nature
structural &
molecular biology

An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells

Gene W Yeo[1,2,4], Nicole G Coufal[2], Tiffany Y Liang[1,3,4], Grace E Peng[2], Xiang-Dong Fu[3] & Fred H Gage[2]

The elucidation of a code for regulated splicing has been a long-standing goal in understanding the control of post-transcriptional gene expression events that are crucial for cell survival, differentiation and development. We decoded functional RNA elements *in vivo* by constructing an RNA map for the cell type–specific splicing regulator FOX2 (also known as RBM9) via cross-linking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) in human embryonic stem cells. The map identified a large cohort of specific FOX2 targets, many of which are themselves splicing regulators, and comparison between the FOX2 binding profile and validated splicing events revealed a general rule for FOX2-regulated exon inclusion or skipping in a position-dependent manner. These findings suggest that FOX2 functions as a critical regulator of a splicing network, and we further show that FOX2 is important for the survival of human embryonic stem cells.

Understanding regulated gene expression is vital to providing insights into disease and development. Whereas much effort has been placed on deciphering transcriptional regulation though interactions with functional DNA elements by the more than a thousand transcription factors encoded in mammalian genomes, little is known about an equally sizable number of RNA binding proteins and their involvement in diverse aspects of RNA metabolism. A dominant function of these RNA binding proteins is to regulate alternative splicing, a major form of post-transcriptional regulation of gene expression that is thought to contribute to the structural and functional diversity of the cellular proteome[1]. One of the ultimate goals in the RNA field is to deduce a set of rules that govern the control of splice-site selection to produce the 'splicing code'. This goal can now be approached due to recent advances in functional genomics and high-throughput sequencing.

Human embryonic stem cells (hESCs) are pluripotent cells that propagate perpetually in culture as undifferentiated cells and can be readily induced to differentiate into various cell types both *in vitro* and *in vivo*[2]. As hESCs can theoretically generate most if not all of the cell types that constitute a human, they serve as an excellent model for understanding early embryonic development. Furthermore, hESCs are a nearly infinite source for generating specialized cells such as neurons and glia for potential therapeutic purposes or for screening small molecules to intervene with specific biological processes[3,4]. Therefore, there has been intense interest in identifying the molecular changes that are important for the survival of hESCs, maintenance of pluripotency and promotion of cell differentiation.

In our previous Affymetrix exon-tiling array analysis, we demonstrated that the FOX binding motif GCAUG was enriched proximal to a set of exons that are alternatively spliced in hESCs, suggesting that FOX splicing factors may have a vital role in the biology of hESCs[5]. Thus, we selected the RNA binding protein FOX2 to identify the functional RNA elements in the human genome in hESCs by deep sequencing. FOX2, a member of the FOX family of RNA binding proteins, was initially identified as a factor involved in dosage compensation in *Caenorhabditis elegans* and was later found to be evolutionarily conserved across mammalian genomes[6]. FOX2 is best known for its tissue-specific expression in muscle and neuronal cells and for its activity in regulated splicing in those highly differentiated cell types[6,7]. Unexpectedly, we found that FOX2 is expressed abundantly in the hESC lines HUES6 and H9, which are positive for the pluripotency markers OCT4, SOX2, NANOG and SSEA4 (**Fig. 1a** and **Supplementary Fig. 1** online). In contrast, FOX1 (also known as A2BP1) is not expressed in any hESCs examined. Consistent with their tissue-specific expression in cells of the neural lineage, both FOX1 and FOX2 are expressed in neural progenitors.

RESULTS

CLIP-seq for mapping functional RNA elements

We began to address the function of FOX2 in hESCs by developing a high-throughput experimental approach to large-scale identification of FOX2 targets *in vivo*, by coupling a modified CLIP technology[8] with high-throughput sequencing, a method we refer to as CLIP-seq (**Fig. 1b**). Key features of CLIP include: stabilization of *in vivo*

[1]Crick-Jacobs Center for Theoretical and Computational Biology, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Laboratory of Genetics, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [3]Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, Calfornia 92093-5004, USA. [4]Present address: Stem Cell Program, Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-5004, USA. Correspondence should be addressed to G.W.Y. (geneyeo@ucsd.edu), F.H.G. (gage@salk.edu) or X.-D.F. (xdfu@ucsd.edu).

Received 27 October 2008; accepted 11 December 2008; published online 11 January 2009; doi:10.1038/nsmb.1545

