THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Control of embryonic stem cells self-renewal and differentiation via coordinated splicing and translation of YY2

OPEN ACCESS

Control of embryonic stem cells self-renewal and differentiation via coordinated splicing and translation of YY2

Soroush Tahmasebi[1,2,3], Seyed Mehdi Jafarnejad[1,2,3], Ingrid S. Tam[2,3], Thomas Gonatopoulos-Pournatzis[4], Edna Matta-Camacho[2,3], Yoshinori Tsukumo[2,3], Akiko Yanagiya[2,3], Wencheng Li[5], Yaser Atlasi[6], Maxime Caron[7,8], Ulrich Braunschweig[4], Dana Pearl[2,3], Arkady Khoutorsky[2,3,9], Christos G. Gkogkas[10], Robert Nadon[7,8], Guillaume Bourque[7,8], Xiang-Jiao Yang[2,3,11], Bin Tian[5], Hendrik G. Stunnenberg[6], Yojiro Yamanaka[2,7], Benjamin J. Blencowe[4,12], Vincent Giguère[2,3,11] and Nahum Sonenberg[2,3,]*

[1] These authors contributed equally to this manuscript

[2] Goodman Cancer Research Center, McGill University, Montreal, QC H3A 1A3, Canada

[3] Department of Biochemistry, McGill University, Montreal, QC H3A 1A3, Canada

[4] Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada

[5] Department of Microbiology, Biochemistry, and Molecular Genetics, Rutgers New Jersey Medical School, Newark, NJ 07103, USA

[6] Radboud University, Faculty of Science, Department of Molecular Biology, Nijmegen, 6525GA, the Netherlands

[7] Department of Human Genetics, McGill University, Montréal, QC H3A 1A3, Canada

[8] McGill University and Genome Quebec Innovation Centre, Montreal, QC H3A 0G1, Canada

[9] Present address: Department of Anesthesia, McGill University, Montreal, QC, H3G 1Y6, Canada

[10] Patrick Wild Centre, Centre for Integrative Physiology, University of Edinburgh, Edinburgh, EH8 9XD, UK

[11] Department of Medicine, McGill University Health Center, Montreal, QC H3A 1A3, Canada

[12] Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada

[*] Correspondence: nahum.sonenberg@mcgill.ca

Summary

Translational control of gene expression plays a key role during early phases of embryonic development. Here we describe a novel regulator of mouse embryonic stem cells (mESCs), Yin-Yang 2 (YY2) that is controlled by the translation inhibitors, 4E-BPs. YY2 plays a critical role in regulating mESC functions through transcriptional control of key pluripotency factors, including *Oct4* and *Esrrb*. Importantly, overexpression of YY2 directs the differentiation of mESCs into cardiovascular lineages. We show that the splicing regulator PTBP1 promotes the retention of an intron in the 5′ UTR of *Yy2* mRNA, which confers sensitivity to 4E-BP-mediated translational suppression. Thus, we conclude that YY2 is a major regulator of mESC self-renewal and lineage commitment, and document a multilayer regulatory mechanism that controls its expression.

Statement of significance

Stringently controlled mRNA translation is critical for embryonic stem cells (ESCs), which in contrast to differentiated cells, rely on low translation rate to maintain stemness by unknown mechanisms. By employing ribosome profiling to unveil the eIF4E-sensitive mRNAs in ESCs, we identified a hitherto not studied ESC-related transcription factor, YY2, which controls self-renewal and differentiation of ESCs. While a basal level of YY2 is essential for ESC self-renewal, increased YY2 expression directs differentiation of mESCs toward cardiovascular lineages. By comprehensive examination of the Yy2 5'UTR, we delineated a novel, multilayer regulatory mechanism in mESCs by which YY2 expression is dictated by the combined actions of the splicing regulator, PTBP1, and the translation inhibitors, 4E-BPs. ChIP-Seq analysis revealed that YY2 directly controls the expression of several pluripotency and development-related genes. This is the first study that describes a synchronized network of alternative splicing, and mRNA translation in controlling self-renewal and differentiation.

Introduction

Stringent control of mRNA translation is critical during early embryonic development, as relatively small changes in expression of developmentally related genes dramatically affect self-renewal and differentiation of stem cells. In fact, a modest ($\leq$ 2 fold) increase or decrease in OCT4 or SOX2 protein levels impairs ESC self-renewal and triggers differentiation (Kopp et al., 2008; Niwa et al., 2000). mRNA translation, which is low in undifferentiated ESCs and multipotent somatic stem cells (e.g. HSCs and skin stem cell), increases significantly during differentiation (Blanco et al., 2016; Sampath et al., 2008; Signer et al., 2014). Importantly, genome-wide analysis of the transcriptome vs. proteome of ESCs during early stages of differentiation demonstrated that protein levels do not correlate with mRNA levels (Pearson's R<0.4), underscoring the importance of post-transcriptional regulation in ESC differentiation (Lu et al., 2009).

mRNA translation can be divided into three steps; initiation, elongation, and termination. Translational control has been documented most extensively at the initiation step, at which ribosomes are recruited to the mRNA by the concerted action of eukaryotic translation initiation factors (eIFs) (Sonenberg and Hinnebusch, 2009). Control of translation is exerted mainly by two key protein complexes: eIF4F (eIF4E-eIF4G-eIF4A) and the ternary complex (eIF2–GTP–Met-tRNA$^{Met}_i$) (Sonenberg and Hinnebusch, 2009). The mammalian target of rapamycin complex 1 (mTORC1) controls the assembly of eIF4F through phosphorylation of eIF4E-binding proteins (4E-BPs) (Beretta et al., 1996; Pelletier et al., 2015). The 4E-BPs consist of a family of small molecular weight translational inhibitors (15-20 kDa; 4E-BP1, 2 and 3 in mammals), that when dephosphorylated, avidly bind eIF4E and block its association with eIF4G to form the

eIF4F complex. Following phosphorylation by mTORC1, 4E-BPs dissociate from eIF4E, allowing eIF4F complex formation and activation of translation (Beretta et al., 1996; Brunn et al., 1997; Hara et al., 1997). 4E-BPs inhibit cap-dependent translation in embryonic and somatic stem cells (Hartman et al., 2013; Sampath et al., 2008; Signer et al., 2014). While eIF4E promotes cap-dependent translation of all cellular mRNAs, the translation of a subset of mRNAs, which generally contains long and highly structured 5′ untranslated region (5′ UTR), is strongly dependent on eIF4E (Koromilas et al., 1992; Pelletier et al., 2015). These mRNAs are known as "eIF4E-sensitive", and encode proteins that control fundamental cellular processes, such as cell proliferation and survival (Bhat et al., 2015).

We showed that 4E-BPs are required for reprogramming of mouse embryonic fibroblasts (MEFs) to induced pluripotent stem cells (iPSCs) (Tahmasebi et al., 2014). In the current study, we describe a tightly coordinated network in mESCs whereby the expression of YY2 transcription factor is controlled via the splicing regulator PTBP1 and the translation inhibitors, 4E-BPs. Our data reveal that tight regulation of YY2 expression by this network is critical for mESC self-renewal and lineage commitment.

Results

Transcriptome and translatome profiling of WT and 4E-BP1/2-null mESCs

To investigate the role of 4E-BPs in mESCs, we first derived mESCs from wild-type (WT) and *Eif4ebp1* and *Eif4ebp2* double-knockout (DKO) mice and examined the eIF4F complex using m^7GTP-agarose pull-down assay. eIF4F amount was elevated, as

demonstrated by increased (~14-fold) eIF4G1 pull-down in DKO mESCs (Figure S1A). However, polysome profiling (Figure S1B) and [^{35}S]methionine/cysteine labeling assays (Figure S1C) did not detect a substantive difference in global mRNA translation between WT and DKO mESCs. These data are consistent with previous findings that the lack of 4E-BPs affects the translation of a subset of mRNAs, rather than global translation (Dowling et al., 2010; Tahmasebi et al., 2014).

To identify 4E-BP-sensitive mRNAs in mESCs, we performed ribosome profiling (Ingolia et al., 2009), which allows for precise measurement of the translation of mRNAs on a genome-wide scale, by deep sequencing of ribosome-protected fragments (ribosome footprints; RFPs). We achieved a high degree of reproducibility between the replicates for mRNA-Seq and footprints ($R^2 > 0.97$ Figure S1D). Metagene analysis confirmed the enrichment of RFP reads in coding sequences and the expected three-nucleotide periodicity (Figure S1E). These analyses validated the quality of the mRNA and RFP libraries (Supplemental Table 1). We used Babel analysis (Olshen et al., 2013) to compute changes in abundance of RFPs (Figure 1A), independent of changes in the levels of their corresponding mRNAs (Figure 1B). A significant enhancement in translation efficiency (TE) of a small subset of mRNAs was detected in DKO mESCs (FDR<0.1; Figure 1C and Supplemental Table 2), consistent with the lack of global change in translation in the DKO mESCs (Figure S1B and C). Strikingly, mRNA-Seq data revealed downregulation of mRNA levels for several pluripotency factors, such as *Prdm14*, *Eras*, *Esrrb*, and *Nanog* in DKO mESCs (Figure 1D and Supplemental Table 3). Possible reasons for this will be addressed in the Discussion.

