The Chromatin Regulator ZMYM2 Restricts Human Pluripotent Stem Cell Growth and Is Essential for Teratoma Formation

Elyad Lezmi,1,4 Uri Weissbein,1,4 Tamar Golan-Lev,1,2 Malka Nissim-Rafinia,2,3 Eran Meshorer,2,3,* and Nissim Benvenisty1,2,*

1The Azrieli Center for Stem Cells and Genetic Research, The Hebrew University, Jerusalem, Israel
2Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem, Israel
3Edmond and Lily Center for Brain Sciences (ELSC), The Hebrew University, Jerusalem, Israel
4Co-first authors
*Correspondence: eran.meshorer@mail.huji.ac.il (E.M.), nissimb@mail.huji.ac.il (N.B.)
https://doi.org/10.1016/j.stemcr.2020.05.014

SUMMARY

Chromatin regulators play fundamental roles in controlling pluripotency and differentiation. We examined the effect of mutations in 703 genes from nearly 70 chromatin-modifying complexes on human embryonic stem cell (ESC) growth. While the vast majority of chromatin-associated complexes are essential for ESC growth, the only complexes that conferred growth advantage upon mutation of their members were the repressive complexes LSD-CoREST and BHC. Both complexes include the most potent growth-restricting chromatin-related protein, ZMYM2. Interestingly, while ZMYM2 expression is rather low in human blastocysts, its expression peaks in primed ESCs and is again downregulated upon differentiation. ZMYM2-null ESCs overexpress pluripotency genes and show genome-wide promoter-localized histone H3 hyper-acetylation. These mutant cells were also refractory to differentiate in vitro and failed to produce teratomas upon injection into immunodeficient mice. Our results suggest a central role for ZMYM2 in the transcriptional regulation of the undifferentiated state and in the exit-from-pluripotency of human ESCs.

INTRODUCTION

Human embryonic stem cells (ESCs), derived from the inner cell mass (ICM) of the pre-implantation blastocyst, have the capacity for self-renewal in culture as well as the potential to differentiate into virtually any cell type of the developing fetus (Schuldiner et al., 2002; Thomson et al., 1998). Cultured ESCs differ from their embryonic origin in several molecular characteristics (Huang et al., 2014; Nakamura et al., 2016), but in vitro resetting of human primed ESCs to a more “naive” state has been reported to closely mimic ICM-like transcriptional and epigenetic features (Sagi and Benvenisti, 2016; Schlesinger and Meshorer, 2019; Theunissen et al., 2014; Ware et al., 2014; Weinberger et al., 2016; Yilmaz and Benvenisty, 2019).

Due to their hallmarks of self-renewal and pluripotency, human ESCs serve as a model system that has greatly contributed to our understanding of early human development and disease mechanisms (Avior et al., 2016). ESCs also hold a great promise for medical applications, such as cell therapy and tissue engineering (Trounson and DeWitt, 2016), but many obstacles remain in the way to the fulfillment of their clinical potential. One of the adverse characteristics of ESCs is their tumorigenic potential as undifferentiated cells, manifested by the formation of teratoma tumors upon in vivo transplantation (Allison et al., 2018).

A growing number of studies implicated chromatin and its associated proteins as central components regulating ESC pluripotency and differentiation (Meissner, 2010; Surani et al., 2007), as well as the transition from the naive to the primed state (Schlesinger and Meshorer, 2019).

In a recent study, we performed a genome-wide loss-of-function screen in haploid ESCs (Yilmaz et al., 2018). This work focused on cellular essentiality, with approximately 10% of the genes, taking part in various cellular processes, identified as essential. In contrast to essential genes, growth-restricting genes are genes that upon mutation confer a selective advantage in conventional ESC culture conditions. Less than 5% of the screened genes were identified as growth restricting. We hypothesized that these genes could exert their effects by either inducing apoptosis, by slowing down the cell cycle, or by inducing differentiation (as differentiating cells generally proliferate slower).

In this work we re-analyzed the data generated in our screen, specifically addressing the essentiality of chromatin-related factors, as well as whole epigenetic protein complexes. We identified ZMYM2 as the most growth-restricting chromatin-related gene for human ESCs, and generated ZMYM2-null ESCs to study its function. We report significant alterations in the transcription and histone modification of pluripotency-related genes upon ZMYM2 loss. We further show that ZMYM2 is playing a role in early differentiation in vitro, and is essential for the formation of ESC-derived teratomas in vivo, suggesting that it naturally functions to promote the exit-from-pluripotency of human ESCs.
RESULTS

ZMYM2 and LCH Complex Members Are Highly Expressed although Growth Restricting for Human ESCs

