CellPress

Cell Stem Cell
Short Article

Alternative Splicing of MBD2 Supports Self-Renewal in Human Pluripotent Stem Cells

Yu Lu,[1,2,3,13] Yuin-Han Loh,[2,4,5,13] Hu Li,[6,7,8,13] Marcella Cesana,[2,4] Scott B. Ficarro,[1,2,9] Jignesh R. Parikh,[9,10] Nathan Salomonis,[11] Cheng-Xu Delon Toh,[5] Stelios T. Andreadis,[12] C. John Luckey,[3,14] James J. Collins,[6,7,14] George Q. Daley,[2,4,14,*] and Jarrod A. Marto[1,2,9,14,*]

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA
[4]Department of Medicine, Division of Pediatric Hematology Oncology, Howard Hughes Medical Institute, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, MA 02115, USA
[5]A*STAR Institute of Molecular and Cell Biology and Department of Biological Sciences, National University of Singapore, Singapore 138673, Singapore
[6]Howard Hughes Medical Institute, Department of Biomedical Engineering and Center of Synthetic Biology, Boston University, Boston, MA 02215, USA
[7]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA
[8]Center for Individualized Medicine, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA
[9]Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[10]Bioinformatics Program, Boston University, Boston, MA 02115, USA
[11]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
[12]Department of Chemical and Biological Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260, USA
[13]Co-first author
[14]Co-senior author
*Correspondence: george.daley@childrens.harvard.edu (G.Q.D.), jarrod_marto@dfci.harvard.edu (J.A.M.)
http://dx.doi.org/10.1016/j.stem.2014.04.002

SUMMARY

Alternative RNA splicing (AS) regulates proteome diversity, including isoform-specific expression of several pluripotency genes. Here, we integrated global gene expression and proteomic analyses and identified a molecular signature suggesting a central role for AS in maintaining human pluripotent stem cell (hPSC) self-renewal. We demonstrate that the splicing factor SFRS2 is an OCT4 target gene required for pluripotency. SFRS2 regulates AS of the methyl-CpG binding protein MBD2, whose isoforms play opposing roles in maintenance of and reprogramming to pluripotency. Although both MDB2a and MBD2c are enriched at the *OCT4* and *NANOG* promoters, MBD2a preferentially interacts with repressive NuRD chromatin remodeling factors and promotes hPSC differentiation, whereas overexpression of MBD2c enhances reprogramming of fibroblasts to pluripotency. The miR-301 and miR-302 families provide additional regulation by targeting *SFRS2* and *MDB2a*. These data suggest that OCT4, SFRS2, and MBD2 participate in a positive feedback loop, regulating proteome diversity in support of hPSC self-renewal and reprogramming.

INTRODUCTION

The transcription factors OCT4, NANOG, and SOX2 are master regulators of pluripotency in embryonic stem cells (ESCs) (De Los Angeles et al., 2012) and, along with Klf4 and c-Myc, facilitate reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) (Park et al., 2008; Takahashi and Yamanaka, 2006). ESCs are indispensable models of early development, whereas iPSCs hold great promise as cell-based therapeutics that circumvent the immunologic and ethical hurdles of embryo-derived cells. As a result, significant effort has been invested in elucidating the mechanisms that underlie stem cell function, with a particular emphasis on these core pluripotent genes. Despite the requirement of OCT4, SOX2, and NANOG in stem cell function (De Los Angeles et al., 2012), discrepancies between ostensibly identical pluripotent cell lines (Gore et al., 2011), in addition to the divergent lineage commitment properties of iPSCs derived from different adult tissues (Kim et al., 2010), illustrate that the molecular network balancing self-renewal, pluripotency, and lineage commitment is not yet resolved.

Recently, functional genomics and molecular profiling approaches have been used to explore the broader role of the core pluripotent factors in stem cell biology. These studies expanded the set of genes that support pluripotency (Chia et al., 2010) and defined a biochemical network centered around OCT4, NANOG, and SOX2 that is highly enriched for genes essential for development and stem cell function (Kim et al., 2008). Furthermore, use of chromatin immunoprecipitation

(ChIP)-chip (Boyer et al., 2005) and ChIP sequencing (Guenther et al., 2010) has established the landscape of genetic targets for several key pluripotent factors and defined correlations between promoter co-occupancy and transcriptional activation. In parallel, genome-scale molecular measurement technologies have been used to quantify differences in epigenetic modifications (Gifford et al., 2013; Xie et al., 2013), gene expression (Tang et al., 2010), and protein translation (Ingolia et al., 2011) in addition to protein expression and phosphorylation (Brill et al., 2009; Phanstiel et al., 2011) between pluripotent stem cells and other cell types. These data provide a rich resource of molecular information, although it remains challenging to generate specific hypotheses from these disparate data types or establish mechanistic links between these molecular profiles and the core pluripotent factors.

Recently, alternative splicing (AS) has garnered attention as a possible means by which stem cells regulate the expression of gene and protein isoforms in order to support pluripotency and self-renewal. Indeed, functional roles for alternatively spliced gene products of *NANOG*, *FOXP1*, and *Tcf7l1* have been demonstrated (Das et al., 2011; Gabut et al., 2011; Salomonis et al., 2010). In addition, the muscleblind-like family of RNA binding proteins was found to repress pluripotency by mediating expression of several somatic cell-specific protein isoforms, including FOXP1 (Han et al., 2013). These data illustrate a general role for AS in pluripotent cells; however, the specific splicing factors and mechanistic links to the core pluripotent genes, which work in concert to reinforce a ground state of self-renewal, remain unresolved.

The splicing factor SFRS2 (also known as SC35) is essential for embryonic development (Xiao et al., 2007) and regulates transcription (Lin et al., 2008). Although several splicing substrates have been identified (Lin et al., 2008), no pluripotency-specific role has been established for SFRS2.

The methyl-DNA binding protein methyl-CpG binding domain protein 2 (MBD2) comprises two predominant isoforms, *MBD2a* and *MBD2c* (Hendrich and Bird, 1998), that share the same MBD domain but differ in the C-terminal region as a result of AS. MBD2 silences gene expression by binding to methylated DNA and recruiting the nucleosome remodeling and deacetylation (NuRD) complex (Zhang et al., 1999). Although NuRD has well-established roles in development (Reynolds et al., 2012), the function of *MBD2* in stem cells is not well understood. In fact, data from two recent studies are inconsistent with respect to the impact of *MBD2* in somatic cell reprogramming (Lee et al., 2013; Onder et al., 2012), although the possibility of isoform-specific function was not considered.

