Cell Reports, Volume *15*

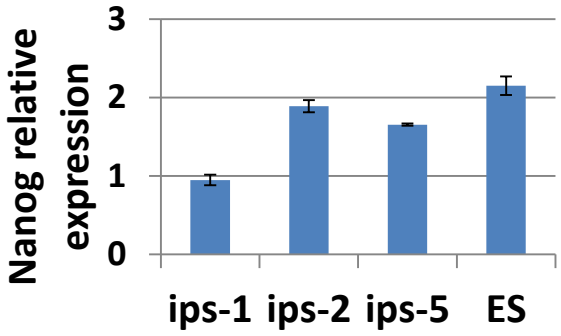
Supplemental Information

Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins

Benjamin Cieply, Juw Won Park, Angela Nakauka-Ddamba, Thomas W. Bebee, Yang Guo, Xuequn Shang, Christopher J. Lengner, Yi Xing, and Russ P. Carstens

Figure S1.

A.


Nanog relative expression
3
2
1
0
ips-1
ips-2
ips-5
ES

B.

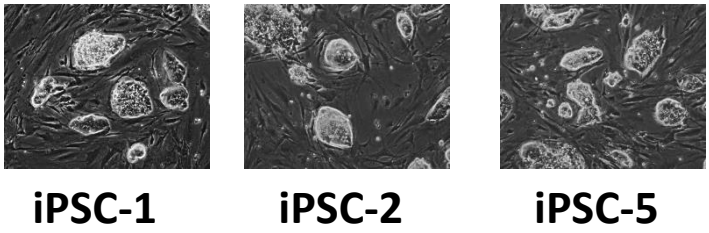

iPSC-1
iPSC-2
iPSC-5

Figure S2

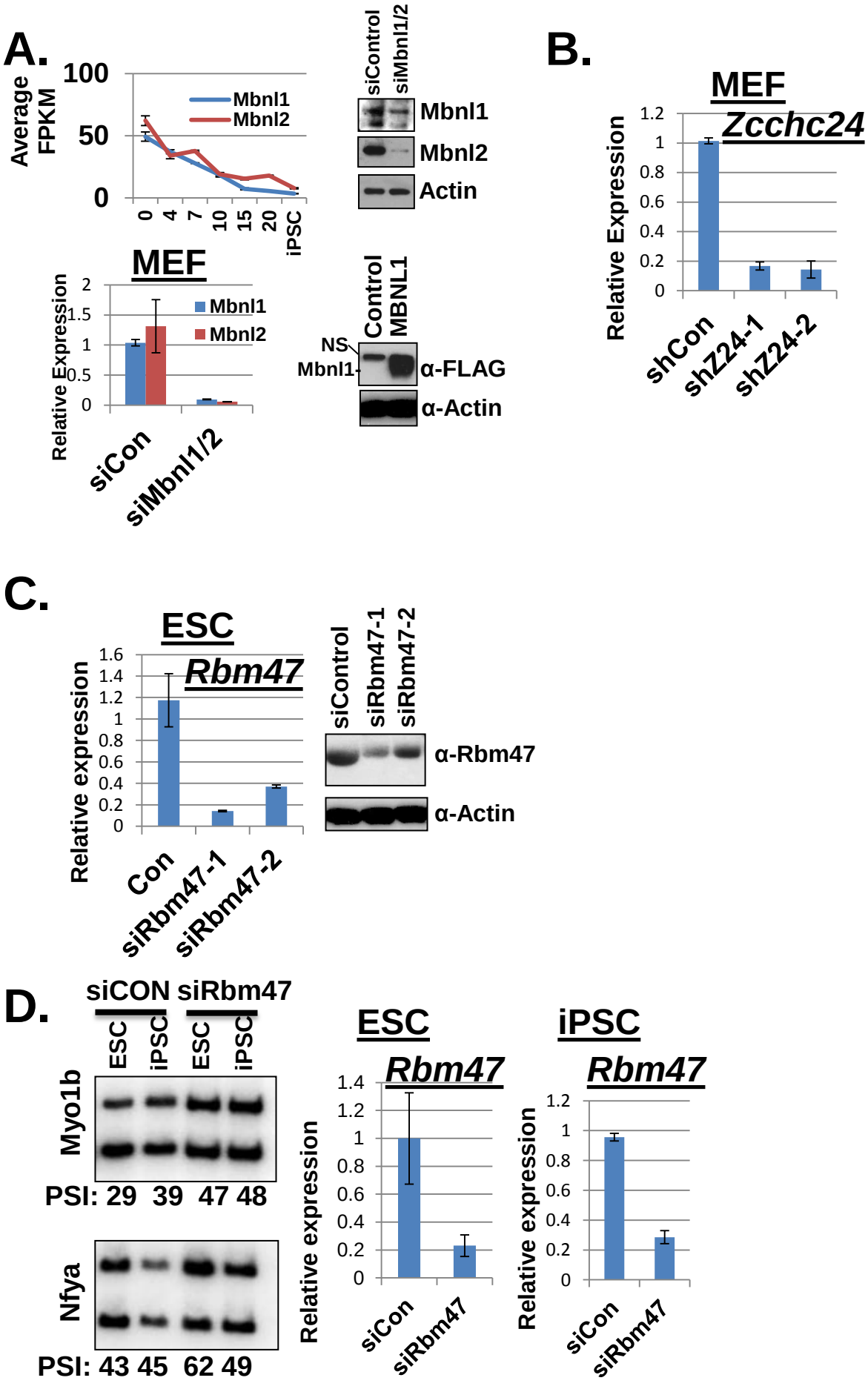
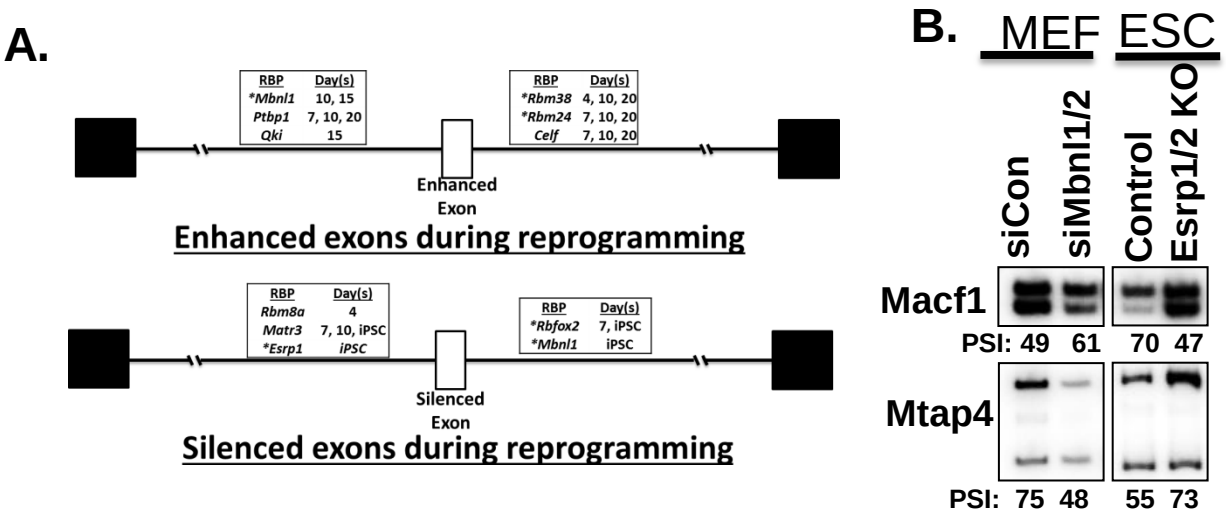
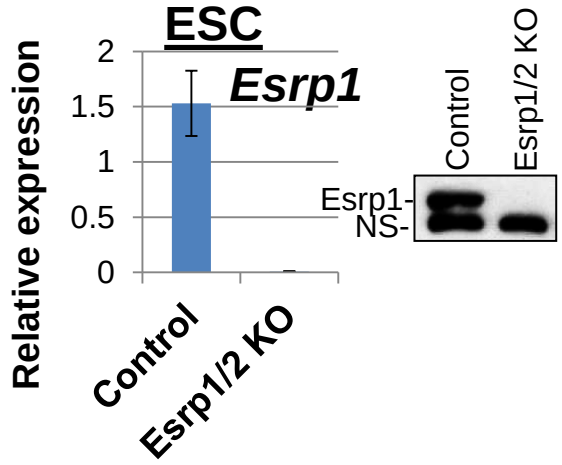

A.
Average FPKM
100
50
0
Mbnl1
Mbnl2
0 4 7 10 15 20 iPSC
siControl
siMbnl1/2
Mbnl1
Mbnl2
Actin
MEF
Relative Expression
2
1.5
1
0.5
0
Mbnl1
Mbnl2
siCon
siMbnl1/2
Control
MBNL1
NS
Mbnl1
α-FLAG
α-Actin
B.
MEF
Zcchc24
Relative Expression
1.2
1
0.8
0.6
0.4
0.2
0
shCon
shZ24-1
shZ24-2
C.
ESC
Rbm47
Relative expression
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Con
siRbm47-1
siRbm47-2
siControl
siRbm47-1
siRbm47-2
α-Rbm47
α-Actin
D.
siCON siRbm47
ESC
iPSC
ESC
iPSC
Myo1b
PSI: 29 39 47 48
Nfya
PSI: 43 45 62 49
ESC
Rbm47
Relative expression
1.4
1.2
1
0.8
0.6
0.4
0.2
0
siCon
siRbm47
iPSC
Rbm47
Relative expression
1.2
1
0.8
0.6
0.4
0.2
0
siCon
siRbm47

Figure S3.


