ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

Yara
Souto Becerra

Tese de doutoramento

Determining the role of the zinc-finger protein ZMYM2 in pluripotency

Santiago de Compostela, 2022

Programa de Doutoramento en Medicina Molecular

TESE DE DOUTORAMENTO

DETERMINING THE ROLE OF THE ZINC-FINGER PROTEIN ZMYM2 IN PLURIPOTENCY

Yara Souto Becerra

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN MEDICINA MOLECULAR

SANTIAGO DE COMPOSTELA

2022

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR/TITOR DA TESE

D./Dna. **Miguel Ángel Fidalgo Pérez**

En condición de: **Titor/a e director/a**

Título da tese: **Determining the role of the zinc-finger protein ZMYM2 in pluripotency**

INFORMA:

Que a presente tese, correspóndese co traballo realizado por D/Dna Yara Souto Becerra, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En **Santiago de Compostela**, **15 de Decembro de 2021**

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

AUTORIZACIÓN DO DIRECTOR/TITOR DA TESE

D./Dna. **Celia María Pombo Ramos**

En condición de: **Director/a**

Título da tese: **Determining the role of the zinc-finger protein ZMYM2 in pluripotency**

INFORMA:

Que a presente tese, correspóndese co traballo realizado por D/Dna Yara Souto Becerra, baixo a miña dirección/titorización, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director/titor desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En **Santiago de Compostela**, **15 de Decembro de 2021**

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DO AUTOR/A DA TESE

D./Dna. **Yara Souto Becerra**

Título da tese: **Determining the role of the zinc-finger protein ZMYM2 in pluripotency**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela**, **15 de Decembro de 2021**.

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

DECLARACIÓN DE CONFLICTO DE INTERÉS

Yara Souto Becerra declara no tener ningún conflicto de interés en relación con la tesis doctoral.

Así mismo, con la presente la doctoranda declara la completa autoría de todas las figuras que en la tesis se muestran.

En Santiago de Compostela, 22 de diciembre de 2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CONTENTS

1. RESUMO 7
2. AGRADECEMENTOS 21
3. ABSTRACT 29
4. INTRODUCTION 33
4.1 The pluripotent cell state 35
4.1.1 The early mouse embryo development 36
4.1.2 *In vitro* derivation of pluripotent stem cells 39
4.2 Alternative pluripotent states 39
4.2.1 The ground-state of pluripotency 41
4.2.2 The primed state of pluripotency 42
4.3 The molecular control of pluripotency 44
4.3.1 The transcriptional network controlling pluripotency 46
4.3.1.1. Core regulatory network 46
4.3.2 Main signaling pathways involved in pluripotency 48
4.3.3 Epigenetic regulation of pluripotency 51
4.3.3.1. DNA methylation 51
4.3.3.2. Histone acetylation 53
4.3.3.3. Histone methylation 54
4.3.3.4. Histone citrullination 56
4.3.4 Metabolic state in pluripotent cells 57
4.3.4.1. Metabolism in early embryo development and pluripotency 57
4.3.4.2. Metabolic differences between pluripotent cell states 60
4.4 From somatic cells to pluripotency 62
4.4.1 Cell fusion for reprogramming pluripotency 63
4.4.2 Induction of pluripotency with transcription factors 65
4.4.3 Changes occurring during the acquisition of pluripotency 67
4.4.3.1. Silencing of the somatic program 68
4.4.3.2. Activation of the pluripotency program 70
4.5 Zinc finger proteins 71

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.5.1 Structural classification and molecular functions 71
4.5.2 Zinc-finger proteins in pluripotency 73
4.5.3 Zinc finger MYM-type containing proteins 76
4.5.4 Zinc finger MYM-type containing 2 (ZMYM2) 78
5. AIMS 81
6. MATERIAL AND METHODS 85
6.1 Reagents and kits 87
6.2 Experimental models 87
6.3 Cell culture 89
6.4 Chemical inhibition of PADI4 activity 90
6.5 Cloning of plasmid vectors 90
6.5.1 Cloning protein-coding regions into PiggyBac transposon system 90
6.5.2 Cloning shRNA oligos into pLKO lentiviral vector 91
6.5.3 Cloning the *Padi4* genomic regulatory region into a *Luciferase* reporter .. 92
6.6 Isolation of plasmid DNA 94
6.6.1 Bacterial competent cells production 94
6.6.2 Bacterial transformation 95
6.6.3 Plasmid DNA extraction 95
6.7 Mammalian cell transfection 96
6.8 Viral production and Transduction 96
6.8.1 Generation of viral particles 96
6.8.2 Mammalian cell transduction 97
6.9 Genomic DNA isolation 98
6.10 Genotyping PCR 98
6.11 Genomic edition using CRISPR/Cas9 technology 99
6.12 Cell proliferation and viability assays 99
6.13 Cell cycle analysis 100
6.14 Primed pluripotent cell fate conversion assay 100
6.15 Alkaline-phosphatase staining 101
6.16 Colony formation assay 102
6.17 Differentiation assays 102
6.17.1 EB differentiation assay 102

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.17.2 Mesendoderm and neural ectoderm differentiation assay 103
6.18 Reprogramming assays 103
6.18.1 OKSM reprogramming assays in MEFs 103
6.18.2 Reprogramming assays in pre-iPSCs 104
6.18.3 Heterokaryon-based fusion reprogramming assay 105
6.19 Assessing the bioenergetic profile of ESCs 106
6.20 Chromatin accessibility assay 107
6.21 Luciferase assay 108
6.22 RNA extraction 109
6.23 Analysis by quantitative PCR 110
6.24 Western blot analyses 111
6.25 Affinity purification of ZMYM2 protein complexes 112
6.26 Immunofluorescence 114
6.27 Nuclear area measurement on ESCs 115
6.28 RNA sequencing and analysis 116
6.29 Chromatin immunoprecipitation followed by sequencing and analysis 117
6.30 ChIP-qPCR assay 120
6.31 Gene Ontology and KEGG pathway analyses 120
6.32 Gene set enrichment analysis (GSEA) 122
6.33 Determining the PADI family conservation tree 122
6.34 Nuclear localization signal prediction 122
6.35 Interaction network analysis 123
6.36 Statistical analysis 123
6.37 Bibliographic resources 123
6.38 Software and algorithms 124
7. RESULTS 125
7.1 Identification of ZMYM2 as a potential novel regulator of pluripotency 127
7.2 Deciphering the role of ZMYM2 in pluripotency 129
7.2.1 Characterization of *Zmym2*-null ESCs 129
7.2.2 ZMYM2 is required for proper control of ESC maintenance 131
7.2.3 ZMYM2 is required for the establishment and maintenance of primed pluripotency 136
7.2.4 ZMYM2 is required for exiting pluripotency *in vitro* 140

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7.3 Analyzing the role of ZMYM2 in reprogramming to induced pluripotency 145
7.3.1 ZMYM2 depletion promotes iPSC generation 145
7.3.2 Absence of ZMYM2 promotes the pre-iPSC to iPSC transition 146
7.3.3 Loss of ZMYM2 enhances heterokaryon-based reprogramming 149
7.4 Elucidating the molecular functions governed by ZMYM2 in pluripotency 151
7.4.1 ZMYM2 influences the transcriptional landscape in murine ESCs 151
7.4.2 ZMYM2 regulates the activation and silencing of different transcriptional networks in ESCs 155
7.4.3 ZMYM2 can act both as a transcriptional activator and a repressor in stem cell pluripotency 157
7.4.4 Identification of ZMYM2-interacting partners in ESCs 161
7.4.5 Genome-wide significance of the ZMYM2-LSD1 interaction 163
7.4.6 Loss of ZMYM2 promotes a metabolic switch in ESCs 168
7.4.7 ZMYM2 regulates histone citrullination 171
7.4.8 Lack of ZMYM2 leads to chromatin decondensation 177
7.4.9 PADI4 is regulated at the transcriptional level by ZMYM2 178
7.4.10 ZMYM2 cooperates with other factors to regulate *Padi4* expression 180
7.4.11 PADI4 inhibition partially restores transcriptional changes of *Zmym2* null ESCs 184
8. DISCUSSION 187
8.1 The relevance of ZMYM2 in self-renewal and differentiation 191
8.2 ZMYM2 functions in alternative pluripotent states 193
8.3 ZMYM2 roles during somatic cell reprogramming 194
8.4 The interplay between ZMYM2 and the cell epigenome 196
8.5 ZMYM2 and cellular metabolism 197
8.6 ZMYM2 in development and disease 199
9. CONCLUSIONS 203
10. APENDIX 207
10.1 Supplementary Table 1. Reagents 209
10.2 Supplementary Table 2. Kits and accessories 211
10.3 Supplementary Table 3. Cell lines 212
10.4 Supplementary Table 4. Recombinant DNA 212
10.5 Supplementary Table 5. Primary and secondary antibodies 213

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

10.6 Supplementary Table 6. Oligonucleotides 214
10.7 Supplementary Table 7. List of DEGs upon *Zmym2* depletion in ESCs 217
10.8 Supplementary Table 8. List of ZMYM2-Associated Proteins 220
11. ABBREVIATIONS LIST 223
12. FIGURES LIST 229
13. BIBLIOGRAPHY 235

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1. RESUMO

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

As células nai pluripotentes (PSCs, do inglés *pluripotent stem cells*), que inclúen as células nai derivadas de embrión (ESCs) e as células pluripotentes inducidas (iPSCs), poden autorrenovarse indefinidamente en cultivo e diferenciarse en tódalas células especializadas do corpo adulto. Dadas estas propiedades celulares únicas, as PSCs considéranse unha ferramenta moi valiosa para a medicina rexenerativa. Ademáis, estas céluas especiais representa un sistema poderoso que outorga novas oportunidades para interrogar os mecanismos moleculares que controlan a determinación do destino celular, o que continúa a ser unha das grandes preguntas abertas da Bioloxía moderna. O control molecular do estado pluripotente pénsase que reside en redes reguladoras conservadas e redes de interacción con proteínas de numerosos factores de transcrición (TFs) e reguladores epixenéticos (ERs), que actúan xuntos para definir a identidade celular pluripotente. Non obstante, a comprensión completa dos fundamentos moleculares que rexen a pluripotencia require a identificación e caracterización de factores adicionais que actúan en conxunto con TFs e ERs clave coñecidos nas PSCs, que son fundamentais para obter, manter e diferenciar as PSCs con fins terapéuticos. Ainda que varias proteínas contedoras de dominios de dedo de zinc (ZNFs) foron implicadas nestes procesos, a función específica da gran maioría das ZNFs agardan por explorada. Por este motivo, nesta tese centrámonos no estudo do papel da proteína con dominios dedo de zinc tipo MYM 2 (ZMYM2) nos complexos mecanismos moleculares que rexen a pluripotencia murina. Para esta fin, usei intervencións xénicas e químicas en células somáticas e pluripotentes en combinación con diferentes aproximacións celulares e bioquímicas, así coma técnicas ómicas e análises *in silico*. Aquí amoso que ZMYM2 constitúe un novo factor clave na rede reguladora central que rexe a pluripotencia. De feito, os meus datos amosan que ZMYM2 é un obstáculo para a reprogramación eficiente de células somáticas e é necesario tanto para afinar a autorrenovación das ESCs como nas estapas iniciais da diferenciación celular.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Mecanísticamente, atopei que ZMYM2 e LSD1 se asocian para o control transcripcional de xenes de pluripotencia e diferenciación, entre os que se atopan o TF NANOG e o ER PADI4. En resumo, os meus resultados identifican un novo modelo de regulación transcripcional adicado á reprogramación é á pluripotencia, e abren novas vías de investigación para explorar a implicación da asociación ZMYM2-LSD1 no desenvolvemento e nas enfermidades.

Coa finalidade de contextualizar e facilitar a comprensión do traballo desenvolto na miña tese doutoral, a continuación exporase o estado do arte no campo da pluripotencia relacionado con este traballo, para continuar cos resultados obtidos e a súa discusión, así como as aproximación metodolóxicas empregadas pra alcanzalos.

A pluripotencia pódese definir como a habilidade que posúen determinadas células para dividirse indefinidamente en cultivo e diferenciarse en tódalas células especializadas do corpo adulto. Estas propiedades emerxen pronto e de forma natural durante o desenvolvemento embrionario de mamíferos nunha masa de células localizada no interior do blastocisto, a masa interna celular (ICM, do inglés *inner cell mass*). Aquí, unha serie de toma de decisións de destino celular que xorde de forma secuencial define a reorganización do blastocisto temprano. Deste xeito, as células que conforman a ICM e expresan *Oct4*, un factor de transcripción (TF, do inglés *transcription factor*) típico de pluripotencia, veranse sometidas a un proceso de diferenciación ou automantemento segundo a expresión diferencial de dous TFs: aquelas células que expresen *Gata6* relocalizaranse na zona superficial da ICM formando o endodermo primitivo (PrE, do inglés *primitive endoderm*), mentres que as células que expresen *Nanog* darán lugar ao epiblasto (EPI) mantendo un estado pluripotente.

Dadas estas propiedades celulares únicas, as células pluripotentes considéranse unha ferramenta valiosa para a medicina rexenerativa. Por este motivo, e durante anos, realizáronse grandes esforzos na derivación de células pluripotentes a partir de células

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

localizadas na ICM ata que finalmente en 1981 dous grupos independentes foron capaces de establecer liñas celulares de ESCs (do inglés, *embryonic stem cells*) a partir de células pluripotentes da ICM de blastocistos murinos. Tras isto, numerosos estudos demostraron que ESCs derivadas de blastocistos en diferentes momentos do desenvolvemento embrionario amosan diferencias nas súas propiedades fundamentais. Así, as ESCs mantidas en condicións convencionais de cultivo celular (en presencia de suero e LIF (do inglés, *leukemia inhibitory factor*)) atópanse nun estado pluripotente coñecido coma *metastable* no que se activan cascadas de señalización contraditorias que fomentan unha elevada heteroxeneidade morfolóxica, transcripcional e funcional. Sen embargo, ESCs derivadas da ICM de blastocistos tempranos e mantidas nun medio de cultivo que fomenta a estabilización molecular grazas á adición de inhibidores químicos específicos das cascadas de ERK1/2 e GSK3 (coñecido coma medio 2i) protexe ás células fronte a sinais inductoras de diferenciación, fomentando un estado pluripotente coñecido coma *naïve*. Na contra, células derivadas do epiblasto de blastocistos tardíos e cultivados en presencia de factores FGF e Activina A (medio FA) atópanse nun estado pluripotente alternativo denominado *primed* e caracterizado por atoparse nun estado de desenvolvemento máis avanzado, así coma por unha disminución do seu potencial de diferenciación. Deste xeito, o concepto clásico de pluripotencia transfórmase, puidéndose ver coma un espectro de estados alternativos no que a capacidade de diferenciación das células se vai restrinxindo a medida que o desenvolvemento embrionario avanza.

O control molecular do estado pluripotente sábese que reside en redes reguladoras conservadas e redes de interacción con proteínas de numerosos TFs e ER (do inglés, *epigenetic regulators*) que actúan de maneira conxunta para definir a identidade celular pluripotente. Así, estudos transcripcionais levados a cabo en ESCs revelaron o papel fundamental de tres TFs no control molecular do estado celular pluripotente. OCT4 (*Octamer-binding transcription factor 4*), SOX2

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(*Sry box-containing gene 2*) e NANOG (*Nanog homeobox*) son proteínas que se saben presentes e necesarias en células pluripotentes. Estudos de pérdida e ganancia de función demostran a necesidade dunha regulación fina dos seus niveis de expresión para o mantemento dun estado pluripotente, además da existencia de bucles recíprocos de regulación entre estos factores para o correcto control transcripcional do circuito pluripotente. Por outro lado, análises dos niveis de ocupación de cromatina destos tres factores amosaron grandes niveis de solapamento entre eles, evidenciando unha intrincada rede de regulación. Además, o estudo das interacións proteicas dos factores centrais de regulación de pluripotencia foi clave para desenmascarar a función de factores adicionais en salvagardar esta identidade celular única, entre os que se atopan numerosas proteínas con dominios dedos de zinc, coma ZFP281, ZFP42 (ou REX1) e membros da familia KLF (KLF2 e KLF4) que actúan coma TFs, además de diferentes ERs encargados de realizar modificacións químicas na cromatina. Ditas modificación poden ocorrer tanto nas secuencias de ADN coma nas proteínas histónicas que conforman os nucleosomas, e en conxunto constitúen unha nova capa de regulación transcripcional denominada epixenética. Estudos realizados nas últimas décadas poñen de manifesto o papel fundamental das modificacións epixenéticas no control da pluripotencia. Así, por exemplo, sábese que a metilación de ADN conforma unha das modificacións epixenéticas más importantes no control do estado transcripcional das células, resultando na represión da expresión de xenes. Dita modificación é levada a cabo por proteínas pertencentes á familia DNMT (*DNA methyltransferases*), mentres que a eliminación dos grupos metilo en nucleótidos de citosina é levada a cabo por proteínas da familia TET (*Ten-eleven translocation enzymes*), todas elas necesarias para o correcto mantemento dos estados *naïve* e *primed* de pluripotencia. Outras modificacións epixenéticas cruciais para a pluripotencia son a acetilación e metilación de histonas, conformando marcas de activación e represión transcripcional respectivamente. Células pluripotentes amosan elevados niveis de acetilación que neutralizan as cargas positivas presentes nas colas

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

histónicas, cambiando as súas propiedades físicas e resultando nunha menor compactación da cromatina permitindo unha maior accesibilidade para a maquinaria de transcripción. Pola contra, a metilación de histonas parece ser relevante durante os procesos de diferenciación celular, de xeito que un incremento nos niveis de metilación resultan críticos para o correcto silenciamento de xenes importantes para a pluripotencia durante este proceso. LSD1 (*Lysine Demethylase 1*), unha proteína importante na eliminación dos grupos metilos de histonas, resulta ser importante durante o desenvolvemento embrionario temprano, actuando de maneira conxunta con diferentes factores de transcripción coma OCT4, SOX2 e NANOG para a regulación da expresión de xenes involucrados en procesos de diferenciación e mantemento dun estado celular pluripotente. Recentemente demostrouse que unha nova marca epixenética, a citrulinación de histonas, resulta importante para a pluripotencia *naïve*. Esta modificación, consistente na conversión dun residuo de arxinina nun residuo de citrulina nun proceso denominado deiminación, é levado a cabo por proteínas da familia PADI (Peptydil Arginine Deiminases) e parece exercer efectos opostos sobre o control transcripcional. Deste xeito, a citrulinación histónica mediada por PADI4 parece ter un papel similar ao da acetilación de histonas, permitindo unha menor compactación da cromatina e unha maior accesibilidade para o control da expresión xénica no contexto do desenvolvemento temprano e da pluripotencia *naïve*.

Alteracións no metabolismo celular sábense importantes nos estados epixenéticos e transcripcionais das células pluripotentes, resultando na adquisición de estados alternativos que son ou non favorables para procesos de diferenciación. Así, as ESCs caracterízanse por unha elevada actividade glucolítica derivada dun incremento nos niveis de expresión de transportadores de glucosa. Sen embargo, os baixos niveis de consumo de osíxeno nestas células resultan nunha elevada producción de lactato derivada dunha elevada actividade glucolítica en condicións anaeróbicas, o chamado efecto Warburg. Este

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

fenómeno permite unha rápida obtención de enerxía necesaria para a rápida proliferación que se observa nestas células. A medida que os procesos de desenvolvemento avanzan e o potencial de diferenciación se restrinxe nas células pluripotentes, a elevada actividade glucolítica das células tórnase progresivamente hacia unha actividade oxidativa dependente de mitocondria, resultando no aumento do consumo de osíxeno característico das células pluripotentes *primed*.

Para comprender completamente os fundamentos moleculares que rexen a pluripotencia requérese a identificación e caracterización de factores adicionais que actúan en conxunto cos TF e ER clave coñecidos nas PSCs (do inglés, *pluripotent stem cells*), o que é fundamental para obter, manter e diferenciar as PSCs con fins terapéuticos. As proteínas con dominios de dedos de zinc (ZNFs) constitúen a maior familia de proteínas existentes con unha amplia variedade de funcións biolóxicas e implicadas en numerosos procesos moleculares. Estas proteínas caracterízanse polo seu contido en dominios de dedos de zinc, estructuras nas que dous residuos de histidina se unen a dous residuos de cisteína por medio dun ión de zinc que lles conferen a habilidade de recoñecer secuencias específicas de ADN e otorgándolles a súa función coma TFs. Numerosas ZNFs sábense implicadas na regulación de procesos moleculares importantes no contexto da pluripotencia. Así, por exemplo, KLF4 revelouse coma un factor esencial na adquisición do estado pluripotente, un proceso denominado reprogramación celular, actuando de forma sinérxica con OCT4, SOX2 e c-MYC. LIN28 e SALL4, outras ZNFs coñecidas, tamén teñen un papel importante en procesos de reprogramación celular, sendo capaces de substituir a OCT4 e KLF4 no cocktail orixinal para a obtención de iPSCs (do inglés, *induced pluripotent stem cells*), mentres que membros da familia GATA tamén son capaces de incrementar a eficiencia de reprogramación e de incluso sustituir a OCT4 no proceso. Recentemente ZFP281 revelouse coma un interactor común dos TFs centrais en pluripotencia OCT4, SOX2 e NANOG, sendo indispensable para a regulación dos estados alternativos de

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pluripotencia. Pese a que o coñecemento sobre os papeis específicos das ZNFs no contexto de pluripotencia se vai acumulando, as funcións moleculares da gran maioría destas proteínas todavía se descoñece.

Tras a análise en detalle dos interactomas dos TFs centrais de pluripotencia (OCT4, SOX2 e NANOG) previamente publicados, a proteína contedora de dedos de zinc tipo MYM 2 (ZMYM2) revelouse coma un interactor común de OCT4 e NANOG, evidenciando o seu posible papel na regulación da pluripotencia. Por isto, nesta tese propuxémonos como obxetivo principal o de descubrir o papel de ZMYM2 nos complexos mecanismos moleculares que rexen a pluripotencia murina.

Para abordar isto, empreguei varias cobinacións de intervencións xénicas e químicas en células pluripotentes (ex., ESCs e iPSCs) e somáticas (ex., fibroblastos) de orixe murino elixidas como principais modelos experimentáis deste estudo. A longo do traballo urilizáronse unha amplia gama de ensaios celulares e bioquímicos de uso xeral no campo da pluripotencia,. Ademáis empregáronse varias técnicas ómicas incluindo aquelas necesarias para o estudo de cambios transcripcionais globais (ex., RNA-seq), a identificación de localización na cromatina (ex., ChIP-seq) e desenmarcarar redes de interacción proteicas (ex., LC/MS-MS). Os datos obtenidos integráronse e analizáronse *in silico* empregrando unha amplia variedades de recursos bioinformáticos. Deste xeito descubrín que ZMYM2 é un novo compoñente clave da rede reguladora central que rexe a pluripotencia, interaccionando con TFs claves para o control do estado pluripotente tales coma OCT4 e NANOG (**Figura R1**). Ensaios de pérdida de función amosaron que pese a ser un factor dispensable para o mantemento da pluripotencia convencional (en condicións de presencia de suero e LIF), a súa ausencia resulta nunha derregulación de marcadores de diferenciación (*Gata4*, *Gata6*) e pluripotencia (*Nanog*, *Zfp42*), así coma nun incremento da autorrenovación celular (**Figuras R2-R4**). En concreto, e de xeito similar á función doutra ZNF coñecida (ZFP281), a función de ZMYM2 é necesaria para o correcto

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

establecemento e mantemento dun estado pluripotente *primed* (**Figuras R5-R6**). Así, a ausencia de ZMYM2 impide a activación transcripcional de marcadores específicos de pluripotencia primed (*Fgf5*) durante a transición de ESCs a EpiLSCs, resultando nun descenso da viabilidade celular. Ademáis, empregando un sistema de rescate mediante expresión exóxena dunha isoforma de ZMYM2 en células *Zmym2 knockout* puiden confirmar que o debandito efecto sobre a pluripotencia primed é específica desta proteína. En relación con isto, os meus datos amosan que ZMYM2 é necesaria para a correcta diferenciación das células pluripotentes (**Figuras R7-R8**). Así, empregando dous modelos de diferenciación celular distintos puiden demostrar que a ausencia de ZMYM2 resulta en alteracións transcripcionais que impiden a correcta saída de pluripotencia cando se empregan condicións de cultivo que fomentan a diferenciación e especificación celular.

Por outra parte, os meus datos tamén revelaron que ZMYM2 é un obstáculo para a reprogramación eficiente de células somáticas (**Figuras R9-R11**). Así, o silenciamento de *Zmym2* en fibroblastos murinos embrionarios (MEFs) resultou nun incremento da eficiencia de reprogramación destas células, así coma dun acurtamento de proceso de obtención de iPSCs. Tendo en conta os meus datos que indican que na ausencia de ZMYM2 en ESCs NANOG se atopa regulado a alza, e que este facttor de transcripción sábese que ten un papel fundamental nas últimas fases da reprogramación celular somática, decidín empregar intermediarios de reprogramación (pre-iPSCs) coma modelo para estudar o efecto de ZMYM2 na adquisición dun estado pluripotente. De xeito destacable, atopei que ZMYM2 tamén actúa como barreira de reprogramación celular incluso nas fases tardías do proceso. Ademáis, o silenciamento de *Zmym2* en pre-iPSCs resultou nun incremento da expresión de *Nanog*, polo que non se pode descartar a interacción funcional destes factores no proceso de adquisición dun estado celular pluripotente. Por último, ensaios de reprogramación basados en heterocariontes derivados de fusión celular tamén demostraron un

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

incremento da eficiencia de reprogramación de células humanas adultas en ausencia de ZMYM2.

Coa fin de determinar o efecto específico de ZMYM2 no control da identidade celular pluripotente levei a cabo un perfil transcripcional de ESCs en presencia e ausencia de ZMYM2 mediante o emprego de sequenciación profunda de ARN (RNA-seq). Os meus datos amosaron que ZMYM2 exerce un importante papel na regulación transcripcional de xenes involucrados en metabolismo celular, desenvolvemento e organización da cromatina (**Figuras R12-R14**). Dun total de 790 xenes diferencialmente expresados tras a deleción de ZMYM2, 270 atópanse regulados á baixa (FDR<0.05, FC>2). Entre estes puideron atoparse marcadores de especificación celular e reguladores epixenéticos con importantes funcións no mantemento e adquisición de pluripotencia. Pola contra, a depleción de ZMYM2 resultou na regulación á alza (FDR<0.05, FC>2) de 520 xenes importantes en procesos metabólicos e con funcións no mantemento dunha pluripotencia *naïve*. Ademáis, a análise dos niveis de ocupación de ZMYM2 na cromatina mediante ChIP-seq amosaron que este factor se une dunha maneira específica de secuencia ao ADN, resultando na regulación directa dun 60% dos xenes diferencialmente expresados (**Figuras R15-R16**). Estes análises tamén revelaron que ZMYM2 pode actuar coma un activador e represor transcripcional de xenes implicados en procesos metabólicos e de diferenciación celular. Mecanísticamente, mediante análisis proteomicos por LC-MS/MS, atopei que ZMYM2 se asocia con numerosas proteínas en ESCs formando redes de interación proteicas, nas que se atopan tanto factores de transcripción coma reguladores epixenéticos (**Figura R17**). De entre estos, LSD1 aparece coma o candidato coa interacción máis estable. A análise dos niveis de ocupación de LSD1 na cromatina mediante datos dispoñibles de ChIP-seq en ESCs con presencia e ausencia de ZMYM2 revelaron que a distribución dos picos de unión deste factor na cromatina está sometida a unha profunda remodelación en ausencia de ZMYM2 (**Figuras R18-R19**). Ademáis, ambos factores amosan un elevado grao de

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

solapamento nos seus niveis de ocupación da cromatina, suxerindo un papel sinérxico no control transcripcional de xenes diana en común. De entre un total de 468 xenes diana diferencialmente expresados en ausencia de ZMYM2, un 45% dos mesmos atópanse unidos por LSD1. Dos mesmos, tan só os xenes que se atopan regulados á alza en ausencia de ZMYM2 tamén se atopan regulados á alza en ausencia de LSD1, o que suxire unha función conxunta na represión transcripcional dun grupo de xenes involucrados en pluripotencia e diferenicación. Ademáis, entre os xenes diana controlados polo eixe ZMYM2-LSD1 atópanse xenes implicados no control de diferentes cascadas de sinalización importantes na regulación do estado metabólico celular. Por este motivo, e aproveitándome da tecnoloxía SeaHorse, puiden observar que a ausencia de ZMYM2 en células pluripotentes resulta nun aumento da actividade glicolítica e un descenso no seu consumo de osíxeno, o que concorda con unha incompatibilidade do mantemento do estado *primed* de pluripotencia observado (**Figura R20**). Estes datos foron confirmados seguendo a mesma aproximación a apartir doutra línea knockout de Zmym2 en ESCs xerada pola tecnoloxía CRISPR/Cas9 (**Figura R21**).

Coa fin de descifrar os mecanismos moleculares nos que ZMYM2 se encontra involucrado para a regulación do estado pluripotente, decidín realizar unha análise transcripcional de reguladores epixenéticos coñecidos (**Figura R22**). Así, os resultados amosaron que membros da familia de remodeladores epixenéticos *Padi* se atopan regulados á alza en ausencia de ZMYM2. Análises dos niveis transcripcionais destos factores en diferentes liñas celulares puxeron de manifestó unha regulación global á alza dos mesmos en células pluripotentes, suxerindo un papel importante no contexto da pluriptoencia murina revelada recentemente nun artigo publicado (**Figura R23**). En concreto, de entre os membros desta familia de proteínas, PADI4 é único que presenta unha sinal de localización nuclear (NLS), permitindo levar a cabo a citrulinación de residuos de arxinina en proteínas histónicas no interior do núcleo dun xeito

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

dinámico (**Figura R24**). Mediante o emprego de técnicas de inmufluorescencia de célula única, Western blot e inmunoprecipitación de cromatina seguida de secuenciación masiva (ChIP-seq) puiden demostrar que un aumento nos niveis de expresión de *Padi4* tras a depleción de ZMYM2 corresponden con un aumento nos niveis de H3Cit no interior nuclear (**Figura R24**). De acordo cun reciente artigo no que se amosa que os niveis elevados de H3Cit observados en células pluripotentes *naïve* corresponden cunha decompactación da cromatina e unha maior accesibilidade, ensaios de inmunofluorescencia demostraron un maior radio nuclear en células *knockout* para *Zmym2* (**Figura R25**). Ademáis, a dixestión de fragmentos de cromatina empregando a nucleasa micrococal revelaron unha maior accesibilidade cromatínica en ausencia de ZMYM2. O papel de ZMYM2 como represor directo da expresión de *Padi4* foi finalmente demostrado mediante o emprego dun ensaio *Luciferasa* baixo o control da rexión promotora deste xen o que se une ZMYM2 (**Figura R26**).

Coa finalidade de revelar o mecanismo molecular polo cal ZMYM2 exerce o seu papel de represor transcripcional do xen *Padi4*, primeiro exameinei se a modificación por SUMO de ZMYM2 podía estar mediando este efecto transcripcional e non atopei evidencias que así o suxeriran (**Figura R27**). Logo realizei unha análise conxunta de datos procedentes de técnicas de secuenciación (**Figura R28**). Deste xeito, mediante ChIP-seq demostrei que ZMYM2 se encarga do reclutamento do factor de represión LSD1 na rexión promotora de *Padi4*, resultando no silenciamento transcripcional deste xen. Ademáis, ensaios de pérdida de función de interactores directos de ZMYM2, nos que se inclúen ademáis de LSD1 a ZFP281 e as HDAC1/2, tamén demostraron un aumento significativo nos niveis de expresión de *Padi4* cando se fixeron ensaios de pérdida de función en ESCs WT mediante o uso de shRNAs específicos (**Figura R29**).

Por último, para entender cómo PADI4 contribúe nos cambios transcripcionais producidos pola ausencia de ZMYM2, realicei unha comparación do meu RNA-seq con un *microarray* dispoñible en células

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pluripotentes tratadas con un inhibidor específico de PADI4, demostrando un solapamento parcial nos cambios transcripcionais en ambas condicións (**Figura R30**). Para estudar con máis detalle os efectos específicos da regulación á alza de *Padi4* en ausencia de ZMYM2 realicei un ensaio de rescate mediante o emprego de inhibidores selectivos e reversibles de PADI4 seguido dun perfil transcripcional empregando RNA-seq (**Figura R30**). Datos derivados deste ensaio permitiron identificar un rescate dun 25% e 32% dos xenes reprimidos e activados tras a ausencia de ZMYM2, respectivamente. Además, estudos bioinformáticos permitiron revelar que parte destes xenes cuxa alteración transcripcional tras a depleción de ZMYM2 é consecuencia indirecta dun aumento nos niveis de expresión de *Padi4* e un incremento nos niveis de citrulinación de histonas se atopan implicados en procesos de diferenciación e control do estado metabólico celular.

En conclusión, os meus resultados identifican a ZMYM2 coma un factor esencial no correcto control do mantemento e saída de pluripotencia. Además, dito factor tamén actúa coma unha barreira de reprogramación debido á súa función en salvagardar unha identidade celular pluipotente. Mecanísticamente, ZMYM2 coopera con LSD1 na represión transcripcional de xenes importantes para a bioloxía pluripotente, tales coma *Nanog* e *Padi4*, controlando o grao de compactación da cromatina mediante a regulación da citrulinación de histonas (**Figura D1**). No seu conxunto, os datos amosados na miña tese doutoral identifican un novo modo de regulación transcripcional dedicado á reprogramación e á pluripotencia, e abren novas vías para explorar a implicación da asociación ZMYM2-LSD1 nos eidos do desenvolvemento e enfermidades.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2. AGRADECEMENTOS

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

O camiño da tese foi moi longo, e a verdade é que moitas veces dudei de que este momento chegase. Pero aquí estamos finalmente, escribindo o que levo tantos anos agardando: a sección de agradecementos da miña tese doutoral. Quen me coñeza un mínimo (se estás lendo esto, probablemente o fagas) saberá que son moi intensa, así que ponte cómodo e prepárate para ler á Yara máis visceral (de feito xa estou chorando con este parágrafo inicial, "*se vienen cositas*").

En primeiro lugar debo agradecer a Chito e Celia por darme a oportunidade de realizar unha tese doutoral e abrirme as portas do CiMUS. Dende o día da entrevista inicial, e no momento no que dixen que me quedaba, a miña vida cambiaría dunha forma da que todavía non era consciente. Todavía recordo a Chito ensinándome cousas de clonación de plásmidos cunha paixón desmedida mentres eu trataba de tan sequera entender o significado das palabras que pronunciaba e pensaba "oxalá algún día poder ter a capacidade de procesamento de información que ten este home". E a Celia, a miña co-directora de tese. Moitas grazas por facer que este día sexa posible.

En segundo lugar, agradecer a Miguel e Diana. Grazas a vós os dous esta tese doutoral foi posible (tanto polo económico de construir unha Unidade coma por facer as mellores fichaxes para o equipo). Miguel, non sabes o eternamente agradecida que estou de poder dicir que es o meu mentor. Certo é que houbo discusións e momentos difíciles, pero sempre soubeches qué dicir para impulsarme e que non abandonase. Es unha das mentes máis brillantes que coñecín e non podo sentir máis que admiración pola túa dedicación incansable. De verdade, moitísimas grazas por todo. E Diana, a capitana do laboratorio irmán, científica coa que puiden traballar man a man nos inicios compostelanos de ADAR, e xefa á que recurrir en caso de dúbida. Grazas polas túas valiosas aportacións. E por certo, noraboa aos dous por traer ao mundo a un científico en miniatura, Gael.

Parda, compañeiro nos inicios do laboratorio e o meu primeiro amigo en Santiago. Innumerables as horas de longas conversas sobre

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

gran variedade de teimas, dende física cuántica ata o significado da existencia humana, pasando por altos niveis de testosterona e moito alcol. Incontables tamén as noites de festa, os kilos de churrasco de merenda, as ensaladas sen verde e, sobre todo, as risas. Es a persoa á que sempre acudir cando necesito despexarme, deixar de pensar e retornar á vida normal. Mentiría se disexe que sempre foron bos momentos, pero unha amizade basease niso, altos e baixos, e os nosos altos sempre nos prepararon para permanecer xuntos nos baixos máis baixos. Quérote moito.

Vera, Adriana e Tomás, os "*fraichers*". Cada un de vós por separado é incrible, pero xuntos sodes (somos) insuperables. Vera, a ti adopteite coma á miña versión en pequeno coa intención de que non cometeras os mesmos erros que cometín eu (un pouco maternal, a dicir verdade), e terminaches adiantándome pola esquerda. As horas da comida constituíron un espazo seguro no que debatir sobre temas polémicos, e grazas a eles conseguín elaborar un discurso coherente i estructurar a miña propia ideoloxía. Nestes anos vinte crecer moitísimo, tanto a nivel persoal como académico/científico, e pese a que ningunha de nós as dúas é demasiado afectiva, creo que deberías saber que me sinto moi orgullosa de ti e de poder dicir que es a miña amiga. Tes moito esforzo e traballo ás costas, e non me cabe dúbida de que che espera un futuro brillante i exitoso. Adriana, *dancing queen* e futura *medical doctor*. A nosa amizade foise forxando de a poucos. Podería dicir que todavía recordo o momento exacto no que pensei "ollo con Adriana que ten moito potencial": tomando un batido postlaboratorio no Cangaceiro. Dende ese intre comecei a comprender mellor cómo ves o mundo, e creo que todos estaríamos de acordo en dicir que es un ser de luz. Por favor, non deixes que nada nin ninguén te corrompa (aínda que para iso *fraichers* quizais non sexa unha moi boa influencia). Tomás. Pensar que a nosa amizade comezou dun odio compartido. Fuches o meu compañeiro na época de horarios descompensados saíndo do laboratorio de madrugada. Tivemos discusións científicas terribles (momento plásmidos "*eso que dices es una gilipollez*"), pero tamén

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

fuches (i es) un gran apoio fóra. Persoa lóxica e racional donde as haiba. Estoico e samurai. Modelo a seguir, sempre un paso por diante. Sodes un pilar fundamental i espero que *fraichers* non morra nunca.

Tiago, portugués de nacemento, galego de corazón. As nosas vidas sempre foron un drama constante, o que non axudou moito na nosa misión de deixar de fumar. "*Este va a ser mi año*" é a frase prohibida, porque sempre o dicimos e nunca o é. Pero xa sabes, unha vez tocas fondo só queda a subida, e sabes que sempre poderás contar comigo para o que necesites.

Cristina e Álex, premio parella do lustro doutoral. Bos foron os inicios con esas sesións de cine nocturnas seguidas de ceas no Cambalache ou no Udon. Falo en nome do laboratorio cando digo que estamos agardando o momento voda. Prometemos facer un bó regalo.

Cristina e Alba, fichaxes do laboratorio irmán. Trouxestes o aire reivindicativo-sindical e de debate que facía falla no grupo. Visceralidade e paixón, pero tamén sobriedade, poden definiros a ambas as dúas.

Helena, Lucía, David e Fernando. Chegastes xusto no meu momento de transición, pero non me fixeron falla máis de un par de cañas para saber que sodes un bó legado para a Unidade. Helena, noraboa por esa Margarita Salas e a ver se tés éxito na túa misión de ser propietaria. Lucía, sigo esperando polas miñas catro patacas de Coristanco (#RedFlag). David, éiche deixar que labregues no meu ducado. Fernando, vémonos en Bloom.

Nesta sección tampouco me podo esquecer dos meus anfitrións en NY. Jianlong e Hongwei. Menudo equipo tedes montado aló. Nunca vin xente traballar tan duro coma a que hai nese laboratorio, con sesións continuadas de 24 horas e microsiestas diante do ordenador. Dan, a persoa que me dou cama durante os primeiros días de estancia, compañeira de bench e coa que podía falar de salseos. Xianle, persoa noble e pura igual de intensa ca min (pero por dentro). Tarik, o meu

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

bro. Case catro meses foron suficientes para saber que teño un amigo agardando por mín nas Américas. A persoa *emotionally unavailable* que se volcou en darme un fogar na miña estadía, que me integrou no seu día a día e ó que lle ofrecín a peor influencia que puiden ("*Let's get drunk on a Monday*!"). De seguro os nosos camiños volverán cruzarse. Vémonos pronto.

E como non, ás novas incorporacións grazas a Gliatime. María, Paula e Nathalia. María, levamos só oito meses xuntas pero xa temos falado moito. Non sei canto tempo me queda polo CiMUS, pero agora que me liberei da carga da tese éime de encargar de ofrecerche as mellores festas de fin de semana para que esquezas as penurias do día a día. E cando marche, sempre terás as portas da miña casa abertas. Paula, grazas por traer tan bó ambiente ao laboratorio. As túas anécdotas son o mellor da hora do café. Nathalia, qué dicir. Resonancia e sintonía. Con taras, pero unidas.

Pero non todo é ciencia. De feito, agradécese que non o sexa. An-Droide Toretto, aquí vas ti. Qué dicir, madre mía. Amigas dende os dezaséis, xa pasamos por moitas xuntas, pero a etapa da tese… Fíxonos crecer a ambas as dúas, creo eu. Todo comezou co noso costume dos martes de Impro (sen comentarios, por favor), continuou cunha viaxe a Óns na que che din plantón e rematou con un "*qué te parece si compartimos piso*" e un "*ya solo tenemos que ir a firmar*" dous días máis tarde. Este último ano foi todo un carrusel. Costou adaptarse a vivir xuntas, máis que nada por min, a miña nula xestión emocional, os meus encerros na habitación e a miña cara de cheirar merda todo o rato. E ti sempre estivestes aí, respetando a miña cuestionable forma de vida e axudándome cando me deixaba. Déboche moitas cousas, moitísimas, pero a máis importante é a túa amizade incondicional. Todo o que crecín neste último ano é grazas a ti. Eu transferinche por debaixo da porta a miña intensidade, pero ti escoitastes os meus dramas, secaches as miñas bágoas, regaláchesme flores nos peores días, fixéchesme sopa cando estiven enferma, empurráchesme a ir a cross e poñerme durísima coma

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

unha pedra… Esta tese é tanto miña coma túa, e sairás no book de fotos con ela como madriña. Quérote.

Por último, ás persoas máis importantes da miña vida. Ós meus pais. Non fun a mellor filla que uns pais poden desexar, pero vós sí sodes os mellores pais que unha filla pode ter. Houbo épocas moi complicadas, pero sempre estivestedes e sempre estaredes ó meu lado, apoiándome incondicionalmente, i esta tese non sería posible de non ser polos vosos esforzos. Mamá, papa e Grey, sodes inmortais. Quérovos.

E bueno, creo que tamén hai unha persoa que o merece. Eu mesma. Porque só eu sei o moito que costou chegar ata aquí. Moitas foron as veces nas que pensei en tirar a toalla por non sentirme suficiente, pero son demasiado cabezota coma para deixar algo a medias. Por non abandonar, grazas Yara.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3. ABSTRACT

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Pluripotent stem cells (PSCs), including embryo-derived ESCs and somatic cell-derived iPSCs, can self-renew indefinitely in culture and differentiate in all specialized cells of the adult body. Given these unique cellular properties, PSCs are considered a valuable tool for regenerative medicine and have proven to be a powerful system creating new opportunities to interrogate the molecular mechanisms controlling cell fate determination. Molecular control of the pluripotent state is thought to reside in conserved regulatory networks and protein interaction networks of numerous transcription factors (TFs) and epigenetic regulators (ERs), which act together to define pluripotent cell identity. However, the full understanding of the molecular underpinnings governing pluripotency requires the identification and characterization of additional factors that act in conjunction with known key TFs and ERs in PSCs, which are critical for obtaining, maintaining, and differentiating PSCs for therapeutic purposes. Although a handful of zinc-finger proteins (ZNFs) have been implicated in these processes, the vast majority of ZNFs have not been analyzed yet. In this thesis, I focused on discovering the role of the zinc zinger MYM-type containing 2 (ZMYM2) in the complex molecular mechanisms that govern mouse pluripotency. To address this, I used genetic and chemical interventions in somatic and pluripotent cells in combination with several cellular and biochemical assays, omic techniques, and *in silico* analyses. Here, I found that ZMYM2 is a novel key component of the core regulatory network that governs pluripotency. Indeed, my data revealed that ZMYM2 is a roadblock to efficient somatic cell reprogramming and is required to fine-tune the self-renewal of ESCs and the early differentiation process. Mechanistically, I found that ZMYM2 and LSD1 associate for the transcriptional control of pluripotency and differentiation genes, amongst others of the TF NANOG and the ER PADI4, which are essential to regulate pluripotency or drive reprogramming. In sum, my results identify a novel transcriptional regulatory mode dedicated to reprogramming and pluripotency and open new avenues for exploring the involvement of ZMYM2-LSD1 partnership in development and disease.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4. INTRODUCTION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4.1 THE PLURIPOTENT CELL STATE

Pluripotency is defined as the ability of a cell to develop into the three primary germ layers (i.e., ectoderm, mesoderm, and endoderm) of the early embryo and therefore into all cells of an adult organism[1-3] (**Figure I1**). This unique cellular state appears transiently during the early development of mammals (e.g., ~E3.5 in the mouse), in the epiblast cells within the inner cell mass (ICM) of the blastocyst[2]. Importantly, epiblast cells can be derived and maintained *in vitro* under specific culture conditions and are referred to as Embryonic Stem Cells (ESCs)[4,5]. Under appropriate culture conditions, ESCs can proliferate indefinitely without differentiation, a property that is known as self-renewal, and at the same time retain the developing potential to generate cells of all three primary germ layers[4-7] (**Figure I1**). The unique nature of murine pluripotent stem cells (PSCs) both *in vivo* and *in vitro* will be discussed in the following sections.

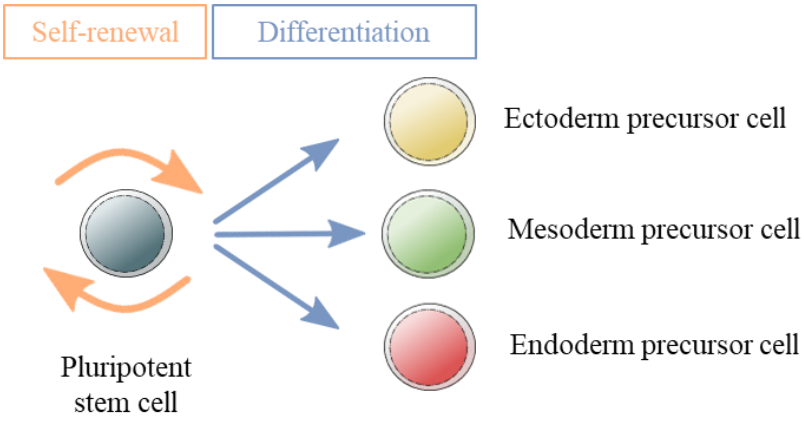


Figure I1. Main characteristics of pluripotent stem cells. Simplified representation showing that each PSC can give rise to an identical daughter cell (self-renewal) and can also differentiate into precursor cells of the three primary germ layers (differentiation).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.1.1 The early mouse embryo development

Unlike humans, rodent embryo development is characterized by rapid cell proliferation that leads to a short developmental process and gestation[8]. However, the structure of the blastocyst, which contains the pluripotent epiblast cells, is generally similar in all mammalian species[9]. Moreover, it has been proposed that the general lineage relationships during early human development are similar to those in the mouse[9,10].

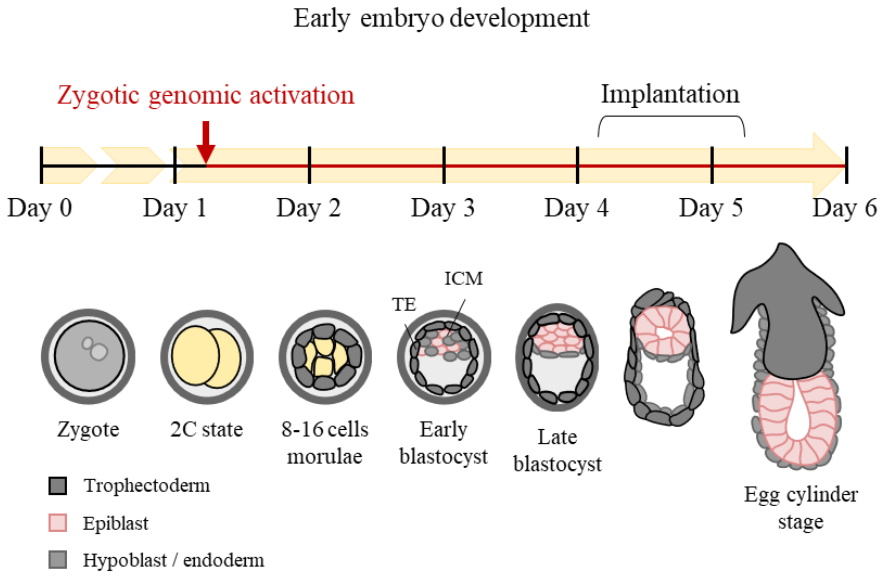


Figure I2. Early stages of embryo development. Simplified depiction showing the early stages of embryo development in mouse. Cells conserving a pluripotent state when cultured *in vitro* are shown in pink.

Shortly after fertilization, pre-inactivated genomes from both the egg and the sperm[11-13] need to become active for the stimulation of the zygote transcriptional program for developmental progress, a process termed as zygotic genomic activation (ZGA) [3,14-16]. At the one-cell stage, a first minor wave of transcriptional activation occurs[3,15] and the zygote undergoes its first cell division giving place to the two-cell embryo (2C)[2,3]. These two cells (i.e., the blastomeres) can give rise to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

any cell type that makes up an adult organism and all extra-embryonic tissues required to sustain its development, a property known as totipotency. At this stage, a second and major wave of ZGA takes place, and severe degradation of maternally inherited mRNAs is needed for the proper activation of the zygotic transcriptional program[15]. Then, the embryo undergoes a series of rapid cell divisions by which it increases in cell number without increasing its overall size (**Figure I2**). In addition, all cells conforming to this early embryo seem to possess similar developmental properties[17].

However, around embryonic day 2.5 (E2.5), the blastocyst emerges, and the first wave of cell segregation occurs in the pre-implantation stage of murine development. At this point, the expression levels of two transcription factors determine decisions about cell fate and lineage specification of two different cell types (**Figures I2** and **I3**). These cells conform to a hollow structure known as a blastocyst, made up of an outer cell layer of trophectoderm (TE) cells characterized by high expression levels of *Cdx2* and an inner cell mass (ICM) of *Oct4*-expressing cells[18,19]. Cells from the TE will give rise to all extra-embryonic tissues that surround and nourish the embryo during the developmental process, whereas specific cells from the ICM will generate the three germ layers that finally will give rise to a whole organism[19,20].

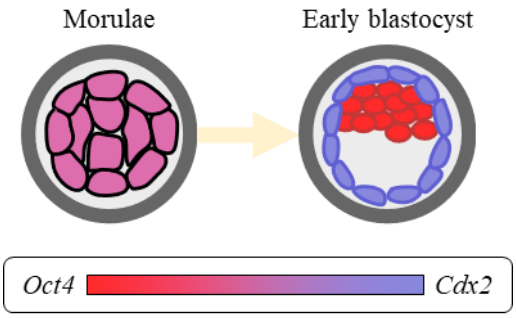


Figure I3. Mutual exclusion of *Oct4* and *Cdx2* specifies cell identity during early mouse development. In the morula, both *Oct4* (red) and *Cdx2* (blue) are homogeneously expressed (pink). As development proceed and the early blastocyst is generated, two different cell populations emerge. In one hand, *Oct4*-expressing

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cells will define the ICM, whereas *Cdx2*-expressing cells will generate the trophectoderm.

Around E3.5, the ICM undergoes a structural reorganization, and a second cell fate determination occurs (**Figure I2**). Cells from the ICM express two different transcription factors (e.g., NANOG and GATA6) in a mutually exclusive manner (**Figure I4**). Thus, the cells that express *Gata6* relocate to the surface of the ICM, conforming the primitive endoderm (PrE) or hypoblast and giving rise to the extraembryonic endoderm. In contrast, *Nanog*-expressing cells remain pluripotent, becoming epiblast (EPI) cells that will differentiate into cells from the three germ layers[21,22].

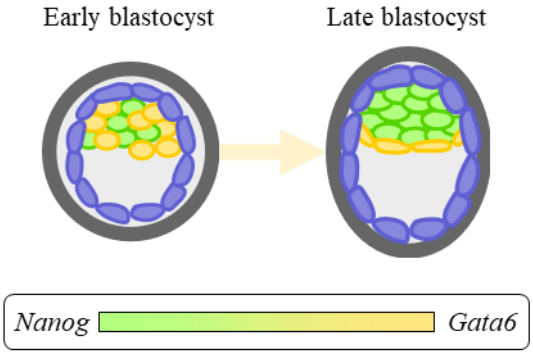


Figure I4. Mutual exclusion of *Nanog* and *Gata6* specifies cell identity during early mouse development. In the established early blastocyst, both *Nanog* and *Gata6* are mutually exclusive in cells conforming the ICM, so they are expressed in a "salt-and-pepper" manner. Thus, *Nanog*-expressing cells will generate the epiblast in the late blastocyst, whereas cells expressing *Gata6* will give rise to the primitive endoderm (PrE) or hypoblast.

When this reorganization is completed, at around E4.0, implantation of the blastocyst in the female uterus takes place. At this moment, cells from the ICM conforming the epiblast start to organize as an epithelial sheet that proliferates until it generates the egg cylinder, which in turn will give rise to the three germ layers of the embryo[2,23]. In contrast, the PrE differentiates into the visceral endoderm, generating the visceral yolk sac after gastrulation. Taken together, pluripotency can

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

be seen as a spectrum in which small cell fate decisions lead to cell determination and potency restriction over time.

4.1.2 *In vitro* derivation of pluripotent stem cells

Originally, *in vitro* pluripotency was captured from the disaggregation of both teratomas[24] and teratocarcinomas[25-28]. These cells, known as embryonal carcinoma cells (ECs), can differentiate into non-tumorous cells from the three germ layers when cultured in suspension[25,27] or injected into mice[24,26,27,29]. For this reason, ECs were considered pluripotent cells. However, since pluripotent cells obtained from ECs have undergone an oncogenic process, they do not exactly recapitulate all features present in the *in vivo* pluripotency, so efforts were made for the direct obtention of pluripotent cells from embryos. Hence, in 1981, two independent groups were able to derive pluripotent cells from the ICM located in the inner region of murine blastocysts [4,5]. Thus, murine ESCs were derived from the ICM using EC-conditioned medium[5]. Moreover, ESCs from the ICM of delayed peri-implantation blastocysts were also directly derived by disaggregating the cells and culturing them into DMEM supplemented with 20% serum[4]. In both cases, cells showed a normal karyotype, and they also presented the self-renewal and both *in vitro* and *in vivo* differentiation capabilities typical of PSCs[4,5]. However, some differences between cell lines might emerge, since the precise developmental timepoint from which cells were derived is different. Furthermore, artificial and forced delay of peri-implantation blastocysts[4] may result in epigenetic and metabolic cellular differences altering the natural pluripotential state of derived cells. Hence, differences in the exact pluripotent state of cells emerging due to differential derivation time windows and the use of specific media for cell culture will be discussed in the section below.

4.2 ALTERNATIVE PLURIPOTENT STATES

The pluripotent state is the building block of embryo development. As mentioned previously, cells that acquire pluripotency have the

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

functional capacity to differentiate into cells of all somatic lineages conforming to an adult organism. *In vivo*, pluripotent cells first emerge in the inner region of early blastocysts, in which *Oct4*-expressing cells conform the ICM (for more detail, see section **4.1.1**, **Figure I3**). As development progresses, the ICM consolidates, and cells expressing *Nanog* conform the epiblast[30]. At this point, epiblast cells enter the "ground-state" of pluripotency. As post-implantation development proceeds, inductive stimuli promote the progression of these cells into "primed" pluripotency. In this state, cells are known to be poised for the rapid activation of the transcriptional programs leading to lineage specification[31].

In vivo and *in vitro* correspondence of the different pluripotency states is illustrated in **Figure I5**, and culture conditions for their *in vitro* self-renewal will be specified in the following sections.

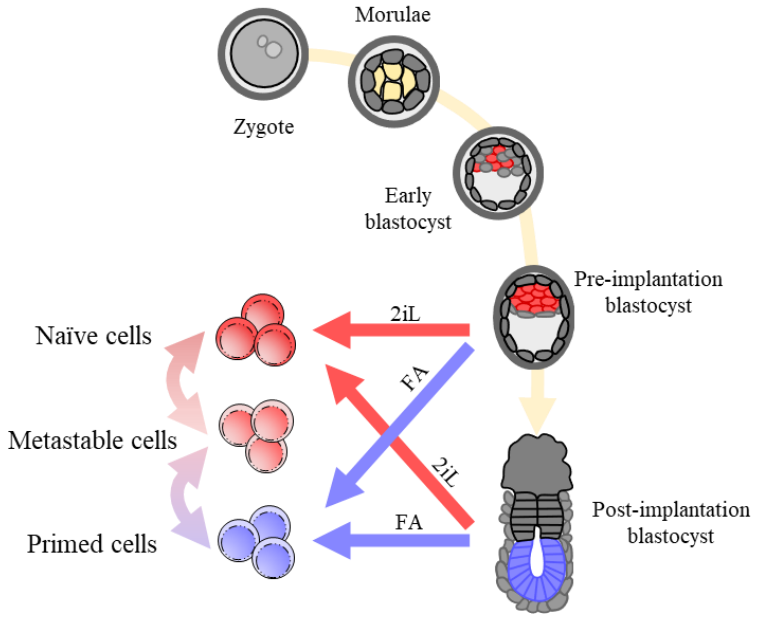


Figure I5. Pluripotent cells during *in vivo* early development and their *in vitro* counterparts. Simplified depiction of pluripotent cells during early development and their *in vitro* representation. Cells from the ICM of pre-implantation blastocysts can be derived for the establishment of naïve pluripotency, whereas cells from the epiblast of post-implantation blastocysts can be used for the establishment of primed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cells. In addition, both naïve and primed cells can be established from *their in vivo* counterparts by using defined culture conditions.

4.2.1 The ground-state of pluripotency

As described in previous sections (see section **2.1.1**), pluripotent cells are present from pre-implantation to post-implantation stages of embryo development. These different stages are characterized by both functional and structural features that can be retained indefinitely in culture under specific conditions. Classically, pluripotent cells were derived from an intact blastocyst or isolated ICM on a layer of mitotically inactive fibroblasts (termed as *feeders*) and in the presence of serum[4,5]. Those feeders are known to endow trophic factors conditioning the media to support ESC self-renewal[32,33]. Supplementation of serum-containing media with leukemia inhibitory factor (LIF) can substitute for feeders[34] in the presence of an extracellular matrix such as gelatin[33,34]. Under serum/LIF conditions, ESCs are known as "conventional" ESCs and retain the ability to differentiate into cells from the three germ layers both *in vitro* and *in vivo* by the generation of embryoid bodies (EBs) and teratomas, respectively. Moreover, high-quality pluripotent cells are tested by their contribution to form chimeras and germline transmission, as well as by tetraploid complementation approaches[35]. Remarkably, conventional ESCs fulfill all these criteria[4,5,36], so they were classically considered to be in a ground-state of pluripotency. Nevertheless, one consequence of maintaining conventional ESCs in the presence of serum is the activation of conflicting signaling pathways, which leads to a heterogeneous population with morphological, transcriptional, and functional differences[37-39]. Therefore, conventional ESCs are known to be in a metastable pluripotent state (**Figure I5**).

The accumulated knowledge of the molecular mechanisms governing the pluripotency state led to the development of specific culture conditions for the stabilization of ESC maintenance in a ground-state of pluripotency. Several studies have shown that inhibition of both ERK1/2[40] and glycogen synthase kinase 3 (GSK3) pathways[41,42] is

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

compatible with pluripotency maintenance independently of LIF supplementation. Hence, the use of specific chemical inhibitors against ERK and GSK3 signaling pathways (i.e., PD0325901 and CHIR99021, respectively) termed as 2i condition, shields ESCs from inductive differentiation signals (for details see section **2.3.2**). Cells cultured in 2i conditions and LIF are referred to as naïve cells[31] (**Figure I5**) and show transcriptional and epigenetic features resembling the pluripotent ground-state *in vivo*[43] (summarized in **Table I1**).

FEATURE	NAÏVE	PRIMED
MEK-ERK pathway activation	No	Yes
Long-term dependence on FGF2 signaling	No	Yes
Long-term dependence on TGFß-Activin A	No	Yes
Global DNA hypomethylation	Yes	No
X chromosome inactivation	No	Yes
Dominant OCT4 enhancer	Distal	Proximal
Dependence on DNMT1, DICER, METTL3	No	Yes
Priming markers (OTX2, ZIC2)	Low	High
Pluripotency markers (NANOG, KLFs, ESRRB)	High	Low
Promotion of pluripotency maintenance via NANOG and PRDM14	Yes	No
Cellular metabolism	OxPhos, glycolytic	Glycolytic
Contribution to form chimeras	High	Low

Table I1. Molecular and functional differences between naive and primed pluripotency. Specific signaling pathways and molecular features are described in both ground-state and primed cells in murine pluripotency.

4.2.2 The primed state of pluripotency

As counterparts from naïve cells recapitulating the *in vivo* ground-state of pluripotency, epiblast cells from the post-implantation blastocyst can also be derived in the presence of Nodal/FGF factors[44,45]. These cells, termed as epiblast-derived stem cells (EpiSCs) are in a more advanced developmental stage, conforming the primed state of pluripotency (**Figure I5**). EpiSCs retain the functional capacity of differentiating into cells from all the three germ layers both *in vitro* and *in vivo*[44,45]. However, when injected in pre-implantation blastocysts, they are inefficient in their contribution to form chimeric animals[44-46] due to defects in their proper integration with the surrounding epiblast

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cells[47]. This suggests that these cells are in a more advanced developmental state than the host blastocyst, undergoing functional reconfigurations leading to a restriction in their differentiation potential. Nevertheless, chimeric animals could be obtained by the injection of EpiSCs into host post-implantation embryos *ex vivo*. Therefore, a suitable environment that leads to the activation of specific signaling pathways can favor the correct integration of these cells in the developing embryo, giving rise to an adequate differentiation capacity[46-48]. Different studies have demonstrated that alternative primed culture conditions can promote the generation of EpiSCs with region-specific features. In this sense, EpiSCs cultivated under the classical FGF2/Activin A (FA) condition yield transcriptional and functional characteristics of cells of the primitive streak located in the anterior late gastrula[46]. Alternatively, EpiSCs cultured in the presence of FGF2 and IWR1 (a specific WNT inhibitor) show a high degree of similarity with cells from the posterior proximal epiblast[49]. Moreover, some degree of heterogeneity also exists in EpiSCs cultured under the classical FA medium, with cells representing different developmental stages[50]. It was also demonstrated that the length of time for which pluripotent cells are maintained under primed conditions may influence their functionality, restricting their differentiation capacity[51].

Direct interconversion between these two alternative pluripotent states (i.e., naïve and primed) was demonstrated, therefore facilitating the obtention of both ground-state and EpiSC-like cells. On one hand, the culture of mouse ESCs under FA conditions yields a progressive conversion of these cells into primed pluripotency[52,53]. Additionally, findings from this year have demonstrated that a stepwise protocol conforms a suitable approach for a highly efficient method for the naïve-to-primed interconversion[54]. In contrast, the reverse conversion (i.e., from primed to naïve) could also be performed, even though at very low rates, suggesting the existence of an epigenetic barrier between these two alternative states of pluripotency[55]. However, it was recently shown that culturing murine cells in the presence of both LIF

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and Activin A led to the obtention of ALSCs, a new cell pluripotent state showing transcriptional and phenotypic features of EpiSCs but with an increased capacity to be reverted to ESCs (rESCs)[56].

Finally, new pluripotent states have also been described. For example, an intermediate state between the naïve and primed states of pluripotency, coined as *formative*[57], was successfully derived from both murine and human embryos[58]. Transcriptomic analyses have revealed that formative pluripotent stem cells (fPSCs) show a transcriptional profile that differs from both naïve and primed pluripotent cells[59]. In addition, these cells also show intermediate epigenetic and transcriptional features between ground-state and primed pluripotency[59]. Taken together, fPSCs are envisioned as mature naïve cells prepared for lineage differentiation[57,60].

