LETTER

doi:10.1038/nature12270

MBNL proteins repress ES-cell-specific alternative splicing and reprogramming

Hong Han[1,2]*, Manuel Irimia[1]*, P. Joel Ross[3], Hoon-Ki Sung[4], Babak Alipanahi[5], Laurent David[6], Azadeh Golipour[2,6], Mathieu Gabut[1], Iacovos P. Michael[4], Emil N. Nachman[1,2], Eric Wang[7], Dan Trcka[6], Tadeo Thompson[3], Dave O'Hanlon[1], Valentina Slobodeniuc[1], Nuno L. Barbosa-Morais[1,8], Christopher B. Burge[7], Jason Moffat[1,2], Brendan J. Frey[1,5], Andras Nagy[4,9], James Ellis[2,3], Jeffrey L. Wrana[2,6] & Benjamin J. Blencowe[1,2]

Previous investigations of the core gene regulatory circuitry that controls the pluripotency of embryonic stem (ES) cells have largely focused on the roles of transcription, chromatin and non-coding RNA regulators[1–3]. Alternative splicing represents a widely acting mode of gene regulation[4–8], yet its role in regulating ES-cell pluripotency and differentiation is poorly understood. Here we identify the muscleblind-like RNA binding proteins, MBNL1 and MBNL2, as conserved and direct negative regulators of a large program of cassette exon alternative splicing events that are differentially regulated between ES cells and other cell types. Knockdown of MBNL proteins in differentiated cells causes switching to an ES-cell-like alternative splicing pattern for approximately half of these events, whereas overexpression of MBNL proteins in ES cells promotes differentiated-cell-like alternative splicing patterns. Among the MBNL-regulated events is an ES-cell-specific alternative splicing switch in the forkhead family transcription factor FOXP1 that controls pluripotency[9]. Consistent with a central and negative regulatory role for MBNL proteins in pluripotency, their knockdown significantly enhances the expression of key pluripotency genes and the formation of induced pluripotent stem cells during somatic cell reprogramming.

A core set of transcription factors that includes OCT4 (also called POU5F1), NANOG and SOX2, together with specific microRNAs and long non-coding RNAs, control the expression of genes required for the establishment and maintenance of ES-cell pluripotency[1–3,10–12]. Alternative splicing, the process by which splice sites in primary transcripts are differentially selected to produce structurally and functionally distinct messenger RNA and protein isoforms, provides a powerful additional mechanism with which to control cell fate[7,8,13], yet its role in the regulation of pluripotency has only recently begun to emerge. In particular, the inclusion of a highly conserved ES-cell-specific 'switch' exon in the FOXP1 transcription factor changes its DNA binding specificity such that it stimulates the expression of pluripotency transcription factors, including OCT4 and NANOG, while repressing genes required for differentiation[9]. However, the *trans*-acting regulators of this and other alternative splicing events[14–16] implicated in ES-cell biology are not known. These factors are important to identify, as they may control regulatory cascades that direct cell fate, and likewise they may also control the efficiency and kinetics of somatic cell reprogramming.

To identify such factors, we used high-throughput RNA sequencing (RNA-seq) data to define human and mouse cassette alternative exons that are differentially spliced between ES cells and induced pluripotent stem cells (iPSCs), and diverse differentiated cells and tissues, referred to below as 'ES-cell-differential alternative splicing'. A splicing code analysis[17] was then performed to identify *cis*-elements that may promote or repress these exons. The RNA-seq data used to profile alternative splicing were also used to detect human and mouse splicing factor genes that are differentially expressed between ES cells/iPSCs and non-ES cells/tissues. By integrating these data sources, we sought to identify differentially expressed splicing regulators with defined binding sites that match *cis*-elements predicted by the code analysis to function in ES-cell-differential alternative splicing.

We identified 181 human and 103 mouse ES-cell-differential alternative splicing events, with comparable proportions of exons that are ≥25% more included or more skipped in ES cells versus the other profiled cells and tissues (Fig. 1a, Supplementary Figs 1a and 2, and Supplementary Tables 1 and 2). When comparing orthologous exons in both species, 25 of the human and mouse ES-cell-differential alternative splicing events overlapped ($P < 2.2 \times 10^{-16}$; hypergeometric test). The human and mouse ES-cell-differential alternative splicing events are significantly enriched in genes associated with the cytoskeleton (for example, *DST*, *ADD3*), plasma membrane (for example, *DNM2*, *ITGA6*) and kinase activity (for example, *CASK*, *MARK2* and *MAP2K7*) (Supplementary Table 3). They also include the aforementioned FOXP1 ES-cell-switch alternative splicing event, and previously unknown alternative splicing events in other transcription or chromatin regulatory factor genes (for example, *TEAD1* and *MTA1*) that have been implicated in controlling pluripotency[18,19]. These results suggest a considerably more extensive role for regulated alternative splicing in ES-cell biology than previously appreciated.

The splicing code analysis revealed that motifs corresponding to consensus binding sites of the conserved MBNL proteins are the most strongly associated with ES-cell-differential alternative splicing in human and mouse. The presence of MBNL motifs in downstream flanking intronic sequences is associated with exon skipping in ES cells, whereas their presence in upstream flanking intronic sequences is associated with exon inclusion in ES cells (Fig. 1b, human code; Supplementary Fig. 1b, mouse code). To a lesser extent, features resembling binding sites for other splicing regulators, including polypyrimidine tract binding protein (PTBP) and RNA-binding fox (RBFOX) proteins, may also be associated with ES-cell-differential alternative splicing.

From RNA-seq expression profiling of 221 known or putative splicing factors, 11 genes showed significant differential expression between human ES cells/iPSCs and other cells and tissues (Bonferroni-corrected $P < 0.05$, Wilcoxon rank-sum test) (Fig. 1c and Supplementary Table 4). Notably, *MBNL1* and *MBNL2* had the lowest relative mRNA levels in ES cells/iPSCs compared to other cells and tissues

[1]Banting and Best Department of Medical Research and Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Developmental and Stem Cell Biology, The Hospital for Sick Children, 101 College Street, Toronto, Ontario M5G 1L7, Canada. [4]Center for Stem Cells and Tissue Engineering, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [5]Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada. [6]Center for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [7]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [8]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal. [9]Department of Obstetrics and Gynecology, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.
*These authors contributed equally to this work.