A.
RBP Day(s)
*Mbnl1 10, 15
Ptbp1 7, 10, 20
Qki 15
RBP Day(s)
*Rbm38 4, 10, 20
*Rbm24 7, 10, 20
Celf 7, 10, 20
Enhanced Exon
Enhanced exons during reprogramming
RBP Day(s)
Rbm8a 4
Matr3 7, 10, iPSC
*Esrp1 iPSC
RBP Day(s)
*Rbfox2 7, iPSC
*Mbnl1 iPSC
Silenced Exon
Silenced exons during reprogramming
B.
MEF ESC
siCon
siMbnl1/2
Control
Esrp1/2 KO
Macf1
PSI: 49 61 70 47
Mtap4
PSI: 75 48 55 73


C.
ESC
Esrp1
Relative expression
2
1.5
1
0.5
0
Control
Esrp1/2 KO
Control
Esrp1/2 KO
Esrp1-
NS-

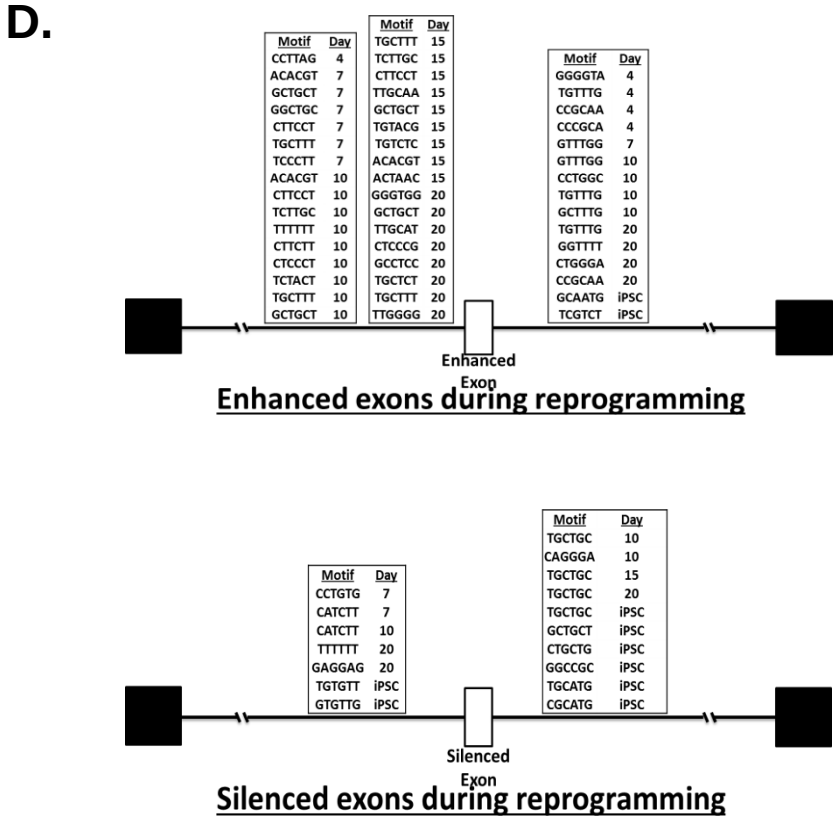
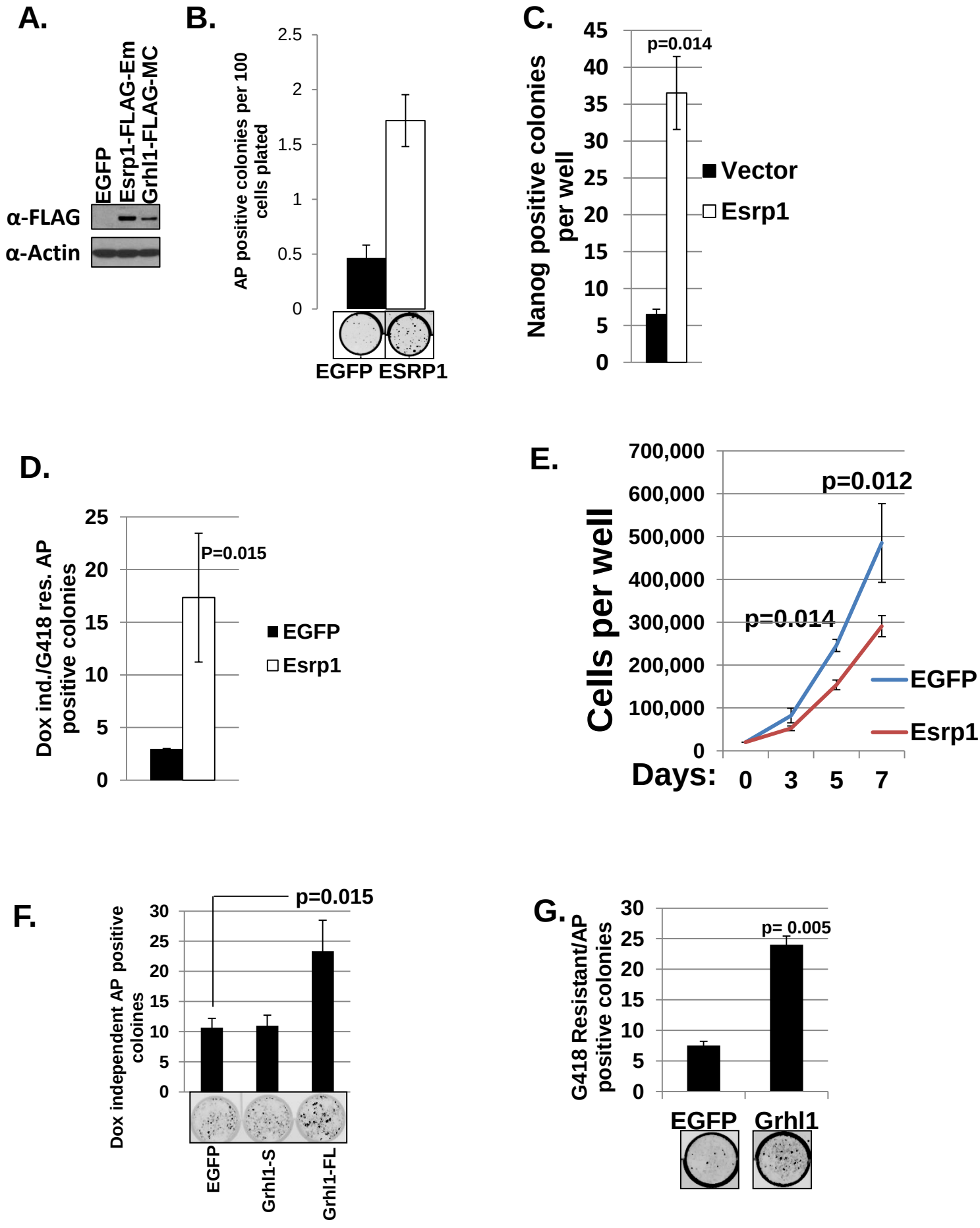