4.3 THE MOLECULAR CONTROL OF PLURIPOTENCY

Pluripotency can be understood as the bypass from the totipotent zygote to differentiated tissues conforming an adult organism. As seen above, PSCs can be obtained from the derivation of cells conforming the ICM and/or the epiblast during the early stages of embryo development (ESCs and EpiSCs, respectively). The pluripotent state is known to be governed by an intricate cooperative network of transcription factors (TFs), epigenetic regulators (ERs) and signaling pathways for the regulation of gene expression[61-63]. In addition, this specific gene expression needs to be fine-tuned for proper maintenance of the pluripotent state. However, despite the growing knowledge accumulated in recent years, the molecular underpinnings that govern pluripotency (e.g., acquisition, maintenance and exiting pluripotency) are far from being understood. In the following sections, I will provide a brief description of main transcriptional and epigenetic regulatory layers that govern the different cell states of pluripotency related to my PhD thesis work. Furthermore, I will highlight the main signaling pathways involved in crosstalk between these regulatory networks, as well as their relationship with the metabolic state of PSCs.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

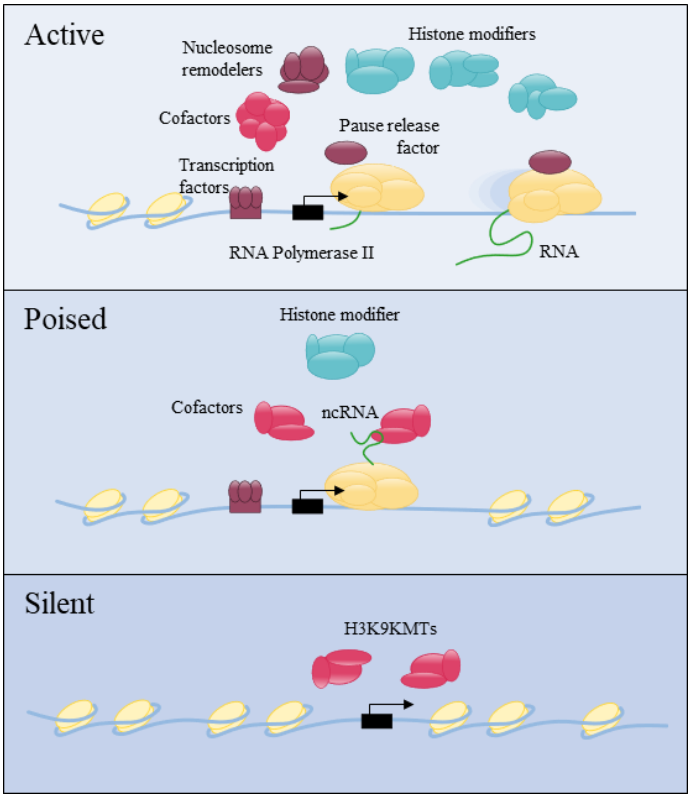


Figure I6. Representation of transcriptionally active, poised, and silent genes. Transcription factors, cofactors, and epigenetic factors can cooperate for the activation, posing, or silencing of genes. At active genes, regulatory regions, such as promoters and enhancers, are bound by TFs, which in turn can recruit cofactors that directly interact with the RNA polymerase II for transcription initiation. RNA polymerase II generates a short transcript and pauses until pause release factors allow the transcriptional elongation of the nascent mRNA. Epigenetic factors, such as nucleosome remodelers and histone modifiers, can also be recruited to the chromatin for its remodeling, facilitating the recruitment, mobilization, and stabilization of the transcriptional machinery. Poised genes are rapidly activated in ESCs under the induction of the correct extrinsic signals. At poised genes, transcription initiation can occur, but pause release and elongation of the transcript is blocked. Silent genes do not show any evidence of transcription initiation or elongation, being occupied by repressive histone marks, such as H3K9me3 (a histone mark for heterochromatin), or DNA methylation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.3.1 The transcriptional network controlling pluripotency

In mammals, transcription factors (TFs) conform the largest class of proteins encoded by the genome, with more than 10% of all protein-coding genes making for these transcriptional regulators[64,65]. TFs are proteins with the ability to recognize specific DNA sequences and bind to chromatin for the regulation of the transcriptional machinery. Hence, TFs may lead to different functions in the regulation of gene transcription by contributing to gene activation, establishing a poised state in which cells rapidly respond to environmental cues for gene expression, or transcriptional silencing (**Figure I6**).

TFs may promote gene activation by either the recruitment or release of the transcriptional machinery from pause sites[66]. Moreover, some of them called pioneer TFs can also bind to their target sequence within nucleosomal DNA, recruiting different epigenetic factors for the remodeling of chromatin configuration, which allows chromatin accessibility to the transcriptional machinery for the regulation of gene expression[67,68].

Next, I provide a brief overview of core regulators of pluripotency along with other well-known TFs involved in the regulation of pluripotent cell identity.

4.3.1.1. Core regulatory network

Pluripotency is governed by a cluster of three master regulators: OCT4 (*Octamer-binding transcription factor 4*), SOX2 (*Sry box-containing gene 2*), and NANOG (*Nanog homeobox*)[6,35,69]. Initially, both OCT4 and NANOG were highlighted as key regulators of pluripotency based on their specific expression patterns during development[19,70,71]. These two factors are known to be expressed in pluripotent cells[19,21,70,72-74]. Murine embryos deficient in *Oct4* expression were shown to die at the peri-implantation stage due to an impediment in the proper establishment of the ICM[19]. In line with this, genetic ablation of *Nanog* was also demonstrated to impede the proper

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

proliferation of this structure[21]. Genetic manipulations for loss- or gain-of-function experiments in ESCs have demonstrated the overlapping functions of these TFs. Hence, transient inhibition of both *Oct4* and *Nanog* expression pushes ESCs towards extraembryonic differentiation[21,75]. In contrast, overexpression of OCT4 leads to endoderm differentiation[21,76], thus phenocopying *Nanog* deletion[21,75] and leading to the idea of an existing regulatory loop in the transcriptional control of the central pluripotency circuitry (**Figure I7**).

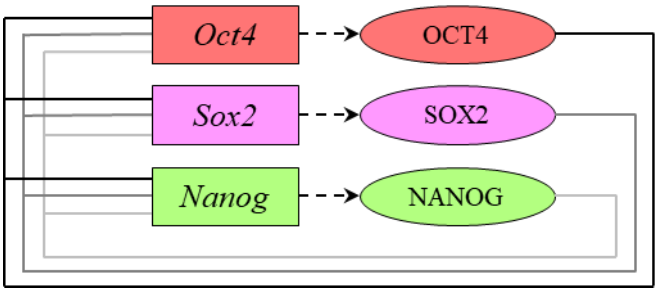


Figure I7. Autoregulatory feeding loops in the core transcriptional network regulating pluripotency. The expression levels of these core transcription factors need to be fine-tuned regulated for the proper maintenance of the pluripotent state. Therefore, a negative loop governed by OCT4 exists in the regulation of *Nanog* expression. In addition, both OCT4 and SOX2 interact for the positive regulation of *Nanog* levels, along with *Sox2* and *Oct4* themselves.

Chromatin occupancy analyses of these factors have demonstrated a high degree of overlap in their binding regions[7], suggestive of strong crosstalk between them in the regulation of pluripotency. Moreover, protein interaction analyses of OCT4[77], NANOG[78,79], and SOX2[80] provided evidence of an intricate network of factors safeguarding the pluripotent cell identity. Among them, several zinc finger proteins such as ZFP281, ZFP42 (also known as REX1), and members of the KLF protein family (i.e., KLF2 and KLF4) were demonstrated to be part of an extended transcriptional network for ESCs[81] (**Table I2**). Hence, KLF4 was originally demonstrated to facilitate the direct conversion of somatic cells into pluripotent cells when co-transfected with other reprogramming factors (e.g., OCT4, SOX2, and c-MYC)[82] (see section **4.4.2**). In addition, KLF4 can act by

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

integrating into the NANOG transcriptional network to specify the ESC gene expression[83]. Furthermore, this factor was also demonstrated to be a downstream effector of the LIF/STAT3 pathway for the ESC self-renewal[84]. Recently, authors have suggested that KLF4 might be orchestrating chromatin reorganization by forming condensates at its genomic loci[85], therefore facilitating the activation of the pluripotent transcriptional program. In contrast, ZFP281 was demonstrated to play a role in the regulation of alternative pluripotent states of cells[86,87]. In parallel, recent reports have shown the direct implication of this factor in safeguarding pluripotency by recruiting chromatin remodelers such as PRC2 for the transcriptional repression of extraembryonic endoderm markers[88]. Lastly, KLF2 was implicated in the maintenance of an extended pluripotency state[89,90].

TRANSCRIPTION FACTOR	FUNCTION	REFERENCE
OCT4	Core circuitry	19,21,70,76,91-93
SOX2	Core circuitry	91 94
NANOG	Core circuitry	21,70,92
KLF2	Pluripotency maintenance	89,90,95,96
KLF4	LIF signaling	83 84
ZFP281	Pluripotent cell identity	86,87
ZFP42	Extended pluripotency network	81

Table I2. Examples of transcription factors involved in the control of pluripotency. Well-known characterized transcription factors with functional relevance in the molecular control of pluripotency.

4.3.2 Main signaling pathways involved in pluripotency

Cells recognize and integrate environmental cues to maintain their homeostasis. In this sense, in parallel with TFs, signaling pathways act through two alternative mechanisms: activation and promotion of the pluripotency state, or inhibition of differentiation cues[97]. Moreover, a tight connection between signaling pathways and transcriptional regulation exists in the pluripotent context (**Figure I8**). Classically, ESCs were cultured on a monolayer of irradiated fibroblasts for the proper maintenance of an undifferentiated state[98,99]. Later, it was demonstrated that fibroblasts produce molecules responsible for the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

activation of the pluripotency response. One example is FGF proteins, which act through the extracellular signal-regulated kinase (ERK) signaling pathway promoting cell differentiation[100]. This signaling pathway is endogenously controlled by pluripotency TFs (e.g., OCT4 and SOX2 are required for the proper activation of *Fgf4*)[101]. *Leukaemia inhibitory factor* (LIF) is another molecule contained in fibroblasts-conditioned media promoting ESC self-renewal[102,103] that can completely bypass the requirement for fibroblasts in the *in vitro* maintenance of ESCs. This factor activates the JAK/STAT3 signaling pathway promoting JAK-dependent phosphorylation of STAT3[102]. Among LIF-STAT3 targets are the pluripotency regulators KLF4 and TFCP2L1[84,104,105]. Overexpression of these factors enables self-renewal in ESCs in a LIF-independent manner through upregulation of *Nanog*[104,105]. In addition, it was further demonstrated that *Nanog* overexpression leads to the proper maintenance of self-renewal in ESCs even in the absence of LIF[84], further validating the strong interplay between this signaling pathway and the transcriptional regulation of pluripotency. BMP proteins present in serum-containing media act through SMAD transcription factors that activate the expression of *Id* genes inhibiting differentiation programs[106]. The cooperative function of both BMP and LIF pathways is required for the proper maintenance of self-renewal in ESCs[106]. In addition, chromatin binding analyses also demonstrated that both STAT3 and SMAD1 co-occupy the vast majority of chromatin regions with OCT4, SOX2, and NANOG[107]. Moreover, it seems likely that this co-occupancy is highly dependent on OCT4 since depletion of this pluripotency regulator reduces the occupancy levels of those factors[107]. Remarkably, LIF and BMP pathways also cooperate counteracting the pro-differentiation cues provided by the activation of the FGF signaling pathway, further stabilizing the pluripotent state.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

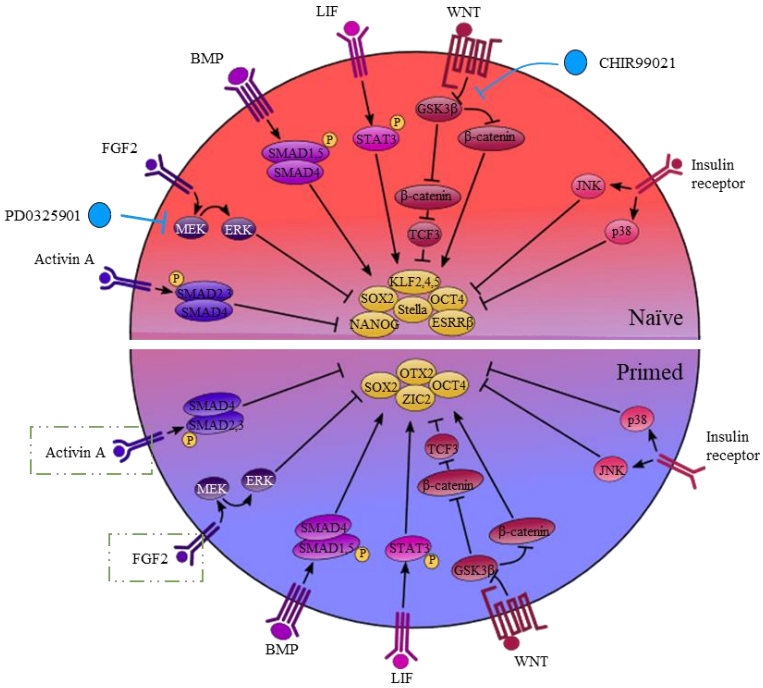


Figure I8. Signaling pathways that are involved in the control of pluripotent cell identity. (*Top*) A simplified depiction of the main signaling pathways that regulate the naïve pluripotent state. The MEK-specific inhibitor (PD0325901) blocks the ERK signaling pathway, facilitating the expression of genes important for the ground-state of pluripotency. Moreover, the GSK3β chemical inhibitor (CHIR99021) leads to the blockage of the WNT signaling pathway involved in the repression of pluripotency factors. The use of these two small molecules in a defined serum-free medium promotes a naïve pluripotent state *in vitro*. (*Bottom*). Simplified depiction of main signaling pathways regulating the primed state of pluripotency. Activation of the ERK and Nodal signaling pathways using the cytokines Fgf2 and Activin A, respectively, in a defined serum-free is required for the primed pluripotent state *in vitro*. Other important signaling pathways are shown in these two alternative pluripotent states. Arrows indicate activation, while bars indicate inhibition.

As mentioned before, fine-tuning of the steady-state amounts of factors involved in these signaling pathways are required for the maintenance of the self-renewal capacity of ESCs. Low activation of differentiating pathways pushes cells towards a more responsive state to differentiation cues. However, complete inhibition of some signaling pathways is compatible with an extended potency of ESCs. For

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

example, blocking the WNT pathway through chemical inhibition of GSK3 leads to the maintenance of pluripotency in both human and mouse ESCs[41]. Remarkably, inhibition of MAPK/ERK and GSK3signaling pathways with small compounds (PD0325901 and CHIR99021, respectively) has been shown to promote a naïve pluripotent state[40,89,96,100,108] (**Figure I8**, top panel). On the other hand, exogenous induction of FGF-ERK signaling pathway by adding FGF2 to the medium, in parallel with activation of the NODAL pathway with Activin A supplementation leads to a more advanced state of pluripotency, the so-called primed pluripotent state[44,45] (**Figure I8**, bottom panel).

4.3.3 Epigenetic regulation of pluripotency

Epigenetics is defined as changes in gene expression that occur without alteration in the DNA sequence and are mediated by multiple epigenetic regulators including writers, readers, erasers of histone posttranslational modifications, DNA and RNA modifications, as well as histone variants and non-coding RNAs[109,110]. Over the past years, it has become clear that pluripotency is controlled by global epigenetic mechanisms such as DNA methylation, histone modifications, and regulation of ATP-dependent remodeling of chromatin[62]. Pluripotent cells show open and accessible chromatin with global higher acetylation and lower methylation levels than differentiated cells[111,112].

Here, I will briefly summarize the main epigenetic regulatory layers involved in transcriptional control in pluripotency, including the chemical modifications most related to this thesis work that are deposited on DNA (i.e., methylation) and the N-terminal “tails” of histones (i.e., acetylation, methylation, and citrullination).

4.3.3.1. DNA methylation

DNA methylation on the 5th position of cytosine, known as 5mC, is one of the best-studied epigenetic marks in mammals, acting primarily as a stable repressive regulator of promoter activity and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

essential in mammalian development[62,113,114]. DNA methylation is specifically catalyzed by various types of DNA methyltransferase (DNMT) enzymes involved in two different methylation processes: maintenance of methylation of previously established 5mC patterns during each round of DNA replication (i.e., DNMT1), and *de novo* methylation marks (i.e., DNMT3a and DNMT3b) mainly on CpG islands[113,115]. While 5mC can be passively diluted during DNA replication, the active removal of this mark can occur by the iterative oxidation by the ten-eleven-translocation family of enzymes (i.e., TET1, TET2, and TET3) from 5mC to 5-hydroxymethylcytosine (5hmC) to 5-formylcytosine (5fC) to 5-carboxylcytosine (5caC) and the action of thymine-DNA glycosylase (TDG) on 5fC and 5caC to generate unmodified cytosine (C) via base excision repair (BER) mechanism[116-118] (**Figure I9**).

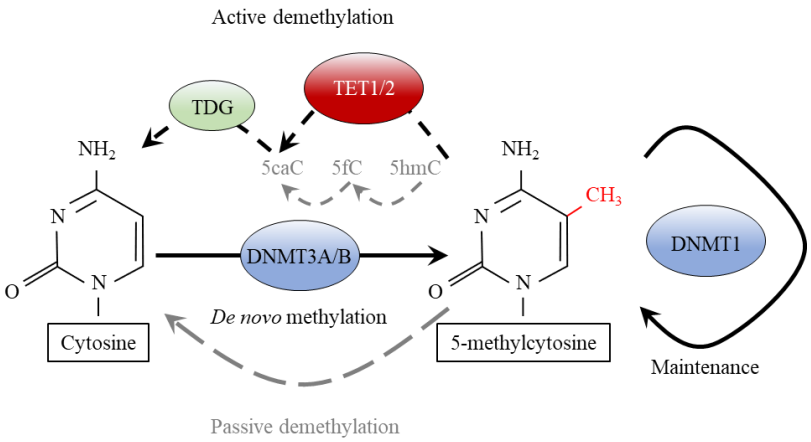


Figure I9. Enzymatic activities of cytosine methylation and demethylation machineries. In mammals, *de novo* deposition of a methyl group at the fifth carbon of a cytosine residue is carried out by DNMT3A and DNMT3B (straight black arrow). DNA methylation patterns are maintained by DNMT1 during cell replication. Passive demethylation (grey dotted arrow) occurs during serial cell divisions. Active demethylation (black dotted arrow) is carried out by members of the TET family (i.e., TET1, TET2, or TET3), leading to the conversion of 5mC to 5hmC/5fC/5caC intermediates. Finally, the base excision repair of 5fC/5CaC completes the DNA demethylation process.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

During the last years, numerous evidences have demonstrated a crucial role of DNA methylation in the regulation of gene expression for proper cell fate decisions during early development and pluripotency (i.e., the acquisition and maintenance of pluripotency, and differentiation)[119-121]. Pluripotent cells are characterized by a global DNA hypomethylation level in comparison to somatic cells[62,122,123]. Although the absence of all three catalytic active DNMTs or all three TETs is compatible with pluripotent maintenance of mouse PSCs under standard culture conditions (i.e., presence of serum and LIF), their actions are crucial for naïve and primed pluripotency as well as for acquisition and exiting of the pluripotent state[86,117,119,120,124,125].

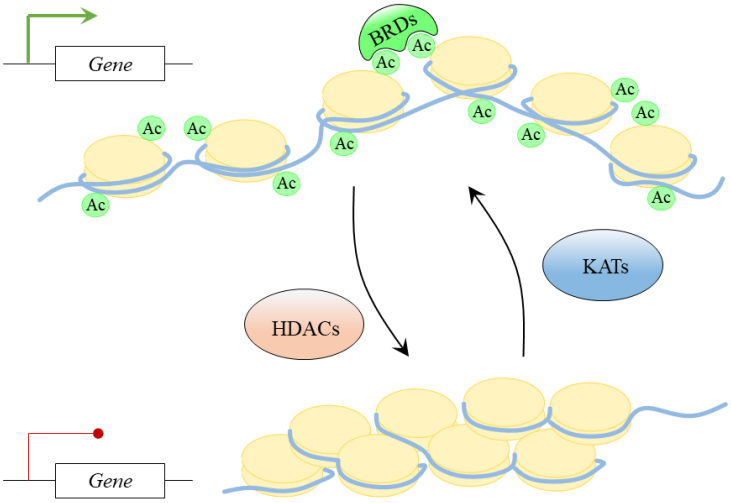


Figure I10. Histone acetylation dynamics modulates chromatin conformation. Histone acetylation (Ac) levels depend on the action of lysine acetyltransferase (KATs) and histone deacetylases (HDACs). High levels of histone acetylation lead to a more open chromatin, allowing the active transcription of adjacent genes. In contrast, low levels of histone acetylation translate in a compact chromatin configuration, impeding the expression of genes.

4.3.3.2. Histone acetylation

Acetylation of histones is a key epigenetic modification involved in transcriptional regulation playing a crucial role in determining cell fates in development and pluripotency[62,126]. Histone

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

acetylation works by neutralizing a positive charge in histone tails, preventing compaction of chromatin between adjacent nucleosomes, and allowing a more open chromatin structure that facilitates transcriptional activation at the surrounding regions[127] (**Figure I10**).

Among the histone acetylation marks, H3K27ac is of particular interest because it distinguishes active enhancers from poised ones[128]. In addition, recent reports have demonstrated the relevance of this histone mark in the rapid resetting of the pluripotent transcriptional program soon after cell division, suggestive of its role in the ESC self-renewal[129]. Importantly, global levels of histone acetylation are regulated by both acetyltransferases (KATs) and deacetylases (HDACs) through the addition and removal of acetyl groups to lysine residues on histone tails, respectively, which are recognized by proteins containing bromodomains (BRDs) to further promote regulation of gene expression[130,131] (**Figure I10**).

4.3.3.3. Histone methylation

Histone methylation has also become recognized as a crucial player in development and pluripotency through the regulation of gene activity[132]. The major methylation sites within the histone tails are the basic amino acid chains of lysine and arginine residues by the action of histone methyltransferases (HMTs) that use S-adenosyl-methionine (SAM) as the methyl group donor[133]. Methylated lysines can be found either in a mono-, di-, or trimethylated state, whereas arginines can be either mono- or demethylated[134]. For example, active enhancers and promoters carry H3K4me1 and H3K4m3, respectively, while the presence of H3K9me2/3 and H3K27me3 is associated with transcriptional silencing[62].

Increased levels of these repressive histone marks appear upon cell differentiation[135,136], facilitating the recruitment of DNMTs for proper gene silencing[137]. The impact of histone methylation on development and pluripotency depends not only on the recruitment of

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the HMTs that catalyze these chemical modifications but also on de level of demethylases (erasers) and proteins that recognize these posttranslational histone modifications (readers)[138] (**Table I3**). Of special interest throughout the results of my thesis work, Lysine Demethylase 1 (LSD1), also known by its official gene name *Kdm1a*, can demethylate both H3K4me1/2 and H3K9me1/2 acting as a coactivator or a corepressor of gene expression[139]. LSD1 is essential during early embryogenesis[140,141]. Genome-wide analysis of LSD1 occupancy in ESCs shows its presence in regulatory regions (e.g., enhancers and promoters) along with pluripotency TFs such as OCT4, SOX2, and NANOG[142]. Although LSD1 is not essential for the maintenance of murine ESC identity under standard culture conditions, its genetic ablation results in differentiation deficiencies[141-143]. Remarkably, there is also a strong correlation between histone and DNA methylation, and histone acetylation. For example, the absence of LSD1 leads to DNA hypomethylation[143] and increased levels of histone acetylation[141], reinforcing the idea of a tight interplay between epigenetic marks for transcriptional regulation.

	WRITERS	READERS	ERASERS
H3K4me3	MLL1-4, SETD1A/B, SETD7, SMYD1/2, PRDM9	CHD1/2/7/8, BPTF, TAF3, ING2	LSD1/2, JARID1A-D, NO66
H3K9me3	GLP, PRDM2, Suv39H1/2, G9a, SETDB1	HP1α/β/γ, UHRF1	JMJD1A-C, JHDM1D, JMJD2A/B/C/D, PHF2/8
H3K27me3	EZH1/2	CBX7, CDY/L/L2, MPP8	JHDM1D, JMJD3, UTX, UTY
H3K36me3	NSD1/2/3, ASH1L, SETD2/3, SETMAR	PHF2/PCL1, PHF19/PCL2, EAF3	FBXL10/11
H3R17me2	CARM1	Unknown	Unknown
H4R3me2	PRMT1/5/6/8	Unknown	Unknown

Table I3. Epigenetic regulators of the histone methylation status in pluripotency. The global levels of histone marks depend on the interplay between different

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

epigenetic factors depositing them (writers), recognizing them (readers), and eliminating them (erasers)[144].

4.3.3.4. Histone citrullination

In contrast to the extensive characterization of recent years on the role of histone acetylation and methylation in the regulation of gene activity in development and pluripotency, the specific function of histone citrullination in these cellular contexts has almost been unexplored. This covalent modification in histones and non-histone proteins is mediated by a small family of hydrolases known as peptidyl arginine deiminases (PADIs) that convert protein-encoded arginines into citrullines through a calcium-dependent reaction called citrullination or deimination[145] (**Figure I11**). There are five members (i.e., *Padi1*, *Padi2*, *Padi3*, *Padi4*, and *Padi6*) of the PADI family in the murine genome. Increased levels of some of these PADI proteins have been implicated in human diseases such as multiple sclerosis, rheumatoid arthritis, and cancer progression[146-148]. It is noteworthy that, to date, PADI4 is the only member of the PADI family that has been reported to be localized in the nucleus of cells[149], thus it is reasonable to speculate that histone citrullination is likely mediated by this enzyme. Histone citrullination can occur at different arginine residues in the same histone protein, resulting in transcriptional activation or repression. Hence, for example, H3R26Cit promotes gene activation, whereas H3R17Cit results in gene silencing[150]. Furthermore, histone citrullination has been shown to counteract the effects of arginine methylation[151,152]. Recently, the role of PADI4-induced citrullination in the regulation of mouse pluripotency and chromatin compaction has been highlighted, demonstrating that PADI4 resides within the pluripotency transcriptional network and disruption of citrullination reduces acquisition of pluripotency and impairs early development and maintenance of pluripotency[153].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

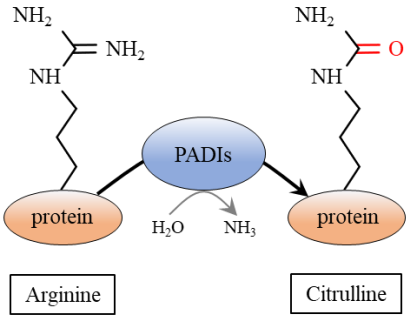


Figure I11. Protein deimination is mediated by members of the PADI family. Members of the PADI family (denoted in blue) carry out the conversion of an arginine residue to citrulline in a reaction termed deimination or citrullination.

4.3.4 Metabolic state in pluripotent cells

In recent years, technological advances in the study of metabolomics and computational analyses have highlighted the relevance of metabolic pathways and their correct counterbalance in the regulation of cell fate decisions and the maintenance of cell identity. Therefore, metabolism can be interpreted as the centerpiece of the intricated puzzle formed by the epigenetic and transcriptional regulatory networks that govern gene expression. In the following sections, I will summarize various evidence for the active role of metabolism in early development and pluripotency.

4.3.4.1. Metabolism in early embryo development and pluripotency

As the zygote develops into a full organism, cells must accommodate different requirements for cell growth, proliferation, and differentiation[154]. Totipotent cells show reduced glycolysis given the reduced expression levels of hexokinase (HK) and phosphofructokinase (PFK1), the two main glycolytic enzymes, and the only carbon source for energy production is pyruvate[155,156]. Thus, this two-cell stage shows a reduction in citrate and α-ketoglutarate (αKG), along with reduced levels of acetylation of histones (H4K4ac and H3K27ac) and transcriptional activity[157].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

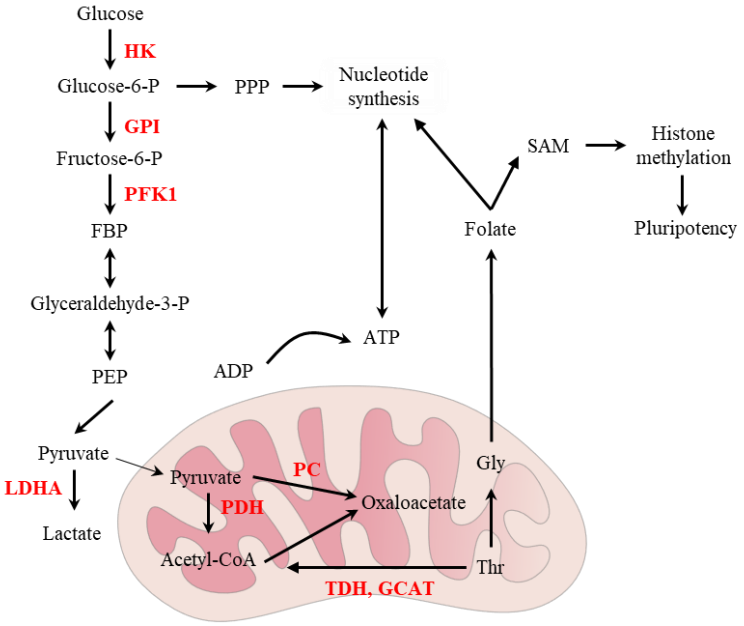


Figure I12. Cell metabolism in pluripotent cells. As development progresses, totipotent cells become pluripotent, which is accompanied by an increase in the expression levels of glycolytic genes allowing an increase in the glycolytic rate of these cells. Moreover, an increase in the oxidative phosphorylation occurring in the mitochondria takes place, leading to an increase in the production of energy that is required for the synthesis of different metabolites.

Furthermore, under conditions of high concentrations of glucose, embryo development is inhibited[158]. Even though the oxidation of pyruvate is the preferred source of energy in totipotent cells, the oxygen consumption rate (OCR) in these cells is very low due to an increased cellular content on immature mitochondrias[157]. In contrast, as the embryo develops towards the morula stage, glucose oxidation increases to finally compete against pyruvate metabolism[159]. Accordingly, single-cell RNA transcriptomes of pre-implantation embryos have demonstrated an increase in the expression levels of genes governing oxidative phosphorylation during the blastocyst stage[160,161]. At later stages, when the morula develops into the blastocyst, the first wave of differentiation occurs (for more details, see

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sections **4.1.1** and **4.3.1**). Here, a change in the cellular metabolic demands occurs due to an increase in the proliferation rate of these cells[154]. Glucose transporters (GLUT1 and GLUT3) become highly expressed with the consequent increase in glucose uptake and flux[162]. This phenomenon leads to an increase in the production of lactate[159] (see **figure I12**). Moreover, an increase in the OCR is also observed in the blastocyst, followed by an increase in the oxidative phosphorylation metabolic pathway. However, these changes are mainly due to the metabolic switch occurring in trophectoderm cells[163].

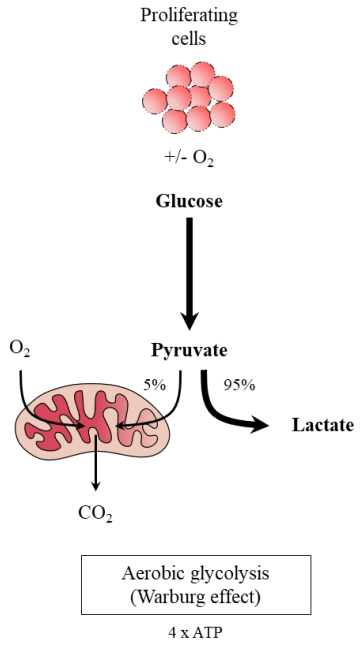


Figure I13. Highly proliferating cells present a metabolic profile reminiscent of the Warburg effect. Pluripotent cells show a metabolic profile reminiscent of the Warburg effect observed in cancer stem cells, in which an increase in the anaerobic glycolytic pathway is observed in the presence and/or absence of O_2. Hence, glucose is mostly converted into lactate, with a decrease in the production of pyruvate for the activation of the oxidative reactions taking place in the mitochondria.

In this sense, both cells from the ICM and ESCs under culture conditions are characterized by an increase in glucose consumption[164,165]. This phenomenon is known to be reminiscent of the Warburg effect seen in cancer stem cells[165], which is characterized by rapid glycolysis followed by lactate production in anaerobic conditions (**Figure I13**). This unique metabolism offers a rapid rate of energy

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

obtention with minimal reactive oxygen species (ROS) production, which is demonstrated to be detrimental for cells. Furthermore, glycolytic intermediates are used as substrates in the pentose phosphate pathway (PPP), allowing a rapid synthesis of nucleotides[165,166] (**Figure I12**). Although O_2 consumption is important in ESCs, ATP production seems to be decoupled from it, so it highly depends on glycolysis[167]. Hence, it is hypothesized that O_2 is used in the electron transport chain (ETC) occurring in mitochondria for the restoring of NAD^+ levels and maintaining the Krebs cycle flux[154].

4.3.4.2. Metabolic differences between pluripotent cell states

As mentioned in the previous section, ESCs exhibit a high glycolytic flux to sustain rapid proliferation[167]. However, metabolic differences exist between alternative pluripotent states. Naïve cells are more dependent on respiration than primed cells[168] (**Figure I14**). This is in accordance with their lower proliferation rates[157]. Furthermore, these distinct metabolic requirements translate into remarkable epigenetic differences.

Metabolites produced by different metabolic pathways reverts in the availability for substrates to generate epigenetic marks. Hence, naïve cells show remarked hypomethylation levels due to the decreased levels of S-adenosylmethionine (SAM) intermediate[157]. In addition, ground-state cells devoid of *c-Myc* and *N-Myc* enter a quiescent state, similar to diapause embryos[156,169] with lowered expression levels of glycolytic genes[156]. Also, the expression of different pluripotency factors influences the metabolic state of cells. *Stat3*, previously highlighted as sustaining naïve pluripotency (for details, see section **2.3.2**), directly regulates the expression levels of genes promoting respiration[170], whereas c-Myc directly trans-activates glycolytic genes[171,172]. *Esrrb* also induces OxPhos during the primed-to-naïve transition[173] and it acts synergistically with *Zic3* for the reinforcement of the glycolytic pathway during the pluripotency acquisition[173]. In

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

addition, metabolic intermediates can act as substrates for different epigenetic marks. αKG, a TCA intermediate, is required as a co-factor for TET proteins, impacting methylation levels of H3K9me3 and H3K27me3 marks in naïve, metastable, and primed cells[174,175]. Moreover, SAM can influence the H3K4me3 levels in ES cells acting as the main donor of methyl groups[176], whereas acetyl-CoA produced in the mitochondrial oxidation of pyruvate affects global acetylation levels in murine ESCs[177].

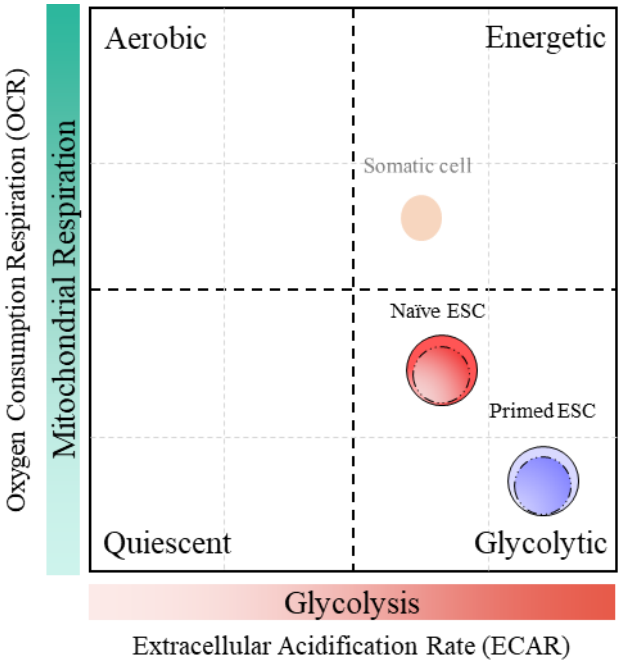


Figure I14. Cells with alternative pluripotent states show different bioenergetic requirements. Both mitochondrial respiration (determined by the oxygen consumption rate, OCR) and glycolysis (determined by the extracellular acidification rate, ECAR) defined four different metabolic states (i.e., aerobic, energetic, quiescent, and glycolytic). ESCs are known to be in a glycolytic state, whereas differentiated somatic cells show an energetic state with high levels of mitochondrial respiration. However, some differences in the metabolic requirements of pluripotent cells define the alternative states of pluripotency. Hence, naïve cells showed an increased OCR when compared to primed cells, which are defined by lower mitochondrial respiration and an increased glycolytic rate.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.4 FROM SOMATIC CELLS TO PLURIPOTENCY

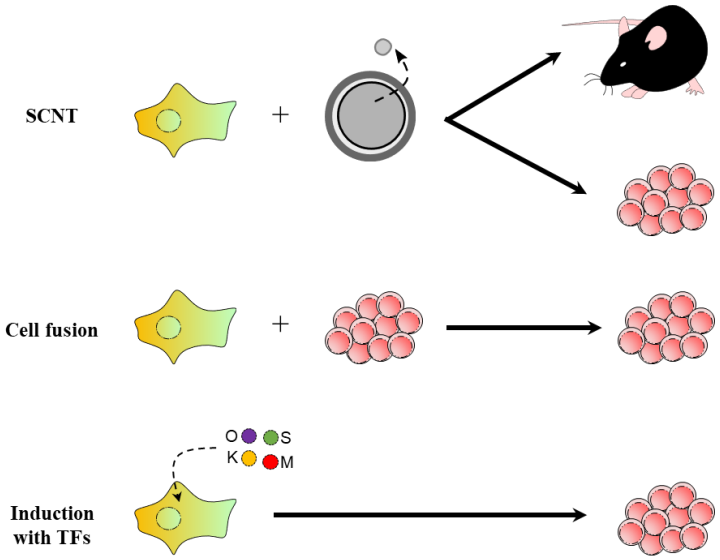


Figure I15. Different approaches for somatic cell reprogramming. (*Top*). Somatic cell nuclear transfer consists of the injection of a somatic nucleus into an enucleated egg, giving rise to a clone after transplantation into a foster mother, or to cells resembling ESCs. (*Middle*). Cell fusion consists of the physical union of a recipient cell with a donor cell, leading to the transcriptional rewiring of the recipient cell. (*Bottom*). Transfection of reprogramming factors (OSKM) into somatic cells leads to the generation of iPSCs.

During the early development, cells progressively lose their potency until they acquire a differentiated state (for more details see section **4.1.1**). However, back in 1,962, Sir John Gurdon demonstrated that adult somatic cells contain all the genetic information required for the generation of an adult organism[178]. Moreover, several experiments have shown that ESCs possess molecular contents necessary for the reversion of the cell state back to pluripotency when fused to differentiated adult cells[179,180]. Finally, seminal works reveal that the forced expression of defined transcription factors can completely rewire the transcriptional program of cells[82,181]. Thus, three different approaches have been developed and widely used for cell fate reversion into pluripotency:

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

somatic cell nuclear transfer (SCNT), cell fusion, and induction of pluripotency mediated by transcription factors (**Figure I15**). In the following sections, a brief review of the main highlights from the last two approaches that will be used during my thesis work will be discussed. In addition, transcriptional, epigenetic, and metabolic changes occurring during the reprogramming process will be also summarized.

4.4.1 Cell fusion for reprogramming pluripotency

Cell fusion is an approach to reprogramming somatic genomes involving fusion between two or more cells to form hybrids and heterokaryons[182]. On one hand, hybrids emerge from the cytoplasmatic and nuclear fusion of both cell types participating in the cell fusion approach. These cells can synthesize DNA for cell division and proliferation. However, loss of nuclear contents is typical, so genetic instability may lead to aberrant phenotypes and cell death. On the other hand, heterokaryons appear early after the fusion of the parental cell types, and cytoplasmic contents are mixed, whereas nuclei remain intact. Hence, these cells do not properly differentiate.

Classically, it was demonstrated that fusion between different cell types leads to the alteration of the transcriptional program of the recipient cell. Hence, pieces of evidence for this alteration were reported after inter-specific fusion of two distinct mammalian cell types leading to the transcriptional rewiring of the recipient cell[183]. Cell fusion was also used for the molecular characterization of tumor cells, so the fusion of malignant cells with a non-malignant cell type led to the generation of malignant hybrids[184,185], suggesting that the malignancy program somehow has a dominant character. Both tumorigenic transformation and pluripotency share common features, so a dominant role of the pluripotency network could be hypothesized. For this reason, different types of pluripotent cells have been used in cell fusion approaches to study the molecular underpinnings regulating the pluripotent program (**Figure I16**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

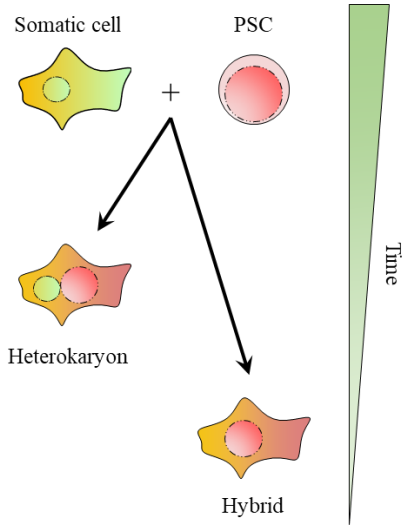


Figure I16. Directed fusion of somatic cells with PSCs. Cell fusion-based reprogramming consists of the fusion of an adult somatic cell with a pluripotent stem cell (PSC). Soon after the fusion, the mixing of cytoplasmic contents of both cell types takes place, leading to the generation of a heterokaryon with double DNA content. These cells are characterized by a high degree of genome instability and very low proliferation rates. When cytoplasm and karyoplasm contents from both cell types are mixed, hybrids are formed, which can replicate their DNA content for cell division and proliferation.

In this sense, embryonic germ cells (EGCs) showing extensive demethylating patterns at both imprinted and non-imprinted loci became a useful tool for the study of this transcriptional and epigenetic rewiring[186]. When fused to somatic cells, EGCs can erase the DNA-methylome of differentiated cells[187], leading to their rejuvenation and activation of their pluripotent transcriptional program. On the other hand, ESCs have also been used to decipher the molecular relevance of different factors during cell reprogramming. OCT4[188], NANOG[189], and SALL4[190], for example, were demonstrated to be relevant in the activation of the pluripotency network by using cell fusion approaches. Furthermore, it was also demonstrated that hybrids formed by the fusion between ESCs and somatic cells can restore the epigenome of the recipient cell[191], indicative of the great capacity of these cells in promoting the proper reprogramming of adult differentiated cells. In addition, complete reprogramming of human fibroblasts was achieved by using human ESCs as donors[180], providing evidence that the reprogramming capacity of ESCs is conserved across species. Furthermore, the high quality of reprogrammed cells obtained from cell fusion using ESCs as donors was reported, including the generation of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

embryoid bodies (EBs)[192] and the ability to contribute to the formation of chimeras in mice[179,193,194].

Importantly, cell fusion approaches can be used to decipher the specific role of nuclear versus cytoplasmatic components during somatic cell reprogramming. In this sense, in a very elegant study, Do and Schöler induced the reprogramming of neural stem cells (NSCs) by fusion with the cytoplast or karyoplast of ESCs, demonstrating that only the nuclei extracted from these pluripotent cells are capable of reprogramming NSCs[195]. Thus, a key role of specific components located in the nucleus could be determined, highlighting the functional relevance of the cell fusion assay to decipher the molecular mechanisms that govern pluripotent cell identity.

4.4.2 Induction of pluripotency with transcription factors

In 2006, Dr. Kazutoshi Takahashi and Dr. Shinya Yamanaka reported that mature somatic cells can be reprogrammed to immature cells capable of developing into all tissues of the body[82]. In their pioneering work, they demonstrate that is possible to reprogram mouse adult somatic cells (i.e., fibroblasts) into pluripotent cells by the simple transfection of four transcription factors (i.e., OCT4, SOX2, c-MYC, and KLF4)[82]. One year later, the same authors also showed that induced pluripotent cells (iPSCs) can also be generated from human somatic cells using the same combination of TF[196]. Subsequent studies have revealed that this and other combinations of factors can successfully rewire the transcriptional landscape of murine and human somatic cells, leading to their rejuvenation and reprogramming to pluripotency[197-200]. For example, the substitution of KLF4 and c-MYC with NANOG and LIN28 from the original reprogramming cocktail, respectively, was demonstrated to efficiently generate human iPSCs[201], whereas a combination of SALL4, ESRRB, NANOG, and LIN28 was used for the generation of murine iPSCs[202]. These findings demonstrate that even OCT4, originally seen as the essential factor for the acquisition of pluripotency, is dispensable for the reprogramming of somatic cells and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

that activation of the pluripotency program does not depend on specific factors but in a synergic function of an intricate transcriptional network.

Comparative studies have demonstrated that murine iPSCs are almost indistinguishable from mouse ESCs[203,204]. Indeed, they are morphologically identical and show a high degree of similarity in their differentiation potential[203-206]. Moreover, the same reports have demonstrated the ability of iPSCs to integrate into the ICM of blastocysts and in the generation of fully termed mice by tetraploid complementation[203-206], suggestive of the ability to obtain high-quality iPSCs.

Early studies have demonstrated how slow and inefficient the induction of pluripotency is. Evidence for the molecular underpinnings regulating this process has indicated the existence of alternative cellular trajectories during the process. Therefore, both permissive and refractory trajectories can be achieved during the early phases of the reprogramming to pluripotency[82,207-214], suggestive of an initial stochastic model (**Figure I17**) in which cell fate decisions are taken in a non-deterministic manner[215]. In this sense, some reprogrammed somatic mesenchymal may undergo morphological changes that lead to a mesenchymal-to-epithelial transition (MET), thus acquiring a permissive path for adequate reprogramming process[216-222]. In contrast, other cells can manifest a senescent phenotype along with genomic aberrations, leading to a cellular trajectory that is refractory for the induction of pluripotency[211]. Additionally, during the final stages of the reprogramming process, a hierarchical and sequential model has been evidenced in which SOX2 acts as a leading factor inducing the activation of the pluripotency network[215] (**Figure I17**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

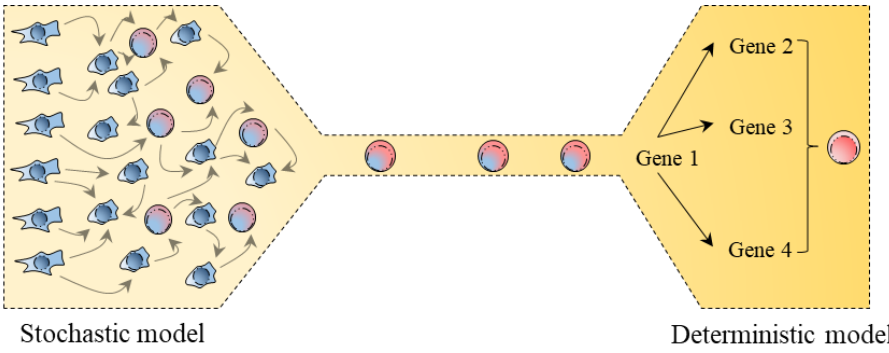


Figure I17. Probabilistic models of somatic cell reprogramming. The OSKM reprogramming process from fibroblasts to iPSCs can broadly be divided into a long stochastic first phase of gene activation and a short deterministic second phase that begins with the activation of the *Sox2* gene facilitating the sequential and hierarchical activation of different pluripotent regulators.

4.4.3 Changes occurring during the acquisition of pluripotency

During somatic cell reprogramming, cells are submitted to different molecular changes occurring sequentially. Characterization of these molecular mechanisms has demonstrated that cells can undergo different trajectories leading to the generation of both permissive and refractory cellular states for the acquisition of pluripotency[207-213] (**Figure I18**). Refractory reprogramming intermediates are characterized by genomic aberrations leading to a secretory phenotype and cell senescence[211]. In contrast, permissive trajectories allow the overall repression of the somatic program and the activation of the pluripotency network, allowing the correct generation of iPSCs. In this section, I will discuss the major transcriptional, epigenetic, and metabolic changes that occur during reprogramming.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

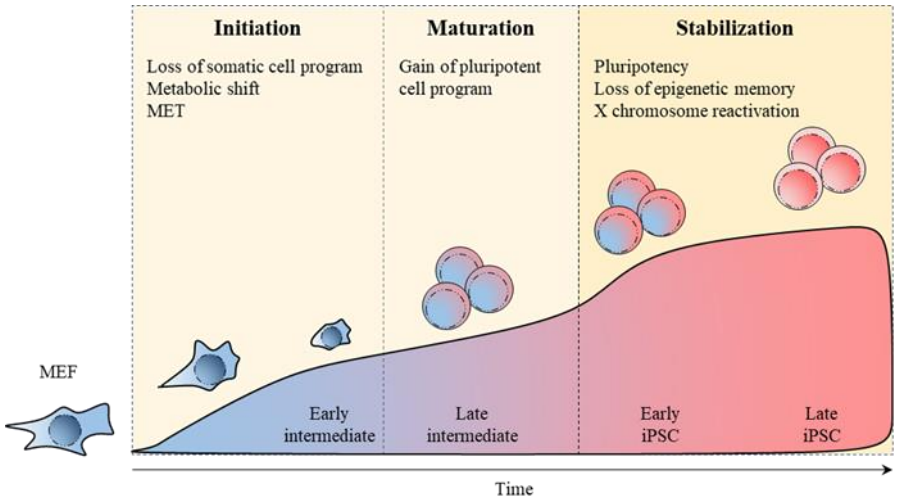


Figure I18. Permissive phases of the reprogramming process. Schematic representation of successful reprogramming phases for iPSC formation after somatic mesenchymal fibroblasts (i.e., MEFs) are induced with the OSKM cocktail. In the early phase (i.e., initiation), reprogrammable cells initiate the mesenchymal-to-epithelial transition. These cells will then enter in a long intermediate phase (i.e., maturation) before beginning the conversion to a pluripotent state (i.e., stabilization). Some important characteristics of each phase are shown.

4.4.3.1. Silencing of the somatic program

Processes occurring during the first steps of reprogramming are well characterized[223]. Here, the first wave of transcriptional changes occurs[224] in which repression of the somatic program conforms the first requirement for an efficient reprogramming[225]. Loss of accessibility of regulators governing the activation of specific genes controlling the somatic program accounts for this transcriptional repression[226-228]. In this sense, the TFs OCT4, SOX2, and KLF4 were demonstrated to actively participate in this transcriptional silencing[229]. Remarkable repression of several mesenchymal genes (e.g., *Snai1*, *Snai2*, *Twist1*, *Twist2*, *Zeb1*, *Zeb2*, *Thy1*, and *Cd44*) can be observed[210,230,231], whereas a transcriptional activation of epithelial genes occurs. This transcriptional shift leads to morphological changes reminiscent of a mesenchymal-to-epithelial transition (MET)[230,232]. However, transient induction of the epithelial-to-mesenchymal transition (EMT) programs

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

before pluripotency were demonstrated to further enhance the reprogramming efficiency[216,222,233,234], which can be promoted by a sequential induction of the Yamanaka factors[222].

Although TFs are important for this initial transcriptional remodeling, epigenetic factors are also key for the proper repression of the somatic program[207]. For example, the forced expression of OSKM leads to early induction of *Sap30*, a component of the SIN3A/HDAC2 repressive complex[207] recently demonstrated to interact with NANOG[235]. Here, NANOG-dependent SIN3A/HDAC2 recruitment leads to the deacetylation of reprogramming barriers, which finally turns in their genetic repression enhancing the reprogramming efficiency[235].

Somatic cellular metabolism heavily relies on the oxidative phosphorylation (OxPhos) pathway, whereas pluripotent cells are highly glycolytic (for more details, see section **4.3.4**). Metabolic changes are also observed before the activation of the pluripotency transcriptional program[236], evidencing the requirement for a metabolic shift during the initiation phase of somatic cell reprogramming. Studies have demonstrated that cells showing an increased glycolytic rate are more efficient in the acquisition of the pluripotent program[237], suggestive of the existence of a "metabolic memory". Additionally, it was demonstrated that c-MYC, one of the factors conforming the Yamanaka's reprogramming cocktail, directly regulates the expression levels of some metabolic genes[236], indicative of an interplay between the transcriptional and metabolic profile of cells. The absence of this gene in the reprogramming cocktail does not impair the metabolic shift occurring during the generation of iPSCs[236], suggesting that this process can also be achieved by alternative pathways. Chemical inhibition of key enzymes regulating the glycolytic pathway in cells by using small molecules has resulted in the impediment of an efficient iPSCs generation[236-238]. In contrast, chemical activation of glycolysis or upregulation of the glycolytic gene expression was demonstrated to accelerate the reprogramming of somatic cells[172,236,239]. Taken together,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

these findings indicate the relevance of this metabolic shift during the process.

4.4.3.2. Activation of the pluripotency program

After the silencing of the somatic program, a second transcriptional wave takes place[224] characterized by a global activation of the pluripotency program. This activation is known to occur in a stochastic fashion, leading to a "bottleneck" effect during the acquisition of pluripotency[207] (**Figure I17**). Reprogramming barriers might lead to a decrease or total absence of reprogramming factors to genomic regions responsible for the activation of the pluripotency transcriptional program, therefore reducing the efficiency of this process[207].

Several studies have demonstrated that the early binding of reprogramming factors (i.e., OSKM) preferentially occurs at regulatory regions[67,227]. These pioneering factors act by facilitating the posterior binding of factors and co-factors leading to a more open chromatin structure[68] for the proper activation of the pluripotency program[67,227]. The ability of these pioneering factors to bind to chromatin depends on their capacity to partially recognize specific chromatin motifs[67,68]. In this sense, the epigenome might confer a permissive or refractory state for the binding of pioneering factors. Bisulfite sequencing analyses demonstrated that global DNA hypomethylation is required at the promoter regions of pluripotency genes for their transcriptional activation[203-205]. Accordingly, overexpression of TET proteins increases the reprogramming efficiency in a cellular context-dependent manner[207,240]. Numerous pieces of evidence have highlighted the thigh interplay between epigenetics and chromatin accessibility and configuration for transcriptional regulation (reviewed in REF 241). For example, it is known that changes in the deposition of activating and repressing histone marks are required for efficient reprogramming to pluripotency. Hence, genetic manipulation for the induction of active histone marks, such as H3K4me3 at specific promoters, or silencing of

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

methyltransferases depositing the H3K9me3 repressive mark translates into decondensed chromatin conforming a permissive environment for transcriptional rewiring[242-244]. However, knowledge of the specific molecular mechanisms governing chromatin configuration and how this affects cell fate decisions is still missing.

4.5 ZINC FINGER PROTEINS

4.5.1 Structural classification and molecular functions

Zinc finger proteins (ZNFs) are one of the most abundant groups of proteins characterized by their structural self-stabilization through a zinc ion[245] and their ability to recognize and bind to specific DNA sequences[245]. Structurally, ZNFs can be classified into two different groups depending on the specific amino acids conforming the zinc finger domain. Thus, classical ZNFs show a C2H2 configuration, so two cysteine residues from one aminoacidic chain bind to two histidine residues from another chain through a zinc ion[246] (**Figure I19**). In contrast, alternative configurations that differ in the cysteine/histidine composition are common in non-classical ZNFs, giving place to C2-H2, C2-CH, or C2-C2 domains[246]. Nowadays, 30 different types of ZNFs exist based on their domain structure[247] (**Table I4**), which allows the recognition of several different DNA sequences and a wide range of molecular functions[248,249]. Classical C2H2 ZNFs, for example, mainly act as transcription factors regulating the expression levels of genes[250], whereas non-classical ZFNs show a great amount of functional variety, ranging from ligase activity to epigenetic remodeling or the regulation of cytoskeleton interactions and focal adhesions[246,251-253]. In addition, researchers have developed a very useful tool based on the ability of these proteins in recognizing specific DNA sequences, generating zinc finger-containing constructs fused to sequence-independent endonucleases for genome editing[254,255]. This technology has later evolved in precise methodologies for genomic engineering such as the CRISPR/Cas9 technology, with applications in regenerative medicine.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

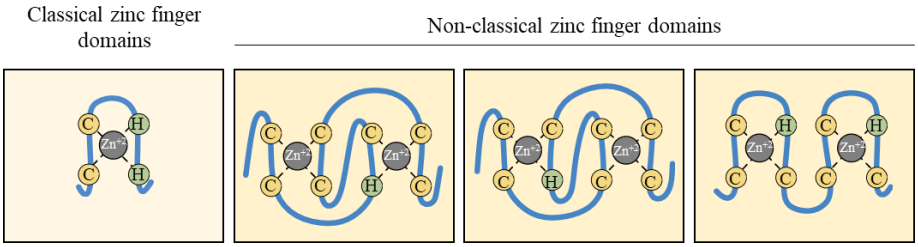


Figure I19. Structure of the main zinc finger domains. Schematic representation of classical and non-classical zinc finger domains. The basic zinc finger domain of C2H2-type ZNFs consists of two cysteine residues that interact with two histidines through a zinc ion. Non-classical ZNFs possess alternative structures in their zinc finger domains with different cysteine/histidine compositions.

PROTEIN FAMILY	STRUCTURE	MEMBERS
Zinc fingers C2H2-type (ZNF)	C-x-C-x-H-x-H	720
Ring finger proteins (RFN)	C-x-C-x-C-x-H-xxx-C-x-C-x-C-x-C	275
PHD finger proteins (PHF)	C-x-C-x-C-x-C-xxx-H-x-C-x-C-x-C	90
LIM domain-containing	C-x-C-x-H-x-C-x-C-x-C-x-C-(CHD)	53
Nuclear hormone receptors (NR)	C-x-C-x-C-x-C-xxx-C-x-C-x-C	50
Zinc fingers CCCH-type (ZC3H)	C-x-C-x-C-x-H	35
Zinc fingers FYVE-type (ZFYVE)	C-x-C-x-C-x-C-xxx-C-x-C-x-C	31
Zinc fingers CCHC-type (ZCCHC)	C-x-C-x-H-x-C	25
Zinc fingers DHHC-type (ZDHHC)	C-x-C-x-H-x-C-xxx-C-x-C-x-H-x-C	24
Zinc fingers MYND-type (ZMYND)	C-x-C-x-C-x-C-xxx-C-x-C-x-H-x-C	21
Zinc fingers RANBP2-type (ZRANB)	C-x-C-x-C-x-C	21
Zinc fingers ZZ-type (ZZZ)	C-x-C-x-C-x-C	18
Zinc fingers C2HC-type (ZC2CH)	C-x-C-x-H-x-C	16
GATA zinc finger domain-containing (GATAD)	C-x-C-x-C-x-C	15
ZF class homeoboxes and pseudogenes	C-x-C-x-H-x-H	15
THAP domain containing (THAP)	C-x-C-x-C-x-H	12
Zinc fingers CXXC-type (CXXC)	C-x-C-x-C-x-C-xxx- C-x-C-x-C-x-C	12
Zinc fingers SWIM-type (ZSWIM)	C-x-C-x-C-x-H	9
Zinc fingers AN1-type (ZFAND)	C-x-C-x-C-x-C-xxx-C-x-H-x-H-x-C	8
Zinc fingers 3CxxC-type (Z3CXXC)	C-x-C-x-H-x-C	8
Zinc fingers CW-type (ZCW)	C-x-C-x-C-x-C	7
Zinc fingers GRF-type (ZGRF)	C-x-C-x-C-x-C	7
Zinc fingers MIZ-type (ZMIZ)	C-x-C-x-H-x-C	7
Zinc fingers BED-type (ZBED)	C-x-C-x-H-x-H	6
Zinc fingers HIT/type (ZNHIT)	C-x-C-x-C-x-C-xxx-C-x-C-x-H-x-C	6
Zinc fingers MYM-type (ZMYM)	C-x-C-x-C-x-C	6
Zinc fingers matrin-type (ZMAT)	C-x-C-x-H-x-H	5

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Zinc fingers C2H2C-type	C-x-C-x-H-x-H	3
Zinc fingers DBF-type (ZDBF)	C-x-C-x-H-x-H	3
Zinc fingers PARP-type	C-x-C-x-H-x-C	2

Table I4. Types of zinc finger proteins. Table showing the classification of zinc finger protein families depending on their zinc finger domains. Additionally, the structure of the defined zinc finger domains is shown, along with the number of members conforming the corresponding protein family.

4.5.2 Zinc-finger proteins in pluripotency

Cell proliferation and differentiation are crucial for maintaining the stem cell homeostasis[256,257] and several zinc-finger proteins (ZNFs) have been implicated in the regulation of these processes[258-274]. Classically, specific ZNF proteins were demonstrated to play an important role in the maintenance of the ESC identity. Thus, for example, *Klf4*-null mice show defects during early development leading to premature death soon after birth[275], suggestive of a possible role of KLF4 in regulating embryo development and pluripotency. Remarkably, this protein was included among Yamanaka's reprogramming factors[82] (i.e., OCT4, SOX2, KLF4, and c-MYC; for more details, see section **4.4.2**). Moreover, a specific role for KLF4 in ESC self-renewal was evidenced by gain-of-function experiments during embryoid body formation[276]. Here, *Klf4*-overexpressing cells maintained increased levels of *Oct4* impeding their proper differentiation under specific signals[276]. In addition, tight cooperation among KLF4, OCT4, and SOX2 was demonstrated in pluripotent cells[277], and a KLF circuitry regulating the pluripotent transcriptional program was also reported[83]. LIN28 and SALL4, known zinc finger-containing proteins, were also implicated in the acquisition of pluripotency, bypassing the requirement of the presence of OCT4 and KLF4 in the cocktail of reprogramming factors during the generation of iPSCs[201,202] (for more details, see section **4.4.2**). Proteins from the GATA-type family were also widely characterized in the context of pluripotency. Thus, it is known that GATA4 directly blocks the expression of endogenous *Nanog* when co-transfected with the Yamanaka factors impeding the proper reprogramming of cells[278]. In

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

addition, overexpression of this factor in ESCs promotes their differentiation towards extraembryonic endoderm[279]. However, recent reports showed that the forced expression of any of the GATA family members (i.e., GATA1, GATA2, GATA3, GATA4, GATA5, and GATA6) can act as reprogramming enhancers[280,281], being able to replace OCT4 in the reprogramming cocktail[280]. These opposing effects of members of the GATA-type protein family on the pluripotent acquisition and exit suggest the wide range of molecular functions that these factors might exert, and more studies are needed to decipher the exact molecular implications of these ZNF factors in pluripotency.

Emerging evidence suggests that many other ZNFs could be involved in the regulation of pluripotency (examples of this were summarized in **Table I5**). Transcriptomic analyses of TFs during reprogramming of human somatic cells show that a total of 198 ZNFs are upregulated in the early stages of this process[282]. Furthermore, interactome analyses of pluripotency master regulators (i.e., OCT4, SOX2, and NANOG) have identified several ZNFs as their interacting partners[77-80], suggesting their potential role in the transcriptional regulation of the pluripotent cellular state. Among them, ZNFs can be found that act as epigenetic factors such as TET1 and TET3, two DNA demethylating enzymes containing CXXC-binding domains[283,284], which are known to play important roles in development and pluripotency (extensively reviewed in [124]). It was also shown that PRDM14, a protein containing the SET domain, is essential for naïve pluripotency in mouse ESCs[285]. In addition to their roles as epigenetic factors, ZNFs can also i) compete with epigenetic remodelers for the occupation of specific loci, or ii) act as cofactors by recruiting chromatin remodelers for regulation of their genomic targets. For example, ZFP281, a C2H2-type zinc finger protein, was found to be upregulated in undifferentiated cells compared to adult cells[286]. Furthermore, the interactome analyses showed that this transcription interacts with NANOG[287], OCT4 and SOX2[286], and c-MYC[288]. Although ZFP281 was shown to be dispensable for the maintenance of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ESCs[289], a key role of this protein in the regulation of alternative pluripotent states was reported[86]. In this regard, the direct interaction between ZFP281 and TET1 is essential for the establishment and maintenance of primed pluripotency, while its interaction with HDAC2 represses the expression of the naïve regulator TET2[86]. Another example, YY1, a GLI-Krüpple zinc finger protein, was reported to be essential in maintaining the self-renewal of ESC[290]. YY1 interacts directly with components of the BAF complex and together co-occupy promoters and super-enhancers important for cell proliferation and pluripotency[290]. ZNFs may also have important functions in epitranscriptomic networks for the governance of pluripotency. For example, ZFP217, another C2H3-type zinc finger protein, interacts with METLL3 and prevents the deposition of m^6A in RNA transcripts including pluripotent ones such as *Nanog* and *Sox2*[291]. ZFP217 also participates in the regulation of chromatin gene expression, and its absence compromises the maintenance of ESCs and the formation of iPSCs[291].

PROTEIN	FUNCTION	REFERENCE
ZFP206	ESC maintenance	292,293
ZFP42	ESC maintenance	294
ZFP217	ESC maintenance	291
YY1	ESC self-renewal	290
TET1	Primed pluripotency	295-300
TET3	Early development and differentiation	301
KLF2	Naïve pluripotency	89
KLF5	Naïve pluripotency	302
PRDM14	Naïve pluripotency	285
ZFP281	Primed pluripotency	86
GATA1/6	Reprogramming facilitator	280,281
LIN28	Reprogramming facilitator	201
SALL4	Reprogramming facilitator	202
KLF4	Reprogramming facilitator and ESC maintenance	82,83

Table I5. Role of specific ZNFs in the context of pluripotency. Examples of relevant ZNFs involved in the regulation of pluripotency.

Although studies of several ZNFs have been published in the context of pluripotency, knowledge of the specific molecular functions of most ZNFs remains obscure due to the large number of members that make up this family of proteins. Therefore, further studies are needed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

to analyze the mechanisms by which ZNFs control and regulate cell identity and homeostasis in the context of pluripotency.

4.5.3 Zinc finger MYM-type containing proteins

More than 20 years ago it was shown that a mutation in the "without children" (*woc*) protein of *Drosophila melanogaster* produces a sterile phenotype[303]. In addition, mutant flies also showed defects during early larvae development and alterations in their metabolism[303-305]. Structural characterization of this protein identified the presence of four zinc finger domains of the C2-C2 type, which have a high level of homology with the human form of zinc finger MYM-type containing proteins ZMYM2, ZMYM3, and ZMYM4[303].

