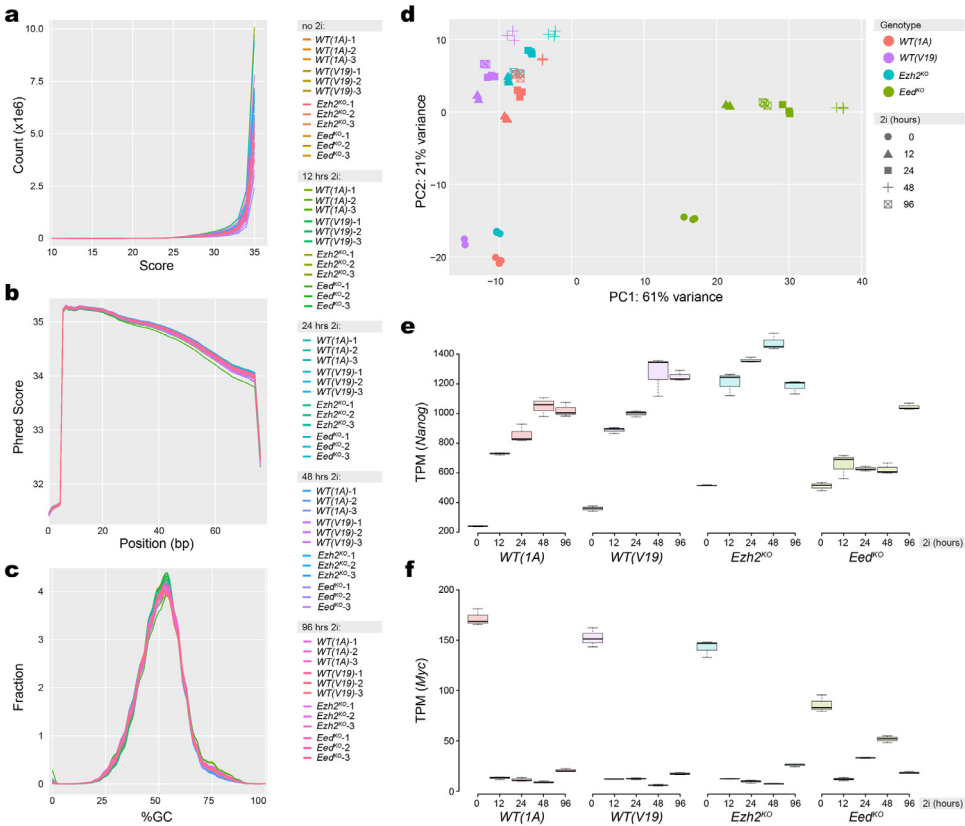


Fig. 4. Technical validation of the RNA-Seq data. (a–c) Demultiplexed samples were processed using MultiQC pipeline; quality indicators, including mean sequence quality score (a), Phred quality score per basepair (b) and GC content percentage (c) are summarized for each sample. (d) Exploratory analyses of data using principal components. Note expected segregation of all samples with 2i treatment along PC2, and distinct appearance of *Ezh2*^{KO} and *Eed*^{KO} along PC1, suggesting non-equivalent effects of two knock-outs. (e,f) TPM counts for two well-characterized responder genes *Nanog* (e) and *Myc* (f) across the 2i time-course for each genotype, shown as boxplots; note increased *Nanog* expression in PRC2 mutants corroborated by Western analyses shown in Fig. 3.

pairs corresponding to chr6: 47,517,725 - 47,517,732 (mm39) interval. *Eed*^{KO} clone contains a homozygous deletion of ten base pairs corresponding to chr7: 89,626,190 - 89,626,199 (mm39) interval. Both clones show no defects in viability compared to wild type parental line, and do not accumulate detectable amounts of respective PRC2 subunits and core histone H3 K27me2/3 (Fig. 2, Supplemental Fig. 1).

Western blot. Equivalent of 20,000 cells was loaded per lane on a 4–20 % gradient Tris-Glycine SDS-PAGE (Novex XP04205BOX) and transferred on 0.2 μm PVDF membrane, incubated with DirectBlue71 to assess protein transfer efficiency, blocked with 5 % non-fat milk in PBS-Tween-20, and incubated with primary antibodies to Eed (Cell Signaling E4L6E, #85322), Ezh2 (Cell Signaling AC22, #3147), H3 K27me2 (Cell Signaling D18C8 #9728), H3 K27me3 (Cell Signaling C36B11, #9733), GAPDH (Cell Signaling D16H11, #5174), and Nanog (BioLegend SER211, #678102) at 1:1000. Secondary HRP-conjugated anti-rat (Jackson 712–035–153), anti-mouse (Jackson 615–035–214) and anti-rabbit (Jackson 611–035–215) antibodies were used at 1:10,000 dilution. After final washes with TBS-Tween-20, membranes were incubated with West Pico PLUS ECL substrate (Thermo 34580) and imaged using BioRad ChemiDoc MP system. Unadjusted, uncropped images of membranes are shown in Supplemental Fig. 1.