Ablation of *Eif4ebp1* and *Eif4ebp2* results in reduced expression of mESCs markers

To validate the mRNA-Seq results, we examined the expression of pluripotency factors in undifferentiated WT and DKO mESCs, as well as after differentiation, by WB and RT-qPCR. While all DKO mESC lines maintained normal morphology under standard mESC culture conditions (in the presence of LIF and feeder layer [irradiated MEFs]), expression of the ESC marker NANOG was markedly reduced (Figure 1E). These changes were not due to unintended consequences of the knockout procedure, as RNAi-mediated depletion of 4E-BP1 and 2 (DKD; double knockdown) resulted in a similar reduction in NANOG (Figure S2A). DKO mESCs proliferated slower than WT mESCs (Figure S2B), and when cultured in the absence of feeder layer, they exhibited flattened morphology, which is indicative of cellular differentiation, while WT mESCs preserved their normal morphology (Figure S2C). Similar morphological changes were observed in DKD ESCs (Figure S2D). Notably, when cultured in the absence of LIF and feeder layer, NANOG and OCT4 expression was dramatically suppressed in DKO mESCs, while WT mESCs maintained higher expression of these proteins (Figure S2E). In addition to *Nanog*, mRNA levels of *Oct4* and *Sox2* were lower in DKO mESCs on day-0 of differentiation (Figure S2F). Moreover, mRNA levels, as determined by RT-qPCR of several other ESC-related factors, such as *Lin28a*, *Eras*, and *Tets* (*Tet1*, *2* and *3*), were reduced in DKO mESCs (Figure S2G). Therefore, 4E-BPs are required for normal expression of pluripotency factors. Analysis of embryoid bodies (EB) derived from WT and DKO ESCs revealed that the lack of 4E-BP1 and 4E-BP2 resulted in differentiation of mESCs toward mesodermal and endodermal lineages, as indicated by significant upregulation of

Bmp4, an early mesoderm marker, and *Gata4* and *Gata6*, early endoderm markers (Figure S2H). This coincides with the downregulation of the neuroectoderm markers, *Map2* and *Sox17* mRNAs, in DKO EBs.

The decrease in mRNA levels for several pluripotency factors, such as *Nanog*, *Eras*, and *Esrrb*, in 4E-BP DKO mESCs, as compared to WT mESCs (Figures 1D, S2F and S2G) (Supplemental Table 3), suggests that 4E-BP-dependent translational regulation of one or more factor(s) effects the changes in the transcriptome.

Stringent control of YY2 expression in mESCs

One of the mRNAs exhibiting the most significant increase in translation efficiency in DKO mESCs versus WT cells is the Yin-Yang 2 (*Yy2*) transcription factor (Babel p-value = 0.0001; Figure 1C and Supplemental Table 2). YY2 exhibits considerable sequence homology (55% identity) with the YY1 transcription factor (Nguyen et al., 2004). Similar to Rex1 (Zfp42), a well-known ESC marker (Kim et al., 2007; Rogers et al., 1991), YY2 is a retroposed copy of the YY1 gene, which evolved only in placental mammals (Luo et al., 2006). YY1 is a pleiotropic transcription factor that regulates diverse cellular processes and plays a critical role in early embryonic development, ESC biology, and reprogramming (Donohoe et al., 1999; Onder et al., 2012; Vella et al., 2012). Although their N-terminal domains differ significantly, the C-terminal DNA binding domain of YY1 is highly conserved in Rex1 and YY2 (Kim et al., 2007). This raises the possibility that YY2 plays an important role in regulation of gene expression in mESCs. While there is no significant change in the level of *Yy2* mRNA in DKO mESCs as compared to WT cells (Figure S3A), YY2 protein, but not YY1, is elevated in DKO mESCs (Figure 1F).

Importantly, expression of a phosphorylation-resistant 4E-BP1-4A mutant (Thoreen et al., 2012) in DKO mESCs reduced YY2 protein levels (Figure 1G). These data demonstrate that the translation of *Yy2* mRNA is controlled by 4E-BPs.

To further study the functional consequence of YY2 upregulation in mESCs, we generated a mESC line carrying a doxycycline-inducible YY2 construct (dox-YY2). Consistent with the DKO mESC data (Figure 1D and S2E-G), overexpression of YY2 caused a reduction in expression of pluripotency factors, *Nanog*, *c-Myc*, and *Oct4* mRNAs (Figure 2A and B), indicating a negative role for YY2 in mESC self-renewal. However, constitutively expressed shRNA against *Yy2* in mESCs caused a dramatic depletion of mESCs in culture (Figure S3B). This coincided with increased levels (8.4 ± 2.4 fold) of the apoptosis marker, cleaved Caspase-3, in *Yy2* knockdown cells (Figure S3C) suggesting that mESCs require a basal level of YY2 expression for survival. To study the impact of a moderate knockdown, we generated an mESC line carrying doxycycline-inducible shRNA construct against *Yy2* (shYy2). While slight induction of shYy2 (0.2 µg/ml doxycycline) enhanced expression of the pluripotency factors *Nanog* and *Sox2* (Figure 2C), higher doses of doxycycline failed to do so (Figure 2C), indicating a dose-sensitive effect of YY2 on mESC pluripotency. Importantly, the deleterious effect of complete depletion of YY2 was not limited to mESCs, as CRISPR/Cas9-mediated *Yy2* knockout blastocysts were unable to maintain their inner cell mass, as demonstrated by blastocyst outgrowth assays (Figures 2D, S3D and E). A similar defective outgrowth has been previously described for $Yy1^{-/-}$ blastocysts (Donohoe et al., 2007), suggesting that lack of YY2 may cause peri-implantation lethality and demonstrating that YY1 and YY2 fulfill non-redundant functions in blastocysts growth.

Recent studies showed that YY1 directly regulates Nkx2.5 expression and promote cardiogenesis uncovered a novel function for YY1 in cardiomyocyte differentiation and cardiac morphogenesis (Beketaev et al., 2015; Gregoire et al., 2013). To examine the effect of YY2 on mESC differentiation toward cardiovascular lineages, embryoid bodies (EBs) derived from dox-YY2 ESCs were exposed to doxycycline. One week after induction of YY2 expression, foci of beating cardiomyocytes began to appear in the plates with the highest level of YY2 induction (0.2 μg/ml doxycycline) (Supplemental Video 1), and the number of foci continued to increase in the following 2 weeks. No beating foci appeared in non-induced embryoid bodies up to the 3rd week of differentiation. Consistently, expression of several cardiovascular-specific markers, such as: *Nkx2.5*, *Bnp*, *αMHC*, *MLC2α*, and *MLC2v* mRNAs was increased in the YY2-overexpressing embryoid bodies in a dose-dependent manner (Figure 2E). These data suggest that YY1 and YY2 have overlapping function in directing differentiation of mESC toward cardiovascular lineages.

YY2 binds to regulatory regions of key genes for ESC pluripotency and differentiation

Genomic targets of YY1, but not YY2, in mESCs have been documented (Sigova et al., 2015; Vella et al., 2012). We determined the genome-wide binding sites of YY2 in mESCs by ChIP-Seq in cells overexpressing YY2. Due to the high degree of similarity between the C-terminal domains of YY1 and YY2, we used a monoclonal antibody that specifically recognizes the N-terminal domain of YY2 (Figures 2A and S3F, see Supplementary Materials and Methods). YY2 binding sites (Supplemental Table 4)

exhibit enrichment for the genomic loci of coding genes (exons or introns, symbolized by Gene, 43%; Figure 3A) and a preference for the promoter regions surrounding the transcription start sites (Figure 3B). Nearly half of the peaks contained the consensus YY1-binding motif (Figure 3C), which is consistent with similar sequence preference for YY1 and YY2 (Kim et al., 2007; Nguyen et al., 2004). Motif distribution across binding peaks revealed enrichment for the known consensus YY1 binding motif directly at YY2 peak centers (Figure 3D), indicating specific recognition of these binding sites by YY2.

Pathway analysis of genes associated with YY2 binding peaks by Ingenuity Pathway Analysis program (IPA) revealed significant enrichment for genes related to embryonic stem cell pluripotency (Figure 3E), such as *Oct4*, *Tdgf1, Esrrb,* and *FoxD3*. We also found enrichment for genes involved in the activation of the retinoic acid receptor (RAR)-signaling pathway. Previous studies showed that activation of the RAR pathway promotes differentiation of ESC to cardiomyocytes, particularly MLC2v$^+$ ventricular cardiomyocytes, and that the RAR-signaling pathway plays a critical role in cardiogenesis (Pan and Baker, 2007; Wobus et al., 1997). These findings indicate that activation of this pathway, along with other cardiovascular-related YY2 targets (e.g. *Bnp*, *Mesp2* and *Mkl1*), is responsible for engendering differentiation of mESCs toward cardiovascular lineage by YY2 (Figure 2E). We validated the ChIP-Seq results for a selected number of genes with ChIP-qPCR (Figures 3F, 3G and S3G).

RT-qPCR analysis of a selected number of pluripotency factors in the ChIP-Seq dataset, such as: *Oct4*, *Esrrb, Tet1*, and *Tet2* mRNAs, revealed that YY2 overexpression suppresses their expression in mESCs (Figure 4A). The ChIP-Seq analysis did not identify *Nanog* as a target of YY2. Thus, the decrease in *Nanog* mRNA expression in

DKO mESC (Figure 1D and E) or upon overexpression of YY2 (Figure 2A and B) is likely secondary to downregulation of other pluripotency factors. In agreement with our previous observation (Figure 2C), doxycycline-inducible *Yy2* knockdown in mESCs revealed a dose-sensitive regulation of its target genes (Figure 4B).

We analyzed the target genes identified by the YY2 ChIP-Seq assay relative to those of YY1 in mESCs (Sigova et al., 2015). 27.8% of YY2 targets are shared with YY1 (Figure 4C and Supplementary Table 4). Conversly, a large portion of the YY2 targets (~72%) are not present among the YY1 targets (Figure 4C), which supports a previous genome-wide mRNA expression study in HeLa cells that demonstrated that YY1 and YY2 regulate some shared, but mostly unique sets of genes (Chen et al., 2010). While some pluripotency-related genes, such as *Tdgf1,* are unique YY2 targets, *Oct4*, *Klf5,* and *Foxd3* are common targets between YY1 and YY2 (Figure 4C). Notably, *Yy1* is a target of both YY1 and YY2 and our data showed that YY2 has a dose-sensitive effect on YY1 expression (Figure 2A and C). The similar consensus-binding motifs of YY1 and YY2 are consistent with these factors exhibiting overlapping or competing effects on common target genes (Klar, 2010; Klar and Bode, 2005; Lee et al., 2015; Nguyen et al., 2004). One plausible explanation for their distinct activities is the considerable divergence in their N-terminal domains (Figure S4A). A recent report showed that binding of YY1 to active promoters/enhancers in ESCs through its C-terminal DNA-binding domain is facilitated by simultaneous binding of its N-terminal domain to RNA species transcribed from these regulatory elements (Sigova et al., 2015). We hypothesized that the N-terminal domain of YY2 cannot interact with RNA. We examined this possibility by Electrophoretic Mobility Shift Assay (EMSA) using purified mouse YY1 and YY2

proteins (Figure S4B and C), and DNA and RNA probes derived from the *ARID1A* promoter, which interact with YY1 (Sigova et al., 2015). Notably, *Arid1a* is among the YY2 targets in our ChIP-Seq analysis (Supplemental Table 4). While YY2 interacts with the DNA probe with only slightly less efficiency than YY1 (Figure 4D), only YY1 binds to the RNA probe (Figure 4E), as no visible binding was detected for YY2 to the *Arid1a* promoter-derived RNA, even after prolonged exposure (Figure S4D). In agreement with these results, analysis of the N-terminal domains of mouse YY1 and YY2 proteins by BindN, an RNA-binding prediction server (Wang and Brown, 2006), identified two distinct RNA-binding motifs for YY1, which are not conserved in YY2 (Figure S4E). Our data suggest a novel mechanism by which YY2 may act differently from YY1, whereby differential affinity for the promoter-derived RNAs may underlie the opposing effects of YY1 and YY2 on certain shared promoters.