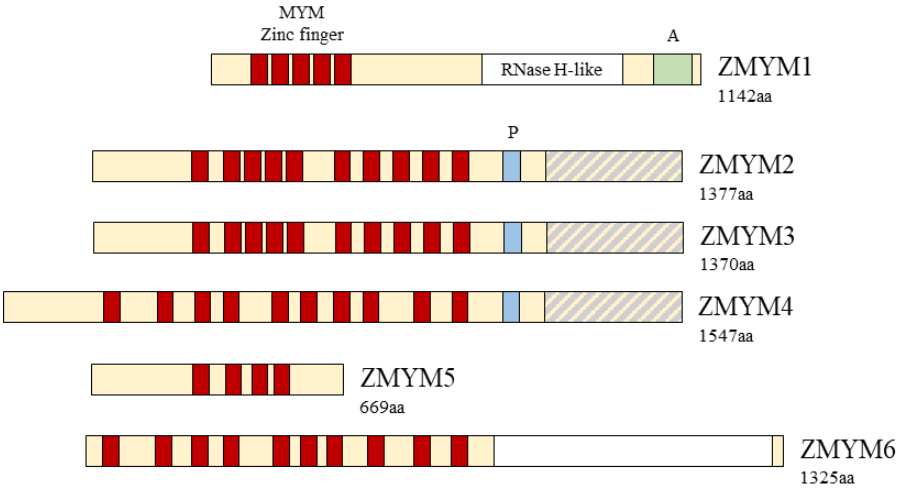


Figure I20. Family of zinc finger MYM-type containing proteins. Schematic representation of the conserved members of the MYM-type zinc finger containing proteins family. A, AT-hook DNA binding motifs; P, proline repeat domains. The hatched part at the C-terminal domain represents a highly conserved region.

The zinc finger MYM-type containing proteins family consist of six members (i.e., ZMYM1, ZMYM2, ZMYM3, ZMYM4, ZMYM5 and ZMYM6), which show a high degree of homology in their structure with the exception of ZMYM5[306] (**Figure I20**). The mapping of their nucleotide sequences facilitated their functional characterization as

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

transcription factors due to their ability to recognize and bind to specific DNA sequences[307].

Both ZMYM2 and ZMYM3 are the most characterized members of this family of proteins, and their alterations are linked to several syndromes and diseases[307-309]. ZMYM3 was identified as a candidate gene involved in X-linked intellectual disabilities (XLID)[307,308]. Furthermore, ZMYM3 was also found to be one of the three main genes associated with the progression of Alzheimer's disease[310]. ZMYM2 mutations were found to play a crucial role in the development of myeloproliferative syndromes[309,311,312]. Therefore, *ZMYM2*, located in chromosome 13, can be translocated to chromosome 8, where the fibroblast growth factor receptor 1 (FGFR1) gene is found[309]. This translocation, termed as t(8;13)(p11;q11-12), produces an 87kD ZMYM2-FGFR1 fused protein that is located in the cytoplasm[309]. Patients afflicted with this specific mutation develop rare leukemia/lymphoma syndrome with lymphoblastic and myeloproliferative disorders[309]. Two alternative explanations were proposed for the evolution of this syndrome. First, a constitutively active form of the FGFR1 receptor can lead to an uncontrolled proliferation of hematopoietic stem cells[311,312], and second, heterodimerization between wild-type ZMYM2 and the fused ZMYM2-FGFR1 might impede its normal function in cells[309]. Remarkably, a promising treatment for this syndrome has been proposed that leads to degradation of the fused protein[313].

Since last year, a growing number of infectious disease research has been published due to the COVID situation affecting our lives. Surprisingly, a direct relation between miRNAs and ZMYM5 has been evidenced[314]. In this study, COVID patients showed downregulation of specific miRNAs (miR-31, miR-29, miR-126, and miR-17), resulting in the upregulation of their target mRNAs[314]. Among them, the expression levels of *ZMYM5* were related to the aggressiveness of the infection[314], highlighting a possible role of the ZMYM proteins in the inflammatory response.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Collectively, mutations in members of the ZMYM family have been linked to pathological alterations in humans, thus affecting the onset, progression, and aggressiveness of various diseases. However, studies on the role of these proteins in physiological conditions are lacking and the implications of their specific molecular mechanisms are still unknown.

4.5.4 Zinc finger MYM-type containing 2 (ZMYM2)

As mentioned above, ZMYM2 mutations have been widely linked to myeloproliferative syndromes[309,311,312]. In recent years some studies have been published that also describe the molecular functions of ZMYM2 under physiological conditions. Here, I will summarize some of these relevant discoveries.

Circular RNAs (circRNAs) are RNA structures known to play a role during tumorigenesis[315,316]. Aberrant overexpression of circZMYM2 was found in pancreatic cancer cells, promoting their proliferation and invasiveness of non-tumor tissues[317]. circZMYM2 acts as a sponge of miR-335-5p, thus preventing the posttranscriptional repression of *Jmjdc2* and indirectly affecting the epigenetic state of tumor cells[317].

Other studies have reported the involvement of ZMYM2 in the integration of epigenetic and transcriptional signals[318,319]. BHC110, a component of the BRAF35-HDAC complex (a.k.a. BHC complex), was shown to interact with ZMYM2 in human cells[318]. This interaction leads to the formation of a complex, termed XFIM, which showed HDAC activity towards core histones[318]. Specifically, the XFIM complex is recruited into the c-*fos* promoter in an EGF-dependent manner, resulting in its transcriptional repression[318]. The LCH complex, formed by the direct interaction of LSD1-CoREST-HDAC1 proteins, was also shown to associate with ZMYM2[319]. Indeed, ZMYM2 interacts with an intact LCH complex but not with its subunits, thus acting as a scaffolding protein for its stabilization and allowing its molecular

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

function as a transcriptional repressor of neural genes[319]. Furthermore, both ZMYM1 and ZMYM3 were also shown to interact directly with this histone modifier complex[306], suggesting a possible redundant molecular role of MYM-type proteins in epigenome regulation.

Post-translational modifications (PTMs) of proteins consist of reversible and irreversible covalent modifications that affect their molecular functions[320,321]. For example, SUMOylation involves the covalent attachment of SUMO (small ubiquitin-like modifier) to a lysine residue on the target protein[322,323]. This chemical modification occurs in a sequential manner that resembles the ubiquitination of proteins and is known to alter the physical properties of modified proteins, ultimately affecting their interaction properties[322,323]. ZMYM2 was shown to be SUMOylated by direct interaction with members of the SUMO protein family such as SUMO-1 and SUMO-2[324,325]. SUMOylation of ZMYM2 in non-tumor cells alters its ability to associate with other proteins and leads to cell transformation[324].

Although the role of ZMYM2 in pluripotency and reprogramming is largely unknown, previous studies reported that ZMYM2 can interact with NANOG and OCT4 in ESCs[77,287]. Furthermore, an RNAi-based genome-wide screening identified ZMYM2 as a possible reprogramming barrier during the generation of human iPSCs[326]. Taken together, it is tempting to speculate that ZMYM2 may play important roles in pluripotency and reprogramming. To this end, in this work, I will combine several classical approaches in the pluripotent field together with new methodologies to decipher the molecular mechanisms by which ZMYM2 could be controlling pluripotent cell identity. It is noteworthy that in the course of this thesis work, several studies on this topic have been published and will be discussed concerning my results in the Discussion section.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

5. AIMS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The basis of pluripotency resides in conserved transcriptional regulatory and protein interacting networks of numerous transcription factors and epigenetic regulators. The zinc finger protein ZMYM2 was previously characterized as an interacting partner of OCT4 and NANOG, which are well-known key components of the core regulatory network that governs pluripotency and somatic cell reprogramming. With this in mind, this thesis project will be focused on testing the hypothesis that *ZMYM2 could play important regulatory activities in maintaining pluripotency as well as differentiating and obtaining pluripotent cells.* Therefore, the main goal of my work will be to address the functional significance of ZMYM2 in the processes of self-renewal and differentiation of pluripotent stem cells as well as in the reprogramming of somatic cells to pluripotency. Using mouse pluripotency as a well-defined and tractable model system, this thesis project will pursue the next three specific aims:

Aim 1. Dissecting the impact of ZMYM2 in the maintenance and differentiation of pluripotent stem cells.

Aim 2. Elucidating the contribution of ZMYM2 to somatic cell reprogramming.

Aim 3. Determining the molecular mechanisms underlying the functional importance of ZMYM2 in pluripotency.

I hope that the fundamental insights that will be gained in my thesis can contribute to a better understating of the basis of pluripotency and represent new angles to overcome the current challenges such as heterogeneity and tumorigenicity of ESCs and iPSCs for therapeutic applications. Furthermore, the expected results from my study will provide novel insights into the mechanism of ZMYM2 that could be relevant in other contexts including development and disease.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6. MATERIAL AND METHODS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6.1 REAGENTS AND KITS

All the reagents used in this thesis were of research grade and are listed in **Supplementary Table 1**. The specific concentrations for each reagent are indicated in their respective section. Kits that were used in my thesis are listed in **Supplementary Table 2**.

6.2 EXPERIMENTAL MODELS

Cell lines used in this thesis are listed in the next Table:

Experimental Models: Cell Lines		
Name	Source	Identifier
CJ7 (mouse ESCs)	Dr. Jianlong Wang (Columbia Univ., USA)	n/a
CCE (mouse ESCs)	Dr. Jianlong Wang (Columbia Univ., USA)	n/a
Zmym2$^{+/+}$ (mouse ESCs)	Dr. Jianlong Wang (Columbia Univ., USA)	n/a
Zmym2$^{GT/GT}$ (mouse ESCs)	Dr. Jianlong Wang (Columbia Univ., USA)	n/a
Zmym2 WT (mouse ESCs)	This study	n/a
Zmym2 KO (mouse ESCs)	This study	n/a
Zmym2$^{GT/GT}$ + ^{3xFL}*Zmym2* (mouse ESCs)	This study	n/a
TNGA (mouse pre-iPSCs)	Dr. Jose C.R. Silva (Cambridge Univ., UK)	n/a
O4G MEFs (mouse fibroblasts)	Dr. Jianlong Wang (Columbia Univ., USA)	n/a
hB (human B cells)	Dr. Carlos-Filipe Pereira (Lund Univ., Sweden)	n/a
HEK293T	ATCC	CRL-3216
Plat-E	Cell Biolabs	RV-101

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CJ7 ESCs: mouse ESC line derived from the ICM of 129/SvlmJ male embryos.

CCE ESCs: mouse ESC line derived from the ICM of 129/Sv male embryos.

***Zmym2* wild-type (WT) and mutant ESC lines**: *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ are ESCs derived from the intercrosses of heterozygous (*Zmym2*$^{GT/+}$) mice. GT refers to the gene-trap *Zmym2* strategy. *Zmym2* WT and *Zmym2* KO are specific subclones obtained from the CRISPR-Cas9 knockout strategy using CCE ESCs.

O4G MEFs: primary mouse embryonic fibroblasts (MEFs) derived from E13.5 mouse embryos (mixed C57BL/6:129S4Sv/Jae genetic background) harboring an IRES-EGFP fusion cassette downstream of the stop codon of the *Oct4* (*Pou5f1*) gene.

TNGA pre-iPSCs: clonal line of MEF-proliferative reprogramming intermediates (pre-iPSCs), which harbor an IRES-EGFP fusion cassette downstream of the stop codon of the *Nanog* gene, obtained from pMX-based retroviral reprogramming factors OCT4, SOX2, KLF4, and c-MYC (OSKM) and *Nanog*-GFP negative cell sorted cells.

hB cells: human B lymphocytes transformed with Epstein-Barr virus (EBV). In this thesis, these proliferative somatic cells were used in cell fusion assays.

HEK293T cells: immortalized human embryonic kidney cell line, containing SV40 T-antigen that facilitates propagation of vectors bearing the SV40 replication region. These cells were used for the generation of lentiviral particles containing shRNAs using pLKO vectors.

Plat-E cells: the platinum retroviral packaging cell line is based on the HEK293T cell line, bearing *gag*, *pol*, and *env* genes allowing retroviral packaging with a single plasmid transfection. I used this cell line for the generation of retroviral particles using pMX vectors.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.3 CELL CULTURE

Feeder-free mouse ESCs and TNGA pre-iPSCs were cultured on 0.1% gelatin-coated plates under pluripotent standard conditions [high-glucose Dulbecco's Modified Eagle Medium (DMEM; Corning, 10-013-CV), supplemented with 15% heat-inactivated fetal bovine serum (FBS; Gibco, 10270106), 100U/mL penicillin/streptomycin (Gibco, 15070-063), 1% nucleoside mix [uridine (U; Sigma-Aldrich, U3003), adenosine (A; Sigma-Aldrich, A4036), guanosine (G; Sigma-Aldrich, G6264), cytidine (C; Sigma-Aldrich, C4654)], 200mM L-glutamine (Gibco, 25030-081), 100mM non-essential amino acids (NEAA; Gibco, 11140-050), 0.1mM beta-mercaptoethanol (Sigma-Aldrich, M-6250) and 10^3U/mL Leukemia Inhibitory Factor (LIF; homemade)].

Gelatin-coated plates were prepared using 0.1% (w/v) gelatin (Sigma-Aldrich, G9391) in milli-Q water. The surface of the plastic plates was completely covered and incubated for at least 20–30 min in a 37ºC cell incubator. After incubation, the gelatin was aspirated from the plate before adding the cells.

O4G MEFs, HEK293T, and Plat-E cells were cultured directly on plastic plates and maintained in culture at 70%–80% confluency to avoid growth inhibition by cell contact. These were cultured in somatic medium [DMEM (Corning, 10-013-CV) supplemented with 10% heat-inactivated FBS (Gibco, 10270106), 200mM L-glutamine (Gibco, 25030-081), and 100U/mL penicillin/streptomycin (Gibco, 15070-063)].

Human hB lymphocytes used for cell fusion assays were cultured in suspension using 175cm^2 tissue culture flasks (Falcon, 353112) with hB medium [RPMI-1640 (Sigma-Aldrich, R0883) with 10% heat-inactivated FBS (Gibco, 10270106), 2mM L-glutamine (Gibco, 25030-081), 50U/mL penicillin and 50µg/mL streptomycin (Gibco, 15070-063)].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

For adherent cultures, cell detachment was performed with trypsin-EDTA solution (Sigma-Aldrich, T3924) or accutase solution (Sigma-Aldrich, A6964). Suspension cultures were passaged by diluting some cells in a larger volume of fresh media. Cell culture was performed in a sterile tissue culture hood under aseptic conditions and cells were maintained in a humidified 37ºC incubator with 5% CO_2.

For long-term preservation of cells, they were harvested and pelleted by centrifugation at 300g for 5 min at room temperature (RT). Then, the cells were resuspended in 1X freezing solution [90% proper cell type culture medium and 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich, 75-21-8)] aliquoted in cryovials and maintained at -80ºC for a maximum of 48 h. Finally, frozen cells were transferred to tanks containing liquid nitrogen for long-term storage.

6.4 CHEMICAL INHIBITION OF PADI4 ACTIVITY

Two different reversible chemical inhibitors were tested to inhibit PADI4 enzymatic activity (see below). Each chemical was tested at different concentrations based on efficiency to reduce global citrullination at histone 3 by western blot assay (see section **6.24**) using a specific antibody (**Supplementary Table 5**).

CHEMICAL INHIBITORS		
INHIBITOR	TARGET	SOURCE/REFERENCE
GSK199	PADI4	Cayman, CAY-17489
GSK484	PADI4	Cayman, CAY-17488

6.5 CLONING OF PLASMID VECTORS

6.5.1 Cloning protein-coding regions into PiggyBac transposon system

To investigate the consequences of ZMYM2 gain-of-function and transgenic rescue in this thesis work, a PiggyBac (PB) transposon system was used as a tool for transgenesis[327]. Briefly, the coding DNA sequence (CDS) of the *Zmym2* gene was obtained by reverse

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

transcription (see section **6.23**) of mouse ESC RNA using specific oligonucleotides that contain the restriction endonuclease recognition sites of *NheI* (Thermo Scientific, 10755851) and *PacI* (Thermo Scientific, 10715271). Next, the CDS of *Zmym2* was cloned into pJET1.2 vector using the CloneJet PCR cloning kit (Thermo Scientific, K1231) following the manufacturer's recommendations, and validated by SANGER sequencing. Then, the backbone vector (pPB-CAG-3xFLAG-pA-pgk-hph; Addgene, #48753) carrying the ampicillin and hygromycin resistant genes for bacterial and mammalian selection, respectively, and the pJET1.2-*Zmym2* construct were digested with *NheI* and *PacI* restriction enzymes. Finally, a ligation reaction using the Rapid DNA Dephos & Ligation kit (Roche, 4898117001) following the manufacturer's recommendations was performed between the linearized PB vector and the isolated DNA fragment containing *Zmym2* CDS. The final construct named PB-flox-^{3xFl}*Zmym2*-Pgk-Hygro was validated by SANGER sequencing and protein expression by Western blot assay (see section **6.24**).

To investigate the consequences of SUMOylation in the ZMYM2 protein on the transcriptional control of the *Padi4* gene (see Results section **7.4.10**), a strategy similar to the previous one was carried out to subclone the human ZMYM2 ortholog that contains specific mutations in different SIM motifs and the WT version from pCS2 vectors (kindly provided by Dr. Andrew Sharrows, Manchester University, UK) into a pPB-CAG-3xFLAG-pA-pgk-hph plasmid.

The primer sequences for CDS amplification used in this thesis are provided in **Supplementary Table 6**.

6.5.2 Cloning shRNA oligos into pLKO lentiviral vector

The specific shRNAs for knockdown of *Zmym2*, *Lsd1*, *Zfp281*, *Hdac1*, *Hdac2* and one control shRNA against *Luciferase* were performed as previously described[328]. Briefly, specific oligonucleotides containing the target sequences of the shRNAs were designed,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

synthesized, and subcloned into pLKO.1 (Addgene, #32682) vector that contains the ampicillin and hygromycin resistant genes for bacterial and mammalian selection, respectively. Constructs were then introduced into *Stbl3* competent *E. coli* cells following a heat shock-based protocol. Next, cells were plated onto ampicillin containing LB agar plates and incubated in a 37ºC incubator at 225 rpm for 16 h. The initial identification of bacterial colonies of pLKO containing the shRNA sequences was performed by PCR following a strategy recently published by us[329]. Briefly, a PCR reaction mix was prepared using the reagents listed below, and colonies were picked and resuspended in the PCR reaction mix.

Reagent	SOURCE/REFERENCE
Forward oligo [20µM]	n/a
Reverse oligo [20µM]	n/a
10X NEBuffer 2	New England BioLabs, B7002S

Then, PCR was performed using a thermocycler and the following program: initial denaturing 2 min at 95ºC, 30 cycles of amplification (95ºC, 30 sec; 60ºC, 30 sec; 68ºC, 30 sec) and a final extension 5 min 68ºC. Next, PCR amplicons were analyzed by electrophoresis in 2.5% (w/v) agarose with Synergel™ (Diversified Biotech, DIVESYN-100) gel and visualized using the BioRad Gel Doc™ XR+ system. Finally, positive pLKO-shRNA constructs were validated by SANGER sequencing and knockdown efficiency using RT-qPCR (see section **6.23**).

The primer sequences containing the target sequences of the shRNAs used in this thesis are provided in **Supplementary Table 6**.

6.5.3 Cloning the *Padi4* genomic regulatory region into a *Luciferase* reporter

To investigate the repressive effect exerted by ZMYM2 on the expression levels of the *Padi4* gene, a *Luciferase* reporter assay using a pGL3 vector was performed. Briefly, the CDS of the promoter region of *Padi4* gene was obtained by reverse transcription (see section **6.23**)

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

of mouse ESC RNA using specific oligonucleotides (**Supplementary Table 6**) and a thermocycler following the next amplification conditions: initial denaturing 30sec at 98ºC, 5 cycles of initial amplification (98ºC, 10 sec; 63ºC, 10 sec; 72ºC, 66 sec), 25 cycles of amplification (98ºC, 10 sec; 65ºC, 10 sec; 72ºC, 66 sec), and a final extension during 5min at 72ºC. Next, samples were loaded in a 2.5% (w/v) agarose TAE gel with Synergel, and the amplicon was visualized with the SafeVIEW-Mini2 LED transilluminator (Fisherbrand, 15502506) system and extracted using the GenElute Gel Extraction Kit (Sigma-Aldrich, NA1111). Then, the CDS containing the promoter region of *Padi4* was cloned into the pJET1.2 vector using the CloneJet PCR cloning kit (Thermo Scientific, K1231) following the manufacturer's recommendations. The constructs were introduced into DH5α competent *E. coli* cells following a protocol based on heat shock and then the cells were plated on LB agar plates containing kanamycin and shaken at 225rpm for 16h at 37ºC. The initial identification of bacterial colonies containing the pJET.*Padi4*-promoter construct was performed by PCR with specific oligonucleotides (**Supplementary Table 6**) in a thermocycler following the next amplification conditions: initial denaturing 30sec at 98ºC, 30 cycles of amplification (98ºC, 10 sec; 64ºC, 30 sec; 72ºC, 1 min), and a final extension during 5 min at 72ºC. Next, samples were loaded in a 2.5% agarose TAE gel with Synergel, and amplicon sizes were visualized. Positive pJET.*Padi4*-promoter constructs were validated by SANGER sequencing. Both the backbone vector (pGL3; Addgene, #64784) carrying the ampicillin resistant gene for bacterial selection and the pJET.*Padi4*-promoter construct was digested with *XhoI* (Thermo Scientific, 10880041) and *XbaI* (Thermo Scientific, 10880591) restriction enzymes, and linearized plasmids were loaded in a 0.7% (w/v) agarose TAE gel. Specific bands corresponding with the linearized pGL3 vector and the *Padi4* promoter were extracted using the GenElute Gel Extraction Kit following the manufacturer's instructions. Finally, a ligation reaction using the Rapid DNA Dephos & Ligation kit (Roche, 4898117001) following the manufacturer's recommendations was performed. The

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

final construct named pGL3.*Padi4*-promoter was validated by SANGER sequencing.

6.6 ISOLATION OF PLASMID DNA

6.6.1 Bacterial competent cells production

To generate competent bacterial cells for plasmid transformation, we follow a previously reported protocol ["*Molecular Cloning 3rd edition*", by Sambrook y Russell (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2001)] with some modifications. First, DH5α and *Stabl3 Escherichia coli* strains to clone non-viral and viral constructs, respectively, were incubated on Luria Broth (LB)-agar plate at 37ºC and 225rpm for 16-20 h. Second, 5-7mL of LB medium was inoculated with a selected colony and incubated overnight at 37ºC in a shaker at 225rpm. Third, 0.5, 1, and 2mL of growth were used to inoculate 100mL of fresh LB medium and incubated at 37ºC until the absorbance at 600 nm was 0.55. From here, the selected growth was transferred to iced water for 10min and cells were collected by spinning them for 5 min at 2,500g, 4ºC in a 50mL Falcon tube. The supernatant was discarded and cells were resuspended in cold *Inoue* buffer (see recipe in the following Table) and then collected again by centrifugation (10 min at 2,500g, 4ºC). Another resuspension step was performed using the same iced-cold *Inoue* buffer, and DMSO (Sigma-Aldrich, 75-21-8) was added to a final concentration of 7%. Resuspended cells were then homogenized by swirling and chilled on ice for 10 min. Finally, homogenized cells were distributed in 100µL aliquots in 1.5mL microcentrifuge tubes, and flash-frozen in a cube of dry ice and 100% ethanol mix. For long-term storage, aliquots were maintained at -80ºC.

***Inoue* BUFFER RECIPE**		
SOLUTION	FINAL CONCENTRATION	SOURCE/REFERENCE
$MnCl_2 \cdot 4H_2O$	55mM	Thermo Scientific, 205891000
$CaCl_2 \cdot 2H_2O$	15mM	Fisher Chemical, VL2151000

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

KCl	250mM	Fisher Chemical, P/4240/65
PIPES (pH 6.7)	10mM	Sigma-Aldrich, P6757100G

6.6.2 Bacterial transformation

First, plasmid DNA was added to competent bacterial cells (10% v/v maximum) and homogenized. Second, the bacterial solution containing the plasmid was incubated on ice for 20 min and a heat shock was applied at 42ºC for 1 min or at 37ºC for 45 sec for *Stbl3* and DH5α strains, respectively, to promote entry of the plasmid into competent cells. Third, cells were incubated on ice for 5 min. Then, 400µL of S.O.C. (for 1L, add 20g Tryptone, 5g yeast extract, and 0.5g NaCl) were added, and cells were pre-incubated for 1 h at 37ºC. Finally, 100-200µL of cells were seeded on an antibiotic-containing LB-agar plate for growth selection and incubated at 37ºC for ~16 h.

6.6.3 Plasmid DNA extraction

A single transformed colony containing the recombinant plasmid was picked and inoculated into 8mL of liquid LB medium containing the appropriate antibiotic (see the table below) to select for the chosen plasmid. Then, the culture was grown overnight (16-18 h) at 37ºC and 225rpm. The next day, bacterial cultures were pelleted, and plasmid DNA was isolated using the GenElute™ Plasmid MiniPrep kit (Sigma-Aldrich, NA0150) following the manufacturer's recommendations. If larger amounts of plasmid DNA were needed, 100mL of growth was prepared and the PureLink™ HiPure Plasmid Filter MidiPrep kit (Invitrogen, K210015) was used for isolation following the manufacturer's recommendations. Finally, the plasmid DNA concentration and purity were determined using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, ND-2000).

ANTIBIOTIC	CONCENTRATION	SOURCE/REFERENCE
Ampicillin	50µg/mL	Fisher BioReagents, BP1760-25
Kanamycin	50µg/mL	Fisher, BioReagents, BP906-5

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.7 MAMMALIAN CELL TRANSFECTION

PiggyBac (PB) based plasmids for overexpression were co-transfected with pBASE transposase (kindly provided by Jose C. Silva, Cambridge University, UK) in pre-iPSCs and ESCs using the jetPRIME kit (Polyplus, 114-15). Briefly, fifty thousand cells were seeded in a 0.1% gelatin-coated 12-well plate (Corning, 3513) and incubated overnight (ON) under standard pluripotent culture conditions. The next day, the pluripotent standard medium was changed, and cells were transfected with 0.3µg of PB construct and 0.5µg of pBASE with transfection reagent following manufacturer's protocol. First, 0.8µg of recombinant DNA was added to 75µL of jetPRIME buffer in a 1.5mL tube, gently mixed, and centrifuged. Second, 1.6µL of jetPRIME reagent was added, homogenized, and incubated for 10 min at RT. Third, the transfection mix was added dropwise to the cells and gently mixed by swirling. Four hours after incubation, the medium was replaced with a fresh standard pluripotent medium and incubated ON. The following day transfected pre-iPSCs or ESCs were selected using 250µg/mL of Hygromycin (Corning, 30-240-CR). The medium with antibiotics was refreshed every day until a control with non-transfected cells showed 100% cell death.

A similar transfection protocol was used for genomic edition using CRISPR/Cas9 technology (see section **6.11**) and luciferase assays (see section **6.21**).

6.8 VIRAL PRODUCTION AND TRANSDUCTION

6.8.1 Generation of viral particles

Lentiviral pHAGE2-EF1aFull-hOct4-F2A-hKlf4-IRES-hSox2-P2A-hcMyc-W-loxP vector and pLKO shRNA-expressing vectors were co-transfected with psPAX2 and pMD2.G packaging vector into HEK293T cells. Retroviral pMX transgene-overexpressing constructs (i.e., pMX-Oct4, pMX-Sox2, pMX-Klf4, and pMX-cMyc) were transfected into Plat-E cells. Briefly, cells were plated at density of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$8x10^6$ per 150 mm dish (Corning, 430599) and maintained in somatic medium. Twenty-four hours later the medium was changed, and cells were transfected with PEI (Polyethylenimine; Sigma-Aldrich, 408727) and either 20µg of lentiviral plasmids along with 20µg of packaging mix (i.e., 10µg psPAX2 and 10µg pMD2.G) or 20µg of pMX vectors alone. Briefly, the transfection solution was prepared as follows: a) 16µL of 5X PEI was added to 1.6mL of 150mM NaCl solution; b) lentiviral or retroviral recombinant DNA was added to 1.6 mL of 150mM solution; c) PEI-containing solution was added dropwise to the DNA-containing solution, homogenized, and incubated 30 min at RT. Finally, transfection solution containing PEI and recombinant DNA was added dropwise to HEK293T cell or Plat-E cells. Eighteen hours later medium was changed, and virus-containing supernatants were collected 48 hours and 72 hours post-transfection and concentrated using Amicon ultra-centrifugal filter units with 0.22µm pore size (Millipore, UFC9030) by centrifugation 3,500g on a swinging bucket rotor during 30 min at 4ºC. For long-term storage, the concentrated supernatants were aliquoted and stored at -80ºC until use.

6.8.2 Mammalian cell transduction

Lentiviral (i.e., pHAGE2-EF1aFull-hOct4-F2A-hKlf4-IRES-hSox2-P2A-hcMyc-W-loxP) and retroviral infection (i.e., pMX-Oct4, pMX-Sox2, pMX-Klf4, and pMX-cMyc) of somatic cells were performed as previously described[330]. Briefly, the day before transduction, 5 x 10^4 fibroblasts were seeded per single well of a 6-well plate and incubated ON. The next day, the medium was replaced with a somatic medium containing 6µg/mL polybrene (Sigma Aldrich, H9268), and cells were infected with 100µL of concentrated viral particles. Twenty-four hours after infection, the somatic medium was changed. Lentiviral-mediated transduction for loss-of-function assays using pLKO.shRNAs in pre-iPSCs and ESCs were performed as we previously described[329].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.9 GENOMIC DNA ISOLATION

Genomic DNA (gDNA) was extracted and purified using the Centra Puregene Cell Kit (Qiagen, 158388) following the manufacturer's instructions. Briefly, 1 x 10^6 ESCs were resuspended in 200µL Cell Lysis Solution (Qiagen, 158908) and incubated for 1 hour at RT to disrupt cell aggregates. Then, 3µL RNAse A (1mg/mL) (Thermo Scientific, EN0531) was added, and samples were incubated for 1 hour at 37ºC. Next, a cooldown of resuspensions was performed by incubating samples on ice 5-10 min and a 200µL Protein Precipitation Solution (Qiagen, 158912) was added. Proteins were then precipitated by centrifuging samples at RT, 14,000 x g for 3 min, and 200µL isopropanol (Fisher Bioreagents, BP2618-1) was added. Homogenized gDNAs were then centrifuged again at RT, 14,000 x g for 5 min, and the supernatant was discarded. 70% ethanol was added to rinse the DNA pellets for 5 min. After this time, ethanol was discarded, and pellets were completely air-dried. Finally, 66µL DNA Hydration Solution (Qiagen, 158916) was added to pellets, and DNA was rehydrated by incubating samples at 65ºC for 1-2 h and then at RT overnight. The final concentration of gDNA samples was determined by spectrophotometry measuring in a Nanodrop 2000 Spectrophotometer (Thermo Scientific, ND-2000).

6.10 GENOTYPING PCR

The genotypes of *Zmym2* ESC lines were confirmed by Polymerase Chain Reaction (PCR) approach using genomic DNA. Briefly, genotyping PCR was conducted on the genomic DNA using a combination of three specific oligonucleotides (see **Figure R2**). Briefly, a forward primer was combined with two different reverse primers, each of which binds to the wild-type or knockout *Zmym2* allele and results in amplicons that differ in their final sizes. Then, genotyping was performed by following the indicated thermal cycler settings: an initial denaturing step of 2 min at 95ºC, 30 cycles of amplification (94ºC, 30 sec; 55ºC, 30 sec; 72ºC, 30 sec), and a final extension step for

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5 min at 72ºC. Finally, PCR amplicons were analyzed by electrophoresis in 2.5% (w/v) agarose with Synergel™ (Diversified Biotech, DIVESYN-100) gel and visualized using the BioRad Gel Doc™ XR+ system.

The primer sequences for genotyping PCR used in this thesis are provided in **Supplementary Table 6**.

6.11 Genomic edition using CRISPR/Cas9 technology

Zmym2 knockout (KO) ESCs were generated with CRISPR/Cas9 editing tool as previously described[331]. Briefly, two different sgRNAs targeting the first exon of the *Zmym2* gene (**Supplementary Table 6**) were designed using the webtool available at https://zlab.bio/guide-design-resources and were then cloned into the pSpCas9(BB)-2A-Puro (PX459) V2.0 vector (Addgene, #62988) carrying the puromycin resistant gene for mammalian selection. After verifying the correct sequences by SANGER sequencing, CCE ESCs were transfected with the sgRNA-containing plasmids using the jetPRIME kit following the manufacturer's instructions (see section **6.7**). Transfected cells were seeded at clonal density and selected with 2µg/mL puromycin (Fisher Bioreagents, BP2956-100) for 24 h to obtain single colonies for manual picking. Several clones were isolated, and the identity of a *Zmym2* KO cell line was confirmed by SANGER sequencing, genotyping (see section **6.10**), RT-qPCR (see section **6.23**), and western blot analysis (see section **6.24**).

6.12 Cell proliferation and viability assays

Cell counting was performed to investigate the function of ZMYM2 in the ability of pluripotent cells to proliferate under standard cultured conditions or primed pluripotent conversion (see section **6.14**). Briefly, cell detachment of pluripotent adherent cultures was performed with trypsin-EDTA solution (Sigma- Aldrich, T3924) or accutase solution (Sigma- Aldrich, A6964) for standard or primed conditions, respectively. The pluripotent cells were then resuspended in Dulbecco's

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

phosphate-buffered saline (DPBS; Corning, 21-031-CV) and an aliquot was counted using a Neubauer chamber (Hirschmann, 8100103) after the addition of trypan blue 0.4% solution (VWR Chemicals, K940) for analysis of cell proliferation of living cells. On the other hand, determination of cell viability was performed using an MTT assay[332]. Briefly, 5 x 10^3 pluripotent cells were seeded per single well of a 96-well plate (Corning, 3474). The next day, 10% of 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide; thiazolyl blue (MTT; Sigma-Aldrich, M5655) was added to the cells and incubated at 37ºC for 4 hours. Then, a 1:1 ratio isopropanol (Fisher Bioreagents, BP2618-1):DMSO (Sigma-Aldrich, 75-21-8) solution was prepared and added to the cells. Each sample was mixed by pipetting up and down. Then, the redox potential in viable cells promoted by the conversion of water-soluble MTT to insoluble formazan was determined by measuring the absorbance at 570 nm using a Mithras LB 940 microplate reader (Berthol Technologies).

6.13 CELL CYCLE ANALYSIS

For cell cycle analysis, an equal number of proliferating ESCs (e.g., 1x10^5 cells) cultured in standard pluripotent conditions were washed with 1X DPBS and permeabilized by using 0.1% Triton™ X-100 (v/v) (Sigma-Aldrich, G9391) in 1X DPBS. DNA content was stained with 10µM 4'-6-diamidine-2-phenylindol (DAPI; Sigma-Aldrich, D9542) at RT for 10min. The determination of the percentage of cells in specific phases of the cell cycle (e.g., G1, S, and G2-M) was performed by flow cytometry using an Accuri C6 instrument (BD Biosciences), and data were analyzed with FlowJo software (v7.6.1) using the Dean-Jett-Fox model.

6.14 PRIMED PLURIPOTENT CELL FATE CONVERSION ASSAY

To explore the role of ZMYM2 for conversion of mouse ESCs cultured in presence of serum and LIF (hereafter "SL") under standard pluripotent conditions (a.k.a. metastable pluripotent state) into primed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

pluripotent cells, ESC transgenic lines were cultured and subjected to primed conversion assays as previously described[86]. Briefly, 3 x 10^4 cells/cm^2 were seeded on fibronectin-coated plates (10 μg/mL/cm^2; Millipore, FC010) and reprogrammed using chemically defined primed culture conditions (hereafter "FA") [DMEM/F12 (Sigma-Aldrich, D6421) and Neurobasal medium (Gibco, 21103049) mixed at a ratio of 1:1, 1x B27 supplement (Gibco, 17504044), 1x N2 supplement (Gibco, 17502048), 2mM L-glutamine, 0.1mM 2-mercaptoethanol, and 50U/mL Penicillin/Streptomycin, supplemented with ACTIVIN A (20ng/mL; Invitrogen, PHC9564) and bFGF (12ng/mL; Invitrogen, PHG0026)]. FA medium was changed every 24 h.

6.15 ALKALINE-PHOSPHATASE STAINING

The presence of alkaline-phosphatase (AP) enzyme is typical of undifferentiated cells[333] and is therefore used routinely as a marker of pluripotency. AP staining was measured using an alkaline phosphatase kit (Sigma-Aldrich, 86R-1KT) following the manufacturer's recommendations. Briefly, both fixation and staining solutions were prepared as indicated in the table. Then, plates containing cells were washed with DPBS with Ca^{+2} and Mg^{+} (Sigma-Aldrich, D8662), and fixed by incubating them with the fixation solution for 30 seconds at RT. Next, fixed cells were washed with deionized water for 1 minute at RT and a staining solution was added. Staining was allowed for 15min in total darkness at RT. Finally, cells were washed with deionized water for 2 min, and the plates were air-dried and kept at RT.

FIXATION SOLUTION		
REAGENT	SOURCE/REFERENCE	VOLUME (mL)
Citrate	Sigma-Aldrich, 915	2.5
Acetone	Fisher Chemical, A/0560/17	6.5
37% Formaldehyde	Sigma-Aldrich, F8775	0.8
STAINING SOLUTION		
REAGENT	REFERENCE	VOLUME
Sodium Nitrite Solution	Sigma-Aldrich, 86R	0.2
FRV-Alkaline Solution	Sigma-Aldrich, 86R	0.2
Mix by gentle inversion. Chill for 2min and add it to the water		

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Deionized water	-	9
Naphtol AS-Bl Alkaline Solution	Mix Sigma-Aldrich, 86R	0.2

6.16 COLONY FORMATION ASSAY

To examine the consequences for self-renewal of *Zmym2*-depleted cells in metastable SL or primed FA pluripotent culture conditions, colony formation assays (CFAs) were performed. Briefly, 600 cells were plated on 0.1% gelatin-coated 6-well plates in SL culture conditions, or on fibronectin-coated 6-well plates (10μg/mL/cm^2) in FA conditions. SL or FA pluripotent medium was changed every other day. Colonies were stained for AP activity as described above. Colonies were counted and scored manually for AP staining under the microscope. The classification of each colony according to their pluripotent state was performed under i) SL culture conditions: undifferentiated (AP positive), partially differentiated (partially AP positive), or differentiated (AP negative); or ii) FA culture conditions: naïve compact dome-shaped or primed diffuse-shaped AP positive colonies.

6.17 DIFFERENTIATION ASSAYS

To determine if depletion of *Zmym2* in ESCs can affect cellular differentiation, the next *in vitro* differentiation approaches were carried out:

6.17.1 EB differentiation assay

ESCs were differentiated *in vitro* by spontaneously self-assembling in suspension culture into 3D aggregates called embryoid bodies (EBs), which mimic many of the hallmarks of early development[334]. To this end, EBs were performed as previously described[334]. Briefly, ESCs were seeded at a density of 155,000 cells/mL in low attachment Petri dishes (Corning, 3472), and cultured in a standard pluripotent medium without LIF for 9 days. The medium was refreshed every other day and RNA samples were collected (see

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

section **6.22**) at days 0, 3, 6 and 9 of EB formation. The expression levels of pluripotency and differentiation markers were analyzed by RT-qPCR (see section **6.23**). EBs were visually analyzed under microscope and the relative diameter of EBs was quantified using Image J software (v1.33).

6.17.2 Mesendoderm and neural ectoderm differentiation assay

Mesendoderm and neural ectoderm differentiation assays were performed as previously described[335]. Briefly, ESCs were cultured in N2B27 medium [DMEM/F12 (Sigma-Aldrich, D6421) and Neurobasal medium (Gibco, 21103049) mixed at a ratio of 1:1, 1x B27 supplement (Gibco, 17504044), 1x N2 supplement (Gibco, 17502048), 2mM L-glutamine, 0.1mM 2-mercaptoethanol, and 50U/mL Penicillin/Streptomycin) supplemented with 10^3U/mL LIF (homemade), 10ng/mL BMP4 (Gibco, 11330852), and 1μM PD0325901 inhibitor (Sigma-Aldrich, PZ0162) for two days. Then, $1.5x10^4$ cells per cm^2 were cultured on 0.1% gelatin-coated 12-well plates (Corning, 3513) during 48 h in N2B27 medium alone. Finally, cells were cultured during 36 h in N2B27 medium supplemented with 3μM CHIR99021(Sigma-Aldrich, SML1046) or 500nM retinoic acid (RA; Santa Cruz Biotechnology, sc-200898) to induce mesendoderm and neural ectoderm specification, respectively. RNA samples were collected (see section **6.22**) and analyzed by RT-qPCR (see section **6.23**).

6.18 REPROGRAMMING ASSAYS

6.18.1 OKSM reprogramming assays in MEFs

To investigate the consequences of lack of function of ZMYM2 during the somatic cell reprogramming (SCR) process to form iPSCs, mouse *Oct4*-GFP embryonic fibroblasts (O4G-MEFs) carrying a GFP reporter under the control of the *Oct4* promoter were reprogrammed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

according to a published procedure with modifications[330]. Briefly, O4G-MEFs were infected with pHAGE2-based lentiviral reprogramming factors OCT4, KLF4, SOX, and c-MYC (OKSM) and pLKO-based lentiviral shRNAs against *Zmym2* or control *Luciferase* (see section **6.8**). The next day, transduced O4G-MEFs were selected with 2μg/mL puromycin (Fisher Bioreagents, BP2956-100) until one of the wells of noninfected cells (i.e., selection control) showed 100% of cell death. Then, transduced and drug-selected O4G-MEFs were seeded at 2 x 10^4 cells per single 0.1% gelatin-coated well of a 12-well plate. The next day, shRNA transduced O4G-MEFs were incubated with concentrated lentiviruses containing four reprogramming factors (OKSM) in the presence of 6μg/mL polybrene for 16 h. One day after transduction, the MEF medium was changed to a standard pluripotent medium, which was changed every other day. The number of iPSCs colonies was determined at the indicated times by staining with a commercial alkaline phosphatase kit (see section **6.15**) and scoring the GFP-positive colonies under the microscope.

6.18.2 Reprogramming assays in pre-iPSCs

Reprogramming intermediates were obtained as previously described[79]. Briefly, TNGA fibroblasts carrying a GFP reporter under the control of the *Nanog* promoter were infected with pMX-based retroviral reprogramming factors OCT4, SOX2, KLF4, and cMYC (OSKM) in the presence of 6μg/mL polybrene. After 20 days of SCR using the standard pluripotent medium, a clonal line of proliferative *Nanog*-GFP-negative cells (TNGA pre-iPSCs) was sorted by flow cytometry. Then, TNGA pre-iPSC were infected with lentiviral particles containing a control shRNA (sh*Luciferase*) or independent shRNAs against *Zmym2* (sh*Zmym2*) in a standard pluripotent medium with the presence of 6μg/mL polybrene (Sigma-Aldrich, H9268). The next day, transduced pre-iPSCs were selected with 2μg/mL puromycin until one of the wells of non-infected cells (i.e., selection control) showed 100% of cell death. After drug selection, 5 x 10^4 cells stably

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

transduced pre-iPSCs were seeded per single 0.1% gelatin-coated well of a 12-well plate in standard pluripotent medium for 24 h before switching to N2B27-supplemented medium containing LIF plus inhibitors of MEK (PD0325901, 1 µM) and GSK3β (CHIR99021, 3 µM) (hereafter "2iL" medium). *Nanog*-GFP-positive colonies from TNGA MEF reprogramming intermediates were scored at day 10 of reprogramming. AP staining was also performed using a commercial alkaline phosphatase kit (see section **6.15**).

6.18.3 Heterokaryon-based fusion reprogramming assay

Experimental heterokaryons were generated by fusing mouse ESCs and human B (hB) lymphocytes according to a published protocol[336]. Briefly, 100 x 10^6 human B lymphocytes and 100 x 10^6 *Zmym2*$^{+/+}$ or *Zmym2*$^{GT/GT}$ mouse ESCs were resuspended as single-cell suspensions in 15mL of prewarmed hB cell medium and standard pluripotent medium (SL), respectively. Then, cells were harvested by centrifugation for 5 min at 300g and resuspended in 15mL of CMF-PBS medium [3% FBS (REF) in 1X DPBS (Corning, 21-031-CV)]. Next, cell fusion was performed as follows: 1) hB cells and *Zmym2*$^{+/+}$ or *Zmym2*$^{GT/GT}$ cells were mixed in a 1:1 ratio using a canonical 30-mL universal tube, 2) cell mixtures were centrifuged for 5min at 200g and the supernatant was discarded, 3) cell pellets were disrupted by gently tapping the bottom of the tube and cells were then warmed for 2min in a 37ºC water bath, 4) 1mL of polyethylene glycol (PEG) solution [50% PEG 1500 (Sigma-Aldrich, 10783641001), 50% HEPES pH8.0 (Fisher Bioreagents, BP310-100)] was added dropwise to the cell mixtures over a 60-second period while tapping the bottom of the tube, 5) cells were placed in a 37ºC water bath and cell pellets were continuously disrupted by tapping the tube for an additional 90-second period, 6) 1mL of serum-free KO-DMEM (Gibco, 10829018) was added dropwise over a 60-second period and then an extra 10mL of KO-DMEM medium was slowly added, 7) cells were mixed by gently inverting the tube 6 times and incubated for 3 min at 37ºC, 8) cells were centrifuged for 5min at

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

300g and supernatant was discarded, 9) 10mL of ES/heterokaryon medium (see recipe below) was added without disrupting the cell pellet and then incubated for 3min at 37ºC, 10) cells were carefully resuspended, transferred to 90mm gelatin-coated dishes (Sigma-Aldrich, G9391), and incubated for 6 h at 37ºC in a CO_2 incubator, 11) 10^{-5}M Ara-C (Sigma-Aldrich, C1768) and 10^{-5}M ouabain (Sigma-Aldrich, O3125) were added to cell mixtures to eliminate non-fused proliferating ESCs and non-fused hB lymphocytes, respectively, and incubated at 37ºC ON. Finally, the medium was changed daily with a previous washing step with a pre-warmed CMF-PBS medium. Ouabain treatment was maintained for 48h. The reprogramming of hB cells was monitored for 3 days by RT-qPCR (see section **6.23**) analyzing human-specific gene expression. The sequences of human gene-specific primers are provided in **Supplementary Table 6**.

MOUSE ES/HETEROKARYON MEDIUM		
REAGENT	AMOUNT	REFERENCE
KO-DMEM	90%	Gibco, 10829018
FBS	10%	Gibco, 10270106
L-glutamine	2mM	Gibco, 25030-081
Non-essential amino acids (NEAA)	100mM	Gibco, 11140-050
2-Mercaptoethanol	5µM	Sigma-Aldrich, M-6250
Penicillin/streptomycin	50U/mL	Gibco, 15070-063
LIF	10^3U/mL	homemade

6.19 ASSESSING THE BIOENERGETIC PROFILE OF ESCS

To evaluate the role of ZMYM2 in cellular bioenergetics, we measured both the basal oxygen consumption rate (OCR) as a measure of mitochondrial respiration, and the extracellular acidification rate (ECAR) as a measure of glycolysis in WT and *Zmym2* knockout ESCs. To this end, XFp Cell Mito Stress Test Kit (Agilent, 103010) and XFp Glycolytic Rate Assay Kit (Agilent, 103346) were used respectively, and assays were performed in a Seahorse XFp Bioscience instrument (Agilent). Briefly, 4.5 x 10^4 ESCs were seeded in extracellular flux 6-well culture mini plates (Agilent, 103025) coated with 0.1% gelatin and then incubated in the standard pluripotent medium at 37ºC for 24 hours.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Next, the standard pluripotent medium was replaced by SeaHorse XF DMEM Medium, pH7.4 (Agilent, 103575-100), and specific drugs were added to the cells following the manufacturer's instructions. Thus, 50µM oligomycin, 50µM carbonyl cyanide-4-phenylhydrazone (FCCP) and 25µM rotenone/antimycin AA were added to check mitochondrial respiration whereas 25µM rotenone and antimycin AA, and 500mM 2-deoxy-D-glucose (2-DG) were added to check the glycolysis dependence after inhibiting mitochondrial respiration. The cells were then incubated for 45-60 min in a CO_2-free incubator and metabolic parameters were determined using the Seahorse XFp Bioscience instrument.

6.20 CHROMATIN ACCESSIBILITY ASSAY

To investigate the consequences of the absence of ZMYM2 in chromatin biology, a chromatin accessibility assay was performed using the EZ Nucleosomal DNA Prep Kit (Zymo Research, D5220) and following the manufacturer's instructions. Briefly, 1 x 10^6 ESCs were seeded onto 10cm dishes (Corning, 3548) coated with 0.1% gelatin in standard pluripotent medium (i.e., SL) at 37ºC for 48 h. Then, cells were detached using Trypsin-EDTA (Sigma-Aldrich, T3924), split in aliquots containing 1 x 10^6 cells, and washed with 1X DPBS. To isolate the nucleosomes, cell pellets were resuspended in 100µL Nuclei Prep Buffer and incubated on ice for 5min. Then, the supernatant was removed avoiding nuclear pellet disruption. Pellets were washed, resuspended in 100µL Atlantis Digestion Buffer by gentle inversion, and centrifuged. The precipitated nuclei were resuspended in 100µL MN Digestion Buffer, and 0.2 units of micrococcal nuclease enzyme or DMSO as control were added. The final mix was homogenized by inverting the tubes and incubated for 5 min at RT. To stop the reaction, 5X Stop Buffer was added, and nucleosomal DNA isolation was performed. To this end, a 5:1 ratio of DNA Binding Buffer was added, and a homogenization step was performed by a vortex. Then, samples were transferred to Zymo-Spin™ IICR Columns into collection tubes and centrifuged for 30 sec. Flow-through was discarded and two serial

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

washing steps were performed by adding Wash Buffer into the columns followed by a 30 sec centrifugation step. Columns were then transferred to clean collection tubes and nucleosomal DNA was eluted by adding RNase-free water and centrifuged for 30 sec. Lastly, 100ng of DNA for each sample was run on a 1.5% (w/v) agarose (Fisher Bioreagents, BP1356) gel in 0.1 x TBE buffer at a voltage of 90V, which was finally stained with ethidium bromide (Fisher Bioreagents, BP1302-10).

6.21 LUCIFERASE ASSAY

A luciferase reporter system containing the *Padi4* regulatory promoter region (see section **6.5.3**) was employed to confirm the ZMYM2-mediated repression of the *Padi4* gene. Thus, a luciferase assay of the *Padi4* reporter (i.e., pGL3-*Padi4*.promoter) was performed in ZMYM2-expressing ESCs compared to *Zmym2* null ESCs under standard pluripotent culture conditions. Cells were co-transfected with pGL3-*Padi4*.promoter and a pCMV-β-Galactosidase plasmid (kindly provided by Dr. Anxo Vidal, CIMUS/USC) as a control reporter using jetPRIME kit (Polyplus, 114-0715) as previously indicated (see section **6.5.3**). The next day, cells were washed twice with 1X DPBS, and protein was extracted from a confluent well of a 6-well plate by scrapping in the presence of 100µL lysis solution (see recipe below). The lysates were then transferred to microcentrifuge tubes and centrifuged at 10,000g for 2 min at 4ºC. The supernatants were transferred to clean microcentrifuge tubes and for long-term use stored at -80ºC. To determine the luciferase activity, 55µL of Luciferase Assay Buffer (LAB buffer; see below) and 1-5µL of sample diluted in 23µL lysis solution were added into a single well of a clear flat-bottom 96-well plate (Falcon, 353377). Then, mixtures were homogenized by rocking in a microplate shaker for 5min at RT, and 22µL of D-luciferin substrate (Thermo Scientific, 88293) was added to start the reaction. Bioluminescence indicating luciferase activity was measured at 560nm using a Mithras LB 940 microplate reader (Berthol Technologies). In addition, a colorimetric β-Galactosidase assay was performed in parallel to normalize bioluminescence data. Thus, the same samples

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

used in the luciferase assay were added into a 96-well plate (Corning, 3595) into a final volume of 100µL with a Z-buffer (see recipe in the following table), and 2mM O-nitrophenyl-β-D-galactopyranoside (ONPG; Thermo Scientific, 34055) was added. Mixtures were incubated at 37ºC until color turns yellow, and the reaction was stopped by adding 33.34µL of 1M sodium carbonate (Na_2CO_3; Fisher Chemical, S-2880-53). Finally, absorbance was measured at 420nm using a Mithras LB 940 microplate reader (Berthol Technologies).

LYSIS SOLUTION		
REAGENT	FINAL CONCENTRATION	SOURCE/REFERENCE
Phosphate buffer pH 7.8	100mM	Homemade
Triton™ X-100	0.2%	Sigma-Aldrich, T8787
LAB BUFFER		
REAGENT	FINAL CONCENTRATION	SOURCE/REFERENCE
ATP	2mM	Thermo Scientific, R0441
DTT	1mM	Fisher BioReagents, BP172-5
EGTA pH 8.8	4mM	Alfa Aesar, J60767
Glycylglycine pH 8.0	25mM	Alfa Aesar, J62920
Magnesium sulfate	15mM	Fisher Chemical, M/1100/53
Phosphate buffer pH7.8	15mM	Homemade
Z-BUFFER		
REAGENT	FINAL CONCENTRATION	SOURCE/REFERENCE
2-Mercaptoethanol	50mM	Sigma-Aldrich, M-6250
Magnesium sulfate	1mM	Fisher Chemical M/1100/53
Potassium chloride	10mM	Fisher Chemical, P/4240/60
Sodium phosphate solution pH 7.4		Alfa Aesar, J62152

6.22 RNA EXTRACTION

Total RNA from different cell types used in this thesis was extracted with E.Z.N.A.® Total RNA Kit I (Omega Bio-Tek, R6834-02) following the manufacturer's instructions. For RNA sequencing (RNA-seq; see section **6.28**), the RNA samples were extracted using Trizol (Invitrogen, 15596018) and Phasemaker tubes (Phasemaker, A33248) following the manufacturer's indications. RNA concentration and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

quality were assessed by NanoDrop 2000 Spectrophotometer (Thermo Scientific, ND2000) or by Qubit 4 Fluorometer (Invitrogen, Q33226) using Qubit RNA HS Assay Kit (Invitrogen, Q32855) and Qubit RNA IQ Assay Kit (Invitrogen, Q33222) and following the manufacturer's protocols.

6.23 ANALYSIS BY QUANTITATIVE PCR

For the obtention of complementary DNA (cDNA), total RNA was converted into cDNA using the qScript cDNA SuperMix Kit (Quantabio, 95048) following the manufacturer's instructions. Reverse transcription (5 min at 25ºC, 30 min at 42ºC, and 5 min at 85ºC) was performed using a SimpliAmp Thermocycler System (Applied Biosystems, A24811). Quantitative PCR (hereafter "RT-qPCR") was performed using PowerUp SYBR Green Master Mix (Applied Biosystems, A25742), and a mix of forward and reverse gene-specific primers was added to cDNA samples. Then, mixes were transferred to MicroAMP Fast Optical 96-well Reaction plates (Applied Biosystems, 4346907) and sealed with Optical Adhesive Films (Applied Biosystems, 4360954). The qPCR was performed on the StepOnePlus Real-Time PCR System (Applied Biosystem) following specific amplification conditions (see below). Average threshold cycles were determined from triplicate reactions and the data was analyzed using the delta-delta CT method, which was normalized to the values of a housekeeping gene and the relative expression values of indicated control samples. Gene-specific primers used in this thesis are provided in **Supplementary Table 6**. All data representation was performed with GraphPad Prism (v8.0.2) or Microsoft Excel.

Temperature (ºC)	Time	Number of cycles
98	10 min	1
95	15 sec	40
55	15 sec	
72	40 sec	
95	5 sec	1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

65	1 min	Increase 5ºC/min until reach 97ºC
97	15 sec	1
4	hold	-

6.24 WESTERN BLOT ANALYSES

Whole-cell extracts were obtained by cell lysis with RIPA Lysis Buffer System (Santa Cruz Biotechnology, sc-24948) following the manufacturer's instructions with freshly added protease and phosphatase inhibitors (Santa Cruz Biotechnology, sc-24948). Then, lysates were homogenized by shearing them through a narrow 21G needle (Thermo Scientific, 8AN2138R1), incubated for 30 min at 4ºC, and centrifuged at 14,000g for 15 min at 4ºC to remove cellular debris. The supernatant was transferred to new microcentrifuge tubes and samples were stored at -80ºC until use.

Histones were purified using the Histone Extraction kit (Abcam, ab113476) following the manufacturer's instructions. Briefly, cell pellets were resuspended in Triton Extraction Buffer (TEB) containing 0.5% Triton X-100, 2mM phenilmethylsulfonil, and 0.02% NaN_3. First, suspensions were incubated on ice for 10 min. Second, cells were pellet by centrifugation, and another lysis step was performed. Third, cell pellets were resuspended in 0.2N HCl, and acidic extraction of histones was performed at 4ºC for several hours. Finally, lysates were centrifuged, and supernatants containing the purified histones were transferred to new microcentrifuge tubes and stored at -80ºC until use.

Protein concentration was measured with Bradford reagent (Sigma-Aldrich, B6916) following the manufacturer's instructions.

Proteins for Western blot analysis were prepared by adding Laemmli buffer (homemade) and denatured for 5min at 95ºC. Samples were loaded for resolution in a Novex WedgeWell 4–20% tris-Glycine Gel (Invitrogen, XP04205BOX) or homemade acrylamide/bis-acrylamide (Sigma-Aldrich, A3574) gels. Proteins were then separated by

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

electrophoresis and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, IPFL00005) at 4ºC, ON. Then, membranes were washed with miliQ water and stained with Ponceau S reagent (Sigma-Aldrich, P3504) for 5 min at RT. Membranes were washed again in miliQ water and scanned. Next, membranes were rinsed 3 times for 5min each in 1X TBS buffer (Fisher Bioreagents, BP2471-1) supplemented with 0.1% TWEEN® 20 (Sigma-Aldrich, P1379) (hereafter " TBS-T") and blocked to avoid unspecific binding of antibody by incubating them in 5% powder milk (LabScientific, M-0841) or 3% BSA (Fisher Bioreagents, BP9702-100) in TBS-T for 1h at RT. Membranes were washed again 3 times with gentle agitation for 15 min each in TBS-T and incubated with specific primary antibodies diluted in 0.5% powder milk or 0.3% BSA in TBS-T. Then, the blot was washed 3 times and incubated with fluorescence-conjugated secondary antibodies in a 0.5% powder milk or 0.3% BSA TBS-T solution for 45 min at 4ºC. Finally, the membranes were washed 3 times for 15 min each in TBS-T, and Western blot signals were detected using the LI-COR Odyssey 9120 Digital Imaging System (LI-COR), and images were processed and analyzed using the Odyssey Infrared Imaging System (v3.0.29). The antibodies and dilutions used in this thesis for western blot are provided in **Supplementary Table 5**.

6.25 AFFINITY PURIFICATION OF ZMYM2 PROTEIN COMPLEXES

To isolate endogenous ZMYM2-associated protein complexes, nuclear extracts from four large square dishes (245 x 245mm) of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were prepared as previously described[79]. Briefly, cell pellets were first resuspended in 5x packed cell volume (PCV) of ice-cold Nuclear Extract Buffer A (see below) containing 0.5mM DTT (Fisher Scientific, 10592945), 0.2mM PMSF (Santa Cruz Biotechnology, sc-482875), and 0.1mM protease inhibitor cocktail (Sigma-Aldrich, P8340) and lysates were centrifuged at 2,000 x g for 5min at 4ºC. Nuclear pellets were resuspended again in 3x PCV of ice-cold inhibitors-containing Nuclear Extract Buffer A and incubated for 10min on ice. Then, samples were sheared using an 18G ½ narrow

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

needle (Termo, 8AN1838S1) and transferred to 50mL tubes (Falcon, 352070), centrifuged at 4,300g for 5 min at 4ºC, and resuspended in 1.5mL of ice-cold Nuclear Extract Buffer C (see below) containing protease inhibitors without DTT. Nuclear pellets were sheared using a 21G ½ narrow needle (Termo, 8AN2138R1) and rotated for 30min at 4ºC in presence of Benzonase Nuclease (Sigma-Aldrich, E8263) to degrade all forms of DNA and RNA. Homogenates were then centrifuged at 20,000g for 30 min at 4ºC to obtain nuclear extracts. Before immunoprecipitation, 20mg of nuclear extracts were pre-cleared with 100µL of Dynabeads™ Protein G (Invitrogen, 10004D) for 1h at 4ºC in the presence of 750 units of Benzonase Nuclease and rotated for 6h at 4ºC. In parallel, 20µg of ZMYM2 antibody (Abcam, ab30783) were pre-bound to 100µL of Dynabeads™ Protein G and rotated for 4h at 4ºC. Next, pre-bound ZMYM2 and IgG antibodies were each incubated with 15mg of previously pre-cleared nuclear extracts for 8h with gentle rotation. Then, immunoprecipitates were washed five times with Nuclear Extract Buffer D (see below) supplemented with 0.02% NP-40 (Thermo Scientific, 85124) and bound material was eluted by boiling for 5 min in 2x Elution buffer (3% SDS, 1% Glycerol, 500mM Tris HCl pH 6.8, bromophenol blue). Eluted samples were then fractionated on a Novex WedgeWell 4–20% tris-Glycine Gel (Invitrogen, XP04200BOX), and stained with GelCode Blue Safe Protein Stain buffer (Thermo Scientific, 24594). Afterward, each whole lane was cut into four pieces and subjected to quantitative nano-scale liquid chromatography-tandem mass spectrometry (nLC-MS/MS) sequencing. MS data analysis was processed by Thermo Proteome Discoverer software with SEQUEST engine against Swiss-Prot mouse protein sequence database. Proteins were filtered by identified unique peptide≥ 2, and $\log_2$ spectrum count ratio of ($Zmym2^{+/+}/Zmym2^{GT/GT}$) ≥ 3. The list of ZMYM2-associated proteins I provided in **Supplementary Table 8.**

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

NUCLEAR EXTRACT BUFFER A		
REAGENT	FINAL CONCENTRATION	REFERENCE
HEPES pH 7.9	10mM	Cytiva, SH30237.01
KCl	10mM	Fisher Chemical, P/4240/60
$MgCl_2$	1.5mM	Acros Organics, 223210010
NUCLEAR EXTRACT BUFFER C		
REAGENT	FINAL CONCENTRATION	REFERENCE
EDTA	0.2mM	Sigma-Aldrich, E7889
Glycerol	25%	Sigma-Aldrich, G5516
HEPES pH 7.9	20mM	Cytiva, SH30237.01
$MgCl_2$	1.5mM	Acros Organics, 223210010
NaCl	0.42M	Fisher Bioreagents, BP358-1
NUCLEAR EXTRACT BUFFER D		
REAGENT	FINAL CONCENTRATION	REFERENCE
EDTA	0.2mM	Sigma-Aldrich, E7889
Glycerol	25%	Sigma-Aldrich, G5516
HEPES pH 7.9	20mM	Cytiva, SH30237.01
$MgCl_2$	1.5mM	Acros Organics, 223210010

6.26 IMMUNOFLUORESCENCE

To analyze both the localization and expression levels of different proteins, immunostaining was performed using specific antibodies. Briefly, $2x10^3$ ESCs were seeded on 96-well plates (Corning, 3595) coated with 0.1% gelatin (Sigma-Aldrich, G9391) and incubated at 37ºC in a CO_2 incubator ON. The next day, cells were fixed with a 4% paraformaldehyde (PFA; Sigma-Aldrich, P6148) solution for 15 min in darkness at RT or with 100% ice-cold methanol (MetOH; Fisher Chemical, M-4056-17) for 10 min at -20ºC, depending on the antibody used, and washed 2 times with 1X DPBS. Then, cells were permeabilizated with 0.25% Triton™ X-100 (Sigma-Aldrich, G9391) diluted in 1X DPBS for 5 min at RT followed by 2 times of 1X DPBS washes, and a blocking step was performed by incubating 10% BSA (Fisher BioReagents, BP1605) diluted in 1X DPBS for 30min at RT. Next, primary antibodies diluted in 3% BSA were incubated ON at 4ºC. After 5 washes of 15 min each using 1X DPBS, fluorescent secondary antibodies and 1µM 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich, D9542) for nuclei stained were added in 3% BSA diluted in 1X DPBS and incubated for 1 h in darkness at RT. Immunofluorescence

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

images were taken with a Leica DMI 6000 inverted microscope (Leica Microsystems) and then analyzed using the LAS X software (Leica Microsystems) and CellProfiler™ 2.1.1 (https://cellprofiler.org/) software. For the quantification of each marker, the images were loaded into the CellProfiler™ 2.1.1 software and grouped according to the experimental condition, and the intensities of the signals were transformed into grayscale ranges. Therefore, single-cell measurements were defined based on previously determined nuclear size (20-70 pixels), and specific signal intensities were associated with individual cells. The results were then imported into Microsoft Excel (Microsoft) and different parameters were analyzed.