D.
Motif Day
CCTTAG 4
ACACGT 7
GCTGCT 7
GGCTGC 7
CTTCCT 7
TGCTTT 7
TCCCTT 7
ACACGT 10
CTTCCT 10
TCTTGC 10
TTTTTT 10
CTTCTT 10
CTCCCT 10
TCTACT 10
TGCTTT 10
GCTGCT 10
Motif Day
TGCTTT 15
TCTTGC 15
CTTCCT 15
TTGCAA 15
GCTGCT 15
TGTACG 15
TGTCTC 15
ACACGT 15
ACTAAC 15
GGGTGG 20
GCTGCT 20
TTGCAT 20
CTCCCG 20
GCCTCC 20
TGCTCT 20
TGCTTT 20
TTGGGG 20
Motif Day
GGGGTA 4
TGTTTG 4
CCGCAA 4
CCCGCA 4
GTTTGG 7
GTTTGG 10
CCTGGC 10
TGTTTG 10
GCTTTG 10
TGTTTG 20
GGTTTT 20
CTGGGA 20
CCGCAA 20
GCAATG iPSC
TCGTCT iPSC
Enhanced Exon
Enhanced exons during reprogramming
Motif Day
CCTGTG 7
CATCTT 7
CATCTT 10
TTTTTT 20
GAGGAG 20
TGTGTT iPSC
GTGTTG iPSC
Motif Day
TGCTGC 10
CAGGGA 10
TGCTGC 15
TGCTGC 20
TGCTGC iPSC
GCTGCT iPSC
CTGCTG iPSC
GGCCGC iPSC
TGCATG iPSC
CGCATG iPSC
Silenced Exon
Silenced exons during reprogramming

Figure S4.