A retained 5′ UTR intron renders *Yy2* mRNA sensitive to 4E-BP-dependent translational repression

To search for eIF4E-sensitive elements in the 5′ UTR of *Yy2* mRNA, we first used 5′ RACE (5′-rapid amplification of cDNA ends) to annotate the sequence in mESCs. In addition to the annotated mRNA sequence of 217 nt 5′ UTR (RefSeq: NM_001098723.1), henceforth referred to as variant A, we uncovered two additional variants, possessing 290 nt and 100 nt long 5′ UTRs (Figure 5A and B). While the 290 nt variant, designated as variant B, is a 5' extension of variant A, the 100 nt variant is a spliced version of variant A, lacking 117 nt. The spliced region harbors all of the features of a canonical intron, including a GU dinucleotide at the 5′ splice site (SS), AG

dinucleotide at the 3′ SS, and a polypyrimidine tract and putative branch site A nucleotide within 20 nucleotides of the 3′ SS (Figure 5A). Considering that the intronic sequence is shared between variants A and B, we termed their corresponding spliced variants ΔA and ΔB, respectively (Figure 5B). Retroposed genes are generally intronless, as they are generated through the reverse transcription of mature mRNAs (Fablet et al., 2009; Vinckenbosch et al., 2006). Therefore, the intron acquisition by *Yy2* is most likely a recent evolutionary event that occurred after the retroposition of *Yy2* from *Yy1* in placental mammals.

To measure intron retention (IR) in *Yy2* mRNA during differentiation of mESC by RT-PCR, we used two forward (Fw1 and Fw2) primers and a common reverse primer (Rv) to target the flanking exons (Supplemental Table 6). Primers Fw1 and Rv only detect variants B and ΔB (297 bp and 180 bp PCR products, respectively), whereas Fw2 and Rv amplify a 209 bp PCR product for variants A and B, and a 92 bp PCR product for variants ΔA and ΔB. We examined Yy2 IR events in mESCs and EBs at day 4 and 6 post-differentiation and measured the degree of IR using the metric of percent intron retention (PIR) (Braunschweig et al., 2014). There is a higher percentage of non-spliced variants (A and B) in mESCs than in EBs, demonstrating that differentiation coincided with a marked reduction in intron retention (Figures 5C and S5A). Notably, this alternative splicing event is not restricted to mESCs, as various degrees of intron retention were detected at different embryonic stages and across different tissues (Figures 5D and S5B), demonstrating developmental and tissue-specific regulation of *Yy2* alternative splicing.

To identify the *trans*-acting factor(s) responsible for *Yy2* alternative splicing, we

used the RBPmap web server to predict consensus motifs for RNA-binding proteins (Paz et al., 2014). We identified two canonical PTBP (polypyrimidine tract-binding protein) recognition motifs (CUCUCU) flanking the *Yy2* 5´ UTR intron (Figures 5A, 5B and S5C). PTBP1 and PTBP2 (a neural and testis enriched paralog) are RNA-binding proteins implicated in several aspects of mRNA metabolism, including: alternative splicing, stability, localization, and polyadenylation (Oberstrass et al., 2005). PTBP1 is essential for embryonic growth before gastrulation, as *Ptbp1*[-/-] mESCs have severe proliferation defects (Shibayama et al., 2009; Suckale et al., 2011). To explore its role in *Yy2* 5´ UTR intron retention, PTBP1 was depleted in mESCs using shRNA (Figure S5D). PTBP1 knockdown resulted in reduced *Yy2* intron retention, demonstrating that it acts as an inhibitor of *Yy2* splicing in mESCs (Figure 5E). This was associated with reduced growth and smaller colonies of mESCs (Figure S5E). To demonstrate that PTBP1 directly affects *Yy2* 5´ UTR splicing, we employed an *in vitro* splicing assay using a WERI retinoblastoma (WERI-Rb) cell extract, purified recombinant PTBP1 protein, and the three *Yy2* 5´ UTR variants (A, B and ΔA). Incubation with cell extract resulted in splicing of the introns from A and B variants, with no effect on the ΔA variant (Figure S5F and G). Importantly, addition of recombinant PTBP1 protein (Figure S5H) dramatically suppressed splicing and resulted in the complete retention of the intron in a dose-dependent manner (Figures 5F and S5I).

The length and complexity of the 5´ UTR play a critical role in mRNA translation, as mRNAs with long and structured 5´ UTRs are more sensitive to eIF4E activity (Colina et al., 2008; Koromilas et al., 1992). The combined use of alternative transcription start sites and alternative splicing determine the complexity of the *Yy2* 5´

UTR. To examine the effect of 5′ UTR variants on mRNA translation, we constructed luciferase reporters containing the 5′ UTRs of the A, B and ΔA variants (Figure S5J), which were used to generate mRNAs that were transfected into WT and DKO mESCs. The B, and to a lesser extent A, variant mRNAs were poorly translated, while the intron-less ΔA variant translation was dramatically better (>6 compared with control) in WT mESCs (Figure 5G, left panel). These data demonstrate that the 5′ UTR containing the optional intron sequence inhibits translation. The translation of A and B variant mRNAs was significantly elevated in the DKO mESCs, whereas the ΔA variant remained insensitive to 4E-BPs levels (Figure 5G, right panel). These results demonstrate that intron retention in *Yy2* 5′ UTR, in combination with 4E-BPs activity, determine the outcome of *Yy2* mRNA translation. Intron retention adds an extra 117 nt to the *Yy2* 5′ UTR, which increases its secondary structure complexity (Figure S5K), rendering it sensitive to 4E-BP-mediated translation repression. This double-layered control mechanism, consisting of the retention of the *Yy2* 5′ UTR intron by PTBP1 and suppression of translation of the resulting mRNA variant by 4E-BPs, allows for the adjustment of *Yy2* mRNA translation.

Consistent with the increased translation of the *Yy2* spliced variant (Figure 5G), the highest expression of YY2 protein was detected in the heart and muscle (Figure S5L) tissues, in which the lowest degree of *Yy2* 5′ UTR intron retention is observed (Figure 5D and S5B).

Discussion

Many mammalian mRNAs species contain heterogeneous 5'UTRs (Mele et al., 2015). Length, structure, and sequence elements in the 5'UTRs strongly impact

17

translation (de Klerk and t Hoen, 2015; Floor and Doudna, 2016). We found that the combination of alternative transcription start sites and splicing produces four *Yy2* mRNA variants with different translation efficiencies. Thus, the relative level of each variant, in combination with 4E-BP activity, dictates the rate of YY2 protein synthesis. ESC differentiation, which is concomitant with downregulation of PTBP1 expression (Linares et al., 2015), triggers the splicing of the *Yy2* 5′ UTR intron. However, there is considerable variation in the rate of *Yy2* 5′ UTR intron retention (IR) among adult mouse tissues, with heart and skeletal muscle displaying the lowest rate of IR. Different degrees of IR among mouse tissues, most of which express very low levels of PTBP1 (Figure S6), implies the existence of additional regulatory mechanism(s) that augments splicing of the *Yy2* 5′ UTR intron in heart and skeletal muscles. We also found consensus binding motifs for the MBNL1 splicing factor in the *Yy2* 5'UTR (Figure S5C). MBNL1 is highly expressed in cardiac and skeletal muscles (Miller et al., 2000) and is a known regulator of mRNA splicing in these tissues (Han et al., 2013; Kino et al., 2009). IR has emerged as a widespread mechanism to regulate gene expression in different cell and tissue types as well as during stem cell differentiation (Braunschweig et al., 2014). The finding that a retained intron at the 5′ UTR of *Yy2* mRNA controls its translation underscores the intricate interplay between alternative splicing and translational control (Maslon et al., 2014; Sterne-Weiler et al., 2013).

We established *Yy2* mRNA as a target of 4E-BP-dependent translation suppression in mESCs, and demonstrated that a basal level of YY2 expression is essential for mESC survival. Similar to *Yy1* knockout embryos, CRISPR-mediated *Yy2* knockout embryos survived the pre-implantation period, but the growth of their inner cell mass was

impaired, as revealed by blastocyst outgrowth assays (Figure 2D). These observations exclude the possibility of redundant functions for YY1 and YY2 in early stages of embryonic development. The critical function of YY2 in mESCs is most likely mediated by its direct transcriptional regulation of ESC-related genes, such as *Oct4*, *Eras*, *Tet1*, *Tet2*, and *Tdgf1*. We demonstrated that mESCs are highly sensitive to YY2, and a modest manipulation of *Yy2* amount significantly affects expression of pluripotency factors, ESC self-renewal, and differentiation (Figures 2A-C and 4A-B).