The antibodies and dilutions used in this thesis for Immunofluorescence are provided in **Supplementary Table 5**.

6.27 NUCLEAR AREA MEASUREMENT ON ESCS

Nuclear area measurement on ESCs was performed using DAPI staining to visualize nuclear DNA. Briefly, $2x10^3$ *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were seeded on 96-well plates (Corning, 3595) coated with 0.1% gelatin (Sigma-Aldrich, G9391) and incubated at 37ºC in a CO_2 incubator overnight. The next day, cells were fixed with a 4% paraformaldehyde (PFA; Sigma-Aldrich, P6148) solution for 15 min in darkness at RT and washed 2 times with 1X DPBS. Then, cells were permeabilizated with 0.25% Triton™ X-100 (Sigma-Aldrich, G9391) diluted in 1X DPBS for 5min at RT followed by 2 times of 1X DPBS washes, and a blocking step was performed by incubating 10% BSA (Fisher BioReagents™, BP1605) diluted in 1X DPBS for 30min at RT. Next, 1µM 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich, D9542) for nuclei stained was added in 3% BSA diluted in 1X DPBS and incubated for 1 h in darkness at RT. The stained nuclei were then visualized with a Leica DMI 6000 inverted microscope (Leica Microsystems) and the images were analyzed with CellProfiler™ 2.1.1 software (https://cellprofiler.org/) following an approach similar to that

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

described above in section **6.26**. Finally, the nuclear areas were inferred from the Nuclear Mean Radius parameter recorded in the analysis.

6.28 RNA SEQUENCING AND ANALYSIS

To investigate how ZMYM2 orchestrates the transcriptional programs involved in maintaining pluripotency, total RNA was extracted using TRIzol reagent (see section **6.22**). RNA quality was evaluated by Agilent 2100 BioAnalyzer and subjected to a RiboZero selection protocol and paired-end sequencing (RNA-seq) was carried out with the Illumina HiSeq4000 platform with a read length of 150bp on Novogene (https://en.novogene.com). Four biological replicates of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs grown under standard pluripotent culture conditions (i.e., SL) were subjected to sequencing.

Bioinformatic analyses to identify differentially expressed genes between null and control *Zmym2* ESCs were performed by my labmate Alejandro Fuentes-Iglesias as we previously described[337]. Briefly, paired-end 150bp-length reads were aligned against the GRCm38 mouse genome (mm10) using the TopHat (v2.1.1) and Bowtie (v2.3.4.1) software with the default parameters settings. Assembly of novel transcripts was not allowed (-G) and other parameters were used following default settings. Then, transcript assembly and differential expression analysis were performed using the Cufflinks (v2.2.1) software. Expression shared by transcripts assigned to the same identifier (Gene_ID) was quantified as fragments per million mapped reads (FPKM) and the significance of differential expression was determined using the Benjamin-Hochberg correction for multiple testing. Hierarchical clustering of samples was performed using GENE-E 3.0.215, and datasets were grouped using the Pearson correlation coefficient (R). The unexpressed genes were filtered out.

For the published LSD1 RNA-seq analysis (see table below), additional steps were performed in the bioinformatic pipeline. The quality of single-end reads was tested using the FastQC (v0.11.5) software. To

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

increase both the quality and validation of the RNA-seq analysis, short-length reads with less than 40bp were removed from the analysis using the Trimmomatic (v0.39) software. A quality confidence scan of reads was performed using a 4bp sliding window and reads with a quality value of less than 15 were discarded from the analysis.

Other public RNA-seq datasets used in this thesis are listed below. The alignment, assembling, and differential analysis of transcripts of all RNA-seq data were processed with the same settings.

GSE	DATASET	REFERENCE
GSE33923	LSD1 RNA-seq	338
GSE106332	MEFs, iPSCs RNA-seq	339
GSE53212	MEFs, iPSCs, ESCs RNA-seq	340
GSE93027	MEFs, iPSCs, ESCs RNA-seq	226

6.29 CHROMATIN IMMUNOPRECIPITATION FOLLOWED BY SEQUENCING AND ANALYSIS

H3Cit chromatin immunoprecipitation (ChIP) was performed as previously described [341]. Briefly, 10 x 10^6 *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were crosslinked by adding 1/10 volume of fresh 11% formaldehyde solution [50mM HEPES-KOH pH7.5 (Fisher Bioreagents, BP299), 100mM NaCl (Invitrogen, AM9760G), 1mM EDTA (Fisher Bioreagents, BP2482), 0.5mM EGTA (Alfa Aesar, J61721.AD), 11% formaldehyde (Sigma-Aldrich, F8775)] to the plates and incubated for 10 min at RT. Then, 1/20 volume of 2.5M glycine was added to quench formaldehyde, and cells were rinsed twice with 10mL of 1X DPBS (Corning, 21-031-CV) and then harvested by scrapping. Next, cells were transferred to 50mL conical tubes and spined at 700g for 5min at 4ºC, and the supernatant was discarded. Cell pellets were resuspended in 5mL of 1X DPBS, centrifuged again at 700g for 5 min at 4ºC, supernatants were discarded, and cell pellets were flash-frozen in liquid nitrogen and stored at -80ºC. For the chromatin immunoprecipitation, magnetic beads were prepared at 4ºC

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

by adding 100µL of Dynabeads™ Protein G (Invitrogen, 10004D) to a 1.5mL microfuge tubes containing 1mL of Block Solution [1X DPBS. 0.5% BSA (Fisher BioReagents, BP1605-100)]. Magnetic beads were then washed by placing the Dynabeads™ Protein G-containing tubes in a Dynal MPC (Invitrogen, 12321D) and allowing their collection on the side of the tubes. Then, supernatants were discarded, and 1.5mL of fresh Block Solution was added to the tubes. The washing step was repeated a total of three times. Finally, beads were resuspended in 250µL of Block Solution, and 10µg of ChIP-grade antibody against H3Cit (**Supplementary Table 5**) was added and incubated on a rotator at 4ºC ON. The next day, beads were washed three times in 1mL of Block Solution as previously described and were finally resuspended in 100µL of Block Solution. To proceed with the immunoprecipitation, cell pellets were thawed on ice, resuspended in 10mL of Lysis Buffer 1 [50mM HEPE-KOH pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol (Fisher BioReagents, BP229-4), 0.5% NP-40 (Thermo Scientific, 85125), 0.25% Triton X-100 (Sigma-Aldrich, T8787) and protease inhibitors (Sigma-Aldrich, P8340)] and rocked at 4ºC for 10min. Then, samples were spined at 1,350g for 5 min at 4ºC, and supernatants were discarded. Pellets were resuspended in 5mL of Lysis Buffer 2 [10mM Tris-HCl pH8.0 (Thermo Scientific, J60080.AK), 200mM NaCl, 1mM EDTA, 0.5mM EGTA, and protease inhibitors] and rocked at RT for 10 min. Nuclei were then pelleted by centrifuging at 1,350g for 5 min at 4ºC, and the supernatants were discarded. Pellets were then resuspended in 1.5mL of Lysis Buffer 3 [10mM Tris-HCl pH8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate (Sigma-Aldrich, D6750), 0.5% N-lauroylsarcosine (Sigma-Aldrich, L9150) and protease inhibitors], placed in an iced-water bath and sonicated [20% amplitude, 15 pulses (1sec ON, 0.5sec OFF) and 2 min pause between pulses]. Next, 1/10 volume of 10% Triton X-100 was added to sonicated lysates, which were split into two 1.5mL centrifuge tubes for a centrifuging step at 20,000g for 10 min at 4ºC. Supernatants from the two tubes were then combined in a new 15mL conical tube, 50µL of cell lysates were saved and stored at -20ºC, and nuclei lysates

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

were stored at -80ºC until use. To proceed with the immunoprecipitation, nuclei lysates were thawed on ice, and 100µL of antibody/magnetic beads mix was added and incubated at 4ºC ON. The next day, half of the immunoprecipitated nuclei lysates were transferred to 1.5mL microfuge tubes and beads were collected using a Dynal MPC as indicated above. Supernatants were discarded, and the second half of the immunoprecipitation was added to continue with the beads collection. Then, supernatants were discarded and a wash step with 1mL Wash Buffer (50mM HEPE-KOH pKa7.55, 500mM LiCl, 1mM EDTA, 1% NP-40, and 0.7% Na-Deoxycholate) was performed, followed by another washing step with 1mL TE + 50mM NaCl. Then, beads were spined at 960g for 3 min at 4ºC, residual TE was removed and 210µL of Elution Buffer was added. Elution of material from beads was performed by incubating the tubes in a 65ºC water bath for 15 min with frequent vortex. Next, samples were spined at 16,000g for 1 min at RT, and supernatants were transferred to new tubes. For the crosslinking reversal, supernatants were incubated at 65ºC for a minimum of 6 h, and 200µL of TE was added to each tube to dilute the SDS present in the Elution Buffer. Then, 0.2µg/mL of Proteinase K (Amresco, E195) was added and incubated at 55ºC for 2 h to purify DNA. Next, 400µL phenol:chlorophorm:isoamyl alcohol (Fisher Bioreagents, BP1752I) was added, samples were vortexed and phases were separated by using 2mL Heavy Phaselock tubes (Phasemaker, A33248). The aqueous phases were transferred to new tubes containing 16µL of 5M NaCl and 1.5µL of 20µg/µL glycogen (Roche, 10901393001). Finally, 800µL EtOH (Sigma-Aldrich, 108543) was added and incubated for 30min at -20ºC, samples were then spined at 20,000g for 10 min at 4ºC, and DNA pellets were washed by adding 500µL of 80% EtOH and centrifuging them at 20,000g for 5min at 4ºC. DNA pellets were then resuspended in 70µL of 10mM Tris-HCl, pH8.0. Biological duplicates were used for library construction, followed by massive parallel sequencing with the Illumina HiSeq4000 according to the manufacturer's protocol.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

After sequencing, paired-end 150 bp-length reads were generated and analyzed using FastQC (v0.11.5) to check read quality. Reads were combined and aligned to the GRCm38 mouse genome (mm10) using the Bowtie (v1.2.2) software, with parameters -M 1–chunkmbs 200. Aligned reads were then converted to binary BAM file, sorted, PCR duplicates removed and indexed with Samtools (v1.7), followed by visualization using IGV software. H3Cit ChIP-seq peaks were determined by the MACS2 program (v2.1.4), using Input ChIP-seq as the control data and with default parameters. Public ZMYM23 and LSD1 ChIP-seq data were downloaded (see table below) and processed with the same settings. The ChIP-seq peaks with significantly enriched intensities (FDR < 0.01) were exported. ZMYM2 binding motifs were determined using the findMotifs.pl script in HOMER tools.

GSE	DATASET	REFERENCE
GSE119819	ZMYM2 ChIP-seq	342
GSE119819	LSD1 ChIP-seq	342

6.30 ChIP-qPCR assay

ZMYM2 and LSD1 chromatin immunoprecipitation were performed as described above in **section 6.29**. Purified ZMYM2 and LSD1 ChIPed DNA were analyzed by RT-qPCR using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, A25742) in a SimpliAmp (Applied Biosystems, A24811) thermocycler (see section **6.23**). The percentage of input recovery was calculated for each locus. The primers used for qPCR are provided in **Supplementary Table 6**.

6.31 Gene Ontology and KEGG pathway analyses

Gene ontology analysis (GO) and Kyoto Encyclopedia of Genes and genomes (KEGG) pathways were performed using DAVID functional annotation tool (https://david.ncifcrf.gov/tools.jsp)[343] and/or STRING protein-protein interaction networks functional enrichment analysis tool (https://string-db.org/) with a reference list including all *Mus*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

musculus genes from NCBI. The selected gene ontologies used for comparative transcriptional analysis in this thesis are listed below.

GO ID	TERM
GO:0000165	MAPK signaling pathway
GO:0000187	Activation of MAPK activity
GO:0001666	Response to hypoxia
GO:0001709	Cell fate determination
GO:0006520	Cellular amino acid
GO:0006536	Glutamine metabolic process
GO:0006629	Lipid metabolic process
GO:0006636	Biosynthesis of unsaturated fatty acids
GO:0006730	One-carbon metabolic process
GO:0007275	Multicellular organism development
GO:0008652	Biosynthesis of amino acids
GO:0009058	Biosynthetic process
GO:0009749	Response to glucose
GO:0010628	Positive regulation of gene expression
GO:0010862	Positive regulation of SMAD protein phosphorylation
GO:0014065	PI3K-Akt signaling pathway
GO:0016055	WNT signaling pathway
GO:0016055	Canonical WNT pathway
GO:0018101	Protein citrullination
GO:0019827	Stem cell population maintenance
GO:0030154	Cell differentiation
GO:0030178	Negative regulation of WNT
GO:0032526	Response to retinoic acid
GO:0043408	Regulation of MAPK cascade
GO:0043410	Positive regulation of MAPK
GO:0045165	Cell fate commitment
GO:0055114	Oxidation-reduction process
GO:0070371	Positive regulation of ERK
GO:0071773	Cellular response to BMP stimulus

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.32 GENE SET ENRICHMENT ANALYSIS (GSEA)

GSEA_4.0.3 (http://software.broadinstitute.org/gsea/index.jsp) software was used in this thesis to analyze the expression enrichment differences between *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. Briefly, the analysis was performed running 1,000 permutations and a gene_set permutation type was used. Enrichment analysis was weighted _p2 for the enrichment score calculation, and t-test was used to rank genes. Enrichment plot, normalized enrichment score (NES), statistical significance (*P*-value), and False Discovery Rate (FDR) were calculated using the software.

6.33 DETERMINING THE PADI FAMILY CONSERVATION TREE

The amino acid sequences of each PADI protein in *Mus musculus* were obtained from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/) and are listed below. The evolutionary history was inferred by using the Maximum Likelihood method[344] and Jones-Taylor-Thornton (JTT) model[345]. Next, a Bootstrap method was performed with 500 repetitions and the tree with the highest logarithmic likelihood was elaborated.

PROTEIN	PROTEIN ID
PADI1	NP_035189.1
PADI2	NP_032838.2
PADI3	NP_035190.3
PADI4	NP_035191.2
PADI6	NP_694746.2

6.34 NUCLEAR LOCALIZATION SIGNAL PREDICTION

To predict nuclear localization signal (NLS) sequences in PADI proteins, the fasta amino acidic sequences file for each protein downloaded from the NCBI protein database (see the section above) was uploaded to the online prediction tool cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi). Putative NLS sequences with a cut-off greater than 5 are shown, indicating an exclusive or partially nuclear localization.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.35 INTERACTION NETWORK ANALYSIS

Cytoscape (v3.8.2) software[346] and its applications including STRING DB, Cluster ONE, and ClueGO were applied to analyze protein-protein (PPI) networks in this thesis. For the visualization of the common interactors of NANOG and OCT4, data from protein interactions were taken from publications listed below. Data was loaded into Cytoscape (v3.8.2) software[346] for final visualization.

INTERACTOME	REFERENCE
ZMYM2	This study
NANOG	79,287
OCT4	77

For the visualization of GSEA analysis, output was loaded into Cytoscape (v3.8.2) using the EnrichmentMap tool[347]. Briefly, single circles represent specific gene ontologies, whereas color represent enrichment levels (blue tones are low levels of enrichment and red tones higher levels of enrichment). The size of the circles depends on the number of genes composing the respective gene ontology. Connecting lines are shown between gene ontologies sharing specific genes, and the width of connecting lines represents the number of shared genes between gene ontologies.

6.36 STATISTICAL ANALYSIS

All statistical analyses were performed using GraphPad Prism (GraphPad Software, Inc.), Excel or R (http://www.r-project.org). Statistical significances were calculated using two-tailed t-Test, Mann-Whitney test, one-way ANOVA, or Kruskal-Wallis test as indicated in the figure legends. *P values* lower than 0.05 were deemed statistically significant.

6.37 BIBLIOGRAPHIC RESOURCES

In this thesis, Google Scholar (https://scholar.google.es/) and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) were used for bibliographic

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

search. The bibliography was generated using EndNote X9 software (https://endnote.com/).

6.38 SOFTWARE AND ALGORITHMS

The software and algorithms used in this thesis are listed below.

NAME	SOURCE	IDENTIFIER/LINK
Bowtie v2.3.4.1	Johns Hopkins Univ.	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
CellProfiler	Broad Institute	https://cellprofiler.org
Cufflinks v2.1.1	Washington Univ.	http://cole-trapnell-lab.github.io/cufflinks/
DAVID	Laboratory of HRI	https://david.ncifcrf.gov/tools.jsp
EndNote X9	The EndNote Team	https://endnote.com/
FlowJo v7.6.1	FlowJo, LLC	https://www.flowjo.com/
GENE-E v3.0.215	Broad Institute	https://software.broadinstitute.org/GENE-E/
GraphPad Prism v8.0.2	GraphPad	https://www.graphpad.com/company/
GSEA v4.0.3 and MSigDB	Broad Institute	http://software.broadinstitute.org/gsea/index.jsp
ImageJ	NIH	https://imagej.nih.gov/ij/
IGV v2.7.2	Broad Institute	https://software.broadinstitute.org/software/igv/download
LAS X Life Science	Leica Microsystems	https://www.leica-microsystems.com/products/microscope-software/p/leica-las-x-ls/
Leica Application Suite v3.4.2.18368	Leica Microsystems	https://www.leica-microsystems.com/
Molecular Imager® Gel DocTM XR Image LabTM	BIO-RAD	https://www.bio-rad.com/es-es/sku/1708195-molecular-imager-gel-doc-xr-system-with-image-lab-software?ID=1708195
Morpheus	Broad Institute	https://software.broadinstitute.org/morpheus/
Odyssey Infrared Imaging System (v3.0.29)	LI-COR	n/a
Samtools v0.1.18	(Li et al., 2009)	https://github.com/samtools/samtools
Serial Cloner v2.6.1	Serial Cloner	http://serialbasics.free.fr/Serial_Cloner.html
StepOne v2.3	Thermo Fisher	https://www.thermofisher.com/es/es/home/life-science/pcr/real-time-pcr/
TopHat v2.1.1	Johns Hopkins Univ.	https://ccb.jhu.edu/software/tophat/index.shtml

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7. RESULTS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7.1 IDENTIFICATION OF ZMYM2 AS A POTENTIAL NOVEL REGULATOR OF PLURIPOTENCY

While several zinc finger proteins have been involved in regulating gene expression programs that define pluripotent cell identity (see Introduction section), the functional roles of zinc finger MYM-type (a.k.a., ZMYM) proteins in pluripotency are largely unexplored. Importantly, a protein member of this family, ZMYM2, was previously identified as an interacting partner of NANOG[287] and OCT4[77], which are central to the maintenance of ESC pluripotency. Therefore, these pieces of evidence suggest that ZMYM2 could be involved in the molecular control of pluripotency. Hence, I was interested to explore the potential role of ZMYM2 in the pluripotent cellular state. I first determined the expression levels of each ZMYM member (i.e., *Zmym1*, *Zmym2*, *Zmym3*, *Zmym4*, *Zmym5,* and *Zmym6*) in somatic cells and pluripotent cells. To this end, I took advantage of available RNA-seq datasets of mouse embryonic fibroblasts (a.k.a., MEFs) as well as mouse ESCs and iPSCs[226,339,340]. Notably, I found that *Zmym1*, *Zmym2,* and *Zmym3* showed a similar gene expression pattern, including much higher enrichment in pluripotent cells than in somatic MEFs (**Figure R1a**), suggesting a potential role of these 3 members of the ZMYM family on pluripotent cell identity.

Given that the molecular control of the pluripotent state has been shown to reside in a core of master transcription factors including OCT4, SOX2, and NANOG[6,35,69], I next investigated the existence of a physical association between the ZMYM proteins and the OSN core pluripotency network. Taking advantage of previous studies reporting high confident interactomes of OCT4[77], SOX2[80] and NANOG[78,79] in mouse ESCs (**Figure R1b,** left), I was able to detect as expected the physical association between ZMYM2 with OCT4 and NANOG, but I did not observe a partnership with SOX2 (**Figure R1b,** right). Importantly, ZMYM2 was the only member of the ZMYM family able to interact with some members (i.e., OCT4 and NANOG) of the core pluripotency network (**Figure R1b,** right).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

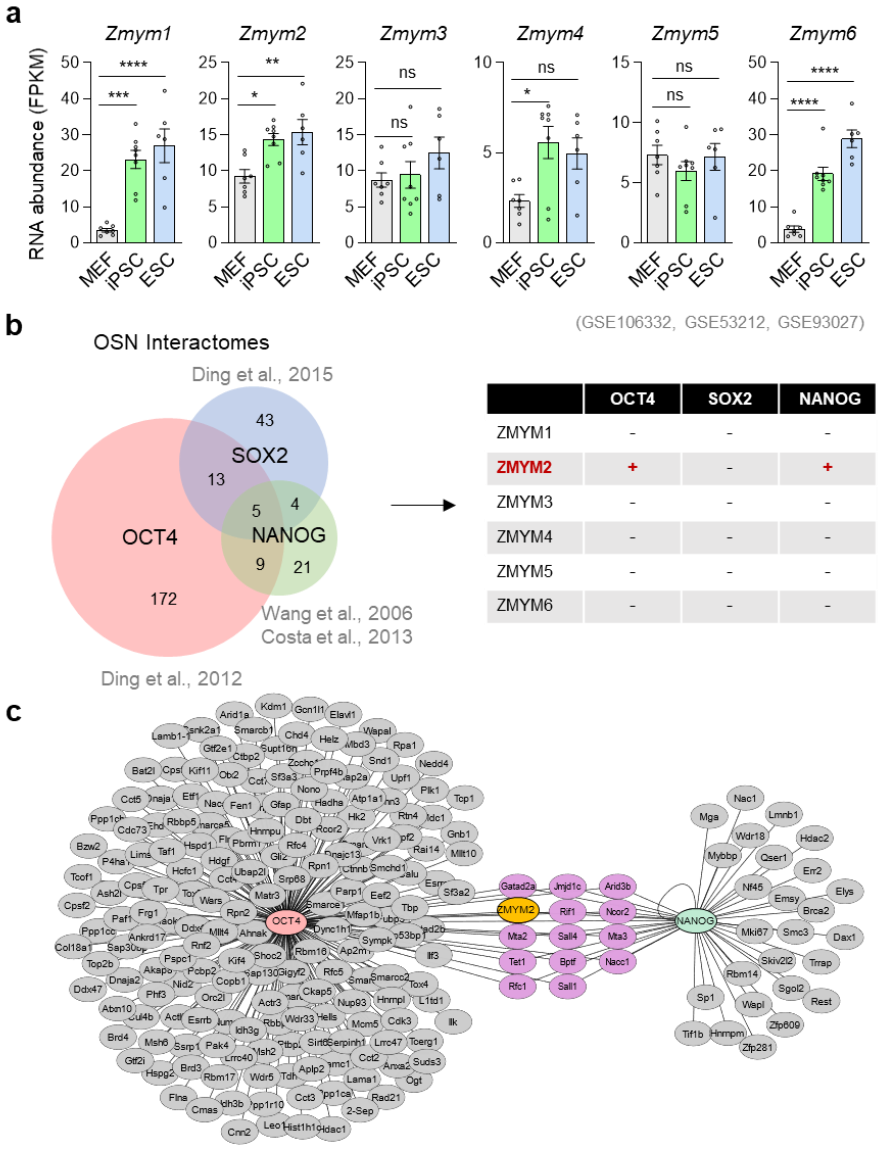


Figure R1. ZMYM2 is present in the OCT4-NANOG protein interaction network in ESCs. a. RNA levels of the MYM-type zinc finger family members in somatic cells (MEFs) and different pluripotent cell lines (iPSCs and ESCs) expressed as FPKM. RNA-seq data were obtained from published datasets (GSE106332, GSE532112, GSE93027). One-way ANOVA test with Kruskal-Wallis correction for non-parametric distributions was performed. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; n.s., non-significant. Data is shown as the mean ± SD and each data point represents biological replicates.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

b. (*Left*) Venn diagram showing the overlap among published interactomes of master regulators of pluripotency. (*Right*) Table summarizing the interactions of each member of the zinc finger MYM-type family with OCT4, SOX2, and NANOG. **c**. Interactome showing common binding partners of OCT4 (red) and NANOG (green). Data were taken from published interactomes for both OCT4 and NANOG. Interactome was done using Cytoscape software (see Material & Methods section). The interactomes for OCT4, SOX2, and NANOG used in **b** and **c** were obtained from the previous studies[77,79,80,287].

Then, I performed the integration of OCT4 and NANOG interactomes using the Cytoscape computational tool and I found ZMYM2 together with other 13 proteins, including transcription factors and epigenetic regulators (i.e., ARIB3b, BPTF, GATAD2a, JMJD1c, MTA2, MTA3, NACC1, NCOR2, RFC1, RIF1, SALL1, SALL4, and TET1). The presence of ZMYM2 among well-documented regulators that play important roles in pluripotency and reprogramming, such as the methylcytosine hydroxylase TET1[79,348], and MTA2/3 and GATAD2b members of the NuRD repressor complex[349,350] among others, prompted me to consider ZMYM2 as a potential novel pluripotent regulator.

Taken together, in this Ph.D. work I decided to focus on ZMYM2 as the ZMYM family member with potential regulatory functions in pluripotent cell identity.

7.2 DECIPHERING THE ROLE OF ZMYM2 IN PLURIPOTENCY

7.2.1 Characterization of *Zmym2*-null ESCs

To investigate the functional significance of ZMYM2 in regulating pluripotency, I first performed loss-of-function approaches. To this end, we obtained from Dr. Wang's group the homozygous gene-trap *Zmym2* (hereafter "*Zmym2*$^{GT/GT}$") mutant mouse ESCs and wild-type (WT) mouse ESCs (hereafter "*Zmym2*$^{+/+}$"), derived from the intercrosses of heterozygous *Zmym2*$^{GT/+}$ mice. The mutant *Zmym2* ESC line contains a gene trap (GT) β-galactosidase/neomycin phosphotransferase (termed βGeo) reporter/selection cassette between exons nine and ten of both *Zmym2* alleles (**Figure R2a**). Thus, the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

encoded transcript leads to a potential premature truncated form of ZMYM2 protein with disruption of the MYM zinc finger-containing domain (**Figure R2a**). First, I wanted to confirm the status of these WT and mutant ESC lines using a specific PCR genotyping strategy (**Figure R2b**). Next, I analyzed whether *Zmym2* expression could be affected in the *Zmym2*$^{GT/GT}$ mutant ESC line. Importantly, the *Zmym2* transcripts were absent in *Zmym2*$^{GT/GT}$ ESCs using an RT-qPCR assay (**Figure R2c**). To further confirm the knockout status of these mutant ESCs, I analyzed the protein expression levels of ZMYM2 by western blot using a commercial antibody against its N-terminal domain. In line with my RNA expression data, analysis of nuclear (**Figure R2d**) and whole-cell protein extracts (**Figure R2e**) from WT and mutant ESCs confirmed that *Zmym2*$^{GT/GT}$ ESCs are null for ZMYM2 protein expression.

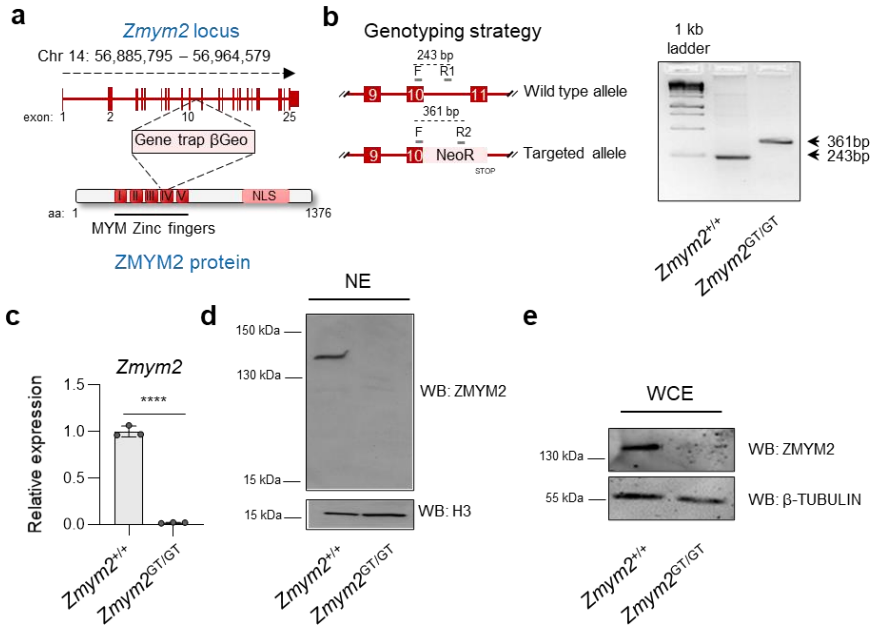


Figure R2. Strategy for *Zmym2*$^{GT/GT}$ ESC line validation. a. Schematic depiction of the *Zmym2* allele for the gene trap (GT) knockout strategy. A genetrap and neomycin cassette was introduced between exons 10 and 11 (*top*) resulting in a truncated protein in the zinc-finger domain (*bottom*). **b.** (*Left*) Schematic depiction of the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

genotyping strategy by PCR. A common forward primer was designed against the exon 10, whereas a specific reverse primer was designed against the intronic region located between exons 10 and 11 for both wild-type and knockout conditions. Expected band sizes for each condition are shown. (*Right*) DNA fragments of amplified alleles by PCR confirming the genetrap insertion for the generation of the *Zmym2*$^{GT/GT}$ cell line. **c**. Relative expression of *Zmym2* normalized against U6 in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cell lines as measured by RT-qPCR. One-tailed Student's t test was performed. ****p<0.0001. Data is shown as the mean ± SD and each data point represents technical replicates. **d**. ZMYM2 protein levels of nuclear extracts from *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESC lines measured by western blot. H3 was used as a loading control. e. ZMYM2 protein levels of whole cell extracts from *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESC lines measured by western blot. ß-TUBULIN was used as a loading control.

7.2.2 ZMYM2 is required for proper control of ESC maintenance

To address if the loss of ZMYM2 affects the self-renewal capacity of pluripotent cells I employed several classical approaches in the field. First, I analyzed the ESC morphology and performed staining against alkaline phosphatase (AP), which is a typical marker of pluripotency[333]. I found that similar to WT ESCs, *Zmym2*$^{GT/GT}$ null ESCs maintain characteristic ESC morphology and are stained positive for AP activity under standard (i.e., presence of serum and LIF) and feeder-free pluripotent culture conditions (**Figure R3a**), indicating null ESCs maintain an undifferentiated self-renewal state. Second, I analyzed the proliferation of ESCs over a period of 8 days and I did not observe a significant difference in the proliferation rates of *Zmym2*$^{GT/GT}$ ESCs relative to WT ESCs (**Figure R3b**). Third, cell cycle profile analysis showed no significant differences in the cell cycle distribution phases (i.e., G1, S, and G2-M) of WT and null ESCs (**Figure R3c-d**). Furthermore, I assessed cell viability by MTT assay. Again, I observed that the percentage of viable cells for 3 days was similar to that of WT and null ESCs (**Figure R3e**), suggesting that the loss of ZMYM2 does not affect ESC survival. Taken together, these results suggest that ZMYM2 is dispensable for the maintenance of murine pluripotency.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

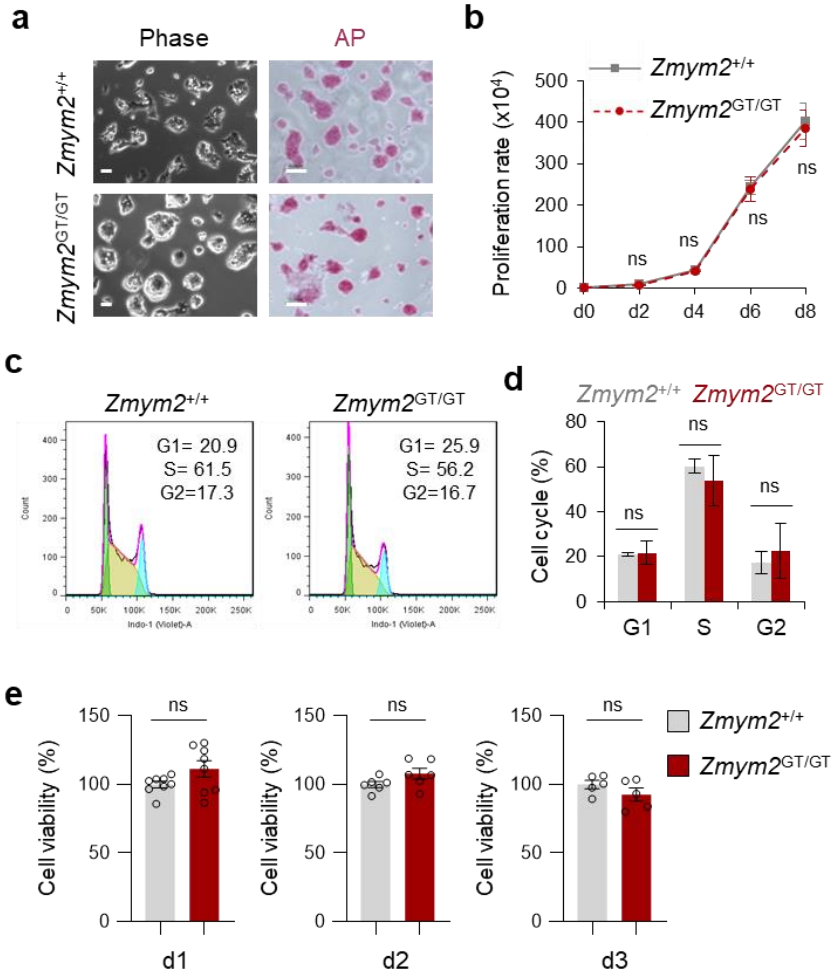


Figure R3. Loss-of-function of ZMYM2 is dispensable for ESC maintenance. a. Representative images of visible microscopy in phase contrast in ESCs in the presence (*Zmym2*$^{+/+}$) and absence of ZMYM2 (*Zmym2*$^{GT/GT}$). Pictures in bright field (*left*) and after AP staining of cells (*right*) are shown. Scale bar indicates 100µm. **b**. Results from a proliferation assay performed in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (red) cells for 8 days shown as number of cells (x10^4) measured at the indicated time points. **c**. Cell cycle profile performed by flow-cytometry in *Zmym2*$^{+/+}$ (*left*) and *Zmym2*$^{GT/GT}$ cells (*right*). Percentage of cells in each phase (G1, S and G2) are shown. **d**. Bar graph showing the percentage of cells present in each phase of cell cycle (G1, S and G2) in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells determined as in **c**. **e**. Percentage of cell viability as measured by MTT assay in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (red) cells at the indicated time points. In b, Student's t test for multiple comparisons was performed. In d, two-way ANOVA test was performed. In e, two-tailed Student's t

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

test was performed, and each data point represents technical replicates. ns, non-significant. In b, data is shown as the mean ± SEM of three independent experiments, whereas in d-e, as the mean ± SD.

To address whether loss of ZMYM2 might affect the expression of pluripotency and lineage-specific genes, I analyzed the expression levels by RT-qPCR of selected pluripotency genes and several germ layers (i.e., ectoderm, mesoderm, and endoderm) markers in WT and *Zmym2*$^{GT/GT}$ ESCs (**Figure R4a**). I found the expression of pluripotency factors (i.e., *Oct4*, *Sox2*, *Klf4*, and *Esrrb*) to be unaffected except for *Nanog* and *Zfp42* (a.k.a. *Rex1*) that were significantly upregulated and downregulated, respectively, in *Zmym2*$^{GT/GT}$ null ESCs (**Figure R4a**). Consistent with previous reports showing that NANOG acts by restricting the endoderm lineage specification in mouse ESCs[21,22], while ZFP42 favors this specific lineage[351], I found robust downregulation of several primitive endoderm markers (i.e., *Gata4*, *Gata6*, and *Sox17*), but no significant differences were observed in the markers of ectoderm (i.e., *Otx2*, *Fgf5*, and *Pax3*) or mesoderm (i.e., *Mixl1*, *Cxcl12*, and *T*) in absence of ZMYM2 (**Figure R4a**). These findings suggest that loss of ZMYM2 leads to aberrant downregulation of primitive endoderm markers in ESCs that could be due to a secondary effect upon ZMYM2 loss (e.g., the dysregulation of pluripotency genes such as *Nanog* and *Zfp42* among others).

The pronounced upregulation of *Nanog* RNA expression (>2-fold) in *Zmym2*$^{GT/GT}$ null ESCs (**Figure R4a**) and since this homeobox gene was discovered based on its ability to drive LIF-independent self-renewal of pluripotent mouse ESCs[21,70], prompted me to examine the self-renewal properties of these knockout pluripotent cells in more detail. I first confirmed the upregulation of NANOG by western blot analysis in the knockout cell line (**Figure R4b**). Then, taking in consideration that NANOG is heterogeneously expressed in ESCs and its levels in individual mouse pluripotent cells influence their self-renewal versus differentiation propensity[37], I performed NANOG immunostaining in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. In line with

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

previous reports, I also observed the heterogeneous expression pattern of NANOG in WT ESCs (**Figure R4c**, top panels), and more importantly, I found that ZMYM2 depletion did not alter this heterogeneous pattern of NANOG (**Figure R4c**, bottom panels). Quantitative immunofluorescence analysis of single cells confirmed significant upregulation of NANOG upon ZMYM2 depletion in ESCs cultured under standard conditions (i.e., presence of serum and LIF) (**Figure R4d**). These data suggest that ZMYM2 likely fine-tunes *Nanog* gene activation in preexisting *Nanog*-expressing cells for proper control of ESC maintenance *in vitro*. Then, I explored this possibility by measuring the self-renewal of ESCs at the single-cell level. To this end, a reduced number of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were plated at clonal density and cultured in the presence and absence of LIF (**Figure R4e**). After 6 days of culture, AP staining was performed and colonies were classified as undifferentiated, partially differentiated, and fully differentiated according to AP staining patterns. In the presence of LIF, I observed statistically more undifferentiated colonies and less partially differentiated and fully differentiated colonies in *Zmym2*$^{GT/GT}$ ESCs than *Zmym2*$^{+/+}$ ESCs (**Figure R4f**, left panel). This result suggests that self-renewal of *Zmym2* null ESCs were maintained and even enhanced under standard ESC culture conditions. Remarkably, colony formation assay indicated that *Zmym2* null ESCs yielded significantly less fully differentiated colonies and, more importantly, showed the presence of undifferentiated colonies in differentiation-promoting (i.e., absence of LIF) culture conditions (**Figure R4f**, right panel). This data suggests that ZMYM2 might play a pivotal role during the exiting of pluripotency.

Together, my results demonstrate that loss of ZMYM2 is dispensable for maintenance of ESCs under standard culture conditions (i.e., presence of serum and LIF) and may negatively regulate the self-renewal state of pluripotent cells. In sum, my data support that ZMYM2 represents a novel factor involved in the proper control of ESC maintenance.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

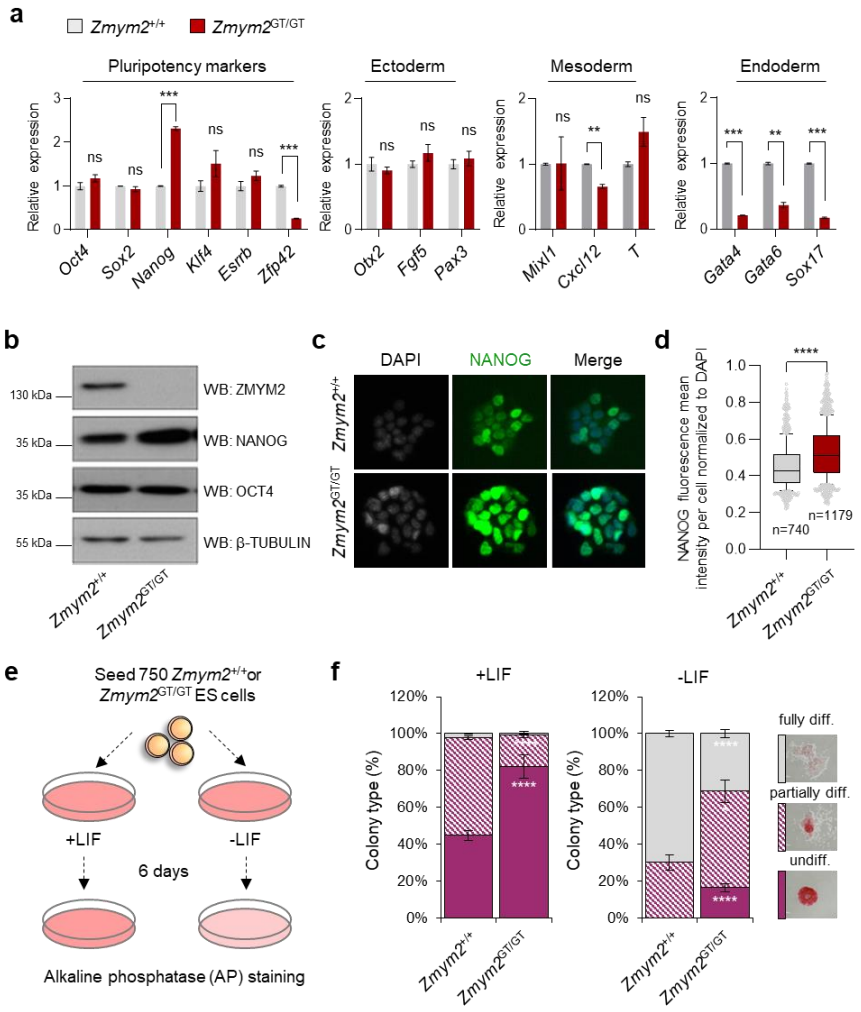


Figure R4. Depletion of *Zmym2* in ESCs causes dysregulation of pluripotency and lineage-specific markers and enhances self-renewal. a. RT-qPCR analysis for pluripotent and germ layer (ectoderm, mesoderm, and endoderm) markers in indicated ESC lines. Values were normalized against *U6*. **b**. Protein levels of two master regulators of pluripotency (NANOG and OCT4) in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells were measured by western blot. ß-TUBULIN was used as a loading control and ZMYM2 as cell line validation. **c**. Representative images of single-cell NANOG immunofluorescence performed in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. **d**. Quantification of the NANOG mean fluorescence intensity from **c**. Values were normalized with DAPI mean fluorescence. **e**. A schematic overview of the colony formation assay performed in **f**. 750 wild-type (*Zmym2*$^{+/+}$) or knockout (*Zmym2*$^{GT/GT}$) cells were

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

seeded in the presence or absence of LIF and colonies were stained for AP activity at day 6 of culture in specified conditions. **f**. Percentage of colonies that arose after culturing indicated ESC lines in the presence (*left*) or absence of LIF (*right*) for 6 days. Colonies were classified in differentiated (grey), partially differentiated (purple and white pattern), or undifferentiated (purple). In **a** and **f**, two-tailed Student's t-test was performed, and biological triplicates were analyzed. Data are represented as mean ± SD. In **d**, a one-tailed Student's t-test was performed (n=number of cells used for each condition). Whiskers represent the 10^{th}-90^{th} percentiles. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; n.s., non-significant.

7.2.3 ZMYM2 is required for the establishment and maintenance of primed pluripotency

Others have demonstrated distinct transcriptional landscapes in both naïve and primed pluripotent states[86,352,353]. My results demonstrate that the absence of ZMYM2 led to increased levels of *Nanog* (**Figure R4a-R4d**), along with an increase in the proportion of cells retaining an undifferentiated state even under differentiating cues (**Figure R4f**), typical of the ground-state of pluripotency. Since Epiblast Stem Cells (termed EpiSCs) are known to be in a more advanced stage of development, I asked if *Zmym2* depletion is compatible with the primed state of pluripotency. First, I characterized the expression levels of *Zmym2* in EpiSCs compared to ESCs. No significant differences could be observed between these cell lines (**Figure R5a**, top left), suggesting that ZMYM2 may be important for the regulation of the primed pluripotency. Additionally, expression levels of several pluripotency markers were analyzed. Previous reports have identified transcriptional changes in the expression of pluripotency markers, with *Oct4* being consistently expressed[76], whereas a clear downregulation of *Nanog* levels is concomitant with the restriction in cellular potency in the primed pluripotent state[70]. Hence, as expected, *Oct4* levels were maintained at similar levels in both ESCs and EpiSCs (**Figure R5a**, top right panel), whereas expression levels of specific markers changed between these two conditions. A severe downregulation of *Nanog* levels was accompanied by an increase in the expression levels of *Fgf5* (a typical marker for the primed pluripotency) (**Figure R5a**, bottom panel). To directly test the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

function of ZMYM2 during the transition from metastable to the alternative primed pluripotent state, *Zmym2* WT, and null ESCs were cultured in a serum-free medium containing FGF2 and Activin A (FA medium), which allowed them to transit into a primed epiblast-like cell (termed EpiLC) state (**Figure R5b**). Then, expression levels of naïve and primed markers were analyzed by RT-qPCR. As expected, WT ESCs maintained similar levels of the naïve and primed marker *Oct4* during the acquisition of the primed pluripotency (**Figure R5c**, top panel). Furthermore, gradual downregulation of the naïve marker *Nanog* could be observed, in accordance with a restriction in the pluripotency capacity of cells cultured under these primed defined culture conditions (**Figure R5c**, middle panel). Concomitant with this, a gradual upregulation of the primed marker *Fgf5* could also be observed (**Figure R5c**, bottom panel). Remarkably, within 3 days upon ZMYM2 depletion, EpiLCs upregulate *Nanog* naïve marker and downregulate *Fgf5* primed marker (**Figure R5c**). Taken together, these results suggest that ZMYM2 is important for the establishment of primed pluripotency.

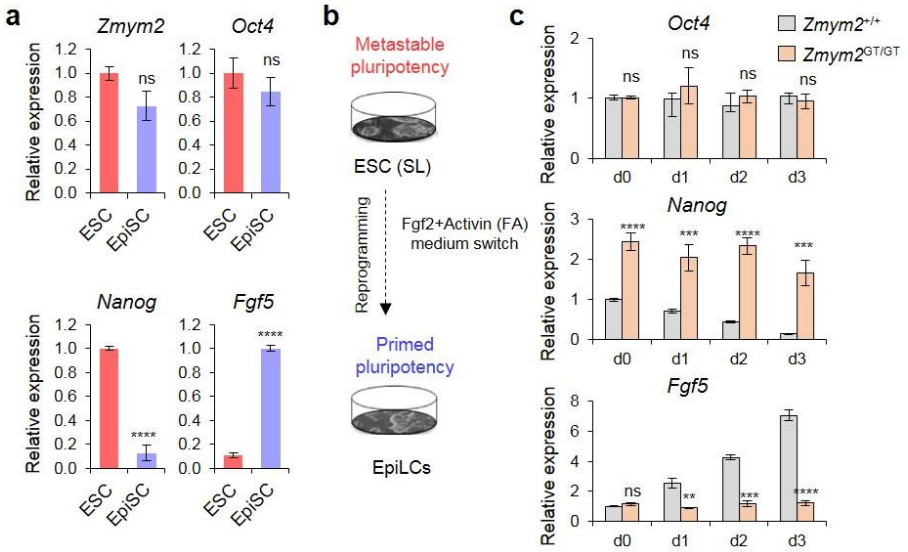


Figure R5. Absence of ZMYM2 compromises the proper transition from ESCs to EpiLCs. a. Relative expression levels of *Zmym2*, different pluripotency markers

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(*Oct4*, *Nanog*) and the primed pluripotency marker *Fgf5* in ESCs and EpiSCs measured by RT-qPCR. Values were normalized against *Gapdh*. **b**. Scheme summarizing in vitro metastable to primed pluripotent cell fate conversion by culturing ESCs in serum-free medium containing FGF2 and Activin A (FA medium). **c**. Relative expression levels of *Oct4*, *Nanog*, and *Fgf5* in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells were measured by RT-qPCR at the indicated time points. Values were normalized against *Gapdh*. In **a**, two-tailed Student's t-test was performed. In **c**, t-test for multiple comparisons was performed. Data are represented as mean ± SD. ***p*<0.01; ****p*<0.001; *****p*<0.0001; n.s., non-significant.

Next, to directly test the function of ZMYM2 in the primed pluripotent transition, I cultured *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs at low density in FA medium for 6 days and examined whether self-renewal properties of primed pluripotent state were affected by the absence of ZMYM2 (**Figure R6a**). Colony-formation assay indicated that *Zmym2*$^{GT/GT}$ cells yielded significantly more naïve compact dome-shaped colonies under FA culture conditions (**Figure R6b** and **R6c**). In contrast, the formation of flat diffuse-shaped colonies characteristic of primed pluripotency was severely compromised by *Zmym2* depletion (**Figure R6b** and **R6c**). Moreover, *Zmym2*-null ESCs showed a significantly reduced cellular metabolic activity examined with MTT assay, which is an indicator of cell viability and/or proliferation, under FA conditions compared with their wild-type counterparts (**Figure R6d**). I also found that *Zmym2*$^{GT/GT}$ cells showed significant cell cycle changes compared to *Zmym2*$^{+/+}$ cells cultured under primed pluripotent conditions (**Figure R6f**). In this sense, an elongated G1 phase was accompanied by a shortening of the S and G2-M phases (**Figure R6f**). Consequently, an increase in the expression levels of p16 and p21, two known cell cycle inhibitors that are involved in cell cycle phase transitions[354,355], was detected in absence of *Zmym2* (**Figure R6g**). Consistently, this elongation of cell cycle was in line with the reduction in *Zmym2*-null cell proliferation rates under FA culture conditions (**Figure R6h**). Consistently, I found that EpiLCs could not be stably maintained in the primed state in absence of ZMYM2 (**Figure R6i**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

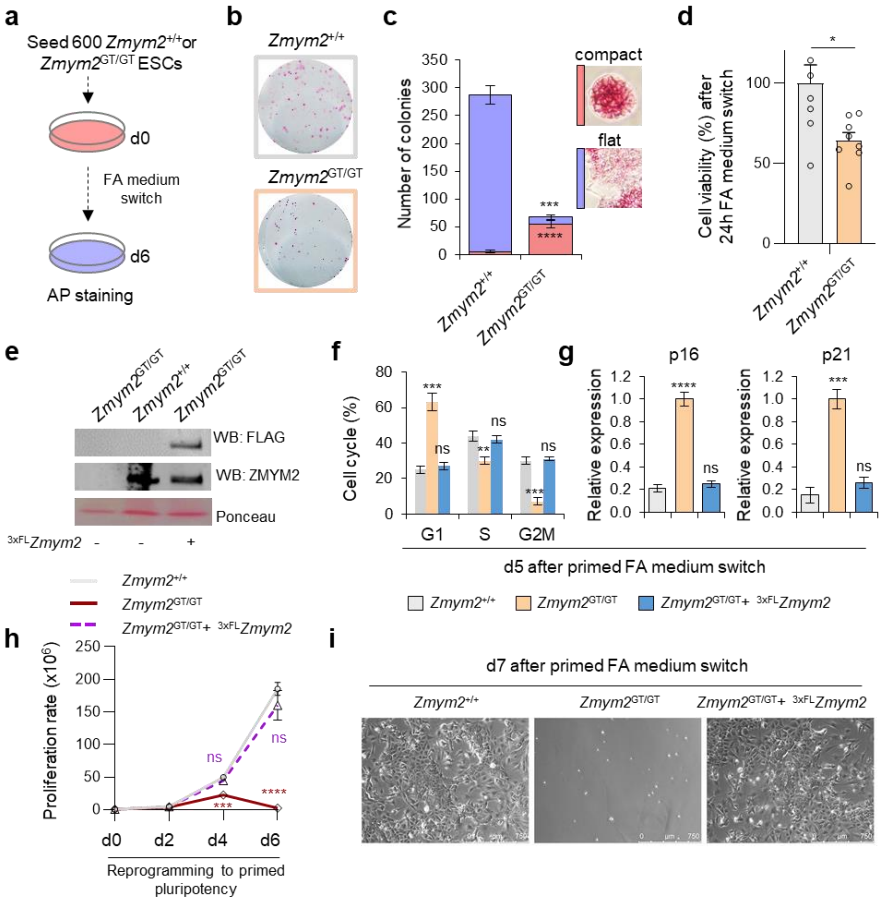


Figure R6. ZMYM2 is critical for primed pluripotency. a. Scheme summarizing the colony-formation assay protocol followed for the analysis of self-renewal properties under primed pluripotent (FA) conditions. 600 *Zmym2*$^{+/+}$ ESCs or *Zmym2*$^{GT/GT}$ ESCs were cultured in FA medium and AP staining was performed at day 6. **b**. Representative images of AP-stained colonies formed after culture conditions specified in **a**. **c**. The number of AP+ colonies obtained after culturing cells as indicated in **a**. Colonies were classified as naïve compact (pink) and primed flat (light blue). **d**. Percentage of cell viability as measured by MTT assay of *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ cells (light orange) after 24 hours of culture under primed conditions. Each data point represents technical replicates. **e**. Western blot against FLAG and ZMYM2 in knock-out cells transfected with a ^{3xFL}*Zmym2* vector. Wild-type (*Zmym2*$^{+/+}$) and knock-out (*Zmym2*$^{GT/GT}$) extracts were used as controls, and Ponceau staining was performed as a loading control. **f**. Percentage of cells present at each phase of the cell cycle (G1, S, G2) as measured by flow-cytometry in *Zmym2*$^{+/+}$ (grey), *Zmym2*$^{GT/GT}$ (light orange), and ^{3xFL}*Zmym2*$^{GT/GT}$ (blue) cells cultured under primed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

conditions for 5 days. **g**. Relative expression levels of cell cycle regulators (i.e., p16 and p21) in *Zmym2*$^{+/+}$ (grey), *Zmym2*$^{GT/GT}$ (light orange), and ^{3xFL}*Zmym2*$^{GT/GT}$ (blue) cells were cultured under primed conditions for 5 days. **h**. Results from a proliferation assay performed in *Zmym2*$^{+/+}$ (grey), *Zmym2*$^{GT/GT}$ (red) and ^{3xFL}*Zmym2*$^{GT/GT}$ cells (purple) after 6 days of culture under primed conditions. Data are shown as the number of cells (x10^6) measured at the indicated time points. **i**. Contrast phase microscopy representative images of *Zmym2*$^{+/+}$, *Zmym2*$^{GT/GT,}$ and ^{3xFL}*Zmym2*$^{GT/GT}$ cells cultured in FA medium for 7 days. Scale bar indicates 750µm. In **c**, **f,** and **h**, a two-way ANOVA test was performed. In **d**, two-tailed Student's t-test was performed, and in **g**, one-way ANOVA was performed. All data are represented as mean ± SD from biological triplicates. **p*<0.05; ***p*<0.01; ****p*<0.001; *****p*<0.0001; n.s., non-significant.

To test if these results are a direct consequence of the absence of ZMYM2 in ESCs, a rescue experiment was performed. Thus, a rescue cell line (hereafter, *Zmym2*$^{GT/GT}$ + ^{3xFL}*Zmym2*) was generated by the transfection of a full-length ZMYM2 protein into *Zmym2*-null ESCs. Validation of the rescue cell line was performed by analyzing ZMYM2 protein levels by western blot using antibodies against FLAG tag and ZMYM2 (**Figure R6e**). Importantly, I observed that ectopic expression of ZMYM2 was sufficient to promote primed features during the standard (i.e., SL) to primed (i.e., FA) transition (**Figures R6f-i)**. First, cell cycle profile analyses showed their complete restoration of the cell cycle distribution phases (i.e., G1, S and G2-M) when compared to WT ESCs (**Figure R6f**), along with the restoration of the expression levels of p16 and p21 cell cycle inhibitors (**Figure R6g**). Second, significant differences in the proliferation rates of the *Zmym2*$^{GT/GT}$ + ^{3xFL}*Zmym2* and *Zmym2*$^{+/+}$ cell lines could not be observed (**Figure R6h**). Finally, ectopic expression of ZMYM2 allowed the maintenance of EpiLCs under primed conditions, even for extended periods (**Figure R6i**). Taken together, my data demonstrate that ZMYM2 functions as a key regulator for the establishment and maintenance of primed pluripotency.

7.2.4 ZMYM2 is required for exiting pluripotency *in vitro*

As mentioned before, primed pluripotent cells are characterized by a decreased potency in comparison to ground-state cells[31]. Since my

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

data show that ZMYM2 is important for both establishment and maintenance of a more advanced developmental state (i.e., the primed state), I asked whether *Zmym2* depletion could affect the cellular differentiation process of ESCs gown under standard pluripotent culture conditions (i.e., SL). To this end, I first differentiated *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs to embryoid bodies (EBs) (**Figure R7a**), which mimic early embryo development in a dish[334]. In contrast to the expression of *Oct4* over a nine-day period of EB formation, I observed that the *Zmym2* expression was upregulated (**Figure R7b**), suggesting that the function of ZMYM2 could be important to exit pluripotency. Notably, my results showed that *Zmym2* null ESCs were able to form EBs (**Figure R7c**), however, these EBs were significantly smaller than their wild-type (WT) counterparts (**Figure R7c** and **R7d**). When looked in more detail, *Zmym2* null ESCs produced EBs that appeared to be developmentally delayed compared to EBs from WT ESCs (**Figure R7c**). Since proper downregulation of self-renewal regulators such as NANOG and SOX2 is necessary for ESCs to differentiate into all cell types of the three germ layers (i.e., endoderm, mesoderm, and ectoderm), I asked if the delayed differentiation of *Zmym2* null ESCs might be due to insufficient downregulation of *Nanog* and *Sox2* in these cells. Indeed, my data revealed that the expression levels of these two core pluripotent regulators were higher in *Zmym2*$^{GT/GT}$ compared to *Zmym2*$^{+/+}$ EBs over the 9 days of EB differentiation (**Figure R7e**). It is also noteworthy that endoderm (i.e., *Afp* and *Gat4*) and mesoderm (i.e., *T* and *Mixl1*) markers were not adequately induced in *Zmym2* null EBs throughout the time course of differentiation (**Figure R7e**). In contrast, the absence of ZMYM2 leads to the upregulation of ectoderm markers (i.e., *Fgf5* and *Nestin*) during EB formation (**Figure R7e**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

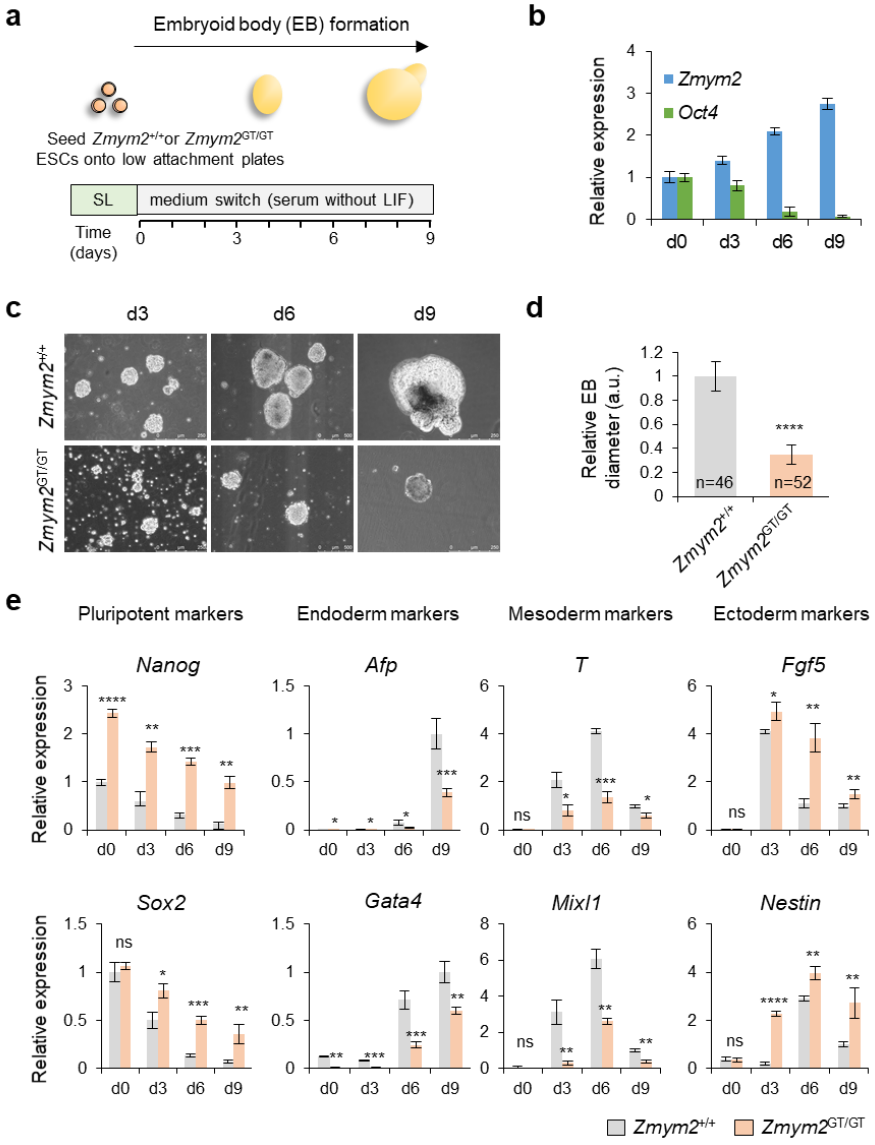


Figure R7. ZMYM2 deficient ESCs show abnormal EB differentiation. a. Scheme summarizing the EB formation assay for *in vitro* differentiation. Wild-type (*Zmym2*$^{+/+}$) and knockout cells (*Zmym2*$^{GT/GT}$) were cultured in suspension with serum-containing medium without LIF for 9 days. **b.** Relative expression analysis of *Zmym2* (blue) and *Oct4* (green) during the EB formation assay in wild-type cells as measured by RT-qPCR at the indicated timepoints. **c.** Representative contrast images of EBs formed under wild-type (*Zmym2*$^{+/+}$) and knock-out (*Zmym2*$^{GT/GT}$) conditions at the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

indicated timepoints. **d**. Relative EB diameter in arbitrary units (a.u.) measured as the area under the curve in *Zmym2*$^{+/+}$ (light grey) and *Zmym2*$^{GT/GT}$ (light red) ESCs. **e**. Relative expression of pluripotent (*Nanog* and *Sox2*) and differentiation markers (*Afp, Gata4, T, Mixl1, Fgf5,* and *Nestin*) during EB differentiation assay in wild-type (grey) and *Zmym2* knockout (light red) conditions at the indicated timepoints. Values were normalized against *Gapdh*. Biological triplicates were analyzed. In **d**, a two-tailed Student's t-test was performed, and in **e**, a t-test for multiple comparisons was performed. **p*<0.05; ***p*<0.01; ****p*<0.001; *****p*<0.0001; n.s., non-significant.

Next, I explored the possibility that *Zmym2* null ESCs may be skewed toward ectoderm differentiation and defective in mesendoderm differentiation during *in vitro* differentiation. For this purpose, I used an early differentiation *in vitro* assay (**Figure R8a**) to rapidly induce the expression of genes involved in mesendoderm (ME) or neural ectoderm (NE) lineage choice[335]. To this end, I first cultured *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs in N2B27 medium for a couple of passages to homogenize the cell cultures. Then, cells were cultured for 2 days in N2B27 basal medium supplemented with LIF, BMP, and MEK inhibitor. Finally, cells were seeded and cultured in the presence of CHIR inhibitor for ME specification or retinoic acid (RA) for NE specification, respectively (**Figure R8a**). First, my data showed that the relative expression levels of *Zmym2* in WT cells were constant over the 5-day period of ME and NE specification, supporting its potential role for lineage choice specification (**Figure R8b**). Notably, although I did not observe differences in *Oct4* expression levels between *Zmym2* WT and null ESCs during ME and NE specification, the lack of ZMYM2 led to sustained upregulation of *Nanog* and *Sox2*, which are reported to repress mesendoderm[356,357] and activate neural ectoderm[358] cell fate choices, respectively (**Figure R8c**). Consistent with this and in line with my previous EB differentiation assay data (**Figure 7e**), I found that regulators for commitment into mesendoderm (i.e., *Eomes* and *Foxa2*) were downregulated in *Zmym2*$^{GT/GT}$ cells (**Figure R8d**). In contrast, on day 5 of NE specification, I observed a significant upregulation of regulators that favor neural ectoderm cell fate commitment (i.e., *Nestin* and *Pax3*) (**Figure R8d**). Taken together, my data further suggest that ZMYM2 influences early cell fate specification during *in vitro*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

differentiation, revealing an important functional role of ZMYM2 in controlling neural ectoderm and mesendoderm differentiation in ESCs.