A.
EGFP
Esrp1-FLAG-Em
Grhl1-FLAG-MC
α-FLAG
α-Actin
B.
AP positive colonies per 100 cells plated
2.5
2
1.5
1
0.5
0
EGFP ESRP1
C.
Nanog positive colonies per well
45
40
35
30
25
20
15
10
5
0
p=0.014
Vector
Esrp1
D.
Dox ind./G418 res. AP positive colonies
25
20
15
10
5
0
P=0.015
EGFP
Esrp1
E.
Cells per well
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
p=0.012
p=0.014
EGFP
Esrp1
Days:
0
3
5
7
F.
Dox independent AP positive coloines
30
25
20
15
10
5
0
p=0.015
EGFP
Grhl1-S
Grhl1-FL
G.
G418 Resistant/AP positive colonies
30
25
20
15
10
5
0
p= 0.005
EGFP
Grhl1

Supplemental Figure and Table Legends:

Figure S1, Related to Figure 1: A) *Nanog* qRT-PCR in the three iPSC clones conducted in technical triplicate that were used for RNA sequencing as well ESCs. B) Phase contrast images for each iPSC clone.

Figure S2, Related to Figure 2: A) Average FPKM of Mbnl1 and 2 during reprogramming timecourse (upper left); qRT-PCR for *Mbnl1* and *Mbnl2* in MEFs after control or *Mbnl1* and *Mbnl2* targeting siRNA transfections (lower left) and western blot (upper right); FLAG western blot demonstrating expression of MBNL1-2x-FLAG-Emerald-GFP (or Actin) in ES cells after transfection (lower right). B) qRT-PCR analysis of *Zcchc24* in MEFs after transduction with control or two different *Zcchc24* targeting shRNA's. C) qRT-PCR for *Rbm47* in ES cells after transfection with control or two different *Rbm47* targeting siRNAs (left graph) and western blot (right). D) si*Rbm47* pool transfected into ESCs or iPSCs followed by RT-PCR to assess splicing changes in Myo1b and Nfya (left) and qRT-PCR to validate knock down (right graphs). Error bars for all qRT-PCRs represent SDM of technical duplicate.

Figure S3, Related to Figure 2: A) Enrichment for known RBP motifs in the introns up or downstream of exons that change during reprogramming. The reprogramming timpoint(s) where the enrichment was identified is listed to the right of each RBP gene symbol. (*change in expression by at least 2 fold). B) RT-PCR analysis of two AS events that are reciprocally regulated by Mbnl1/2 knock down in MEFs or Esrp1/2 knock out in ES cells (the Macf1 results are also shown in figure 2B and 3B). C) Control or Esrp1/2 knock out ES cells assayed for Esrp1 expression via qRT-PCR (left graph) or western blot (right), NS is non- specific band. D) Unbiased analysis of all 6mers enriched near exons that change during reprogramming.

Figure S4, Related to Figure 3 and 4: A) Western blot analysis of protein from MEFs that were infected with EGFP, Esrp1-FLAG-Emerald, or Grhl1-FLAG-mCherry. B) Dox reprogrammable MEFs transduced with EGFP or Esrp1-emerald retrovirus and assayed for AP positive colonies after 10 days of dox treatment conducted in biological triplicate. C) Dox reprogrammable MEFs transduced with either empty vector or Esrp1-2xFLAG and then assayed for Nanog positive colonies after 12 days of dox treatment via immunofluorescence conducted in biological duplicate. D) Dox reprogrammable MEFs transduced with either EGFP or Esrp1-emerald and treated with dox for 10 days upon which dox was removed and G418 was added and grown for an additional 10 days then fixed and stained for AP, conducted in biological triplicate. E) Doxycycline reprogrammable MEFs were infected with EGFP or Esrp1-Emerald and assayed, in the presence of Dox, for cell proliferation by cell counting over 7 days in triplicate. F) Dox reprogrammable MEFs transduced with either EGFP, Grhl1-S, or Grhl1-FL and treated with Dox for 10 days upon which it was removed and then AP positive colonies were quantifed 7 days later, conducted in biological triplicate. G) Dox reprogrammable MEFs transduced with either EGFP or Grhl1-FL and treated with dox for 10 days upon which G418 was added; then fixed and stained for AP 10 days later, conducted in biological duplicate. All p-values calculated by students t-test. Error bars represent SDM in all experiments shown.

Table S1, Related to Figure 1: All cassette exons with significant changes in alternative splicing during reprogramming identified by rMATS as having an FDR of less than 5% and delta percent spliced in (PSI) of at least 5% at one or more timepoints relative to time zero. ID in heat map, gene ID, gene symbol and exon coordinates are indicated in columns A-D; iPS-0 is MEFs; iPS-4, iPS-7, iPS-10, iPS-15, iPS-20 are 4, 7, 10, 15, and 20 days of Dox treatment respectively and iPSC are transgene independent clones. SC is skipped counts; IC is inclusion counts (columns E-BF); PSI of triplicates for each time point (Columns BG-CB).

Table S2, Related to Figure 1 and 2: Tab A) List of cassette exons arranged in temporal clusters based on changes in alternative splicing; minimum of 5 exons per cluster. The cluster ID is listed in column A. Tab B) List of RNA binding proteins arranged in temporal clusters based on expression changes. The cluster ID is listed in column A. Tab C) Correlation coefficients between alternative exons and RBP expression changes across the time course.