We documented that increased expression of YY2 directs differentiation of mouse embryoid bodies (mEBs) toward the cardiovascular lineage. The importance of YY2 in cardiomyocytes is likely not limited to early differentiation, as the heart expresses the highest level of YY2, as compared to other tissues of adult mouse (Figure S5L). This function is likely mediated by transcriptional control of the RAR pathway and multiple mesoderm/cardiomyocyte-related YY2 targets, such as *Bmp4*, *Nodal*, *Bnp*, *Mesp2*, and *Mkl1*, as evident by the ChIP-Seq analysis. Recent studies showed binding of YY1 to the promoters of highly expressed ribosomal proteins and the nuclear encoded mitochondrial membrane, enzymes, and ribosomal proteins (Chen et al., 2010), and highlighted the importance of YY1 in cardiomyocyte differentiation and heart morphogenesis (Beketaev et al., 2015; Gregoire et al., 2013). Notably, cardiac specific ablation of Yy1 causes severe abnormalities in the heart indicating that YY2 in the heart does not compensate for YY1 deletion (Xi et al., 2007). Comparison of YY2 and YY1 ChIP-Seq data (Figure 4C and Supplementary Table 4) demonstrated overlapping sets of nuclear encoded mitochondrial proteins, such as aminoacyl tRNA synthetase *Tars2*, and *Lars2* as well as mitochondrial ribosomal proteins *Mrps18c*, *Mrpl1*, *Mrpl44* and *Mrpl53* (Figure 4C and

Supplementary Table 4). Existence of unique as well as overlapping target genes among YY2 and YY1 targets explains the convergent, yet non-redundant, function of YY1 and YY2 in development.

In summary, we described a novel mechanism by which mESC self-renewal and lineage commitment is controlled via stringent regulation of YY2 expression at two stages; a, by repressing the splicing of the *Yy2* 5´ UTR intron via *Ptbp1*, and b, through suppressing the translation of the resulting mRNA variant by of *Eif4ebp1* and *2*. These two layers of control (Figure 6) coalesce to limit the expression of YY2 protein to a low basal level in mESCs, thereby maintaining their self-renewal and pluripotency.

Materials and Methods

ESC Cell Culture and differentiation

Mouse ESCs were maintained in Dulbecco's modified Eagle's medium (DMEM, Wisent Inc.), 1% nonessential amino acids (Gibco), 1% L-Glutamine (Wisent Inc.), 1% sodium pyruvate (100× stock from Invitrogen), 0.1 mM β-mercaptoethanol, 15% fetal bovine serum, 1,000 U mouse LIF/ml (ESGRO, Millipore), penicillin (50 μg/ml), and streptomycin (50 μg/ml) and expanded on the feeder layer or gelatin. For mESC differentiation (Wobus et al., 2002), 800-2000 mESCs were cultured for 2 days in hanging drops containing differentiation medium [DMEM, 20% fetal bovine serum, 1% nonessential amino acids (Gibco), 1% L-Glutamine (Wisent Inc.), penicillin (50 μg/ml) and streptomycin (50 μg/ml). The resulting embryoid bodies were transferred to bacteriological dishes and cultured in suspension for an additional 3 days. Next, they

were platted onto gelatin-coated tissue cultured plates for the rest of the differentiation process.

sgRNA synthesis

The DNA template for *Yy2* sgRNA was synthesized by PCR reactions using px330 (Addgene) as a template. Two primers were used: T7-*Yy2*-sgRNA forward (5'-TTAATACGACTCACTATAGGTTCGATGGTTTGGCCTACGCGTTTTAGAGCTAG AAATAGC-3') and sgRNA reverse (5'- AAAAGCACCGACTCGGTGCC -3'). The PCR product was purified with the PCR DNA fragments extraction kit (Geneaid), and was used as a template for sgRNA synthesis with the T7 MAXIscript kit (Ambion). The synthesized sgRNA was EtOH precipitated and dissolved in RNase free water.

Cytoplasmic microinjection, embryo culture and blastocyst outgrowth

50ng/µl Cas9 mRNA (Sigma) and 50ng/µl *Yy2*-sgRNA in 10 mM KCl was injected into CD1 zygotes in M2 media (Zenith Biotech). Cytoplasmic injection was performed with Femtojet (Eppendorf) and Cyto721 intracellular amplifier (WPI) for the Tickler's oscillation to penetrate the zygote's membrane. The injected zygotes were cultured for 4 days in KSOM drops covered with mineral oil (Zenith Biotech) in a 5% CO_2 incubator at 37^OC. The zona pellucidae of the developed blastocysts were removed with acid-tyrode's solution (Millipore). The blastocysts were plated on gelatin-coated 24-well plates and cultured 5-7 days in DMEM (Wisent) with ES-FBS (Wisent).

Cycloheximide treatment and hypotonic cell lysis

Cells were pretreated with cycloheximide (Bioshop Canada Cat#CYC003) (100 µg/ml) for 5 min, and lysed in hypotonic buffer; 5 mM Tris-HCl (pH 7.5), 2.5 mM MgCl2, 1.5 mM KCl, 1x protease inhibitor cocktail (EDTA-free), 100µg/ml cycloheximide, 2 mM DTT, 200 U/ml RNaseIn), 0.5% (v/w) Triton X-100, and 0.5% (v/w) Sodium Deoxycholate, to isolate the polysomes.

Polysome fractionation

250 µg of the polysomes were separated on a 10%–50% sucrose gradient by ultracentrifugation at 36,000 RPM for 2 h in an SW40 rotor (Beckman Coulter) at 4°C, and fractionated using an ISCO gradient fractionation system and optical density at 254 nm was continuously recorded with a FOXO JR Fractionator (Teledyne ISCO).

Collection of ribosome footprints (RFPs)

Ribosome profiling assay was performed as described (Ingolia et al., 2012), with minor modifications. Briefly, 500 µg of the RNPs (2 biological replicates) were subjected to ribosome footprinting by RNase I treatment at 4°C for 50 min with gentle mixing. Monosomes were pelleted by ultracentrifugation in a 34% sucrose cushion at 70,000 RPM for 3h and RNA fragments were extracted twice with acid phenol, once with chloroform, and precipitated with isopropanol in the presence of NaOAc and GlycoBlue. Purified RNA was resolved on a denaturing 15% polyacrylamide urea gel and the section corresponding to 28-32 nucleotides containing the RFPs was excised, eluted, and

precipitated by isopropanol.

Random RNA fragmentation and mRNA-Seq

150 μg of cytoplasmic RNA was used for mRNA-Seq analysis. Poly (A)+ mRNAs were purified using magnetic oligo-dT DynaBeads (Invitrogen) according to the manufacturer's instructions. Purified RNA was eluted from the beads and mixed with an equal volume of 2X alkaline fragmentation solution (2 mM EDTA, 10 mM Na2CO3, 90 mM NaHCO3, pH 9.2) and incubated for 20 min at 95°C. Fragmentation reactions were mixed with stop/precipitation solution (300 mM NaOAc pH 5.5 and GlycoBlue), followed by isopropanol precipitation. Fragmented mRNA was size-selected on a denaturing 10% polyacrylamide urea gel and the area corresponding to 35-50 nucleotides was excised, eluted, and precipitated with isopropanol.

Author Contributions

ST and SMJ conceived the project, designed and performed most experiments and interpreted the data with input from NS, YY, CG, AK, BT, XJY, HGS, YT, BJB and VG. IST and ST performed the ChIP assay. TGP performed the *in vitro* splicing assay and RT-PCR in mouse tissues. EMC performed the EMSA. AY performed the luciferase assay. WL, YA, MC, and UB performed the bioinformatics analyses. DP assisted with western blot experiments. SMJ, ST and NS wrote the manuscript with input from all authors.

Acknowledgments

<div align="center">References</div>

Beketaev, I., Zhang, Y., Kim, E.Y., Yu, W., Qian, L., and Wang, J. (2015). Critical role of YY1 in cardiac morphogenesis. Developmental dynamics : an official publication of the American Association of Anatomists *244*, 669-680.

Beretta, L., Gingras, A.C., Svitkin, Y.V., Hall, M.N., and Sonenberg, N. (1996). Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. Embo J *15*, 658-664.

Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J., and Topisirovic, I. (2015). Targeting the translation machinery in cancer. Nat Rev Drug Discov *14*, 261-278.

Blanco, S., Bandiera, R., Popis, M., Hussain, S., Lombard, P., Aleksic, J., Sajini, A., Tanna, H., Cortes-Garrido, R., Gkatza, N*., et al.* (2016). Stem cell function and stress response are controlled by protein synthesis. Nature *534*, 335-340.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome research *24*, 1774-1786.

Brunn, G.J., Hudson, C.C., Sekulic, A., Williams, J.M., Hosoi, H., Houghton, P.J., Lawrence, J.C., Jr., and Abraham, R.T. (1997). Phosphorylation of the translational repressor PHAS-I by the mammalian target of rapamycin. Science *277*, 99-101.

Chen, L., Shioda, T., Coser, K.R., Lynch, M.C., Yang, C., and Schmidt, E.V. (2010). Genome-wide analysis of YY2 versus YY1 target genes. Nucleic acids research *38*, 4011-4026.

Colina, R., Costa-Mattioli, M., Dowling, R.J., Jaramillo, M., Tai, L.H., Breitbach, C.J., Martineau, Y., Larsson, O., Rong, L., Svitkin, Y.V.*, et al.* (2008). Translational control of the innate immune response through IRF-7. Nature *452*, 323-328.

de Klerk, E., and t Hoen, P.A. (2015). Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends in genetics : TIG *31*, 128-139.

Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Molecular cell *25*, 43-56.

Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., and Shi, Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Molecular and cellular biology *19*, 7237-7244.

Dowling, R.J., Topisirovic, I., Alain, T., Bidinosti, M., Fonseca, B.D., Petroulakis, E., Wang, X., Larsson, O., Selvaraj, A., Liu, Y.*, et al.* (2010). mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. Science *328*, 1172-1176.

Fablet, M., Bueno, M., Potrzebowski, L., and Kaessmann, H. (2009). Evolutionary origin and functions of retrogene introns. Molecular biology and evolution *26*, 2147-2156.

Floor, S.N., and Doudna, J.A. (2016). Tunable protein synthesis by transcript isoforms in human cells. eLife *5*.

Gregoire, S., Karra, R., Passer, D., Deutsch, M.A., Krane, M., Feistritzer, R., Sturzu, A., Domian, I., Saga, Y., and Wu, S.M. (2013). Essential and unexpected role of Yin Yang 1 to promote mesodermal cardiac differentiation. Circulation research *112*, 900-910.

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.

Hara, K., Yonezawa, K., Kozlowski, M.T., Sugimoto, T., Andrabi, K., Weng, Q.P., Kasuga, M., Nishimoto, I., and Avruch, J. (1997). Regulation of eIF-4E BP1 phosphorylation by mTOR. The Journal of biological chemistry *272*, 26457-26463.