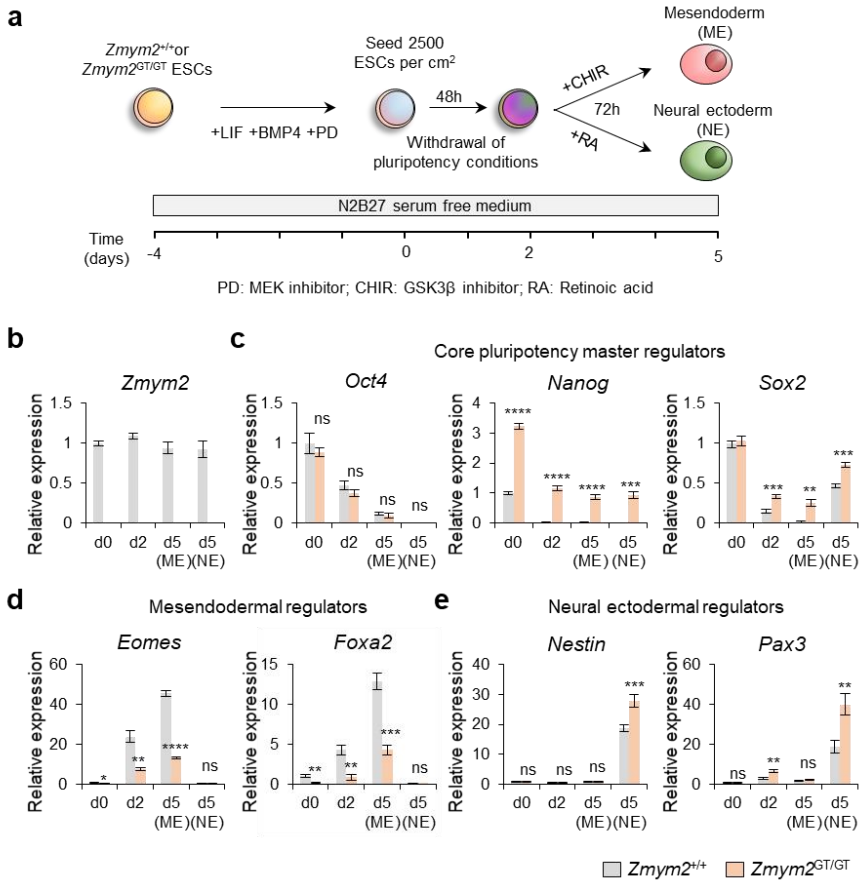


Figure R8. ZMYM2 depleted ESCs are skewed toward neural ectoderm differentiation and defective in mesendoderm differentiation. a. Scheme summarizing the *in vitro* mesendoderm (ME) and neural ectoderm (NE) specification assay. *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were cultured in N2B27 medium in the presence of BMP4, LIF, and MEK inhibitor (PD0325901). After 4 days, the medium was changed to N2B27 medium and maintained for an extra two-day period. Then, germline specification was promoted by the addition of RA or GSK3 inhibitor (CHIR99021) for 5 days for neural ectoderm and mesendoderm specification, respectively. **b**. *Zmym2* relative expression analysis during the germ layer specification assay at the different timepoints as measured by RT-PCR. **c**-**e**. Relative expression of core pluripotency (**c**), mesendodermal (**d**), and neural ectoderm (**e**) markers during germ layer specification in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells as

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

measured by RT-qPCR at the indicated timepoints. Values from **b**, **c**, and **d** were normalized against *Gapdh*. Biological triplicates were analyzed. Data are represented as mean ± SD, and a t-test for multiple comparisons was performed. **p*<0.05; ***p*<0.01; ****p*<0.001; *****p*<0.0001; n.s., non-significant difference.

7.3 ANALYZING THE ROLE OF ZMYM2 IN REPROGRAMMING TO INDUCED PLURIPOTENCY

7.3.1 ZMYM2 depletion promotes iPSC generation

Given that my results show that lack of ZMYM2 enhances the self-renewal of ESCs and compromises the exit from pluripotency, I hypothesized that it may also be involved in the re-acquisition of pluripotency during somatic cell reprogramming (SCR). To investigate whether ZMYM2 participates in the SCR process, I examined the effect of transient suppression of ZMYM2 in the acquisition of the pluripotent cell identity using MEFs harboring a green fluorescent protein (GFP) reporter under the control of the *Oct4* gene transduced with the reprogramming factor cocktail (OCT4, SOX2, KLF4, and cMYC; OSKM) (**Figure R9a**). In line with my hypothesis, I found that downregulation of *Zmym2* using three independent validated shRNAs (**Figure R9b**) led to a significant increase in the number of iPSC colonies, scored both by morphology and GFP expression, compared with a control shRNA targeting luciferase (sh*Luci*) (**Figure R9b**). Furthermore, knockdown of ZMYM2 resulted in accelerated kinetics of iPSC formation over the 21-day period of the SCR process, as shown by increasing percentages of GFP-positive colonies at earlier stages (**Figure R9c**). Downregulation of *Zmym2* could potentially affect the proliferation of MEFs, which would confound the calculation of reprogramming efficiency. However, no significant changes in MEF growth rates were found between sh*Luci* and shRNAs against *Zmym2* (**Figure R9d**). In summary, my data show that ZMYM2 functions as a reprogramming barrier during the SCR process to form iPSCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

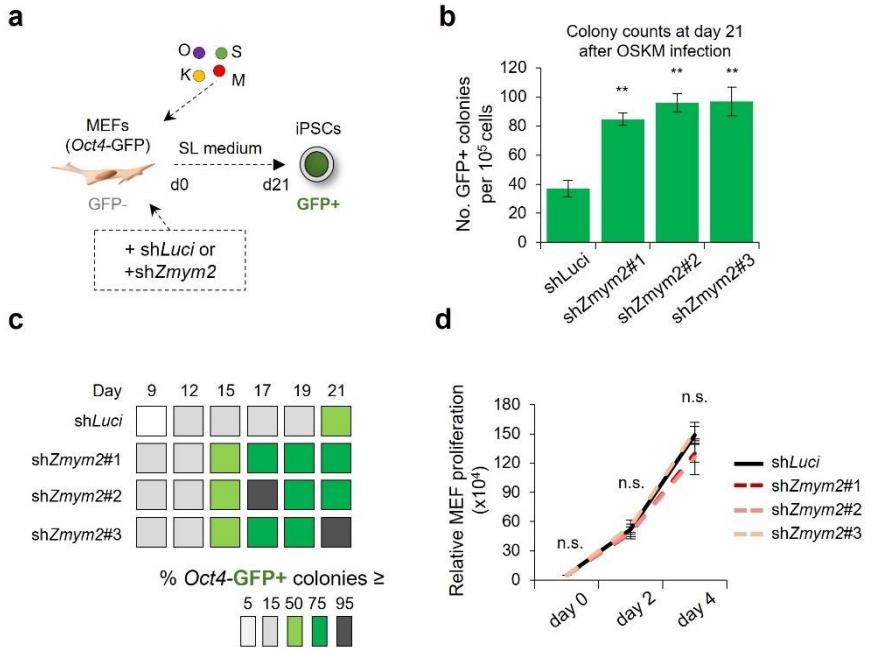


Figure R9. *Zmym2* depletion increases iPSC reprogramming efficiency. a. Scheme summarizing the somatic cell reprogramming process. *Oct4*-GFP MEFs were transfected with the OSKM reprogramming cocktail along with shRNAs against *Zmym2* or a control shLuci. Then, cells were cultured in a standard pluripotent medium (i.e., SL medium) for 21 days and GFP-positive colonies were counted to estimate the reprogramming efficiency. **b**. The number of GFP-positive colonies at day 21 was obtained after reprogramming of MEFs with an OSKM cocktail. Biological triplicates were analyzed. **c**. Percentage of *Oct4*-GFP positive colonies under the indicated conditions over the 21-day period of SCR process. **d**. The relative proliferation of *Oct4*-GFP MEFs infected with indicated shRNAs. Data are shown as the number of cells (x10^6) measured at the indicated time points. In **b**, **b**, one-way ANOVA with the Kruskal-Wallis correction for non-parametric distributions was performed. In **d**, a two-way ANOVA test was performed. ***p*<0.01; n.s., non-significant.

7.3.2 Absence of ZMYM2 promotes the pre-iPSC to iPSC transition

SCR consists of sequential phases showing specific transcriptional programs and three different phases (i.e., initiation, maturation, and stabilization) can be distinguished throughout the successful reprogramming process to obtain iPSCs[223,359,360].

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Importantly, NANOG plays an important role during the late stabilization phase, when intermediate reprogrammed cells or so-called "pre-iPSCs" have activated the pluripotency gene program and can successfully transit to iPSC[30,359]. Given my previous results indicating that ZMYM2 is important for transcriptional repression of the *Nanog* gene in pluripotency and differentiation, I directly address whether ZMYM2 depletion can promote the pre-iPSC to iPSC transition using a published reprogramming system that allows direct investigation of the pre-iPSC to iPSC transition[361]. Therefore, I first generated a clonal line of OSKM reprogramming intermediates from MEFs harboring a *Nanog*-GFP reporter transgene (**Figure R10a**). The pre-iPSCs (GFP-negative sorted cells) were maintained in standard pluripotent culture conditions (i.e., SL medium) and transduced with two independent shRNAs against *Zmym2* and sh*Luci* as control. Expression of *Zmym2* was analyzed in these cells by RT-qPCR to check knock-down efficiency. Hence, > 70% reduction in the expression levels of *Zmym2* was found in pre-iPSCs after shRNA-mediated *Zmym2* knockdown when compared to sh*Luci* (**Figure R10b**). More importantly, I detected increased levels of *Nanog* and other factors (i.e., *Esrrb*, *Sall4*, and *Tet2*) that promote pre-iPSC to iPSC transition after knockdown of *Zmym2* (**Figure R10c**). Taking into account that only a minority of pre-iPSCs will become iPSCs after switching to the 2iL medium condition, and the reprogramming efficiency can be greatly enhanced if exogenous NANOG is provided[362], I asked whether knockdown of *Zmym2* could replace the requirement for exogenous NANOG to promote the pre-iPSC to iPSC transition (**Figure R10d**). Indeed, I found that loss-of-function of ZMYM2 resulted in a significant increase in the number of AP-positive colonies after 10 days of 2iL medium switch (**Figure R10e**), and more than a 10-fold increase in the number of GFP-positive colonies compared to control sh*Luci* (**Figure R10f**). Taken together, these results suggest that ZMYM2 restricts the reactivation of *Nanog* and other pluripotent factors during the reprogramming process and, thus, functions as a molecular barrier to the transition from pre-iPSCs into pluripotent iPSCs.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

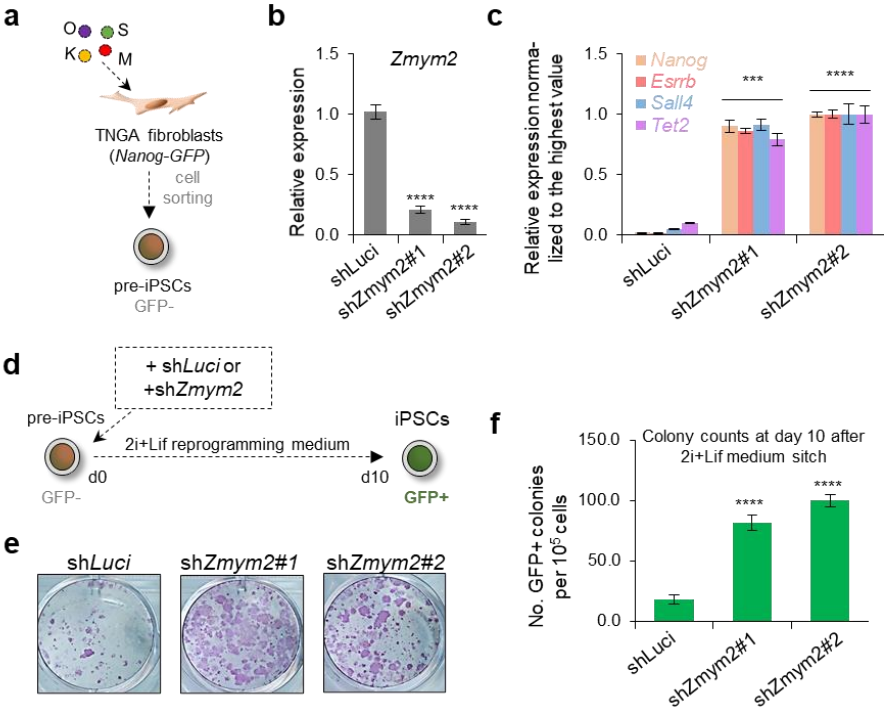


Figure R10. *Zmym2* depletion enhances the reprogramming efficiency of pre-iPSCs. a. Scheme summarizing the flow-cytometry protocol for pre-iPSCs generation. *Nanog*-GFP TNGA MEFs transfected with the reprogramming cocktail (OSKM) were cultured for reprogramming. Then, GFP-negative pre-iPSCs were sorted by flow-cytometry. **b**. *Zmym2* relative expression of pre-iPSCs infected with two different shRNAs against *Zmym2*. Values were normalized against *Gapdh*. Biological duplicates were analyzed. **c**. Relative expression of pluripotency-specific genes in pre-iPSCs infected with shRNAs against *Zmym2* compared to those infected with a control shRNA. Values were normalized against *Gapdh* and then against the highest value. **d**. Scheme summarizing the pre-iPSCs reprogramming protocol. Cells were infected with two independent shRNAs against *Zmym2* and then cultured in 2i + LIF medium for 10 days. AP staining and counting of GFP-positive colonies were performed to estimate reprogramming efficiency. **e**. Representative images from AP staining performed at day 10 of pre-iPSC reprogramming, as described in **d**. **f**. The number of GFP-positive colonies at day 10 of pre-iPSC reprogramming after control and *Zmym2* shRNA infection. Biological triplicates were analyzed. In **b**, **c**, and **f**, a one-way ANOVA test with the Kruskal-Wallis correction for non-parametric distributions was performed. Data are represented as mean ± SEM. ***p<0.001; ****p<0.0001.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7.3.3 Loss of ZMYM2 enhances heterokaryon-based reprogramming

Enforced expression of NANOG not only ensures direct reprogramming of somatic cells to the pluripotent ground state[30] but also promotes transfer of pluripotency after cell fusion from ESCs to somatic cells[189]. Given my previous results showing *Nanog* upregulation and enhanced self-renewal in *Zmym2* null ESCs when compared with their wild-type counterparts (**Figure R4**), I asked whether the absence of ZMYM2 could also enhance mouse ESC and human B (hB) cell heterokaryon reprogramming[188] (**Figure R11a**). Briefly, human somatic B lymphocytes were fused with *Zmym2*$^{+/+}$ or *Zmym2*$^{GT/GT}$ murine ESCs, heterokaryons were selected in presence of Ara-C and Ouabin drugs, and reprogramming efficiency was measured by the pluripotent human-specific gene expression profile (**Figure R11a**). Notably, my results showed that reprogramming of hB cells was obvious in both *Zmym2*$^{+/+}$ ESC/hB and *Zmym2*$^{GT/GT}$ ESC/hB heterokaryons (**Figure R11b)**. Related to this, although early upregulation of human B somatic markers (h*CD19* and h*CD45*) was found in *Zmym2*$^{GT/GT}$ ESC/hB heterokaryons at day 1, this early activation was rapidly suppressed and there were no significant changes in their expression levels at days 2 and 3 of the cell fusion assay between WT and null ESC/hB heterokaryons (**Figure R11b**). Importantly, my data revealed an enhanced human ES-specific gene expression profile (i.e., h*OCT4*, h*NANOG*, and h*CRIPTO*) in *Zmym2*$^{GT/GT}$ ESC/hB heterokaryons compared to in *Zmym2*$^{+/+}$ ESC/hB heterokaryons (**Figure R11c**). Taken together, these results indicate that the lack of ZMYM2 improves the reprogramming efficiency of the cell fusion system, providing additional validation of the functional implication of ZMYM2 as a reprogramming barrier.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

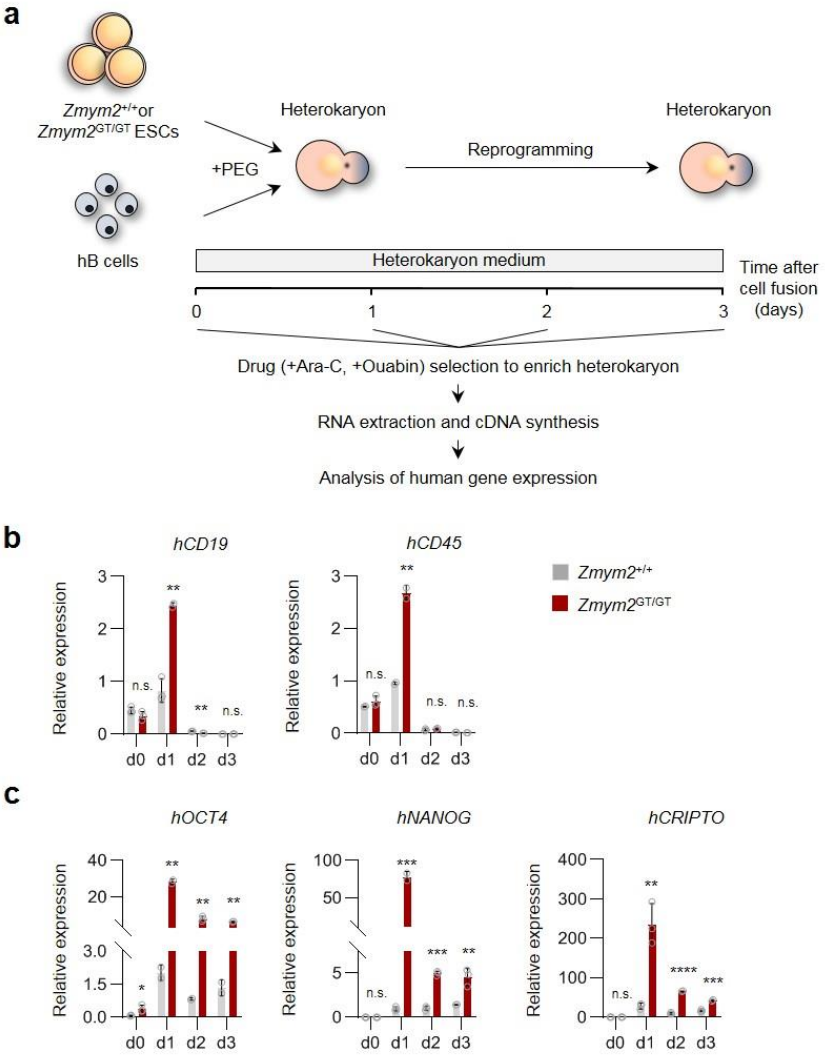


Figure R11. *Zmym2* depletion in ESCs improves heterokaryon-based fusion reprogramming of human B cells. a. Scheme summarizing the cell fusion reprogramming system. Murine ESCs with (*Zmym2*$^{+/+}$) or without *Zmym2* (*Zmym2*$^{GT/GT}$) were fused to human B cells. Heterokaryons were selected using drugs (Ouabin and Ara-C) and transcriptional changes were analyzed by RT-qPCR at the indicated timepoints. **b-c**. Relative expression of human B somatic markers (**b**) and human pluripotent markers (**c**) measured by RT-qPCR. Values from heterokaryons were normalized against human *Gapdh*. Technical triplicates were analyzed. Statistical significance was calculated using the unpaired Student's t-test with Holm-

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Sidak correction for multiple comparisons. Representative data from three independent experiments are shown as mean ± SD. **p<0.01; ***p<0.001; ****p<0.0001; n.s., non-significant.

7.4 ELUCIDATING THE MOLECULAR FUNCTIONS GOVERNED BY ZMYM2 IN PLURIPOTENCY

7.4.1 ZMYM2 influences the transcriptional landscape in murine ESCs

ZMYM2 is essential to determine the pluripotent state of murine ESCs (**Figure R6**), and its depletion affects the processes of cell differentiation and reprogramming (**Figures R7-R11**), suggesting an important role for ZMYM2 in coordinating gene regulation on cell fate decisions. To explore the molecular mechanisms by which ZMYM2 regulates pluripotent cell identity, transcriptional profiling of ESCs was performed in the presence and absence of ZMYM2. To this end, four biological replicates of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells were cultured in a standard pluripotent SL medium for two days (**Figure R12a**, top panel). After this time, extraction of total RNA was performed followed by massive sequencing. All four replicates for each condition showed a very similar transcriptional profile, as seen in the correlation matrix (**Figure R12a**, bottom panel). Hence, all *Zmym2*$^{GT/GT}$ biological replicates clustered together and independently from the WT cluster (**Figure R12a**, bottom panel). Then, a Gene Set Enrichment Analysis (GSEA) was performed to determine the enrichment in a total of 10,185 gene sets derived from Gene Ontology (GO; http://geneontology.org/) between null and WT transcriptomes (**Figure R12b**). The network visualization of GSEA results was performed using Cytoscape, and GOs positively enriched in *Zmym2*-null ESCs were represented as red circles, whereas GOs negatively enriched were represented in blue. Additionally, enriched GOs sharing high numbers of genes were represented as clusters. Notably, a cluster of development-related pathways appeared as negatively enriched in *Zmym2*$^{GT/GT}$ cells, concomitant with the delay in cell differentiation observed during the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

generation of EBs (**Figure R7**) and the overall transcriptional deregulation during germ layer specification (**Figure R8**).

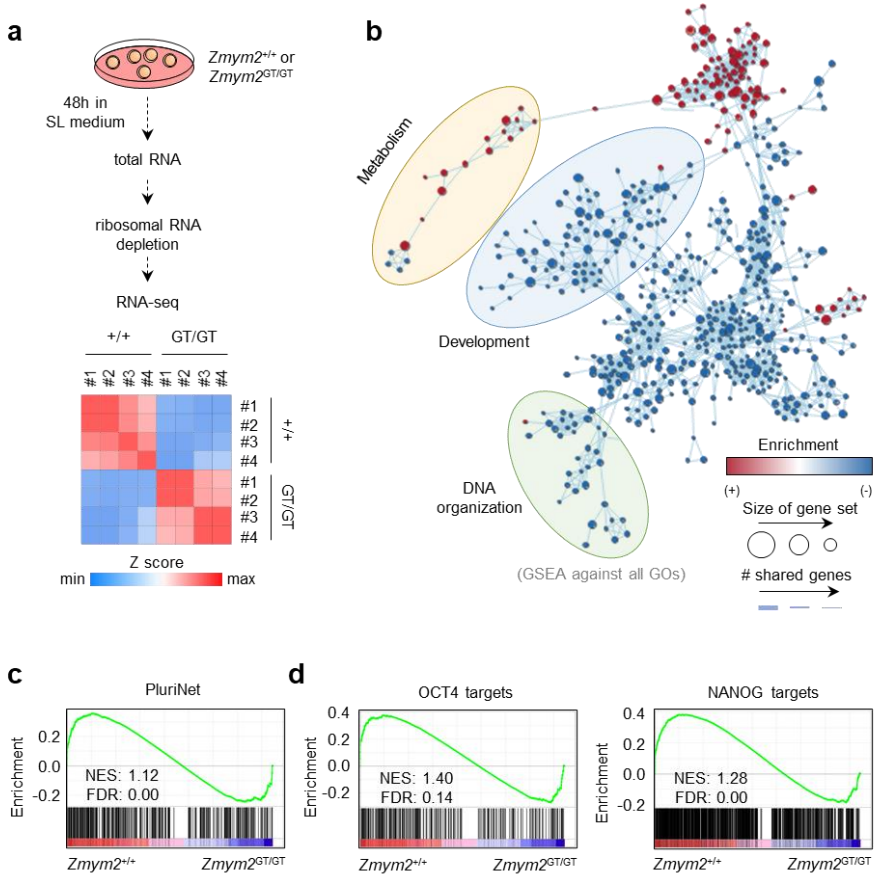


Figure R12. Absence of ZMYM2 alters the transcriptional landscape of murine ESCs. a. (*Top*) Depiction of the followed steps for the transcriptional profiling of murine ESCs under wild-type (*Zmym2*$^{+/+}$) and *Zmym2* knockout (*Zmym2*$^{GT/GT}$) conditions. (*Bottom*) Correlation matrix of the four biological replicates from *Zmym2*$^{+/+}$and *Zmym2*$^{GT/GT}$ cells. **b** Representation of the interconnection among gene ontologies (GOs) with a significant enrichment when comparing transcriptional profiles between *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cell lines. The color indicates enrichment (red: positive enrichment; blue: negative enrichment), the size of bubbles indicates the number of genes composing the specific gene set, and line thickness indicate the number of shared genes between adjacent gene sets. Analyses were performed using GSEA and representation was done using Cytoscape Inc. software. **c-d.** Gene set enrichment analysis from wild-type and *Zmym2* knockout

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RNA-seq data against gene sets of pluripotency genes (**c**), OCT4 targets (**d**; *left*) and NANOG targets (**d**; *right*). Normalized enrichment score (NES) and FDR are shown.

Moreover, a cluster of GOs related to cellular metabolism appeared as positively enriched in *Zmym2*-null ESCs, whereas pathways involved in DNA organization were negatively enriched in those cells, among others (**Figure R12b**). Remarkably, detailed GSEA analyses confirmed an altered pluripotency network after *Zmym2* depletion denoted by a significant differential enrichment in the expression of metastable pluripotent genes (designed as PluriNet) (**Figure R12c**). Furthermore, since ZMYM2 was confirmed to physically interact with the pluripotent master regulators OCT4 and NANOG (**Figure R1b**), enrichment analyses also identified the deregulation of OCT4 and NANOG targets in the absence of ZMYM2 when compared with their wild-type counterparts (**Figure R12d**). Thus, these results show that *Zmym2* depletion leads to transcriptional changes in genes related to development, cellular metabolism, and DNA organization by direct or indirect mechanisms, resulting in the alteration of the pluripotent state of murine ESCs.

To gain further insight into the molecular underpinnings of ZMYM2 function in regulating pluripotent cell identity, I performed transcriptional profiling of wild-type and *Zmym2*-null ESCs. I identified a total of 790 differentially expressed genes (fold-change>2; FDR<0.05) (**Figure R13a** and **Supplementary Table 7**). Then, global GSEA of these misregulated genes upon ZMYM2 depletion showed that 293 out of 792 GOs were significantly altered in *Zmym2*-null ESCs when compared with their wild-type counterparts (**Figure R13b**). Remarkably, gene signatures related to embryo development showed downregulation in *Zmym2*$^{GT/GT}$ cells (**Figures R13b** and **R13c**, top panel), in agreement with the general disruption in cell differentiation seen *in vitro* (**Figures R7** and **R8**). Furthermore, regulatory region nucleic acid binding genes were also significantly downregulated after the depletion of *Zmym2* (**Figure R13b** and **R13c**, bottom panel), in line

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

with a previously described ZMYM2 function as a transcription factor binding to DNA in a sequence-specific manner[319].

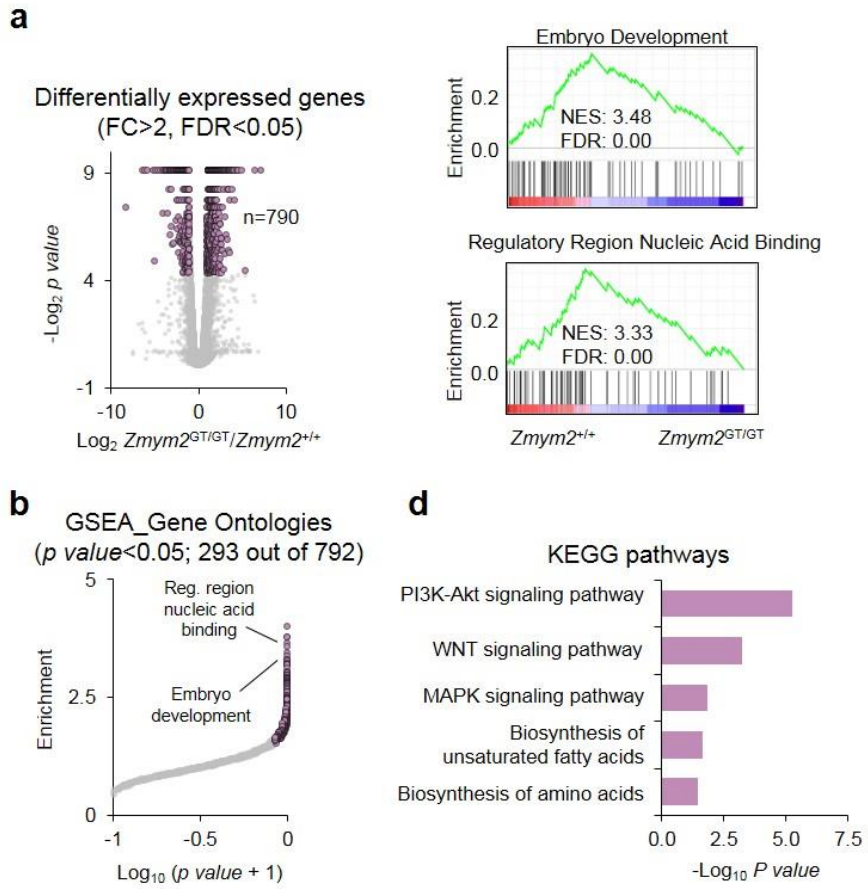


Figure R13. Absence of ZMYM2 leads to the misregulation of genes implicated in signaling pathways. a. Volcano plot showing gene expression levels (in log_2 FPKM) in *Zmym2*$^{GT/GT}$ versus *Zmym2*$^{+/+}$ ESCs determined by RNA-seq. Four biological replicates were analyzed. Genes differentially expressed (FDR < 0.05) more than 2-fold are shown in purple. FDR denotes false discovery rate. **b**. Scatter plot showing the enrichment scores at gene ontologies determined by GSEA analysis from RNA-seq data of *Zmym2*$^{GT/GT}$ versus *Zmym2*$^{+/+}$ cells. Statistically significant gene ontologies (*p*-value < 0.05) are shown in purple. **c**. Gene set enrichment analysis (GSEA) from wild-type and *Zmym2* knockout RNA-seq data against an embryo development (*top*) and a regulatory region nucleic acid binding (*bottom*) gene set. Normalized enrichment score (NES) and FDR are shown. **d**. Gene ontology (GO) analysis of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

differentially expressed genes in *Zmym2* knock-out ES cells. The x-axis corresponds to the negative $\log_{10}$ *p*-values.

Next, Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of these 790 differentially expressed genes in absence of ZMYM2 showed enrichment for biosynthetic processes of different metabolic macromolecules (e.g., fatty acids and amino acids) (**Figure R13d**). In addition, enrichment of several signaling pathways, such as the PI3K-AKT, WNT, or MAPK pathways (**Figure R13d**), which were extensively related to the regulation of the pluripotency state[363-366], could also be observed (**Figure R13d**). Taken together, these data suggest a feasible role of ZMYM2 in regulating cell identity through the control of both metabolome and transcriptional landscapes.

7.4.2 ZMYM2 regulates the activation and silencing of different transcriptional networks in ESCs

Loss of function of ZMYM2 in ESCs caused a global dysregulation of the genes involved in cellular metabolism and the control of signaling pathways that act to maintain the pluripotent state (**Figure R13d**). Therefore, I next further dissect the specific transcriptional changes promoted by the absence of ZMYM2 in ESCs. I identified 520 upregulated and 270 downregulated genes that were significantly changed with a difference in abundance greater than two-fold in the absence of ZMYM2 (fold-change>2; FDR<0.05) (**Figures R14a** and **R14b** and **Supplementary Table 7**). In accordance with my previous results, specification markers such as *Eomes* are significantly downregulated in the absence of ZMYM2 (**Figure R14b** and **R14c**), which could explain the observed defects during cell specification (**Figure R8**). *Lin28b*, a gene implicated in the regulation of cellular metabolism and primed conversion[175], was also downregulated in *Zmym2*$^{GT/GT}$ ESCs (**Figure R14b** and **R14c**). Moreover, epigenetic factors such as *Dnmt3b* and *Dnmt3l*, whose downregulation improves the acquisition of pluripotency[367], were significantly less expressed in

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Zmym2 null ESCs compared with their wild-type counterparts (**Figure R14b** and **R14c**).

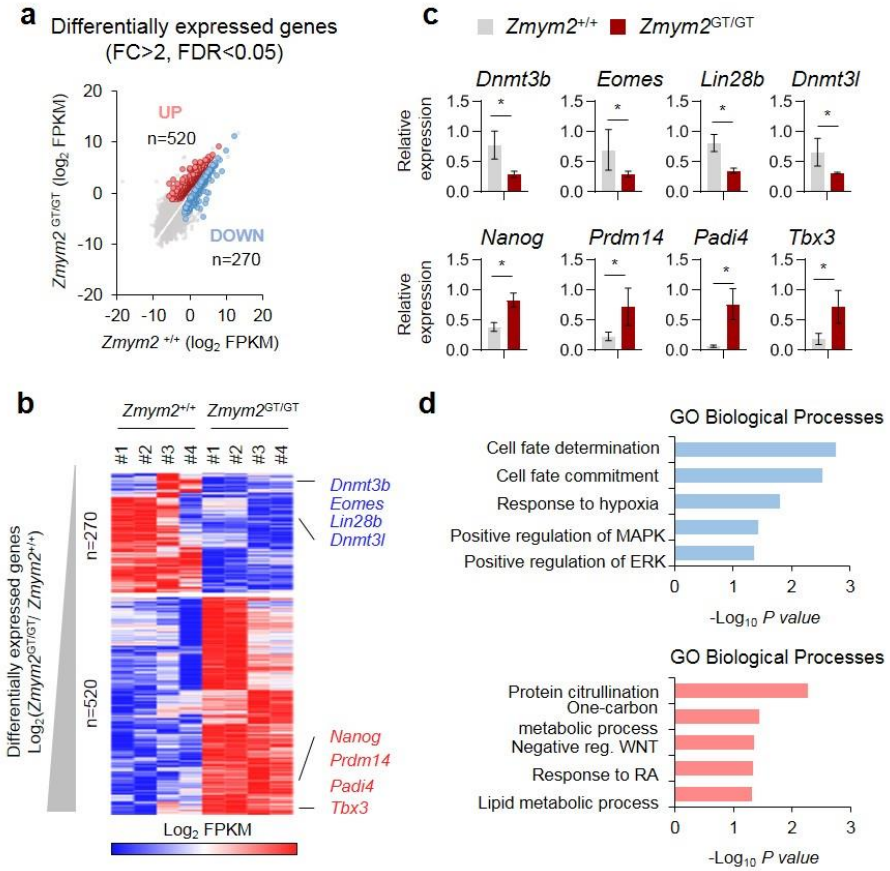


Figure R14. ZMYM2 depletion leads to activation and silencing of gene expression in ESCs. a. Scatter plot of gene expression levels (in $\log_2$ FPKM) in *Zmym2*$^{GT/GT}$ versus control mouse ESCs determined by RNA-seq. Four biological replicates were analyzed. Genes significantly upregulated (red) and downregulated (blue) more than 2-fold (FC>2) and significantly changed (FDR<0.05) are shown. **b**. Heatmap of more than 2-fold differentially expressed genes in *Zmym2*$^{GT/GT}$ vs *Zmym2*$^{+/+}$ cells. Four biological replicates from each cell line were analyzed. Data is represented as $\log_2$ FPKM as determined by RNA-seq. **c**. Relative expression levels of highlighted genes in **b** as determined by RT-qPCR. *U6* was used as housekeeping. Technical triplicates were analyzed. Data are represented as mean ± SD, and a two-tailed t-test was performed. **p*<0.05. **d**. Gene ontology (GO) analyses for biological processes of downregulated (*top*) and upregulated (*bottom*) genes in *Zmym2*$^{GT/GT}$ versus *Zmym2*$^{+/+}$ ESCs as determined by RNA-seq.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

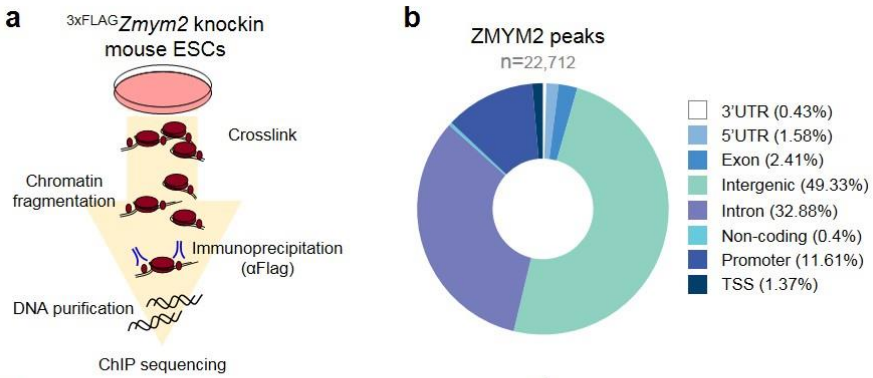
In contrast, genes involved in the transcriptional regulation of ESC self-renewal such as *Nanog*, *Padi4*, *Prdm14* and *Tbx3*[71,153,368,369] were significantly upregulated in *Zmym2* null ESCs (**Figure R14b** and **R14c**). Remarkably, gene ontology (GO) analyses confirmed the implication of downregulated genes in cell fate determination and commitment (**Figure R14d**, top panel). On the other hand, genes upregulated after the depletion of *Zmym2* appeared to be related to the regulation of different metabolic processes (e.g., one-carbon and lipid metabolic processes) and in post-translational modification of proteins (e.g., protein citrullination) (**Figure R14d**, bottom panel). Collectively, these data further suggest the feasible role of ZMYM2 in governing cell identity and cell fate decisions by interconnecting gene expression and different metabolic processes.

7.4.3 ZMYM2 can act both as a transcriptional activator and a repressor in stem cell pluripotency

My results indicate that the absence of ZMYM2 in murine ESCs leads to transcriptional changes affecting pluripotent cell identity. In addition, others have demonstrated that ZMYM2 can bind to DNA in a sequence-specific manner[319]. To further investigate the exact role of ZMYM2 in regulating the transcriptional landscape of murine ESCs, I identified global targets of ZMYM2 by combining chromatin immunoprecipitation assay with sequencing (CHIP-seq). To this end, I took advantage of the recent ZMYM2 ChIP-seq dataset generated by our collaborator Dr. Jianlong Wang (Columbia University, USA) in mouse ESCs[342] (**Figure R15a**). Using *in silico* analysis I was able to identify a total of 22,712 ZMYM2 binding regions from which 65% (14,607 out of 22,712) were located at regulatory (i.e., UTRs, promoters, and TSSs) and intergenic regions (**Figure R15b**). Furthermore, when a *de novo* motif analysis was performed using HOMER tools, several DNA motifs were significantly revealed (**Figure R15c**). Among them, a highly significant *de novo* binding motif was detected at 363 binding regions, therefore conforming the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

most feasible ZMYM2 DNA binding motif (**Figure R15d**). Collectively, these data suggest that ZMYM2 may occupy the regulatory regions of genes for the control of their transcriptional levels in the pluripotent context.



c

Logo	N sites	*P value*
TGTGTCCGGA	363	1e-16
GTGTGCCGTT	290	1e-15
GCTCGAGCCTGT	35	1e-14
CATTCCTCCC	289	1e-14

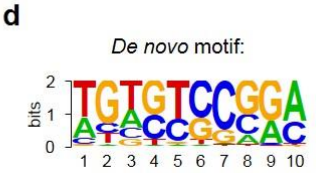


Figure R15. ZMYM2 binds to DNA in a sequence-specific manner. a. Scheme summarizing the followed steps for ZMYM2 chromatin immunoprecipitation followed by sequencing (ZMYM2 ChIP-seq) performed by Yang and collaborators (REF 342). **b**. Pie chart showing the distribution of ZMYM2 peaks in the chromatin as determined by ChIP-seq. UTRs, untranslated regions; TSS, transcription start sites. **c**. Table showing the most significant ZMYM2 binding motifs as determined by ChIP-seq analysis. **d**. *De novo* ZMYM2 binding motif for chromatin occupancy as determined by ChIP-seq and analyzed using the HOMERO online tool.

Next, to further investigate how ZMYM2 regulates its transcriptional targets in ESCs cultured under pluripotent standard conditions (i.e., SL medium) I performed a combinatorial analysis *in silico* by integrating the ChIP-seq and RNA-seq datasets. Hence, from the overall 790 differentially expressed genes after the depletion of *Zmym2* (fold-change>2*;* FDR<0.05) (**Figure R13a**), I identified a total of 468 genes being directly regulated by ZMYM2 (**Figure R16a**). Thus, almost 60% of differentially expressed genes appeared to be targets of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ZMYM2 (468 out of 790 genes), in contrast with 40% of genes being indirectly regulated by this transcription factor (**Figure R16a**). This suggests the existence of both ZMYM2 direct and indirect mechanisms in the regulation of the transcriptional landscape in murine ESCs. When analyzed separately, I found that almost 60% of significantly down- and upregulated genes in the absence of ZMYM2 were also direct targets of this protein (**Figure R16b**). In this sense, ZMYM2 peaks were identified in 157 out of 270 and 311 out of 520 downregulated and upregulated genes, respectively (**Figure R16b**). Gene ontology analyses showed that downregulated targets of ZMYM2 are enriched in genes implicated in both the positive and negative regulation of the canonical WNT signaling pathway (**Figure R16c**, left panel). Activation of this specific signaling pathway leads to GSK3 inactivation, which finally promotes an undifferentiated state in murine ESCs[41]. Additionally, CHIR99021, a specific chemical inhibitor for GSK3, was demonstrated to strongly support the *in vitro* self-renewal of naïve pluripotent cells[40]. Taking this into consideration, it is plausible that the absence of ZMYM2 leads to aberrant hyperactivity of this signaling pathway, promoting an increased self-renewal capacity of cells as seen in **Figure R4f**. Furthermore, genes implicated in the regulation of cell differentiation were also found to be directly downregulated after *Zmym2* depletion (**Figure R16c**, left panel), in agreement with my results of altered *in vitro* differentiation of *Zmym2*$^{GT/GT}$ cells (**Figures R7** and **R8**). Moreover, genes related to the cellular response to glucose were also directly activated by ZMYM2 (**Figure R16c**, left panel), suggesting a possible role of this protein in the regulation of the metabolic state of cells. In contrast, genes regulating the MAPK signaling pathway were also directly disrupted by ZMYM2 (**Figure R16c**, right panel). Inactivation of this pathway with the PD0325901 chemical inhibitor was demonstrated to support an extended potency in murine ESCs[40]. Additionally, components of the MAPK pathway are known to be downstream effectors of the FGF signaling pathway, which activation is required for the maintenance of a primed state of pluripotency[370]. Since ZMYM2 was demonstrated to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

fuse with FGFR1 protein in myeloproliferative syndrome[303,371], it would be possible that a direct interplay between ZMYM2 and components of this signaling pathway exist for the proper regulation of an undifferentiated state. Lastly, genes implicated in multicellular organism development and cell differentiation were found to be ZMYM2 targets significantly upregulated in knockout cells (**Figure R16b**, right panel), further supporting that ZMYM2 is essential for the proper differentiation of cells.

Taken together, these results demonstrate that ZMYM2 governs both direct and indirect mechanisms for the regulation of the transcriptional landscape in murine pluripotency. Furthermore, ZMYM2 directly orchestrates the expression levels of genes involved in the regulation of signaling pathways that have a key role in maintaining pluripotent cell identity.

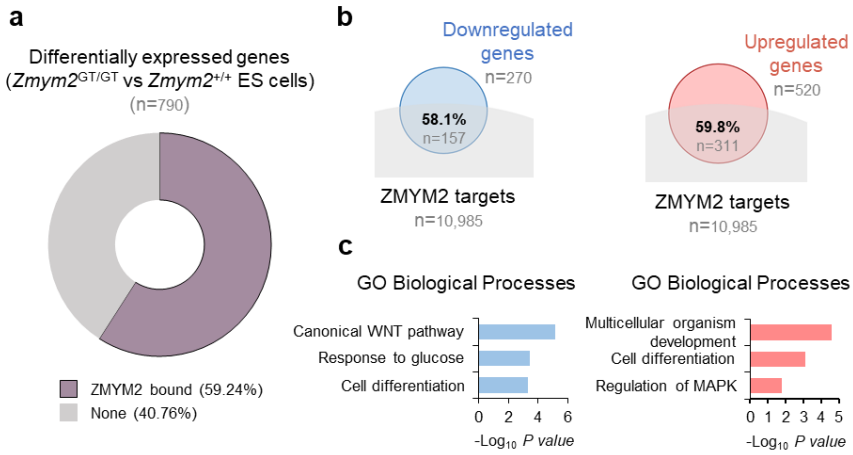


Figure R16. ZMYM2 can function as both an activator and a repressor of gene transcription. a. Pie chart showing the percentage of differentially expressed genes that are bound by ZMYM2 as estimated by ChIP-seq. **b.** Venn diagram showing the percentage of downregulated (blue; *left*) and upregulated genes (red; *right*). **c.** Gene ontology analysis of downregulated (blue; *left*) and upregulated (red; *right*) targets of ZMYM2 determined in **b**. The x-axis corresponds to the negative $\log_{10}$ p-value.

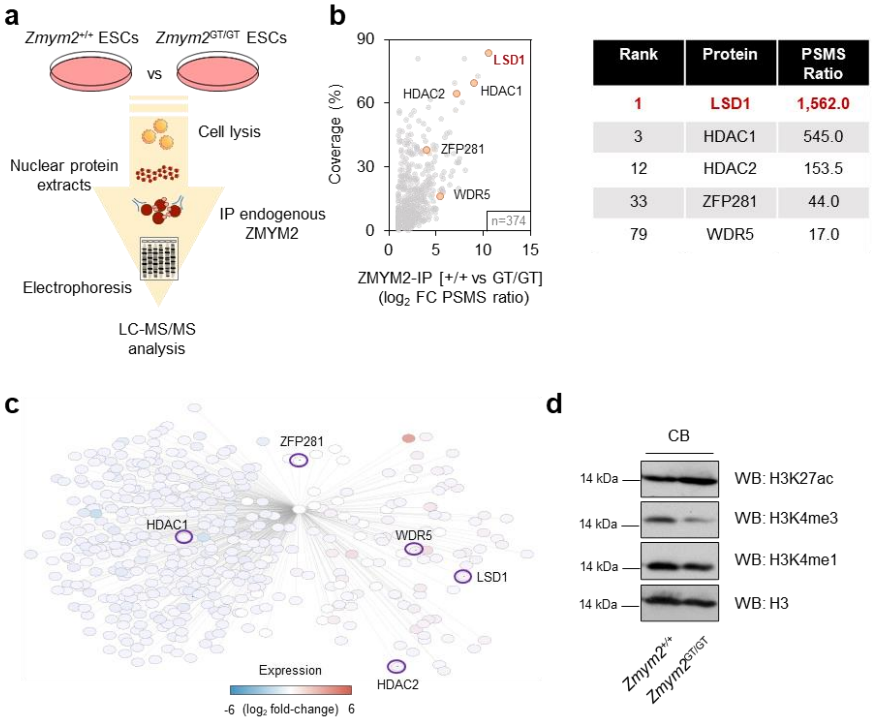
USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7.4.4 Identification of ZMYM2-interacting partners in ESCs

To understand how ZMYM2 regulates the transcriptional landscape in pluripotency, I investigated the ZMYM2 interactome in ESCs. To this end, we used affinity purification followed by LC-MS/MS as previously described[86] to identify ZMYM2-interacting proteins (**Figure R17a**). Notably, I identified a total of 374 high-confidence ZMYM2-interacting partners (**Figure R17b**). These contain many transcription factors (TFs), RNA processing factors, protein folding factors, epigenetic regulators, and others (**Supplementary Table 8**). Consistent with previous findings demonstrating the interaction between ZMYM2 and HDAC proteins composing the LSD1-RCOR1/2 corepressor complex in ESCs[372], I also found HDAC1, HDAC2, and LSD1 as interacting partners of ZMYM2 (**Figure R17b**). Additionally, ZFP281, another zinc finger protein having a pivotal role in the regulation of both naïve and primed states of pluripotency[86] was also found to directly interact with ZMYM2 (**Figure R17b**). Both LSD1 and ZFP281 have previously been implicated in the repression of genes regulating the undifferentiated state of cells[142,143,289,373,374]. Interestingly, the epigenetic factor WDR5, which was involved in the maintenance of the ESC state[375], was also enriched in my ZMYM2 interactome (**Figure R17b**).

To further decipher the molecular mechanisms that implicate ZMYM2 and its directly interacting partners in the regulation of ESC identity, I first examined the possible transcriptional changes occurring in their protein-coding genes as a consequence of the loss-of-function of ZMYM2 in ESCs. Taking advantage of our RNA-seq data, I was able to observe that most of the genes that encode proteins that partner with ZMYM2 did not have significant changes in their expression levels in the absence of ZMYM2 (**Figure R17c**). Specifically, LSD1, HDAC1/2, ZFP281 and WDR5 protein interacting partners are among the genes whose expression levels are not significantly altered in *Zmym2* null ESCs (**Figure R17c**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



Rank	Protein	PSMS Ratio
1	LSD1	1,562.0
3	HDAC1	545.0
12	HDAC2	153.5
33	ZFP281	44.0
79	WDR5	17.0

Figure R17. The ZMYM2 interactome in ESCs. a. Scheme summarizing the followed steps for the affinity purification and mass spectrometry (AP-MS) identification of ZMYM2-interacting partners. **b**. (*Left*) Scatter plot showing the 374 high-confidence ZMYM2-interacting partners ordered by the percentage of coverage (y-axis) and by fold change in the ratio of the number of identified peptide spectrum matches (PSMs) between *Zmym2* WT and null ESCs (x-axis). (*Right*). A table highlighting main ZMYM2 interactors, along with their PSMs ratio between *Zmym2* WT and null ESCs. **c**. Interactome of all 374 ZMYM2-interacting partners identified in **b**. The color shows differential expression levels when comparing *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cell lines as determined by RNA-seq. Interactome was performed using Cytoscape Inc. software. **d**. Western-blot showing relative levels of indicated histone marks. Chromatin bound fractions from *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ nuclear extracts were used. H3 was used as a loading control.

My results shown that the ZMYM2 interactome is enriched in several histone modifiers regulating ESC identity. Previous reports have highlighted the essential role of WDR5 in the trimethylation of H3K4[376]. For this reason, I asked whether H3K4me3 levels may be affected by the absence of ZMYM2, so chromatin-bound fractions of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

both *Zmym2* WT and null ESCs were analyzed by western blot for this epigenetic mark. Remarkably, profound decreased levels of H3K4me3 were observed after *Zmym2* depletion (**Figure R17d**), in line with my findings of ZMYM2 directly associating with WDR5. In addition, a central role for LSD1 in the decommission of ESCs enhancers through its demethylating activity on H3K4me1 was demonstrated[142]. Hence, a slight decrease in this epigenetic mark in the absence of ZMYM2 could also be observed (**Figure R17d**), highlighting a possible partnership between ZMYM2 and LSD1 in the regulation of enhancer activity in ESCs. Lastly, H3K7 acetylation levels are known to be regulated by the interplay between KATs and HDACs[130,131]. When levels of this epigenetic mark were analyzed, an increase in H3K27ac was seen in *Zmym2* null chromatin-bound fractions (**Figure R17d**), in accordance with a previous report highlighting the direct association between ZMYM2 and HDAC1/2[319].

Together, these results show that ZMYM2 physically interacts with different ESC regulators that play important roles in pluripotency and reprogramming.

7.4.5 Genome-wide significance of the ZMYM2-LSD1 interaction

Our results highlight LSD1 as the top high-confidence ZMYM2-interacting partner (**Figure R17b**). This interaction was previously reported in HeLa and Hek293T cells[319] and also in ESCs. Previous reports have demonstrated LSD1 to be essential during embryo development[140,143]. Moreover, LSD1 is a key histone modifier, demethylating histone H3 on Lys 4 or Lys 9 (H3K4/K9), which is reported to be essential in the balance between self-renewal and differentiation in ESCs by regulating bivalent domains[374] and in regulating enhancers of the pluripotency program during differentiation[377]. Remarkably, the group of our collaborator Dr. Jianlong Wang (Columbia University, USA) has recently reported that the ZMYM2-LSD1 axis is involved in the transcriptional repression of

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

endogenous retrovirus in ESCs[342]. However, it is not yet fully understood how LSD1 is recruited to chromatin. Therefore, I interrogated the potential role of ZMYM2 in the recruitment of LSD1 for transcriptional regulation. To this end, I took advantage of a published LSD1 ChIP-seq dataset in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs[342] to analyze LSD1 chromatin occupancy. In presence of ZMYM2, I identified 11,428 LSD1 genome-wide peaks, and more than 65% of them were located at regulatory regions (i.e., UTRs, intergenic regions, promoters, and TSSs) (**Figure R18a**, top panel). Notably, although when I determined the occupation of LSD1 in *Zmym2* null ESCs I detected only a slight decrease in the number of genome-wide peaks (i.e., 10,959 peaks in null ESCs versus 11,428 in WT ESCs), I could observe important alterations in their distribution in regulatory regions such as intergenic and promoter region (**Figure R18a**, bottom panel). These data suggest that the partnership between ZMYM2 and LSD1 might be important for fine-tuning the proper localization of LSD1 on chromatin. Next, to understand how LSD1 contributes to transcriptional regulatory functions of ZMYM2, I compared ZMYM2 and LSD1 ChIP-seq data in ESCs under the same standard pluripotent culture conditions (**Figure R18b**). Importantly, I observed a great overlap of ZMYM2 and LSD1 in gene occupancy. Remarkably, 70.8% of target genes (4,518 out of 6,382) that are bound by LSD1 are directly bound by ZMYM2 (**Figure R18b**). In contrast, only 41.1% of target genes (4,518 out of 10,985) that are bound by ZMYM2 are also bound by LSD1 (**Figure R18b**), suggesting a much wider genome distribution of ZMYM2 than LSD1. Interestingly, the absence of ZMYM2 led to both the gain (n= 3,014) and loss (n= 2,618) of target genes occupied by LSD1 (**Figure R18c**). Among the targets in which LSD1 occupation was preserved in *Zmym2* null ESCs (n=3764), 75.98% (2,860 out of 3,764) of those genes were also co-occupied by ZMYM2 (**Figure R18d**). Importantly, when occupancy levels at those genes (e.g., *Nanog*) were analyzed after loss of *Zmym2*, an obvious decrease in LSD1 binding could be observed (**Figure R18e**). Collectively, these results suggest that ZMYM2 may recruit LSD1 to a subset of shared

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

target genes for modulating ZMYM2 transcriptional regulatory functions through histone demethylase activity of LSD1.

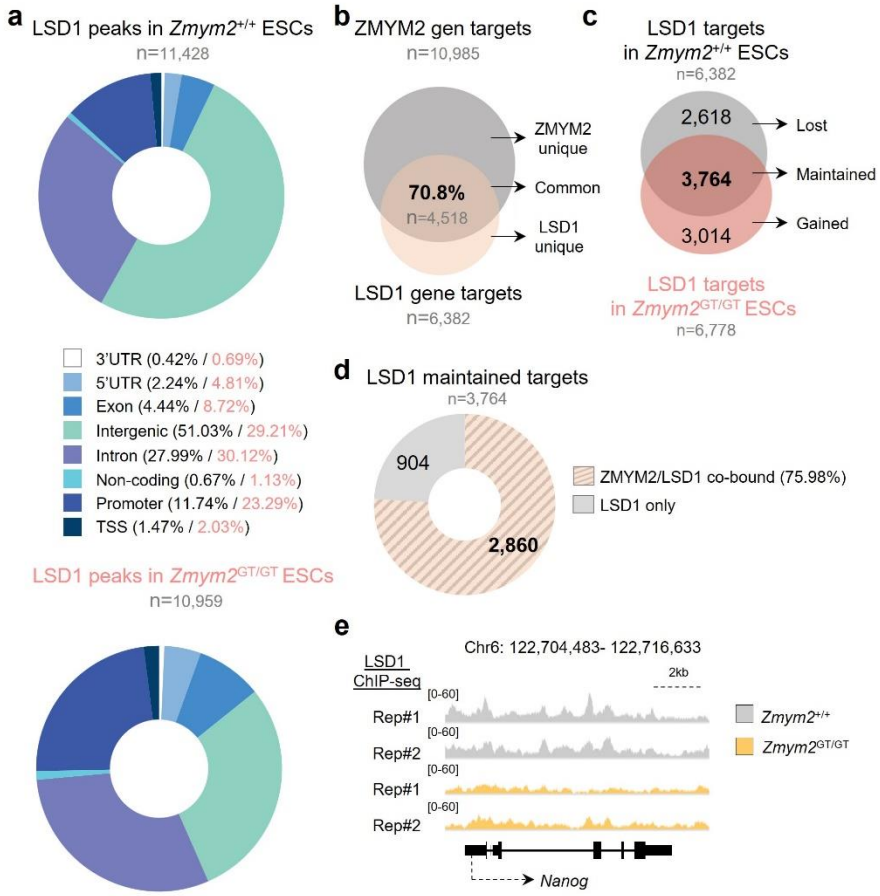


Figure R18. Analysis of genome-wide occupancy between ZMYM2 and LSD1 in ESCs. a. Pie charts showing the distribution of LSD1 chromatin binding peaks in *Zmym2*$^{+/+}$ (*top*) and *Zmym2*$^{GT/GT}$ ESCs (*bottom*) as determined by ChIP-seq. UTRs, untranslated regions; TSS, transcriptional start sites. **b**. Venn diagram showing the overlap of ZMYM2 (grey) and LSD1 (orange) target genes as determined by ChIP-seq. **c**. Venn diagram showing the overlap of LSD1 targets in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (red) ESCs as determined by ChIP-seq. **d**. Pie chart showing the proportion of common LSD1 peaks in *Zmym2* wild-type and null ESCs that are co-occupied by ZMYM2. e. ChIP-seq binding profiles (reads per million per base pair) for LSD1 at the Nanog locus in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (yellow) ESCs. Data were visualized using the IGV software.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

To better understand how ZMYM2 cooperates with LSD1 for transcriptional regulation in ESCs, I asked whether differentially expressed genes after *Zmym2* depletion are co-occupied by ZMYM2 and LSD1. Notably, I found that among the 468 differentially expressed genes between *Zmym2* WT and null ESCs that are also targets of ZMYM2 (**Figure R16a**), 44.87% are also co-occupied by LSD1 (210 out of 468 genes) (**Figure R19a**). When I analyzed the transcriptional changes of these 210 ZMYM2-LSD1 target genes in more detail, I found that 63.33% of them (133 out of 210) were significantly upregulated, and 36.67% of them (77 out of 210) were downregulated in *Zmym2* null cells compared with their wild-type counterparts (**Figure R19b**). Collectively, these results suggest a cooperative role of ZMYM2 and LSD1 in the transcriptional regulation of a subset of shared target genes. To further confirm this hypothesis, I examined gene regulation by LSD1 using a published RNA-seq dataset from *Lsd1* WT and null murine ESCs grown under the same standard pluripotent conditions as our *Zmym2* ESC lines[338]. Notably, among the 210 ZMYM2 and LSD1 shared target genes, while in the absence of LSD1 I did not observe significant gene expression changes in those 77 genes downregulated in *Zmym2* null ESCs (**Figure R19c**, left panel), I also found a significant upregulation in *Lsd1*$^{-/-}$ ESCs in those 133 genes that are activated after loss-of-function of ZMYM2 (**Figure R19c**, right panel). Moreover, this upregulation was consistent among the three independent replicates of *Lsd1* null ESCs compared with their wild-type counterparts (**Figure R19d**). Gene ontology analyses of these 133 genes revealed their role in the development and maintenance of the stem cell population, as well in different signaling pathways (i.e., SMAD, BMP, and MAPK signaling pathways) known for their relevance in pluripotency (i.e., acquisition, maintenance, and differentiation), and more importantly, are known to be key in the connection between the metabolism of these cell fate processes and their epigenetic status for the control of gene expression[107,366,378-383] (**Figure R19e**). Taken together, these data suggest that LSD1

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

cooperates with ZMYM2 for the transcriptional repression of a subset of shared target genes involved in pluripotency and reprogramming.

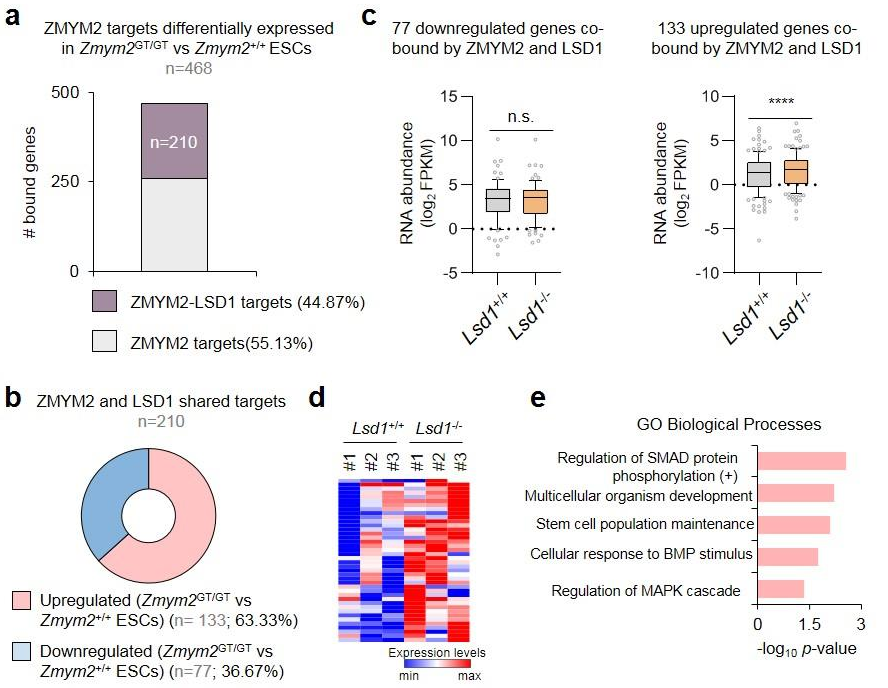


Figure R19. ZMYM2 cooperates with LSD1 to act as a transcriptional repressor. a. Comparison of ZMYM2-bound genes with LSD1 targets that are differentially expressed upon depletion of ZMYM2 in ESCs. **b**. Pie chart of differentially expressed genes that are co-bound by ZMYM2 and LSD1 as determined in **a**. Genes significantly upregulated (light red) and downregulated (blue) are shown when comparing *Zmym2*$^{GT/GT}$ vs *Zmym2*$^{+/+}$ transcriptional profiles determined by RNA-seq. **c**-**d**. mRNA expression levels (shown as $\log_2$ FPKM) in *Lsd1*$^{+/+}$ (grey) and *Lsd1*$^{-/-}$ ESCs (orange) of downregulated (**c**) and upregulated (**d**) LSD1 and ZMYM2 co-bound targets determined in **a**. Biological triplicates were analyzed. Data was obtained from REF 338. Statistical significance was calculated using the Wilcox test for paired samples. ****p<0.0001; n.s. non-significant. **e**. Heatmap representing expression levels of upregulated ZMYM2 and LSD1 shared targets in *Lsd1*$^{+/+}$ and *Lsd1*$^{-/-}$ ESCs as determined by RNA-seq. **f**. Gene ontology analysis for biological processes of upregulated genes determined in **d**. The x-axis is shown as -$\log_{10}$ p-value.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7.4.6 Loss of ZMYM2 promotes a metabolic switch in ESCs

Due to the functional cooperation between ZMYM2 and LSD1 for the transcriptional repression of a subset of shared genes involved in the regulation of different signaling cascades (e.g., SMAD and MAPK pathways) important in metabolic-associated processes that influence several pluripotent features, I next examined how the loss of ZMYM2 in ESCs affected the expression of genes related to the SMAD and MAPK signaling pathways. Notably, I found that most of those genes were activated in *Zmym2*$^{GT/GT}$ when compared with their wild-type counterparts (**Figure R20a**). Furthermore, altered protein levels of key regulators of these pathways could be observed by western blot (**Figure R20b**). In particular, along with an increase in the undifferentiated potency of ESCs in the absence of ZMYM2, reduced levels of AKT were observed in whole-cell extracts from *Zmym2*$^{GT/GT}$ compared to *Zmym2*$^{+/+}$ ESCs (**Figure R20b**). In contrast, increased activation of the MAPK/ERK pathway, exemplified by increased phosphorylation of both ERK1 and MEK1 kinases, was found in *Zmym2* null ESCs (**Figure R20b**). Remarkably, the ERK pathway, a downstream effector of the FGF signaling pathway, has been implicated in the activation of aerobic glycolysis in mammalian cells[384,385]. Therefore, I first analyzed whether the absence of ZMYM2 can alter the expression of metabolic gene networks that govern the balance between glycolytic and oxidative phosphorylation (OxPhos) activity (**Figure R20c**). Indeed, I observed that ZMYM2 contributes to the proper transcriptional regulation of glycolytic- and OxPhos-related pathways in ESCs (**Figure R20c**). Then, to evaluate possible bioenergetic changes after ZMYM2 depletion, I used the Glycolysis Stress and Cell Mito Stress kits and a SeaHorse XFp analyzer (Agilent Technologies) to quantify the extracellular acidification rate (ECAR) as an index of the glycolytic activity and the oxygen consumption rate (OCR) as indicative of the respiratory capacity. In agreement with my previous data, I found that, even before glucose addition, *Zmym2* null ESCs showed a higher rate of basal glycolysis activity than *Zmym2*$^{+/+}$

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ESCs (**Figure R20d**). Consistent with this, basal oxygen consumption rate (OCR) was significantly reduced in *Zmym2*$^{GT/GT}$ ESCs (**Figure R20e**), indicating reduced basal respiration in these cells when compared with their wild-type counterparts. Furthermore, the addition of both rotenone and antimycin A (Rot/AA), specific chemical inhibitors of mitochondrial complex I and III, respectively, induced a less profound reduction of the OCR levels after genetic ablation of *Zmym2* in ESCs, suggesting a decreased maximal respiratory capacity in *Zmym2* null ESCs when compared to *Zmym2*$^{+/+}$ ESCs (**Figure R20e**). Subsequently, the OCR/ECAR ratio was reduced in *Zmym2*$^{GT/GT}$ ESCs (**Figure R20f**), further supporting that *Zmym2* depletion shifted mitochondrial metabolism toward glycolytic activity.