To characterize the epigenetic “essentialome” of human ESCs, we analyzed the data generated in our previous loss-of-function genetic screen in human (primed) ESCs (Yilmaz et al., 2018), focusing only on chromatin-related factors and the various complexes they form, as defined in the EpiFactors database (Medvedeva et al., 2015). Out of 703 epigenetic factors examined, 233 genes (~33%) were significantly essential for ESC growth, while only 68 genes (~9.7%) were growth restricting. These numbers represent an over 3-fold enrichment of essential genes, and a 2-fold enrichment of growth-restricting genes in chromatin-related genes, when compared with the entire gene pool. ZMYM2 (zinc finger MYM-type containing 2, also known as ZNF198) was identified as the most growth-restricting gene (false discovery rate [FDR] = 6.18 × 10⁻⁷, Figure 1A). When we assessed the total essentiality of epigenetic complexes, we found, as expected, that most complexes are essential. However, BHC and LSD-CoREST complexes were exceptionally enriched with growth-restricting genes (Figure 1B). Interestingly, both complexes are transcriptional repressors, based on the same core LSD1/KDM1A-CoREST/RCOR1-HDAC1/2 (LCH) repressive complex, and both include ZMYM2. To examine the expression levels of the nine genes common to both BHC and LSD-CoREST, we mined publicly available datasets, comparing their expression between ESCs (median expression from 57 independent RNA sequencing [RNA-seq] experiment) and 30 other different cell and tissue types taken from the GTEx dataset (Carithers et al., 2015). We found that all four core members of the LCH complex, as well as ZMYM2, are highly expressed in ESCs, and that ESCs cluster separately from all other cell types over these nine genes (Figure 1C). To analyze the expression of ZMYM2 during differentiation, we compared RNA-seq data of ESCs, to in-vitro-differentiated embryoid bodies (EBs), and to in-vivo-differentiated teratomas, and found that ZMYM2 expression is highest in undifferentiated ESCs (Figure 1D). We additionally performed a western blot analysis on wild-type ESCs and their teratoma derivative, validating that ZMYM2 protein levels significantly drop upon differentiation (Figure 1A). To extend our analysis to different states of pluripotency, we gathered expression data from four studies that performed primed to naive conversion of ESCs (Guo et al., 2016; Messmer et al., 2019; Pastor et al., 2016; Theunissen et al., 2014), as well as two studies that provided RNA-seq data from human blastocysts and their (primed) ESC derivatives (Warrier et al., 2018; Yan et al., 2013). In all studies, ZMYM2 expression was significantly lower in naive ESCs compared with primed cells (Figure 1E). Similarly, ZMYM2 was expressed at low levels in human blastocysts and then upregulated upon generation of primed ESCs (Figure 1F).

Taken together, ZMYM2 is the major growth-restricting chromatin-related gene in ESCs, and its expression peaks at the primed state in comparison with naive ESCs or differentiated cells, both in vitro and in vivo.

ZMYM2-Null Mutation Leads to Adverse Transcriptional Effects

To further investigate the role of ZMYM2 in the transcriptional regulation of ESCs, we have generated, via a lentiviral CRISPR/Cas9 system, three independent ZMYM2-mutant clones (named Z6, Z15, and Z22) and three matching control clones bearing the same CRISPR vector without guide RNA (Emp1, Emp2, and Emp3). RNA-seq alignment showed lack of aligned reads downstream to the Cas9 cut site in exon 17, in all mutant clones, with Z6 showing overall low ZMYM2 expression levels, possibly via nonsense-mediated decay initiation in this clone (Figure S1B). ZMYM2 protein was not detected in any of the mutant clones by western blot analysis (Figure 2A) and mutation of the gene was validated at the DNA level (Figure S1C).

We further ruled out the possibility of off-target effects generated by our CRISPR system, via analysis of the genomic sequence of the top 10 off-target loci predicted by the CRISTA algorithm (Abadi et al., 2017) (Figure S1C; Table S4), and finally validated that all of our cells have a normal intact karyotype via e-karyotyping (Ben-David et al., 2013) (Figure S1D).

The morphology of ZMYM2 mutants appears similar to the control cells, presenting typical human ESC morphology (Figure S1E) and they express normal levels of the core pluripotency factors OCT4 (POUSF1), NANOG, and LIN28A (Adewumi et al., 2007) (Figure S1F).