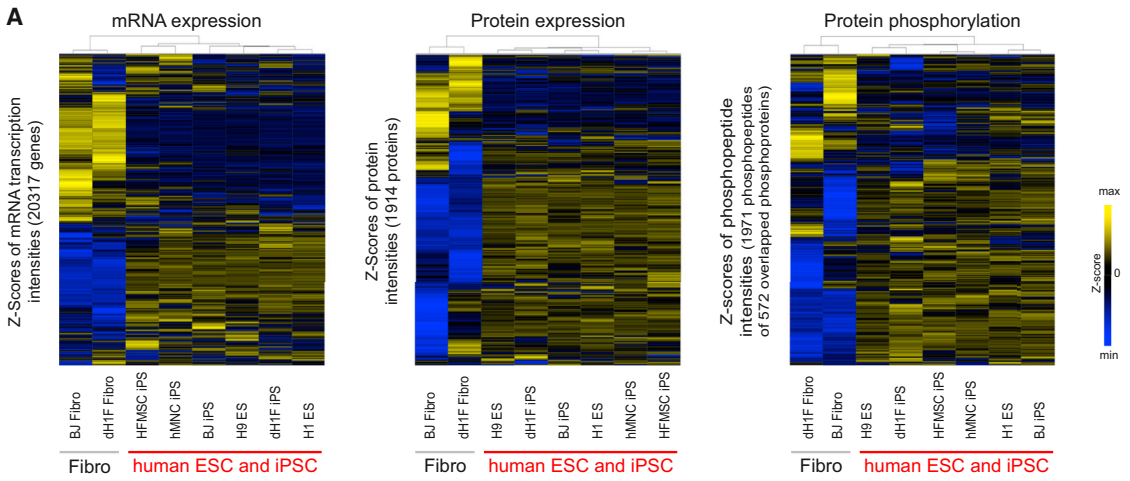
In this study, we establish mechanistic links between OCT4 and SFRS2 and demonstrate that these factors work in concert to regulate AS of *MBD2*. Expression of specific *MBD2* isoforms is further regulated by the microRNA (miRNA) machinery, and we find that the resulting gene products play opposing functional roles with respect to self-renewal of human PSCs (hPSCs) and reprogramming of fibroblasts. Consistent with these observations, MBD2 isoforms target the promoters of *OCT4* and *NANOG* in human ESCs (hESCs) but differ dramatically in their ability to biochemically interact with chromatin remodeling proteins. Collectively, our results suggest a positive feedback loop comprised of OCT4, SFRS2, and splice products of MBD2 that regulates proteome diversity in order to support a self-renewing ground state.

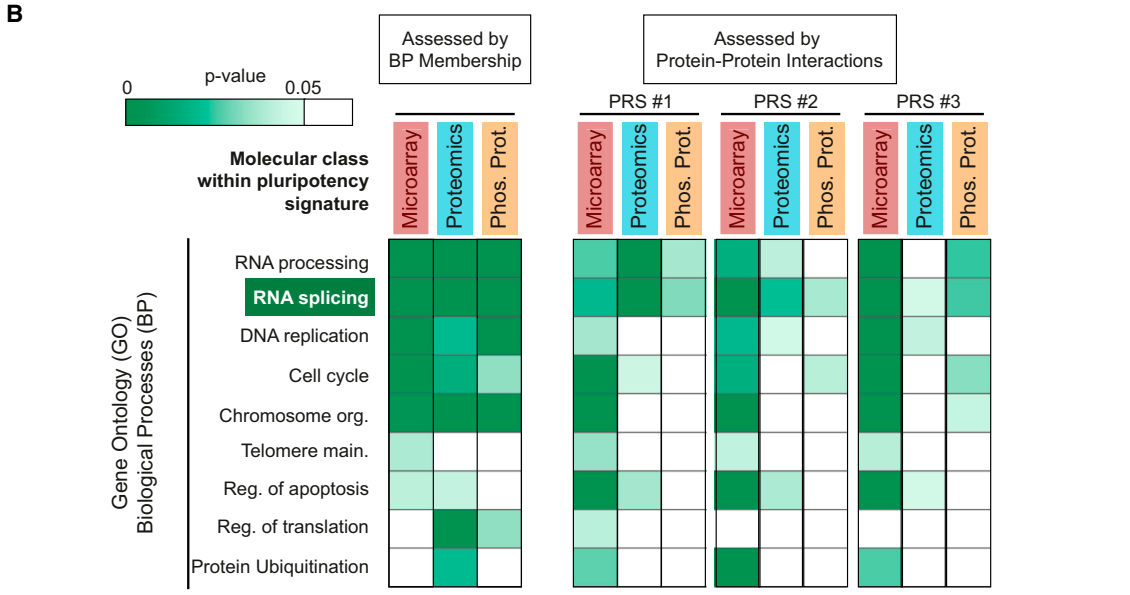
RESULTS

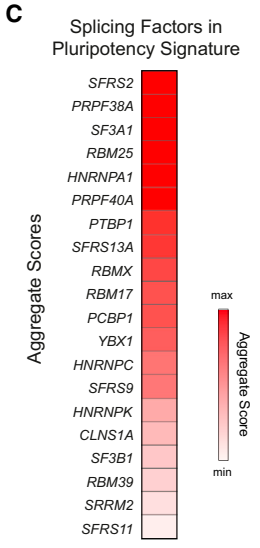
First, we sought to identify a molecular signature for pluripotency that integrated gene and protein expression, in addition to protein phosphorylation in cells representing a broad range of genetic backgrounds and cell fates (Figures S1A and S2 and Table S1 available online). Independent hierarchical clustering of each data type revealed that hPSCs from different tissue types exhibit protein phosphorylation, gene transcription, and protein expression profiles that are clearly distinct from differentiated fibroblasts (DFs; Figure 1A), and each molecular class contributes a subset of unique genes to the signature (Figure S1B). Notably, the molecular divergence observed between pluripotent cells and DFs was considerably higher than it was in hPSCs (Figure S1C); in addition we confirmed that the phosphorylation signature was strongly linked to cell type rather than specific culture conditions (Figure S1D). As is typical of high-throughput measurements (Brill et al., 2009; Phanstiel et al., 2011; Tang et al., 2010), classification of gene function within the pluripotency signature based on gene ontology (GO) biological process revealed enrichment of several disparate pathways (Figure 1B, left).