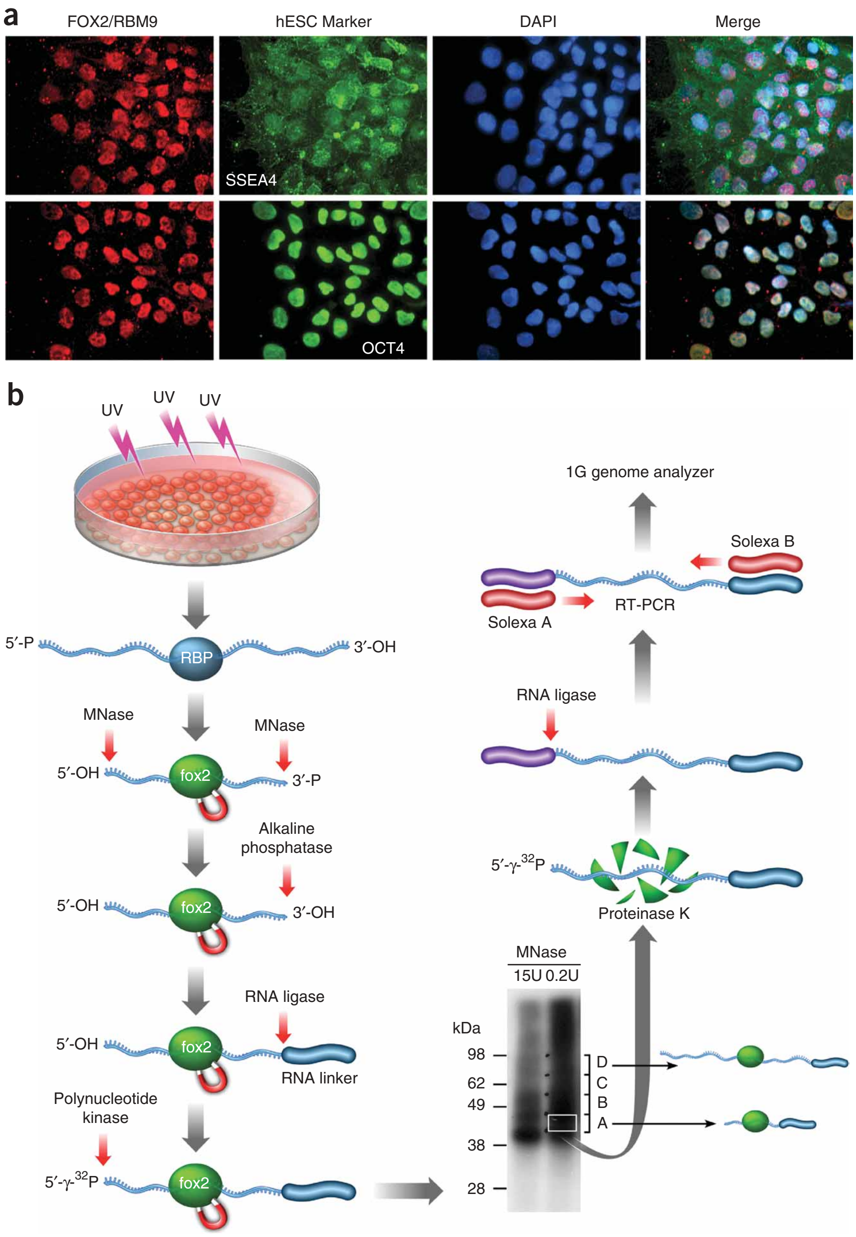


Figure 1 CLIP-seq of FOX2 in hESCs. (**a**) FOX2 is expressed in hESCs positive for pluripotent markers such as cytoplasmic SSEA4 and nuclear OCT4. Nuclei indicated by DAPI staining. (**b**) Flow chart of CLIP-Seq. RNA in complex with RNA binding proteins from UV-irradiated HUES6 hESCs was subjected to enrichment using anti-FOX2 rabbit polyclonal antibody. RNA in the complex was trimmed by MNase at two different concentrations, followed by autoradiography, as illustrated. Protein-RNA covalent complexes corresponding to bands A and B were recovered following SDS-PAGE, RT-PCR amplified and sequenced by the Illumina 1G genome analyzer.

protein-RNA interactions by UV irradiation, antibody-mediated enrichment of specific RNA-protein complexes, SDS-PAGE to isolate protein-RNA adducts after RNA trimming by nuclease, 3′ RNA linker ligation and 5′ labeling using ^{32}P-γATP. To prevent continuous RNA trimming by the RNase A used in the original protocol, we used micrococcal nuclease (MNase), which can be inactivated by EGTA, a modification that improves RNA recovery. Titration of the MNase allowed controlled trimming, resulting in short RNA molecules in the range of 50 nucleotides (nt) to 100 nt that remain attached to the protein (bands A and B in **Fig. 1b**). Recovered RNA was ligated to a 5′ linker before amplification by reverse-transcription PCR (RT-PCR). We designed both linkers to be compatible for sequencing on the Illumina 1G Genome Analyzer.

We obtained 5.3 million 36-nt sequence reads from anti-FOX2–enriched RNA from HUES6 hESCs, 83% of which (4.4 million) were uniquely mapped to the repeat-masked human genome (data available at hg17 and hg18 genome browsers (http://genome.ucsc.edu/) under ‘Regulation’). Our comparisons between genes containing CLIP reads in smaller-scale sequencing runs in HUES6 and H9 cells indicated a high overlap ranging from 70% to 90% (**Supplementary Fig. 2** online), indicating that FOX2 binds to similar targets in both cell lines. As expected from a splicing regulator that interacts primarily with transcribed mRNA, we found that the FOX2 binding sites were largely confined within protein-coding genes (∼3.7 million or 80% of total tags), 97% of which are oriented in the direction of transcription (sense-strand reads) (**Supplementary Fig. 3** online), confirming that DNA contamination was not a major issue with our preparation. Among annotated human genes, 16,642 (75%) contain one read within exonic or intronic regions, 3,598 (22%) have up to 10 reads and 543 (3%) harbored 10 to more than 1,000 reads. This distribution probably reflects the abundance of individual gene transcripts expressed in HUES6 cells, an assumption that was confirmed by the observation that the read density was positively correlated with gene expression measured on Affymetrix exon arrays in general (**Supplementary Fig. 4** online). This observation indicates that we cannot identify preferred targets for FOX2 by simply rank ordering the reads that map to individual transcripts.

Genomic distribution of *in vivo* FOX2 binding sites

To distinguish enriched FOX2 binding sites from background binding, we established gene-specific thresholds based on the assumption that FOX2 may prefer to bind to specific loci, rather than binding randomly to distributed sites along individual transcripts. We therefore computationally extended each genome-aligned sequence read in the 5′-to-3′ direction by 100 nt—the average length of RNA fragments after MNase treatment. The height at each position indicates the number of reads that overlap with that position.

To identify enriched FOX2 binding in clusters, we determined the false-discovery rate (FDR) for each position by computing the ‘background’ frequency after randomly placing the same number of extended reads within the gene for 100 iterations, similar to an approach that has been described for finding DNA-protein interaction clusters[9]. For a particular height, our modified FDR was computed as the ratio of the probability of observing background positions of at least that height to one standard deviation above the average probability of observing actual positions of at least that height. We

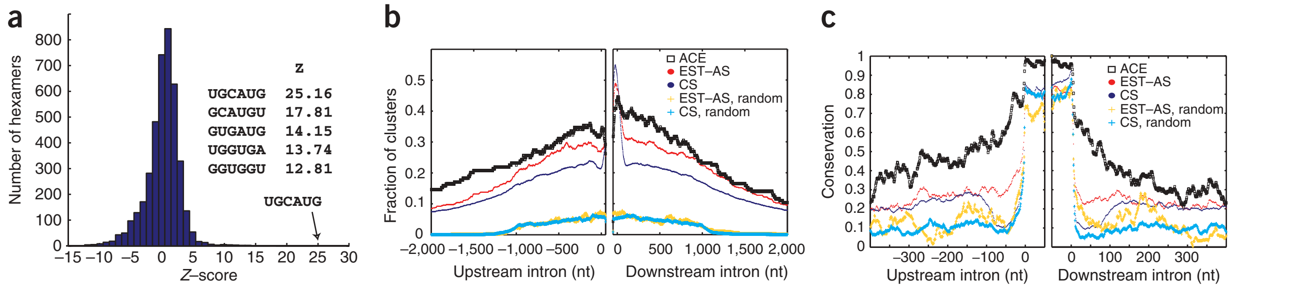


Figure 2 Genomic mapping and analysis of FOX2 CLIP-seq reads. (**a**) Consensus *in vivo* FOX2 binding sites identified by CLIP-seq. Histogram of *Z*-scores indicating the enrichment of hexamers in CLIP-seq clusters compared to randomly chosen regions of similar sizes in the same genes. *Z*-scores of the top five hexamers were indicated. (**b**) Enrichment of FOX2 CLIP-seq clusters within both constitutive and regulated exons and flanking intronic regions, particularly in the 3′ half of exons and downstream intronic regions. The FOX2 CLIP-seq clusters mapped most frequently to alternative conserved exons (ACEs) predicted by ACEScan, followed by EST-verified AS exons (EST-AS), compared to constitutively spliced exons (CS). Randomly chosen regions of similar sizes in the same genes were not distributed near EST-AS exons (EST-AS, random) and CS exons (CS, random). The *x* axis indicates a composite intron-exon-intron structure, containing sequences from 2,000 nt in the upstream intron and the first 50 nt of the exon (left), and the last 50 nt of the exon and 2,000 nt in the downstream intron (right). The *y* axis indicates the frequency of FOX2 CLIP-seq clusters. (**c**) Sequence conservation of FOX2 CLIP-seq clusters associated with different classes of exons. The average Phastcons scores were used to compute the extent of conservation[17].

identified FOX2 binding clusters by grouping positions that satisfied the condition FDR $<$ 0.001 and occurred within 50 nt of each other. This analysis identified 6,123 FOX2 binding clusters throughout the human genome. The median distance between clusters within protein-coding genes was 813 nt, whereas the median distance between randomly chosen regions of similar sizes was 7,978 nt (**Supplementary Fig. 5** online). This result demonstrated that true FOX2 binding loci are indeed distributed non-randomly in protein-coding genes.