Table 1
RNA-Seq sample details. 2i, CHIR99021 and PD0325901.

Accession	Treatment	mESC line	Sample Count
GSE237656	none	wild type (1A)	3
		wild type (V19)	3
		1A-Ezh2 ^{KO}	3
		1A-Eed ^{KO}	3
	2i-12 h	wild type (1A)	3
		wild type (V19)	3
		1A-Ezh2 ^{KO}	3
		1A-Eed ^{KO}	3
	2i-24 h	wild type (1A)	3
		wild type (V19)	3
		1A-Ezh2 ^{KO}	3
		1A-Eed ^{KO}	3
	2i-48 h	wild type (1A)	3
		wild type (V19)	3
		1A-Ezh2 ^{KO}	3
		1A-Eed ^{KO}	3
	2i-96 h	wild type (1A)	3
		wild type (V19)	3
		1A-Ezh2 ^{KO}	3
		1A-Eed ^{KO}	3

RNA and cDNA integrity, and library validation. RNA purity was assured using Nanodrop spectrophotometer (Thermo) A260/A230 and A260/A280 ratios both before and after DNaseI treatment. Quantity and integrity of NGS libraries was validated using Tapestation automated electrophoresis system (Agilent) and Qubit fluorometer (Thermo).

RNA-Seq data quality. Quality assessment of the bulk RNA-Seq data was performed using MultiQC v1.14 [17] (Fig 4a–c). Samples were aligned to the GRCh39 genome using STAR (v. 2.7.10b) [18]. Raw counts matrix was generated using featureCounts (v. 2.0.3) [19]. Transformed counts matrix for exploratory analyses of the data such as principal component analysis (Fig. 4d) was created in DESeq2 (v1.40.1) using variance-stabilizing-transformation [20]. Samples were filtered to keep gene rows with at least five total counts or higher. Data were pre-processed using R and visualized using ggplot2 v. 3.4.2.

Biological replicates. Experiment was designed to ensure extensive biological replication. First, three independent biological replicates, per genotype, per time point were analysed. Second, two independently derived wild type lines were profiled. Third, two PRC2 null mutants represented by *Eed* and *Ezh2* knock-out lines, were generated – together corresponding to six biological replicates per time point for wild type and PRC2 mutant conditions each.

Limitations

Limitations of the study stem from both experimental design and biological factors. First, while cell lines used here have been extensively referenced in publications, we acknowledge that many additional mES lineages are available. While to our knowledge no comprehensive transcriptomic and proteomic comparison of multiple mouse mES cell lines exists, we recognize the possibility that cell lines derived from other mouse lineages may have a distinct response to both 2i cocktail and PRC2 loss. Second, our time-course experiment was limited to 96 h, roughly corresponding to six consecutive cell divisions [12]. While similar or shorter conversion times are standard, we acknowledge that long-term effects of 2i treatment have not been investigated in our study. Third, our study of PRC2 mutants only included two core complex subunits. While *Eed* is essential for PRC2 function and *Ezh2* represents the major and perhaps best-characterized subunit, we acknowledge that studies of additional subunits could identify additional phenotypes. Finally, it is important to note that RNA-Seq assesses abundance of processed mRNA tran-

scripts but does not directly test transcriptional activity or protein levels. While correlations between the three are strong, additional experiments are necessary to make specific conclusions. We hope our data can provide useful clues to such future investigations.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Cameron M. Chapa: methodology, investigation, data curation, visualization, writing - original draft preparation. **Dustin R. Fetch:** investigation, visualization. **Amina Jumamyradova:** investigation, visualization. **Alexey A. Soshnev:** conceptualization, writing - original draft preparation, writing - reviewing and editing.

Data Availability

RNA-Seq data were deposited in Gene Expression Omnibus (GEO) under accession number GSE237656. The dataset contains 60 samples, corresponding to five time points across four genotypes, and three biological replicates per sample. Detailed sample information is listed in [Table 1](#).