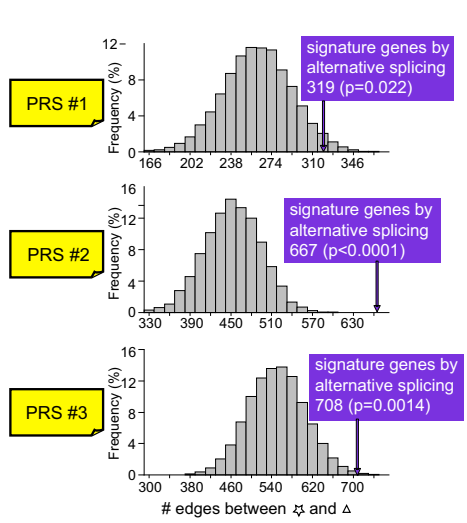
There is growing appreciation that the principles of network theory are applicable to human physiology whereby extended physical, genetic, or metabolic relationships between biomolecules may have predictive power with respect to biological outcomes (Balázsi et al., 2011; Vidal et al., 2011). Consistent with this notion, we next asked whether interpretation of our molecular signature data within the context of physical interaction networks would highlight specific cellular functions that support self-renewal. Accordingly, we assessed the number of physical interactions between constituent genes of the pluripotency signature and three positive reference sets (PRSs) of pluripotent factors derived from (1) literature survey, (2) a recent functional genomics study, and (3) proteins defined as biochemical interactors of Oct4 or Nanog (Figure S1E, Supplemental Experimental Procedures, and Table S2). This analysis revealed that only members of the RNA splicing pathway are consistently enriched across each measurement class (Figure 1B, right, and Table S3). Additional analysis (Supplemental Experimental Procedures) of the splicing factors in our pluripotent molecular signature suggested that the splicing factor *SFRS2* might be an important mediator of pluripotency (Figure 1C and Table S3).

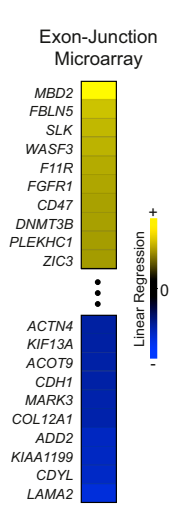
Given the role of SFRS2 in AS, we next compared the levels of spliced isoforms for 16,084 genes in hESCs and DFs and found that the spliced products from 2,974 genes differed between these cell types (Figure S3A and Table S4). Strikingly, we observed that 1,424 of these were not otherwise represented in the set of pluripotency signature genes (Figure S3A). As with other molecular classes of the pluripotent signature (Figure 1B), gene products subject to AS in hESCs are enriched for physical interactions with the PRSs (Figure 1D and Table S3). Extension of this analysis to GO annotation revealed a consistent enrichment of factors related to transcription regulation and chromatin modification (Figure S3B and Table S3); in total, we observed











(legend on next page)

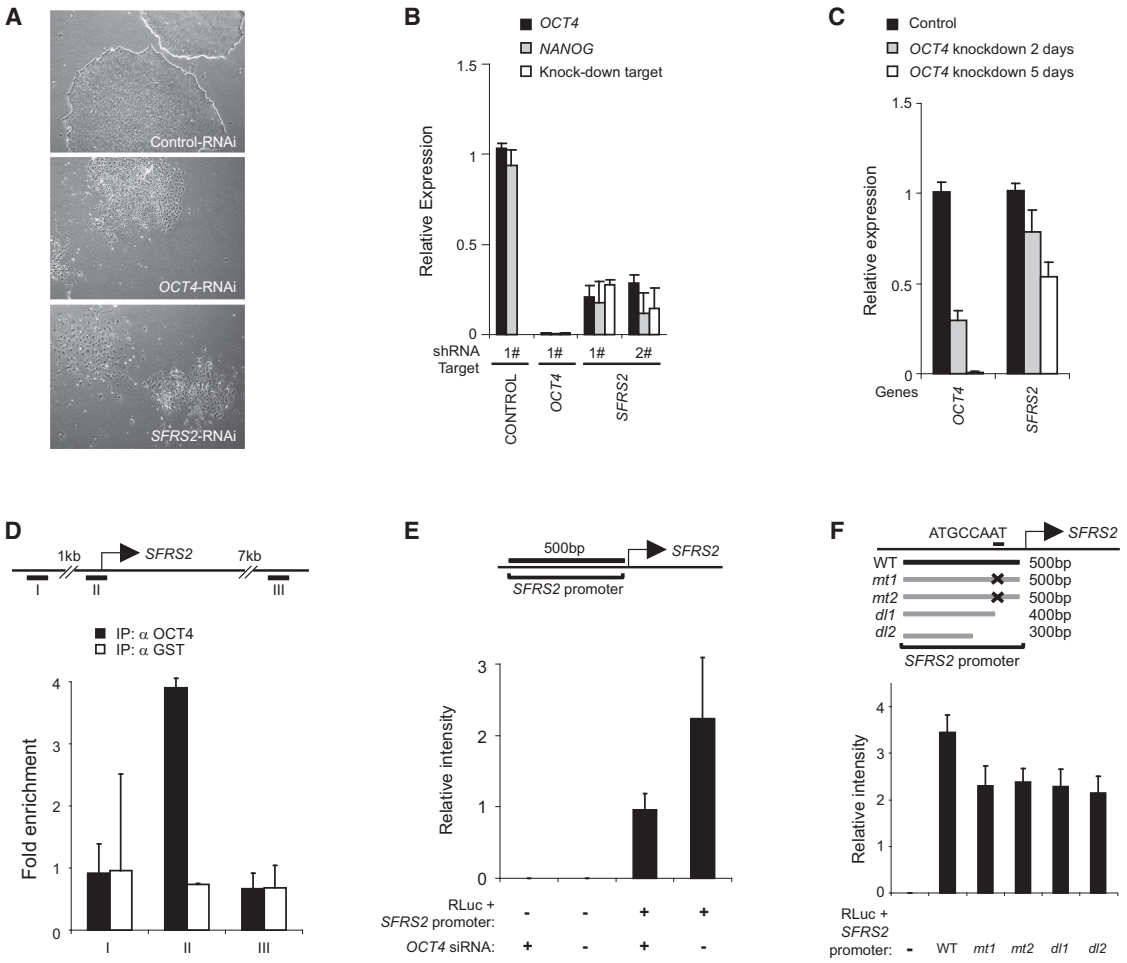


Figure 2. *OCT4* and *SFRS2* Display Interdependent Functional Links in hPSCs

(A and B) *SFRS2* is required to support self-renewal. Lentiviral small hairpin RNA (shRNA)-mediated depletion of *SFRS2* disrupted pluripotency in H1 ESCs as monitored by colony morphology (A) and expression of *OCT4* and *NANOG* (B).