Since our CRISPR screen identified ZMYM2 as growth restricting (its loss confers the cells with a selective advantage in culture), we sought to explore growth-related phenotypes in ZMYM2-null ESCs. We examined the baseline rate of apoptosis by annexin V and propidium iodide (PI) staining followed by fluorescence-activated cell sorting (FACS) analysis, to find no difference in the apoptosis rate of ZMYM2 mutants (Figure 2B, left). Next we analyzed the cell-cycle profile, by live DNA staining followed by FACS, of ZMYM2 mutant and control populations to find that ZMYM2 mutation did not affect the proliferative capacity of the cells (Figure 2B, right). To further characterize the mutant cells, we performed RNA-seq and a principal-component analysis over the expression of all genes, showing that ZMYM2-mutant cells cluster closer to their matched controls than to other ESC lines examined and...
Figure 1. Loss-of-Function Genetic Screening Identifies ZMYM2 as the Major Growth-Restricting Chromatin-Related Gene in Human ESCs

(A) Analysis of loss-of-function growth phenotype in human ESCs of all epigenetic factors (data taken from Yilmaz et al., 2018). The volcano plot shows the CRISPR scores (log_2FC) versus –log(p value) of genes encoding all epigenetic factors (dots in tricolor), all protein-coding genes are depicted by gray dots. The pie chart demonstrates the division of the genes to the different categories.

(B) Bar histograms describing the essential and growth-restricting gene fractions comprising each chromatin-modifying complex. The left panel shows a schematic representation of the gene members of the BHC and LSD-CoREST complexes. Heatmap and hierarchical clustering depict the expression profile of all genes common to both BHC and LSD-CoREST complexes, in ESCs, and across 31 tissues and transformed cell lines. ZMYM2 and members of the core LSD1-CoREST-HDAC1/2 complex (in bold) are mostly upregulated in ESCs.

(C) The expression profile of all epigenetic factors (dots in tricolor), all protein-coding genes are depicted by gray dots. The pie chart demonstrates the division of the genes to the different categories.

(D) Bar charts depicting relative ZMYM2 expression in various cell states: (D) ZMYM2 expression is downregulated upon differentiation in vitro into embryoid bodies (EBs) and in vivo into teratomas. (E) ZMYM2 expression is downregulated upon primed to naive conversion of human ESCs. (F) ZMYM2 expression is upregulated in (primed) ESCs compared with their blastocyst of origin. Error bars indicate mean ± SE; n, number of samples; df, degrees of freedom; p, p value as calculated by two-tailed Student’s t test. Data in (D) originate from independent samples, in (E and F) the x axis indicates the study that generated the data.
are overall similar to the cells carrying the empty vector (Figure S1G).

Although initially no visible phenotype was detected upon ZMYM2 loss, differential expression analysis revealed 1,072 differentially expressed genes upon ZMYM2 loss (FDR < 0.05). Of those, 683 were upregulated and 389 were downregulated (Figure 2C; Table S1), in accordance with the speculated repressive role of the LCH-ZMYM2 chromatin-modifying complex. Among the differentially expressed upregulated genes, we observed known pluripotency factors, such as UTF1, FGF4, and DPPA2 (John et al., 2008; Li and Belmonte, 2017). The naive markers TFCP2L1...
and DPPA5 (Messmer et al., 2019; Theunissen et al., 2014) were also upregulated while the primed pluripotency markers ZIC2, ZIC5, and SFRP2 (Guo et al., 2016; Messmer et al., 2019; Pastor et al., 2016) were all downregulated (Figure 2C). To understand whether ZMYM2 loss caused a transition of the transcriptional program between different pluripotency states, we compared our data with the expression data from primed to naive conversion (Guo et al., 2016; Messmer et al., 2019; Pastor et al., 2016; Theunissen et al., 2014) and ESC derivation from blastocysts (Warrier et al., 2018; Yan et al., 2013), establishing a high confidence list of differentially expressed genes between primed and naive ESCs (considering only genes significantly differentially expressed in at least five out of the six studies, Table S2). Interestingly, there is a significant overlap between genes that are differentially expressed in both naive compared with primed cells and ZMYM2-null cells compared with matched controls (p < 10^-16, χ^2 test, Figure 2D). To validate that ZMYM2 loss indeed drives the cells toward an earlier developmental stage, we analyzed the expression values of the 100 most significant differentially expressed genes upon ZMYM2 depletion, in ZMYM2-null ESCs, matched controls, primed ESCs, matched naive ESCs, human blastocysts, and matched-derived (primed) ESCs (Warrier et al., 2018; Yan et al., 2013). In a hierarchical clustering, our control ESCs formed an isolated group alongside all of the primed ESCs, while ZMYM2 mutants, the two blastocysts, and naive cells clustered outside the primed group, with ZMYM2-null ESCs clustering outside the dataset together with one of the naive samples (Figure 2E). Our results suggest a clear link between ZMYM2 and the pluripotency transcriptional network, and its loss leads to transcriptional similarities to cells in an earlier developmental state of pluripotency.