Hartman, N.W., Lin, T.V., Zhang, L., Paquelet, G.E., Feliciano, D.M., and Bordey, A. (2013). mTORC1 targets the translational repressor 4E-BP2, but not S6 kinase 1/2, to regulate neural stem cell self-renewal in vivo. Cell reports *5*, 433-444.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Kim, J.D., Faulk, C., and Kim, J. (2007). Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. Nucleic acids research *35*, 3442-3452.

Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., and Ishiura, S. (2009). MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. Nucleic acids research *37*, 6477-6490.

Klar, M. (2010). It is not necessarily YY1--the frequently forgotten Yin-Yang-2 transcription factor. Proceedings of the National Academy of Sciences of the United States of America *107*, E190; author reply E191.

Klar, M., and Bode, J. (2005). Enhanceosome formation over the beta interferon promoter underlies a remote-control mechanism mediated by YY1 and YY2. Molecular and cellular biology *25*, 10159-10170.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells *26*, 903-911.

Koromilas, A.E., Lazaris-Karatzas, A., and Sonenberg, N. (1992). mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. Embo J *11*, 4153-4158.

Lee, S.H., Cho, Y.C., Jeong, H.M., Kim, K.H., Choi, H.J., Lee, K.Y., and Kang, B.Y. (2015). Yin-Yang 1 and Yin-Yang 2 exert opposing effects on the promoter activity of interleukin 4. Archives of pharmacal research.

Linares, A.J., Lin, C.H., Damianov, A., Adams, K.L., Novitch, B.G., and Black, D.L. (2015). The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. eLife *4*.

Lu, R., Markowetz, F., Unwin, R.D., Leek, J.T., Airoldi, E.M., MacArthur, B.D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L.A*., et al.* (2009). Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature *462*, 358-362.

Luo, C., Lu, X., Stubbs, L., and Kim, J. (2006). Rapid evolution of a recently retroposed transcription factor YY2 in mammalian genomes. Genomics *87*, 348-355.

Maslon, M.M., Heras, S.R., Bellora, N., Eyras, E., and Caceres, J.F. (2014). The translational landscape of the splicing factor SRSF1 and its role in mitosis. eLife, e02028.

Mele, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J*., et al.* (2015). Human genomics. The human transcriptome across tissues and individuals. Science *348*, 660-665.

Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A., and Swanson, M.S. (2000). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. Embo J *19*, 4439-4448.

Nguyen, N., Zhang, X., Olashaw, N., and Seto, E. (2004). Molecular cloning and functional characterization of the transcription factor YY2. The Journal of biological chemistry *279*, 25927-25934.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L*., et al.* (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. Science *309*, 2054-2057.

Olshen, A.B., Hsieh, A.C., Stumpf, C.R., Olshen, R.A., Ruggero, D., and Taylor, B.S. (2013). Assessing gene-level translational control from ribosome profiling. Bioinformatics *29*, 2995-3002.

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.

Pan, J., and Baker, K.M. (2007). Retinoic acid and the heart. Vitamins and hormones *75*, 257-283.

Paz, I., Kosti, I., Ares, M., Jr., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic acids research *42*, W361-367.

Pelletier, J., Graff, J., Ruggero, D., and Sonenberg, N. (2015). Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. Cancer Res *75*, 250-263.

Rogers, M.B., Hosler, B.A., and Gudas, L.J. (1991). Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. Development *113*, 815-824.

Sampath, P., Pritchard, D.K., Pabon, L., Reinecke, H., Schwartz, S.M., Morris, D.R., and Murry, C.E. (2008). A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. Cell Stem Cell *2*, 448-460.

Shibayama, M., Ohno, S., Osaka, T., Sakamoto, R., Tokunaga, A., Nakatake, Y., Sato, M., and Yoshida, N. (2009). Polypyrimidine tract-binding protein is essential for early mouse development and embryonic stem cell proliferation. The FEBS journal *276*, 6658-6668.

Signer, R.A., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature *509*, 49-54.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978-981.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M.A., Pourmand, N., and Sanford, J.R. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. Genome research *23*, 1615-1623.

Suckale, J., Wendling, O., Masjkur, J., Jager, M., Munster, C., Anastassiadis, K., Stewart, A.F., and Solimena, M. (2011). PTBP1 is required for embryonic development before gastrulation. PloS one *6*, e16992.

Tahmasebi, S., Alain, T., Rajasekhar, V.K., Zhang, J.P., Prager-Khoutorsky, M., Khoutorsky, A., Dogan, Y., Gkogkas, C.G., Petroulakis, E., Sylvestre, A.*, et al.* (2014). Multifaceted regulation of somatic cell reprogramming by mRNA translational control. Cell stem cell *14*, 606-616.

Thomsen, S., Anders, S., Janga, S.C., Huber, W., and Alonso, C.R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome Biol *11*, R93.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109-113.

Vella, P., Barozzi, I., Cuomo, A., Bonaldi, T., and Pasini, D. (2012). Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. Nucleic acids research *40*, 3403-3418.

Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed gene copies in the human genome. Proceedings of the National Academy of Sciences of the United States of America *103*, 3220-3225.

Wang, L., and Brown, S.J. (2006). BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic acids research *34*, W243-248.

Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C., *et al.* (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle *8*, 498-504.

Wobus, A.M., Kaomei, G., Shan, J., Wellner, M.C., Rohwedel, J., Ji, G., Fleischmann, B., Katus, H.A., Hescheler, J., and Franz, W.M. (1997). Retinoic acid accelerates embryonic stem cell-derived cardiac differentiation and enhances development of ventricular cardiomyocytes. Journal of molecular and cellular cardiology *29*, 1525-1539.

Xi, H., Yu, Y., Fu, Y., Foley, J., Halees, A., and Weng, Z. (2007). Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. Genome research *17*, 798-806.

Figure Legends

Figure 1. Lack of 4E-BPs deregulates expression of pluripotency factors in mESCs.

Log2 abundance of (**A**) ribosome footprints (RFP) and (**B**) mRNA abundance (RNA-Seq) is plotted for wild-type (WT) and *Eif4ebp1* and *Eif4ebp2* double-knockout (DKO) mESCs. (**C**) Babel analysis of transcripts with significant change in ribosome footprints independent of the corresponding change in mRNA abundance (black dots; FDR<0.1). *Triml2* and *Trmt61b*, respectively, are mRNAs with the highest and the lowest RFP ratio in DKO as compared to WT mESCs. (**D**) Heat-map representation of log2 mRNA expression levels of selected pluripotency factors in DKO over WT mESCs, identified by mRNA-Seq analysis. (**E**) Western blot analysis of NANOG expression in a WT and two independent DKO mESC clones (1 and 2). Numbers indicate the ratio of NANOG expression in each DKO clone compared with WT after normalization with ß-actin. (**F**) WB analysis of YY2 and YY1 expression in a WT and two independent DKO mESC

28

clones. Numbers indicate the ratio of NANOG expression in each DKO clone compared with WT after normalization with α-tubulin. (**G**) DKO mESCs carrying doxycycline-inducible 4E-BP1-4A mutant construct were treated with 0, 0.005, 0.01, 0.05, 0.1 or 0.2 µg/ml doxycycline for 24 h and subjected to WB analysis.

Figure 2. Stringent regulation of YY2 levels is critical for mESC survival and differentiation. (**A**) WB and (**B**) RT-PCR analysis of WT mESCs carrying doxycycline-inducible YY2 construct and treated with 0, 0.2, 1 or 4 µg/ml doxycycline for 24 h. (**C**) RT-qPCR analysis of Yy2 and pluripotency factors expression in mESCs transduced with a doxycycline-inducible shYy2 lentivirus. Resistant colonies were selected with puromycin (5 µg/ml) and incubated with increasing amount of doxycycline (0, 0.2, 1 and 4 µg/ml per day) for 72 h. Values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3). $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns= non-significant. (**D**) Blastocyst-outgrowth assay in a WT and 2 independent CRISPR/CAS9-mediated *Yy2* knockout embryos. Cas9 mRNA and sgRNAs targeting *Yy2* were injected into zygotes. The blastocysts derived from injected embryos were subjected to blastocyst-outgrowth assay. The mutagenesis strategy and the sequence of mutant alleles are provided in figure S3D and E. (**E**) RT-PCR analysis of embryoid bodies carrying doxycycline-inducible YY2 construct, treated with 0, 0.002, 0.02 or 0.2 µg/ml doxycycline every other day for 4 weeks.

Figure 3. YY2 controls ESC transcriptional regulatory network and developmentally related genes. (**A**) Pie-chart displaying the distribution of YY2 ChIP-

Seq peaks across the genome based on the distance of the peaks to the nearest RefSeq gene: Proximal (<2 kb upstream of Transcription Start Site; TSS), Gene (exon or intron), Distal (2 - 10 kb upstream of TSS), 5 d (10 - 100 kb upstream of TSS), Gene desert (>100 kb from a RefSeq gene) and Other (anything not included in the above categories). (**B**) Histogram depicting the distance of YY2 ChIP-Seq peaks relative to the TSS of the nearest gene. (**C**) *De novo* motifs enriched in YY2-binding events. Enrichment *p*-values and percentage of targets containing each motif are displayed, as generated by HOMER software. (**D**) Plots showing average density of selected motifs in a 2 kb window from YY2 peak center. (**E**) Most significantly enriched canonical pathways in genes associated with YY2 ChIP-Seq peaks, as identified by Ingenuity Pathway Analysis. (**F**) Standard ChIP-qPCR validation of YY2 binding regions. Data are normalized to IgG. *Lmnb2* gene was used as a negative control. (**G**) Graphical representation of selected YY2 binding peaks, obtained from the UCSC browser. 20 kb windows are displayed.