(C) Depletion of *OCT4* in H1 ESCs for 2 and 5 days led to a coordinate decrease in the expression of *SFRS2*.

(D) OCT4 selectively binds the proximal promoter region of *SFRS2* in H1 ESCs.

(E) OCT4 depletion in human clone 9 iPSCs disrupts luciferase expression downstream of the native *SFRS2* promoter.

(F) Mutation or deletion of the predicted OCT4 binding motif (ATGCCAAT) in the proximal *SFRS2* promoter region decreased downstream luciferase expression in clone 9 iPSCs. See also Figure S4. Error bars represent the mean ± SEM.

236 alternatively spliced genes that spanned these pathways. Within the exon junction microarray data (Table S4), *MBD2* had the highest prediction score for AS between hESCs and DFs (Figure 1E and Table S4).

Next, we sought to establish specific links between *SFRS2*, *MBD2*, and the machinery supporting pluripotency. Depletion of endogenous *SFRS2* disrupted self-renewal in hESCs as gauged by cell morphology (Figure 2A), expression of *OCT4*

Figure 1. Analysis of the Molecular Signature Associated with hPSCs Suggests a Central Regulatory Role for RNA Splicing

(A) Independent hierarchical clustering of microarray and proteomic data demonstrated that hPSCs are molecularly distinct in comparison to DFs at the level of gene expression (left), protein expression (middle), and protein phosphorylation (right).

(B) Left, analysis of pluripotency signature genes according to their membership in gene ontology (GO) biological processes (BP) revealed enrichment of multiple pathways. Right, further analysis of genes within each GO-BP pathway and measurement class (gene, protein, and phosphoprotein) on the basis of physical interactions with three positive reference sets (PRSs) of pluripotent factors (see Figure S1E, Supplemental Experimental Procedures, and Table S2) suggested that the RNA splicing pathway is strongly associated with pluripotency. Main., maintenance; Org., organization; Reg., regulation.

(C) Splicing factors within the pluripotent molecular signature were individually ranked (see the Supplemental Experimental Procedures and Table S3) with SFRS2 as the top candidate.

(D) Alternatively spliced genes associated with hESCs are enriched for physical interactions with PRSs of pluripotent factors. Null distributions (gray bars) were created by random selection (10,000 iterations) of identically sized gene sets from the background of all genes detected by exon-junction microarray.

(E) The methyl-DNA binding protein *MBD2* displays the strongest alternative splicing pattern between hESCs and DFs on the basis of the linear regression analysis (see the Supplemental Experimental Procedures).

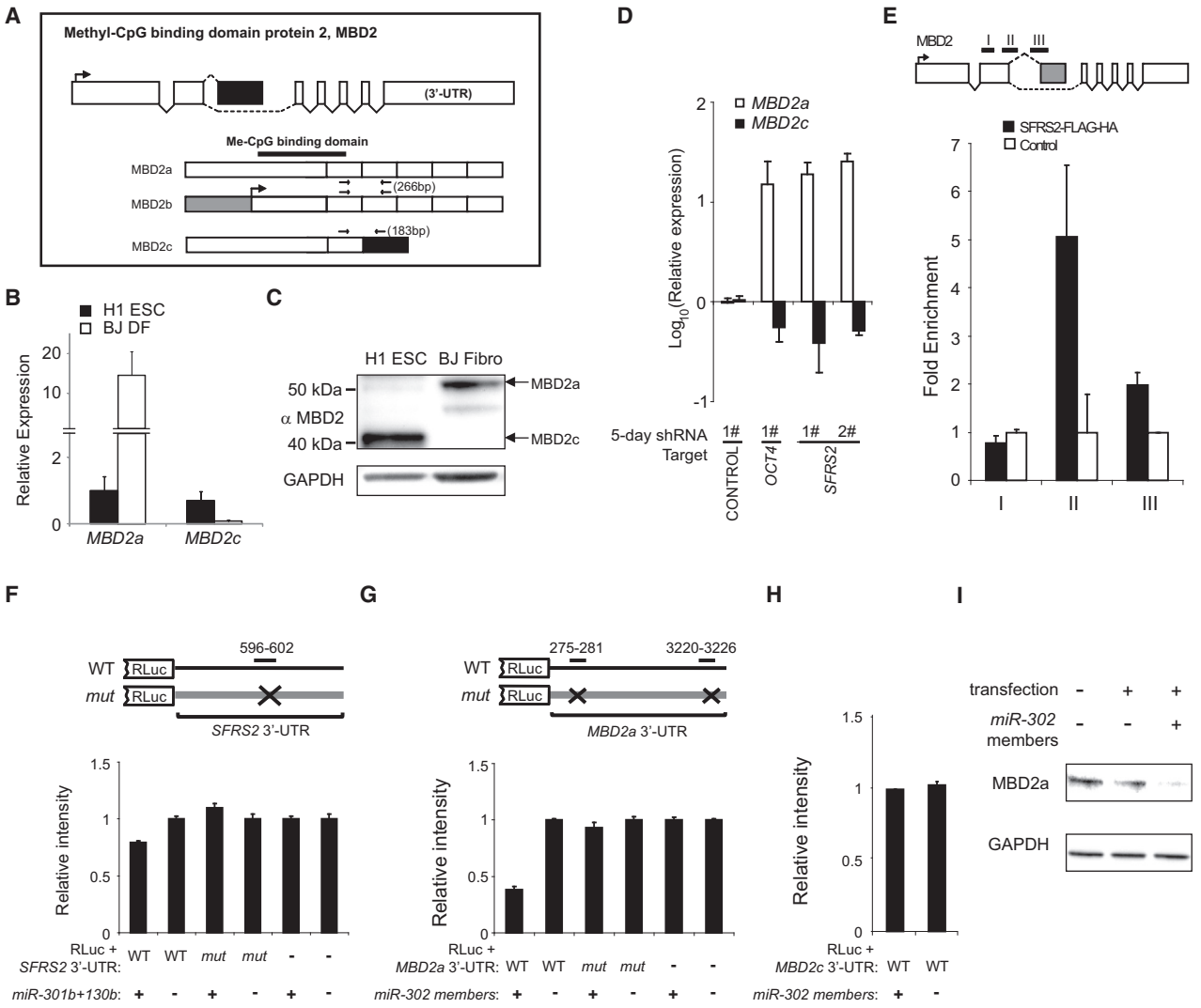


Figure 3. *MBD2* Isoform Expression Is Independently Regulated by the Splicing Factor SFRS2 and the *miR-302* Family of miRNAs in hPSCs

(A) Exon and protein graph for the methyl-CpG binding protein *MBD2*. Dashed lines indicate splice sites. Protein segments corresponding to each exon are annotated with predicted functional domains and primer locations.