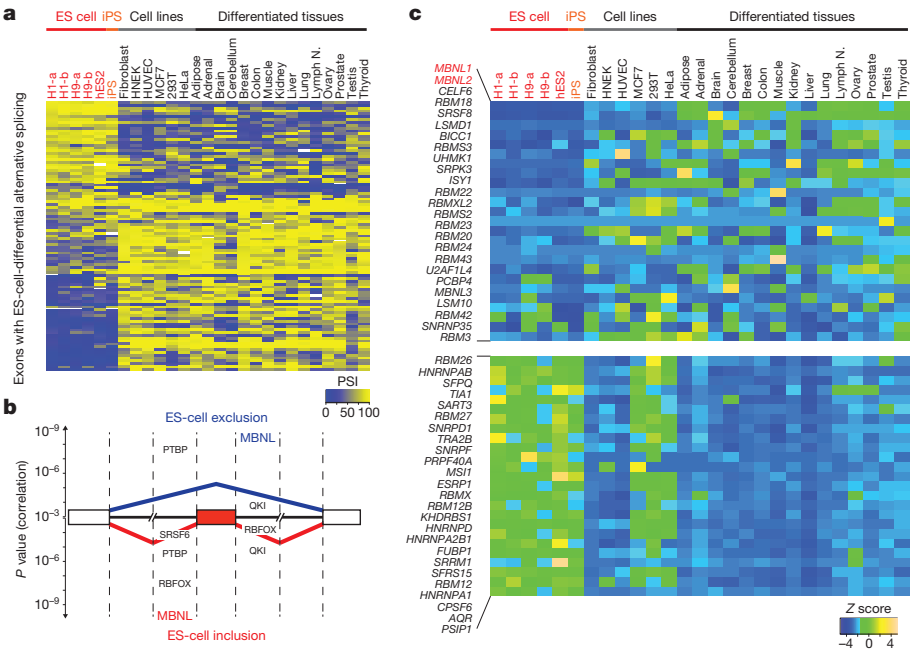

Figure 1 | Identification of regulators of ES-cell-differential alternative splicing. **a**, Heat map of per cent spliced in (PSI) values for 95 representative ES-cell-differential alternative splicing events in transcripts that are widely expressed across human ES cells/iPSCs, non-ES-cell lines and differentiated tissues. **b**, Splicing code features that are significantly associated with ES-cell-differential alternative splicing. Features are ranked according to Pearson correlation *P* values (*y* axis) for alternative exons with either lower (top) or higher (bottom) inclusion in ES cells/iPSCs. Dashed lines indicate 300 nucleotide intervals from splice sites. **c**, Heat map of *Z* scores of mRNA expression (cRPKM) levels for splicing factors. Twenty-five splicing factors with the lowest or highest relative mRNA expression levels in ES cells/iPSCs compared to non-ES cells/tissues are shown. cRPKM, corrected reads per kilobase exon model per million reads[30].

(Fig. 1c, Supplementary Fig. 3a and Methods). Quantitative RT–PCR (polymerase chain reaction with reverse transcription; qRT–PCR) assays confirmed this observation (Supplementary Fig. 3b). Similar results were obtained when analysing mouse expression data (Supplementary Fig. 3c–e and Supplementary Table 4). PTBP, RBFOX and other splicing factors potentially associated with ES-cell-differential alternative splicing by the splicing code analysis did not exhibit significant differences in mRNA levels between ES cells/iPSCs and other cells or tissues. Collectively, these results suggest a conserved and prominent role for MBNL1 and MBNL2 in ES-cell-differential alternative splicing.

Because MBNL proteins are expressed at minimal levels in ES cells compared to other cell types, we proposed that they may repress ES-cell-differential exons in non-ES cells, and/or activate the inclusion of exons in non-ES cells that are skipped in ES cells. Indeed, previous studies have shown that in differentiated cells, MBNL proteins suppress exon inclusion when they bind upstream flanking intronic sequences, and they promote inclusion when binding to downstream flanking intronic sequences[20,21]. The results of the splicing code analysis are consistent with this mode of regulation, when taking into account that MBNL proteins are depleted in ES cells relative to differentiated cells and tissues (Fig. 1b and Supplementary Fig. 1b).

To test the above hypothesis, we used short interfering RNAs (siRNAs) to knock down MBNL1 and MBNL2 (to ~10% of their endogenous levels), individually or together, in human (293T and HeLa) and mouse (neuro2A (N2A)) cells (Fig. 2a and Supplementary Fig. 4a; see below). For comparison, knockdowns were performed in human (H9) and mouse (CGR8) ES cells. RT–PCR assays were used to monitor the ES-cell-switch exon of *FOXP1*/*Foxp1* (human exon 18b/mouse exon 16b), which is partially included in ES cells and fully skipped in differentiated cell types[9]. The splicing code analysis suggested that this exon is associated with conserved regulation by MBNL proteins, through possible direct disruption of splice-site recognition (Fig. 2b; see legend and below). Knockdown of MBNL2 in 293T or HeLa cells resulted in a <1% increase in *FOXP1* exon 18b inclusion, whereas

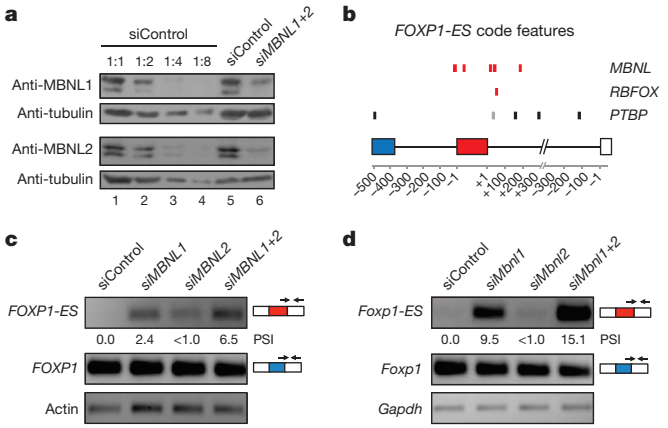

Figure 2 | MBNL proteins regulate ES-cell-specific alternative splicing. **a**, Western blots confirming efficient knockdown of MBNL1 and MBNL2 proteins in human 293T cells transfected with siRNA pools targeting these factors (*siMBNL1+2*, lane 6). Lane 5, lysate from cells transfected with a non-targeting siRNA pool (siControl). Lanes 1–4, serial dilutions (1:1, 1:2, 1:4 and 1:8) of lysate from cells transfected with siControl. **b**, Splicing code map highlighting genomic locations of *MBNL*, *RBFOX* and *PTBP* motifs associated with ES-cell-specific alternative splicing of *FOXP1*/*Foxp1* exon 18b/16b, the inclusion of which forms the *FOXP1-ES*/*Foxp1-ES* isoform. Human (black), mouse (grey) or conserved features (red) are indicated. Note that conserved MBNL motifs are associated with possible direct interference of exon 18b/16b splice site regulation. **c**, RT–PCR assays monitoring mRNA levels of *FOXP1* canonical (blue exon) and *FOXP1-ES* (red exon) isoforms in 293T cells transfected with siControl, *siMBNL1*, *siMBNL2* or *siMBNL1+2*. RT–PCR used splice-junction-specific primers, as indicated. Expression levels of actin are shown as loading controls. **d**, mRNA levels of murine *Foxp1*-canonical and *Foxp1-ES* isoforms were assayed as in **c** in N2A cells. Expression levels of *Gapdh* are shown as loading controls.

knockdown of MBNL1 alone, or together with MBNL2, resulted in increases in PSI (per cent spliced in), from zero to 2–2.4% and 6.5–7.1%, respectively (Figs 2c and 3c and Supplementary Figs 4b and 5). More pronounced effects were observed for *Foxp1* exon 16b in N2A cells (PSI shift from 0 to 15.1 for the double knockdown; Fig. 2d and Supplementary Fig. 4c). Knockdowns in ES cells had modest effects on exon 18b/16b splicing, consistent with the low levels of MBNL expression in these cells (Supplementary Fig. 4d, e). Knockdown of a third MBNL family member, MBNL3, which has a more restricted cell-type distribution compared to MBNL1 and MBNL2 (ref. 22), had no detectable effect on exon 18b splicing (Supplementary Fig. 5). These results suggest that MBNL1 and MBNL2 proteins have conserved and partially redundant roles in the negative regulation of *FOXP1*/*Foxp1* exon 18b/16b inclusion.