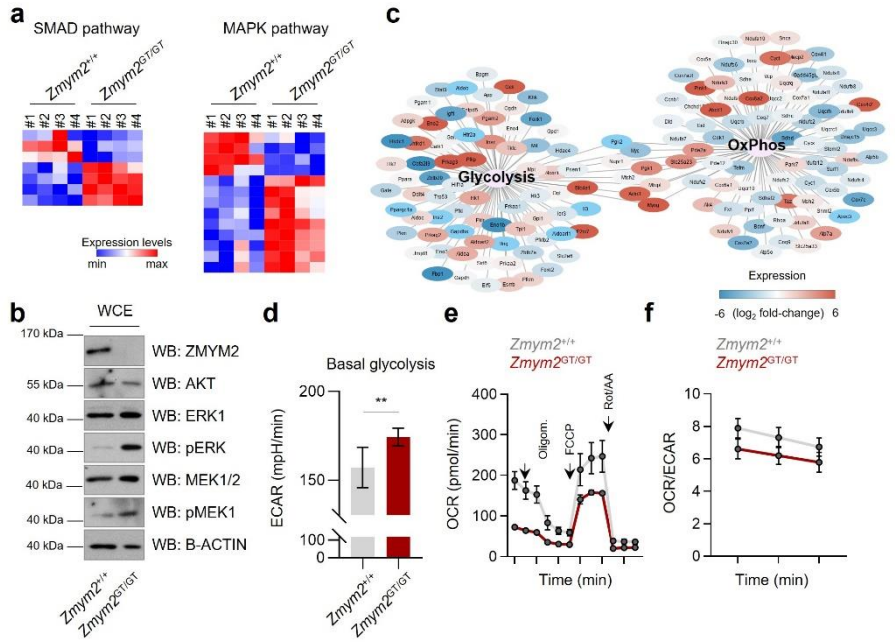


Figure R20. *Zmym2* depletion promotes an increase in glycolytic activity in ESCs. **a.** Heatmap showing the expression levels (represented as $\log_2$ FPKM) of genes related to SMAD (*left*) and MAPK (*right*) signaling pathways in indicated ESC lines. **b.** Western blot analysis of indicated proteins using whole-cell extracts (WCE) from *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. **c.** Analysis of metabolic gene networks that govern the balance between glycolytic and oxidative phosphorylation (OXPHOS) activity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Scale bar represents the transcriptional changes caused by ZMYM2 depletion obtained from RBA-seq datasets (log_2 fold-change values from *Zmym2*$^{GT/GT}$ vs *Zmym2*$^{+/+}$). **d**. Measured basal glycolysis activity determined as extracellular acidification rate (ECAR) expressed as mpH/min in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. **e**. Bioenergetic assays in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (red) ESCs. The different drugs used in the assay to determine the oxygen consumption rate (OCR) are indicated. **f**. OCR/ECAR ratio under basal conditions for the indicated ESC lines at three different time points. Bioenergetic assays were performed using a SeaHorse XFp analyzer. Three technical replicates were analyzed, and representative data from two independent experiments is shown. In **d,** a Wilcox test for paired samples was performed. $^{**}p<0.01$.

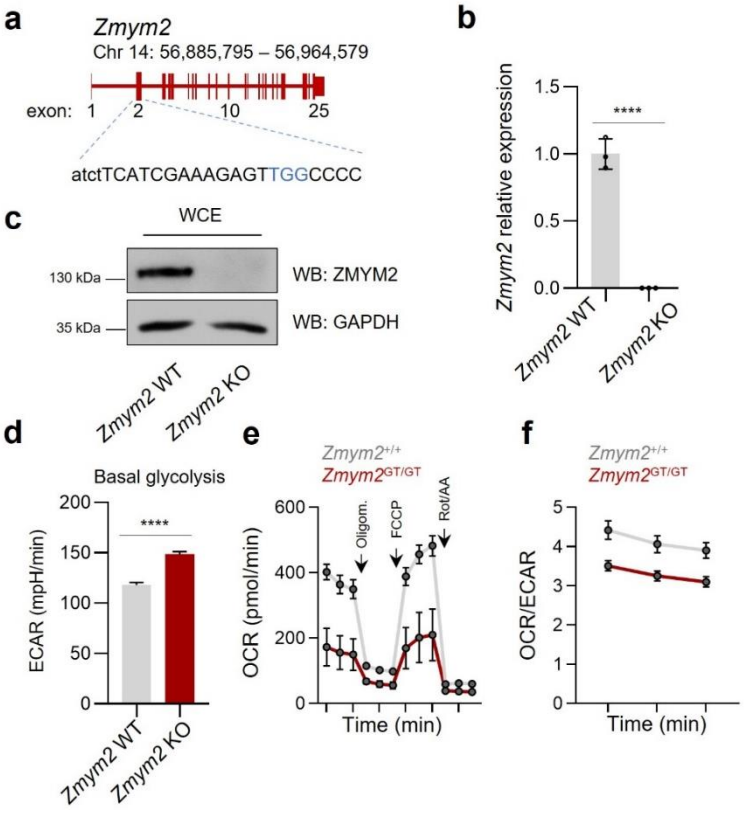


Figure R21. Alternative strategy using the CRISPR/Cas9 system to validate the role of ZMYM2 in modulating energy metabolism in ESCs. a. Scheme of the strategy followed for the generation of a second *Zmym2* knockout cell line using the CRISPR/Cas9 nuclease system. **b**. Relative expression of *Zmym2* by RT-qPCR in indicated ESC clones generated in **a**. **c**. Western blot analysis of ZMYM2 protein level from whole-cell extracts (WCE) in *Zmym2* wild-type (WT) and knockout (KO) ESCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GAPDH was used as the loading control. **d**. Basal glycolysis levels determined as ECAR (mpH/min) in WT and *Zmym2* KO ESCs. **e**. Bioenergetic assays in *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (red) ESCs. The different drugs used in the assay to determine the OCR are indicated. **f**. OCR/ECAR ratio under basal conditions for the indicated ESC lines at three different time points. Bioenergetic assays were performed using a SeaHorse XFp analyzer. In **b** and **d**, Student's t-test was performed. Representative data from two independent experiments are represented, and three technical replicates were analyzed. ****p<0.0001; n.s., non-significant.

Importantly, these results were validated using an independent *Zmym2* null ESC line (**Figure R21**). To this end, I used the CRISPR/Cas9 nuclease system to generate a *Zmym2* KO ESC line. First, I designed a sgRNA specifically targeting the start codon site of the *Zmym2* gene, located at Exon 2 (**Figure R21a**). RT-qPCR (**Figure R21b**) and western blot (**Figure R21c**) analysis confirmed complete depletion of ZMYM2 in the selected ESC clone. Notably, consistent with the gene trap data, depletion of ZMYM2 caused an increase in glycolytic activity and a reduction in OxPhos activity in ESCs (**Figures R21d- R21f**). Collectively, these results support an important function of ZMYM2 in the regulation of the cellular metabolic state for the reinforcement of pluripotent cell identity.

7.4.7 ZMYM2 regulates histone citrullination

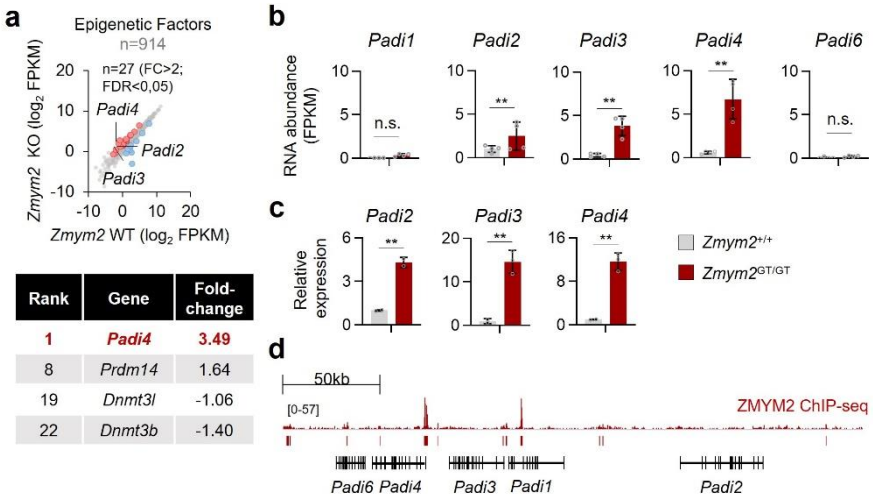
My ZMYM2 interactome contained several ESCs regulators that play important roles in pluripotency and reprogramming (**Figure R17**). Among them, similar to the functions of ZMYM2 described in this thesis, the zinc finger protein ZFP281 was reported to be a key transcriptional regulator for pluripotency *in vitro* and a roadblock to efficient pluripotent acquisition during reprogramming[289,373]. Interestingly, ZFP281 was also shown to interact with the DNA dioxygenase TET1 and negatively regulate *Tet2* expression through the action of HDACs to establish and maintain primed pluripotency[86]. Taken into account, I hypothesized the existence of a parallel molecular mechanism for ZMYM2 in the control of the expression of some epigenetic regulators that could be important for the transcriptional regulation of both ZMYM2 targets and non-targets in ESCs. To test this

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

hypothesis, I first analyzed the expression changes of 914 known epigenetic factors in *Zmym2* null ESCs. Notably, we found 27 genes encoding epigenetic regulators that were significantly misregulated (fold-change>2; FDR<0.05) in *Zmym2*$^{GT/GT}$ ESCs when compared with their wild-type counterparts (**Figure R22a**, top panel). Among them, there were several epigenetic factors with important functions in pluripotency such as PRDM14, DNMT3b, DNMT3l, and PADI4 (**Figure R22a**, bottom panel). For example, a dual role for PRDM14 was highlighted both in the activation of genes associated with self-renewal of murine ESCs and in the repression of differentiation markers[368]. DNMT3B and DNMT3Lwere both found to be downregulated in ESCs cultured under naïve pluripotent conditions[386] and silencing of *Dnmt3b* enhances reprogramming efficiency[367]. PADI4 was recently reported to be important for naïve pluripotency and reprogramming[153]. Considering that the *Padi4* gene was the most upregulated hit in *Zmym2* null ESCs (**Figure R22a**, bottom panel), along with how this gene is transcriptionally regulated remains largely unexplored, prompted me to examine whether this epigenetic regulator is contributing in part to some of the transcriptional changes observed in this thesis under depletion of ZMYM2 in ESCs. Interestingly, I found that other members of the PADI family, such as *Padi2* and *Padi3*, were also significantly upregulated in *Zmym2* null ESCs (**Figures R22a** top panel and **R22b**). As highlighted in the introductory section, the peptidylarginine deiminases (PADI) convert positively charged arginine residues to neutrally charged citrulline on histone and non-histone protein substrates in a process that is known as citrullination or deimination[152,387]. In line with this, the significant presence of GO term linked to protein citrullination was already revealed in my previous GO analysis for biological processes in differentially expressed genes between *Zmym2* null and WT ESCs (**Figure R14d**). Then, taking advantage of the RNA-seq datasets, I confirmed that *Padi2*, *Padi3* and *Padi4* genes, but not the other two members of the *Padi* family (i.e., *Padi1* and *Padi6*) were significantly upregulated in *Zmym2* null ESCs (**Figure R22b**). These results were further confirmed by RT-qPCR

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

(**Figure R22c**). Furthermore, when I analyze the presence of ZMYM2 in the genomic region containing *Padi* genes, I was able to detect ZMYM2 peaks near the regulatory regions of *Padi* loci, suggesting a direct transcriptional repressive role of ZMYM2 regulating the expression of these genes (**Figure R22d**).



Rank	Gene	Fold-change
1	*Padi4*	3.49
8	*Prdm14*	1.64
19	*Dnmt3l*	-1.06
22	*Dnmt3b*	-1.40

Figure R22. *Padi2*, *Padi3*, and *Padi4* are regulated at the transcriptional level by ZMYM2. a. (*Top*) Scatter plot of gene expression (in $\log_2$ FPKM) in *Zmym2*$^{GT/GT}$ versus *Zmym2*$^{+/+}$ ESCs determined by RNA-seq. Four biological replicates were analyzed. Genes significantly (FDR<0.05) upregulated and downregulated are shown in red and blue, respectively. (*Bottom*) Table showing ranked genes encoding epigenetic factors along with their fold-change expression. **b**. RNA abundance levels (expressed in FPKM) of the five *Padi* family genes in wild-type (grey) and *Zmym2* knockout (red) ESCs as determined by RNA-seq. Four biological replicates were analyzed. **c**. Relative expression levels of indicated *Padi* genes as determined by RT-qPCR. Three technical replicates were analyzed. **d**. ZMYM2 peaks are present in the genomic region containing *Padi* genes as determined by ZMYM2 ChIP-seq. Data were visualized using the IGV software. In **b** and **c**, Student's t-test was performed. Data are represented as mean ± SD. ***p*<0.01; n.s., non-significant.

The association of ZMYM2 with the transcriptional repression of *Padi* genes prompted me to consider whether citrullination may modulate some of the effects caused by the lack of ZMYM2 in ESCs. First, I analyzed the expression levels of *Padi* genes in somatic cells (i.e., MEFs) and pluripotent cells (i.e., iPSCs and ESCs) using available

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RNA-seq datasets[226,339,340]. Notably, among the five *Padi* genes, I found a differential expression pattern of *Padi2*, *Padi3*, and *Padi4* between somatic and pluripotent cells with significantly high expression levels of these genes in ESCs and iPSCs (**Figure R23a**), suggesting that in addition to the reported role of PADI4 in pluripotency[153], PADI2 and PADI3 may also be relevant epigenetic factors in this cellular context. This is in line with the high conservation level of the members of the PADI family, as indicated by comparative analyses of their amino acid sequences (**Figure R23b**), which also suggest the possibility of redundant physiological functions.

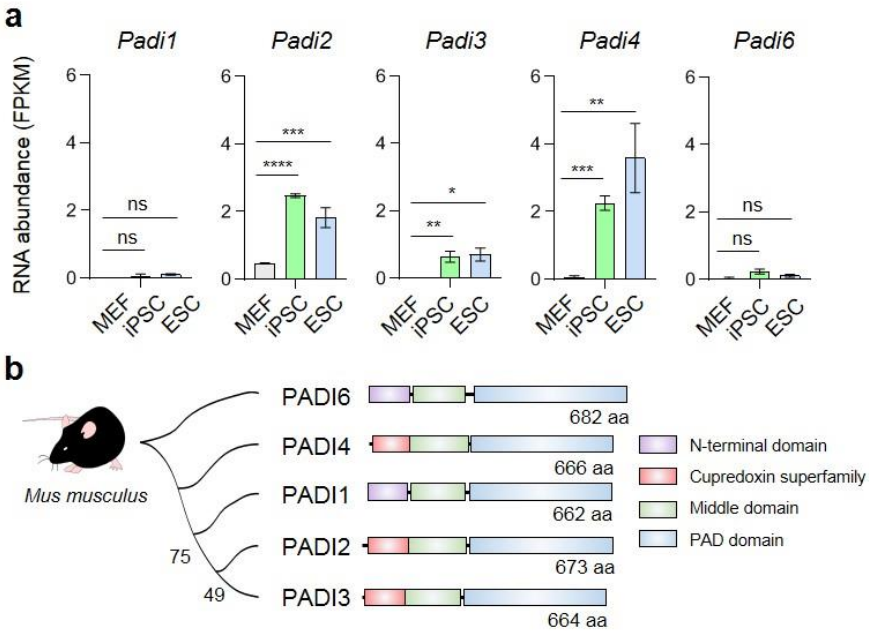


Figure R23. Analysis of *Padi* genes expression profiles and evolutionary conservation. a. RNA abundance levels (expressed in FPKM) of *Padi* genes in somatic cells (MEFs) and pluripotent cells (i.e., iPSCs and ESCs) as determined by RNA-seq. Data are represented as mean ± SD. The student's t-test was performed. **p*<0.05; ***p*<0.01; ****p*<0.001; *****p*<0.0001; n.s., non-significant.**b**. Analysis of evolutionary conservation of the protein members of PADI family in *Mus musculus*. A phylogenetic tree is shown accompanied by the scheme of conserved protein domains.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Given that my results showed significant upregulation of *Padi2*, *Padi3*, and *Padi4* genes in *Zmym2* null ESCs (**Figures R22a** and **R22b**), I next asked whether the lack of ZMYM2 could alter histone citrullination levels in ESCs. Taking into account that only PADI proteins located in the nucleus would be expected to citrullinate the histone proteins, I first performed an *in silico* approach[388] to predict the existence of nuclear localization signal (NLS) in PADI proteins. I found that among the PADI members, only PAD1, PADI3, and PADI4 are predicted to possess mono- and/or bipartite NLSs in their amino acid sequences (**Figures R24a** and **R24b**). Notably, among them, PADI4 showed the highest confidence prediction score (>7) for a monopartite NLS sequence (**Figure R24a**). Consistent with my data suggesting that PADI4 is likely involved in post-translational modification of histones in chromatin, previous reports state that PADI4 is the only member of the PADI family capable of histone citrullination[389 149,153,390]. Taking everything into consideration, I turned my efforts to address whether the absence of ZMYM2 that leads to upregulation of the *Padi4* gene could be sufficient to alter the chromatin citrullination status in ESCs. Thus, I first analyzed the citrullination levels of arginines R2, R8, and R17 on histone H3 (hereafter "H3Cit"), a histone modification shown previously to be carried out specifically by PADI4[153,390], in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. Importantly, following my predictions, quantitative immunofluorescence analysis of single cells revealed a significant increase of H3Cit levels upon ZMYM2 depletion under standard pluripotent culture conditions (**Figures R24c** and **R24d**). Notably, similar results were obtained by conducting transient loss-of-function experiments using two independent shRNAs against *Zmym2* to reduce the likelihood of off-target effects in CJ7 ESCs (**Figure R24e**). Then, I also confirmed the role of ZMYM2 by modulating histone citrullination levels in ESCs using western blot as a complementary technical approach (**Figure R24f**). I next examined regions of H3Cit enrichment across the entire genome of *Zmym2* WT and null ESCs by performing chromatin immunoprecipitation followed by sequencing (ChIP-seq). Interestingly, among the genome regulatory regions with

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the highest increase of H3Cit in *Zmym2* null ESCs when compared with their wild-type counterparts was the *Nanog* locus (**Figure R24g**), which has previously been reported to be positively regulated by PADI4 and H3Cit[153], and here I have shown that it is upregulated in absence of ZMYM2 in different pluripotency contexts (e.g., acquisition, maintenance, and exit of pluripotency). Altogether, my data support an important role for PADI4-mediated histone citrullination in modulating the transcriptional differences detected between *Zmym2* WT and null ESCs.

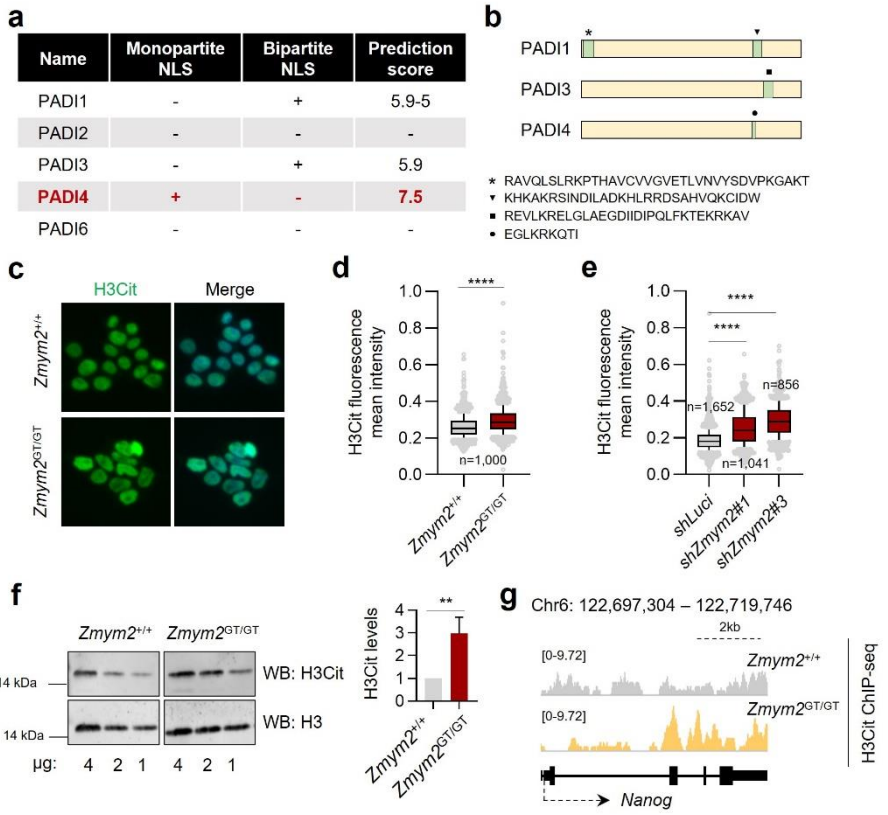


Figure R24. Absence of ZMYM2 increases H3Cit levels in ESCs. a. Table showing predicted NLS sequences in PADI proteins. **b.** Specific aminoacidic sequences of predicted NLS domains determined in **a**. **c.** H3Cit levels in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs as determined by single-cell quantitative immunofluorescence. H3Cit signals were normalized against DAPI nuclei staining. **d.** Quantification of H3Cit levels per

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

single cell as determined in **c**. **e**. H3Cit levels in CJ7 ESCs transduced with indicated shRNAs and determined as in **c**-**d**. **f**. (*Left*) Western blot analysis of H3Cit levels from histone fractions of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. H3 was used as the loading control. (*Right*) Quantification of H3Cit levels. Data are represented as mean ± SEM. **g**. H3Cit levels at the *Nanog* locus in *Zmym2* WT (grey) and null ESCs (orange) as determined by H3Cit ChIP-seq. Two biological replicates were analyzed. Data were visualized using the IGV software. In **d** and **e**, whiskers extend to the 10th–90th percentile range, and ''n'' indicates the number of cells analyzed per condition. In **d** and **e**, one-way ANOVA was performed, whereas in **f**, a two-tailed Student's t-test was performed. ***p*<0.01; *****p*<0.0001.

7.4.8 Lack of ZMYM2 leads to chromatin decondensation

In this work, I found that *Zmym2* depletion leads to the upregulation of the *Padi4* gene and increased H3Cit levels. Given that PADI4-mediated histone citrullination has been shown to facilitate chromatin decondensation in ESCs[153], I speculated that loss-of-function of ZMYM2 could influence the chromatin structure and facilitate its decompaction in ESCs. Therefore, I first performed nuclear visualization using 4',6-diamidino-2-phenylindole (DAPI) staining in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs (**Figure R25a**, left panel). Single-cell quantitative immunofluorescence of the nuclear mean radius revealed a significant increase of nuclear size in *Zmym2* null ESCs when compared with their wild-type counterparts, indicating a change in chromatin structure (**Figure R25a**, right panel). Then, to corroborate whether a larger nuclear size correlates with a more open global chromatin structure, I performed micrococcal nuclease susceptibility assays in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. To this end, chromatin fractions from these ESCs were isolated and then treated with a micrococcal nuclease enzyme. Remarkably, depletion of ZMYM2 increased digestion by micrococcal nuclease, indicating a more open overall chromatin structure (**Figure R25b**). Furthermore, I observed that *Zmym2* null ESCs even without treatment showed evident nucleosome fragments, probably due to the chromatin extraction process of these cells, which have greater chromatin accessibility than WT ESCs (**Figure R25b**). Altogether, these data establish an important role of ZMYM2 in the regulation of chromatin compaction in pluripotent cells.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

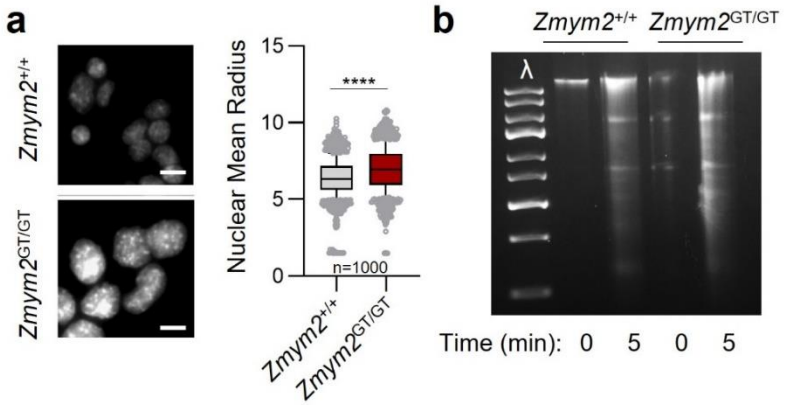


Figure R25. The absence of ZMYM2 promotes chromatin decompaction in ESCs and increases chromatin accessibility. a. (*left*) Representative DAPI staining of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. (*left*) Determination of nuclear size using single-cell quantitative immunofluorescence of indicated ESCs stained with DAPI. Whiskers extend to the 10th–90th percentile range. p-value between null and WT ESCs were calculated using a one-sided Wilcoxon signed-rank test. ****p<0.0001. ''n'' indicates the number of cells analyzed per condition. **b**. Micrococcal nuclease assay of chromatin fractions from *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs. Each cell type was treated with a micrococcal nuclease enzyme for 5 min at RTλ, 1kb DNA ladder.

7.4.9 PADI4 is regulated at the transcriptional level by ZMYM2

My previous data indicated that *Padi4* is significantly upregulated in *Zmym2* null ESCs (**Figures R22b** and **R22c**). Moreover, I found three ChIP-seq peaks of ZMYM2 binding to the *Padi4* locus (**Figure R22d**), suggesting that ZMYM2 is a regulator of *Padi4* expression by direct transcriptional repression. To further support this role of ZMYM2, I first analyzed the *Padi4* locus in more detail and observed two proximal ChIP-seq peaks of ZMYM2 binding to its regulatory regions (**Figure R26a**) bearing promoter marks (i.e., H3K4me3 and H3K27ac; data not shown). Moreover, the enrichment of ZMYM2 was examined by ChIP-qPCR using a set of primers for the amplicon containing the second ChIP-seq peak of ZMYM2 binding to the *Padi4* promoter region (denote as "ii") and another set of primers for an amplicon outside the *Padi4* locus (denote as "i") (**Figure R26a**).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

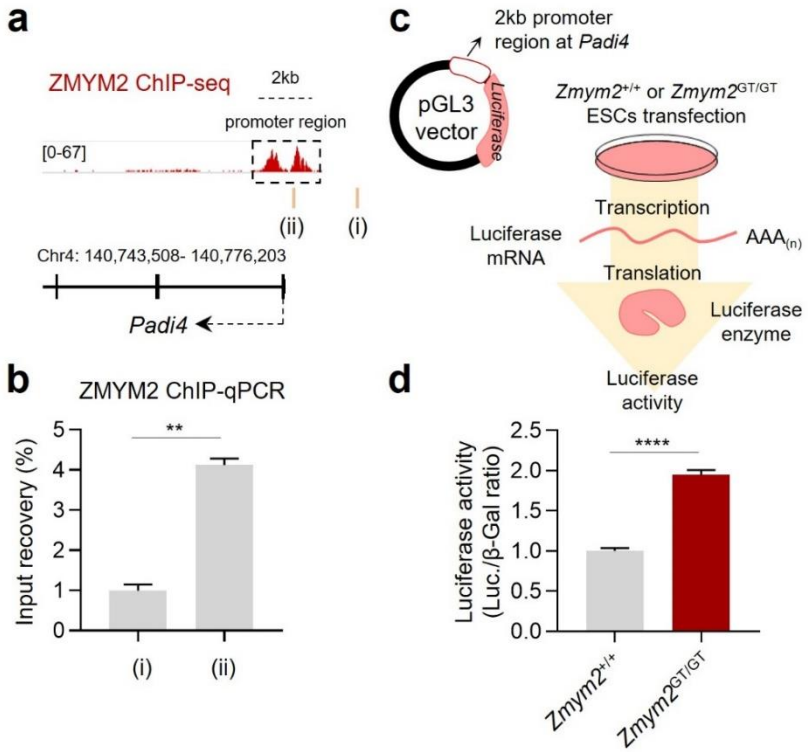


Figure R26. ZMYM2 directly represses the expression of *Padi4*. a. ZMYM2 peaks on the *Padi4* promoter region as determined by ChIP-seq. Two biological replicates were analyzed. Data were visualized using the IGV software. **b**. Enrichment of ZMYM2 at the regulatory locus of *Padi4* by ChIP-qPCR validating *Padi4* as a direct transcriptional target of ZMYM2. (i) denote amplicon for the negative control, and (ii) denote amplicon for the second ChIP-seq peak of ZMYM2 binding to the Padi4 promoter region. Technical triplicates were analyzed. **c**. Scheme showing the *Luciferase* reporter construct (*left*) used in the luciferase assay (*right*). *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs were co-transfected with the luciferase reporter and a ß-galactosidase reporter. c**d**. Levels of luciferase activity in indicated ESC lines. B-Galactosidase reporter was included for normalization (Luc./ß-Gal ratio). All error bars in this figure indicate average ± SD. *p* values were determined by Student's unpaired t-test. ***p*<0.01; *****p*<0.0001.

After chromatin immunoprecipitation using a ZMYM2 antibody, DNA enrichment was analyzed by RT-qPCR. As expected, I was able to confirm the enrichment of ZMYM2 at the *Padi4* promoter, but not outside this locus (**Figure R26b**). Finally, I further tested the transcriptional regulation of *Padi4* expression by ZMYM2 using a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

luciferase reporter assay whereby the luciferase gene is driven by the specific *Padi4* promoter region containing the two ChIP-seq peaks of ZMYM2 (**Figure 26c**). To this end, the ZMYM2 binding region in the *Padi4* promoter (highlighted in **Figure R26a**) was amplified and cloned into a pGL3 vector containing a luciferase reporter. This construct was then co-transfected with a β-galactosidase reporter for normalization in *Zmym2*$^{+/+}$ or *Zmym2*$^{GT/GT}$ to analyze differential luciferase activity. As expected, a significant increase in luciferase activity was observed in the absence of ZMYM2 (**Figure R26d**), further confirming the repressive role of ZMYM2 in this gene. Taken together, these data establish ZMYM2 as a major regulator of *Padi4* expression in pluripotency control by direct transcriptional repression.

7.4.10 ZMYM2 cooperates with other factors to regulate *Padi4* expression

To unravel the underlying mechanism by which ZMYM2 modulates *Padi4* expression, I first looked at previous studies on how ZMYM2 functions can be modulated. Interestingly, the recruitment of ZMYM2 to the chromatin in human transformed cells was reported to be dependent on its sumoylation[391], and recently hyposumolylation has been shown to enhance reprogramming and favor less compact chromatin in pluripotent cells[392], which are phenotypes similar to those due to the loss-of-function of ZMYM2 described in the previous sections. Taking everything into account, I speculated if the role of ZMYM2 in modulating the expression of *Padi4* depends on the SUMOylated state of this protein. To this end, I rescued *Zmym2*$^{GT/GT}$ ESCs with wild-type ZMYM2 or mutant forms of its three previously characterized SUMO-interacting motifs (SIM) (i.e., either dominant SIM2 or all three SIMs mutated)[391]. First, I validated the rescue lines and observed similar expression levels between *Zmym2* WT and mutant forms in *Zmym2* null ESCs (**Figure R27a**). Then, to my surprise, I found that ZMYM2 mutants could efficiently rescue *Padi4* expression (**Figure R27b**), which suggests that the SIMs in ZMYM2 are not required for its recruitment to chromatin in pluripotent cells in contrast to what was reported for human transformed cells[391]. Then, as expected

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

given the previous results regarding the expression of *Padi4*, I observed that both ZMYM2 WT and mutants were able to rescue *Nanog* expression in *Zmym2*$^{GT/GT}$ ESCs (**Figure R27c**). Taken together, these results show that the transcriptional repressor function of ZMYM2 on the *Padi4* locus is not dependent on its SUMOylated state, at least on these previously characterized SUMO-interacting motifs.

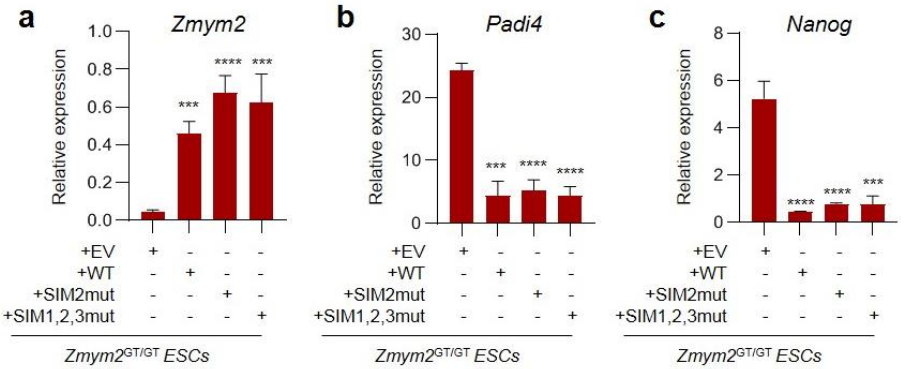


Figure R27. The ZMYM2 SIMs are not essential for the transcriptional regulation of *Padi4* in ESCs. a. Quantitative PCR analysis of *Zmym2* (**a**), *Padi4* (**b**) and *Nanog* (**c**) expression in *Zmym2*$^{GT/GT}$ ESCs rescued with 3×FL-EV (empty vector), 3xFL-Zmym2 WT (wild-type), 3xFL-Zmym2 SIM2mut or 3xFL-Zmym2 SIM1,2,3mut. The data are relative to expression in *Zmym2*$^{+/+}$ ESCs and *U6* as a housekeeping gene. All error bars in this figure indicate average ± SD. Technical triplicates were analyzed. p values were determined by one-way ANOVA with the Kruskal-Wallis correction for non-parametric distributions. ***p<0.001; ****p<0.0001.

Next, given my previous results showing that the ZMYM2-LSD1 partnership cooperates in the transcriptional repression of a subset of shared targets genes (**Figure R19**), I investigated whether LSD1 also participates in the regulation of *Padi4* expression. First, I used the LSD1 ChIP-seq dataset to examine its occupancy at the *Padi4* locus. Notably, in *Zmym2*$^{+/+}$ ESCs I observed two ChIP-seq peaks of LSD1 binding to the *Padi4* promoter (**Figure R28a**), following the same binding pattern that I also detected for ZMYM2 (**Figure R26a**). Interestingly, the overall LSD1 ChIP-seq signal at the binding *Padi4* promoter decreased in *Zmym2*$^{GT/GT}$ ESCs compared to that in WT ESCs (**Figure R28a**), suggesting that ZMYM2 is necessary for correct recruitment of LSD1 in the *Padi4* promoter. This data was further

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

validated by LSD1 ChIP-qPCR in *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ ESCs (**Figure R28b**). I then analyzed the expression of *Padi4* in *Lsd1*$^{+/+}$ and *Lsd1*$^{-/-}$ ESCs using published RNA-seq datasets[338]. Importantly, in accordance with my results showing that ZMYM2 directly represses the *Padi4* gene, I observed upregulation of *Padi4* in the absence of LSD1 (**Figure R28c**). Together, my results establish a negative role of ZMYM2 in restricting *Padi4* expression in ESCs by recruiting LSD1 for direct transcriptional repression of this gene.

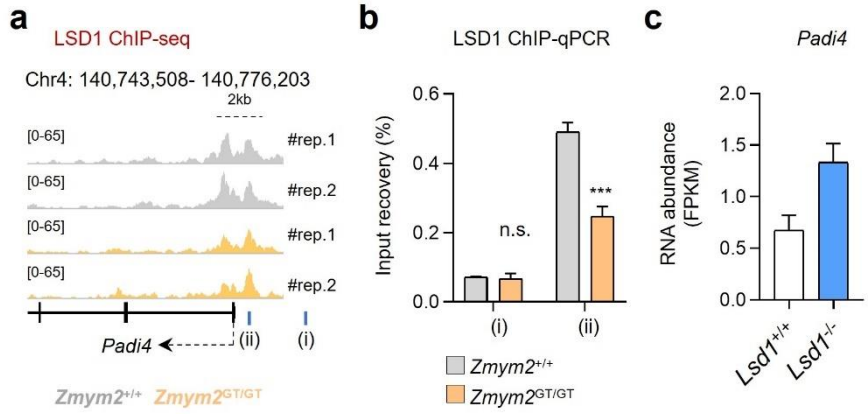


Figure R28. ZMYM2-depending binding of LSD1 to the *Padi4* promoter for its transcriptional control. a. LSD1 peaks on the *Padi4* promoter region as determined by ChIP-seq. Two biological replicates from *Zmym2*$^{+/+}$ (grey) and *Zmym2*$^{GT/GT}$ (orange) ESCs were analyzed. Data were visualized using the IGV software. **b**. Enrichment of LSD1 at the regulatory locus of *Padi4* by ChIP-qPCR validating *Padi4* as a direct transcriptional target of LSD1. (i) denote amplicon for the negative control, and (ii) denote amplicon for the second ChIP-seq peak of LSD1 binding to the *Padi4* promoter region. Technical triplicates were analyzed in indicated cell lines. p values were determined by Student's unpaired t-test. ***p<0.001; n.s., non-significant. **c**. Levels of *Padi4* transcripts in *Lsd1*$^{+/+}$ (white) and *Lsd1*$^{-/-}$ (blue) ESCs as determined by RNA-seq. Three biological replicates were analyzed. All error bars in this figure indicate average ± SEM.

To further explore whether other factors cooperate with ZMYM2 and LSD1 for direct repression of the *Padi4* gene, I examined my ZMYM2 interactome in ESCs (**Figure R17**). Among the high-confidence, ZMYM2-interacting partners were HDAC1 and HDAC2, previously reported to form the CoREST repressor complex with LSD1

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

in ESCs[393,394], and the zinc-finger protein ZFP281, which performs similar functions in pluripotency and reprogramming to those described in this thesis for ZMYM2[86]. To examine this possibility, I performed loss-of-function assays using previously validated shRNAs against *Hdac1*, *Hdac2,* and *Zfp281*, as well as sh*Lsd1* as a positive control and sh*Luci* as a negative control in *Zmym2* WT ESCs. As I expected, knockdown of *Lsd1* led to an increase in the expression of *Padi4* compared to ESCs infected with sh*Luci* (**Figure R29a**). Notably, I also observed significant upregulation of *Padi4* after the silencing of *Zfp281*, *Hdac1,* and *Hdac2* (**Figure R29a**), suggesting an interplay of ZMYM2-LSD1 with these factors for the transcriptional repression of *Padi4* in pluripotent cells.

In sum, I propose a working model for transcriptional control of the *Padi4* gene in pluripotent cells (**Figure 29b**). Mechanistically, transcriptional regulation of the *Padi4* locus is mediated by the ZMYM2-dependent binding of LSD1 to the *Padi4* promoter region. Furthermore, I speculated that the ZMYM2-LSD1 complex may facilitate the binding of HDAC1-HDAC2-ZFP281 to direct transcriptional repression of *Padi4* locus, but future studies are warranted to validate this mechanistic model.

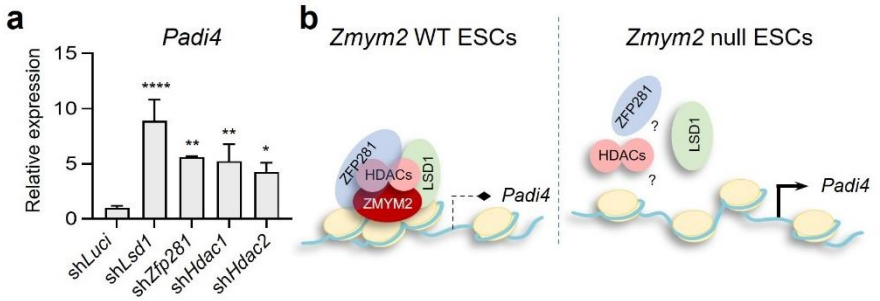


Figure R29. ZMYM2 may recruit LSD1-HDAC1/2-ZFP281 complex for direct repression of *Padi4* gene. a. Relative expression levels of *Padi4* in *Zmym2*$^{+/+}$ cells (grey) transduced with shRNAs against *Lsd1*, *Zfp281*, *Hdac1*, *Hdac2*, or sh*Luci* as control. *U6* was used as housekeeping. Three technical replicates were analyzed. One-way ANOVA was performed. *p<0.05; **p<0.01; ****p<0.0001; n.s., non-significant. **b.** A schematic hypothetical model of how ZMYM2 and its interacting partners mediate direct transcriptional repression of *Padi4* in ESCs.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

7.4.11 PADI4 inhibition partially restores transcriptional changes of *Zmym2* null ESCs

To understand how PADI4 contributes to the transcriptional changes promoted by the lack of ZMYM2 in ESCs (**Figures R13** and **R14**), I took advantage of a published microarray dataset where the authors identified genes under the control of PADI4 enzymatic activity using the chemical Cl-amidine, which disrupts citrullination by PADI4, in mouse ESCs under standard pluripotent culture conditions[153] Notably, among the 1,117 significant downregulated genes in Cl-amidine treated ESCs compared to vehicle control, I was able to identify 520 genes that were upregulated in *Zmym2* null ESCs compared to WT ESCs (**Figure R30a**, top panel). In contrast, among the 869 activated genes after chemical inhibition of PADI4, I found 270 genes that were downregulated in the absence of ZMYM2 (**Figure R30a**, bottom panel). With this in mind, I next focused my analysis on only those non-target genes of ZMYM2 that were differentially misregulated in *Zmym2*$^{GT/GT}$ compared to *Zmym2*$^{+/+}$ ESCs and differentially expressed under chemical inhibition of PADI4 in ESCs. Differential expression analysis revealed a set of 288 genes with specific expression patterns comprising two major groups: (1) 194 genes upregulated in *Zmym2* null ESCs whose activation is probably mediated by PADI4, and (2) 94 genes downregulated in *Zmym2* null ESCs whose repression is probably mediated by PADI4 (**Figure R30b**). Remarkably, Gen Ontology (GO) analysis of these misregulated genes associated with PADI4 enzymatic activity showed enrichment for metabolic processes and cell fate commitment (**Figure R30c**). Next, to better define how upregulation of *Padi4* in absence of ZMYM2 coordinates globally or locally gene regulation in pluripotency, I performed chemical inhibition of PADI4 in *Zmym2* null ESCs. To this end, I employed two new optimized and selective reversible inhibitors (i.e., GSK199 and GSK488) of PADI4 enzymatic activity[395]. Notably, I found that the presence of GSK484 at 10 µM for 24 hours in

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Zmym2$^{GT/GT}$ ESCs was sufficient to restore H3cit levels similar to the histone citrullination levels present in WT ESCs (**Figure R30d**).

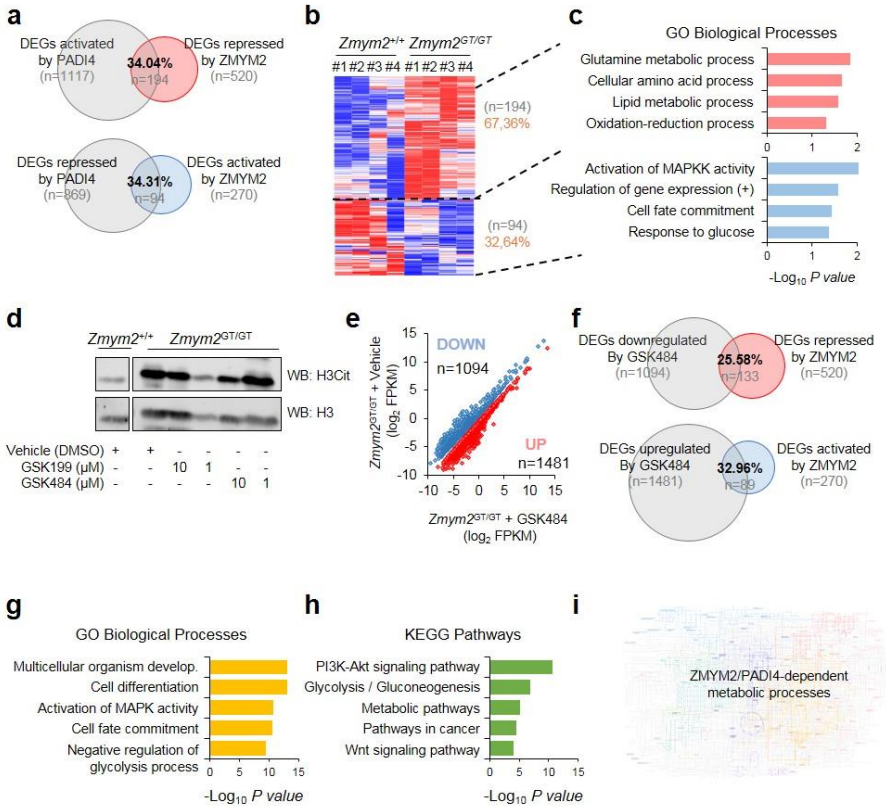


Figure R30. PADI4 is involved in part of the transcriptional changes promoted by the lack of ZMYM2. a. Ven diagrams showing the overlap between DEGs i) activated by PADI4 and repressed by ZMYM2 (*up*), and ii) repressed by PADI4 and activated by ZMYM2 (*bottom*). Data from a published microarray performed in mouse ESCs treated with Cl-amidine[153], a chemical inhibitor of citrullination, was used to determine the transcriptional impact of PADI4. **b.** Heatmap showing the expression levels in *Zmym2* WT and null cells of the 194 and 94 DEGs activated and repressed dependent on the PADI4 activity identified in **a**. Four biological replicates of *Zmym2*$^{+/+}$ and *Zmym2*$^{GT/GT}$ cells were used. **c.** Gene ontology analyses of the 194 upregulated (*top*) and 94 downregulated (*bottom*) genes associated with PADI4 activity **d.** Western blot showing the H3Cit levels in *Zmym2*$^{GT/GT}$ cells after the treatment with two different reversible inhibitors of PADI4 (GSK199 and GSK484). *Zmym2*$^{+/+}$ cells treated with DMSO vehicle were used as control and H3 levels were used as a loading control. **e.**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Scatter plot of gene expression levels (in $\log_2$ FPKM) in GSK484-treated versus DMSO-treated *Zmym2*$^{GT/GT}$ ESCs determined by RNA-seq. Two biological replicates were analyzed. Genes significantly upregulated (red) and downregulated (blue) (FDR<0.05) are shown. **f**. Ven diagrams show the overlap between i) commonly downregulated genes in ESCs treated with GSK484 and *Zmym2*$^{GT/GT}$ ESCs (*up*), and ii) activated genes in ESCs treated with GSK484 and *Zmym2*$^{GT/GT}$ ESCs (*bottom*). **g-h.** Analysis of gene ontology (GO) (**g**) and KEGG pathway (**h**) of the 222 DEGs by ZMYM2 (133 up- and 89 downregulated) whose expression levels were restored after chemical inhibition of PADI4 determined in **f**. **i**. Depiction of main metabolic pathways and processes that depend on the molecular functions of ZMYM2/PADI4. Restored metabolic pathways and processes upon PADI4 inhibition in *Zmym2* null ESCs are highlighted in color.

Then, I performed transcriptional profiling of *Zmym2*$^{GT/GT}$ ESCs treated with GSK484 and with DMSO as a control vehicle. I identified 1,481 upregulated and 1,094 downregulated genes that were significantly changed upon GSK484 treatment in *Zmym2* null ESCs (**Figure R30e**). Importantly, the restoration of histone citrullination levels in the *Zmym2*$^{GT/GT}$ ESCs was able to restore 25.58% and 32.96% of repressed and activated genes, respectively, by ZMYM2 (**Figure R30f**). Furthermore, in line with the previous results, GO and KEGG pathway analysis of these 222 genes revealed an enrichment of differentiation and metabolic categories (**Figures R30g** and **R30h**). Taken together, my data indicate that the reduction of PADI4 activity in *Zmym2*$^{GT/GT}$ ESCs partially restores proper expression of the transcription networks related to pluripotency and energy metabolism in ESCs, including metabolic programs that are closely related to the functions of ZMYM2 and PADI4 (**Figure R30i**).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8. DISCUSSION

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Although a handful of zinc-finger proteins (ZNFs) have been implicated in the molecular underpinnings that govern the acquisition, maintenance, and differentiation of pluripotency, the specific role of zinc finger MYM-type containing proteins in the molecular control of pluripotency has been largely unexplored. Here, whereas I found that the expression patterns of several members of this ZNF family (i.e., *Zmym1*, *Zmym2*, and *Zmym6*) are upregulated in pluripotent cells (i.e., ESCs and iPSCs) compared to somatic cells (i.e., MEFs), ZMYM2 was the only ZMYM member capable to interact with OCT4 and NANOG, suggesting that this factor may play important roles in controlling pluripotency and reprogramming. Therefore, in this thesis, I have performed a detailed functional characterization of the ZMYM2 for its roles in the maintenance and differentiation of mouse ESCs, as well as during the acquisition of pluripotency. Importantly, in this study, I provide several lines of evidence that highlight the central role of this protein in controlling pluripotent cell identity. First, my results show that while ZMYM2 is dispensable for the maintenance of ESCs under standard culture conditions (i.e., presence of serum and LIF), it negatively regulates the self-renewal state of pluripotent cells. Depletion of *Zmym2* in ESCs causes dysregulation of pluripotency and lineage-specific markers and enhances self-renewal even under differentiating clues (e.g., absence of LIF). Second, I show that ZMYM2 is essential for the establishment and maintenance of primed pluripotent state resembling post-implantation epiblast cells. Indeed, *Zmym2* null ESCs reprogrammed into primed epiblast-like cells (EpiLCs) using a serum-free medium containing the cytokines fibroblast growth factor 2 and Activin A, and the absence of LIF yielded significantly more naïve compact dome-shaped colony numbers than the typical primed diffuse-shaped colonies. Furthermore, the expression of lineage-primed markers was severely compromised in *Zmym2* null cells during the primed transition and these cells could not be maintained under primed culture conditions for more than six days. Third, my data reveal that the presence of ZMYM2 is required for the proper differentiation of ESCs *in vitro*, demonstrating that the lack of

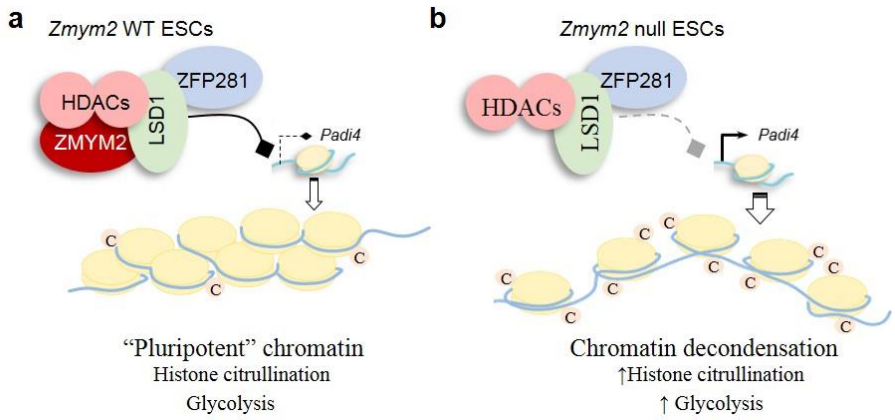
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ZMYM2 compromises early cell fate specification of the neural ectoderm and mesendoderm lineages, mainly due to a failure in the silencing of pluripotency regulators in these processes. Fourth, I identify that ZMYM2 functions as a molecular barrier during somatic cell reprogramming for the acquisition of pluripotency. In particular, the loss-of-function of ZMYM2 during somatic cell reprogramming leads to higher iPSC generation efficiencies and accelerates the kinetics of the reprogramming process.

At the molecular level, I show that depletion of ZMYM2 leads to transcriptional changes in genes related to development, cellular metabolism, and DNA organization, resulting in the alteration of the pluripotent cellular identity of murine ESCs. Mechanistically, my study demonstrates that ZMYM2 can act as a transcriptional activator and repressor of genes involved in the molecular control of pluripotency and somatic cell reprogramming. Furthermore, I identify that ZMYM2 physically interacts with different transcription factors and epigenetic regulators with known prominent roles in gene expression regulation in controlling both pluripotency and reprogramming. Notably, I find a physical association of Lysine-specific demethylase 1 (LSD1) and ZMYM2. Here, my results support that ZMYM2 recruits LSD1 to a subset of shared target genes for modulating ZMYM2 transcriptional regulatory functions that are related to the maintenance of pluripotency and energy metabolism, among others. The absence of ZMYM2 induces a shift in the basal metabolism of ESCs involving an increase of glycolytic activity accompanied by a reduction of OxPhos activity. Moreover, my data demonstrate a cooperative role between ZMYM2 and LSD1 in the transcriptional repression of the *Peptidyl arginine deiminase 4* (*Padi4*) gene, which was previously shown to be part of the pluripotency transcriptional network and mediates profound chromatin decondensation through histone citrullination[153]. In accordance with this, my results evidence that activation of *Padi4* in the absence of ZMYM2 is accompanied by higher histone citrullination levels and chromatin decondensation than WT ESCs. Remarkably, my

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

data show that reducing PADI4 activity in *Zmym2* null ESCs partially restores proper expression of transcriptional networks in ESCs, including those related to cell differentiation and metabolic processes. These results uncover a new molecular mechanism underlying the ZMYM2-LSD1 partnership for fine-tuning *Padi4* expression that affects the epigenetic configuration and chromatin organization of pluripotent cells and links energy metabolism with its potential to influence reprogramming and pluripotent cell fate outcomes (**Figure D1**).



FigureD1. A model illustrating the transcriptional repression role of ZMYM2 and interacting partners in regulating *Padi4* expression that influences the epigenetic and metabolic features of ESC identity. a. In wild-type (WT) ESCs, ZMYM2 cooperates with LSD1 repressor complex and ZFP281 for fine-tuning *Padi4* expression that controls chromatin compaction and energy metabolism in pluripotent cells. **b.** In the absence of ZMYM2, *Padi4* is upregulated, leading to decondensed chromatin through increased levels of histone citrullination and contributing to a shift in the basal metabolism of ESCs.

8.1 THE RELEVANCE OF ZMYM2 IN SELF-RENEWAL AND DIFFERENTIATION

Molecular control of the pluripotent state is thought to reside in a core circuitry of master transcription factors (TFs) including OCT4, SOX2, and NANOG[61,396]. Whereas the roles of these TFs in pluripotency have been extensively investigated, comparatively little is known about their interacting partners and their contribution to self-renewal and early

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

differentiation processes. In this thesis, I was particularly interested in dissecting the unknown functions of ZMYM2, the only member of the zinc finger MYM-type containing proteins capable of interacting with OCT4 and NANOG, in pluripotency (**Figure R1**). My results show that while ZMYM2 is dispensable for maintenance of murine ESCs under standard culture conditions (i.e., presence of serum and LIF) (**Figure R3**), it negatively regulates the self-renewal state of pluripotent cells (**Figure R4**), and it is essential for early differentiation process (**Figures R7** and **R8**). In agreement with my findings, two recent studies also report that ZMYM2 inhibits mouse ESC self-renewal[397] and it is required for exiting pluripotency in mouse ESCs[397] and human ESCs[398].

Here, I observed that *Zmym2* depletion significantly affected the expression of OCT4 and NANOG targets in ESCs (**Figure R12d**), suggesting that ZMYM2 not only physically but also functionally interacts with OCT4 and NANOG for proper control of ESC maintenance. Although I did not observe *Oct4* transcriptional changes in *Zmym2* null ESCs (**Figure R4),** the genomic occupancy of ZMYM2 reveals that it can bind to the *Nanog* locus for its transcriptional repression (**Figures R4** and **R16**). Given that upregulation of *Nanog* relieves ESCs from their leukemia inhibitory factor (LIF) requirement and blocks ESC differentiation[30,70], it is tempting to hypothesize that ZMYM2-mediated *Nanog* repression is preventing some of the effects that I find in *Zmym2* null ESCs compared to their WT counterparts. However, the data reported by Lawrence and colleagues in mouse ESCs is still controversial as to whether the effect of ZMYM2 reducing ESC self-renewal and its requirement for proper exiting of pluripotency is through the regulation of *Nanog* expression[397]. Thus, future studies are warranted to test my hypothesis and also unravel this *Nanog* repression mechanism for the control of pluripotency. Related to this, among the ZMYM2-interacting partners with the highest confidence in ESCs was Lysine-specific demethylase 1 (LSD1) (**Figure R17**), which has also been shown to form part of the core of TFs that govern pluripotency[399].

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

My data support that ZMYM2 recruits LSD1 to a subset of shared target genes including *Nanog* for its transcriptional control (**Figure R18**). Furthermore, the zinc finger protein ZFP281 was another top interacting partner of ZMYM2 (**Figure R17**). Interestingly, ZFP281 was reported to mediate NANOG autorepression and its depletion showed phenotypes similar to those observed here in the absence of ZMYM2 in self-renewal and differentiation of ESCs[86,289,373]. Future studies will be needed to dissect to what extent the physical interaction between ZMYM2 (and/or LSD1) and ZFP281 contributes to the molecular underpinnings that govern pluripotent cell identity, and whether ZMYM2 is also required to direct OCT4 and NANOG to shared targets genes that are critically involved in the regulation of pluripotency.

8.2 ZMYM2 FUNCTIONS IN ALTERNATIVE PLURIPOTENT STATES

Murine pluripotent stem cells (PSCs), including ESCs and iPSCs, maintained under standard culture conditions (i.e., presence of serum and LIF) show heterogeneity among individual cells. For example, these PSCs spontaneously transit between the so-called "naïve" and "primed" pluripotent states, which largely resemble the *in vivo* pre- and post-implantation epiblast, respectively[31,61]. With lower frequency, another pluripotent subpopulation can also emerge spontaneously and resembles a more primitive two-cell (2C) stage of development[338,400]. Each PSC identity (i.e., naïve, primed, and 2C-like stages) differ in their capacity of self-renewal and developmental propensity and potency for differentiation into different lineages[31,61]. The existence of these interconvertible states with relevance for the study of developmental biology and regenerative medicine provides another powerful system to dissect the molecular underpinnings controlling cell identity and function *in vitro*. While Yang and colleagues recently reported that the lack of ZMYM2 and LSD1/HDAC corepressor complex enhances 2C-like cell population in ESCs through the activation of MERVL retrotransposon elements[342], I identify that ZMYM2 is essential for the establishment and maintenance of primed pluripotency (**Figures R5**

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and **R6**). Indeed, my data show that *Zmym2* null ESCs retain naïve features during reprogramming towards primed pluripotency (**Figures R5** and **R6**). In line with this, it has also been shown that ZMYM2 loss in human primed ESCs leads to transcriptional similarities to cells in the early naïve developmental state of pluripotency[398]. Furthermore, the knockdown of *Zmym2* has also been reported to have a pronounced positive effect on NANOG-induced reprogramming of mouse primed epiblast stem cells (EpiSCs) into naïve pluripotent cells[397]. In addition, previous studies investigating regulators of naïve and primed pluripotency found among their hits that ZMYM2 may play an important role in exiting naïve and establishment primed pluripotency[86,401]. However, it remains to be determined what are the key effectors of ZMYM2 function in the establishment and safeguarding primed pluripotency. In this regard, besides MERVL silencing by ZMYM2-LSD1 partnership[342], another recent independent study also shows that ZMYM2 participates with SETDB1 in the efficient silencing of endogenous retroviruses (ERVs) in mouse ESCs[402]. Given that LSD1 and SETDB1 are among the confidence interacting partners of ZMYM2 that I identified here (**Figure R17** and **Supplementary Table 8**), future work in dissecting the mechanistic action of ZMYM2-dependent ERV expression in these interconvertible naïve/primed/2C-like stages should provide additional means to manipulate PSC identity for therapeutic applications.

8.3 ZMYM2 ROLES DURING SOMATIC CELL REPROGRAMMING

My study uncovers that ZMYM2 depletion increases the efficiency of iPSC formation from mouse somatic mesenchymal cells using the OSKM reprogramming cocktail (**Figure R9b**). Furthermore, the kinetics of iPSC generation is accelerated in the absence of ZMYM2 (**Figure R9c**). Throughout the successful reprogramming process to obtain iPSCs from mesenchymal cells can be distinguished three trajectories (i.e., initiation, maturation, and stabilization)[223,359,360]. Remarkably, an earlier study using genome-wide RNAi screen to determine global effectors during the initiation stage of human

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

reprogramming (i.e., mesenchymal-to-epithelial transition, MET) also identified ZMYM2 as a repressor of the reprogramming process[326]. In particular, the authors show that combinatorial knockdown of five factors, including ZMYM2, was sufficient to enhance the reprograming efficiencies several folds using multiple reprogramming systems and parental somatic cell lines[326]. Taken together, the results of my study and Toh and colleagues support a conserved role for ZMYM2 as a reprogramming barrier in both mouse and human settings, respectively.

I also identified that ZMYM2 functions as a molecular barrier during the late stabilization phase (**Figure R10**), when intermediate reprogrammed cells or so-called "pre-iPSCs" have activated the pluripotency gene program and can successfully transit to iPSC[30,359]. Notably, my data reveal that *Zmym2* knockdown leads to upregulation of pluripotent factors, including *Nanog* and *Tet2* (**Figure R10c**), whose forced expression in pre-iPSCs have been shown to promote the transition from pre-iPSC to iPSC[30,79,362]. In particular, NANOG and TET1/2 cooperation increase DNA 5-hydroxymethylcytosine levels at common target loci (e.g., *Esrrb*) and facilitates the establishment of pluripotency from pre-iPSCs and also heterokaryon-based reprogramming[79]. In line with this, my results also demonstrate that the lack of ZMYM2 improves the reprogramming efficiency of the cell fusion system (**Figure R11**), providing additional support for the functional implication of ZMYM2 as a reprogramming barrier. Therefore, it is tempting to speculate whether ZMYM2 is negatively affecting NANOG-TET2 functions in active DNA demethylation to facilitate pluripotency acquisition. In this regard, a recent study also reported that forced expression of ZMYM2 completely prevents NANOG-mediated reprogramming of pre-iPSCs[397]. In addition, *Zmym2* knockdown could not overcome the lack of endogenous *Nanog* in the reprogramming of pre-iPSCs[397]. Future work will be necessary to delineate the molecular mechanism of ZMYM2 acting as a roadblock during the reprogramming process.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

8.4 THE INTERPLAY BETWEEN ZMYM2 AND THE CELL EPIGENOME

Accumulating evidence demonstrates that the mechanistic relationships between chromatin modifications and TFs underlie the regulation of transcription in pluripotency and reprogramming[62,403]. Here, to further elucidate the molecular mechanisms by which ZMYM2 regulates these processes, I identified 27 epigenetic regulators that are differentially expressed by the lack of ZMYM2 in ESCs (**Figure R22**). Among them, I focused in PADI4 because it was recently reported to be part of the pluripotency transcriptional network that regulates histone citrullination that favors nucleosome accessibility in ESCs[153]. In particular, Christophorou and colleagues demonstrated that citrullination by PADI4 is essential both in naïve self-renewal and reprogramming and in early embryo development[153]. Here, my data reveals that lack of ZMYM2 leads to hyperactivation of *Padi4* locus which is accompanied by increased levels of histone citrullination and a more open global chromatin structure (**Figures R22** and **R24-R25**). Mechanistically, my data support that ZMYM2 cooperates with LSD1-HDAC repressor complex and ZFP281 for the transcriptional repression of *Padi4* locus and therefore regulates the global chromatin decondensation of ESCs that impacts transcriptional regulation (**Figure D1***)*. In this respect, chemical inhibition of PADI4 activity in *Zmym2* null ESCs was sufficient to partially restore almost one-third of transcriptional changes promoted by loss of ZMYM2 function in ESCs (**Figure R30**). However, additional epigenetic mechanisms related to ZMYM2 are likely to operate in pluripotency. In this regard, my data also support that ZMYM2, in addition to the *Padi4* locus, may recruit LSD1 to a subset of shared target genes to modulate the transcriptional repression functions of ZMYM2 through the histone demethylase activity of LSD1 and/or LSD1-associated HDAC activity (**Figures R18** and **R19**). In line with this, Fan and colleagues recently also found that ZMYM2 recruits HDAC-containing complexes (i.e., LSD1-NuRD-HDAC1/2 repressor complex) to directly represses MERVL/2C genes expression and restrict the 2C-like cell population in ESCs and in

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

totipotent 2C embryos[342]. Moreover, another recent study also reported that loss of ZMYM2 in human ESCs led to genome-wide H3 hyper-acetylation at promoter regions, possibly in an LSD1-CoREST-HADAC1/2-mediated manner[398]. On the other hand, my data also suggest that the impact of ZMYM2 in transcriptional regulation can also be mediated by altering DNA methylation status. For example, the expression of *de novo* methylation enzymes (i.e., *Dnmt3b* and *Dnmt3l*) were significantly downregulated in *Zmym2* null ESCs compared with their WT counterparts (**Figure R22a**). In this sense, it is known that *Dnmt3b* and *Dnmt3l* are downregulated in naïve ESCs[386] and DNMT3 acts as a reprogramming barrier during iPSC generation[367], which is in line with my data showing ZMYM2 as a roadblock of the reprogramming process. Furthermore, similarly to ZMYM2, its interaction partner ZFP281 (**Figure R17**) is essential in the establishment and maintenance of primed pluripotency through its interaction with TET1/HDAC2 and its repressive role on *Tet2* by modulating DNA demethylation and histone deacetylation[86]. Nevertheless, considering that I have identified numerous epigenetic regulators (ERs) in my ZMYM2 interactome (**Figure R17** and **Supplementary Table 8**), as well as transcriptional alterations of several ERs in the absence of ZMYM2 in ESCs (**Figure R22a** and **Supplementary Table 7**), future studies are warranted to depict exactly how ZMYM2 acts directly or indirectly through the cell epigenome in pluripotency and reprogramming.