To further group the clusters, we determined the reduction in cluster number as a function of increasing window size. The number of clusters decreased eight-fold as the window size increased until the threshold of 1,500 nt was reached (**Supplementary Fig. 6** online). In contrast, the number of randomly chosen regions of similar sizes remained unaltered at any window size. Using this approach, we identified 3,547 combined clusters within the 1500-nt window, probably representing true FOX2 binding events, occurring either individually or in groups, in the human genome.

Having established grouped clusters, we next determined the motifs enriched in the clusters compared to randomly selected regions of similar sizes within the same protein-coding genes. Using *Z*-score statistics[10], we found that the most significantly enriched hexamer within the clusters was UGCAUG (*Z*-score of 25.16; *P*-value $< 10^{-70}$) (**Fig. 2a**), which exactly matched the biochemically defined consensus FOX1 and FOX2 binding site[11]. We next calculated the fractions of the grouped clusters that contained the consensus, observing that 1,052 (33%) and 704 (22%) of the FOX2 binding clusters harbored the GCAUG and UGCAUG motif, respectively, compared to 23% and 11%, respectively, of randomly located regions. Although this enrichment is highly significant (*P*-value $<10^{-10}$), the observation indicates that FOX2 did not bind to all available consensus-containing sequences and that FOX2 may also recognize other types of sequences in complex with other RNA-processing regulators. Consistent with the previously published bioinformatics analyses showing that composite functional RNA elements tend to be more evolutionarily conserved than other genomic regions that contain just the consensus[12–14], we found that 8% and 5% of FOX2 binding clusters contained one or more GCAUG and UGCAUG, respectively, that were perfectly conserved across four mammalian genomes (human, dog, mouse and rat). In contrast, only 2% (four-fold difference) and 1% (five-fold difference) of randomly selected pre-mRNA regions contained one or more perfectly conserved GCAUG and UGCAUG sites. These findings strongly suggest the functional importance of the FOX2 binding sites identified by CLIP-seq.

Preferential FOX2 action near alternative splice sites

To characterize the FOX2 binding profile relative to known splice sites, we found a median of 1.7 reads per kilobase of nucleotide sequence within protein-coding genes, with 13.5 reads per kilobase in exons, 2.2 reads per kilobase in introns, 0.3 reads per kilobase in promoters and 0.7 reads per kilobase in 3′ untranslated regions (UTRs). This observation suggests that FOX2 binds preferentially to exonic and intronic regions, consistent with its function as a splicing regulator. We observed that FOX2 binding clusters were 20-fold more likely to lie within exons and flanking intronic regions relative to randomly selected regions in the same protein-coding genes (**Fig. 2b**). This enrichment decreases to the background level ~3 kb away from the exons. Notably, the FOX2 binding sites were significantly (*P*-value $<$ 0.001) enriched in the downstream intronic region ~50–100 nt from the 5′ splice site, consistent with several characterized FOX2 binding sites in regulated splicing[7]. The FOX2 binding sites were also enriched in the upstream intronic regions near the 3′ splice site, but at a level 2.5-fold to 3-fold lower than in the downstream regions (**Fig. 2b**).

Preferential FOX2 binding to intronic regions near both 3′ and 5′ splice sites supports a crucial role of FOX2 in splice-site selection. Previous studies showed that intronic regions flanking alternatively spliced exons are more conserved than those flanking constitutive exons[15,16]. To determine whether FOX2 functions through conserved *cis*-acting regulatory RNA elements, we compared the association of mapped FOX2 binding clusters with constitutive and alternative splice sites and found that the highest enrichment occurred around alternative conserved exons (ACE) (**Fig. 2b**). Conversely, using Phastcons scores as a measure of evolutionary sequence conservation (Phastcons scores vary from 0 to 1, with 1 indicating high conservation)[17], we confirmed that FOX2-bound intronic regions flanking alternative exons were approximately two-fold more conserved than those flanking constitutive exons, and four- to seven-fold more conserved than other intronic regions containing randomly selected regions of similar sizes (**Fig. 2c**). These findings are fully consistent with existing examples of FOX2-regulated alternative splicing events[18], where high levels of flanking-sequence conservation were predictive of regulated splicing in mammalian cells[15,16].

FOX2 regulation of RNA targets

Overall, we identified FOX2 binding clusters in 1,876 protein-coding genes, suggesting that ~7% of human genes are subjected to FOX2 regulation in hESCs. To study the function of these FOX2 target genes, we performed gene ontology analysis[16], revealing a surprising enrichment for RNA binding proteins (P-value $< 10^{-8}$; **Supplementary Table 1** online). We also noted enrichment for nuclear mRNA splicing factors (P-value $< 10^{-5}$) and serine/threonine kinase activity (P-value $< 10^{-3}$). Among these FOX2 target genes were heterogeneous ribonucleoproteins (hnRNPs; for example, *A2/B1*, *H1*, *H2*, *PTB* and *R*), known alternative splicing regulators (for example, *A2BP1*, *PTB*, *nPTB*, *QKI*, *SFRS3*, *SFRS5*, *SFRS6*, *SFRS11* and *TRA2A*) and RNA binding proteins important for stem-cell biology (for example, *LIN28* and *MSI2*). This observation suggests that FOX2 may have a crucial role in establishing and maintaining the splicing and signaling programs in hESCs.

Figure 3 presents four examples of FOX2 RNA targets. A total of 962 CLIP-seq reads were localized within the fibroblast growth factor receptor (FGFR) gene *FGFR2* (**Fig. 3a**), which is known to be subject to FOX2 regulation[19]. A substantial number (103) of FOX2 CLIP-seq reads were clustered around one of the mutually exclusive exons (exon 8, which is selected to produce FGFR7 or keratinocyte growth factor (KGF) in epithelial cells, whereas exon 9 is used to produce FGFR2 in fibroblasts). The mapped FOX2 binding sites coincide with three UGCAUG and two GCAUG sites that are conserved across humans, dogs, mice and rats.