(B and C) Verification of the distinct MBD2 isoforms in H1 ESCs and BJ DFs by quantitative RT-PCR (qRT-PCR; B) and western blotting (C).

(D) Lentiviral shRNA-mediated depletion of *OCT4* and *SFRS2* independently led to a significant increase in *MBD2a* expression along with reduced levels of *MBD2c* after 5 days in H1 ESCs.

(E) Exogenously expressed SFRS2-FLAG-HA preferentially binds to *MBD2* pre-mRNA at intron 2 (primer pairs II and III, each spanning into exons 2 and 3, respectively) but not inside exon 2 (primer pair I) in H1 ESCs.

(F) *miR-301b* and *miR-130b* suppress luciferase expression in the context of wild-type, but not mutated, sequences corresponding to the 3′ UTR of *SFRS2* in HeLa cells.

(G–I) *miR-302* cluster members target the 3′ UTR of *MBD2* in an isoform-specific manner.

(G) *miR-302* cluster members specifically suppressed luciferase expression upstream of the wild-type, but not mutated, *MBD2a* 3′ UTR sequence in HeLa cells.

(H) *miR-302* cluster members did not affect luciferase expression upstream of the *MBD2c* 3′ UTR sequence in HeLa cells.

(I) Overexpression of *miR-302* cluster members in 293T cells reduced expression of endogenous MBD2a. Error bars represent the mean ± SEM.

and *NANOG* (Figure 2B), alkaline phosphatase staining (Figure S4A), and cell colony integrity (Figure S4B). We observed a coordinate decrease in expression level of *SFRS2* upon *OCT4* depletion in hESCs (Figure 2C); importantly, this effect was specific to *SFRS2* and not observed for other splicing factors (Figure S4C). Furthermore, we found that OCT4 bound directly to the promoter of *SFRS2* in hESCs (Figure 2D) and drove expression of luciferase downstream of the native *SFRS2* promoter in vitro (Figure 2E). The specificity of this interaction was confirmed by mutation or deletion within the predicted OCT4 binding site of the *SFRS2* promoter (Figure 2F). These data provide evidence for functional and genetic links between OCT4 and *SFRS2* in hPSCs.

MBD2 comprises multiple isoforms (Hendrich and Bird, 1998) (Figure 3A). We detected preferential gene- and protein-level expression of the MBD2c and MBD2a isoforms in H1 ESCs

and BJ DFs, respectively (Figures 3B and 3C). Interestingly, depletion of endogenous *SFRS2* or *OCT4* in hESCs led to a dramatic increase in expression of *MBD2a* and a reduction in *MBD2c* (Figures 3D and S4D). Next, we probed for a direct biochemical interaction between SFRS2 and *MBD2* pre-mRNA by assaying RNA that coprecipitated with exogenously expressed SFRS2-FLAG-HA. We observed that SFRS2 bound to *MBD2* pre-mRNA specifically at intron 2, preceding exon 3, which is unique to the ESC-predominant *MBD2c* isoform (Figure 3E), suggesting that SFRS2 may mediate alternative splicing of this methyl-DNA binding protein in hPSCs.

In addition, close inspection of the 3′ untranslated region (3′ UTR) of *SFRS2* and *MBD2a* (but not *MBD2c*) revealed potential binding motifs for *miR-301* and *miR-302*, miRNA families that are functionally associated with lineage commitment and self-renewal (Figure S4E) (Bar et al., 2008). We confirmed that overexpression of *miR-301b* and *miR-130b* reduced luciferase driven by the wild-type *SFRS2* 3′ UTR, whereas mutation of the *miR-301* motif restored luciferase expression (Figure 3F). Similarly, *miR-302* specifically targeted the 3′ UTR of *MBD2a* (Figure 3G) but not that of *MBD2c* (Figure 3H). Indeed, we confirmed that exogenous expression of *miR-302* reduced levels of MBD2a in vivo (Figure 3I). These data suggest that the *miR-301* and *miR-302* families may independently regulate *SFRS2* and *MBD2* in order to fine-tune the expression of MBD2 isoforms.