**ZMYM2 Loss Leads to H3 Hyper-acetylation at Promoter Regions**

Since the histone modifiers HDAC1/2 and LSD1 are known to interact with ZMYM2, we sought to understand the effect of ZMYM2 loss on histone modifications. We performed a chromatin immunoprecipitation assay followed by high-throughput sequencing (ChIP-seq), using antibodies against pan-acetylated histone H3 (H3Ac) and monomethylated H3 Lys4 (H3K4me1). Our H3K4me1 analysis could not detect differences in the methylation signal between the ZMYM2-null and control samples (Figure S2A), and no differentially methylated sites have passed the FDR threshold (Table S5). In our H3Ac analysis we detected a clear effect of ZMYM2 loss on histone H3 acetylation (Figure 3A), observing robust hyper-acetylation at promoter regions in the mutant cells (Figure 3B), while the expected relationship between the expression level of genes and the amount of H3-acetylation at their transcription start site (Figure S2B) was kept upon ZMYM2 loss (Figure S2C). Our analysis identified H3Ac peaks at the promoters of 14,019 genes overall, out of these 820 promoters were hyper-acetylated and 168 were hypo-acetylated in the mutant samples (FDR < 0.05 and absolute-logFC > 0.5, Figure 3C; Table S6). Gene set enrichment analysis (GSEA) analysis (Subramanian et al., 2005) revealed that the hypo-acetylated genes are enriched with gene ontology (GO) terms related to somatic differentiation while the hyper-acetylated genes are enriched with terms related to germ cell fate (Figure 3C; Table S7), consistent with the idea that ZMYM2 depletion reverts the cells to an earlier developmental state. As expected, GSEA analysis on the genes that were highly acetylated in both conditions and unchanged upon ZMYM2 loss (FDR >0.9, absolute-logFC < 0.1) were enriched with housekeeping-related GO terms, such as “Cell cycle”, “Cytoskeleton organization,” and “RNA processing” (Table S7). We indeed identified a positive correlation (r = 0.57, p < 10^-16) between the fold-change in expression and fold-change in promoter-H3-acetylation (Figure 3D), and observed the expected differential acetylation in the pluripotency factors that were found differentially expressed (Figures 2C and 3D). In addition, we identified the naive pluripotency factors; ZFP42 (REX1) KLF2 and KLF4 (Kalkan and Smith, 2014; Yilmaz and Benvenisty, 2019) to be hyper-acetylated, while the primed marker DUSP6 (Messmer et al., 2019) was hypo-acetylated (Figures 3D and 3E). Overall our results suggest that ZMYM2 functions to repress a large set of germline-related genes, possibly in an HDAC1/2-mediated manner, while on the other hand it normally primes a second smaller set of genes that are related to somatic differentiation, likely as a secondary effect of the first gene set.

**ZMYM2 Loss-of-Function Delays the Exit-from-Pluripotency and Is Essential for the Formation of Human ESC-Derived Teratomas**

To assess the differentiation capacity of ZMYM2-mutant ESCs, we induced differentiation by either omitting basic fibroblast growth factor (bFGF) and/or adding retinoic acid (RA) to the culture medium. Two days after induction of differentiation, ZMYM2-mutant cells retained high expression of the pluripotency markers NANOG and OCT4 in all conditions where bFGF was withdrawn, as observed by quantitative real-time PCR analysis (Figure 4A). Next, we performed a longer differentiation assay, culturing the cells in the presence of RA and the absence of bFGF for 8 days. Here, ZMYM2 mutants showed a delay of approximately 2 days, in the reduction of NANOG and OCT4 expression when compared with the control cells (Figure 4B). These results suggest that ZMYM2 expression is required for the early stages of differentiation.
To assess the tumorigenic and the in vivo differentiation potential of ZMYM2-null ESCs, we performed a teratoma formation assay by subcutaneously injecting undifferentiated cells, from all six mutant and control cell lines, into immune-deficient mice. With only a single exception, all 13 mice injected with control cells developed mature teratomas when examined at 8 weeks after injection. Interestingly, we were unable to detect any tumors in 12 out of the 20 injections of ZMYM2-null ESCs, while in the remaining eight injections extremely small tumors were detected (Figure 4C), achieving a maximal teratoma weight of 48 mg compared with control-derived teratomas weighing up to 2.3 g (Figure 4D). To obtain a deeper understanding of the transcriptional program acting within the small ZMYM2-null-derived teratomas, we performed RNA-seq on two of these tumors and two matched control teratomas. ZMYM2-null-derived tumors showed significantly increased levels of the pluripotency markers OCT4, NANOG, DNMT3B, DPPA4, and LIN28A (Figure 4E), in agreement with the in vitro differentiation experiments (Figures 4A and 4B). We also performed histological analysis on the small tumors generated by ZMYM2-null ESCs. Hematoxylin and eosin staining revealed clear endodermal, mesodermal,
and ectodermal structures in control teratomas, while the ZMYM2-null-derived tumors showed clusters of unidentified cells alongside adipocyte-like structures (Figure 4F, left). We next used immunohistochemical staining to assess the expression of the pluripotency marker OCT4 and found high presence of this marker in ZMYM2-null-derived tumors but not in the controls (Figure 4F, right). Together, these data strongly support a role for ZMYM2 in early differentiation both in vitro and in vivo.
DISCUSSION

The human gene ZMYM2 is known mainly in the context of its genomic fusion with FGFR1, causing myeloproliferative syndromes (Xiao et al., 1998). ZMYM2 encodes for a 155-kDa protein that can undergo phosphorylation and sumoylation at multiple sites (UniProt Consortium, 2019). It has a zinc-finger domain typical of protein-protein interaction (Kunapuli et al., 2003) and was shown to interact with proteins involved in processes ranging from DNA damage response to mRNA splicing (Kasyapa et al., 2005).