Table S3, Related to Figure 2: FPKM of all mouse genes during reprogramming in triplicate; gene ID, gene symbol, and genomic locus are indicated in columns A-C; iPS-0 is MEFs; iPS-4, iPS-7, iPS-10, iPS-15, iPS-20 are 4, 7, 10, 15, and 20 days of Dox treatment respectively and iPSC are transgene independent clones.

Table S4, Related to Figure 3: MATS analysis of Esrp1 ^{WT}/Esrp2 $^{-/-}$ vs. Esrp1 $^{flox/flox}$/Esrp2 $^{-/-}$ ES cells following transfection with Cre to mediate ablation of floxed Esrp1 conducted in ES cells derived from 3 independent blastocysts per group.

Supplemental Experimental Procedures

Cell culture

Mouse embryo fibroblasts were isolated at E12.5 by mincing, incubation with 0.25% Trypsin/EDTA for 20 minutes followed by trituration. They were cultured at 5% CO_2 and 5% O_2 in MEF media (DMEM, 10% FBS, non essential amino acids (Gibco 11140-050); Pen/Strep (Gibco15140-122) 1mM sodium pyruvate (Gibco 11360-070); 0.1 mM beta-mercaptoethanol (Gibco 21985-023). Embryonic stem cells were isolated from E3.5 blastocysts and maintained in ES-media (Knock-Out DMEM (Gibco); 10% FBS ES tested; nonessential amino acids (Gibco 11140-050); Pen/Strep (Gibco15140-122); 0.1 mM beta-mercaptoethanol (Gibco 21985-023) L-Glutamine (Gibco 25030-081); ESGRO-LIF (Millipore)) on irradiated feeder MEFs. Ablation of the Esrp1 floxed region (an Esrp1 allele with LoxP sites flanking exons 7, 8 and 9) (Bebee et al., 2015) was mediated by pLVX-Cre-EGFP-IRES-PURO transfection using X-Fect-ES Reagent (Clonetech).

Antibodies

FLAG-M2, mouse, Sigma; Nanog, rabbit, Bethyl; Rbm47, rabbit, Sigma; Esrp1-27H12, mouse, Rockland; Mbnl1 and Mbnl2 mouse, Santa Cruz Biotechnology.

Generation of expression vectors for Grhl1 and MBNL1

Grhl1 cDNA clone MmCD00312394 was obtained from the Harvard PlasmID Database and amplified using primers: Grhl1-Topo-F: CACCATGACACAGGAGTACGACAA; Grhl1-Topo-R: GGCAATCTCTGTCAGGGTGAGCT and cloned into pENTR-D-Topo (Life Technologies) according to the manufacturer's protocol. The cDNA was then transferred into pMXS-CFFB-mCherry-IRES-Blast2-Gateway or pMXS-CFFB-IRES-Puro-Gateway via Gateway cloning to express the Grhl1-2x-FLAG or Grhl1-2x-FLAG-mCherry fusion-proteins. The Grhl1-Short isoform was generated using primers Grhl1-Topo-F and Grhl1-ex5-Skip-R: GGCAAGTTGTTCCTCCTGAACGCCCTCCTTGA in the same manner.

MBNL1 cDNA clone, accession BC043493 was cloned into the ECORV and NOT1 sites of pIPX-CFFB-Emerald-GFP.

RNA isolation, RNA-sequencing and RT-PCR

RNAs were harvested using Trizol reagent (Life Technologies). cDNA libraries for RNA-sequencing were prepared with poly-A purified mRNAs and the TruSeq™ Stranded mRNA LT Sample Prep Kit (Illumina) according to the manufacturers protocol . We generated ~734 million 101x2 bp paired-end and ~335 million101 bp single-end RNA-seq reads on a HiSeq 2000 sequencer. The RNA-seq data has been deposited into the NCBI Gene Expression Omnibus under the accession number GSE70022.

Quantitative RT-PCR was conducted using either Taqman for *Esrp1* normalized to *Gapdh* or Sybergreen for *Zcchc24*, *Mbnl1*, *Mbnl2*, *Rbm47* and *Nanog* normalized to *B2m* or *Gapdh*. For semi-quantitative radioactive labeled RT-PCR of alternative splicing of cassette exons, primers were designed in the constitutive exons flanking the alternative exon and amplicons that represented the included (larger) or skipped(smaller) form were resolved using 5% polyacrylamide gel electrophoresis. Gels were dried and scanned on a Typhoon FLA9500; band intensity was quantified using Image Quant.