8.5 ZMYM2 AND CELLULAR METABOLISM

Although cellular metabolism has been considered for many years as a passive component of cell fate determination, accumulating evidence indicates that specific metabolic pathways are determinants in guiding cell fate outcomes[404,405]. While epigenetic regulators control the expression of key metabolism genes and their encoded pathways, emerging reciprocal evidence shows that cell metabolism influences gene expression by modulating the epigenome[405-408]. For example, histone methylation and acetylation are based on the use of S-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

adenosylmethionine (SAM) and acetyl-CoA as methyl and acetyl donors, respectively, and DNA methylases and DNA hydroxymethylases require SAM and α-ketoglutarate (α-KG) respectively[408,409]. However, the crosstalk between metabolite availability and epigenome configuration for transcriptional regulation is largely unknown. A robust example of the interdependent interactions between metabolic changes and epigenetics occurs in pluripotency[157,405,410-415]. Indeed, acquisition of a pluripotent cellular state from a somatic cell by OSKM ectopic expression (SCR) involves a drastic restructuration of the chromatin biology and RNA modifications that are accompanied by a metabolic transition from OxPhos to glycolysis[410,412,416]. Conversely, the opposite metabolic transition is observed during cellular differentiation[157,411,412]. Moreover, naïve and primed pluripotent states can be stabilized in culture and can be distinguished by epigenetic and metabolic features[413], and the 2C-like population has a very unique metabolic state showing reduced levels of oxidative and glycolytic activities[417]. However, our current understanding of the molecular underpinnings behind the crosstalk of metabolic signaling and epigenetic regulation is far from being understood.

Here, it is noteworthy that I uncovered an important novel role for ZMYM2 in coordinating energy metabolism in ESCs. Concretely, my data demonstrate that the loss of ZMYM2 leads to a shift in the basal metabolism of ESCs that implies an increase of glycolytic activity accompanied by a reduction of OxPhos activity (**Figures R20** and **R21**). At the molecular level, I found that ZMYM2 cooperates with LSD1 for the transcriptional control of a subset of shared target genes involved in the regulation of cellular metabolism processes activity (**Figures R19** and **R20**). Taking into account that interference with glycolysis or OxPhos processes has been shown to interfere with the transition from naïve to primed pluripotent state and exiting of pluripotency[413,418,419], it is reasonable to speculate that the observed defects of *Zmym2* null ESCs showed in this thesis with respect to these

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

processes are mediated in part by the functions of ZMYM2 in the control of energy metabolism. In this sense, considering the important role of cellular metabolism during iPSC generation[410,412,416], it would not be surprising if a similar metabolic shift occurs in somatic cells after the lack of ZMYM2, which could also contribute to the observed higher efficiency of iPSC formation compared to the reprogramming process where ZMYM2 is present. In this regard, LSD1 inhibition was previously shown to enhance reprogramming by facilitating the same metabolic switch from oxidative to glycolytic metabolism[420] that I observed in *Zmym2* null ESCs (**Figures R20** and **R21**). Although Sun and colleagues show that this metabolic shift is mediated in part by rescuing the expression of *Hif1α*, which is repressed by the H3K4 demethylation activity of LSD1, I did not detect significant alterations of *Hif1α* expression due to lack of ZMYM2. These differences between my work and Sun and colleagues could be due to the use of different cellular systems (i.e., ESCs vs pre-iPSCs, respectively), and future studies will be required to scrutinize this possibility. In addition, I also identified that the upregulation of *Padi4* was partly responsible for the increased glycolytic activity observed in *Zmym2* null ESCs compared to their WT counterparts (**Figures R20**, **R21**, and **R30**). Although my data establish that ZMYM2 is a novel key regulator of energy homeostasis in pluripotency, how the observed metabolic changes affect the epigenome reconfiguration to govern cell fate specification and which metabolites are involved in the crosstalk between epigenetic regulators and transcriptional control will require further investigation.

8.6 ZMYM2 IN DEVELOPMENT AND DISEASE

While important roles of the *woc* zinc finger protein (a.k.a. *without children*), the *Drosophila melanogaster* ortholog of mammalian zinc finger MYM-type proteins, has been established in the development and physiology of fly organisms[303,421,422], the functions of ZMYM2 in mammalian development and disease remain to be addressed.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

My results and recent studies[86,342,397,398,401,402] support that ZMYM2 is acting as a key transcription factor in the regulation of stem cell pluripotency, suggesting that ZMYM2 likely plays diverse roles in mammalian embryogenesis and developmental processes. Indeed, zygotic depletion of ZMYM2 was shown to arrest development at 2C-stage embryos[342], and analysis of the heterozygous mattings, in collaboration with Dr. Jianlong Wang (Columbia University, USA), indicated that *Zmym2* null embryos die around embryonic day 11.5 (data not shown). Although these data further suggest that ZMYM2 is a key regulator of early cell fate specifications during *in vivo* development, whether similar molecular mechanisms governed by ZMYM2 reported by me or others operate in these cellular contexts, and how these specification events can be mechanistically linked to ZMYM2 for proper early mammalian development remains an open question.

On the other hand, a recent report links loss of ZMYM2 functions with congenital anomalies of the kidney and urinary tract (CAKUT), which represents the most common cause of chronic kidney disease in the first three decades of life[423]. In particular, the authors found that *Zmym2* heterozygous mutant mice, which recapitulate human frameshift mutation, show a spectrum of CAKUT-like defects and also suggest that, in addition to the role of ZMYM2 in renal development, this transcription factor can also be important in craniofacial development[423]. Notably, ZMYM2 was initially identified by its oncogenic role in various subtypes of leukemia and lymphoma as a result of a translocation event between the *Zmym2* gene (a.k.a. *Znf198*) and fibroblast growth factor receptor-1 (*Fgfr1*) gene[371]. Since then, the fusion of ZMYM2 with FGFR1 has been associated with several rare and aggressive myeloproliferative syndromes[424,425]. Although this and other fusion events with ZMYM2 (e.g., ZMYM2-FLT3) have been reported as oncoproteins in hematologic malignances[313,426-429], the molecular underpinnings of these cellular transformation processes are far from being understood. Taken together, it is tempting to speculate

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

that alterations in ZMYM2 functions may be involved in oncogenic transformation. In this regard, unlike normal differentiated cells, which rely primarily on mitochondrial oxidative phosphorylation to generate the energy necessary for cell proliferation, most cancer cells rely on anaerobic glycolysis, a phenomenon called the "Warburg effect", which has been associated with chemoresistance in cancer[430]. Given that I here establish a key role for ZMYM2 in controlling energy metabolism in pluripotent cells, perhaps a similar phenomenon is involved in specific types of cancers associated with ZMYM2 mutations. In addition, several interacting partners of ZMYM2 including LSD1 and ZFP281 have been also involved in cancer and various other pathological processes such as obesity and type 2 diabetes[431-433]. Future experimental work is needed to delineate the precise role of ZMYM2 and its interacting partners in cancer and other pathophysiological conditions.

In sum, the molecular programs governed by ZMYM2, discovered in this thesis, may also have functional implications under pathological contexts and shed new light on early development processes.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

9. CONCLUSIONS

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1. ZMYM2 is required for proper control of ESC maintenance.
2. ZMYM2 is critical for the establishment and maintenance of primed pluripotency.
3. The absence of ZMYM2 affects the exit from pluripotency.
4. ZMYM2 depletion promotes iPSC generation and enhances heterokaryon-based reprogramming.
5. ZMYM2 participates in the regulation of transcriptional programs in ESCs.
6. ZMYM2 can act both as a transcriptional activator and a repressor in stem cell pluripotency.
7. LSD1 cooperates with ZMYM2 for the transcriptional repression of shared target genes involved in pluripotency, including *Nanog* and *Padi4*.
8. ZMYM2 regulates histone citrullination through the transcriptional regulation of *Padi4*.
9. Lack of ZMYM2 leads to chromatin decondensation and metabolic shift from oxidative to glycolytic metabolism in ESCs.
10. Reduction of PADI4 activity in *Zmym2* null ESCs partially restores proper expression of transcriptional networks in ESCs.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

10. APENDIX

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

10.1 SUPPLEMENTARY TABLE 1. REAGENTS

REAGENTS		
REAGENT	SOURCE	REFERENCE
2-Mercaptoethanol	Sigma-Aldrich	A1910
5X JetPEI	Sigma-Aldrich	408727
10X NEBuffer 2	NewEnglandBioLabs	B7002S
37% Formaldehyde	Sigma-Aldrich	F8775
Accutase solution	Sigma-Aldrich	A6964
Acetone	Fisher Chemical	A/0560/17
Acrylamide:Bis-acrylamide, 30% solution	Sigma-Aldrich	A3574
Activin A	Invitrogen	PHC9564
Adenosine	Sigma-Aldrich	A4036
Agarose	Fisher BioReagents	BP1356
ATP	Thermo Scientific	R0441
B-27 Supplement (50X)	Gibco	17504044
Benzonase Nuclease	Sigma-Aldrich	E8263
bFGF	Invitrogen	PHG0026
BMP4 recombinant human protein	Gibco	11330852
Bovine Serum Albumin (BSA)	Fisher BioReagents	BP1605
BSA, Fraction V	Fisher BioReagents	BP1605
Bradford reagent	Sigma-Aldrich	B9616
Cell lysis solution	Puregene Accesories	158908
CHIR99021	Sigma-Aldrich	SML1046
Citrate	Sigma-Aldrich	915
Cytidine	Sigma-Aldrich	C4654
Cytosine ß-D-arabinofuranoside (Ara-C)	Sigma-Aldrich	C1768
D-luciferin substrate	Thermo Scientific	88293
DAPI	Sigma-Aldrich	D9542
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D8418
DNA hydration solution	Qiagen	158916
DNase I	Thermo Scientific	EN0521
DreamTaq Green PCR master mix (2X)	Thermo Scientific	K1081
Dulbecco's Modified Eagle's Medium (DMEM) – high glucose	Corning	10-013-CVR
DMEM/Nutrient Mixture F-12 Ham	Sigma-Aldrich	D6421
Dulbecco's Phosphate-Buffered Salt (DPBS)	Corning	21-031-CV
DPBS with $MgCl_2$ and $CaCl_2$	Sigma-Aldrich	D8662
Dynabeads Protein G	Invitrogen	10004D
ECL Select Western Blotting Detection Reagent	GE Healthcare	12644055/ RPN2235
EDTA, 0.5M Solution, pH 8.0	Sigma-Aldrich	E7889
EGTA. 0.5M Solution, pH 8.0	Alfa Aesar	J60767
Ethanol	Sigma-Aldrich	459844
Ethidium bromide	Fisher Bioreagents	BP1302-10
Ethanol	Sigma-Aldrich	108543
Fetal bovine serum (FBS)	Gibco	10270106
Fibronectin	Millipore	FC010

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FRV-Alkaline Solution	Sigma-Aldrich	86R
Gelatin from bovine skin	Sigma-Aldrich	G9391
GelCode Blue Safe Protein Stain Buffer	Thermo Scientific	24594
Glycerol	Sigma-Aldrich	G5516
Glycogen	Roche	10901393001
Glycylclycine pH8.8	Alfa Aesar	J62920
GSK199	Cayman Chemical	CAY-17489
GSK484	Cayman Chemical	CAY-17488
Guanosine	Sigma-Aldrich	G6264
HEPES Buffer, 1M Solution, pH 7.3	Fisher Bioreagents	BP299-100
HEPES Buffer, 1M Solution, pH 7.9	Cytiva	SH30237.01
HEPES	Fisher Bioreagents	BP310-100
Hygromycin B	Corning™	30-240-CR
2-propanol	Fisher Bioreagents	BP2618-212
JetPrime Reagent	Polyplus	114-07
KCl	Fisher Chemical	P/4240/60
KO-DMEM	Gibco	10829018
L-glutamine solution	Sigma Aldrich	G7513
Lipofectamine 3000	Invitrogen™	15282465
Luminata Crescendo Western HRP	Merck	WBLUR0100
MEM non-essential amino acids solution 100X	Gibco	11140050
Methanol	Fisher Chemical	M-4056-17
$MgCl_2$	Acros Organics	223210010
MTT	Sigma-Aldrich	M5655
N-2 supplement (100X)	Gibco	17502048
N-lauroylsarcosine	Sigma-aldrich	L9150
NaCl	Fisher Bioreagents	BP358-1
Naphtol AS-Bl Alkaline Solution	Sigma-Aldrich	86R
NP-40	Thermo Scientific	85124
Neurobasal™ medium	Gibco	21103049
O-nitrophenyl-ß-D-galactopyranoside (ONPG)	Thermo Scientific	34055
Ouabain octahydrate	Sigma-Aldrich	O3125
Paraformaldehyde	Sigma-Aldrich	P6148
PD0325901	Sigma-Aldrich	PZ0162
PEG 1500	Sigma-Aldrich	10783641001
Penicillin-Streptomycin (10,000U/mL)	Gibco	15140122
Phenol:chlorophorm:isoamyl alcohol	Fisher BioReagents	BP1752I
Polybrene	Sigma-Aldrich	H9268
Polyethylenimine (PEI)	Sigma-Aldrich	430599
Ponceau S	Sigma-Aldrich	P3504
Powder milk	LabScientific	M-0841
PowerUp™ SYBR™ Green Master Mix	Applied Biosystems	A25742
Protease inhibitor cocktail	Sigma-Aldrich	P8340
Protein precipitation solution	Qiagen	158912
Proteinase K	Amresco	E195
Puromycin dihydrochloride	Fisher Bioreagents	BP2956-100
Retinoic acid	SantaCruzBiotechnology	sc-200898
RIPA Lysis Buffer System	SantaCruzBiotechnology	sc-24948

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RNase A	Thermo Scientific	EN0531
SeaHorse XF DMEM Medium pH7.4	Agilent	103575-100
Sodium chloride (NaCl) 5M	Invitrogen	AM9760G
Sodium-deoxycholate	Sigma-Aldrich	D6750
SYNERGEL	Diversified Biotech	DIVESYN-100
Thiazolyl blue tetrazolium bromide (MTT)	Sigma-Aldrich	M5655
Thymidine	Sigma-Aldrich	T1895
Tris-HCl pH8.0	Thermo Scientific	J60080.AK
Triton™ X-100	Sigma-Aldrich	T8787
TRIzol reagent	Invitrogen™	15596026
Trypan Blue 0.4& Solution	VWR Chemicals	K940
Trypsin-EDTA Solution	Sigma-Aldrich	T3924
TWEEN® 20	Sigma-Aldrich	P1379
Uridine	Sigma-Aldrich	U3003

10.2 SUPPLEMENTARY TABLE 2. KITS AND ACCESSORIES

COMMERCIAL KITS		
NAME	SOURCE	REFERENCE
Centra Puregene Cell Kit	Qiagen	158388
ChIP DNA Clean & Concentrator	ZymoResearch	D5205
CloneJet PCR cloning kit	Thermo Scientific	K1231
E.Z.N.A.® Total RNA I Kit	Omega Bio-Tek	R6834-02
EZ Nucleosomal DNA Prep Kit	Zymo Research	D5220
E.Z.N.A.® Total RNA Kit I	Omega Bio-Tek	R6834-02
GenElute™ Endotoxin-free Plasmid Midiprep Kit	Sigma-Aldrich	PLED35
GenElute™ Gel Extraction Kit	Sigma-Aldrich	NA1111
GenElute™ HP Plasmid Miniprep Kit	Sigma-Aldrich	NA0150
Gibson Assembly Master Mix	NewEnglandBioLabs	E2611S
Histone Extraction Kit	Abcam	ab113476
jetPRIME® Transfection Kit	Polyplus	114-07
Leukocyte Alkaline Phosphatase Kit	Sigma-Aldrich	86R
PureLink™ HiPure Plasmid MidiPrep kit	Invitrogen	K210015
qSCRIPT® cDNA SuperMix	Quantabio	95048-025
Qubit RNA HS Assay Kit	Invitrogen	Q32855
Qubit RNA IQ Assay Kit	Invitrogen	Q33222
Rapid DNA Depho & Ligation kit	Roche	4898117001
Subcellular Protein Fractionation Kit	Thermo Scientific	78840
PCR Cloning Kit CloneJET™	Thermo Scientific	K1231
TOPO™ TA Cloning Kit for Subcloning	Invitrogen	450641
TOPO™ Zero Blunt™ PCR Cloning Kit	Invitrogen	K280020
XFp Cell Mito Stress Test Kit	Agilent	103010-100
XFp Glycolytic Rate Assay Kit	Agilent	103346-100
Zero Blunt TOPO PCR Cloning kit	Invitrogen	45-0245
ACCESSORIES		
NAME	SOURCE	REFERENCE
Amicon ® Ultra-15 Centrifugal Filter Unit	Millipore	UFC9030

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

0.22µm EMD Millipore Steriflip™	Millipore	SCGP00525
Corning™ Centrifuge Tubes with CentriStar™ Cap	Corning	430791
DynaMag™-2 Magnet	Invitrogen	12321D
Eppendorf® LoBind microcentrifuge tubes	Sigma	Z666505
Falcon® 5mL Round Bottom Polystyrene tubes	Falcon	352235
Immobilon-P PVDF Membrane	Millipore	IPVH00010
Invitrogen™ Qubit™ Assay Tubes	Invitrogen	Q32856
MicroAMP Fast Optical 96-well Reaction plates	Applied Biosystems	4346907
Mini-PROTEAN® TGX™ Precast Protein Gels	BIO-RAD	4561095S
Novex WedgeWell 4–20% tris-Glycine Gel	Invitrogen	XP04205BOX
Optical Adhesive Film	Applied Biosystems	4360954
Phasemaker™ Tubes for RNA	Phasemakers	A33248
Qubit™ 4 Fluorometer	Invitrogen	Q33226

10.3 SUPPLEMENTARY TABLE 3. CELL LINES

CELL LINE	DESCRIPTION	SOURCE
CJ7 ESCs	129S1/SvImJ male derived line	RRID:CVCL_C316
CCE ESCs	129/Sv derived line	N/A
Zmym2$^{+/+}$	CCE ESCs derived from 129/Sv male mice	Yang et al., 2020
Zmym2$^{GT/GT}$	CCE ESCs derived from 129/Sv male mice	Yang et al., 2020
Zmym2$^{GT/GT}$ + ^{3xFL}*Zmym2*	*Zmym2*$^{GT/GT}$ cells expressing an ectopic form of ZMYM2 protein	Yang et al., 2020
Zmym2 WT	CCE ESCs derived from 129/Sv male mice	This study
Zmym2 KO	CCE ESCs derived from 129/Sv male mice	This study
Nanog-GFP TNGA MEFs pre-iPSCs	MEFs pre-iPSCs carrying a GFP reporter under the control of the *Nanog* promoter	This study
Oct4-GFP MEFs	MEFs carrying a GFP reporter under the *Oct4* promoter	Jackson Laboratory, 008204
hB cells	Epstein-Barr virus (EBV)-transformed human B lymphocytes	N/A
HEK293T	Human embryonic kidney cells expressing the T antigen	ATCC, CRL-3216
Plat-E	Human embryonic kidney cells	Cell Biolabs,RV-101

10.4 SUPPLEMENTARY TABLE 4. RECOMBINANT DNA

PLASMID	FUNCTION	SOURCE	IDENTIF.
pCS2-myc-hZMYM2FL^{WT}	Overexpression	Dr. Sharrocks	pF1401
pCS2-myc-hZMYM2FL^{mutSIM2}	Overexpression	(Mancherter	pF1402
pCS2-myc-hZmym2FL	Overexpression	University, UK)	pF1403
pBASE	Transposon system	Jose C. Silva (Cambridge	N/A
pDEST (hygro)		Univ., UK)	N/A
PB-CAG-3XFLAG-EV-pA-pgk-h	Overexpression		N/A

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PB-CAG-3XFLAG- hZMYM2FL-WT-pA-pgk-hph	Overexpression	This study	pF1416
PB-CAG-3XFLAG- hZMYM2FL-mutSIM2-pA-pgk-hph	Overexpression	This study	pF1417
PB-CAG-3XFLAG- hZmym2FL-mutSIM1,2,3-pA-pgk-hph	Overexpression	This study	pF1418
PB-CAG-3XFLAG-mZMYM2-pA-pgk-hph	Overexpression	This study	pF058
PB-CAG-3XFLAG-NANOG-pA-pgk-hph	Overexpression	This study	pF1002
pSpCas9(BB)-2A-Puro (PX459)	CRISPR/Cas9	Addgene	#62988
PX330 + sg*Zmym2*#1	Expression sgRNA	This study	pF1095
PX330 + sg*Zmym2*#2	Expression sgRNA	This study	pF1096
pLKO.pim.shEV	Knockdown	Addgene	#32682
pLKO.pim.shLuci	Knockdown	This study	pF150
pLKO.pim.shHdac1	Knockdown	This study	pF816
pLKO.pim.shHdac2	Knockdown	This study	pF819
pLKO.pim.shLsd1	Knockdown	This study	pF102
pLKO.pim.shZfp281	Knockdown	This study	pF256
pLKO.pim.shZmym2#1	Knockdown	This study	pF680
pLKO.pim.shZmym2#2	Knockdown	This study	pF681
pLKO.pim.shZmym2#3	Knockdown	This study	pF682
pHAGE2-EF1aFull-hOct4-F2a-hKlf4-IRES-hSox2-P2A-hcMyc-W-loxP	Overexpression	Ihor Lemischka (ISMMS, USA)	N/A
pMD2.G	Generation of lentiviral particles	Addgene	#12259
psPAX2		Addgene	#12260
pMXs-cMyc	Overexpression	Addgene	#13375
pMXs-Klf4	Overexpression	Addgene	#13370
pMXs-Oct4	Overexpression	Addgene	#13366
pMXs-Sox2	Overexpression	Addgene	#13367
pCMV-ß-Galactosidase	Luciferase assay	Anxo Vidal (USC, Spain)	N/A
pGL3-Basic	Luciferase assay		N/A
pGL3-*Padi4*.promoter	Luciferase assay	This study	pF1450

10.5 SUPPLEMENTARY TABLE 5. PRIMARY AND SECONDARY ANTIBODIES

ANTIBODY	DILUTION	REFERENCE
Donkey anti-Mouse Alexa Fluor 488	1:10,000	JacksonImmunoResearch, 711-545-150
Donkey anti-Rabbit Alexa Fluor 488	1:10,000	JacksonImmunoResearch, 715-545-152
Mouse monoclonal AKT1/2/3	1:1,000	Santa Cruz Biotechnology, sc-81434
Mouse monoclonal ß-actin	1:2,000	Santa Cruz Biotechnology, sc-47778
Mouse monoclonal anti-ß-Tubuli	1:5,000	Santa Cruz Biotechnology, sc-55529
Mouse monoclonal ERK1/2	1:1,000	Santa Cruz Biotechnology, sc-514302

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Mouse monoclonal p-ERK1/2	1:1,000	Santa Cruz Biotechnology, sc-81492
Mouse monoclonal MEK1/2	1:1,000	Santa Cruz Biotechnology, sc-81504
Mouse monoclonal p-MEK1	1:1,000	Santa Cruz Biotechnology, sc-271914
Rabbit polyclonal anti-NANOG	1:10,000	Bethyl Laboratories, A300-397A-2
Mouse monoclonal anti-NANOG	1:500	Santa Cruz Biotechnology, sc-374103
Rabbit polyclonal anti-ZMYM2	1:2,000	Abcam, ab30783
Mouse monoclonal anti-FLAG	1:1,000	Sigma Aldrich, F1804
Rabbit polyclonal anti-H3Cit (citrulline R2 + R8 + R17)	1:1,000	Abcam, ab5103
Rabbit polyclonal anti-H3K14ac	1:2,500	Abcam, ab218055
Rabbit polyclonal anti-H3K27ac	1:2,000	Abcam, ab4729
Rabbit polyclonal anti-H3K27me3	1:2,000	Millipore, 07-449
Rabbit polyclonal anti-H3K4me1	1:1,000	Abcam, ab8895
Rabbit polyclonal anti-H3K4me3	1:1,000	Abcam, ab8580
Rabbit polyclonal anti-H3	1:20,000	Abcam, ab1791
Rabbit monoclonal anti-LSD1	1:2,000	Abcam, ab129195
Rabbit polyclonal anti-GAPDH	1:10,000	ProteinTech, 10494-1-AP
Rabbit polyclonal anti-ZMYM2	1:2,000	Abcam, ab30783
Rabbit polyclonal anti-OCT4	1:1,000	Abcam, ab19857

10.6 SUPPLEMENTARY TABLE 6. OLIGONUCLEOTIDES

GENE	SEQUENCE
	RT-qPCR
Afp F	CAGCAGCCTGAGAGTCCATA
Afp R	GGCGATGGGTGTTTAGAAAG
Brachyury (*T*) F	GCTTCAAGGAGCTAACTAACGAG
Brachyury (*T*) R	CCAGCAAGAAAGAGTACATGGC
Cxcl12 F	TTTCAGATGCTTGACGTTGG
Cxcl12 R	GCGCTCTGCATCAGTGAC
Dnmt3b F	CTCGCAAGGTGTGGGCTTTTGTAAC
Dnmt3b R	CTGGGCATCTGTCATCTTTGCACC
Eomes F	CCAAAACACGGATATCACCC
Eomes R	GACCTCCAGGGACAATCTGA
Esrrb F	CGATTCATGAAATGCCTCAA
Esrrb R	CCTCCTCGAACTCGGTCA
Fgf5 F	CTCAGGGGATTGTAGGAATACGAGGA
Fgf5 R	GGATCGCGGACGCATAGGTATTATAG
Foxa2 F	CGAGTTAAAGTATGCTGGGAG
Foxa2 R	TATGTGTTCATGCCATTCATCC
Gata4 F	CCCTACCCAGCCTACATGG
Gata4 R	ACATATCGAGATTGGGGTGTCT
Gata6 F	TTGCTCCGGTAACAGCAGTG

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Gata6 R	GTGGTCGCTTGTGTAGAAGGA
hCD19 F	GCTCAAGACGCTGGAAAGTATTATT
hCD19 R	GATAAGCCAAAGTCACAGCTGAGA
hCD45 F	CCCCATGAACGTTACCATTTG
hCD45 R	GATAGTCTCCATTGTGAAAATAGGCC
hCRIPTO F	AGAAGTGTTCCCTGTGTAAATGCTG
hCRIPTO R	CACGAGGTGCTCATCCATCA
hGAPDH F	TCTGCTCCTCCTGTTCGACA
hGAPDH R	AAAAGCAGCCCTGGTGACC
hNANOG F	CCAACATCCTGAACCTCAGCTAC
hNANOG R	GCCTTCTGCGTCACACCATT
hOCT4 F	TCGAGAACCGAGTGAGAGGC
hOCT4 R	CACACTCGGACCACATCCTTC
Klf4 F	AACCTTACCACTGTGACTGG
Klf4 R	AAAAGTGCCTCTTCATGTGT
Mixl1 F	ATCCGCCCGGACCCTCCAAA
Mixl1 R	TCGGTTCTGGAACCACACCTGGA
Nanog F	TTGCTCTTTCTGTGGGAAGG
Nanog R	CCAGGAAGACCCACACTCAT
Negative peak F	ATGGCCCCTGAGTACAGAGA
Negative peak R	ACAAGCGCTGGATTCTCATT
Nestin F	GAGTTCTCAGCCTCCAGCAG
Nestin R	AGATCGCTCAGATCCTGGAA
Oct4 F	TCTTTCCACCAGGCCCCCGGCTC
Oct4 R	TGCGGGCGGACATGGGGAGATCC
Otx2 F	GGCCTCACTTTGTTCTGACC
Otx2 R	AACTTGCCAGAATCCAGGGT
P16 F	ACTCCAAGAGAGGGTTTTCT
P16 R	CATCATCATCACCTGGTCC
P21 F	ATGTCCAATCCTGGTGATGTCCG
P21 R	AAGTCAAAGTTCCACCGTTCTCG
Padi2 F	CCGCCGGGTATGAAATAGTCC
Padi2 R	CGCCGGTGTACTTGACCAC
Padi3 F	CTACAGAGGATTGTGCGTGTG
Padi3 R	AGGAACCGCCCCATAAATGTC
Padi4 F	TCTGCTCCTAAGGGCTACACA
Padi4 R	GTCCAGAGGCCATTTGGAGG
Padi6 F	TGGTAGGCATGGAAATCACCT
Padi6 R	GACGGAGCTAGAGATGTGGAT
Pax3 F	GCAGCGCAGGAGCAGAACCA
Pax3 R	GCACTCGGGCCTCGGTAAGC
Positive peak F	TCTCTTCGGGACTTGCTCTC
Positive peak R	GGGTGTGATGATGTGAGCTG
Prdm14 F	ACAGCCAAGCAATTTGCACTAC
Prdm14 R	TTACCTGGCATTTTCATTGCTC
Rex1 (*Zfp42)* F	CAGCTCCTGCACACAGAAGA
Rex1 (*Zfp42*) R	ACTGATCCGCAAACACCTG
Sall4 F	AGTGATGTGGCTTGTGACCA

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sall4 R	AACCCGCTTCTTTCCAAAAT
Sox17 F	CGCACGGAATTCGAACAGTA
Sox17 R	GTCAAATGTCGGGGTAGTTG
Sox2 F	GGAGTGGAAACTTTTGTCC
Sox2 R	GGGAAGCGTGTACTTATCCT
Tbx3 F	TTATTTCCAGGTCAGGAGATGGC
Tbx3 R	GGTCGTTTGAACCAAGTCCCTC
Tet2 F	ACATCCCACCTCTGTCCATT
Tet2 R	GAACTATACATCACTGTGCGTCA
U6 F	CTCGCTTCGGCAGCACATATACT
U6 R	ACGCTTCACGAATTTGCGTGTC
Zmym2 CRISPR F	TTTGGCATGGACACAAGTTC
Zmym2 CRISPR R	CGTCATCATCATCTTCTACTGATG
Zmym2 GT F	ATGACAGGCTCAGCACCAC
Zmym2 GT R	CACAGGGACAGGAATAGGC

PCR

Padi4 promoter F	TCTACTCGGCACCCACTTTC
Padi4 promoter R	CTCTGGGTGCTCCCTGTTAG
pLKO insert F	TGCATATACGATACAAGGCTGTT
pLKO insert R	TCTTTCCCCTGCACTGTACC
Zmym2 arm1 F	CTGTCCTGTCCCCCTTACATTAG
Zmym2 arm1 R	CCGGAACCTCCTCCGCTCCCTTTTACTAGAAGAACCCGTACAAGC
Zmym2 arm2 F	AGTTCTTCTGATTCGAACATCTGAAGACACAGACTAAAAAGGACCA
Zmym2 arm2 R	ACTGTCTCCCTTCAAGTAGCCTG
Zmym2 F	AGGTTCAGTCAGCCAGTGCT
Zmym2 R1	ACCAAAGCTTGATGCCAGTT
Zmym2 R2	GGGTCTCTTTGTCAGGGTCA

shRNA

sh*Hdac1* F	CCGGGCCAGTCATGTCCAAAGTAATCTCGAGATTACTTTGGACATGACTGGCTTTTTG
sh*Hdac1* R	AATTCAAAAAGCCAGTCATGTCCAAAGTAATCTCGAGATTACTTTGGACATGACTGGC
sh*Hdac2* F	CCGGCGAGCATCAGACAAACGGATACTCGAGTATCCGTTTGTCTGATGCTCGTTTTTG
sh*Hdac2* R	AATTCAAAAACGAGCATCAGACAAACGGATACTCGAGTATCCGTTTGTCTGATGCTCG
sh*Lsd1* F	CCGGGAGTTGAAAGAGCTTCTTAATCTCGAGATTAAGAAGCTCTTTCAACTCTTTTTG
sh*Lsd1* R	AATTCAAAAAGAGTTGAAAGAGCTTCTTAATCTCGAGATTAAGAAGCTCTTTCAACTC
sh*Luciferase* F	CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTTG
sh*Luciferase* R	AATTCAAAAACCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG
sh*Zfp281* F	CCGGGCTGGGAGTCTAATGTAATTTCTCGAGAAATTACATTAGACTCCCAGCTTTTTG
sh*Zfp281* R	AATTCAAAAAGCTGGGAGTCTAATGTAATTTCTCGAGAAATTACATTAGACTCCCAGC
sh*Zmym2*#1 F	CCGGGCAATGGGATAACACAGCATTCTCGAGAATGCTGTGTTATCCCATTGCTTTTTG
sh*Zmym2*#1 R	AATTCAAAAAGCAATGGGATAACACAGCATTCTCGAGAATGCTGTGTTATCCCATTGC
sh*Zmym2*#2 F	CCGGCCATTATTCTATGTAGTGGTTCTCGAGAACCACTACATAGAATAATGGTTTTTG
sh*Zmym2*#2 R	AATTCAAAAACCATTATTCTATGTAGTGGTTCTCGAGAACCACTACATAGAATAATGG
sh*Zmym2*#3 F	CCGGGCGTGGAAACATTGGGTCAAACTCGAGTTTGACCCAATGTTTCCACGCTTTTTG
sh*Zmym2*#3 R	AATTCAAAAAGCGTGGAAACATTGGGTCAAACTCGAGTTTGACCCAATGTTTCCACGC

sgRNAs

TARGET	SEQUENCE	FUNCTION
Zmym2#1 (F)	CACCGCAACTAGTCTCACGAACGT	*Zmym2* KO ESCs
Zmym2#1 (R)	AAACACGTTCGTGAGACTAGTTGC	*Zmym2* KO ESCs

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Zmym2#2 (F)	CACCGCTGTACTTCCTAATAAGAC	*Zmym2* KO ESCs
Zmym2#2 (R)	AAACGTCTTATTAGGAAGTACAGC	*Zmym2* KO ESCs

10.7 SUPPLEMENTARY TABLE 7. LIST OF DEGS UPON *ZMYM2* DEPLETION IN ESCS

RNA-seq analysis comparing *Zmym2*$^{GT/GT}$ ESCs to *Zmym2*$^{+/+}$ ESCs under standard culture conditions containing serum and LIF. In the tables below, the most differentially expressed genes (DEGs) in *Zmym2*$^{GT/GT}$ ESCs (FDR<0.05 and >2-fold change alteration compared to WT ESCs) are shown.

List of the top 50 downregulated genes:

RANK	GENE	LOG_2 FOLD-CHANGE	FDR
1	*Hrnr*	-8.299134462	0.0058
2	*Ddx3y*	-6.312073885	0.0018
3	*Olig2*	-6.056538228	0.0018
4	*Uty*	-5.862759695	0.0018
5	*Eif2s3y*	-5.192739586	0.0018
6	*Eno1b*	-4.78920827	0.0018
7	*Olig1*	-4.603732695	0.0018
8	*Acta2*	-4.245254549	0.0018
9	*C1qtnf1*	-4.04385407	0.0071
10	*Rpph1*	-3.984225012	0.0018
11	*Col3a1*	-3.925964528	0.0018
12	*Insm1*	-3.707914223	0.0018
13	*Ghr*	-3.628471499	0.0018
14	*C3*	-3.614193137	0.0018
15	*Grb10*	-3.606352179	0.0018
16	*Col6a1*	-3.593757745	0.0018
17	*2610037D02Rik*	-3.486540971	0.0018
18	*Lpar6*	-3.465530739	0.0071
19	*Srgn*	-3.336105109	0.0033
20	*Fosb*	-3.305085097	0.0018
21	*Col1a2*	-3.257630891	0.0018

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

22	*Gata4*	-3.243638568	0.0018
23	*C4b*	-3.207946239	0.0018
24	*Pax6*	-3.168480987	0.0143
25	*Thbs1*	-3.15622219	0.0018
26	*Bgn*	-3.143393937	0.0018
27	*Prss23*	-3.122542375	0.0033
28	*Sema3e*	-3.064310603	0.0018
29	*Sfmbt2*	-3.036518661	0.0018
30	*Has2*	-3.028673731	0.0109
31	*Tchh*	-3.025236474	0.0018
32	*Cubn*	-2.979163723	0.0018
33	*Cdx2*	-2.978806941	0.0199
34	*Enpp2*	-2.971909022	0.0018
35	*Gm7120*	-2.843848997	0.0018
36	*Cxcl12*	-2.800855326	0.0191
37	*Sp5*	-2.786286604	0.0018
38	*Sema6a*	-2.773072823	0.0018
39	*Kbtbd11*	-2.73926328	0.0018
40	*Gsx1*	-2.735465298	0.0018
41	*Wt1*	-2.616520431	0.0018
42	*Dab2*	-2.596663679	0.0018
43	*Trim36*	-2.592560909	0.0018
44	*Spta1*	-2.585929414	0.0018
45	*Zfp951*	-2.546601931	0.0018
46	*Col6a2*	-2.504531513	0.0090
47	*Spink1*	-2.429837083	0.0018
48	*Foxi3*	-2.419167629	0.0018
49	*Rcbtb2*	-2.417985301	0.0018
50	*Eln*	-2.404623388	0.0018

List of the top 50 upregulated genes:

RANK	GENE	LOG_2 FOLD-CHANGE	FDR
1	*Tex13*	7.208921039	0.0018

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

2	*Usp26*	6.534727927	0.0018
3	*Bfsp2*	5.334334169	0.0452
4	*Rcsd1*	5.029186249	0.0018
5	*Ube2l6*	4.641372026	0.0018
6	*1700108F19Rik*	4.636430234	0.0018
7	*Arl14epl*	4.458665986	0.0018
8	*Gm364*	4.225340944	0.0018
9	*Cd200*	4.17812715	0.0046
10	*Lama3*	4.077479679	0.0018
11	*Smtnl1*	4.005209229	0.0018
12	*Matn1*	3.938876238	0.0033
13	*Hsh2d*	3.880729785	0.0018
14	*Hydin*	3.857527075	0.0018
15	*Txlnb*	3.825294154	0.0018
16	*Tmem108*	3.735368094	0.0018
17	*Ttc39c*	3.672912702	0.0018
18	*Gm5039*	3.651052899	0.0253
19	*Sh3tc2*	3.638563949	0.0018
20	*Klk13*	3.629491383	0.0018
21	*Trpm1*	3.628904503	0.0018
22	*Slc32a1*	3.578964265	0.0046
23	*Sohlh2*	3.549988063	0.0018
24	*Piwil1*	3.507133505	0.0018
25	*Padi4*	3.496037496	0.0018
26	*Grin2c*	3.495873415	0.0033
27	*Ly6k*	3.448353508	0.0018
28	*Hcrtr2*	3.446732089	0.0033
29	*Tktl2*	3.429413476	0.0241
30	*Sytl3*	3.415637991	0.0033
31	*Myoz3*	3.343848505	0.0169
32	*Ceacam20*	3.319887341	0.0018
33	*Ttc29*	3.281493228	0.0018
34	*Arhgef38*	3.214855744	0.0018

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

35	*Spesp1*	3.177390173	0.0143
36	*Tuba3a*	3.147382226	0.0018
37	*LOC105246506*	3.136852694	0.0018
38	*Gsx2*	3.133325809	0.0266
39	*Rhox10*	3.129171259	0.0081
40	*Sh3rf2*	3.128049919	0.0018
41	*Klk10*	3.120889097	0.0018
42	*Padi3*	3.083457531	0.0018
43	*Fxyd6*	3.074661787	0.0018
44	*Tbx3os2*	3.0691567	0.0018
45	*Mslnl*	3.023905892	0.0018
46	*Galnt6*	3.006459206	0.0018
47	*Pla2g4e*	2.989316614	0.0018
48	*Slc12a1*	2.954323929	0.0443
49	*LimS2*	2.938385188	0.0018
50	*Adrb1*	2.88683041	0.0033

10.8 SUPPLEMENTARY TABLE 8. LIST OF ZMYM2-ASSOCIATED PROTEINS

PROTEIN	SCORE	RANK
KDM1A	10.6	1
RCOR2	9.5	2
HDAC1	9.1	3
ZMYM4	8.2	4
MDN1	8.1	5
TUBB4B	8.0	6
TUBB2B	7.8	7
ZMYM3	7.5	8
RIF1	7.4	9
POGZ	7.3	10
4732456N10RIK	7.3	11
HDAC2	7.3	12
GTF3C1	7.1	13

PROTEIN	SCORE	RANK
TUBA1C	7.1	14
DYNLL1	7.0	15
TUBB3	6.9	16
ADNP	6.7	17
TUBB6	6.6	18
CHAMP1	6.5	19
NUMA1	6.5	20
GTF3C2	6.4	21
CHD4	6.1	22
HNRNPF	6.1	23
RCOR1	5.9	24
GTF3C4	5.8	25
GTF2IRD1	5.8	26

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROTEIN	SCORE	RANK
ATF7IP	5.8	27
PELP1	5.7	28
HSPA2	5.6	29
GTF3C3	5.6	30
EXOC4	5.6	31
CBX3	5.6	32
ZFP281	5.5	33
SMARCAD1	5.4	34
DYNLL2	5.2	35
RBBP4	5.0	36
NUP98	5.0	37
ATR	5.0	38
INTS1	5.0	39
EIF4A1	5.0	40
ZFP462	5.0	41
LAS1L	4.9	42
PPP1CA	4.9	43
L1TD1	4.9	44
ILF2	4.8	45
CHD8	4.8	46
PCBP2	4.8	47
GTF3C5	4.8	48
HNRNPK	4.8	49
EXOC2	4.8	50
ILF3	4.7	51
PPP1CC	4.6	52
COPG1	4.6	53
SALL4	4.6	54
PTBP1	4.5	55
MGA	4.5	56
EXOC1	4.5	57
COPE	4.4	58

PROTEIN	SCORE	RANK
UBB	4.4	59
WDR82	4.3	60
TMPO	4.3	61
GM8994	4.3	62
RBM14	4.3	63
HSP90AA1	4.2	64
INTS4	4.2	65
GTF2I	4.2	66
WDR61	4.2	67
RFC4	4.2	68
RFC3	4.2	69
CDK12	4.2	70
RCOR3	4.2	71
TRRAP	4.2	72
PPP1R10	4.1	73
WDR5	4.1	74
WDR18	4.1	75
RFC5	4.1	76
DDX42	4.1	77
H2AFZ	4.1	78
TET1	4.1	79
DDX21	4.1	80
SLC25A5	4.0	81
WIZ	4.0	82
DDX39A	3.9	83
UCHL5	3.9	84
SOX2	3.9	85
ELAVL1	3.9	86
HNRNPA0	3.9	87
PRRC2C	3.9	88
PRPF8	3.8	89
PHC1	3.8	90

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROTEIN	SCORE	RANK
SRSF5	3.8	91
JMJD1C	3.8	92
PCBP1	3.7	93
CBX1	3.7	94
ACIN1	3.7	95

PROTEIN	SCORE	RANK
RAE1	3.7	96
INTS3	3.7	97
LIN54	3.7	98
PRPF19	3.7	99
HSPD1	3.7	100

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

11. ABBREVIATIONS LIST

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

2C: 2-cell stage embryo
2CLC: 2C-like cell
2iL: serum-free medium with MEK and GSK3 inhibitors, and LIF
5mC: 5-methyl cytosine
αKG: α-ketoglutarate
aa: amino acids
ADP: adenosine diphosphate
AP: alkaline phosphatase
ATP: adenosine triphosphate
BF: bright field
BMP: bone morphogenetic protein
bp: bases pairs
Cas9: CRISPR associated protein 9
CDK: cyclin-dependent kinase
cDNA: complementary DNA
ChIP: chromatin immunoprecipitation
CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
DAPI: 4 ',6-diamidino-2-fenilindol
DMSO: dimethyl sulfoxide
DNMT: DNA methyltransferase
EB: embryoid body
EC: embryonal carcinoma
ECAR: extracellular acidification rate
EGC: embryonic germ cell
EMT: epithelial-to-mesenchymal transition
EPI: epiblast
EpiLCs: epiblast-like cells
EpiSCs: epiblast-derived stem cells
ERK: extracellular signal-regulated kinase
ESC: embryonic stem cell
ET: extra terminal
ETC: electron transport chain
EV: empty vector
FA: serum-free medium containing FGF2 and Activin A
FGF: fibroblasts growth factor
FGFR1: FGF receptor 1

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FPKM: fragments per kilo base per million mapped reads
GLUT1: glucose transporter 1
GLUT3: glucose transporter 3
GFP-: absence of GFP
GFP: green fluorescent protein
GO: gene ontology
GSEA: gene set enrichment analysis
GSK3: glycogen synthase kinase 3
GT: gene trap
HDAC: histone deacetylase
HK: hexokinase
ICM: inner cell mass
IL-6: interleukin 6
iPSC: induced pluripotent stem cell
ISWI: imitation SWI from *D. melanogaster*
Kac: acetylated lysine residue
KAT: lysine acetyltransferase
Kb: kilobases
kDa: kilo Dalton
KDAC: lysine deacetylase
KDM: lysine demethylase
KEGG: Kyoto encyclopedia genes and genomes
LCH: LSD1-CoREST-HDAC1 ternary complex
LIF: leukemia inhibitory factor
MAPK: mitogen-activated protein kinase
MEF: mouse embryonic fibroblast
MEK: MAPK/ERK kinase
MET: mesenchymal-to-epithelial transition
MLL: mixed-lineage leukemia
mRNA: messenger RNA
MS: mass spectrometry
NF-κB: nuclear factor kappa-light-chain-enhancer of activated B cells
NSCs: neural stem cells
NT-ESC: nuclear transfer ESC
OCR: oxygen consumption rate
OXPHOS: oxidative phosphorylation

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PADI: protein arginine deiminase
PcG: polycomb group
PFK1: phosphofructokinase
PHD: plant homeodomain
PPP: pentose phosphate pathway
pre-iPSC: pre-induced pluripotent stem cell
PRC: polycomb repressive complex
PrE: primitive endoderm
PRMT: protein arginine methyltransferase
PTM: post-translational modification
RCOR1: REST corepressor 1
RING: *Really Interesting New Gene*
ROS: reactive oxygen species
SAM: S-adenosylmethionine
SCNT: somatic cell nuclear transfer
SCR: somatic cell reprogramming
SD: standard deviation
SEM: standard error of the mean
TCA: tricarboxylic acid
TE: trophectoderm
TF: transcription factor
TSS: transcription start site
VPA: valproic acid
WT: wild type
YFP: yellow fluorescent protein
ZGA: zygotic genomic activation
ZNF: zinc finger

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

12. FIGURES LIST

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Figure I1. Main characteristics of pluripotent stem cells.

Figure I2. Early stages of embryo development.

Figure I3. Mutual exclusion of *Oct4* and *Cdx2* specifies cell identity during early mouse development.

Figure I4. Mutual exclusion of *Nanog* and *Gata6* specifies cell identity during early mouse development.

Figure I5. Pluripotent cells during *in vivo* early development and their *in vitro* counterparts.

Table I1. Molecular and functional differences between naive and primed pluripotency.

Figure I6. Representation of transcriptionally active, poised, and silent genes.

Figure I7. Autoregulatory feeding loops in the core transcriptional network regulating pluripotency.

Figure I8. Signaling pathways that are involved in the control of pluripotent cell identity.

Figure I9. Enzymatic activities of cytosine methylation and demethylation machineries.

Figure I10. Histone acetylation dynamics modulate chromatin conformation.

Figure I11. Protein deimination is mediated by members of the PADI family

Figure I12. Cell metabolism in pluripotent cells.

Figure I13. Highly proliferating cells present a metabolic profile reminiscent of the Warburg effect.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure I14. Cells with alternative pluripotent states show different bioenergetic requirements.

Figure I15. Different approaches for somatic cell reprogramming.

Figure I16. Directed fusion of somatic cells with PSCs.

Figure I17. Probabilistic models of somatic cell reprogramming.

Figure I18. Permissive phases of the reprogramming process.

Figure I19. Structure of the main zinc finger domains.

Figure I20. Family of zinc finger MYM-type containing proteins.

Figure R1. ZMYM2 is present in the OCT4-NANOG protein interaction network in ESCs.

Figure R2. Strategy for *Zmym2*$^{GT/GT}$ ESC line validation.

Figure R3. Loss-of-function of ZMYM2 is dispensable for ESC maintenance.

Figure R4. Depletion of *Zmym2* in ESCs causes dysregulation of pluripotency and lineage-specific markers and enhances self-renewal.

Figure R5. The absence of ZMYM2 compromises the proper transition from ESCs to EpiLCs.

Figure R6. ZMYM2 is critical for primed pluripotency.

Figure R7. ZMYM2 deficient ESCs show abnormal EB differentiation.

Figure R8. ZMYM2 depleted ESCs are skewed toward neural ectoderm differentiation and defective in mesendoderm differentiation.

Figure R9. *Zmym2* depletion increases iPSC reprogramming efficiency.

Figure R10. *Zmym2* depletion enhances the reprogramming efficiency of pre-iPSCs.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure R11. *Zmym2* depletion in ESCs improves heterokaryon-based fusion reprogramming of human B cells.

Figure R12. The absence of ZMYM2 alters the transcriptional landscape of murine ESCs.

Figure R13. The absence of ZMYM2 leads to the misregulation of genes implicated in signaling pathways.

Figure R14. ZMYM2 depletion leads to activation and silencing of gene expression in ESCs.

Figure R15. ZMYM2 binds to DNA in a sequence-specific manner.

Figure R16. ZMYM2 can function as both an activator and a repressor of gene transcription.

Figure R17. The ZMYM2 interactome in ESCs.

Figure R18. Analysis of genome-wide occupancy between ZMYM2 and LSD1 in ESCs.

Figure R19. ZMYM2 cooperates with LSD1 to act as a transcriptional repressor.

Figure R20. *Zmym2* depletion promotes an increase in glycolytic activity in ESCs.

Figure R21. Alternative strategy using the CRISPR/Cas9 system to validate the role of ZMYM2 in modulating energy metabolism in ESCs.

Figure R22. *Padi2*, *Padi3*, and *Padi4* are regulated at the transcriptional level by ZMYM2.

Figure R23. Analysis of *Padi* genes expression profiles and evolutionary conservation.

Figure R24. Absence of ZMYM2 increases H3Cit levels in ESCs.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure R25. Absence of ZMYM2 promotes chromatin decompaction in ESCs and increases chromatin accessibility

Figure R26. ZMYM2 directly represses the expression of *Padi4*.

Figure R27. The ZMYM2 SIMs are not essential for the transcriptional regulation of *Padi4* in ESCs

Figure R28. ZMYM2-depending binding of LSD1 to the *Padi4* promoter for its transcriptional control.

Figure R29. ZMYM2 may recruit LSD1-HDAC1/2-ZFP281 complex for direct repression of the *Padi4* gene.

Figure R30. PADI4 is involved in part of the transcriptional changes promoted by the lack of ZMYM2.

Figure D1. A model illustrating the transcriptional repression role of ZMYM2 and interacting partners in regulating *Padi4* expression that influences the epigenetic and metabolic features of ESC identity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

13. BIBLIOGRAPHY

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1 Posfai, E., Tam, O. H. & Rossant, J. in *Current Topics in Developmental Biology* Vol. 107 (ed Michael Rendl) 1-37 (Academic Press, 2014).
2 Rossant, J. & Tam, P. P. L. New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* 20, 18-28, doi:10.1016/j.stem.2016.12.004 (2017).
3 Rebuzzini, P., Zuccotti, M. & Garagna, S. Building Pluripotency Identity in the Early Embryo and Derived Stem Cells. *Cells* 10, doi:10.3390/cells10082049 (2021).
4 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-156, doi:10.1038/292154a0 (1981).
5 Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* 78, 7634-7638, doi:10.1073/pnas.78.12.7634 (1981).
6 Young, R. A. Control of the embryonic stem cell state. *Cell* 144, 940-954, doi:10.1016/j.cell.2011.01.032 (2011).
7 Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics* 38, 431-440, doi:10.1038/ng1760 (2006).
8 Ciemerych, M. A. & Sicinski, P. Cell cycle in mouse development. *Oncogene* 24, 2877-2898, doi:10.1038/sj.onc.1208608 (2005).
9 Cockburn, K. & Rossant, J. Making the blastocyst: lessons from the mouse. *J Clin Invest* 120, 995-1003, doi:10.1172/JCI41229 (2010).
10 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* 142, 3151-3165, doi:10.1242/dev.123547 (2015).
11 Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* 3, 662-673, doi:10.1038/nrg887 (2002).
12 Grewal Shiv, I. S. & Moazed, D. Heterochromatin and Epigenetic Control of Gene Expression. *Science* 301, 798-802, doi:10.1126/science.1086887 (2003).
13 Reik, W. *et al.* Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *J Physiol* 547, 35-44, doi:10.1113/jphysiol.2002.033274 (2003).
14 Evsikov, A. V. *et al.* Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes Dev* 20, 2713-2727, doi:10.1101/gad.1471006 (2006).
15 Ko, M. S. H. in *Current Topics in Developmental Biology* Vol. 120 (ed Melvin L. DePamphilis) 103-124 (Academic Press, 2016).
16 Svoboda, P., Franke, V. & Schultz, R. M. in *Current Topics in Developmental Biology* Vol. 113 (ed Howard D. Lipshitz) 305-349 (Academic Press, 2015).
17 Johnson, M. H. & McConnell, J. M. L. Lineage allocation and cell polarity during mouse embryogenesis. *Seminars in Cell & Developmental Biology* 15, 583-597, doi:https://doi.org/10.1016/j.semcdb.2004.04.002 (2004).
18 Chawengsaksophak, K., James, R., Hammond, V. E., Köntgen, F. & Beck, F. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* 386, 84-87, doi:10.1038/386084a0 (1997).
19 Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* 95, 379-391, doi:https://doi.org/10.1016/S0092-8674(00)81769-9 (1998).
20 Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* 132, 2093-2102, doi:10.1242/dev.01801 (2005).
21 Mitsui, K. *et al.* The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* 113, 631-642, doi:10.1016/S0092-8674(03)00393-3 (2003).
22 Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

MAPK Pathway. *Developmental Cell* 10, 615-624, doi:10.1016/j.devcel.2006.02.020 (2006).
23 O'Shea, K. S. Self-renewal vs. Differentiation of Mouse Embryonic Stem Cells1. *Biology of Reproduction* 71, 1755-1765, doi:10.1095/biolreprod.104.028100 (2004).
24 Kahan, B. W. & Ephrussi, B. Developmental Potentialities of Clonal In Vitro Cultures of Mouse Testicular Teratoma2. *JNCI: Journal of the National Cancer Institute* 44, 1015-1036, doi:10.1093/jnci/44.5.1015 (1970).
25 Kleinsmith, L. J. & Pierce, G. B. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Research* 24, 1544 (1964).
26 Finch, B. W. & Ephrussi, B. RETENTION OF MULTIPLE DEVELOPMENTAL POTENTIALITIES BY CELLS OF A MOUSE TESTICULAR TERATOCARCINOMA DURING PROLONGED CULTURE in vitro AND THEIR EXTINCTION UPON HYBRIDIZATION WITH CELLS OF PERMANENT LINES. *Proc Natl Acad Sci U S A* 57, 615-621, doi:10.1073/pnas.57.3.615 (1967).
27 Rosenthal, M. D., Wishnow, R. M. & Sato, G. H. In Vitro Growth and Differentiation of Clonal Populations of Multipotential Mouse Cells Derived From a Transplantable Testicular Teratocarcinoma2. *JNCI: Journal of the National Cancer Institute* 44, 1001-1014, doi:10.1093/jnci/44.5.1001 (1970).
28 Martin Gail, R. Teratocarcinomas and Mammalian Embryogenesis. *Science* 209, 768-776, doi:10.1126/science.6250214 (1980).
29 Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* 309, 255-256, doi:10.1038/309255a0 (1984).
30 Silva, J. *et al*. Nanog Is the Gateway to the Pluripotent Ground State. *Cell* 138, 722-737, doi:10.1016/j.cell.2009.07.039 (2009).
31 Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* 4, 487-492, doi:10.1016/j.stem.2009.05.015 (2009).
32 Prowse, A. B. J., McQuade, L. R., Bryant, K. J., Marcal, H. & Gray, P. P. Identification of Potential Pluripotency Determinants for Human Embryonic Stem Cells Following Proteomic Analysis of Human and Mouse Fibroblast Conditioned Media. *Journal of Proteome Research* 6, 3796-3807, doi:10.1021/pr0702262 (2007).
33 Smith, A. G. *et al*. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* 336, 688-690, doi:10.1038/336688a0 (1988).
34 Williams, R. L. *et al*. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* 336, 684-687, doi:10.1038/336684a0 (1988).
35 Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics* 18, 180-191, doi:10.1038/nrg.2016.156 (2017).
36 Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* 90, 8424-8428, doi:10.1073/pnas.90.18.8424 (1993).
37 Chambers, I. *et al*. Nanog safeguards pluripotency and mediates germline development. *Nature* 450, 1230-1234, doi:10.1038/nature06403 (2007).
38 Hayashi, K., de Sousa Lopes, S. M. C., Tang, F., Lao, K. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell stem cell* 3, 391-401, doi:10.1016/j.stem.2008.07.027 (2008).
39 Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* 135, 909-918, doi:10.1242/dev.017400 (2008).
40 Ying, Q.-L. *et al*. The ground state of embryonic stem cell self-renewal. *Nature* 453, 519-523, doi:10.1038/nature06968 (2008).
41 Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

signaling by a pharmacological GSK-3-specific inhibitor. *Nature Medicine* 10, 55-63, doi:10.1038/nm979 (2004).
42 Wray, J. *et al*. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nature Cell Biology* 13, 838-845, doi:10.1038/ncb2267 (2011).
43 Marks, H. *et al*. The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* 149, 590-604, doi:10.1016/j.cell.2012.03.026 (2012).
44 Brons, I. G. M. *et al*. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* 448, 191-195, doi:10.1038/nature05950 (2007).
45 Tesar, P. J. *et al*. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448, 196-199, doi:10.1038/nature05972 (2007).
46 Kojima, Y. *et al*. The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak. *Cell Stem Cell* 14, 107-120, doi:10.1016/j.stem.2013.09.014 (2014).
47 Huang, Y., Osorno, R., Tsakiridis, A. & Wilson, V. In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation. *Cell Reports* 2, 1571-1578, doi:10.1016/j.celrep.2012.10.022 (2012).
48 Tsakiridis, A. *et al*. Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development (Cambridge, England)* 141, 1209-1221, doi:10.1242/dev.101014 (2014).
49 Wu, J. *et al*. An alternative pluripotent state confers interspecies chimaeric competency. *Nature* 521, 316-321, doi:10.1038/nature14413 (2015).
50 Han, D. W. *et al*. Epiblast Stem Cell Subpopulations Represent Mouse Embryos of Distinct Pregastrulation Stages. *Cell* 143, 617-627, doi:10.1016/j.cell.2010.10.015 (2010).
51 Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. *Cell* 146, 519-532, doi:10.1016/j.cell.2011.06.052 (2011).
52 Guo, G. *et al*. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development (Cambridge, England)* 136, 1063-1069, doi:10.1242/dev.030957 (2009).
53 Hanna, J. *et al*. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell stem cell* 4, 513-524, doi:10.1016/j.stem.2009.04.015 (2009).
54 Okamura, D. *et al*. Stepwise conversion methods between ground states pluripotency from naïve to primed. *Biochemical and Biophysical Research Communications* 574, 70-77, doi:https://doi.org/10.1016/j.bbrc.2021.07.097 (2021).
55 Bao, S. *et al*. Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* 461, 1292-1295, doi:10.1038/nature08534 (2009).
56 Wei, M. *et al*. Establishment of Mouse Primed Stem Cells by Combination of Activin and LIF Signaling. *Front Cell Dev Biol* 9, 713503-713503, doi:10.3389/fcell.2021.713503 (2021).
57 Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development (Cambridge, England)* 144, 365-373, doi:10.1242/dev.142679 (2017).
58 Kinoshita, M. *et al*. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell stem cell* 28, 453-471.e458, doi:10.1016/j.stem.2020.11.005 (2021).
59 Wang, X. *et al*. Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell research* 31, 526-541, doi:10.1038/s41422-021-00477-x (2021).
60 Hoogland, S. H. A. & Marks, H. Developments in pluripotency: a new formative state. *Cell Research* 31, 493-494, doi:10.1038/s41422-021-00494-w (2021).
61 Li, M. & Izpisua Belmonte, J. C. Deconstructing the pluripotency gene regulatory network. *Nature Cell Biology* 20, 382-392, doi:10.1038/s41556-018-0067-6 (2018).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

62 Gökbuget, D. & Blelloch, R. Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development (Cambridge, England)* 146, dev164772, doi:10.1242/dev.164772 (2019).
63 Boyer, L. A., Mathur, D. & Jaenisch, R. Molecular control of pluripotency. *Current Opinion in Genetics & Development* 16, 455-462, doi:https://doi.org/10.1016/j.gde.2006.08.009 (2006).
64 Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* 424, 147-151, doi:10.1038/nature01763 (2003).
65 Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* 10, 252-263, doi:10.1038/nrg2538 (2009).
66 Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186-192, doi:10.1038/nature08449 (2009).
67 Soufi, A., Donahue, G. & Zaret, Kenneth S. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* 151, 994-1004, doi:10.1016/j.cell.2012.09.045 (2012).
68 Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* 161, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).
69 Ng, H.-H. & Surani, M. A. The transcriptional and signalling networks of pluripotency. *Nature Cell Biology* 13, 490-496, doi:10.1038/ncb0511-490 (2011).
70 Chambers, I. *et al.* Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* 113, 643-655, doi:10.1016/S0092-8674(03)00392-1 (2003).
71 Hatano, S.-y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mechanisms of Development* 122, 67-79, doi:https://doi.org/10.1016/j.mod.2004.08.008 (2005).
72 Schöler, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. & Gruss, P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J* 9, 2185-2195, doi:https://doi.org/10.1002/j.1460-2075.1990.tb07388.x (1990).
73 Pesce, M. & Schöler, H. R. Oct-4: Control of totipotency and germline determination. *Molecular Reproduction and Development* 55, 452-457, doi:https://doi.org/10.1002/(SICI)1098-2795(200004)55:4<452::AID-MRD14>3.0.CO;2-S (2000).
74 Niwa, H. Molecular Mechanism to Maintain Stem Cell Renewal of ES Cells. *Cell Structure and Function* 26, 137-148, doi:10.1247/csf.26.137 (2001).
75 Hough, S. R., Clements, I., Welch, P. J. & Wiederholt, K. A. Differentiation of Mouse Embryonic Stem Cells after RNA Interference-Mediated Silencing of OCT4 and Nanog. *STEM CELLS* 24, 1467-1475, doi:https://doi.org/10.1634/stemcells.2005-0475 (2006).
76 Niwa, H., Miyazaki, J.-i. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics* 24, 372-376, doi:10.1038/74199 (2000).
77 Ding, J., Xu, H., Faiola, F., Ma'ayan, A. & Wang, J. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Research* 22, 155-167, doi:10.1038/cr.2011.179 (2012).
78 Wang, J., Levasseur, D. N. & Orkin, S. H. Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* 105, 6326-6331, doi:10.1073/pnas.0802288105 (2008).
79 Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495, 370-374, doi:10.1038/nature11925 (2013).
80 Ding, J. *et al.* Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. *Cell stem cell* 16, 653-668, doi:10.1016/j.stem.2015.04.001 (2015).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