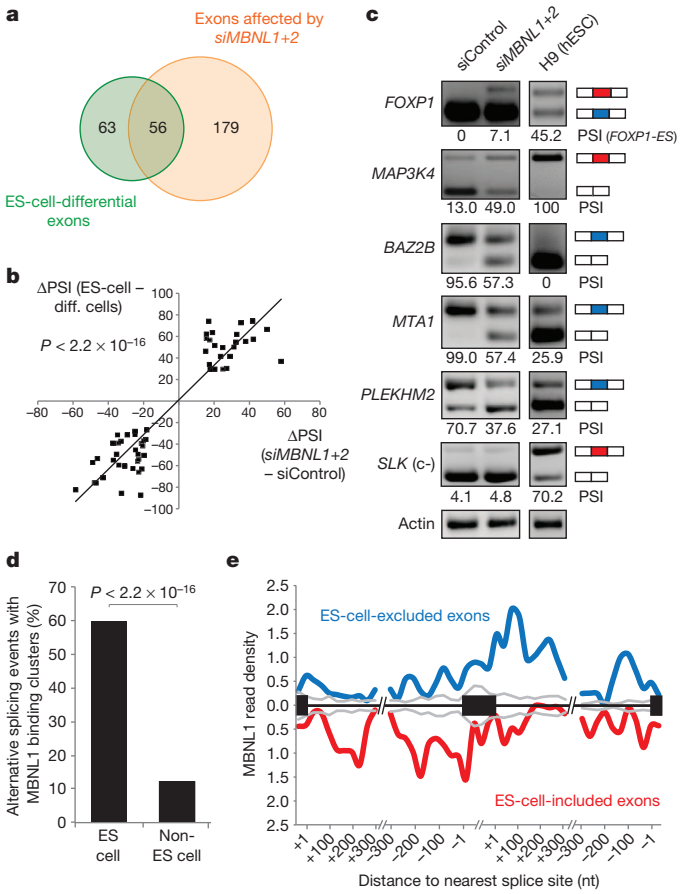

Figure 3 | MBNL proteins regulate approximately half of ES-cell-differential alternative splicing events. **a**, Venn diagram showing the proportion of ES-cell-differential alternative splicing events (green) that display ≥15 PSI change between HeLa cells transfected with siRNA pools targeting *MBNL1* and *MBNL2* (*siMBNL1+2*) versus siControl pool (orange). **b**, High association ($P < 2.2 \times 10^{-16}$, one-sided binomial test between quadrants) between differences in PSIs of ES cells versus differentiated cells/tissues, and differences in PSIs of *siMBNL1+2* knockdown versus siControl treatments. **c**, Representative RT–PCR validations for ES-cell-differential alternative splicing events that have PSI changes in HeLa cells after *siMBNL1+2* transfection and for ES-cell-differential alternative splicing events that do not change upon *siMBNL1+2* knockdown (c-); splicing patterns in human H9 ES cells are shown for comparison. **d**, Percentage of alternative splicing events with overlapping MBNL1 CLIP-seq binding clusters[20] in C2C12 cells for ES-cell-differential or non-ES-cell-regulated alternative exons. **e**, Merged map of MBNL binding clusters in transcripts with ES-cell-differential alternative splicing events. Maps of MBNL1 binding sites with respect to exons that have higher or lower inclusion in ES cells/iPSCs, relative to non-ES cells/tissues.

To assess the extent to which ES-cell-differential alternative splicing events are controlled by MBNL proteins, MBNL1 and MBNL2 were knocked down in HeLa cells, and RNA-seq profiling was used to detect alternative splicing changes (Fig. 3). Of 119 profiled ES-cell-differentially spliced exons, nearly half are affected by knockdown of MBNL proteins, with a ≥15 PSI change towards an ES-cell-like alternative splicing pattern (Fig. 3a). A strong overall association ($P < 2.2 \times 10^{-16}$, one-sided binomial test) was observed when comparing PSI changes for exons differentially spliced between ES cells and non-ES cells/tissues, and PSI changes for the same exons following knockdown of MBNL proteins (Fig. 3b). RT–PCR experiments confirmed all analysed MBNL knockdown-dependent and -independent PSI changes (Fig. 3c and Supplementary Fig. 6a). The specificity of the knockdown experiments was further demonstrated by comparing individual siRNAs that target different sequences within MBNL1 transcripts (Supplementary Fig. 6b). Comparable results were observed when MBNL1 and MBNL2 proteins were simultaneously knocked down in 293T cells, and in undifferentiated C2C12 mouse myoblast cells (Supplementary Fig. 7). Conversely, overexpression of MBNL1 and MBNL2 proteins in mouse ES cells promoted differentiated-cell-like patterns for all analysed ES-cell-differential alternative splicing events (Supplementary Fig. 8), including a switch to the exclusive use of the canonical (that is, non-ES cell) exon 16 in *Foxp1* transcripts. Consistent with this observation, overexpression of MBNL proteins in ES cells also led to increased kinetics of silencing of core pluripotency factors upon differentiation, and further promoted the expression of specific lineage markers representative of all three germ layers (Supplementary Fig. 9).

Mapping of MBNL protein binding to endogenous transcripts using ultraviolet crosslinking coupled to immunoprecipitation and sequencing (CLIP-seq or HITS-CLIP[23]) in undifferentiated C2C12 myoblast cells[20] confirmed that these proteins directly target ES-cell-differential alternative splicing events, including *Foxp1* exon 16b (Fig. 3d and Supplementary Fig. 10a). Of 57 mouse ES-cell-differential exons expressed in C2C12 cells, ~34 (60%) are associated with overlapping or proximal clusters of MBNL CLIP-seq tags ('binding clusters'), whereas binding clusters are associated with 72 out of 601 (12%) of exons that are not differentially regulated in ES cells ($P < 2.2 \times 10^{-16}$, proportion test; Fig. 3d). The binding clusters associated with ES-cell-differential alternative splicing are significantly enriched in consensus binding sites for MBNL proteins (Supplementary Fig. 10b)[20,21,24]. Moreover, consistent with the splicing code analysis (Fig. 1b and Supplementary Fig. 1b) and previous results[20,21], the locations of MBNL binding clusters correlate with whether the target exons are more or less included in ES cells compared to other cells and tissues (Fig. 3e). Collectively, the results so far demonstrate that MBNL proteins act widely and directly to regulate ES-cell-differential alternative splicing, and consequently pluripotency factor expression.

We next asked whether MBNL proteins have an impact on somatic cell reprogramming (Fig. 4a). Secondary mouse embryonic fibroblasts (MEFs)[25] expressing the 'OKSM' transcription factors (OCT4, KLF4, SOX2, c-MYC)[26] from transgenes under doxycycline-inducible control were transfected with siRNA pools to knock down MBNL1 and MBNL2 (*siMbnl1+2*), or with a control, non-targeting siRNA pool (siControl). At days 3 and 5 after doxycycline induction, mRNA expression of endogenous pluripotency genes, including *Oct4*, *Nanog*, *Sall4* and *Alpl*, were assayed by qRT–PCR (Fig. 4b and Supplementary Fig. 11a). None of these genes displayed significant changes in expression at day 3 (Fig. 4a, b); however, at day 5, MBNL knockdown stimulated their expression by approximately twofold over the siControl treatment (Fig. 4b and Supplementary Fig. 11a). MBNL knockdown also resulted in a ~30% increase in the colony area immunostained for SSEA1, a pluripotency-associated marker expressed early during reprogramming (Fig. 4c). In contrast, knockdown of OCT4 (*siOct4*) resulted in significant reductions in endogenous pluripotency gene expression and in SSEA1-positive colonies (Fig. 4b, c).