We previously identified in the STE20-like kinase (*SLK*) gene a 93-nt alternative exon that was included in hESCs but excluded in differentiated cells or tissues[5]. We mapped a total of 495 FOX2 CLIP-seq reads around three conserved (U)GCAUG elements upstream of the alternative exon (**Fig. 3b**). Indeed, FOX2 knockdown resulted in exon skipping of the alternative exon.

A total of 2,563 CLIP-seq reads were mapped to the polypyrimidine tract binding protein 2 (*nPTB*) gene, which is crucial for many regulated splicing events in neurons[20,21]. We identified 15 FOX2 binding clusters that could be aggregated into four groups (**Fig. 3c**),

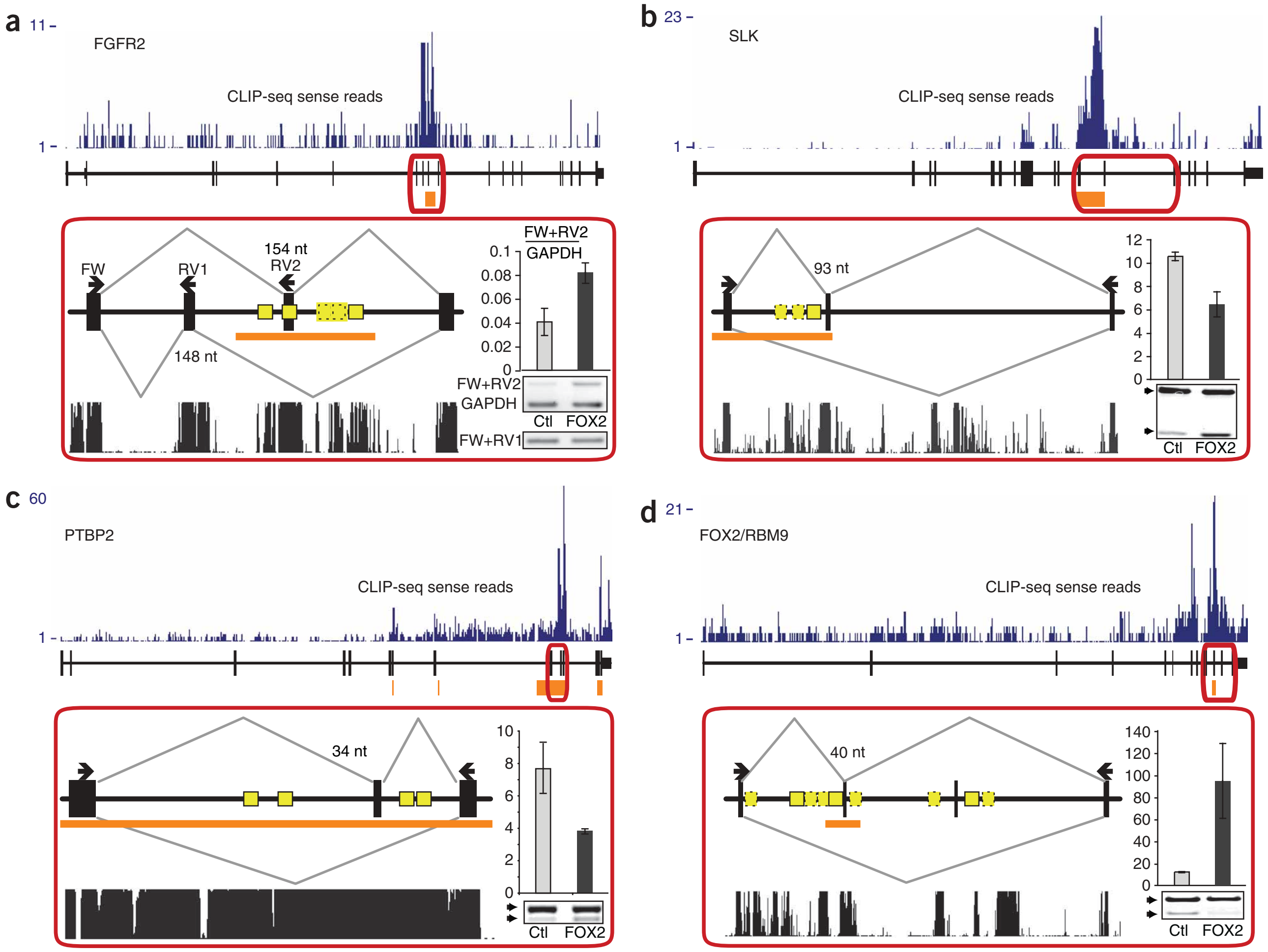
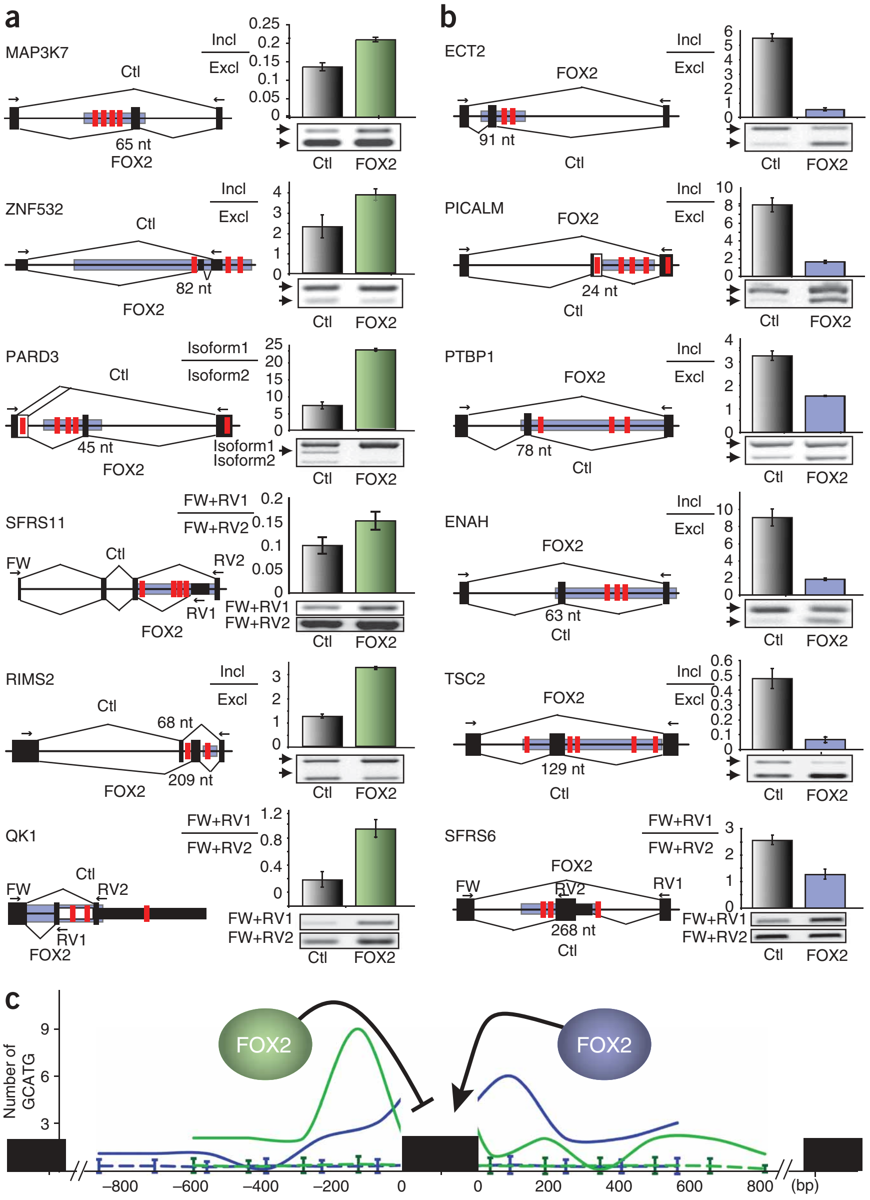