RNA-sequencing analysis

We first mapped RNA-seq reads to the mouse genome (mm10) and transcriptome (Ensembl, release 72) using the software TopHat (v1.4.1) allowing up to 3 bp mismatches per read and up to 2 bp mismatches per 25 bp seed. We mapped single-end and paired-end reads separately then merged the mapping results for the downstream analysis. We used Cuffdiff (v2.2.0) to calculate RNA-seq based gene expression levels using the FPKM metric (fragments per kilobase of exon per million fragments mapped) then identified differential gene expression between the two time points at FDR<5%, >2 fold difference in gene expression based on average FPKM, and minFPKM>0.1 . To identify differential AS events between DAY0 and other time points (DAY4 through DAY20 and iPS clone), we used a modified version of rMATS v3.0.8 (http://rnaseq-mats.sourceforge.net) to identify differential AS events

from strand-specific RNA-seq data corresponding to all five basic types of AS patterns. Briefly, rMATS uses a modified version of the generalized linear mixed model to detect differential AS from RNA-seq data with replicates (Shen et al., 2014). It accounts for exon-specific sequencing coverage in individual samples as well as variation in exon splicing levels among replicates. For each AS event, we used both the reads mapped to the splice junctions and the reads mapped to the exon body as the input for rMATS.

Temporal cluster analysis of time-course iPSC RNA-seq data

We compiled a list of 226 genes encoding RNA binding proteins (RBPs) involved in splicing regulation. From this list, we identified 95 RBPs that met the following three criteria: (1) expressed in at least one of the 7 time points (average FPKM gene expression value > 5.0); (2) significant change in FPKM values across the 7 time points (ANOVA $p<0.01$); (3) at least two-fold change in FPKM values between the two time points with the highest and lowest average expression levels. We considered these 95 RBPs as candidate RBPs that may drive alternative splicing changes during iPSC reprogramming. To separate these 95 RBPs into subgroups representing distinct temporal patterns of gene expression during iPSC reprogramming, we performed cluster analysis of the time-course RNA-seq data followed by cutting of the cluster dendrogram at the appropriate height. Specifically, for every pair of RBPs we calculated the Jackknife version of Pearson correlation coefficient using their expression levels across the entire time course (7 time points x 3 replicates). Using a permutation procedure, we estimated an FDR of <0.01 at a correlation coefficient threshold of 0.5. We then cut the clustering dendrogram at a height of 0.5 (i.e. 1- correlation coefficient threshold of 0.5) and separated the 95 RBPs into 9 distinct clusters. Using a similar approach, we performed temporal cluster analysis of differential exon skipping events during iPSC reprogramming identifying 284 exons in 18 clusters, at a Jackknife correlation coefficient threshold of 0.5 and a minimum of 5 exons per cluster (FDR=0.02 based on permutation test).

Motif Enrichment Analysis

We sought to identify binding sites of splicing factors and other RNA binding proteins that were significantly enriched in differential exon skipping events between two time points as compared to control (non-regulated) alternative exons. We collected 115 known binding sites of RNA binding proteins including many well-characterized splicing factors from the literature (Ray et al., 2013). For each motif, we scanned for motif occurrences separately in exons or their 250 bp upstream or downstream intronic sequences. For intronic sequences, we excluded the 20 bp sequence within the 3' splice site and the 6 bp sequence within the 5' splice site. Alternative exons without splicing changes (rMATS FDR>50%, maxPSI>15%, minPSI<85%) in highly expressed genes (average FPKM>5.0 in at least one time point) were treated as control exons. For each motif, after we counted the number of occurrences in the differentially spliced exons and the control exons, we calculated p-value for motif enrichment via the Fisher's exact test (right-sided) and used Benjamini-Hochberg FDR correction to adjust for multiple testing for exons, upstream intronic sequences, and downstream intronic sequences separately. A motif can be counted multiple times in a given sequence (by-NT) or a motif can only be counted zero or one time in a given sequence (by-Sequence). We conducted both motif counting approaches and identified enriched motifs at FDR(by-NT)<5% and p-value(by-Sequence)<0.01.

Esrp RNA map anaylsis

We used top twelve GU-rich Esrp binding sites previously identified by the SELEX-Seq (Dittmar et al., 2012) to identify the RNA binding map of the Esrps for the differential SE events between two time points as compared to control alternative exons. We assigned motif scores based on the overall percentage of nucleotides covered by any of twelve Esrp binding sites within a 50-nt window. We slid this window by 1-nt across the exon body and 250 nt of upstream and downstream intron as well as the upstream and downstream exons.