Figure 4. Regulatory network and distinct mode of action of YY2 compared with YY1. (**A**) RT-qPCR analysis of selected YY2 targets in control and YY2-overexpressing mESCs. mESCs carrying doxycycline-inducible YY2 construct treated with 0 or 0.2 μg/ml doxycycline for 24 h. Values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3). $^{**}p < 0.01$, $^{***}p < 0.001$, ns= non-significant. (**B**) RT-qPCR analysis of YY2 targets in mESCs transduced with a doxycycline-inducible sh*Yy2* lentivirus as described in Figure 2C. Values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns= non-significant. (**C**) Comparison of YY2 and YY1 ChIP-Seq targets in mESCs. Only peaks with at least 1 nt overlap were considered as common

targets. (**D**) Electrophoretic mobility shift assay (EMSA) with radioactive labeled double-stranded DNA oligonucleotide probe derived from the promoter region of the mouse *Arid1a* gene (Sigova et. al. 2015) and purified recombinant mouse 3xF-YY1 and 3xF-YY2 proteins. The probes were incubated in the presence of increasing amounts of recombinant proteins in the presence or absence of antibodies as indicated. (**E**) EMSA with radioactive labeled single-stranded RNA oligonucleotide probe derived from the promoter region of the mouse *Arid1a* gene (Sigova et. al. 2015) and purified recombinant mouse 3xF-YY1 and 3xF-YY2 proteins.

Figure 5. Retention of *Yy2* 5'UTR intron renders it sensitive to 4E-BP-mediated translation suppression. (**A**) Sequence of the promoter region of mouse *Yy2* gene. The 2 alternative transcription start sites (TSS) are marked by arrowheads, the boxed sequence shows the retained intron, the two highlighted hexamers are the consensus PTBP-binding motifs, and the underlined ATG is the translation start codon for *Yy2* mRNA. (**B**) A cartoon depicting the 4 variants of *Yy2* 5'UTR. vB and vΔB represent long variant with and without intron retention, respectively; vA and vΔA represent short variant with and without intron retention, respectively. CD: Coding DNA sequence (**C**) RT–PCR using primer pairs designed to recognize all 4 possible variants (Fw2 and Rv) to estimate the splicing efficiency of the 5' UTR intron in mESC and mouse embryoid body (mEB) in days 4 and 6 post-differentiation. (PIR: Percentage of Intron Retention). *Gapdh* mRNA was used as control. (**D**) RT-PCR analysis of intron retention (IR) in *Yy2* 5'UTR in different mouse embryonic stages and adult tissues using primers described in **C.** *Gapdh* mRNA was used as control. (**E**) RT-PCR analysis of IR in *Yy2* 5'UTR using primers

described in **C** upon depletion of PTBP1 expression in mESCs by two independent shRNAs. *Yy2*-ORF primers amplifying a segment of the coding region of *Yy2* transcript were used to demonstrate the change in overall expression of *Yy2* mRNA. *β-actin* mRNA was used as internal control, (**F**) RT-PCR amplification (primers Fw2 and Rv) of the *in vitro* splicing products of the A, B and ΔA variants in WERI retinoblastoma cell extracts with different amounts of recombinant PTBP1 protein. Recombinant BSA was used as a negative control. (**G**) Luciferase reporter assay with Firefly (Fluc) and *Renilla* (Rluc) luciferase reporter mRNAs, as described in Fig. S5J. The *in vitro*-transcribed mRNAs were purified and transfected into WT and DKO mESCs. The left panel depicts the normalized luciferase activity of each construct in WT mESCs. The right panel compares the luciferase activity of each construct in DKO versus WT mESCs. *Fluc* mRNA was co-transfected with *Rluc* mRNA as a transfection control. Data are mean ± s.d. (n=3). $^{**}p <$ 0.01, $^{***}p < 0.001$, ns= non-significant.

Figure 6. Regulation of YY2 expression by splicing and mRNA translational control.
Proposed model depicting the regulation of YY2 expression at two different steps: 1-alternative splicing (via PTBP1) and 2-mRNA translation (via 4E-BP1/2). Basal level of YY2 expression is required for maintenance of mESC self-renewal, whereas increased translation of *Yy2* mRNA directs cardiovascular lineage commitment.

Supplementary Figure Legends

Supplementary Figure 1. Translation control in WT and 4E-BP1/2-null mESCs. (**A**) Lysates from WT and *Eif4ebp1* and *Eif4ebp2* double-knockout (DKO) mESCs were subjected to m^{7}GTP pull-downs and analyzed for indicated proteins. (**B**) Polysome profiles of WT and DKO mESCs treated with 100 μg/ml cycloheximide for 5 min. Absorbance light was set at 254 nm. (**C**) [^{35}S] methionine/cysteine incorporation into newly synthesized proteins from WT and DKO mESCs grown in 10% dialyzed fetal bovine serum and pulsed for 30 minutes with [^{35}S] methionine/cysteine. Data are mean ± s.d. (*n*=3). (**D**) Correlation between replicates in mRNA-Seq and ribosome profiling datasets. R^2 indicates Pearson correlation. (**E**) Metagene analysis of randomly fragmented mRNAs and ribosome footprints (RFPs) in mESCs. Normalized read counts are averaged across the entire transcriptome and aligned at the annotated start codons and stop codons.

Supplementary Figure 2. Characterization of 4E-BP1/2-null mESCs. (**A**) Western blot analysis of mESCs transduced by lentivirus expressing shRNAs for 4E-BP1 and 4E-BP2 (sh4EBP1/2) or control (shCTR). SE: short exposure, LE: long exposure. (**B**) Cell growth assay for WT and *Eif4ebp1* and *Eif4ebp2* double-knockout (DKO) mESCs. Data are mean ± s.d. (*n*=3). $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. (**C**) WT and DKO mESCs cultured in the absence of feeder layer. (**D**) Control and 4E-BP1 and 2 double knockdown (DKD) mESCs cultured in the absence of feeder layer. (E) and (**F**) Time-course WB and RT-qPCR analysis of expression of selected pluripotency factors in WT and DKO mESCs cultured in the absence of feeder layer and LIF for the indicated time. RT-qPCR values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3). (**G**) RT-qPCR analysis of selected pluripotency markers in WT and DKO mESCs. Values are normalized to *ß-*

actin. Data are mean ± s.d. (*n*=3). $^{**}p < 0.01$, $^{***}p < 0.001$. (**H**) RT-qPCR analysis of WT and DKO embryoid bodies 2 weeks post-differentiation for expression of selected differentiation markers. Values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3). $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, ns= non-significant.

Supplementary Figure 3. Role of YY2 in mESCs and blastocysts formation. (**A**) RT-qPCR analysis of *Yy2* mRNA expression in WT and DKO mESCs. Values are normalized to *ß-actin*. Data are mean ± s.d. (*n*=3) ns= non-significant. (**B**) WT and DKO mESCs were transduced with 3 independent shRNAs against *Yy2* or control shRNA. Resistant colonies were selected with puromycin (5 µg/ml) for 2 days and subjected to alkaline-phosphatase staining. (**C**) WB analysis of mESC transduced with 3 independent shRNAs against *Yy2* or control shRNA. (**D**) Schematic presentation of CRISPR-Cas9 target site in mouse *Yy2* gene. The guide RNA was specifically designed to avoid regions with high homology to mouse *Yy1*. PAM; the protospacer-adjacent motif. Numbers indicate the distance from start codon. (**E**) The Sequence of the *Yy2* alleles in the WT and 2 mutant male blastocysts depicted in figure 2D. (**F**) WB analysis of mESC overexpressing GFP or v5-YY1. (**G**) Graphical representation of selected YY2 binding peaks, obtained from the UCSC browser. 20 kb windows are displayed.

Supplementary Figure 4. YY1, but not YY2, binds RNA through its N-terminal domain. (**A**) Pairwise alignment of mouse YY1 and YY2 protein sequences using Geneious software. The protein motifs are clustered according to their similarity. (**B**)

Purification of triple Flag-tagged (3xF)-YY1 and 3xF-YY2 proteins from 293T cell lysate under non-denaturing conditions with anti-Flag M2 Agarose beads and elution with 3xFlag-peptide. (**C**) Coomassie-stained 10% SDS-PAGE gel of purified 3xF-YY1 and 3xF-YY2 proteins. (**D**) Prolonged exposure (3 weeks) of the EMSA with radioactive labeled single-stranded RNA oligonucleotide probe described in Figure 4E. (**E**) RNA binding prediction at N-terminal domains of YY1 and YY2 using BindN server (http://bioinfo.ggc.org/cgi-bin/bindn/bindn.pl). Predicted binding residues are labeled with "+" and in red; non-binding residues labeled with "-" and in green. Numbers denote "Confidence" from level 0 (lowest) to level 9 (highest).

Supplementary Figure 5. PTBP1 regulates *Yy2* 5'UTR intron retention. (**A**) RT–PCR analysis of mESCs, and day 4 (D4) and day 6 (D6) mouse embryoid bodies (mEB) using primer pairs designed to recognize the B and ΔB variants (Fw1 and Rv) to estimate the splicing efficiency of *Yy2* 5' UTR intron; PIR: Percentage of Intron Retention. (**B**) RT–PCR analysis (primers Fw1 and Rv) of intron retention in variant B of *Yy2* 5'UTR in different embryonic days or mouse tissues. (**C**) Consensus PTBP1 or MBNL1 binding motifs in 5'UTR of mouse *Yy2* mRNA, identified by RBPmap web server. (**D**) WB analysis of PTBP1 expression in mESCs stably expressing two independent shRNAs against *Ptbp1* or control shRNA (shGFP); related to Figure 5E and S5E. (**E**) Representative images of mESC colonies described in **D**. (**F**) and (**G**) RT-PCR analysis of the *in vitro* splicing products of the A, B and ΔA variants in presence (+) or absence (-) of the WERI retinoblastoma cell extract. Primers Fw2 and Rv were used in **F** to recognize all 4 possible variants and primers Fw1 and Rv were used in **G** to recognize B

and ΔB variants. (**H**) Coomassie-stained 10% SDS-PAGE to detect the level of recombinant PTBP1 protein used for the *in vitro* splicing assay in Figures 5F and S5I. 1 μg of BSA was loaded as control. (**I**) RT- PCR analysis (primer Fw1 and Rv) of the *in vitro* splicing products in WERI retinoblastoma cell extracts with different amounts of recombinant PTBP1 protein. Recombinant BSA was used as a negative control. (**J**) Schematic diagrams of the *Renilla* (*Rluc*) and Firefly (*Fluc*) luciferase reporter mRNA constructs used in Figure 5G. (**K**) Effect of intron retention on sequence/structure of *Yy2* 5' UTR. The mfold web server (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form) was used to predict the secondary mRNA structures and ΔG. Nucleotides are numbered starting from the relevant transcription start site. The intronic sequence is highlighted in red. (**L**) Expression of YY2, PTBP1, and PTBP2 proteins in different tissues of a 4-week old female mouse. SE: short exposure, LE: long exposure.