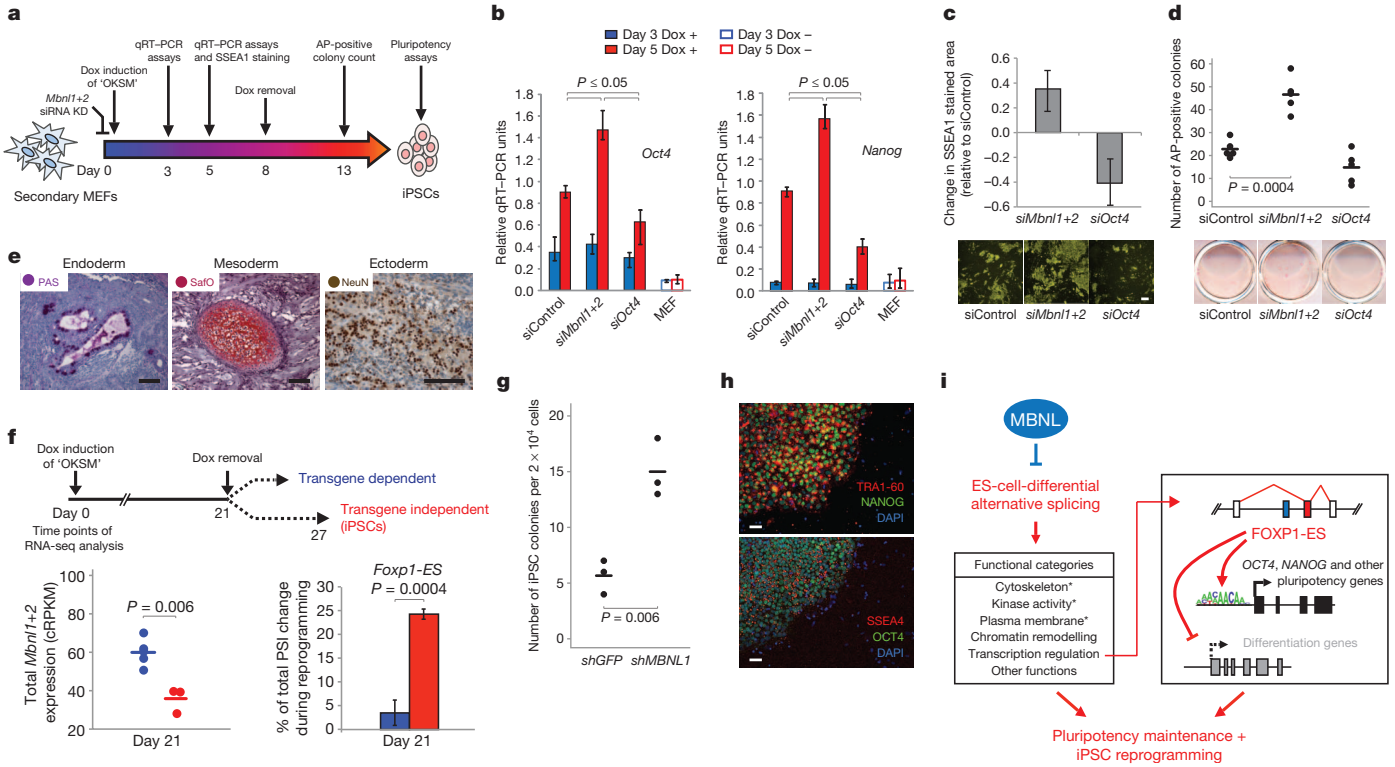

Figure 4 | Knockdown of MBNL proteins enhances reprogramming efficiency and kinetics. **a**, Experimental scheme. **b**, qRT–PCR quantification of mRNA expression levels of endogenous *Oct4* and *Nanog* (data for additional genes in Supplementary Fig. 11a). Secondary MEFs were transfected with control siRNAs (siControl), siRNAs targeting *Mbnl1* and *Mbnl2* (*siMbnl1+2*) or *Oct4* (*siOct4*) and treated with doxycycline (Dox) for 3 days (blue bars) or 5 days (red bars) before analysis. Empty bars, secondary MEFs without doxycycline induction. Values represent means ± range ($n = 3$). **c**, Top: quantification of SSEA1-stained area change relative to siControl at day 5 after doxycycline induction; values represent means ± range ($n = 3$). Bottom: representative images of SSEA1 staining. Scale bar, 100 μm. **d**, Top: quantification of doxycycline-independent iPSC colony formation. Secondary MEFs were treated with doxycycline for 8 days followed by 5 days of doxycycline withdrawal and counting of alkaline-phosphatase-positive colonies. Bottom: representative images of alkaline phosphatase staining. **e**, Teratoma assay assessing the pluripotency potential of iPSCs derived from secondary MEFs after knockdown of MBNL proteins. Haematoxylin and eosin staining, with additional staining/immunolabelling using periodic acid-Schiff (PAS; for detection of glycogen or glycoprotein producing cells), safranin O (SafO; for detection of cartilage), or antibody to neuronal nuclear antigen (NeuN) is shown (see Supplementary Figs 12 and 13 for additional teratoma analysis and chimaera testing of the pluripotency potential of *siMbnl* iPSCs). Scale bar, 100 μm. **f**, Top: experimental scheme for clonal analysis. Upon doxycycline removal at day 21, clones derived from single cells either survive and form iPSCs (transgene independent) or do not survive (transgene dependent). Bottom: analysis of total *Mbnl1*/*Mbnl2* mRNA expression (left) and percentage of total PSI change (right) for *Foxp1* exon 16b in transgene-independent (red) and transgene-dependent (blue) clones, at day 21, where total PSI change is the PSI difference between MEFs and iPSCs during reprogramming. **g**, Quantification (by morphological examination) of human iPSC colonies formed by reprogramming BJ fibroblasts expressing shRNA targeting GFP (shGFP) or *MBNL1* (*shMBNL1*). **h**, Immunostaining of human iPSCs derived from *shMBNL*-expressing BJ fibroblasts for TRA1-60, NANOG, SSEA4 and OCT4 pluripotency markers. Scale bar, 50 μm. See Supplementary Fig. 15 for additional characterization of human iPSCs. *P* values of one-sided *t*-tests shown for all comparisons in this figure. **i**, Model for the role of MBNL proteins in the regulation of ES-cell-differential alternative splicing, pluripotency and iPSC reprogramming. Asterisks indicate significantly enriched gene-function categories.