Figure 3 Clustering of FOX2 CLIP-seq reads around regulated splicing events. (**a**–**d**) The distribution of FOX2 CLIP-seq clusters in four examples of FOX2-regulated genes. The CLIP-seq reads are shown above each gene, with the *y* axis indicating the read density at each position. Each gene is diagrammed by vertical black bars (exons) and thin horizontal lines (introns), with arrows representing specific RT-PCR primers. Identified clusters are marked by horizontal orange bars. Exons encased by the red box in each case are illustrated in an expanded view below in which yellow boxes indicate the location of conserved GCAUG (dashed outlines) and UGCAUG (filled outlines) FOX2 binding motifs. Sequence conservation as measured by Phastcons scores is shown below. The insert in each expanded view shows RT-PCR analysis of alternative splicing in response to FOX2 knockdown by shRNA from triplicate experiments, with a representative gel image and s.d. indicated by error bars. FW, RV1 and RV2 represent forward and reverse primers. Ctl, control.

Figure 4 RNA map of FOX2-regulated alternative splicing. (**a**) FOX2-dependent exon skipping. (**b**) FOX2-dependent exon inclusion. Each gene is diagrammed by vertical black bars (exons) and thin horizontal lines (introns) with arrows representing specific RT-PCR primers. The conserved GCAUG FOX2 binding motifs (red vertical bars) generally overlap with mapped FOX2 binding sites by CLIP-seq (blue horizontal bars). Regulated splicing in control (Ctl) shRNA– and FOX2 (FOX2) shRNA–treated hESCs was analyzed by RT-PCR in triplicate, and s.d. is indicated by error bars. Changes in alternative splicing were significant in all cases, as determined by the Student's *t*-test (*P*-value < 0.05). (**c**) Number of conserved GCAUG sites proximal to the RT-PCR–validated FOX2-regulated alternative splicing, showing that conserved FOX2 binding motifs upstream or downstream of the alternative exon correlate with FOX2-dependent exon skipping (green) or inclusion (blue), respectively. Dashed lines and error bars indicate average number and s.d. of GCAUG sites in 100 independent versions of shuffled CLIP-seq binding sites.



one that contained 958 reads overlapping precisely with the known alternative exon and its flanking introns, which contain four UGCAUG elements in the ultraconserved introns[20,21]. Such dense FOX2 binding may indicate an unexpected mode of regulation, such as cooperative action, that cannot be explained by simple FOX2 recognition of its consensus binding sites.

In the fourth example, 1,576 reads were located on the FOX2 transcript itself, with 198 reads overlapping with six conserved (U)GCAUG elements proximal to the 40-nt alternative exon 11 (**Fig. 3d**), a finding that is consistent with the reported autoregulation of the gene crucial for homeostatic FOX2 expression[19,22]. These and many other examples (see below) show that FOX2 functions as a regulator of other splicing factors, including itself, in hESCs.

Exon inclusion and repression by FOX2 binding in hESCs

Our mapping results revealed preferential FOX2 binding to intronic regions either upstream or downstream or on both sides of the alternative exons. To determine the functional impact of these physical binding events, we selected 23 FOX2 target genes for functional validation in HUES6 cells treated with a lentivirus expressing a short hairpin RNA (shRNA) against FOX2 (**Fig. 4**). Western blotting 36 h after infection indicated specific downregulation of FOX2, relative to a control shRNA against enhanced green fluorescent protein (EGFP) (**Fig. 4**). RT-PCR analysis showed that FOX2 depletion indeed induced differential alternative splicing in 17 out of 23 (73%) tested genes (**Figs. 3** and **4** and **Supplementary Table 2** online).

Notably, we observed a general trend with respect to FOX2-regulated exon inclusion or skipping, depending on the location of FOX2 binding sites in the upstream or downstream intronic regions. Depletion of FOX2 tended to lead to exon inclusion if FOX2 binding sites were located in the upstream intron, as seen in *MAP3K7*, *ZNF532*, *PARD3* and *SFRS11* (**Fig. 4a**). In contrast, depletion of FOX2 resulted in exon skipping if FOX2 binding sites were located in the downstream intron, for instance, in *ECT2*, *PICALM*, *PTBP1* and *ENAH* (**Fig. 4b**). In several cases, such as in *PTBP2* (*nPTB*), *TSC2*, *SFRS6* and *RIMS2*, FOX2 binding sites were present in both upstream and downstream introns (**Figs. 3** and **4**); here, depletion of FOX2 resulted in either exon skipping (*PTBP2*, *TSC2*, *SFRS6*) or inclusion (*RIMS2*), probably reflecting a dominant effect of one binding site over the other(s). Notably, we also observed FOX2-dependent alternative 3′ end formation in the *QK1* gene (**Fig. 4a**).

On the basis of the results from the experimental validation, we generated a general splicing model by compiling consensus FOX2 binding motifs that are associated with FOX2 depletion–induced exon inclusion (green) or skipping (blue) (**Fig. 4c**). Compared to shuffled versions of the regions bound by FOX2, we observed an enrichment of seven-fold to nine-fold of the conserved GCAUG motifs within 1,000 nt of the alternatively spliced exons (*P*-value < 0.01). In fact, this enrichment peaks at 29-fold ~100 nt upstream (repression) and