RT-PCR primers flanking alternative exons:

Sequence Name	Sequence
Azi2-F	CTG GGA ACT GAA GAG GGA GAT G
Azi2-R	TTT AGG AGG GCT TGG GAA TAC C
Bc037034-F	TGC CCT GGA GGA ACA CAA CT

Bc037034-r	TCC GGG AGA AGC CAA GAT TAT
Fgfr1-F	GGG ACC GCA GCG CCA AGT GAG AGT
Fgfr1-R	AGC TAC AGG CCT ACG GTT TGG TTT
Map3k4-F	AGT TTG CAA GCC CTG ATG AAT GAG
Map3k4-R	CGC ACA CTT GGC CGA TGA TA
Tead1-F	GCC GTG TGG GAG GAG AAA AAT CAT
Tead1-R	TCC TGG CTG TCC TGT CTG TAT CAT
Mta1-F	GGT AGT ACG GAA GCC CCT GGA G
Mta1-R	CAT TCG CCG CCG CTT GAC TGG A
Map4-F	GAG GGG AAA CCT GCT GAT GTC AAA
Map4-R	CGC CCT CAG TCT TCA CGG CAC CTC
Nfya-F	GCT GCC TGG GAT CTG TAG AGT GAA
Nfya-R	AGC CTG CCC ACC CTG AAT CTG GAT
Ubap2l-F	GCC CCC TTT ATG CAT ATT CTG ACC
Ubap2l-R	ATC GCC CCA CAT TCT TTT CT
Kansl1-F	ACACCATGCCTCCTGAAATACACT
Kansl1-R	CTCTCCCCTTCTCCTCCTCACT
Exoc1-F	ACTCCGACTATATGAAAGAGAAAT
Exoc1-R	GGACGACTGGGACCTGCGACTG
Map3k7-F	TTCCTGCCACAAACGACAC
Map3k7-R	ATTCTGACACTAGGGCTGGATGAC
Myo1b-F	AAGCGGTACCAGCAGATAAAGAGT
Myo1b-R	AGTCGAATCCAAGAACAGGTAAGG
Lsm14b-F	CCTCGGCGACACAGCTCAATGGT
Lsm14b-R	ATCACAGCTGGGTCCTTCTCTTCC
Epb4.1-F	TTAACATCAACGGGCAAGTC
Epb4.1-R	ATATCGGCATCTCCTGTGA
Arhgef11-F	TCAAGCTGAGAACCAGCAGGAAGT
Arhgef11-R	TGCTCGATGGTGTGGAAGATGCACA
Grhl1-F	ACGGCTTTGCTGTGGGAATC
Grhl1-R	ACCTCTGTTTGGCCGTGTGCT
Fn1-F	GCAAGCCAGTTTCCATCAAT
Fn1-R	TCCTGTCTTCTCTTTCGGGTTCAC
Ssbp3-F	TATTTTGGGACCTTTACTGTGC
Ssbp3-R	CATGTTGGGGTGACCTTGTTGT
Spag9-F	ATTCCACGCTCCGCCTCTATCA
Spag9-R	ATCTGCCCCACCACTGCTACTTTG

qRT-PCR primers

Rbm47: F-TGAGGCCGTAGCGCAGCAACC; R-CCTCTCCCTCGGCCTCTTATG

Mbnl1: F-CGACCAGACACGGAATGTAA ;R-TCTCTGGAGCAACGACCTTT

Mbnl2: F-CTCCGACGAAGAATGCAAGT; R-GTTCTCTCTTGAACAGCGGC

Gapdh (internal control): F-TCGTCCCGTAGACAAAATGG; R-TTGAGGTCAATGAAGGGGTC

B2m (internal control): F-TGGTCTTTCTGGTGCTTGTC; R-GGGTGGAACTGTGTTACGTAG

Nanog: F-CCTCCAGCAGATGCAAGAACTC; R-CTTCAACCACTGGTTTTTCTGCC

Zcchc24: F-GGCCTCCTCCGAACTACC; R-GCATTTGGGGCACTTGTA

Esrp1: Taqman assay: Mm01220936_g1

Gapdh: Taqman assay: Mm99999915_g1

shRNA and siRNA

Target sequences for Zcchc24 were: GCTGCTTTGGCGAGTACAAGT and AGCTAGCTGCAGGGTAGATTT. Control shRNA sequence was: CAACAAGATGAAGAGCACCAA. siMBNL1/2 pools were described previously (Han et al. 2013). Rbm47 siRNAs were from Qiagen, siRbm47-1 (SI00917826) and siRbm47-2 (SI00917833). Control siRNA was the All Stars Negative Controls (SI03650318).

Supplemental References

Bebee, T.W., Park, J.W., Sheridan, K.I., Warzecha, C.C., Cieply, B.W., Rohacek, A.M., Xing, Y., and Carstens, R.P. (2015). The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. eLife *4*.

Dittmar, K.A., Jiang, P., Park, J.W., Amirikian, K., Wan, J., Shen, S., Xing, Y., and Carstens, R.P. (2012). Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. Mol Cell Biol *32*, 1468-1482.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., *et al.* (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172-177.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U. S. A. *111*, E5593-5601.