While the function of human ZMYM2 is largely unknown, mutation in its close Drosophila homolog, WOC (without children), can lead to both male and female sterility (Wismar et al., 2000). In the context of human ESCs, several studies found that ZMYM2 interacts with both OCT4 and NANOG (van den Berg et al., 2010; Ding et al., 2012; Wang et al., 2006), and a genetic screen found that the knockdown of ZMYM2 can enhance the reprogramming capacity of somatic cells into induced pluripotent stem cells (Toh et al., 2016).

ZMYM2 was also reported to have a direct chromatin binding ability in HeLa cells, and it has been shown to associate with the repressive LCH histone-modifying complex and to stabilized on the chromatin (Gocke and Yu, 2008). The murine gene RE1-silencing transcription factor (REST) is known to recruit LCH to chromatin, where it silences neuronal promoters in non-neuronal tissues and maintains self-renewal and pluripotency in mouse ESCs (Singh et al., 2008). Interestingly, ZMYM2 and REST binding to LCH is mutually exclusive, and ZMYM2-like proteins are required for the repression of E-cadherin (a gene known to be repressed by LSD1), but not REST-responsive genes (Gocke and Yu, 2008), suggesting that ZMYM2 might perform a similar task to the one performed by REST, albeit in silencing pluripotency genes by associating with the LCH complex. Supporting its anti-pluripotency role, Zmym2 expression in mouse naive pluripotent stem cells was found to anti-correlate with the expression of bona fide naive pluripotency genes (Kumar et al., 2014). When considering our initial classification of ZMYM2 as growth restricting in ESCs, and in the light of the above mentioned, we suggest a direct link between ZMYM2 and the pluripotency transcriptional network.

Our results show adverse upregulation of transcription upon ZMYM2 loss, in agreement with its speculated mode of action via the repressive LCH complex. We accordingly observed robust promoter-localized histone H3 hyper-acetylation upon ZMYM2 loss, comparable with hyper-acetylation observed upon treatment with HDAC inhibitors (Hezroni et al., 2011b; Karmodiya et al., 2012).

Interestingly, HDAC inhibitors were previously used in the conversion of primed ESCs to a naive state (Ware et al., 2014), emphasizing the link of histone acetylation to naive pluripotency.

The transcriptional landscape of ZMYM2-null ESCs is defined as pluripotent according to all molecular markers, thus we suggest that all phenotypes observed are due to transcriptional changes within the broad pluripotency framework and that cell identity is shifted between different pluripotency states. We show that the directionality of differential expression between ZMYM2-mutant and controls is comparable with the differential expression between blastocyst or naive ESCs and primed ESCs.

In accordance with our findings, murine Zmym2 expression in naive ESCs was found to be highly upregulated upon the induction of epiblast-like differentiation, and Zmym2 loss of function was found to delay and reduce epiblast marker expression upon induction of epiblast cell fate (Hackett et al., 2018). Most recently, Zmym2 ablation in mouse ESCs was shown to promote Nanog-mediated reprogramming of somatic cells into naive cells, while its overexpression seems to promote the spontaneous differentiation of mouse ESCs (Lawrence et al., 2019).

The most striking phenotype in our ZMYM2 mutants is their inability to generate mature teratomas. This feature appears to be linked to their difficulty to exit the pluripotent state upon differentiation induction in vitro, and suggests a non-trivial change in cell identity upon mutation of a single gene. Teratomas with undifferentiated cell foci were previously described as teratocarcinomas, and are considered malignant compared with benign teratomas (Damjanov and Andrews, 2007). Our ZMYM2-null-derived teratomas retained an undifferentiated population, but were extremely small in mass. Since teratoma, as well as teratocarcinoma tumors, are predominantly composed of differentiated cells, we speculate that a block of exit-from-pluripotency upon ZMYM2 loss could explain this small teratoma phenotype. Interestingly, a recent work reported that disruption of LSD1 inhibits teratoma formation (Osada et al., 2018), reinforcing our suggested role of LCH-ZMYM2 interaction in cellular fate acquisition. We suggest that ZMYM2 repression is mediated by the repressive LCH complex, but we do not rule out the possibility that ZMYM2 acts by another unknown mechanism.