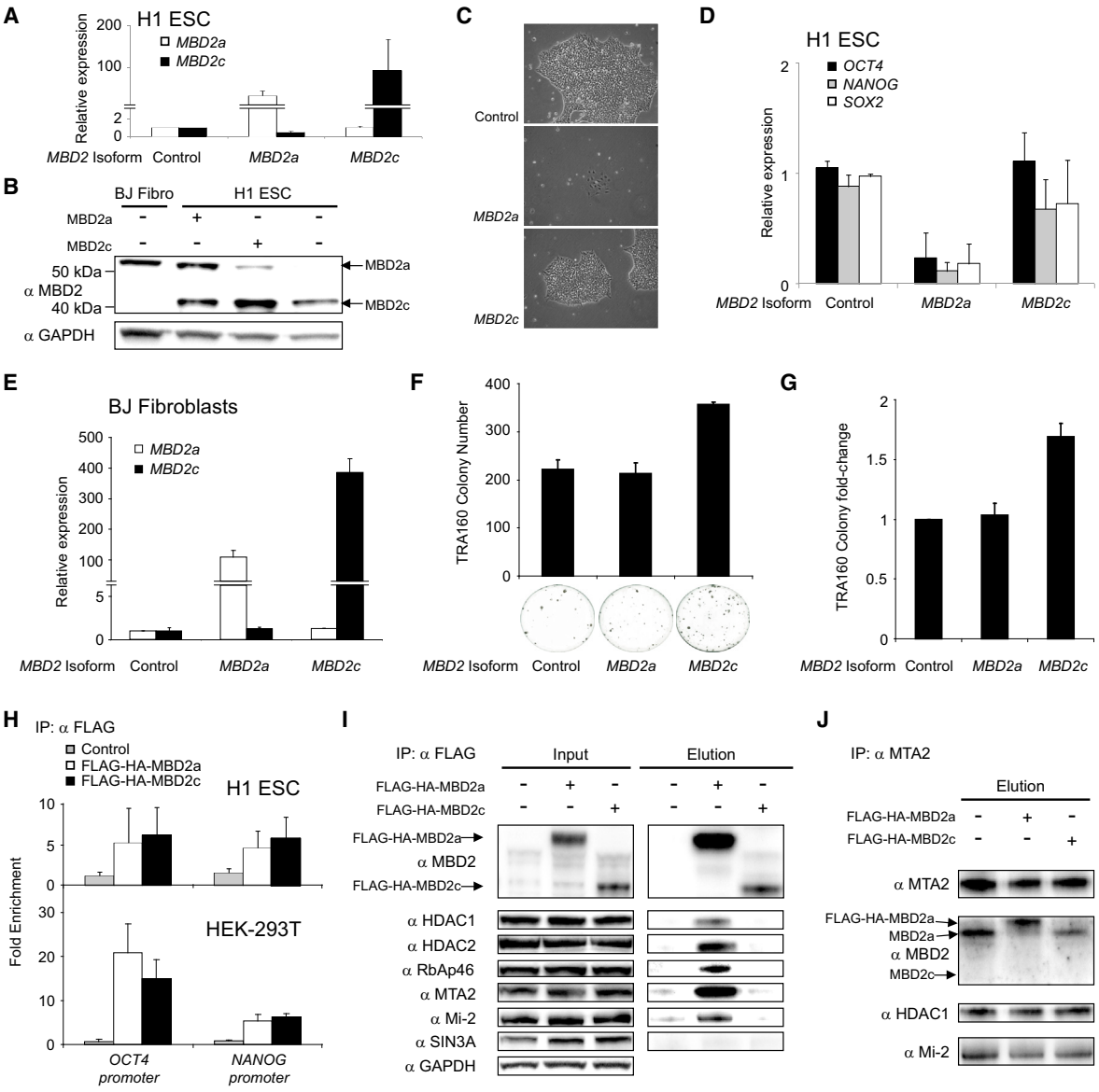
Next, we investigated the functional roles of MBD2 isoforms in hPSCs. Overexpression of MBD2a (Figures 4A and 4B) disrupted pluripotency, as evidenced by cell morphology (Figure 4C) in addition to reduced expression of *OCT4*, *NANOG*, and *SOX2* (Figure 4D). In contrast, increased MBD2c levels had no effect in hESC on the basis of these measures. However, the addition of the ESC-specific *MBD2c* isoform (Figure 4E) to a cocktail of reprogramming factors enhanced reprogramming efficiency in BJ DFs, whereas exogenous expression of *MBD2a* had no effect (Figures 4F and 4G). These data suggested that MBD2a and MBD2c play opposing roles in pluripotency. ChIP indicated that MBD2a and MBD2c were enriched at *OCT4* and *NANOG* promoter regions in 293T cells as well as H1 ESCs (Figure 4H). Interestingly, co- and reverse immunoprecipitation followed by western blotting (Figures 4I and 4J) revealed that the somatic cell-specific MBD2a isoform exhibits much higher affinity for interaction with members of the transcriptionally repressive NuRD complex, including HDAC1, HDAC2, RbAp46, MTA2, and Mi-2 (Zhang et al., 1999). The specificity of the MBD2a-NURD interaction was further confirmed by probing for SIN3A, a corepressor (Zhang et al., 2005) independent of NuRD that did not biochemically interact with either MBD2 isoform (Figure 4I).

DISCUSSION

PSCs are phenotypically well defined but exhibit significant molecular heterogeneity (Cahan and Daley, 2013). These observations suggest that the core pluripotent factors OCT4, SOX2, and NANOG must balance a stochastic transcriptional ground state but yet respond rapidly to exogenous cues in order to properly orchestrate the cell lineages required for life, all from a relatively modest number of protein-coding genes (Wu et al., 2010). Alternative splicing represents a likely pathway whereby the core pluripotency factors can dynamically regulate proteome diversity to support high-fidelity lineage commitment (Wang et al., 2008). Although several examples of alternatively spliced gene products have been functionally validated in pluripotent cells (Das et al., 2011; Gabut et al., 2011; Han et al., 2013; Salomonis et al., 2010), a general framework that mechanistically links OCT4, NANOG, or SOX2 with specific splicing factors, pre-mRNA substrates, and canonical regulators of gene transcription has yet to be described.

We found that the splicing factor SFRS2 was strongly represented within the pluripotent molecular signature and, moreover, that OCT4 bound to *SFRS2* promoters in vivo and drove luciferase expression in vitro. These data establish interdependent genetic and functional links between *OCT4* and *SFRS2* in hPSCs. We confirmed a cell-type-specific expression pattern for *MBD2* isoforms, and found that SFRS2 biochemically targets the pre-mRNA of this methyl-DNA binding protein. We also observed a reciprocal link between *OCT4* and *MBD2a*, manifested at the level of gene expression and pluripotent phenotype. Interestingly, hESCs displayed distinct morphologies in response to depletion of *SRFR2* or overexpression of *MBD2a*, suggesting that the splicing factor most likely targets additional gene products; indeed, it is intriguing to speculate that the pool of pluripotent-specific, alternatively spliced transcripts in our exon-junction microarray data may be rich in previously unrecognized gene isoforms that support self-renewal. Similarly, use of next-generation DNA sequencing technologies may provide an exhaustive set of pluripotent-specific gene isoforms and splicing factor gene targets. Notwithstanding a comprehensive analysis of SFRS2 gene targets, our current results provide compelling mechanistic evidence that the functional role of OCT4 in pluripotent cells extends to the pathways that regulate gene splicing.

Although the editing of pre-mRNA transcripts can be reconstituted in vitro, it has become clear that gene splicing in vivo is intimately linked to transcription, chromatin structure, and histone modifications (Braunschweig et al., 2013; Lin et al., 2008). NuRD is a chromatin remodeling complex that is thought to promote lineage commitment of ESCs via silencing of pluripotency genes (Reynolds et al., 2012). Although previous work suggested that NuRD was recruited to methylated DNA by MBD2 (Zhang et al., 1999), we found that, while both MBD2a and MBD2c bound to the promoter regions of *OCT4* and *NANOG* in hPSCs, only the somatic cell-specific MBD2a isoform biochemically interacts with NuRD. Such isoform-specific recruitment of NuRD may enable pluripotent cells to rapidly regulate their transcriptional profiles in response to specific differentiation cues. Our finding was recently corroborated in murine ESCs along with data suggesting that the region of Mbd2a immediately C-terminal to the MBD domain, but absent in Mbd2c, mediates interaction with NuRD members (Baubec et al., 2013). Interestingly, two recent computational studies suggested that tissue-specific alternative splicing may mediate protein-protein interactions en masse in order to support distinct phenotypes (Buljan et al., 2012; Ellis et al., 2012). Our data provide a specific example that fits this model whereby the activity of a chromatin remodeling factor (NuRD) is mediated through interactions with protein isoforms (MBD2) expressed in a cell-type-specific manner. Our analysis further revealed that the pluripotent-specific MBD2c