Acknowledgements

We are grateful to the personnel of Flow Cytometry Resource Center (RRID:SCR_017694) and Genomics Resource Center (RRID:SCR_020986) at the Rockefeller University for the cell sorting and sequencing services. This project was initiated in the Laboratory of Chromatin Biology and Epigenetics at the Rockefeller University, and the authors are grateful to Bryce Carey, Richard Phillips, and C. David Allis for technical and conceptual insights. Research in the Soshnev laboratory is supported by NCI R01 CA234561, American Cancer Society IRG-21-147-01-IRG, CPRIT awards RP240446 and 240068, and institutional funds from the University of Texas at San Antonio.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.dib.2025.111957](https://doi.org/10.1016/j.dib.2025.111957).

References

- [1] G. Martello, A. Smith, The nature of embryonic stem cells, *Annu Rev. Cell Dev. Biol.* 30 (2014) 647–675.
- [2] R.L. Williams, D.J. Hilton, S. Pease, T.A. Willson, C.L. Stewart, D.P. Gearing, E.F. Wagner, D. Metcalf, N.A. Nicola, N.M. Gough, Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells, *Nature* 336 (1988) 684–687.
- [3] A. Smith, Formative pluripotency: the executive phase in a developmental continuum, *Development* 144 (2017) 365–373.
- [4] J.J. Kim, R.E. Kingston, Context-specific polycomb mechanisms in development, *Nat. Rev. Genet.* 23 (2022) 680–695.
- [5] C.H. Lee, M. Holder, D. Grau, R. Saldana-Meyer, J.R. Yu, R.A. Ganai, J. Zhang, M. Wang, G. LeRoy, M.W. Dobenecker, et al., Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2, *Mol. Cell* 70 (2018) 435–448 e435.
- [6] X. Shen, Y. Liu, Y.J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G.C. Yuan, S.H. Orkin, EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency, *Mol. Cell* 32 (2008) 491–502.
- [7] E. Healy, M. Mucha, E. Glancy, D.J. Fitzpatrick, E. Conway, H.K. Neikes, C. Monger, G. Van Mierlo, M.P. Baltissen, Y. Koseki, et al., PRC2.1 and PRC2.2 synergize to coordinate H3K27 trimethylation, *Mol. Cell* 76 (2019) 437–452 e436.
- [8] C.H. Loh, S. van Genesen, M. Perino, M.R. Bark, G.J.C. Veenstra, Loss of PRC2 subunits primes lineage choice during exit of pluripotency, *Nat. Commun.* 12 (2021) 6985.
- [9] R.S. Illingworth, J.J. Holzspies, F.V. Roske, W.A. Bickmore, J.M. Brickman, Polycomb enables primitive endoderm lineage priming in embryonic stem cells, *Elife* (2016) 5.
- [10] C. Mulas, T. Kalkan, F. von Meyenn, H.G. Leitch, J. Nichols, A. Smith, Defined conditions for propagation and manipulation of mouse embryonic stem cells, *Development* (2019) 146.
- [11] L. Zhong, M. Gordillo, X. Wang, Y. Qin, Y. Huang, A. Soshnev, R. Kumar, G. Nanjangud, D. James, C. David Allis, et al., Dual role of lipids for genome stability and pluripotency facilitates full potency of mouse embryonic stem cells, *Protein Cell* (2023).
- [12] B.W. Carey, L.W. Finley, J.R. Cross, C.D. Allis, C.B. Thompson, Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells, *Nature* 518 (2015) 413–416.
- [13] H. Marks, T. Kalkan, R. Menafrá, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A.F. Stewart, A. Smith, et al., The transcriptional and epigenomic foundations of ground state pluripotency, *Cell* 149 (2012) 590–604.
- [14] G. van Mierlo, R.A.M. Dirks, L. De Clerck, A.B. Brinkman, M. Huth, S.L. Kloet, N. Saksouk, L.I. Kroeze, S. Willems, M. Farlik, et al., Integrative proteomic profiling reveals PRC2-dependent epigenetic crosstalk maintains ground-State pluripotency, *Cell Stem Cell* 24 (2019) 123–137 e128.
- [15] J.P. Concordet, M. Haeussler, CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens, *Nucleic. Acids. Res.* 46 (2018) W242–W245.
- [16] J. Shi, E. Wang, J.P. Milazzo, Z. Wang, J.B. Kinney, C.R. Vakoc, Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains, *Nat. Biotechnol.* 33 (2015) 661–667.
- [17] P. Ewels, M. Magnusson, S. Lundin, M. Kaller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics* 32 (2016) 3047–3048.
- [18] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics* 29 (2013) 15–21.
- [19] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* 30 (2014) 923–930.
- [20] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.* 15 (2014) 550.