Supplementary Figure 6. Expression of *Ptbp1*, *Eif4ebp1*, and *Eif4ebp2* in differentiated tissues. Expression profile of *Ptbp1*, *Eif4ebp1*, *Eif4ebp2*, and *Sox2* in a cohort of embryonic and adult mouse samples.

Supplementary Tables

Supplementary Table 1: Number of uniquely mapped reads for each mRNA-Seq and footprint (RFP) samples (related to Figure 1).

Supplementary Table 2: Differentially translated mRNAs in DKO vs WT mESCs identified by ribosome profiling assay (related to Figure 1).

Supplementary Table 3: Differentially expressed mRNAs in DKO vs WT mESCs identified by mRNA-Seq assay (related to Figure 1).

Supplementary Table 4: YY2 ChIP-Seq target peaks in mESCs and the peaks with overlap in YY2 and YY1 ChIP-Seq assays (related to Figure 3).

Supplementary Table 5: List of a cohort of publically available RNA-Seq datasets (related to Figure S6).

Supplementary Table 6: List of primers used in this study.

Supplementary Items:

Video 1 (Related to Figure 2E): Beating foci in differentiated mESC after induction of YY2 expression (0.2 μg/ml doxycycline).

Supplementary Materials and Methods

List of Antibodies, shRNAs and reagents

Rabbit anti-PTBP1 (Cell Signaling Cat# 8776), rabbit anti-PTBP2 (abcam Cat# ab154853), rabbit anti-eIF4G1 (Cell Signaling Cat# 2469), rabbit anti-YY1 (N-terminal; Sigma Cat# SAB4200303), mouse anti-YY2 (Santa Cruz Cat# sc-377008), rabbit anti-4E-BP1 (Cell Signaling Cat# 9644), rabbit anti-4E-BP2 (Cell Signaling Cat# 2845), mouse anti-eIF4E (BD Biosciences Cat# 610270), mouse anti-Alpha Tubulin (Santa Cruz Cat# sc-23948), mouse anti-β-actin (Sigma Cat# A5441), rabbit anti-NANOG (Bethyl

Laboratories Cat# A300-397A), goat anti-OCT-3/4 (Santa Cruz Cat# sc-8628), goat anti-SOX2 (Y-17), rabbit anti-SOX2 (Cell Signaling Cat# 4195S), mouse anti-c-MYC (Santa Cruz Cat# sc-40), mouse anti-Flag (Sigma Cat# F3165), rabbit anti-Phospho-4E-BP1 Thr37/46 (Cell Signaling Cat#2855), rabbit anti-Phospho-RPS6 (Ser240/244) (Cell Signaling Cat#2215), mouse anti-ribosomal protein S6 (C-8), rabbit anti-Caspase-3 (Cell Signaling Cat#9665). The following shRNAs were obtained from Sigma Mission library: Anti-*Ptbp1* shRNA#1; TRCN0000295113, shRNA#2; TRCN0000295168. Anti-*Eif4ebp1* and *Eif4ebp2* shRNAs (TRCN0000075612 and TRCN0000075614, respectively).

Generation of inducible shRNA construct

The "all-in-one" system (Wiederschain et al., 2009) for the inducible expression of shRNA against mouse *Yy2* (shYy2) was generated by cloning the following oligos into Tet-pLKO-puro lentiviral vector:

mYy2shRNA-T:

CCGGGGCCAAACCATCGAAGTATATCTCGAGATATACTTCGATGGTTTGGCCT TTTT

mYy2shRNA-B:

AATTAAAAAGGCCAAACCATCGAAGTATATCTCGAGATATACTTCGATGGTT TGGCC

Lentiviral particles generated by this vector were used to create stable doxycycline inducible shRNA-expressing mESCs.

Lentivirus Production

293FT cells (Invitrogen) were cultured in DMEM/10% FBS medium containing 400 μg/ml neomycin (BioShop) per the manufacturer's instructions. Medium was replaced by antibiotic free medium at least 8 h prior to transfection with Lipofectamine 2000. 10 μg of expression plasmid, 6.5 μg of psPAX2 and 3.5 μg of pMD2.G packaging plasmids were used to transfect $8X10^6$ 293FT cells in a 10-cm dish. 24 h post-transfection, the medium was collected and replaced with fresh medium. Supernatant was collected for additional 2 days, pooled, and subjected to centrifugation in a SW32-Ti rotor at 25,000 RPM for 1.5 h. The viral pellet was resuspended in DMEM, 10% heat-inactivated FBS and rotated overnight at 4°C. The resulting concentrated virus solution was used to infect the cells directly in the presence of polybrene (6 μg/ml) or stored at -80°C.

Cap pull-down assay using m^{7}GTP-Sepharose

7-methyl-GTP (m^{7}GTP)-Sepharose 4B (Jena Biosciences) was incubated with WT or DKO mESC lysate (500 ug) in buffer containing 50 mM MOPS/KOH (pH 7.4), 100mM KCl, 0.02% NaN3 and 0.5 mM EDTA for 30 min at 4°C and washed five times (5 min each) with the same buffer and eluted with 0.2 mM m^{7}GTP for 15 min at 4°C. Eluted proteins were subjected to SDS-PAGE followed by western blotting using the indicated antibodies.

Analysis of global protein synthesis

Protein synthesis was measured by [^{35}S]methionine/cysteine metabolic labeling. Briefly, mESCs were seeded in a 24-well plate the day before the experiment. Cells were incubated in methionine/cysteine-free DMEM supplemented with 10% dialyzed FBS (GIBCO) and [^{35}S]methionine/cysteine labeling mixture (10 µCi/ml) at 37°C for 30 min, followed by lysis in Laemmli buffer. [^{35}S]methionine/cysteine incorporation was determined by trichloroacetic acid (TCA) precipitation followed by scintillation counting.

Library preparation and sequencing

Fragmented mRNAs and RFPs were dephosphorylated using T4 polynucleotide kinase (New England Biolabs). Denatured fragments were resuspended in 10 mM Tris (pH 7) and quantified using the Bio-Analyzer Small RNA assay (Agilent). A sample of 10 pmol of RNA was ligated to the 3′-adaptor with T4 RNA ligase 1 (New England Biolabs) for 2 h at 37°C. Reverse transcription was carried out using oNTI223 adapter (Illumina) and SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. Products were separated from the empty adaptor on a 10% polyacrylamide Tris/Borate/EDTA-urea (TBE-urea) gel and circularized by CircLigase (Epicentre). Ribosomal RNA amounts were reduced by subtractive hybridization using biotinylated rDNA complementary oligos (Ingolia et al., 2012). The mRNA and ribosome-footprint libraries were amplified by PCR (10 cycles) using indexed primers and quantified using the Agilent BioAnalyzer High-Sensitivity assay. DNA was then sequenced on the HiSeq

2000 platform with read length of 50 nucleotides (SR50) according to the manufacturer's instructions, with sequencing primer oNTI202 (5CGACAGGTTCAGAGTTCTACAGTCCGACGATC).

Analysis of Ribosome profiling data

Reads were trimmed of known adaptor sequences using the FASTX-toolkit. Clipped-reads shorter than 24 nucleotides were discarded and the remaining reads were aligned against mouse rRNA, mtRNA, and tRNA, and unaligned reads were then mapped to the mouse genome (mm9) using bowtie2 (local mode) (http://bowtie-bio.sourceforge.net/bowtie2). Uniquely mapped reads with Mapping Quality (MAPQ) score ≥10 were used for further analysis. mRNA translation efficiency within and between samples was quantified using the Babel analytical method (Olshen et al. 2013). Briefly, a regression-model based approach was used to measure the negative binomial distribution to model the abundance of both mRNA and ribosome footprints. Parametric bootstrap approach was used to test the null hypothesis that the ribosome footprint-counts are similar to mRNA counts. This generated a high-confidence list of genes, whose mRNA translation efficiency significantly deviated from global transcript population. We used a cut-off of 40-counts for mRNAs to be included in this analysis and Benjamini–Hochberg method was used to adjust the p-values for multiple comparisons. For metagene analysis of read distribution around start or stop codons, reads mapped to RefSeq transcripts were used. For a given region, only genes with at least 10 reads whose 5' end was within the region were used. The 5' end position of a read was used for

plotting. The read number was normalized within each gene first, and then across all genes. The final normalized read number is shown as read number per gene per nucleotide.

Analysis of differential mRNA expression

Differential RNA expression was determined using the mapped mRNA-counts and the DESeq method as described (Thomsen et al., 2010). DESeq model count data with negative binomial distribution and normalizes read counts based on library size. Genes were considered differentially expressed if they exhibited a fold change of >=1.5 and FDR<0.05 (based on Benjamini-Hochberg multiple testing adjustment).

5' rapid amplification of cDNA ends (5′ RACE) analysis

5′ RACE was performed with the SMARTer RACE 5′/3′ kit (Clontech, Cat # 634858). Briefly, 1 μg of total RNA extracted from mESCs was treated with DNase I (Fermentas) and cDNA was generated by the SMARTScribe Reverse Transcriptase (Clontech), according to the manufacturer's instructions. The resultant cDNA was used for PCR amplification using the mouse *Yy2* gene-specific forward primers (GSPs) (Suppl. Table 6) together with a common Universal Reverse Primer (UPM), provided by the manufacturer. PCR products were resolved by agarose gel electrophoresis and all visible bands were excised and digested by restriction enzymes followed by cloning into the PUC19 vector provided by the manufacturer and sequenced by Sanger sequencing.