Successful reprogramming requires that cells undergo a transition to an OKSM transgene-independent state[27]. We therefore asked whether suppression of MBNL proteins promotes transgene independence. OKSM transgenes were induced with doxycycline for 8 days, then the cells were cultured for 5 days without doxycycline (Fig. 4a). Whereas knockdown of OCT4 reduced colony formation, knockdown of MBNL proteins resulted in an approximate twofold increase in transgene-independent colonies, as detected by alkaline phosphatase staining ($P = 0.0004$; one-sided *t*-test) (Fig. 4d and Supplementary Fig. 11b). iPSC lines derived from transgene-independent colonies after MBNL knockdown were pluripotent and contributed to all three germ layers in both teratoma and chimaera assays (Fig. 4e and Supplementary Figs 11–13). Consistent with these results, *Mbnl* expression is significantly reduced in secondary MEF clones[27] cultured in the presence of doxycycline that are competent to achieve transgene independence (when doxycycline is removed) versus those that are not ($P = 0.006$; one-sided *t*-test) (Fig. 4f, left). Moreover, the PSI levels of ES-cell-differential alternative splicing events, including *Foxp1* exon 16b, significantly correlate with ES cell/iPSC alternative splicing patterns only in those clones that are competent to transition to transgene independence (Fig. 4f, right; Supplementary Fig. 14 and Supplementary Table 5; $r = 0.80$, $P = 3.2 \times 10^{-12}$). Notably, knockdown of MBNL1 in human fibroblasts expressing OKSM also resulted in an approximate twofold increase in the appearance of iPSC colonies (Fig. 4g, h and Supplementary Fig. 15). MBNL proteins thus have a conserved, negative regulatory role in somatic cell reprogramming.

The results of this study reveal that MBNL proteins negatively regulate an ES-cell-differential alternative splicing network that controls pluripotency and reprogramming (Fig. 4i). These proteins probably act, in part, by directly repressing the ES-cell-specific splicing switch in FOXP1, which promotes the expression of core pluripotency genes. However, additional genes with MBNL-regulated alternative splicing events have been linked to the control of pluripotency, indicating a more extensive role for the alternative splicing network in ES-cell biology (Fig. 4i). These observations represent the first evidence that *trans*-acting splicing regulators have a central role in the core circuitry

required for ES-cell pluripotency and reprogramming. Our results further offer a potential new approach for enhancing the production of iPSCs for research and therapeutic applications.

METHODS SUMMARY

siRNA knockdown and RNA analysis. Cells were transfected with SMART-pool siRNAs (Dharmacon) using DharmaFECT1 reagent and collected 48 or 72 h after transfection. Secondary MEFs were transfected with siRNA pools using Lipofectamine RNAiMAX (Invitrogen), as described previously[28], and OKSM transgenes were induced 24 h after with doxycycline. Semi-quantitative RT–PCR assays were performed using the OneStep RT–PCR kit (Qiagen), with modifications as described previously[29]. Quantitative RT–PCR assays were performed as previously described[9]. Primer sequences are available upon request.

iPSC colony formation assays and characterization. Cells were transfected with siRNAs pools and treated with doxycycline for 5 days before imaging using an IN Cell Analyzer 2000 (GE Healthcare). To assay formation of doxycycline-independent colonies, secondary MEFs transfected with siRNA pools were treated with doxycycline for 8 days. Cell counting was performed before and at each passage after siRNA transfection. At day 8, the same number of cells were passaged in doxycycline-free media in 12-well plates and cultured until day 13, when they were fixed and stained with alkaline phosphatase for colony counting. For single-cell assays, secondary MEFs were induced by doxycycline treatment and clonal derivatives were cultured for 21 days. Removal of doxycycline at day 21 revealed alkaline-phosphatase-positive colonies (transgene-independent clones) and failed colonies (transgene-dependent clones). RNA-seq analysis was performed on three transgene-independent and five transgene-dependent clones at day 21 after doxycycline induction, as previously described[27]. Details of human iPSC generation and characterization are available in the Methods.

Teratoma and chimaera analysis. ES cells were injected subcutaneously into dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) and resulting teratomas were analysed using immunohistochemistry or cell-specific staining 4 to 5 weeks after injection. Chimaera aggregation and whole-mount staining were performed as previously described[27].

Full Methods and any associated references are available in the online version of the paper.

Received 12 September 2012; accepted 7 May 2013.
Published online 5 June 2013.

1. Young, R. A. Control of the embryonic stem cell state. *Cell* **144,** 940–954 (2011).
2. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81,** 145–166 (2012).
3. Bao, X. *et al.* MicroRNAs in somatic cell reprogramming. *Curr. Opin. Cell Biol.* **25,** 208–214 (2013).
4. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40,** 1413–1415 (2008).
5. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
6. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R. & Blencowe, B. J. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152,** 1252–1269 (2013).
7. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463,** 457–463 (2010).
8. Kalsotra, A. & Cooper, T. A. Functional consequences of developmentally regulated alternative splicing. *Nature Rev. Genet.* **12,** 715–729 (2011).
9. Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147,** 132–146 (2011).
10. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
11. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
12. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
13. Irimia, M. & Blencowe, B. J. Alternative splicing: decoding an expansive regulatory layer. *Curr. Opin. Cell Biol.* **24,** 323–332 (2012).
14. Rao, S. *et al.* Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol. Cell. Biol.* **30,** 5364–5380 (2010).
15. Salomonis, N. *et al.* Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl Acad. Sci. USA* **107,** 10514–10519 (2010).
16. Mayshar, Y. *et al.* Fibroblast growth factor 4 and its novel splice isoform have opposing effects on the maintenance of human embryonic stem cell self-renewal. *Stem Cells* **26,** 767–774 (2008).
17. Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465,** 53–59 (2010).
18. Liang, J. *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature Cell Biol.* **10,** 731–739 (2008).
19. Lian, I. *et al.* The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes Dev.* **24,** 1106–1118 (2010).
20. Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150,** 710–724 (2012).
21. Charizanis, K. *et al.* Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* **75,** 437–450 (2012).
22. Pascual, M., Vicente, M., Monferrer, L. & Artero, R. The Muscleblind family of proteins: an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **74,** 65–80 (2006).
23. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456,** 464–469 (2008).
24. Fernandez-Costa, J. M., Llamusi, M. B., Garcia-Lopez, A. & Artero, R. Alternative splicing regulation by Muscleblind proteins: from development to disease. *Biol. Rev. Camb. Philos. Soc.* **86,** 947–958 (2011).
25. Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458,** 766–770 (2009).
26. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
27. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11,** 769–782 (2012).
28. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
29. Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21,** 2963–2975 (2007).
30. Labbé, R. M. *et al.* A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30,** 1734–1745 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank U. Braunschweig, J. Ellis, S. Gueroussov and B. Raj for comments on the manuscript. We acknowledge D. Torti in the Donnelly Sequencing Centre for sequencing samples; L. Lee for assisting with the splicing code analysis; J. Garner (Hospital for Sick Children Embryonic Stem Cell Facility) for preparing feeder cells; A. Piekna for morphological examination of human iPSC colonies; M. Narimatsu for assisting with chimaerism analysis; and P. Mero for assisting with cell imaging. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) (to B.J.B., J.L.W., A.N., J.E. and B.J.F.), the Ontario Research Fund (to J.L.W., B.J.B., A.N. and others), the Canadian Stem Cell Network (to A.N. and B.J.B.), and by a grant from the National Institutes of Health (R33MH087908) to J.E. H.H. was supported by a University of Toronto Open Fellowship. P.J.R., M.I. and N.L.B.-M. were supported by postdoctoral fellowships from the Ontario Stem Cell Initiative, Human Frontiers Science Program Organization, and the Marie Curie Actions, respectively.