(legend on next page)

isoform augmented reprogramming efficiency of somatic cells, whereas MBD2a had no effect. This observation is consistent with a strongly repressive role for endogenous MBD2a-NuRD complexes in somatic cells and potentially reconciles discrepancies reported for the role of MBD2 in pluripotent cells (Lee et al., 2013; Onder et al., 2012). Systematic titration of MBD2a levels in the context of enforced MBD2c expression in somatic cells may fully delineate the interplay of these isoforms and reveal whether MBD2a represents a key hurdle to reprogramming. The function of MBD family proteins and isoforms in NuRD and in reprogramming are most likely complex, as exemplified by a recent report demonstrating that depletion of the MBD3, also a component of NuRD, renders reprogramming deterministic and highly efficient (Rais et al., 2013). Defining the dynamics of these mutually exclusive MBD family-NuRD complexes (Le Guezennec et al., 2006) along with their regulatory target genes in hPSCs should shed further light on the mechanisms of somatic cell reprogramming.

Recent evidence suggests that the repressive activity of NuRD is opposed by signaling pathways that support expression of pluripotent factors, hence maintaining a stochastic ground state in which ESCs self-renew but are transcriptionally poised for lineage-specific differentiation (Hu and Wade, 2012). We found that several serine residues on SFRS2 were preferentially phosphorylated in pluripotent cells (Table S1). The questions of which signaling axis (e.g., AKT and SRPK) (Zhou et al., 2012) mediates phosphorylation on SFRS2 and whether this activity represents a general mechanism to reinforce expression of pluripotent-specific gene isoforms in hPSCs are worthy of future study.

Noncoding RNA has emerged as an important posttranscriptional regulatory pathway in pluripotent cells with functional links established between specific micro- or long noncoding RNAs (lncRNAs) and master regulatory transcription factors (Loewer et al., 2010; Marson et al., 2008). The alternatively spliced MBD2 isoforms harbor differences in both their 3′ UTR and protein-coding sequences. As a result, the somatic-cell-specific *MBD2a* isoform is targeted by miR-302 family members. The 3′ UTR of *MBD2c* does not appear to be subject to miR-mediated suppression, although we did observe modest regulation of *SFRS2* by miR-301 family members in vitro. These results are consistent with the notion that the microRNA machinery may act synergistically with splicing factors and gene isoforms to either enforce a self-renewing ground state or rapidly translate lineage commitment signals into appropriate transcriptional programs. Further analysis will be required in order to determine the full extent of miRNA-mediated regulation of proteome diversity and whether lncRNAs (Wang et al., 2013) or other noncoding sequences are also involved. Collectively, these data may allow a quantitative assessment of the network topology including the relative contribution of each node comprising a putative feedback loop linking the core pluripotent genes with the alternative splicing apparatus and specific gene isoforms.

In summary, we delineate genetic, biochemical, and functional links consistent with a general model (Figure 4K) in which the master regulators of pluripotency (e.g., OCT4) act in concert with splicing factors (e.g., SFRS2) and the miRNA machinery to mediate protein diversity via alternative splicing (e.g., MBD2), ultimately enforcing a pluripotent ground state.

EXPERIMENTAL PROCEDURES

Cell Culture

hPSCs for proteomics and phosphoproteomics were maintained in mTeSR media (STEMCELL Technologies) on six-well plates precoated with matrigel (BD Bioscience) as previously described (Park et al., 2008).

Chromatin Immunoprecipitation

ChIP was performed in H1 ESCs with anti-OCT4 (Santa Cruz Biotechnology).

Coimmunoprecipitation

Protein coimmunoprecipitation was performed in 293T cells with anti-FLAG gel (Sigma-Aldrich).

Genome-wide Expression and Alternative Splicing Data

Gene expression (Table S1) and alternative splicing profiling (Table S4) was performed with Affymetrix arrays.

Proteomic Data

Samples were processed (Ficarro et al., 2009) for protein expression and phosphorylation analyses by 3D reversed phase-strong anion exchange-reversed phase liquid chromatography tandem mass spectrometry (Ficarro et al., 2011). Native mass spectrometry data are available for download at http://blaispathways.dfci.harvard.edu/mz/.

Quantitative RT-PCR

For quantitative RT-PCR assays, relative gene-expression levels in BJ DFs or infected H1 ESCs were calculated on the basis of the internal standard gene *TBP* and normalized to those in either wild-type H1 ESCs or in H1 ESCs with virus infection control, respectively.

Figure 4. A General Model for Regulation of Proteome Diversity that Supports Self-Renewal in hPSCs

(A–D) Lentiviral-mediated expression of MBD2a but not MBD2c disrupts pluripotency in H1 ESCs. Expression of MBD2 isoforms as monitored by qRT-PCR (A) and western blotting (B). Pluripotency in H1 ESCs was assessed by colony morphology (C) and expression of *OCT4*, *NANOG*, and *SOX2* (D).

(E–G) Exogenous expression of *MBD2c*, but not *MBD2a*, enhances reprogramming efficiency in BJ DF. (E) Expression of *MBD2* isoforms in infected BJ DFs as monitored by qRT-PCR. Reprogramming efficiency was assessed by the number (F) and fold change (G) of TRA-1-60 colonies measured across biological replicates.

(H) Exogenous MBD2a and MBD2c independently bind to *OCT4* and *NANOG* promoter regions in H1 ESCs (top) and 293T cells (bottom).

(I and J) MBD2 interacts with the NuRD complex in an isoform-specific manner.

(I) Members of the NuRD transcription repressor complex (HDAC1, HDAC2, MTA2, Mi-2, and RbAp46) coimmunoprecipitate with exogenous FLAG-HA-MBD2a, but not FLAG-HA-MBD2c, in 293T cells. Neither MBD2 isoform interacts with the SIN3A-histone deacetylase complex.

(J) Coimmunoprecipitation of MTA2 in 293T cells overexpressing FLAG-HA-tagged MBD2 isoforms confirmed the preferential interaction between MBD2a and MTA2, a core NuRD complex member.