In conclusion, ZMYM2 peak expression is at the primed state of human ESCs, while its loss halts the exit-from-pluripotency, but also provides ESCs with a growth advantage in culture, by preventing spontaneous differentiation. We identified ZMYM2 as a major component of the network regulating gene expression in human ESCs and suggest that it normally acts to promote exit-from-pluripotency by wide repression of early pluripotency genes, which
can in turn permit the activation and priming of late pluripotency and early differentiation genes.

**EXPERIMENTAL PROCEDURES**

**Cell Culture**
Human ESCs were cultured on a feeder layer of mitomycin-inactivated mouse embryonic fibroblasts in standard human ESC medium: KnockOut Dulbecco’s modified Eagle’s medium supplemented with 15% KnockOut Serum Replacement (Thermo Fisher Scientific), 2 mM L-glutamine, 0.1 mM nonessential amino acids, 50 units/ml penicillin, 50 μg/ml streptomycin, 0.1 mM β-mercaptoethanol, and 8 ng/ml bFGF. ESCs were also alternatively grown in feeder-free conditions on Matrigel-coated plates (Corning) in mTeSR1 medium (STEMCELL Technologies). Cells were routinely checked for the presence of mycoplasma and maintained in a humidified incubator at 37°C and 5% CO2. Passaging was performed with trypsin-EDTA (Biological Industries) and subsequent addition of 10 μM ROCK inhibitor Y-27632 (Stemgent) to the fresh culture medium.

**Generation of ZMYM2-Null ESCs**
CRISPR/Cas9 guide RNA was cloned into LentiCRISPRv2 plasmid and electroploated into CSE54 cells. Two days after electroporation cells were selected with 0.34 μM puromycin treatment and resistant colonies were manually picked and expanded. Control clones received LentiCRISPRv2 plasmid without guide RNA. ZMYM2 knockout was validated using RNA-seq and western blot.

**Apoptosis Assay**
Following dissociation using TrypLE Select (Thermo Fisher Scientific), human ESCs were washed once with PBS and incubated with annexin V and PI using the MEBCYTO Apoptosis Kit according to the manufacturer’s instructions (MBL). Next, cells were filtered through a 70-μm cell strainer and analyzed in BD FACSAria III for the proportion of annexin V- and PI-positive apoptotic cells.

**Cell-Cycle Analysis**
Cells were dissociated using TrypLE Select (Thermo Fisher Scientific), incubated for 30 min with 10 μg/ml Hoechst 33342 (Sigma-Aldrich) in ESC medium. Stained cells were filtered through a 70-μm cell strainer and analyzed in a BD FACSaria II flow cytometer for their cell-cycle profile.

**High-Throughput RNA-Seq**
Cultured cells were harvested four to six passages after mutagenesis (p. 20–22 from original cell line derivation), and total RNA was extracted using RNeasy Kits (QIAGEN), frozen teratomas were homogenized in TRI Reagent (MRC), and RNA was extracted according to the manufacturer’s protocol. Total RNA samples (1 μg RNA) were enriched for mRNAs by pull-down of poly(A)+ RNA. RNA-seq libraries were prepared using the TruSeq RNA Library Prep Kit v.2 (Illumina) according to the manufacturer’s protocol and sequenced using Illumina NextSeq 500 to generate 85 bp single-end reads.

**ChIP-Seq**
ChIP assay was performed as described previously (Alajem et al., 2015; Hezroni et al., 2011a) with minor modifications: samples were sonicated 10 times for 30 s with a 30-s break between each sonication (20 min total). For immunoprecipitation antibodies against histone 3 pan-acetylation (Millipore 06–599) and histone 3 lysine 4 monomethylation (Millipore 07-436) were used. Subsequent library preparation for sequencing was performed as described previously (Blecher-Gonen et al., 2013; Schlesinger et al., 2017).

**ESC Differentiation Assays**
Cells were plated on Matrigel-coated 12-well plates (200k cells per well) and treated for 2 days of with conventional ESC medium lacking bFGF (EB medium), mTeSR medium supplemented with 10 μM RA, and EB medium supplemented with 10 μM RA (the last was repeated also for the 8-day differentiation experiment). Cells were harvested at each time point and stored at −80°C. RNA extraction was performed using the RNeasy Kit (QIAGEN).

**Real-Time qPCR**
Total RNA was reverse-transcribed into first-strand complementary DNA (Quantabio). The quantitative real-time PCR reaction consisted of initial incubation at 50°C for 2 min and denaturation at 95°C for 10 min. The cycling parameters were as follows: 95°C for 15 s and 60°C for 30 s. After 40 cycles, the reactions underwent a final dissociation cycle as follows: 95°C for 15 s, 60°C for 1 min, 95°C for 15 s, and 60°C for 15 s. Expression of each gene was normalized to GAPDH expression.