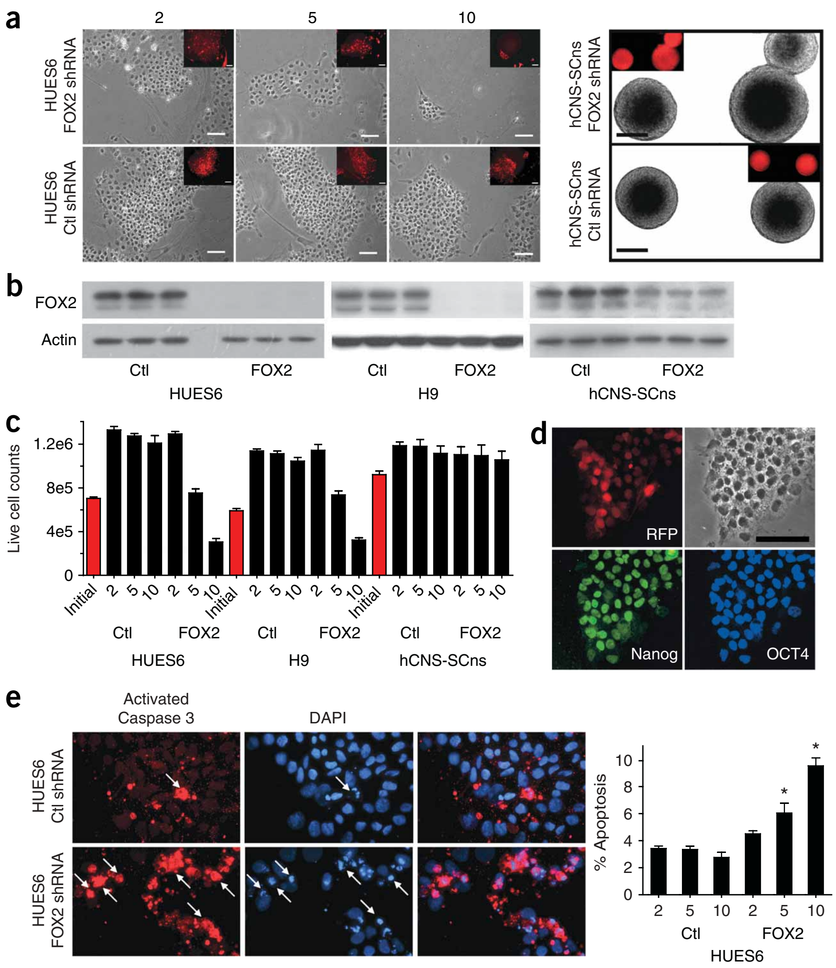


Figure 5 FOX2 is important for hESC survival. (**a**) Left, HUES6 hESC cells underwent rapid cell death in a dosage-dependent manner (2 μl, 5 μl and 10 μl) in response to FOX2 knockdown by lentiviral shRNA. Lentiviruses also expressed RFP, indicated in the inset, demonstrating the extent of infection. Right, similar infection of hCNS-SCns (grown as suspended neurospheres) did not result in a cell-death phenotype. Scale bars for hESC and hCNS-SCns are 25 μm and 200 μm, respectively. (**b**) Efficient FOX2 knockdown determined by western blotting was achieved in all cell types using actin as a loading control. (**c**) Cell counts using trypan blue exclusion indicates that the cell-death phenotype from FOX2 knockdown is specific to hESCs (HUES6 and H9 lines) and occurs in a dose-dependent fashion. (**d**) Infection of hESCs with FOX2 shRNA–RFP virus does not affect expression of pluripotency markers OCT4 and Nanog. (**e**) Knockdown of FOX2 in hESCs resulted in an increase in apoptotic cell death, as indicated by immunocytochemistry toward activated caspase-3 (left) and by FACs analysis using Yo-Pro (right) (* P-value < 0.001). Error bars indicate s.d.

21-fold ~100 nt downstream (activation) of the alternative exons (P-value < 0.001). This splicing model revealed a regulatory RNA map for FOX2 to activate or repress alternative splicing when bound downstream or upstream of the alternative exon, respectively. This RNA map is reminiscent of the trend observed with the neuronal specific splicing regulator Nova[23], suggesting a general splicing code for cell type–specific splicing regulators.

Notably, we observed that the alternative splicing patterns for a randomly selected subset of these exons were different in neural progenitors differentiated from HUES6 cells (HUES6-NP) that were FOX2 depleted by lentiviral shRNA, demonstrating that the splicing patterns were embryonic stem cell specific (**Supplementary Fig. 7** online). Furthermore, the splicing patterns in FOX2-depleted human fetal neural stem cells (human central nervous system stem cells propagated as neurospheres, hCNS-SCns) were similar to FOX2-depleted HUE6-NP cells[24]. HCNS-SCns are primary fate-restricted neural progenitors that, similarly to HUES6-NP cells, also express FOX2, suggesting that upon neural differentiation the targets for FOX2 regulation will be altered. We conclude that our RNA targets of FOX2 identified by CLIP-seq are specific for embryonic stem cells.

FOX2 is an important gene for hESC survival

During the investigation of FOX2 knockdown–induced alternative splicing, we were surprised to observe a rapid cell-death phenotype in response to FOX2 depletion in a dose-dependent manner (**Fig. 5a,b**). We observed the same phenotype using two independent lentiviral shRNAs on two independent hESC lines—HUES6 (**Fig. 5a**) and H9 (**Supplementary Fig. 8** online). In contrast, we knocked down FOX2 in hCNS-SCns and observed no effect on cell survival, with the caveat that the knockdown in hCNS-SCns was of slightly lower efficiency owing to decreased infection in neurospheres[24] (**Fig. 5a,b**). However, knockdowns in embryonic stem cells at comparable efficiencies to hCNS-SCns still recapitulate the cell-death phenotype. Live cell counts using trypan blue exclusion indicated cell death in a dose-dependent fashion, exclusively with FOX2 depletion in hESCs, but not hCNS-SCns (**Fig. 5c**). Furthermore, FOX2 depletion did not affect expression of pluripotency markers in both HUES6 (**Fig. 5d**) and H9 hESC lines (**Supplementary Fig. 9** online). Additionally, knockdown of FOX2 in transformed cell types such as 3T3 and HEK293T cells also did not affect cell viability (**Supplementary Fig. 8**), suggesting overall that FOX2 is selectively required for hESC survival.

To determine the possible cause of cell death, we stained HUES6 cells with the monomeric cyanine dye green fluorescent Yo-Pro-1, a marker of early apoptosis correlated with Annexin V staining[25]. Flow cytometry indicated that a statistically significant portion of FOX2-depleted, but not mock-depleted, cells committed apoptosis in a dosage-dependent manner (Student's t-test, P-value < 0.001) (**Fig. 5e**). This apoptotic death was confirmed by immunocytochemistry for activated caspase 3 (**Fig. 5e**). We also detected the upregulation of numerous genes involved in the necrosis pathway (**Supplementary Fig. 10** online). Together, these results indicate that FOX2-deficient cells underwent both apoptosis and necrosis, independently of cell-cycle arrest (**Supplementary Fig. 11** online).

DISCUSSION

Post-transcriptional gene expression regulation is crucial for many diverse cellular processes, such as development, metabolism and cancer. The fate of hundreds of thousands of mRNA molecules in eukaryotic cells is likely to be coordinated and regulated by hundreds of RNA binding proteins and noncoding RNAs (for example, microRNAs). To shed light on the importance and roles of individual RNA binding proteins, it is necessary to identify the spectrum of targets recognized and associated with these RNA binding proteins. Genome-wide unbiased methods have begun to reveal the plethora of targets and diverse rules by which the post-transcriptional regulatory networks are controlled[26].