81 Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049-1061, doi:10.1016/j.cell.2008.02.039 (2008).
82 Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* 126, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).
83 Jiang, J. *et al*. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biology* 10, 353-360, doi:10.1038/ncb1698 (2008).
84 Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* 460, 118-122, doi:10.1038/nature08113 (2009).
85 Sharma, R. *et al*. Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. *Nature Communications* 12, 5579, doi:10.1038/s41467-021-25761-7 (2021).
86 Fidalgo, M. *et al*. Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. *Cell stem cell* 19, 355-369, doi:10.1016/j.stem.2016.05.025 (2016).
87 Mayer, D. *et al*. Zfp281 orchestrates interconversion of pluripotent states by engaging Ehmt1 and Zic2. *EMBO J* 39, e102591-e102591, doi:10.15252/embj.2019102591 (2020).
88 Huang, X., Bashkenova, N., Yang, J., Li, D. & Wang, J. ZFP281 recruits polycomb repressive complex 2 to restrict extraembryonic endoderm potential in safeguarding embryonic stem cell pluripotency. *Protein Cell* 12, 213-219, doi:10.1007/s13238-020-00775-x (2021).
89 Yeo, J.-C. *et al*. Klf2 Is an Essential Factor that Sustains Ground State Pluripotency. *Cell Stem Cell* 14, 864-872, doi:10.1016/j.stem.2014.04.015 (2014).
90 Brown, H. A. *et al*. An ERK5–KLF2 signalling module regulates early embryonic gene expression and telomere rejuvenation in stem cells. *Biochemical Journal* 478, 4119-4136, doi:10.1042/BCJ20210646 (2021).
91 Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* 23, 7150-7160, doi:10.1038/sj.onc.1207930 (2004).
92 Hart, A. H., Hartley, L., Ibrahim, M. & Robb, L. Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Developmental Dynamics* 230, 187-198, doi:https://doi.org/10.1002/dvdy.20034 (2004).
93 Schöler, H. R., Ruppert, S., Suzuki, N., Chowdhury, K. & Gruss, P. New type of POU domain in germ line-specific protein Oct-4. *Nature* 344, 435-439, doi:10.1038/344435a0 (1990).
94 Masui, S. *et al*. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology* 9, 625-635, doi:10.1038/ncb1589 (2007).
95 Jeon, H. *et al*. Comprehensive Identification of Krüppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming. *PLoS One* 11, e0150715-e0150715, doi:10.1371/journal.pone.0150715 (2016).
96 Qiu, D. *et al*. Klf2 and Tfcp2l1, Two Wnt/β-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Reports* 5, 314-322, doi:10.1016/j.stemcr.2015.07.014 (2015).
97 Hochedlinger, K. & Jaenisch, R. Induced Pluripotency and Epigenetic Reprogramming. *Cold Spring Harb Perspect Biol* 7, a019448, doi:10.1101/cshperspect.a019448 (2015).
98 Smith, A. G. Embryo-Derived Stem Cells: Of Mice and Men. *Annual Review of Cell and Developmental Biology* 17, 435-462, doi:10.1146/annurev.cellbio.17.1.435 (2001).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

99 Smith, T. A. & Hooper, M. L. Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Experimental Cell Research* 145, 458-462, doi:https://doi.org/10.1016/0014-4827(83)90025-3 (1983).
100 Kunath, T. *et al*. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development* 134, 2895-2902, doi:10.1242/dev.02880 (2007).
101 Yuan, H., Corbi, N., Basilico, C. & Dailey, L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 9, 2635-2645 (1995).
102 Niwa, H., Burdon, T., Chambers, I. & Smith, A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* 12, 2048-2060, doi:10.1101/gad.12.13.2048 (1998).
103 Matsuda, T. *et al*. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J* 18, 4261-4269, doi:10.1093/emboj/18.15.4261 (1999).
104 Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J* 32, 2561-2574, doi:10.1038/emboj.2013.177 (2013).
105 Ye, S., Li, P., Tong, C. & Ying, Q.-L. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J* 32, 2548-2560, doi:10.1038/emboj.2013.175 (2013).
106 Ying, Q.-L., Nichols, J., Chambers, I. & Smith, A. BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3. *Cell* 115, 281-292, doi:https://doi.org/10.1016/S0092-8674(03)00847-X (2003).
107 Chen, X. *et al*. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* 133, 1106-1117, doi:10.1016/j.cell.2008.04.043 (2008).
108 Silva, J. & Smith, A. Capturing pluripotency. *Cell* 132, 532-536, doi:10.1016/j.cell.2008.02.006 (2008).
109 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics* 17, 487-500, doi:10.1038/nrg.2016.59 (2016).
110 Roundtree, I. A. & He, C. RNA epigenetics--chemical messages for posttranscriptional gene regulation. *Curr Opin Chem Biol* 30, 46-51, doi:10.1016/j.cbpa.2015.10.024 (2016).
111 Schlesinger, S. & Meshorer, E. Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Developmental Cell* 48, 135-150, doi:10.1016/j.devcel.2019.01.003 (2019).
112 Aung, P. P., Nagarajan, P. & Prieto, V. G. Regression in primary cutaneous melanoma: etiopathogenesis and clinical significance. *Laboratory Investigation* 97, 657-668, doi:10.1038/labinvest.2017.8 (2017).
113 Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science (New York, N.Y.)* 361, 1336-1340, doi:10.1126/science.aat6806 (2018).
114 Shanak, S. & Helms, V. DNA methylation and the core pluripotency network. *Developmental Biology* 464, 145-160, doi:https://doi.org/10.1016/j.ydbio.2020.06.001 (2020).
115 Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nature Biotechnology* 28, 1079-1088, doi:10.1038/nbt.1684 (2010).
116 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472-479, doi:10.1038/nature12750 (2013).
117 Horii, T. & Hatada, I. Regulation of CpG methylation by Dnmt and Tet in pluripotent stem cells. *J Reprod Dev* 62, 331-335, doi:10.1262/jrd.2016-046 (2016).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

118 Shi, D.-Q., Ali, I., Tang, J. & Yang, W.-C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Front Genet* 8, 100-100, doi:10.3389/fgene.2017.00100 (2017).
119 Yang, J., Bashkenova, N., Zang, R., Huang, X. & Wang, J. The roles of TET family proteins in development and stem cells. *Development (Cambridge, England)* 147, dev183129, doi:10.1242/dev.183129 (2020).
120 Garcia-Outeiral, V., de la Parte, C., Fidalgo, M. & Guallar, D. The Complexity of TET2 Functions in Pluripotency and Development. *Front Cell Dev Biol* 8, 630754-630754, doi:10.3389/fcell.2020.630754 (2021).
121 Parry, A., Rulands, S. & Reik, W. Active turnover of DNA methylation during cell fate decisions. *Nature Reviews Genetics* 22, 59-66, doi:10.1038/s41576-020-00287-8 (2021).
122 von Meyenn, F. *et al*. Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Developmental cell* 39, 104-115, doi:10.1016/j.devcel.2016.09.015 (2016).
123 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, doi:10.1016/j.stem.2013.06.002 (2013).
124 Ross, S. E. & Bogdanovic, O. TET enzymes, DNA demethylation and pluripotency. *Biochemical Society Transactions* 47, 875-885, doi:10.1042/BST20180606 (2019).
125 Fang, Y. & Li, X. Metabolic and epigenetic regulation of endoderm differentiation. *Trends in Cell Biology*, doi:10.1016/j.tcb.2021.09.002.
126 He, R., Dantas, A. & Riabowol, K. Histone Acetyltransferases and Stem Cell Identity. *Cancers (Basel)* 13, 2407, doi:10.3390/cancers13102407 (2021).
127 Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nature Reviews Molecular Cell Biology* 16, 258-264, doi:10.1038/nrm3931 (2015).
128 Zhang, T., Zhang, Z., Dong, Q., Xiong, J. & Zhu, B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol* 21, 45-45, doi:10.1186/s13059-020-01957-w (2020).
129 Pelham-Webb, B. *et al*. H3K27ac bookmarking promotes rapid post-mitotic activation of the pluripotent stem cell program without impacting 3D chromatin reorganization. *Molecular Cell* 81, 1732-1748.e1738, doi:10.1016/j.molcel.2021.02.032 (2021).
130 Hirsch, C. L., Wrana, J. L. & Dent, S. Y. R. KATapulting toward Pluripotency and Cancer. *J Mol Biol* 429, 1958-1977, doi:10.1016/j.jmb.2016.09.023 (2017).
131 Dey, A., Uppal, S., Giri, J. & Misra, H. S. Emerging roles of bromodomain protein 4 in regulation of stem cell identity. *STEM CELLS* 39, 1615-1624, doi:https://doi.org/10.1002/stem.3454 (2021).
132 Jambhekar, A., Dhall, A. & Shi, Y. Roles and regulation of histone methylation in animal development. *Nat Rev Mol Cell Biol* 20, 625-641, doi:10.1038/s41580-019-0151-1 (2019).
133 Kouzarides, T. Histone methylation in transcriptional control. *Current Opinion in Genetics & Development* 12, 198-209, doi:https://doi.org/10.1016/S0959-437X(02)00287-3 (2002).
134 Bannister, A. J., Schneider, R. & Kouzarides, T. Histone Methylation: Dynamic or Static? *Cell* 109, 801-806, doi:https://doi.org/10.1016/S0092-8674(02)00798-5 (2002).
135 Lee, J.-H., Hart, S. R. L. & Skalnik, D. G. Histone deacetylase activity is required for embryonic stem cell differentiation. *genesis* 38, 32-38, doi:https://doi.org/10.1002/gene.10250 (2004).
136 Efroni, S. *et al*. Global transcription in pluripotent embryonic stem cells. *Cell stem cell* 2, 437-447, doi:10.1016/j.stem.2008.03.021 (2008).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

137 Epsztejn-Litman, S. *et al*. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* 15, 1176-1183, doi:10.1038/nsmb.1476 (2008).
138 Nicholson, T. B., Veland, N. & Chen, T. in *Epigenetic Cancer Therapy* (ed Steven G. Gray) 31-66 (Academic Press, 2015).
139 Perillo, B., Tramontano, A., Pezone, A. & Migliaccio, A. LSD1: more than demethylation of histone lysine residues. *Exp Mol Med* 52, 1936-1947, doi:10.1038/s12276-020-00542-2 (2020).
140 Wang, J. *et al*. Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 446, 882-887, doi:10.1038/nature05671 (2007).
141 Foster, C. T. *et al*. Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. *Mol Cell Biol* 30, 4851-4863, doi:10.1128/MCB.00521-10 (2010).
142 Whyte, W. A. *et al*. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221-225, doi:10.1038/nature10805 (2012).
143 Wang, J. *et al*. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genetics* 41, 125-129, doi:10.1038/ng.268 (2009).
144 Separovich, R. J., Pang, C. N. I. & Wilkins, M. R. Controlling the Controllers: Regulation of Histone Methylation by Phosphosignalling. *Trends in Biochemical Sciences* 45, 1035-1048, doi:10.1016/j.tibs.2020.08.004 (2020).
145 Bicker, K. L. & Thompson, P. R. The protein arginine deiminases: Structure, function, inhibition, and disease. *Biopolymers* 99, 155-163, doi:10.1002/bip.22127 (2013).
146 De Keyser, J., Schaaf, M. & Teelken, A. Peptidylarginine deiminase activity in postmortem white matter of patients with multiple sclerosis. *Neuroscience Letters* 260, 74-76, doi:https://doi.org/10.1016/S0304-3940(98)00947-1 (1999).
147 Suzuki, A. *et al*. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nature Genetics* 34, 395-402, doi:10.1038/ng1206 (2003).
148 Zhang, Y. *et al*. PADs in cancer: Current and future. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1875, 188492, doi:https://doi.org/10.1016/j.bbcan.2020.188492 (2021).
149 Nakashima, K., Hagiwara, T. & Yamada, M. Nuclear Localization of Peptidylarginine Deiminase V and Histone Deimination in Granulocytes *. *Journal of Biological Chemistry* 277, 49562-49568, doi:10.1074/jbc.M208795200 (2002).
150 Mondal, S. & Thompson, P. R. Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination. *Acc Chem Res* 52, 818-832, doi:10.1021/acs.accounts.9b00024 (2019).
151 Cuthbert, G. L. *et al*. Histone Deimination Antagonizes Arginine Methylation. *Cell* 118, 545-553, doi:10.1016/j.cell.2004.08.020 (2004).
152 Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim Biophys Acta* 1829, 1126-1135, doi:10.1016/j.bbagrm.2013.07.003 (2013).
153 Christophorou, M. A. *et al*. Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 507, 104-108, doi:10.1038/nature12942 (2014).
154 Shyh-Chang, N., Daley, G. Q. & Cantley, L. C. Stem cell metabolism in tissue development and aging. *Development (Cambridge, England)* 140, 2535-2547, doi:10.1242/dev.091777 (2013).
155 Barbehenn Ek Fau - Wales, R. G., Wales Rg Fau - Lowry, O. H. & Lowry, O. H. Measurement of metabolites in single preimplantation embryos; a new means to study metabolic control in early embryos.
156 Brown, J. J. & Whittingham, D. G. The roles of pyruvate, lactate and glucose during preimplantation development of embryos from F1 hybrid mice in vitro. *Development* 112, 99-105, doi:10.1242/dev.112.1.99 (1991).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

157 Zhang, J. *et al*. Metabolism in Pluripotent Stem Cells and Early Mammalian Development. *Cell Metabolism* 27, 332-338, doi:10.1016/j.cmet.2018.01.008 (2018).
158 Seshagiri, P. B. & Bavister, B. D. Glucose Inhibits Development of Hamster 8-Cell Embryos in Vitro1. *Biology of Reproduction* 40, 599-606, doi:10.1095/biolreprod40.3.599 (1989).
159 Leese Hj Fau - Barton, A. M. & Barton, A. M. Pyruvate and glucose uptake by mouse ova and preimplantation embryos.
160 Yan, L. *et al*. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 20, 1131-1139, doi:10.1038/nsmb.2660 (2013).
161 Xue, Z. *et al*. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593-597, doi:10.1038/nature12364 (2013).
162 Pantaleon, M. & Kaye, P. L. Glucose transporters in preimplantation development.
163 Van Blerkom, J. Mitochondria in early mammalian development. *Seminars in Cell & Developmental Biology* 20, 354-364, doi:https://doi.org/10.1016/j.semcdb.2008.12.005 (2009).
164 Kondoh, H. *et al*. A High Glycolytic Flux Supports the Proliferative Potential of Murine Embryonic Stem Cells. *Antioxidants & Redox Signaling* 9, 293-299, doi:10.1089/ars.2006.1467 (2007).
165 Varum, S. *et al*. Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* 6, e20914-e20914, doi:10.1371/journal.pone.0020914 (2011).
166 Filosa, S. *et al*. Failure to increase glucose consumption through the pentose-phosphate pathway results in the death of glucose-6-phosphate dehydrogenase gene-deleted mouse embryonic stem cells subjected to oxidative stress. *Biochem J* 370, 935-943, doi:10.1042/BJ20021614 (2003).
167 Zhang, J. *et al*. UCP2 regulates energy metabolism and differentiation potential of human pluripotent stem cells. *EMBO J* 30, 4860-4873, doi:10.1038/emboj.2011.401 (2011).
168 Zhou, W. *et al*. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J* 31, 2103-2116, doi:10.1038/emboj.2012.71 (2012).
169 Scognamiglio, R. *et al*. Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause. *Cell* 164, 668-680, doi:10.1016/j.cell.2015.12.033 (2016).
170 Carbognin, E., Betto, R. M., Soriano, M. E., Smith, A. G. & Martello, G. Stat3 promotes mitochondrial transcription and oxidative respiration during maintenance and induction of naive pluripotency. *EMBO J* 35, 618-634, doi:10.15252/embj.201592629 (2016).
171 Folmes, C. D. L. *et al*. Nuclear reprogramming with c-Myc potentiates glycolytic capacity of derived induced pluripotent stem cells. *J Cardiovasc Transl Res* 6, 10-21, doi:10.1007/s12265-012-9431-2 (2013).
172 Cao, Y. *et al*. miR-290/371-Mbd2-Myc circuit regulates glycolytic metabolism to promote pluripotency. *EMBO J* 34, 609-623, doi:10.15252/embj.201490441 (2015).
173 Sone, M. *et al*. Hybrid Cellular Metabolism Coordinated by Zic3 and Esrrb Synergistically Enhances Induction of Naive Pluripotency. *Cell Metabolism* 25, 1103-1117.e1106, doi:10.1016/j.cmet.2017.04.017 (2017).
174 Carey, B. W., Finley, L. W. S., Cross, J. R., Allis, C. D. & Thompson, C. B. Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* 518, 413-416, doi:10.1038/nature13981 (2015).
175 Zhang, J. *et al*. LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* 19, 66-80, doi:10.1016/j.stem.2016.05.009 (2016).
176 Shyh-Chang, N. *et al*. Influence of threonine metabolism on S-adenosylmethionine and histone methylation. *Science (New York, N.Y.)* 339, 222-226, doi:10.1126/science.1226603 (2013).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

177 Moussaieff, A. *et al*. Glycolysis-Mediated Changes in Acetyl-CoA and Histone Acetylation Control the Early Differentiation of Embryonic Stem Cells. *Cell Metabolism* 21, 392-402, doi:10.1016/j.cmet.2015.02.002 (2015).
178 Gurdon, J. B. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.
179 Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Current Biology* 11, 1553-1558, doi:https://doi.org/10.1016/S0960-9822(01)00459-6 (2001).
180 Cowan Chad, A., Atienza, J., Melton Douglas, A. & Eggan, K. Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science* 309, 1369-1373, doi:10.1126/science.1116447 (2005).
181 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987-1000, doi:https://doi.org/10.1016/0092-8674(87)90585-X (1987).
182 Pralong, D., Trounson Ao Fau - Verma, P. J. & Verma, P. J. Cell fusion for reprogramming pluripotency: toward elimination of the pluripotent genome.
183 Peterson, J. A. & Weiss, M. C. Expression of differentiated functions in hepatoma cell hybrids: induction of mouse albumin production in rat hepatoma-mouse fibroblast hybrids. *Proc Natl Acad Sci U S A* 69, 571-575, doi:10.1073/pnas.69.3.571 (1972).
184 Klein G Fau - Bregula, U., Bregula U Fau - Wiener, F., Wiener F Fau - Harris, H. & Harris, H. The analysis of malignancy by cell fusion. I. Hybrids between tumour cells and L cell derivatives.
185 Harris, H. Cell fusion and the analysis of malignancy.
186 Hajkova, P. *et al*. Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development* 117, 15-23, doi:https://doi.org/10.1016/S0925-4773(02)00181-8 (2002).
187 Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J* 16, 6510-6520, doi:10.1093/emboj/16.21.6510 (1997).
188 Pereira, C. F. *et al*. Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2. *PLoS Genet* 4, e1000170-e1000170, doi:10.1371/journal.pgen.1000170 (2008).
189 Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* 441, 997-1001, doi:10.1038/nature04914 (2006).
190 Wong, C. C., Gaspar-Maia, A., Ramalho-Santos, M. & Reijo Pera, R. A. High-efficiency stem cell fusion-mediated assay reveals Sall4 as an enhancer of reprogramming. *PLoS One* 3, e1955-e1955, doi:10.1371/journal.pone.0001955 (2008).
191 Kimura, H., Tada, M., Nakatsuji, N. & Tada, T. Histone code modifications on pluripotential nuclei of reprogrammed somatic cells. *Mol Cell Biol* 24, 5710-5720, doi:10.1128/MCB.24.13.5710-5720.2004 (2004).
192 Matveeva, N. M. *et al*. In vitro and in vivo study of pluripotency in intraspecific hybrid cells obtained by fusion of murine embryonic stem cells with splenocytes. *Molecular Reproduction and Development* 50, 128-138, doi:https://doi.org/10.1002/(SICI)1098-2795(199806)50:2<128::AID-MRD2>3.0.CO;2-M (1998).
193 Ying, Q.-L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* 416, 545-548, doi:10.1038/nature729 (2002).
194 Pells, S., Di Domenico, A. I., Gallagher, E. J. & McWhir, J. Multipotentiality of Neuronal Cells after Spontaneous Fusion with Embryonic Stem Cells and Nuclear Reprogramming In Vitro. *Cloning and Stem Cells* 4, 331-338, doi:10.1089/153623002321025005 (2002).
195 Do, J. T. & Schöler, H. R. Nuclei of Embryonic Stem Cells Reprogram Somatic Cells. *STEM CELLS* 22, 941-949, doi:https://doi.org/10.1634/stemcells.22-6-941 (2004).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

196 Takahashi, K. *et al*. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131, 861-872, doi:10.1016/j.cell.2007.11.019 (2007).
197 Yu, J. & Thomson, J. A. Pluripotent stem cell lines. *Genes Dev* 22, 1987-1997, doi:10.1101/gad.1689808 (2008).
198 Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132, 567-582, doi:10.1016/j.cell.2008.01.015 (2008).
199 Lengerke, C. & Daley, G. Q. Disease Models from Pluripotent Stem Cells. *Annals of the New York Academy of Sciences* 1176, 191-196, doi:https://doi.org/10.1111/j.1749-6632.2009.04962.x (2009).
200 Ho, R., Chronis, C. & Plath, K. Mechanistic insights into reprogramming to induced pluripotency. *J Cell Physiol* 226, 868-878, doi:10.1002/jcp.22450 (2011).
201 Yu, J. *et al*. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* 318, 1917-1920, doi:10.1126/science.1151526 (2007).
202 Buganim, Y. *et al*. The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell stem cell* 15, 295-309, doi:10.1016/j.stem.2014.07.003 (2014).
203 Maherali, N. *et al*. Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* 1, 55-70, doi:10.1016/j.stem.2007.05.014 (2007).
204 Wernig, M. *et al*. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 448, 318-324, doi:10.1038/nature05944 (2007).
205 Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* 448, 313-317, doi:10.1038/nature05934 (2007).
206 Blelloch, R., Venere, M., Yen, J. & Ramalho-Santos, M. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell stem cell* 1, 245-247, doi:10.1016/j.stem.2007.08.008 (2007).
207 Apostolou, E. & Stadtfeld, M. Cellular trajectories and molecular mechanisms of iPSC reprogramming. *Current opinion in genetics & development* 52, 77-85, doi:10.1016/j.gde.2018.06.002 (2018).
208 Tran, K. A. *et al*. Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency. *Cell reports* 27, 1726-1741.e1725, doi:10.1016/j.celrep.2019.04.056 (2019).
209 Schwarz, B. A. *et al*. Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming. *Cell stem cell* 23, 289-305.e285, doi:10.1016/j.stem.2018.06.013 (2018).
210 O'Malley, J. *et al*. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* 499, 88-91, doi:10.1038/nature12243 (2013).
211 Schiebinger, G. *et al*. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* 176, 1517-1517, doi:10.1016/j.cell.2019.02.026 (2019).
212 Guo, L. *et al*. Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq. *Molecular Cell* 73, 815-829.e817, doi:10.1016/j.molcel.2019.01.042 (2019).
213 Haridhasapavalan, K. K., Raina, K., Dey, C., Adhikari, P. & Thummer, R. P. An Insight into Reprogramming Barriers to iPSC Generation. *Stem Cell Reviews and Reports* 16, 56-81, doi:10.1007/s12015-019-09931-1 (2020).
214 Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell stem cell* 16, 323-337, doi:10.1016/j.stem.2015.01.015 (2015).
215 Buganim, Y. *et al*. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209-1222, doi:10.1016/j.cell.2012.08.023 (2012).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

216 Unternaehrer, J. J. *et al.* The epithelial-mesenchymal transition factor SNAIL paradoxically enhances reprogramming. *Stem Cell Reports* 3, 691-698, doi:10.1016/j.stemcr.2014.09.008 (2014).
217 Gingold, J. A. *et al.* A genome-wide RNAi screen identifies opposing functions of Snai1 and Snai2 on the Nanog dependency in reprogramming. *Molecular cell* 56, 140-152, doi:10.1016/j.molcel.2014.08.014 (2014).
218 Hawkins, K. E. *et al.* NRF2 Orchestrates the Metabolic Shift during Induced Pluripotent Stem Cell Reprogramming. *Cell reports* 14, 1883-1891, doi:10.1016/j.celrep.2016.02.003 (2016).
219 Kida, Y. S. *et al.* ERRs Mediate a Metabolic Switch Required for Somatic Cell Reprogramming to Pluripotency. *Cell stem cell* 16, 547-555, doi:10.1016/j.stem.2015.03.001 (2015).
220 Zhou, G., Meng, S., Li, Y., Ghebre, Y. T. & Cooke, J. P. Optimal ROS Signaling Is Critical for Nuclear Reprogramming. *Cell reports* 15, 919-925, doi:10.1016/j.celrep.2016.03.084 (2016).
221 Mathieu, J. & Ruohola-Baker, H. Metabolic remodeling during the loss and acquisition of pluripotency. *Development (Cambridge, England)* 144, 541-551, doi:10.1242/dev.128389 (2017).
222 Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT–MET mechanism for optimal reprogramming. *Nature Cell Biology* 15, 829-838, doi:10.1038/ncb2765 (2013).
223 David, L. & Polo, J. M. Phases of reprogramming. *Stem Cell Research* 12, 754-761, doi:https://doi.org/10.1016/j.scr.2014.03.007 (2014).
224 Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151, 1617-1632, doi:10.1016/j.cell.2012.11.039 (2012).
225 Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell stem cell* 2, 230-240, doi:10.1016/j.stem.2008.02.001 (2008).
226 Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* 21, 819-833.e816, doi:10.1016/j.stem.2017.10.012 (2017).
227 Knaupp, A. S. *et al.* Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell* 21, 834-845.e836, doi:10.1016/j.stem.2017.11.007 (2017).
228 Chen, J. *et al.* Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Reports* 14, 1540-1554, doi:10.1016/j.celrep.2016.01.013 (2016).
229 Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* 168, 442-459.e420, doi:10.1016/j.cell.2016.12.016 (2017).
230 Samavarchi-Tehrani, P. *et al.* Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. *Cell Stem Cell* 7, 64-77, doi:10.1016/j.stem.2010.04.015 (2010).
231 Takahashi, K. *et al.* Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. *Nature Communications* 5, 3678, doi:10.1038/ncomms4678 (2014).
232 Li, R. *et al.* A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. *Cell Stem Cell* 7, 51-63, doi:10.1016/j.stem.2010.04.014 (2010).
233 Lai, X. *et al.* Epithelial-Mesenchymal Transition and Metabolic Switching in Cancer: Lessons From Somatic Cell Reprogramming. *Front Cell Dev Biol* 8, 760-760, doi:10.3389/fcell.2020.00760 (2020).
234 Marson, A. *et al.* Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell stem cell* 3, 132-135, doi:10.1016/j.stem.2008.06.019 (2008).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

235 Saunders, A. *et al.* The SIN3A/HDAC Corepressor Complex Functionally Cooperates with NANOG to Promote Pluripotency. *Cell reports* 18, 1713-1726, doi:10.1016/j.celrep.2017.01.055 (2017).
236 Folmes, C. D. L. *et al.* Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell metabolism* 14, 264-271, doi:10.1016/j.cmet.2011.06.011 (2011).
237 Panopoulos, A. D. *et al.* The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell research* 22, 168-177, doi:10.1038/cr.2011.177 (2012).
238 Zhu, S. *et al.* Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell stem cell* 7, 651-655, doi:10.1016/j.stem.2010.11.015 (2010).
239 Prigione, A. *et al.* HIF1α modulates cell fate reprogramming through early glycolytic shift and upregulation of PDK1-3 and PKM2. *Stem cells (Dayton, Ohio)* 32, 364-376, doi:10.1002/stem.1552 (2014).
240 Chen, J. *et al.* Vitamin C modulates TET1 function during somatic cell reprogramming. *Nature Genetics* 45, 1504-1509, doi:10.1038/ng.2807 (2013).
241 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* 20, 207-220, doi:10.1038/s41576-018-0089-8 (2019).
242 Singhal, N. *et al.* Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming. *Cell* 141, 943-955, doi:10.1016/j.cell.2010.04.037 (2010).
243 Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genetics* 45, 34-42, doi:10.1038/ng.2491 (2013).
244 Sridharan, R. *et al.* Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. *Nature cell biology* 15, 872-882, doi:10.1038/ncb2768 (2013).
245 Klug, A. & Schwabe, J. W. R. Zinc fingers. *The FASEB Journal* 9, 597-604, doi:https://doi.org/10.1096/fasebj.9.8.7768350 (1995).
246 Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell Death Discov* 3, 17071-17071, doi:10.1038/cddiscovery.2017.71 (2017).
247 Krishna, S. S., Majumdar, I. & Grishin, N. V. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* 31, 532-550, doi:10.1093/nar/gkg161 (2003).
248 Wang, J. *et al.* Silencing the epigenetic silencer KDM4A for TRAIL and DR5 simultaneous induction and antitumor therapy. *Cell Death Differ* 23, 1886-1896, doi:10.1038/cdd.2016.92 (2016).
249 Huang, F., Abmayr, S. M. & Workman, J. L. Regulation of KAT6 Acetyltransferases and Their Roles in Cell Cycle Progression, Stem Cell Maintenance, and Human Disease. *Mol Cell Biol* 36, 1900-1907, doi:10.1128/MCB.00055-16 (2016).
250 Nunez, N. *et al.* The multi-zinc finger protein ZNF217 contacts DNA through a two-finger domain. *J Biol Chem* 286, 38190-38201, doi:10.1074/jbc.M111.301234 (2011).
251 Borgel, J. *et al.* KDM2A integrates DNA and histone modification signals through a CXXC/PHD module and direct interaction with HP1. *Nucleic Acids Res* 45, 1114-1129, doi:10.1093/nar/gkw979 (2017).
252 Dubrovskyi, O. *et al.* Identification of paxillin domains interacting with ß-catenin. *FEBS Lett* 586, 2294-2299, doi:10.1016/j.febslet.2012.06.016 (2012).
253 Smith, M. A. *et al.* LIM domains target actin regulators paxillin and zyxin to sites of stress fiber strain. *PLoS One* 8, e69378-e69378, doi:10.1371/journal.pone.0069378 (2013).
254 Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases.
255 Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics* 11, 636-646, doi:10.1038/nrg2842 (2010).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

256 De Los Angeles, A. *et al*. Hallmarks of pluripotency. *Nature* 525, 469-478, doi:10.1038/nature15515 (2015).
257 Liu, L., Michowski, W., Kolodziejczyk, A. & Sicinski, P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nature cell biology* 21, 1060-1067, doi:10.1038/s41556-019-0384-4 (2019).
258 Nagandla, H. *et al*. Defective myogenesis in the absence of the muscle-specific lysine methyltransferase SMYD1. *Developmental Biology* 410, 86-97, doi:https://doi.org/10.1016/j.ydbio.2015.12.005 (2016).
259 Li, G. *et al*. CXXC5 regulates differentiation of C2C12 myoblasts into myocytes. *Journal of Muscle Research and Cell Motility* 35, 259-265, doi:10.1007/s10974-014-9400-2 (2014).
260 Li, Q. *et al*. The LIM protein Ajuba promotes adipogenesis by enhancing PPARγ and p300/CBP interaction. *Cell Death Differ* 23, 158-168, doi:10.1038/cdd.2015.83 (2016).
261 Li, K. *et al*. A novel zinc finger protein Zfp637 behaves as a repressive regulator in myogenic cellular differentiation. *Journal of Cellular Biochemistry* 110, 352-362, doi:https://doi.org/10.1002/jcb.22546 (2010).
262 Prenzler, F., Fragasso, A., Schmitt, A. & Munz, B. Functional analysis of ZFP36 proteins in keratinocytes. *European Journal of Cell Biology* 95, 277-284, doi:https://doi.org/10.1016/j.ejcb.2016.04.007 (2016).
263 Bosse, T. *et al*. Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. *Mol Cell Biol* 26, 9060-9070, doi:10.1128/MCB.00124-06 (2006).
264 Beuling, E. *et al*. GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. *Gastroenterology* 140, 1219-1229.e12292, doi:10.1053/j.gastro.2011.01.033 (2011).
265 Tong, Q., Tsai, J., Tan, G., Dalgin, G. & Hotamisligil, G. S. Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. *Mol Cell Biol* 25, 706-715, doi:10.1128/MCB.25.2.706-715.2005 (2005).
266 Boxer, L. D., Barajas, B., Tao, S., Zhang, J. & Khavari, P. A. ZNF750 interacts with KLF4 and RCOR1, KDM1A, and CTBP1/2 chromatin regulators to repress epidermal progenitor genes and induce differentiation genes. *Genes Dev* 28, 2013-2026, doi:10.1101/gad.246579.114 (2014).
267 Shields, J. M., Christy, R. J. & Yang, V. W. Identification and characterization of a gene encoding a gut-enriched Krüppel-like factor expressed during growth arrest. *J Biol Chem* 271, 20009-20017, doi:10.1074/jbc.271.33.20009 (1996).
268 McConnell, B. B., Ghaleb, A. M., Nandan, M. O. & Yang, V. W. The diverse functions of Krüppel-like factors 4 and 5 in epithelial biology and pathobiology. *Bioessays* 29, 549-557, doi:10.1002/bies.20581 (2007).
269 Zhang, W. *et al*. Novel cross talk of Kruppel-like factor 4 and beta-catenin regulates normal intestinal homeostasis and tumor repression. *Mol Cell Biol* 26, 2055-2064, doi:10.1128/MCB.26.6.2055-2064.2006 (2006).
270 Yu, T. *et al*. KLF4 deletion alters gastric cell lineage and induces MUC2 expression. *Cell Death Dis* 7, e2255-e2255, doi:10.1038/cddis.2016.158 (2016).
271 Evans, P. M. & Liu, C. Roles of Krüpel-like factor 4 in normal homeostasis, cancer and stem cells. *Acta Biochim Biophys Sin (Shanghai)* 40, 554-564, doi:10.1111/j.1745-7270.2008.00439.x (2008).
272 Patel, S., Xi, Z. F., Seo, E. Y., McGaughey, D. & Segre, J. A. Klf4 and corticosteroids activate an overlapping set of transcriptional targets to accelerate in utero epidermal barrier acquisition. *Proc Natl Acad Sci U S A* 103, 18668-18673, doi:10.1073/pnas.0608658103 (2006).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

273 Katz, J. P. *et al*. The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development (Cambridge, England)* 129, 2619-2628 (2002).
274 Hayashi, S., Manabe, I., Suzuki, Y., Relaix, F. & Oishi, Y. Klf5 regulates muscle differentiation by directly targeting muscle-specific genes in cooperation with MyoD in mice. *Elife* 5, e17462, doi:10.7554/eLife.17462 (2016).
275 Segre, J. A., Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nature Genetics* 22, 356-360, doi:10.1038/11926 (1999).
276 Li, Y. *et al*. Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood* 105, 635-637, doi:https://doi.org/10.1182/blood-2004-07-2681 (2005).
277 Nakatake, Y. *et al*. Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol* 26, 7772-7782, doi:10.1128/MCB.00468-06 (2006).
278 Serrano, F. *et al*. Gata4 Blocks Somatic Cell Reprogramming By Directly Repressing Nanog. *STEM CELLS* 31, 71-82, doi:https://doi.org/10.1002/stem.1272 (2013).
279 Fujikura, J. *et al*. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev* 16, 784-789, doi:10.1101/gad.968802 (2002).
280 Montserrat, N. *et al*. Reprogramming of Human Fibroblasts to Pluripotency with Lineage Specifiers. *Cell Stem Cell* 13, 341-350, doi:10.1016/j.stem.2013.06.019 (2013).
281 Shu, J. *et al*. GATA family members as inducers for cellular reprogramming to pluripotency. *Cell research* 25, 169-180, doi:10.1038/cr.2015.6 (2015).
282 Cevallos, R. R., Edwards, Y. J. K., Parant, J. M., Yoder, B. K. & Hu, K. Human transcription factors responsive to initial reprogramming predominantly undergo legitimate reprogramming during fibroblast conversion to iPSCs. *Sci Rep* 10, 19710-19710, doi:10.1038/s41598-020-76705-y (2020).
283 Akahori, H., Guindon, S., Yoshizaki, S. & Muto, Y. Molecular Evolution of the TET Gene Family in Mammals. *Int J Mol Sci* 16, 28472-28485, doi:10.3390/ijms161226110 (2015).
284 Melamed, P., Yosefzon, Y., David, C., Tsukerman, A. & Pnueli, L. Tet Enzymes, Variants, and Differential Effects on Function. *Front Cell Dev Biol* 6, 22-22, doi:10.3389/fcell.2018.00022 (2018).
285 Nakaki, F. & Saitou, M. PRDM14: a unique regulator for pluripotency and epigenetic reprogramming. *Trends in Biochemical Sciences* 39, 289-298, doi:10.1016/j.tibs.2014.04.003 (2014).
286 Wang, Z.-X. *et al*. The Transcription Factor Zfp281 Controls Embryonic Stem Cell Pluripotency by Direct Activation and Repression of Target Genes. *STEM CELLS* 26, 2791-2799, doi:https://doi.org/10.1634/stemcells.2008-0443 (2008).
287 Wang, J. *et al*. A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444, 364-368, doi:10.1038/nature05284 (2006).
288 Koch, H. *et al*. Large-Scale Identification of c-MYC-Associated Proteins Using a Combined TAP/MudPIT Approach. *Cell Cycle* 6, 205-217, doi:10.4161/cc.6.2.3742 (2007).
289 Fidalgo, M. *et al*. Zfp281 functions as a transcriptional repressor for pluripotency of mouse embryonic stem cells. *Stem cells (Dayton, Ohio)* 29, 1705-1716, doi:10.1002/stem.736 (2011).
290 Wang, J. *et al*. YY1 Positively Regulates Transcription by Targeting Promoters and Super-Enhancers through the BAF Complex in Embryonic Stem Cells. *Stem Cell Reports* 10, 1324-1339, doi:10.1016/j.stemcr.2018.02.004 (2018).
291 Aguilo, F. *et al*. Coordination of m(6)A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell stem cell* 17, 689-704, doi:10.1016/j.stem.2015.09.005 (2015).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

292 Wang, Z.-X. *et al*. Zfp206 Is a Transcription Factor That Controls Pluripotency of Embryonic Stem Cells. *STEM CELLS* 25, 2173-2182, doi:https://doi.org/10.1634/stemcells.2007-0085 (2007).
293 Zhang, W. *et al*. Zfp206 regulates ES cell gene expression and differentiation. *Nucleic Acids Res* 34, 4780-4790, doi:10.1093/nar/gkl631 (2006).
294 Zhang, J.-Z. *et al*. Screening for Genes Essential for Mouse Embryonic Stem Cell Self-Renewal Using a Subtractive RNA Interference Library. *STEM CELLS* 24, 2661-2668, doi:https://doi.org/10.1634/stemcells.2006-0017 (2006).
295 Ito, S. *et al*. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129-1133, doi:10.1038/nature09303 (2010).
296 Koh, K. P. *et al*. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* 8, 200-213, doi:10.1016/j.stem.2011.01.008 (2011).
297 Ficz, G. *et al*. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398-402, doi:10.1038/nature10008 (2011).
298 Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* 28, 2103-2119, doi:10.1101/gad.248005.114 (2014).
299 Dawlaty, M. M. *et al*. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell stem cell* 9, 166-175, doi:10.1016/j.stem.2011.07.010 (2011).
300 Freudenberg, J. M. *et al*. Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res* 40, 3364-3377, doi:10.1093/nar/gkr1253 (2012).
301 Li, X. *et al*. Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proc Natl Acad Sci U S A* 113, E8267-E8276, doi:10.1073/pnas.1617802113 (2016).
302 Azami, T. *et al*. Klf5 suppresses ERK signaling in mouse pluripotent stem cells. *PLoS One* 13, e0207321-e0207321, doi:10.1371/journal.pone.0207321 (2018).
303 Wismar, J. *et al*. The Mutation without childrenrgl Causes Ecdysteroid Deficiency in Third-Instar Larvae of Drosophila melanogaster. *Developmental Biology* 226, 1-17, doi:https://doi.org/10.1006/dbio.2000.9811 (2000).
304 Maimon, I., Popliker, M. & Gilboa, L. Without children is required for Stat-mediated zfh1 transcription and for germline stem cell differentiation. *Development* 141, 2602-2610, doi:10.1242/dev.109611 (2014).
305 Jin, X., Sun, X. & Song, Q. woc gene mutation causes 20E-dependent α-tubulin detyrosination in Drosophila melanogaster. *Archives of Insect Biochemistry and Physiology* 60, 116-129, doi:https://doi.org/10.1002/arch.20088 (2005).
306 Shapson-Coe, A., Valeiras, B., Wall, C. & Rada, C. Aicardi-Goutières Syndrome associated mutations of RNase H2B impair its interaction with ZMYM3 and the CoREST histone-modifying complex. *PLoS One* 14, e0213553-e0213553, doi:10.1371/journal.pone.0213553 (2019).
307 Smedley, D., Hamoudi, R., Lu, Y.-J., Cooper, C. & Shipley, J. Cloning and Mapping of Members of the MYM Family. *Genomics* 60, 244-247, doi:https://doi.org/10.1006/geno.1999.5918 (1999).
308 Philips Ak Fau - Sirén, A. *et al*. X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes.
309 Xiao, S. *et al*. FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13) leukaemia/lymphoma syndrome.
310 Aubry, S. *et al*. Assembly and interrogation of Alzheimer's disease genetic networks reveal novel regulators of progression.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

311 Popovici, C. *et al*. Fibroblast growth factor receptor 1 is fused to FIM in stem-cell myeloproliferative disorder with t(8;13).
312 Smedley, D. *et al*. The t(8;13)(p11;q11-12) rearrangement associated with an atypical myeloproliferative disorder fuses the fibroblast growth factor receptor 1 gene to a novel gene RAMP.
313 Renneville, A. *et al*. Avadomide induces degradation of ZMYM2 fusion oncoproteins in hematologic malignancies.
314 Keikha, R., Hashemi-Shahri, S. M. & Jebali, A. The relative expression of miR-31, miR-29, miR-126, and miR-17 and their mRNA targets in the serum of COVID-19 patients with different grades during hospitalization.
315 Hansen, T. B. *et al*. Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384-388, doi:10.1038/nature11993 (2013).
316 Zhao, Z.-J. & Shen, J. Circular RNA participates in the carcinogenesis and the malignant behavior of cancer. *RNA Biol* 14, 514-521, doi:10.1080/15476286.2015.1122162 (2017).
317 An, Y. *et al*. circZMYM2 Competed Endogenously with miR-335-5p to Regulate JMJD2C in Pancreatic Cancer. *Cellular Physiology and Biochemistry* 51, 2224-2236, doi:10.1159/000495868 (2018).
318 Hakimi, M.-A., Dong, Y., Lane, W. S., Speicher, D. W. & Shiekhattar, R. A Candidate X-linked Mental Retardation Gene Is a Component of a New Family of Histone Deacetylase-containing Complexes *. *Journal of Biological Chemistry* 278, 7234-7239, doi:10.1074/jbc.M208992200 (2003).
319 Gocke, C. B. & Yu, H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One* 3, e3255-e3255, doi:10.1371/journal.pone.0003255 (2008).
320 Wang, Y.-C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Research* 24, 143-160, doi:10.1038/cr.2013.151 (2014).
321 Ramazi, S. & Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database* 2021, baab012, doi:10.1093/database/baab012 (2021).
322 Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology* 8, 947-956, doi:10.1038/nrm2293 (2007).
323 Wilkinson, Kevin A. & Henley, Jeremy M. Mechanisms, regulation and consequences of protein SUMOylation. *Biochemical Journal* 428, 133-145, doi:10.1042/BJ20100158 (2010).
324 Kunapuli, P., Kasyapa, C. S., Chin, S.-F., Caldas, C. & Cowell, J. K. ZNF198, a zinc finger protein rearranged in myeloproliferative disease, localizes to the PML nuclear bodies and interacts with SUMO-1 and PML. *Experimental Cell Research* 312, 3739-3751, doi:https://doi.org/10.1016/j.yexcr.2006.06.037 (2006).
325 Guzzo, C. M. *et al*. Characterization of the SUMO-Binding Activity of the Myeloproliferative and Mental Retardation (MYM)-Type Zinc Fingers in ZNF261 and ZNF198. *PLoS One* 9, e105271, doi:10.1371/journal.pone.0105271 (2014).
326 Toh, C.-Xu D. *et al*. RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. *Cell Reports* 15, 2597-2607, doi:10.1016/j.celrep.2016.05.049 (2016).
327 Katter, K. *et al*. Transposon-mediated transgenesis, transgenic rescue, and tissue-specific gene expression in rodents and rabbits. *The FASEB Journal* 27, 930-941, doi:https://doi.org/10.1096/fj.12-205526 (2013).
328 Lee, D.-F. *et al*. Combining competition assays with genetic complementation strategies to dissect mouse embryonic stem cell self-renewal and pluripotency. *Nature Protocols* 7, 729-748, doi:10.1038/nprot.2012.018 (2012).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

329 Covelo-Molares, H., Souto, Y., Fidalgo, M. & Guallar, D. A Simple, Rapid, and Cost-Effective Method for Loss-of-Function Assays in Pluripotent Cells. LID - 10.1007/7651_2021_434 [doi].
330 Fuentes-Iglesias, A., Ameneiro, C., Guallar, D. & Fidalgo, M. Detecting and Modulating ER Stress to Improve Generation of Induced Pluripotent Stem Cells. LID - 10.1007/7651_2021_354 [doi].
331 Ran, F. A. *et al*. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* 8, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
332 Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods* 65, 55-63, doi:https://doi.org/10.1016/0022-1759(83)90303-4 (1983).
333 Štefková, K., Procházková, J. & Pachernik, J. Alkaline phosphatase in stem cells. *Stem Cells Int* 2015, 628368-628368, doi:10.1155/2015/628368 (2015).
334 Doetschman Tc Fau - Eistetter, H., Eistetter H Fau - Katz, M., Katz M Fau - Schmidt, W., Schmidt W Fau - Kemler, R. & Kemler, R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium.
335 Thomson, M. *et al*. Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* 145, 875-889, doi:10.1016/j.cell.2011.05.017 (2011).
336 Pereira, C. F. & Fisher, A. G. Heterokaryon-Based Reprogramming for Pluripotency. *Current Protocols in Stem Cell Biology* 9, 4B.1.1-4B.1.14, doi:https://doi.org/10.1002/9780470151808.sc04b01s9 (2009).
337 Guallar, D. *et al*. ADAR1-Dependent RNA Editing Promotes MET and iPSC Reprogramming by Alleviating ER Stress. *Cell Stem Cell* 27, 300-314.e311, doi:10.1016/j.stem.2020.04.016 (2020).
338 Macfarlan, T. S. *et al*. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57-63, doi:10.1038/nature11244 (2012).
339 Amlani, B. *et al*. Nascent Induced Pluripotent Stem Cells Efficiently Generate Entirely iPSC-Derived Mice while Expressing Differentiation-Associated Genes. *Cell Reports* 22, 876-884, doi:10.1016/j.celrep.2017.12.098 (2018).
340 Carter, Ava C., Davis-Dusenbery, Brandi N., Koszka, K., Ichida, Justin K. & Eggan, K. Nanog-Independent Reprogramming to iPSCs with Canonical Factors. *Stem Cell Reports* 2, 119-126, doi:10.1016/j.stemcr.2013.12.010 (2014).
341 Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nature Protocols* 1, 729-748, doi:10.1038/nprot.2006.98 (2006).
342 Yang, F. *et al*. DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* 26, 234-250.e237, doi:10.1016/j.stem.2020.01.004 (2020).
343 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4, 44-57, doi:10.1038/nprot.2008.211 (2009).
344 Guindon, S. & Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52, 696-704, doi:10.1080/10635150390235520 (2003).
345 Jones, D. T., Taylor Wr Fau - Thornton, J. M. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences.
346 Shannon, P. *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498-2504, doi:10.1101/gr.1239303 (2003).
347 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5, e13984-e13984, doi:10.1371/journal.pone.0013984 (2010).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

348 Gao, Y. *et al*. Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA Methylation and Hydroxymethylation in Reprogramming. *Cell Stem Cell* 12, 453-469, doi:10.1016/j.stem.2013.02.005 (2013).
349 You, Y. *et al*. Inhibition of MTA2 and MTA3 induces mesendoderm specification of human embryonic stem cells. *Biochemical and Biophysical Research Communications* 552, 142-149, doi:https://doi.org/10.1016/j.bbrc.2021.03.030 (2021).
350 Bode, D., Yu, L., Tate, P., Pardo, M. & Choudhary, J. Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. *Mol Cell Proteomics* 15, 878-891, doi:10.1074/mcp.M115.053207 (2016).
351 Masui, S. *et al*. Rex1/Zfp42 is dispensable for pluripotency in mouse ES cells. *BMC Developmental Biology* 8, 45, doi:10.1186/1471-213X-8-45 (2008).
352 Hackett, Jamie A. & Surani, M. A. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell* 15, 416-430, doi:10.1016/j.stem.2014.09.015 (2014).
353 Ghimire, S. *et al*. Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Sci Rep* 8, 5884, doi:10.1038/s41598-018-24051-5 (2018).
354 Malumbres, M. & Barbacid, M. To cycle or not to cycle: a critical decision in cancer. *Nature Reviews Cancer* 1, 222-231, doi:10.1038/35106065 (2001).
355 Sherr, C. J. & Roberts, J. M. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev* 18, 2699-2711, doi:10.1101/gad.1256504 (2004).
356 Bourillot, P.-Y. *et al*. Novel STAT3 Target Genes Exert Distinct Roles in the Inhibition of Mesoderm and Endoderm Differentiation in Cooperation with Nanog. *STEM CELLS* 27, 1760-1771, doi:https://doi.org/10.1002/stem.110 (2009).
357 Aksoy, I. *et al*. Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells. *Nature Communications* 5, 3719, doi:10.1038/ncomms4719 (2014).
358 Zhang, S. & Cui, W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J Stem Cells* 6, 305-311, doi:10.4252/wjsc.v6.i3.305 (2014).
359 Plath, K. & Lowry, W. E. Progress in understanding reprogramming to the induced pluripotent state. *Nature Reviews Genetics* 12, 253-265, doi:10.1038/nrg2955 (2011).
360 Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nature Reviews Genetics* 14, 427-439, doi:10.1038/nrg3473 (2013).
361 Silva, J. *et al*. Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLOS Biology* 6, e253, doi:10.1371/journal.pbio.0060253 (2008).
362 Theunissen, T. W. *et al*. Nanog Overcomes Reprogramming Barriers and Induces Pluripotency in Minimal Conditions. *Current Biology* 21, 65-71, doi:10.1016/j.cub.2010.11.074 (2011).
363 Yu, J. S. L. & Cui, W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. *Development* 143, 3050-3060, doi:10.1242/dev.137075 (2016).
364 Sokol, S. Y. Maintaining embryonic stem cell pluripotency with Wnt signaling. *Development* 138, 4341-4350, doi:10.1242/dev.066209 (2011).
365 Xu, Z. *et al*. Wnt/β-catenin signaling promotes self-renewal and inhibits the primed state transition in naïve human embryonic stem cells. *Proceedings of the National Academy of Sciences* 113, E6382, doi:10.1073/pnas.1613849113 (2016).
366 Mossahebi-Mohammadi, M., Quan, M., Zhang, J.-S. & Li, X. FGF Signaling Pathway: A Key Regulator of Stem Cell Pluripotency. *Front Cell Dev Biol* 8, 79-79, doi:10.3389/fcell.2020.00079 (2020).
367 Wongtrakoongate, P., Li, J. & Andrews, P. W. DNMT3B inhibits the re-expression of genes associated with induced pluripotency. *Experimental Cell Research* 321, 231-239, doi:https://doi.org/10.1016/j.yexcr.2013.11.024 (2014).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

368 Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat Struct Mol Biol* 18, 120-127, doi:10.1038/nsmb.2000 (2011).
369 Russell, R. *et al*. A Dynamic Role of TBX3 in the Pluripotency Circuitry. *Stem Cell Reports* 5, 1155-1170, doi:10.1016/j.stemcr.2015.11.003 (2015).
370 Lanner, F. & Rossant, J. The role of FGF/Erk signaling in pluripotent cells. *Development* 137, 3351-3360, doi:10.1242/dev.050146 (2010).
371 Xiao, S. *et al*. FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13) leukaemia/lymphoma syndrome. *Nature Genetics* 18, 84-87, doi:10.1038/ng0198-84 (1998).
372 Yang, P. *et al*. RCOR2 Is a Subunit of the LSD1 Complex That Regulates ESC Property and Substitutes for SOX2 in Reprogramming Somatic Cells to Pluripotency. *STEM CELLS* 29, 791-801, doi:https://doi.org/10.1002/stem.634 (2011).
373 Fidalgo, M. *et al*. Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proceedings of the National Academy of Sciences* 109, 16202, doi:10.1073/pnas.1208533109 (2012).
374 Adamo, A. *et al*. LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nature Cell Biology* 13, 652-659, doi:10.1038/ncb2246 (2011).
375 Ang, Y.-S. *et al*. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145, 183-197, doi:10.1016/j.cell.2011.03.003 (2011).
376 Wysocka, J. *et al*. WDR5 Associates with Histone H3 Methylated at K4 and Is Essential for H3 K4 Methylation and Vertebrate Development. *Cell* 121, 859-872, doi:10.1016/j.cell.2005.03.036 (2005).
377 Whyte, W. A. *et al*. Author Correction: Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 562, E24-E24, doi:10.1038/s41586-018-0487-2 (2018).
378 Hayashi, Y. *et al*. BMP-SMAD-ID promotes reprogramming to pluripotency by inhibiting p16/INK4A-dependent senescence. *Proc Natl Acad Sci U S A* 113, 13057-13062, doi:10.1073/pnas.1603668113 (2016).
379 Fei, T. *et al*. Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Res* 20, 36-44, doi:10.1101/gr.092114.109 (2010).
380 Li, Z. & Chen, Y.-G. Functions of BMP signaling in embryonic stem cell fate determination. *Experimental Cell Research* 319, 113-119, doi:https://doi.org/10.1016/j.yexcr.2012.09.016 (2013).
381 Moussaieff, A., Kogan, N. M. & Aberdam, D. Concise Review: Energy Metabolites: Key Mediators of the Epigenetic State of Pluripotency. *STEM CELLS* 33, 2374-2380, doi:https://doi.org/10.1002/stem.2041 (2015).
382 Gordeeva, O. TGFß Family Signaling Pathways in Pluripotent and Teratocarcinoma Stem Cells' Fate Decisions: Balancing Between Self-Renewal, Differentiation, and Cancer. *Cells* 8, 1500, doi:10.3390/cells8121500 (2019).
383 Tsogtbaatar, E., Landin, C., Minter-Dykhouse, K. & Folmes, C. D. L. Energy Metabolism Regulates Stem Cell Pluripotency. *Front Cell Dev Biol* 8, 87-87, doi:10.3389/fcell.2020.00087 (2020).
384 Parmenter, T. J. *et al*. Response of BRAF-mutant melanoma to BRAF inhibition is mediated by a network of transcriptional regulators of glycolysis. *Cancer Discov* 4, 423-433, doi:10.1158/2159-8290.CD-13-0440 (2014).
385 Papa, S., Choy, P. M. & Bubici, C. The ERK and JNK pathways in the regulation of metabolic reprogramming. *Oncogene* 38, 2223-2240, doi:10.1038/s41388-018-0582-8 (2019).
386 Leitch, H. G. *et al*. Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* 20, 311-316, doi:10.1038/nsmb.2510 (2013).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

387 Vossenaar, E. R., Zendman, A. J. W., van Venrooij, W. J. & Pruijn, G. J. M. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *Bioessays* 25, 1106-1118, doi:https://doi.org/10.1002/bies.10357 (2003).
388 Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* 106, 10171-10176, doi:10.1073/pnas.0900604106 (2009).
389 Asaga, H., Nakashima, K., Senshu, T., Ishigami, A. & Yamada, M. Immunocytochemical localization of peptidylarginine deiminase in human eosinophils and neutrophils. *Journal of Leukocyte Biology* 70, 46-51, doi:https://doi.org/10.1189/jlb.70.1.46 (2001).
390 Darrah, E., Rosen, A., Giles, J. T. & Andrade, F. Peptidylarginine deiminase 2, 3 and 4 have distinct specificities against cellular substrates: novel insights into autoantigen selection in rheumatoid arthritis. *Ann Rheum Dis* 71, 92-98, doi:10.1136/ard.2011.151712 (2012).
391 Aguilar-Martinez, E. *et al*. Screen for multi-SUMO-binding proteins reveals a multi-SIM-binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. *Proc Natl Acad Sci U S A* 112, E4854-E4863, doi:10.1073/pnas.1509716112 (2015).
392 Cossec, J.-C. *et al*. SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* 23, 742-757.e748, doi:10.1016/j.stem.2018.10.001 (2018).
393 Lee, M. G. *et al*. Functional interplay between histone demethylase and deacetylase enzymes. *Mol Cell Biol* 26, 6395-6402, doi:10.1128/MCB.00723-06 (2006).
394 Laugesen, A. & Helin, K. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* 14, 735-751, doi:10.1016/j.stem.2014.05.006 (2014).
395 Lewis, H. D. *et al*. Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nature Chemical Biology* 11, 189-191, doi:10.1038/nchembio.1735 (2015).
396 Saba, J. A., Liakath-Ali, K., Green, R. & Watt, F. M. Translational control of stem cell function. *Nature Reviews Molecular Cell Biology* 22, 671-690, doi:10.1038/s41580-021-00386-2 (2021).
397 Lawrence, M., Theunissen, T. W., Lombard, P., Adams, D. J. & Silva, J. C. R. ZMYM2 inhibits NANOG-mediated reprogramming. *Wellcome Open Res* 4, 88-88, doi:10.12688/wellcomeopenres.15250.1 (2019).
398 Lezmi, E. *et al*. The Chromatin Regulator ZMYM2 Restricts Human Pluripotent Stem Cell Growth and Is Essential for Teratoma Formation. *Stem Cell Reports* 15, 1275-1286, doi:10.1016/j.stemcr.2020.05.014 (2020).
399 Martinez-Gamero, C., Malla, S. & Aguilo, F. LSD1: Expanding Functions in Stem Cells and Differentiation. *Cells* 10, 3252, doi:10.3390/cells10113252 (2021).
400 Morgani, S. M. *et al*. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell reports* 3, 1945-1957, doi:10.1016/j.celrep.2013.04.034 (2013).
401 Hackett, J. A. *et al*. Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nature communications* 9, 4292-4292, doi:10.1038/s41467-018-06230-0 (2018).
402 Tsusaka, T., Fukuda, K., Shimura, C., Kato, M. & Shinkai, Y. The fibronectin type-III (FNIII) domain of ATF7IP contributes to efficient transcriptional silencing mediated by the SETDB1 complex. *Epigenetics Chromatin* 13, 52-52, doi:10.1186/s13072-020-00374-4 (2020).
403 Klein, D. C. & Hainer, S. J. in *Current Topics in Developmental Biology* Vol. 138 (ed Thomas G. Fazzio) 1-71 (Academic Press, 2020).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

404 Ito, K. & Ito, K. Metabolism and the Control of Cell Fate Decisions and Stem Cell Renewal. *Annual review of cell and developmental biology* 32, 399-409, doi:10.1146/annurev-cellbio-111315-125134 (2016).
405 Ly, C. H., Lynch, G. S. & Ryall, J. G. A Metabolic Roadmap for Somatic Stem Cell Fate. *Cell Metabolism* 31, 1052-1067, doi:10.1016/j.cmet.2020.04.022 (2020).
406 Gao, T., Díaz-Hirashi, Z. & Verdeguer, F. Metabolic Signaling into Chromatin Modifications in the Regulation of Gene Expression. *Int J Mol Sci* 19, 4108, doi:10.3390/ijms19124108 (2018).
407 Reid, M. A., Dai, Z. & Locasale, J. W. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nature cell biology* 19, 1298-1306, doi:10.1038/ncb3629 (2017).
408 Schvartzman, J. M., Thompson, C. B. & Finley, L. W. S. Metabolic regulation of chromatin modifications and gene expression. *J Cell Biol* 217, 2247-2259, doi:10.1083/jcb.201803061 (2018).
409 Etchegaray, J.-P. & Mostoslavsky, R. Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Molecular cell* 62, 695-711, doi:10.1016/j.molcel.2016.05.029 (2016).
410 Wu, J., Ocampo, A. & Belmonte, J. C. I. Cellular Metabolism and Induced Pluripotency. *Cell* 166, 1371-1385, doi:10.1016/j.cell.2016.08.008 (2016).
411 Harvey, A., Caretti, G., Moresi, V., Renzini, A. & Adamo, S. Interplay between Metabolites and the Epigenome in Regulating Embryonic and Adult Stem Cell Potency and Maintenance. *Stem Cell Reports* 13, 573-589, doi:10.1016/j.stemcr.2019.09.003 (2019).
412 Spyrou, J., Gardner, D. K. & Harvey, A. J. Metabolism Is a Key Regulator of Induced Pluripotent Stem Cell Reprogramming. *Stem Cells Int* 2019, 7360121-7360121, doi:10.1155/2019/7360121 (2019).
413 D'Aniello, C., Cermola, F., Patriarca, E. J. & Minchiotti, G. Metabolic–Epigenetic Axis in Pluripotent State Transitions. *Epigenomes* 3, doi:10.3390/epigenomes3030013 (2019).
414 Lu, V., Roy, I. J. & Teitell, M. A. Nutrients in the fate of pluripotent stem cells. *Cell Metabolism* 33, 2108-2121, doi:10.1016/j.cmet.2021.09.013 (2021).
415 Liu, W. & Chen, G. Regulation of energy metabolism in human pluripotent stem cells. *Cellular and Molecular Life Sciences* 78, 8097-8108, doi:10.1007/s00018-021-04016-0 (2021).
416 Nishimura, K., Fukuda, A. & Hisatake, K. Mechanisms of the Metabolic Shift during Somatic Cell Reprogramming. *Int J Mol Sci* 20, 2254, doi:10.3390/ijms20092254 (2019).
417 Rodriguez-Terrones, D. *et al*. A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep* 21, e48354-e48354, doi:10.15252/embr.201948354 (2020).
418 Teslaa, T. & Teitell, M. A. Pluripotent stem cell energy metabolism: an update. *EMBO J* 34, 138-153, doi:10.15252/embj.201490446 (2015).
419 Cliff, T. S. & Dalton, S. Metabolic switching and cell fate decisions: implications for pluripotency, reprogramming and development. *Current opinion in genetics & development* 46, 44-49, doi:10.1016/j.gde.2017.06.008 (2017).
420 Sun, H. *et al*. Lysine-specific histone demethylase 1 inhibition promotes reprogramming by facilitating the expression of exogenous transcriptional factors and metabolic switch. *Sci Rep* 6, 30903-30903, doi:10.1038/srep30903 (2016).
421 Warren, J. T., Wismar, J., Subrahmanyam, B. & Gilbert, L. I. Woc (without children) gene control of ecdysone biosynthesis in Drosophila melanogaster. *Molecular and Cellular Endocrinology* 181, 1-14, doi:https://doi.org/10.1016/S0303-7207(01)00404-X (2001).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

422 Quiquand, M. *et al*. New regulators of Drosophila eye development identified from temporal transcriptome changes. *Genetics* 217, iyab007, doi:10.1093/genetics/iyab007 (2021).
423 Connaughton, D. M. *et al*. Mutations of the Transcriptional Corepressor ZMYM2 Cause Syndromic Urinary Tract Malformations. *Am J Hum Genet* 107, 727-742, doi:10.1016/j.ajhg.2020.08.013 (2020).
424 Reiter, A. *et al*. Consistent Fusion of ZNF198 to the Fibroblast Growth Factor Receptor-1 in the t(8;13)(p11;q12) Myeloproliferative Syndrome. *Blood* 92, 1735-1742, doi:10.1182/blood.V92.5.1735 (1998).
425 Kulkarni, S., Reiter, A., Smedley, D., Goldman, J. M. & Cross, N. C. P. The Genomic Structure of ZNF198 and Location of Breakpoints in the t(8;13) Myeloproliferative Syndrome. *Genomics* 55, 118-121, doi:https://doi.org/10.1006/geno.1998.5634 (1999).
426 Ren, M. *et al*. Development of ZMYM2-FGFR1 driven AML in human CD34+ cells in immunocompromised mice. *Int J Cancer* 139, 836-840, doi:10.1002/ijc.30100 (2016).
427 Jawhar, M. *et al*. Cytogenetically cryptic ZMYM2-FLT3 and DIAPH1-PDGFRB gene fusions in myeloid neoplasms with eosinophilia. *Leukemia* 31, 2271-2273, doi:10.1038/leu.2017.240 (2017).
428 Strati, P. *et al*. Myeloid/lymphoid neoplasms with FGFR1 rearrangement. *Leukemia & Lymphoma* 59, 1672-1676, doi:10.1080/10428194.2017.1397663 (2018).
429 Cao, M. *et al*. ZMYM2-FGFR1 fusion as secondary change in acute myeloid leukemia. *Leukemia & Lymphoma* 60, 556-558, doi:10.1080/10428194.2018.1493733 (2019).
430 Liu, C., Jin, Y. & Fan, Z. The Mechanism of Warburg Effect-Induced Chemoresistance in Cancer. *Front Oncol* 11, 698023-698023, doi:10.3389/fonc.2021.698023 (2021).
431 Hahn, S. & Hermeking, H. ZNF281/ZBP-99: a new player in epithelial-mesenchymal transition, stemness, and cancer.
432 Duteil, D. *et al*. LSD1 promotes oxidative metabolism of white adipose tissue. *Nature communications* 5, 4093-4093, doi:10.1038/ncomms5093 (2014).
433 Kim, D., Kim, K. I. & Baek, S. H. Roles of lysine-specific demethylase 1 (LSD1) in homeostasis and diseases. *Journal of Biomedical Science* 28, 41, doi:10.1186/s12929-021-00737-3 (2021).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Mutations of the ZMYM2 protein alter cellular and tissue homeostasis, but its functional roles in pluripotency are largely unexplored. Here, I show that ZMYM2 is a key component of the core regulatory network that governs pluripotency. My data reveal that ZMYM2 is a roadblock to efficient somatic cell reprogramming and is required to fine-tune the self-renewal of ESCs and the early differentiation process. Mechanistically, I found that ZMYM2 and LSD1 associate for the transcriptional control of pluripotency. This study, therefore, establishes a novel transcriptional regulatory mode dedicated to reprogramming and pluripotency and opens new avenues for exploring the involvement of ZMYM2 in development and disease.