Author Contributions H.H. performed experiments in Figs 1–4 and Supplementary Figs 2–9 and 11–13. M.I. performed bioinformatic analyses in Figs 1–4 and Supplementary Figs 1, 3, 7, 10 and 14, with input from N.L.B.-M. L.D. and A.G. assisted with secondary MEF reprogramming experiments and clone characterization, and D.T. generated secondary MEF lines and performed chimaerism testing. P.J.R., T.T. and M.G. performed human reprogramming experiments and iPSC characterization. H-K.S. performed teratoma assays. B.A. and B.J.F. generated splicing code data. I.P.M., H.-K.S. and D.O. assisted with ES-cell overexpression and differentiation experiments. E.W. and C.B.B. generated and analysed CLIP-seq data. E.N.N. and V.S. performed RT–PCR validation experiments. B.J.B., H.H. and M.I. designed the study, with input from J.L.W., J.E., A.N. and J.M. B.J.B., H.H. and M.I. wrote the manuscript, with input from the other authors.

Author Information GEO accession numbers are provided in Supplementary Table 1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.J.B. (b.blencowe@utoronto.ca).

METHODS

siRNA knockdown and RNA analysis. Cells were transfected with SMART-pool siRNAs (Dharmacon) using DharmaFECT1 reagent and collected 48 or 72 h after transfection. Secondary MEFs were transfected with siRNA pools using Lipofectamine RNAiMAX (Invitrogen), as described previously[28], and OKSM transgenes were induced 24 h after with doxycycline. Semi-quantitative RT–PCR assays were performed using the OneStep RT–PCR kit (Qiagen), with modifications as described previously[29]. Quantitative RT–PCR (qRT–PCR) assays were performed as previously described[9]. Primer sequences are available upon request.

Cell lines and cell culture. HeLa, 293T and C2C12 cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and antibiotics (penicillin/streptomycin). Neuro2A (N2A) cells were grown in DMEM supplemented with 10% FBS, sodium pyruvate, MEM non-essential amino acids, and penicillin/streptomycin. H9 human ES cells, CGR8 and R1 mouse ES cells were cultured as described previously[31]. Secondary mouse embryonic fibroblasts (MEFs) were maintained in DMEM supplemented with 10% FBS, L-glutamine and penicillin/streptomycin on 0.1% gelatin-coated plates. During reprogramming, secondary MEFs were grown in mouse ES media and induced to express OKMS factors using 1.5 μg ml^{-1} of doxycycline as described previously[32,33].

Protein extraction and western blotting. Cell pellets were lysed in radioimmunoprecipitation assay (RIPA) buffer by brief sonication. Protein lysate (30–150 μg) was separated on a 10% SDS–polyacrylamide gel and transferred to a PVDF membrane. The membranes were blotted with the following antibodies: anti-Flag M2 (1:1500, Sigma), anti-MBNL1 (1:500, Abcam), anti-MBNL2 (1:200, Santa Cruz Biotechnology) and anti-α-tubulin (1:5000, Sigma). Secondary antibodies (GE Healthcare) and chemiluminescence reagents (Perkin Elmer) were used as per the manufacturer's instructions.

RNA extraction and qRT–PCR assays. Total RNA was extracted using TRI Reagent (Sigma) or RNeasy columns (Qiagen). RT–PCR assays were performed using the OneStep RT–PCR kit (Qiagen), as per the manufacturer's instructions. 20 ng total RNA or 1 ng of polyA+ RNA was used per 10-μl reaction. Radiolabelled reactions contained 0.3 μCi of α-^{32}P-dCTP per 10-μl reaction. The number of amplification cycles was 22 for actin and *Gapdh*, and 27–32 for all other transcripts analysed. Reaction products were separated on 1–3% agarose gels. Quantification of isoform abundance was performed using either ImageQuant (GE Healthcare) or ImageJ software. To amplify the *FOXP1*/*Foxp1* isoforms selectively (Fig. 2c, d), primers specific for splice junctions were used.

For quantitative RT–PCR, first-strand cDNAs were generated from 1–3 μg of total RNA or 100 ng of polyA+ RNA using SuperScript III Reverse Transcriptase (Invitrogen), as per the manufacturer's recommendations, and diluted to 20 ng μl^{-1} and 1 ng μl^{-1}, respectively. qPCR reactions were performed in a 384-well format using 1 μl of each diluted cDNA and FastStart Universal SYBR Green Master (Roche Applied Science). All primers sequences are available upon request.

Immunofluorescence. For immunofluorescence experiments, cells were fixed in 4% PFA for 10 min at room temperature, washed with PBS, and permeabilized for 10 min at 4 °C with 0.1% Triton X-100. After 1 h of blocking, cells were incubated with primary antibodies overnight at 4 °C, and then with secondary antibodies for 1 h at room temperature. Nuclei were stained with Hoechst 33258 (1:5,000, Sigma-Aldrich). Primary antibodies used in this study are: mouse IgM anti-SSEA1 (1:500, BD Biosciences), mouse anti-OCT4 (1:200, BD Biosciences), rabbit anti-NANOG (1:200, Cosmo Bio), goat anti-DPPA4 (1:250, R&D), mouse IgM anti-TRA1-60 (1:100, Invitrogen), rabbit anti-NANOG (1:400, Cell Signaling), mouse anti-SSEA4 (1:100, Invitrogen), rabbit anti-OCT4 (1:200, Abcam), mouse anti-alpha fetoprotein (1:200, R&D), mouse anti-smooth muscle actin (1:200, Invitrogen), and mouse anti-beta-III-tubulin (1:200, Millipore). Secondary antibodies used in this study are: anti-mouse IgM Alexa555 (1:1000, Molecular Probes), anti-mouse IgG Alexa555 (1:1,000, Molecular Probes), anti-rabbit IgG Alexa594 (1:1,000, Molecular Probes), anti-rabbit IgG Alexa488 (1:500, Molecular Probes), and anti-goat IgG Alexa546 (1:1,000, Molecular Probes).

iPSC colony-formation assays and imaging from secondary MEF reprogramming. Secondary MEFs were seeded in 12-well plates, transfected with siRNA pools, and treated with doxycycline for 5 days before fixing and staining. The plates were imaged (for both SSEA1-immunostained and DAPI channels) using an IN Cell Analyzer 2000 (GE Healthcare) with a ×4 objective. For each well, 20 non-overlapping fields were captured and images were analysed using the Columbus System (PerkinElmer). A custom script was generated to identify SSEA1-positive and DAPI-positive colonies. The overall signal in each well was determined using the sum of the overlap area for the 20 fields captured.

To assay the formation of doxycycline-independent colonies, secondary MEFs transfected with siRNA pools were treated with doxycycline for 8 days. Cell counting was performed before and at each passage after siRNA transfection and doubling rates were determined not to change significantly (data not shown). At day 8, the same number of cells were passaged into doxycycline-free mES-cell media on 12-well plates and cultured until day 13, when they were fixed and stained with alkaline phosphatase for colony counting.