Here we have identified the splicing factor FOX2 as being highly expressed in the nuclei of pluripotent hESCs. hESCs constitute an excellent *in vitro* model for survival, self-renewal, differentiation and development. Using a modified CLIP-seq technology and computational analyses that accounted for gene-specific variation in RNA abundance, we have uncovered thousands of FOX2 RNA targets representing ~7% of the human genes in hESCs. Confirming and extending previous computational analyses of human intronic regions[14–16,18], we observed that FOX2 was preferentially bound near alternative splice sites, and the binding sites were located within regions of higher evolutionarily conservation. Experimental validation of targets revealed that FOX2 represses exon usage when bound upstream and enhances exon inclusion when located downstream of the alternative exon, revealing an RNA map for the FOX2-mediated alternative splicing program in hESCs. Our study presenting *in vivo* targets of FOX2 in a biological system strengthens computational predictions from otherwise indistinguishable conserved FOX1 and FOX2 sites[27], as both FOX1 and FOX2 recognize the same RNA element[6,11,28]. The fact that FOX2 is also expressed in differentiated neural progenitors from hESCs and fetal neural stem cells but was not shown to regulate alternative splicing the same way in hESCs, despite having conserved binding sites in the same transcribed pre-mRNA, underscores the importance of experimentally identifying *in vivo* targets in the appropriate cell and tissue context.

The finding that many FOX2 targets are themselves splicing regulators leads to the provocative possibility that FOX2 may function as an upstream regulator of many general and tissue-specific splicing regulators. In addition, we identified FOX2 binding within the FOX2 pre-mRNA itself and, combined with RT-PCR data, demonstrated direct evidence for autoregulation of the *FOX2* gene. The alternative splicing of the FOX2 pre-mRNA may result in unique target pre-mRNA splicing regulation; this possibility deserves further attention in the future.

Last, our preliminary results indicate that FOX2 has an important role in maintaining the viability of hESCs, as depletion of FOX2 led to rapid cell death. Given the many genes controlled by FOX2 in hESCs, it is presently unclear which gene(s) or alternative splicing event(s) is responsible for the lethal phenotype. It is possible that the phenotype is a result of the combined effect of multiple affected genes. Given our observation that FOX2 may function as a master regulator of the alternative splicing program in hESCs and signaling pathways, it may be likely that many events contribute to the phenotype, that is, it may be unrealistic to think that the complex cellular mortality phenotype could be due to a single altered gene product. Nevertheless, the phenotype is remarkably specific to hESCs, and not other cell lines such as 293T or 3T3. More notably, neither neural progenitors derived from hESCs nor primary human fetal neural stem cells were sensitive to FOX2 depletion, suggesting that FOX2 has a different set of targets and, hence, a dissimilar RNA map in other cell types. Our study provides a starting point for the future characterization of the varying target repertoire of the same splicing factor in different biological systems, embracing a need to understand the uniqueness of factor-target relationships throughout biology.

METHODS

Culturing and differentiation of hESCs. We cultured hESC lines HUES6 and H9 as previously described (http://www.mcb.harvard.edu/melton/HUES/)[5]. Briefly, we grew cells on growth factor–reduced (GFR) matrigel–coated plates (BD) in mouse embryonic fibroblast–conditioned medium and FGF2 (20 ng ml^{-1}) in DMEM media (Invitrogen) supplemented with 20% (v/v) Knock Out serum replacement (GIBCO), 1 mM L-glutamine, 50 μM β-mercaptoethanol, 0.1 mM nonessential amino acids (Invitrogen) and 10 ng ml^{-1} FGF2 (R&D Systems), and passaged by manual dissection.

Neural progenitors were derived from hESCs as previously described[5]. Briefly, colonies were removed by treatment with collagenase IV (Sigma) and resuspended in media without FGF2 in nonadherent plates to form embryoid bodies. After 1 week, embryonic bodies were plated on polyornathine/laminin-coated plates in DMEM/F12 supplemented with N2 (1×) and FGF2. Rosette structures were manually dissected and enzymatically dissociated with TryPLE (Invitrogen), plated on polyornathine/laminin-coated plates and grown in DMEM/F12 supplemented with N2, B27 without retinoic acid and 20 ng ml^{-1} FGF2. Progenitors were verified by neuronal differentiation using 20 ng ml^{-1} brain-derived and glial-derived neurotrophic factors (BDNF and GDNF).

Lentiviral short hairpin RNA–mediated knockdown of FOX2. We purchased lentiviral shRNAs constructs toward FOX2 from Open Biosystems in the pLKO.1 vector system (TRCN0000074545 and TRCN0000074546). The control virus used was pLKO.1 containing a shRNA toward GFP (Open Biosystems). Lentivirus production was as previously described[29]. The efficacy of the lentivirus was tested by infection of HUES6 hESCs at varying viral concentrations and subsequent western blotting 36 h after infection with an antibody to FOX2 (1:1,000, Bethyl Laboratories) and actin as a control (1:5,000, Sigma). The FOX2, control and a GFP lentivirus were all made in parallel and concentrated by ultracentrifugation. GFP virus was titered using serial dilutions and infection of HEK293T cells. At 3 d after infection, we analyzed the cells for GFP expression by FACS and determined the viral titer using multiple dilutions, which yielded infections in the linear range. Titers were between 1×10^9 and 3×10^9. We used matched FOX2 and control viruses for hESC infections. Additionally, red fluorescent protein (RFP) was cloned into the PLKO.1 and 74546 lentiviral backbone in place of the puromycin-resistance gene and used in some studies to verify titer and comparable infection rates between the two lentiviruses.