**Teratoma Formation and Analysis**
All experimental procedures were approved by the ethics committee of the Hebrew University of Jerusalem. Human ESCs (2 × 10^6) were resuspended in 100 μl medium and 100 μl Matrigel and subcutaneously injected into NOD-SCID IL2rg−/− immunodeficient mice. Tumors were excised 8 weeks after injection. Hematoxylin and eosin staining was performed as described previously (Kopper et al., 2010). For immunohistochemistry, slides were thawed and fixed in formalin and blocked using 20 mg/ml BSA and 0.1% Triton X. Samples were stained with OCT4 antibody (sc-8626, 2 μg/ml), and a secondary antibody (Jackson ImmunoResearch 705-546-147, 140 μg/ml).

**Data Analysis**
All next-generation sequencing (NGS) data (public and original) underwent quality control and were processed using TrimGalore, aligned to GRCh38 using STAR (Dobin et al., 2013) for RNA samples and bowtie2 (Langmead and Salzberg, 2012) for ChIP samples, and genomic loci were annotated using GENCODE version 29 (Frankish et al., 2019). Coverage files were generated using bamCoverage (deeptools) (Ramirez et al., 2016), and visualizations of alignments were made using IGV (Robinson et al., 2011). Tissue expression medians (TPM) were extracted from the GTEx database (Carithers et al., 2015). Expression values (TPM/MM normalized expression) of ESCs, EBs, and teratomas were calculated out of 56, 27, and 37 independent RNA-seq
samples (respectively) that were retrieved from the SRA (Leinonen et al., 2011) database. All published SRA samples used in this study are detailed in Table S3.

**Statistics**

All experiments were performed in three biological replicates, except for the RNA-seq of teratoma samples (two replicates, due to the lack of ZMYM2-derived teratomas). The statistical tests: two-tailed Student’s t test for independent/paired samples, Pearson correlation, and $\chi^2$ test for goodness of fit, were calculated using R base functions. Sample sizes, degrees of freedom, and p values are detailed in the figures and figure legends of all results. Distance similarity was calculated using the dist R function implementing Euclidian distance calculation. TMM normalization of RNA read counts, as well as differential expression analysis, were performed using EdgeR (Robinson et al., 2009) (Robinson et al., 2009). Peak calling for ChIP-seq was performed using MACS2 (Zhang et al., 2008) and differential binding analysis was done using the DiffBind R package (Ross-Innes et al., 2012). For further information, please refer to the Data and code availability section.

**Data and Code Availability**

Sequencing data performed for this study were deposited in ArrayExpress under the accessions E-MTAB-8169 (RNA-seq) and E-MTAB-8170 (ChiP-seq). All publicly available NGS samples used in this study are detailed in the supplementary information. All code used for statistical analysis and figure generation is available at https://github.com/elyadlezmi/ZMYM2/.

**SUPPLEMENTAL INFORMATION**

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2020.05.014.

**AUTHOR CONTRIBUTIONS**

E.L., U.W., E.M., and N.B. conceived the study and designed the analysis. T.G.-L. assisted with cell culture and performed the immunofluorescence of teratoma samples. M.N.-R. instructed and performed the ChiP-seq experiment. U.W. performed the bioinformatic analysis on epigenetic complexes, generated ZMYM2-null clones, performed the differentiation experiments, and initiated the teratoma formation assay. E.L. performed the western blot analysis, ChiP-seq of ZMYM2-null ESCs, RNA-seq of teratomas, all NGS-related bioinformatic analysis of original and public data, performed the statistical analyses, and generated the figures. E.L. wrote the manuscript with input from U.W., E.M., and N.B. E.M. and N.B. supervised the study.

**ACKNOWLEDGMENTS**

The authors thank Yishai Avior for critically reading the manuscript. N.B. is the Herbert Cohn Chair in Cancer Research. This work was partially supported by the U.S-Israel Binational Science Foundation (grant 2015089), by the Israel Science Foundation (grant 494/17), by the Rosetrees Trust, and by the Azrieli Foundation.