Teratoma analysis. Cells were suspended in PBS and Matrigel (BD Bioscience) mixed solution, and 1×10^6 cells in 100 μl were injected subcutaneously into both dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) anaesthetized with isoflurane. Four to five weeks after injection, mice were killed and teratomas were dissected, fixed overnight in 10% buffered formalin phosphate, and embedded in paraffin. Three-to-four-micrometre-thick sections were deparaffinized and hydrated in distilled water. Sections were stained either with haematoxylin and eosin for regular histological examination, or with the following dyes: 0.1% safranin O solution (cartilage, mesoderm-derived tissue) or 0.5% PAS solution (glycoprotein-producing intestinal cell, endoderm-derived tissue). For immunohistochemistry, sections were deparaffinized and hydrated, and antigen retrieval process was performed. After blocking, sections were incubated overnight at 4 °C with primary monoclonal antibody (1:100, Millipore MAB377, clone A60) specific for neuronal nuclear antigen (NeuN, ectoderm-derived tissue), followed by washing in PBS. After 1 h of incubation with secondary anti-mouse-HRP conjugated antibody (1:500, Jackson ImmunoResearch, 115-035-003), signal was visualized by DAB (3,3′-diaminobenzidine; Vector Laboratories, SK-4100) substrate for 5–20 min. Sections were counter-stained with haematoxylin.

Chimaerism analysis. Chimaera aggregation and whole-mount staining were performed as described previously[34]. Chimeras were obtained through aggregation of siMbnl iPSC clumps with diploid Hsd:ICR(CD-1) embryos. E10.5 embryos were dissected after doxycycline treatment *in utero* via ingestion 24 h before dissection. After dissection, embryos were fixed with 0.25% glutaraldehyde, rinsed in wash buffer (2 mM $MgCl_2$, 0.01% sodium deoxycholate, and 0.02% Nonidet-P40 in PBS) and then stained overnight in LacZ staining solution (20 mM $MgCl_2$, 5 mM K3Fe(CN)6, 5 mM K4Fe(CN)6 and 1 mg ml^{-1} X-gal in PBS). Embryos were embedded in paraffin, sectioned and counterstained with nuclear fast red.

Generation and characterization of human iPSCs. Human BJ foreskin fibroblasts (Stemgent) were reprogrammed using published protocols[35,36], with the following modifications. BJ fibroblasts were first infected with lentivirus vectors encoding both a puromycin resistance gene and doxycycline-inducible shRNA targeting either GFP (negative control, target sequence: 5′-GCAAGCTGACCCTGAAGTTCAT-3′) or *MBNL1* shRNA (target sequence: 5′-GCCTGCTTTGATTCATTGAAA-3′). Lentiviral vector preparations and infections were performed as described[35]. After selection with 1 μg ml^{-1} puromycin, shRNA-encoding BJ fibroblasts were infected with a second lentivirus vector (obtained from Addgene) co-expressing both the mouse retrovirus receptor *mSlc7a1* and the blasticidin resistance gene[37]. During transient selection with 5 μg ml^{-1} blasticidin, puromycin was reduced to 0.5 μg ml^{-1} and maintained at this concentration for 6 days after infection with retroviral reprogramming vectors.

pMXs-based retrovirus vectors encoding the four reprogramming factors hOCT4, hSOX2, hKLF4, hCMYC (OSKM)[37], were obtained from Addgene and packaged exactly as described[35]. Puromycin/blasticidin-resistant BJ fibroblasts were infected in triplicate, using three separate preparations of retrovirus vectors. shRNA expression was induced by treatment with 2 μg ml^{-1} doxycycline, which was initiated contemporaneously with retrovirus vector infection; control cells were treated with vehicle only.

Six days after retrovirus infection, BJ fibroblasts were seeded on a monolayer of feeder cells. Embryonic day 12.5 fibroblasts from Tg(DR4)1Jae/J mice (Jackson Laboratory) were seeded on collagen-coated 6-well plates at a density of 3×10^5 cells per well as described[35]; retrovirus-infected BJ fibroblasts were seeded at a density of 2×10^4 cells per well. At day 28 of reprogramming, quantification (by whole-well morphological examination and by TRA1-60 immunostaining) of human iPSC colonies was performed by investigators who were blinded as to the experimental conditions. To count TRA1-60-positive colonies, the plates were imaged (for both TRA1-60-immunostained and DAPI channels) using an IN Cell Analyzer 2000 (GE Healthcare) with a ×4 objective. For each well, 64 non-overlapping fields were captured and images were analysed using the Columbus System (PerkinElmer). Knockdown of MBNL1 resulted in an approximate twofold increase in TRA1-60 immunostaining colonies over the control knockdown with GFP-targeting shRNA (data not shown). Additional OSKM retrovirus-infected BJ fibroblasts were seeded in parallel; individual colonies from doxycycline-treated plates were manually isolated 4 weeks after infection, and seeded on feeders in collagen-coated 24-well plates[35]. Cells from these colonies were expanded, and subsequently characterized (Supplementary Fig. 15) as described[38].

Clonal analysis by RNA-seq during reprogramming. In a single-cell assay, secondary MEFs were plated in individual wells of a 96-well plate, OKSM factors were induced by doxycycline treatment and the clonal derivatives were cultured for 21 days. Removal of doxycycline at day 21 revealed that approximately 50% of clones produced abundant alkaline-phosphatase-positive colonies (transgene-independent

clones), whereas the rest yielded few or no colonies (transgene-dependent clones). RNA-seq analysis was performed for three transgene-independent and five transgene-dependent clones at day 21 after doxycycline induction (Supplementary Table 1).

Using RNA-seq derived-PSI values (see Supplementary Methods), the inclusion levels of mouse ES-cell-differential cassette alternative exons were quantified for each of these clones. Fifty-one ES-cell-differential alternative splicing events with sufficient read coverage in all samples and with a ≥25 PSI difference between iPSCs and MEFs were compared between the two types of clones (Supplementary Fig. 14 and Supplementary Table 5).

RNA-seq data and analysis. We used RNA-seq data from 36 and 32 different human and mouse samples, respectively. Details and sample sources are provided in Supplementary Table 1. The samples comprise, for human: 5 ES-cell lines (3 different cell lines and 2 replicates), 2 iPSC lines, 7 non-ES-cell lines, and 22 adult tissues (18 different tissues and 4 replicates); for mouse: 6 ES-cell lines, 2 iPSC lines, 8 non-ES-cell lines (5 different cell lines and 3 replicates) and 16 adult tissues (10 different types and 6 replicates). Details of RNA-seq analysis is available in Supplementary Information.

To identify alternative exons differentially regulated in ES cells, we first calculated a single averaged PSI value for tissues of similar origin (see Supplementary Table 1). Only events with enough coverage in at least two ES-cell samples and three distinct tissue types were considered. 'ES-cell-differential alternative splicing events' were defined as those with a mean PSI difference of ≥25 between ES cells and differentiated tissues. To account for alternative splicing events potentially related to cell proliferation, we also required a mean PSI difference of ≥ 25 between ES-cell lines and non-ES-cell lines, when the event had sufficient coverage in at least one cell line sample. The set of background alternative splicing events used throughout the study are alternatively spliced exons (defined here as exons with PSI values of <95% and >5% in at least one sample) that meet the same expression requirement (that is, in ≥2 ES cells and ≥3 differentiated tissue types) and that show an average difference in PSI level of <5% between ES cells and differentiated tissue samples, and between the ES-cell and non-ES-cell lines.

Analyses of splicing factor expression. A total of 221 human and 214 mouse genes were selected for analysis based on literature mining for previously described splicing functions, 'splicing'- and/or 'spliceosome'-associated Gene Ontology (GO) terms, and/or the presence of a PFAM-annotated RNA-binding domain (Supplementary Table 4). To calculate the mRNA expression values for each sample, we used corrected (for mappability) reads per kilobase pair and million mapped reads values (cRPKMs) of the 'stable' (as defined by BioMart) Ensembl transcript for each gene, as previously described[39].

