Supplemental Information

Myc and SAGA Rewire an Alternative Splicing Network During Early Somatic Cell Reprogramming

Calley L. Hirsch[*1], Zeynep Coban Akdemir[*2,3], Li Wang[3,4,5], Gowtham Jayakumaran[1,6], Dan Trcka[1], Alexander Weiss[1], J. Javier Hernandez[1,6], Qun Pan[7], Hong Han[6,7], Xueping Xu[8], Zheng Xia[8,9], Andrew P. Salinger[3,4], Marenda Wilson[3], Frederick Vizeacoumar[1], Alessandro Datti[1], Wei Li[8,9], Austin J. Cooney[8], Michelle C. Barton[2,3,4], Benjamin J. Blencowe[6,7], Jeffrey L. Wrana[1,6,10], Sharon Y.R. Dent[3,4,10]

* These authors contributed equally to this work
[1]Center for Systems Biology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto M5G 1X5, Canada;
[2]Program in Genes and Development, Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA;
[3]Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA;
[4]Department of Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Science Park Research Division, Smithville, Texas 78957, USA;
[5]Program in Molecular Carcinogenesis, Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Science Park Research Division, Smithville, Texas 78957, USA;
[6]Department of Molecular Genetics, University of Toronto, Toronto M5S 1A8, Canada
[7]Banting and Best Department of Medical Research and Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada
[8]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA;
[9]Division of Biostatistics, Dan L. Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA;
[10]Addresses for correspondence: wrana@lunenfeld.ca; sroth@mdanderson.org

Content of Supplemental Information

Supplemental Table S6 (Excel file)
Alternative splicing analysis in reprogramming cells.

Supplemental Table S7 (Excel file)
siRNA and primer lists.

Accession Numbers
RNA-seq and ChIP-seq data are available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE67719.

Supplemental Figure Legends

Figure S1. Functional RNAi screen identifies Gcn5 and SAGA components as regulators of cellular reprogramming.

(*A*) Pie chart representation of the epigenetic classes targeted in the functional RNAi screen.

(*B*) Rank order plot of the RNAi screen DAPI result is displayed using $\log_2$ transformed values from the average DAPI area of two biological replicate experiments, each performed in duplicate. The control values (white), HATs (blue) and SAGA hits (orange) are highlighted. The screen cut-off is denoted by the dotted grey line.

(*C*) Venn diagram displaying 76 reprogramming regulators (purple) identified from the RNAi screen. The lowest 15% of targets (98) from alkaline phosphatase (AP) (red) and DAPI (blue) staining were overlapped to derive the regulator hit list.

(*D*) DAPI single cell area is displayed relative to siControl transfections for various conditions in the RNAi screen. Error bars indicate SE from the two biological replicate experiments, each performed in duplicate.

(*E*) Total cell number of MEFs is displayed following loss of SAGA component expression. MEFs were transfected with indicated siRNA pools. Total cell number was determined by counting transfected cells 48 and 72 hrs following transfection. Values are displayed relative to siControl transfections. Error bars indicate SD from three independent experiments.

(*F*) Knockdown of Gcn5 and SAGA components at different times in the reprogramming process. Top: Experimental schematic is shown. Secondary (2^o)

reprogramming MEFs were transfected with Control, Oct4, Gcn5, Trrap, Ccdc101, Taf12 and Pcaf siRNA pools one (red), five (purple) or nine (light blue) days after initiating Dox treatment. Transfected cells were fixed and stained for AP activity after four days. 2^o iPs cells (dark blue) were transfected with siRNAs, fixed and stained 3 days later. Bottom: Data is plotted as AP ratio relative to siControl transfections. Error bars represent SD of three independent experiments.

(*G*) Reintroduction of Gcn5 partially rescues reprogramming. Secondary reprogramming MEFs were co-transfected with siRNA (siControl, siGcn5-11 or siGcn5-13 that target the 3' UTR of Gcn5) and pCMV6 (control) or Gcn5 plasmids. Cells were stained for AP after 5 days of Dox treatment. Representative images from two independent experiments are shown.

(*H*) Gcn5 is required for human iPS cell reprogramming. BJ fibroblasts were induced to reprogram by daily mRNA transfections with *OCT4*, *SOX2*, *KLF4*, *MYC*, *LIN28* and *eGFP* for ten days. Reprogramming cells were co-transfected with siControl or siGcn5 pools on the third day of reprogramming and cells were fixed and stained for AP activity on Day 12.

Figure S2. Gcn5 bioChIPs in mESCs.

(*A*) Protein levels of bio-Gcn5 are comparable to endogenous Gcn5. Immunoblots were performed with whole cell lysates of wild type, BirAV5-FlBio and BirAV5-FlBioGcn5 stable mESC lines. Ponceau S staining is shown as a loading control.

(*B* and *C*) Examples of Gcn5-bound genes in the Active-a and Active-b clusters. H3K4me3, Gcn5, H3K27ac, H3K9K14ac, Pol II and Input tag numbers (normalized to

10 million tags) are visualized at Gcn5-bound genes from the Active-a (*Eif4a3* and *Dgcr8*) (*B*) and Active-b clusters (*Rsp14* and *Eef1a1*) (*C*) in UCSC genome browser.

(*D*) Gcn5 binds within bivalent domains. H3K4me3, H3K27me3, Gcn5, H3K9K14ac, Pol II and Input tag numbers (normalized to 10 million tags) are viewed across two Gcn5-bound genes (*Hhex* and *Cbx2*) from the Bivalent cluster in UCSC genome browser.

Figure S3. Gcn5 and Myc co-regulate cell cycle genes in mESCs.