(K) Proposed model illustrating a putative positive feedback loop, in which the splicing factor SFRS2, along with miRNAs controlled by the core pluripotency genes, regulates the expression of MBD2 isoforms that either support (MBD2c) or oppose (MBD2a) expression of *OCT4*, *NANOG*, and *SOX2* through recruitment of the NuRD complex. Error bars represent the mean ± SEM.

RNA Immunoprecipitation

Relative occupancy values (fold enrichments) were calculated by determining the immunoprecipitation efficiency (ratios of the amount of immunoprecipitated RNA to that of the input sample) and normalized to the level observed by immunoprecipitation with nonspecific IgG, which was defined as 1.0.

Statistical Methods

The Student's t test was used to estimate statistical significance ($p < 0.05$).

ACCESSION NUMBERS

All microarray data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE55673.

SUPPLEMENTAL INFORMATION

Supplemental Information contains Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.04.002.

AUTHOR CONTRIBUTIONS

Y.L., Y.-H.L., H.L., M.C., C.J.L., J.J.C., G.Q.D., and J.A.M. designed research; Y.L., Y.-H.L., M.C., S.B.F., and C.-X.T. performed experiments; H.L., J.R.P., and N.S. performed bioinformatics analysis; S.T.A. derived the HFMSC-iPSC line; and Y.L. and J.A.M. coordinated the project and wrote the manuscript. All authors provided critical editing of the manuscript.

ACKNOWLEDGMENTS

We thank J. Sahalie for technical assistance, M.W. Lensch for teratoma interpretation, and G. Adelmant for critical reading of the manuscript. Y.-H.L. is supported by the A*STAR Investigator research award. H.L. is supported by Mayo Clinic Center for Individualized Medicine and Mayo Clinic Center for Regenerative Medicine. Generous support for this work was provided (to C.J.L.) by the Pathology Department of the Brigham and Women's Hospital and the National Blood Foundation. J.J.C. is supported by Howard Hughes Medical Institute, SysCODE (Systems-based Consortium for Organ Design & Engineering), and NIH grant RL1DE019021. Generous support for this work was provided (to J.A.M.) by the Dana-Farber Cancer Institute, the Harvard Stem Cell Institute (P01GM099117), and the NIH/NINDS P01NS047572. Y.L. acknowledges support from T32HL66987-11. G.Q.D. is supported by the Howard Hughes Medical Institute and NIGMS 5P50GM099117. G.Q.D., J.J.C., and J.A.M. acknowledge support from NIH/NHLBI 5RC2HL102815-02. G.Q.D. serves on the scientific advisory boards of iPierian, Inc., Verastem, Inc., and MPM Capital.

Received: August 27, 2013
Revised: January 28, 2014
Accepted: April 2, 2014
Published: May 8, 2014

REFERENCES

Balázsi, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. Cell *144*, 910–925.

Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraite, A., Mitchell, P.S., Nelson, A.M., et al. (2008). MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells *26*, 2496–2505.

Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480–492.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. Cell *152*, 1252–1269.

Brill, L.M., Xiong, W., Lee, K.-B., Ficarro, S.B., Crain, A., Xu, Y., Terskikh, A., Snyder, E.Y., and Ding, S. (2009). Phosphoproteomic analysis of human embryonic stem cells. Cell Stem Cell *5*, 204–213.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol. Cell *46*, 871–883.

Cahan, P., and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. Nat. Rev. Mol. Cell Biol. *14*, 357–368.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Das, S., Jena, S., and Levasseur, D.N. (2011). Alternative splicing produces Nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells. J. Biol. Chem. *286*, 42690–42703.

De Los Angeles, A., Loh, Y.-H., Tesar, P.J., and Daley, G.Q. (2012). Accessing naïve human pluripotency. Curr. Opin. Genet. Dev. *22*, 272–282.

Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., et al. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol. Cell *46*, 884–892.

Ficarro, S.B., Adelmant, G., Tomar, M.N., Zhang, Y., Cheng, V.J., and Marto, J.A. (2009). Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. Anal. Chem. *81*, 4566–4575.

Ficarro, S.B., Zhang, Y., Carrasco-Alfonso, M.J., Garg, B., Adelmant, G., Webber, J.T., Luckey, C.J., and Marto, J.A. (2011). Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. Mol. Cell. Proteomics *10*, 011064.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E.O., et al. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell *147*, 132–146.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149–1163.

Gore, A., Li, Z., Fung, H.-L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63–67.

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell *7*, 249–257.

Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241–245.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. *18*, 6538–6547.

Hu, G., and Wade, P.A. (2012). NuRD and pluripotency: a complex balancing act. Cell Stem Cell *10*, 497–503.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049–1061.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A.M., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol. Cell. Biol. *26*, 843–851.

Lee, M.R., Prasain, N., Chae, H.-D., Kim, Y.-J., Mantel, C., Yoder, M.C., and Broxmeyer, H.E. (2013). Epigenetic regulation of NANOG by miR-302 cluster-MBD2 completes induced pluripotent stem cell reprogramming. Stem Cells *31*, 666–681.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. *42*, 1113–1117.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., Probasco, M.D., Bailey, D.J., Swaney, D.L., Tervo, M.A., Bolin, J.M., Ruotti, V., et al. (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat. Methods *8*, 821–827.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S., et al. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. Proc. Natl. Acad. Sci. USA *107*, 10514–10519.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468–478.

Vidal, M., Cusick, M.E., and Barabási, A.-L. (2011). Interactome networks and human disease. Cell *144*, 986–998.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev. Cell *25*, 69–80.

Wu, J.Q., Habegger, L., Noisa, P., Szekely, A., Qiu, C., Hutchison, S., Raha, D., Egholm, M., Lin, H., Weissman, S., et al. (2010). Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. Proc. Natl. Acad. Sci. USA *107*, 5254–5259.

Xiao, R., Sun, Y., Ding, J.-H., Lin, S., Rose, D.W., Rosenfeld, M.G., Fu, X.-D., and Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. Mol. Cell. Biol. *27*, 5393–5402.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924–1935.

Zhang, Y., Fatima, N., and Dufau, M.L. (2005). Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. Mol. Cell. Biol. *25*, 7929–7939.

Zhou, Z., Qiu, J., Liu, W., Zhou, Y., Plocinik, R.M., Li, H., Hu, Q., Ghosh, G., Adams, J.A., Rosenfeld, M.G., and Fu, X.D. (2012). The Akt-SRPK-SR axis constitutes a major pathway in transducing EGF signaling to regulate alternative splicing in the nucleus. Mol. Cell *47*, 422–433.