Supplemental Information

The Chromatin Regulator ZMYM2 Restricts Human Pluripotent Stem Cell Growth and Is Essential for Teratoma Formation

Elyad Lezmi, Uri Weissbein, Tamar Golan-Lev, Malka Nissim-Rafinia, Eran Meshorer, and Nissim Benvenisty
Figure S1

A. ZMYM2 protein expression in ESCs and Teratoma. ZMYM2 expression is higher in Teratoma compared to ESCs.

B. Gene expression analysis of LIN28A, NANOG, and POU5F1 in control and ZMYM2-null cells using microarrays.

C. CRISTA scores for ZMYM2 gene (chr13: 20061175-20061197) and intergenic regions.

D. Expression ratio of ZMYM2-null compared to control cells.

E. Western blot analysis of ZMYM2 and Tubulin proteins in control and ZMYM2-null ESCs.

F. Bar graph showing expression levels of LIN28A, NANOG, and POU5F1 in control and ZMYM2-null cells.

G. Principal component analysis (PCA) plot showing gene expression differences between ESCs, ZMYM2-null, Control (empty vector), and Other ESC lines.
Figure S2

A

B

C

Mean H3K4me1 signal (CPM)

Mean H3Ac signal (CPM)

Distance from TSS (bp)

Distance from TSS (bp)

Distance from TSS (bp)
Supplementary Figure Legends

**Figure S1 | Characterization of ZMYM2 KO cells. Related to Figures 1D, 2B-C**

**A.** A western blot analysis showing reduction in ZMYM2 protein levels upon differentiation of wild-type ESC into teratomas. **B.** A schematic representation of the CRISPR-Cas9 vector used to mutate \textit{ZMYM2} and its effect on RNA sequencing data of the mutated locus. The figure shows the lack of aligned reads downstream to the Cas9 cut site in \textit{ZMYM2}-mutated cell lines. **C.** DNA alignment tracks of ZMYM2-null ESCs and controls, showing the ZMYM2 locus and loci who are the most likely off-targets of the gRNA used for mutating ZMYM2. **D.** e-karyotyping analysis of the ZMYM2-null and empty-vector cell lines used in this study. **E.** Representative morphology of \textit{ZMYM2}-null (left) and wild-type (right) ESCs cultured on matrigel. **F.** Bars depict the expression levels of core pluripotency factors in ZMYM2-null and wild type ESCs. **G.** Total gene expression PCA analysis of ESCs used in this study compared to other ESCs and differentiated cells, showing that ZMYM2-null cells are more similar to their matched control than to other ESC cell lines.

**Figure S2 | Chip-seq of ZMYM2-null cells. Related to Figure 3**

**A.** Meta-gene analysis depicts the accumulative value of histone H3 Lysin-4 monomethylation around the main transcription start site (TSS) of all annotated genes, in ZMYM2-null and control ESCs. **B-C.** Metagene analysis depicts the accumulative value of histone H3 pan-acetylation around the TSS for different levels of gene transcription. In control (\textit{B}) and ZMYM2-null ESCs (\textit{C}).
Supplementary Table 8. Primer sequences.

<table>
<thead>
<tr>
<th>Gene</th>
<th>F</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCT4</td>
<td>GACAGGGGGAGGGGAGGAGCTAGG</td>
<td>CTTCCCTCCAAACCAGTTGCCCAAAAC</td>
</tr>
<tr>
<td>NANOG</td>
<td>CATGAGTGGATCCAGCTTG</td>
<td>CCTGAATAAGCAGATCCATGG</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AGCCACATCGCTCAGACACC</td>
<td>GTACTCAGCGCCAGCATCG</td>
</tr>
</tbody>
</table>

gRNA used for CRISPR KO of ZMYM2

GGATATGATGAGTGAAGACG

qRT-PCR primers
<table>
<thead>
<tr>
<th>Core pluripotency</th>
<th>Naïve pluripotency</th>
<th>Primed pluripotency</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGF4</td>
<td>ESRRB</td>
<td>CER1</td>
</tr>
<tr>
<td>GDF3</td>
<td>DPPA3</td>
<td>FGF5</td>
</tr>
<tr>
<td>POU5F1</td>
<td>KLF2</td>
<td>FOXA2</td>
</tr>
<tr>
<td>SALL4</td>
<td>KLF4</td>
<td>LEF1</td>
</tr>
<tr>
<td>SOX2</td>
<td>KLF5</td>
<td>NODAL</td>
</tr>
<tr>
<td>TDGF1</td>
<td>NANOG</td>
<td>ZIC2</td>
</tr>
<tr>
<td>UTF1</td>
<td>TBX3</td>
<td>ZIC5</td>
</tr>
<tr>
<td>DPPA2</td>
<td>TFCP2L1</td>
<td>CD24</td>
</tr>
<tr>
<td>NANOG</td>
<td>ZFP42</td>
<td>SFRP2</td>
</tr>
<tr>
<td>DPPA4</td>
<td>DPPA5</td>
<td>CD57</td>
</tr>
<tr>
<td>DNMT3B</td>
<td>POU3F1</td>
<td>CD90</td>
</tr>
<tr>
<td>LIN28A</td>
<td>KLF17</td>
<td>OTX2</td>
</tr>
<tr>
<td></td>
<td>CD7</td>
<td>DUSP6</td>
</tr>
<tr>
<td></td>
<td>CD75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CD77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CD130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNMT3L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GATA6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IL6ST</td>
<td></td>
</tr>
</tbody>
</table>