Luciferase reporter with *in* vitro-transcribed mRNAs

PCR products encoding the specified 5'UTR variants of *Yy2* mRNA followed by *Renilla* luciferase coding sequence were generated, and inserted into pBluescript II KS (+) vector (Stratagene) between the SacI and XhoI sites. Resultant plasmids (*pBKS-Yy2-5'UTR-A*, *pBKS-Yy2-5'UTR-B,* and *pBKS-Yy2-5'UTR- ΔA*) were used as templates for production of the luciferase reporter mRNAs by *in vitro* transcription using the MAXIscript T7 *in vitro* Transcription kit (Ambion), according to the manufacturer's protocol in the presence of the anti-reverse cap analog. Poly(A) tail was added using the Poly(A) Tailing kit (Ambion). WT and DKO mESCs were seeded in a 24-well-plate and cultured overnight. The next day cells were transfected with the specified *Renilla* luciferase mRNAs containing *Yy2* 5'UTRs together with firefly luciferase mRNA, as a transfection control, using Lipofectamine 2000 (Invitrogen). Cells were lysed the next day and luciferase activity was determined using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

RT-PCR assays for detection of *Yy2* 5' UTR intron retention

Poly(A) mRNAs from the indicated mouse tissues or cells were monitored for *Yy2* intron retention by RT-PCR, as described previously (Barbosa-Morais NL, 2012). Briefly, 0.5 ng of poly(A)+ RNA was used for cDNA synthesis and amplification with the One-Step RT-PCR kit (Qiagen) according to the manufacturer's recommendations. The number of amplification cycles was 24 for *Gapdh* and 32 for *Yy2*. Reaction products were resolved

using a 2% ethidium bromide agarose gel and imaged using the Gel DocTM XR System (Bio-Rad). Gel densitometry was used to calculate the percent of intron retention (PIR), measured as the amount of intron retained isoform divided by the sum of spliced and retained isoforms.

In vitro splicing assays

In vitro splicing assays were performed as described (Gueroussov et al., 2015). Briefly, *in vitro* splicing assays (20 µl) contained 1.5 mM ATP, 5 mM creatine phosphate, 5 mM DTT, 3 mM MgCl2, 2.6% polyvinyl alcohol (PVA), 30 units of RiboLock RNase inhibitor (Fermentas), 10 ng of *Yy2* RNA splicing substrate, and 50-60 µg of Weri-Rb1 cell extract, in 12 µl of lysis buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, 10% glycerol, 1 mM DTT) with or without recombinant proteins (PTBP1 or BSA). Transcripts used for *in vitro* splicing were prepared using *Yy2* reporter constructs containing a T7 promoter. *In vitro* transcription was performed using the MEGAscript T7 Transcription Kit (Life technologies) according to the manufacturer's recommendations. Amount of recombinant protein used: PTBP1: 250, 500, 1000 ng and BSA: 1000 ng. Reactions were incubated for 1 h at 30°C. RNA was recovered using TRI Reagent and resuspended in 10 µl of nuclease-free water. Intron retention was monitored using 2 µl of the recovered RNA in RT-PCR reactions, as described above; "RT-PCR assays for detection of *Yy2* 5' UTR intron retention".

Chromatin immunoprecipitation (ChIP) and ChIP-Seq assays

Antibodies used for ChIP were anti-RNA Polymerase II (Millipore 05-623), anti-H3K4Me3 (Millipore 07-473), anti-YY2 (Santa Cruz sc-377008), and anti-mouse IgG (ThermoFisher Scientific 10400C). Antibodies were pre-bound overnight at 4°C to 20 µl Dynabeads Protein A (ThermoFisher Scientific 10008D) or 20 µl Dynabeads Protein G (ThermoFisher Scientific 10004D), diluted in ChIP dilution buffer (1% Triton, 10 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA). For each ChIP assay, chromatin prepared from approximately 1×10^7 cells was diluted to 1 ml in ChIP dilution buffer and added to the antibody-bound beads. Chromatin was immunoprecipitated overnight with rotation at 4°C. Beads were washed 5 times at 4°C with wash buffer (50 mM HEPES-KOH pH 7.5, 1 mM EDTA, 0.7% Sodium Deoxycholate (NaDOC), 1% NP-40, 0.5 M LiCl) and once with TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). DNA was eluted in 200 µl elution buffer (1% SDS, 50 mM Tris-HCl pH 8, 10 mM EDTA) and crosslinks were reversed overnight at 65°C, followed by treatment with RNase A (Qiagen 19101) and Proteinase K (Roche 03115879001). DNA was purified with the Qiagen PCR Purification Kit (Qiagen 28106). Quantification of ChIP enrichment was carried out with SYBR-green-based qRT-PCR (Roche 04887352001) and the Roche LightCycler instrument. All primers used in this study are listed in Suppl. Table 6.

Sequencing libraries of ChIP DNA and corresponding 10% chromatin inputs were prepared at the McGill University and Génome Québec Innovation Centre, following quality assessment with the Agilent Bioanalyzer 2100. ChIP libraries were sequenced on the Illumina HiSeq 2000 with 50 base-pair reads. Reads from biological replicates were combined.

Bioinformatics analyses of the ChIP-Seq results

Sequence trimming and quality filtering was performed with Trimmomatic software v.0.32 ([Bolger et al., 2014](#)), retaining reads with phred33 score $\geq$ 30 and length $\geq$ 50. Filtered sequences were aligned to the mm10 assembly with bwa v.0.7.10. Peaks were called with the Model-based Analysis of ChIP-Seq (MACS) algorithm v.2.0 ([Zhang et al., 2008](#)) with p-value 1×10^{-5} using input reads as control. Peaks were annotated based on the nearest TSS with HOMER software v.4.7 ([Heinz et al., 2010](#)). Bedgraph files for ChIP-Seq visualization was performed with HOMER software and uploaded to the UCSC Genome Browser ([Kent et al., 2002](#)). For motif enrichment analysis *de novo* motif finding was performed with HOMER software v.4.7 using default parameters. Annotation of motif occurrences on ChIP-Seq peaks was performed in a 4 kb window from the center of the peak using HOMER software. For Pathway enrichment analysis, gene lists were compiled using binding peaks occurring within 20 kb of the TSS of the nearest gene. Canonical pathway enrichment analysis was performed on these lists with Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, ingenuity.com). P-values to identify significantly enriched biological pathways were calculated using Fisher's exact test.

Expression and purification of YY1 and YY2 proteins

Yy1 and *Yy2* ORFs were amplified from mESC cDNA library using the specific primers described in Supplementary Table 6 and cloned into the pCDNA5.A-3xF plasmid. HEK293H cells (kind gift from Dr. Marcotrigiano) were transfected with the indicated plasmid and cells were collected 30 h post-transfection and resuspended in lysis buffer

(25 mM HEPES-KOH 7.5, 150 mM NaCl, 0.5% NP40, 5% glycerol, 15 µM ZnCl$_2$, 0.5 mM DTT, 1 µg/ml RNase A/DNAse I and protease inhibitors) for 10 min on ice. After centrifugation, the Flag-tagged proteins were affinity purified using anti-Flag M2 Agarose beads (Sigma), and eluted with 5x bead volumes of 3xFlag-peptide 0.1mg/mL. SYBR gold and Coomassie staining were used to measure protein purification and ensure the lack of RNA/DNA impurities. Purified proteins were buffer exchanged to 50 mM HEPES-KOH 7.5, 100 mM NaCl, 15 µM ZnCl$_2$, 0.5 mM DTT buffer and concentrated to 1mM using a 10kDa Amicon (Millipore). Aliquots in 10% glycerol were frozen in liquid nitrogen at -80°C for EMSA experiments.

Electrophoretic mobility shifts assay (EMSA)

Oligonucleotide DNA probe for *Arid1a* (adopted from Sigova et al,) was generated by annealing 30 nt single stranded oligonucleotides (Fw: 5'-CTCTTCTCTCTTAAAATGGCTGCCTGTCTG-3', Rv: 5'-CAGACAGGCAGCCATTTTAAGAGAGAAGAG-3') to obtain 100 µM stock of double stranded DNA probe. 10 pmol were labeled with T4 polynucleotide kinase (PNK, Thermo Scientific) and 20 mCi [γ-32P] ATP (Perkin Elmer) for 30 min at 37°C. PNK was inactivated at 68°C for 10 min and unincorporated label was removed by centrifugation through a spin column (TE-10, Clonthech). The labeling incorporation was verified using 1 mL of labeled probe in scintillation liquid. Binding of YY1 and YY2 proteins to DNA probe was performed according to the Gel-Shift Assay Kit protocol (Promega). For each gel shift reaction 3000 cpm of labeled probe was incubated with

increasing concentrations of 3xFlag-YY2 or 3xFlag-YY1 proteins (10, 100, 300 nM) during 45 min on ice, then 1mL of Flag antibody (Sigma) was added to one sample containing 300 nM protein and all samples were incubated for additional 60 min on ice. Free DNA-probe and DNA-protein complexes were separated on a 6% DNA retardation gel (Novex, Thermo Scientific) at 250V for 25 min in 0.5xTBE at 4°C. Gels were vacuum-dried and exposed to an Amersham Hyperfilm (GE Healthcare Life Science) with a screen at -80°C. The 30 nt single stranded RNA probe for Arid1 (5'-rCrUrCrUrUrCrUrCrUrCrUrUrArArArArUrGrGrCrUrGrCrCrUrGrUrCrUrG-3') (adopted from Sigova et al,) was labeled and purified in the same way as the DNA probe. EMSAs with the recombinant 3xFlag-YY2 or 3xFlag-YY1 proteins were performed in buffer containing 20 mM HEPES-KOH, pH 7.5, 100 mM KCl, 2 mM MgCl$_2$, 15 µM ZnCl$_2$, 0.5 mM DTT, 0.05% NP40, 5% glycerol, 0.1mg/ml acetylated BSA (Sigma) and SUPERase-In RNase inhibitor (Life Technologies). Increasing amounts of 3xFlag-YY2 or 3xFlag-YY1 proteins (10, 100, 300, 600 nM) were incubated for 45 min on ice and applied to a 6% retardation gel (Novex, Thermo Scientific) at 250V for 25 min in 0.5xTBE at 4°C. The gels were vacuum-dried and exposed to an Amersham Hyperfilm (GE Healthcare Life Science) with a screen at -80°C.

PCR genotyping

Genomic DNA from the blastocysts was extracted and amplified using primers listed in Supplementary Table 6. PCR fragments were then sub-cloned into pGEM-T Easy vector for Sanger sequencing. At least 5 clones/outgrowth were sequenced.

Experimental Procedure References

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc *7*, 1534-1550.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Wobus, A.M., Guan, K., Yang, H.T., and Boheler, K.R. (2002). Embryonic stem cells as a model to study cardiac, skeletal muscle, and vascular smooth muscle cell differentiation. Methods Mol Biol *185*, 127-156.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.