Analysis of cross-linking immunoprecipitation reads. The human genome sequence (hg17) and annotations for protein-coding genes were obtained from the University of California, Santa Cruz Genome Browser. Known human genes (knownGene containing 43,401 entries) and known isoforms (knownIsoforms containing 43,286 entries in 21,397 unique isoform clusters) with annotated exon alignments to the human hg17 genomic sequence were processed as follows. Known genes that were mapped to different isoform clusters were discarded. All mRNAs aligned to hg17 that were greater than 300 nt were clustered together with the known isoforms. For the purpose of inferring alternative splicing, genes containing fewer than three exons were removed from further consideration. A total of 2.7 million spliced ESTs were mapped onto the 17,478 high-quality gene clusters to identify alternative splicing. To eliminate redundancies in this analysis, final annotated gene regions were clustered together so that any overlapping portion of these databases was defined by a single genomic position. To determine the number of reads that was contained within protein-coding genes, promoters and intergenic regions, we arbitrarily defined promoter regions as 3 kb upstream of the transcriptional start site of the gene and intergenic regions as unannotated regions in the genome. To identify CLIP clusters, we performed the following steps: (i) CLIP reads were associated with protein-coding genes as defined by the region from the annotated transcriptional start to the end of each gene locus. (ii) CLIP reads were separated into the categories of sense or antisense to the transcriptional direction of the gene. (ii) Sense CLIP reads were extended by 100 nt in the 5′-to-3′ direction. The height of each nucleotide position is the number of reads that overlap that position. (iv) The count distribution of heights is as follows from 1, 2, …h, …H-1, H: $\{n_1, n_2, \ldots n_h, \ldots n_{H\text{-}1,}\ n_H;\ N = \Sigma\ n_i\ (i = 1{:}H)\}$. For a particular height, h, the associated probability of observing a height of at least h is $P_h = \Sigma\ n_i\ (i = h{:}H)\ /\ N$. (v) We computed the background frequency after randomly placing the same number of extended reads within the gene for 100 iterations. This controls for the length of the gene and the number of reads. For each iteration, the count distribution and probabilities for the randomly placed reads ($P_{h,\text{random}}$) was generated as in step (iv). (vi) Our modified FDR for a peak height was computed as FDR(h) = $(\mu_h + \sigma_h)/P_h$, where μ_h and σ_h is the average and s.d., respectively, of $P_{h,\text{random}}$ across the 100 iterations. For each gene loci, we chose a threshold peak height h^* as the smallest height equivalent to FDR(h^*) $<$ 0.001. We identified FOX2 binding clusters by grouping nucleotide positions satisfying $h > h^*$ and occurred within 50 nt of each other. This resulted in 6,123 FOX2 binding clusters. This number varied slightly when repeated for different sets of iterations. As a control for authentic FOX2 clusters, artificial randomly located regions were generated as follows. For each gene that contained one or more FOX2 binding clusters, we randomly picked the same number of regions of the same sizes as the FOX2 clusters in the pre-mRNA. Distances between clusters were measured from the 3′ end of a cluster to the 5′ end of the downstream cluster. Clusters were further grouped, as many clusters were closer than expected when compared to the randomly chosen regions. If a cluster was greater than 50 nt in length and within 1,500 nt to another cluster, we grouped that as a single cluster, resulting in 3,547 clusters.

Cell-cycle and cell-death analysis. We carried out apoptosis staining using Yo-Pro according to the manufacturer's instructions (Invitrogen). Gating for apoptotic cells was determined empirically using a negative control (no Yo-Pro) and a positive control (4-h treatment with 10 μM campthothecin). Cell-cycle staining was performed as previously described[30,31]. Briefly, cells were trypsinized, washed and resuspended in PBS, then fixed by the addition of a 3:1 ratio of ice-cold 100% (v/v) ethanol in PBS overnight at –20 °C. Subsequently, cells were washed and resuspended in a solution containing 50 μg ml^{-1} propidium iodide and 500ng ml^{-1} RNase A for 1 h at 37 °C before analysis by FACS on a Becton-Dickinson FACScan. Immunocytochemistry was performed using the activated caspase-3 antibody (Cell Signaling Technologies, 1:150).

Additional cell culture procedures and antibodies used, RNA extraction, RT-PCR, CLIP library construction and sequencing, Processing of 1G data and Genomic analysis are available in **Supplementary Methods** online.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

The authors would like to acknowledge S. Aigner for technical advice, J. Simon for illustration assistance and R. Keithley and B. Miller for cell culture. G.W.Y. was funded by a Junior Fellowship from the Crick-Jacobs Center for Theoretical and Computational Biology, Salk Institute. F.H.G. is funded by the California Institute of Regenerative Medicine, The Picower Foundation and the Lookout Foundation. Part of this work was supported by US National Institutes of Health grants to X.-D.F. (GM049369 and HG004659) and G.W.Y. (HG004659).

AUTHOR CONTRIBUTIONS

G.W.Y. directed the project; G.W.Y. and F.H.G. designed the project; G.W.Y., N.G.C. and X.-D.F. analyzed the data and wrote the manuscript; G.W.Y., N.G.C., T.Y.L. and G.E.P. performed the experiments; G.W.Y. and T.Y.L. carried out bioinformatics data analysis.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
2. Thomson, J.A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
3. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129–1155 (2005).
4. Sonntag, K.C., Simantov, R. & Isacson, O. Stem cells may reshape the prospect of Parkinson's disease therapy. *Brain Res. Mol. Brain Res.* **134**, 34–51 (2005).
5. Yeo, G.W. *et al.* Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLOS Comput. Biol.* **3**, e196 (2007).
6. Jin, Y. *et al.* A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* **22**, 905–912 (2003).
7. Underwood, J.G., Boutz, P.L., Dougherty, J.D., Stoilov, P. & Black, D.L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25**, 10005–10016 (2005).
8. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
9. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
10. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. & Burge, C.B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
11. Auweter, S.D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163–173 (2006).
12. Kabat, J.L. *et al.* Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLOS Comput. Biol.* **2**, e86 (2006).
13. Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences—the complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).
14. Yeo, G.W., Nostrand, E.L. & Liang, T.Y. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.* **3**, e85 (2007).
15. Sorek, R. & Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**, 1631–1637 (2003).
16. Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C.B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl. Acad. Sci. USA* **102**, 2850–2855 (2005).
17. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
18. Brudno, M. *et al.* Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing. *Nucleic Acids Res.* **29**, 2338–2348 (2001).
19. Baraniak, A.P., Chen, J.R. & Garcia-Blanco, M.A. Fox-2 mediates epithelial cell-specific fibroblast growth factor receptor 2 exon choice. *Mol. Cell. Biol.* **26**, 1209–1222 (2006).
20. Makeyev, E.V., Zhang, J., Carrasco, M.A. & Maniatis, T. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
21. Boutz, P.L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
22. Nakahata, S. & Kawamoto, S. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res.* **33**, 2078–2089 (2005).
23. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
24. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl. Acad. Sci. USA* **97**, 14720–14725 (2000).
25. Idziorek, T., Estaquier, J., De Bels, F. & Ameisen, J.C. YOPRO-1 permits cytofluorometric analysis of programmed cell death (apoptosis) without interfering with cell viability. *J. Immunol. Methods* **185**, 249–258 (1995).
26. Halbeisen, R.E., Galgano, A., Scherrer, T. & Gerber, A.P. Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell. Mol. Life Sci.* **65**, 798–813 (2008).
27. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
28. Ponthier, J.L. *et al.* Fox-2 splicing factor binds to a conserved intron motif to promote inclusion of protein 4.1R alternative exon 16. *J. Biol. Chem.* **281**, 12468–12474 (2006).
29. Singer, O. *et al.* Targeting BACE1 with siRNAs ameliorates Alzheimer disease neuropathology in a transgenic model. *Nat. Neurosci.* **8**, 1343–1349 (2005).
30. Crissman, H.A. & Steinkamp, J.A. Rapid, simultaneous measurement of DNA, protein, and cell volume in single cells from large mammalian cell populations. *J. Cell Biol.* **59**, 766–771 (1973).
31. Krishan, A. Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *J. Cell Biol.* **66**, 188–193 (1975).