Splicing factor genes were ranked according to the relative extent of their differential expression (as determined by cRPKM values) in ES cells and iPSCs versus non-ES-cell lines and tissues by comparing summed ranks of each gene in all ES cell/iPSCs across the full range of samples. On the basis of this approach, human *MBNL1* and *MBNL2* showed the first and second lowest overall rank in ES cells/iPSCs, respectively, and mouse *Mbnl1* and *Mbnl2* showed the first and third lowest overall rank in ES cells/iPSCs, respectively.

To assess the statistical significance of the differential expression of individual splicing factor genes, we compared their cRPKM values in ES cells/iPSCs to the cRPKMs in all other cell lines and differentiated tissues using a Wilcoxon rank-sum test after quantile normalization. Splicing factors with Bonferroni-corrected *P*-values <0.05 were considered significantly differentially expressed (Supplementary Table 4).

Splicing code analyses. The feature vectors for each species were produced by extracting sequence-based features from alternatively spliced exons, their adjacent constitutive exons, and 300 nucleotides of flanking intronic sequence. The features used were a subset of those defined previously[40], with the following differences: (1) all sequence length features are now in the log domain; (2) owing to a lack of comprehensive transcript libraries and the corresponding uncertainty about downstream consequences of frame shifts, premature termination codon features were excluded; and (3) conservation scores and conservation-weighted motifs were excluded from the feature set. In addition, related features (that is, consensus recognition sequences for a given splicing factor inferred by different methods) were combined and included as independent features.

To identify features strongly associated with ES-cell-differential exon inclusion or exclusion, we compared 172 ES-cell-differential exons and 908 background exons for human, along with 102 ES-cell-differential exons and 811 background exons for mouse. Associations between features and ES-cell-differential splicing were detected using Pearson correlation. For each feature, we computed the correlation between its value and the difference in average PSI values in ES cells and non-ES cells, across exons. To obtain more accurate correlation values, we considered two scenarios: (1) a positive scenario in which the differences in average PSI values in ES cells and non-ES cells are larger than 25%; and (2) negative scenario in which the differences in average PSI values in ES cells and non-ES cells are smaller than −25%.

CLIP-seq analysis. We used recently described *Mbnl1* CLIP-seq data from C2C12 cells[41]. To estimate the fractions of ES-cell-differential and background alternative splicing events that are associated with MBNL1 binding, we asked whether CLIP binding clusters overlap the alternative exon and/or flanking intron sequences of each event. CLIP binding clusters were defined as previously described[41]. In short, CLIP-seq tags were trimmed of adapters and then collapsed to remove redundant sequences. These tags were mapped to genome and a database of splice junctions using Bowtie. To identify CLIP clusters lying within genic regions, gene boundaries were first defined using RefSeq, Ensembl and UCSC tables. For each window of 30 nucleotides covered by at least one CLIP-seq tag, a test was performed to assess whether the tag density in the window exceeded that which is predicted by a simple Poisson model which accounts for gene expression and pre-mRNA length. An alternative splicing event was considered to have an overlapping MBNL1 binding cluster if the mid-point of the cluster is located within the alternative exon, within 300 nucleotides of the 5′ or 3′ ends of upstream or downstream flanking introns, and/or within 30 nucleotides within the 3′ end of C1 exon or the 5′ end of C2 exon. Only alternative splicing events that had significant read coverage (see above) in at least one of two C2C12 samples used were analysed. In total, 57 ES-cell-differential alternative splicing events and 601 background alternative splicing events were compared.

To generate an RNA regulatory map[42] highlighting MBNL1 binding sites in relation to ES-cell-differential alternative splicing events with either higher (ES-cell-included) or lower (ES-cell-excluded) exon inclusion levels in ES cells versus other cell lines and tissues, we applied the following procedure: for each nucleotide position from the regions described above and from three sets of alternative splicing events (ES-cell-included, ES-cell-excluded and background), we counted the average number of MBNL1 CLIP-seq tags. To minimize the impact of outliers with extreme read density, we limited the maximum count per event to an average of ten reads per position within each region. To normalize the length of the alternative splicing exon, we divided each exon into 100 bins and uniquely assigned each position to the integer of 100*position/exon_length, with a relative weight inversely related to the length of the alternative splicing exon. To draw the map, we used sliding windows of 30 nucleotides for the intronic regions and 25 nucleotides for the length-corrected exons (total of four windows shown).

Evolutionary conservation of ES-cell-differential events. We analysed three different aspects of conservation of the human and mouse ES-cell-differential alternative exons[43]. To determine whether the alternative exon is conserved at the genomic level, we performed a lift-over of the exon coordinates using Galaxy (https://main.g2.bx.psu.edu/). Exons with a unique lift-over hit in the other species, and with AG (splicing acceptor) and/or GT (splicing donor) sites were considered to be genome-conserved in the other species. In addition, if the orthologous exon has a PSI of <95% and >5% in at least one sample from each species, alternative splicing of the exon was defined as conserved. Finally, to assess whether ES-cell-differential regulation is conserved, we applied two criteria: (1) the exons are independently detected as ES-cell-differential in human and mouse using the criteria as described above (total = 25 alternative splicing events); and (2) the orthologous exons must meet minimal read coverage requirements (also as described above) to afford direct comparison.

Analyses of function and protein domain enrichment. To investigate whether ES-cell-differential events are significantly enriched in genes with specific functional associations and/or protein domains, we used the online tool DAVID (http://david.abcc.ncifcrf.gov/)[44,45] (with annotations from levels 3, 4 and 5 in the GO hierarchy), KEGG pathways and InterPro domains. As background, we used the genes with at least one alternative splicing event that met the minimal expression criteria described above (that is, detection in ≥2 ES cells and ≥3 differentiated cell/tissue types). The main clusters of functionally related genes enriched in both human and mouse (as well as among the conserved) ES-cell-differential events (Supplementary Table 3) are associated with: actin cytoskeleton, plasma membrane (including cell junctions) and protein kinase-associated terms.

31. Gabut, M. *et al.* An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* **147,** 132–146 (2011).
32. Polo, J. M. & Hochedlinger, K. When fibroblasts MET iPSCs. *Cell Stem Cell* **7,** 5–6 (2010).
33. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7,** 64–77 (2010).
34. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11,** 769–782 (2012).
35. Hotta, A. *et al.* EOS lentiviral vector selection system for human induced pluripotent stem cells. *Nature Protocols* **4,** 1828–1844 (2009).

36. Hotta, A. *et al.* Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nature Methods* **6,** 370–376 (2009).
37. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
38. Cheung, A. Y. *et al.* Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Hum. Mol. Genet.* **20,** 2103–2115 (2011).
39. Labbé, R. M. *et al.* A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30,** 1734–1745 (2012).
40. Xiong, H. Y., Barash, Y. & Frey, B. J. Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context. *Bioinformatics* **27,** 2554–2562 (2011).
41. Wang, E. T. *et al.* Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150,** 710–724 (2012).
42. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456,** 464–469 (2008).
43. Irimia, M., Rukov, J. L., Roy, S. W., Vinther, J. & Garcia-Fernandez, J. Quantitative regulation of alternative splicing in evolution and development. *Bioessays* **31,** 40–50 (2009).
44. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).