(*A*) c-Myc, n-Myc and E2f1 ChIP-seq signal densities are enriched in the Active-a and Active-b clusters. Heatmap showing the enrichment of factor binding at 100 bp resolution within ± 5 kb of 7,499 Gcn5 binding sites.

(*B*) Genes from the Active-a and Active-b clusters co-bound by Gcn5, c-Myc, n-Myc and/or E2f1. Tag densities are shown for E2f1, c-Myc, n-Myc, Gcn5, H3K27ac, H3K9K14ac, H3K4me3, Pol II and Input samples (normalized to 10 million tags) across three Gcn5-bound genes (*Eef1a1*, *Rpl37a* and *Rplp0*) in the Integrative Genomics browser (IGV).

(*C* and *D*) c-Myc, n-Myc, Gcn5, H3K27ac, H3K4me3, H3K9K14ac, Pol II and Input tag densities are displayed across the *Gcn5* (B) and *Ccdc101* (C) genes in the UCSC genome browser.

(*E*) Validation of *Gcn5* knockout mESC lines. Gcn5 immunoblots were performed on the Flox(4) and Flox(6) cell lines that had been exposed to Cre recombinant excision of Gcn5 exons 3-18 (KO: 4C2 and 6C4) as well as those that had not (wt: 4-1 and 6-1).

(*F*) Full length *Gcn5* mRNA is not expressed in *Gcn5* knockout mESCs. The number of RNA-seq reads along the *Gcn5* gene is displayed, for the wild type (wt:4-1 and wt:6-1) and knockout (KO:4C2 and KO:6C4) mESC lines in the Integrative Genomics browser (IGV).

(*G*) The top 3 Gene Ontology (GO) terms of Gcn5-induced target genes bound by c-Myc/n-Myc (top), and Gcn5-induced target genes bound by E2f1 (bottom).

(*H*) Representative images of matched wild type (wt:4-1) and knockout (KO:4C2) *Gcn5*$^{-/-}$ mESCs stained for alkaline phosphatase (AP) activity. Similar results were observed with wt:6-1 and KO:6C4 cells. Scale bar = 100 μM.

(*I*) Loss of *Gcn5* does not affect pluripotent gene expression in mESCs. The *Oct4*, *Sox2*, *Nanog* and *Gcn5* mRNA levels of wild type (wt:4-1 and wt:6-1) and knockout (KO:4C2 and KO:6C4) cells were measured. Error bars indicate SD from two independent experiments.

Figure S4. Myc directly targets and induces *Ccdc101* mRNA expression during reprogramming.

(*A*) The mRNA level of *Ccdc101* is up-regulated during reprogramming. The gene expression levels of *Ccdc101* across a time course of Dox-inducible reprogramming in secondary MEFs were quantified by qRT-PCR. D = number of days in Dox. Error bars indicate SD from an average of four independent experiments.

(*B*) Myc stimulates *Ccdc101* expression during early reprogramming. Secondary MEFs were transfected under mock conditions, transfected with siControl, siOct4, siSox2, siKlf4 or siMyc one day prior to Dox exposure. *Ccdc101* mRNA levels were

measured 2 days following Dox induction. Asterisks indicate t-test P value < 0.01 relative to siControl. Error bars indicate SD from average of three independent experiments.

(*C*) qRT-PCR validation of *Oct4*, *Sox2*, *Klf4* and *Myc* mRNA levels following siRNA knockdowns in Day 2 secondary reprogramming MEFs. Error bars indicate SD from the mean of three independent experiments.

(*D*) Myc binds the TSS of *Ccdc101*. ChIP-qPCR was performed, using the c-Myc antibody and primers surrounding the TSS of *Ccdc101*, in mESCs and secondary reprogramming MEFs cultured in the absence or presence Dox for 2 or 3 days. Error bars indicate SD from the mean of two representative data sets.

(*E*) The MEF lentiviral infection scheme is detailed. MEFs were infected with lentivirus six hours after cell plating. Infected cells were imaged or collected for gene expression analysis 3 days following viral introduction. Representative images of EGPF and mCherry infected cell are shown.

(*F*) Immunoblots of lysates from MEF infected cells. Ponceau S staining is shown as a loading control. A representative image of two independent experiments is displayed.

Figure S5. Myc and Gcn5 partner to activate RNA processing genes.

(*A*) Myc binds more genes in early reprogramming cells than in MEFs. Bar plot shows the total number of genes bound by Myc in MEFs and Day 2 (D2) reprogramming cells.

(*B*) Myc and Gcn5 interact in reprogramming cells and mESCs. Co-immunoprecipitations were performed with lysates from Day 2 (D2) reprogramming cells and mESCs using beads only (-) or Myc antibody. Immunoblots for Gcn5, Myc and Max are shown. The vertical bar between the inputs and IPs indicates a different exposure condition for the same membrane. A representative image of three independent experiments is shown.

(*C*) Schematic representation of RNA-seq in early reprogramming cells. Biological replicate experiments were performed.

(*D*) Gcn5 and Myc reprogramming responsive genes. Reprogramming responsive genes (differentially expressed genes at Day 2 (D2) of reprogramming versus MEFs ± 1.4-fold) are plotted as the total number of up-regulated (red) or down-regulated (blue) genes during reprogramming (All). Of the genes that change expression ± 1.4-fold during reprogramming, the number of genes up-regulated or down-regulated by Myc and Gcn5 are plotted as responsive genes.

(*E*) The overlap between Myc and Gcn5-responsive genes (± 1.4-fold) after 2 days of reprogramming is shown as a Venn diagram.

(*F*) Genes directly up-regulated by Gcn5 and Myc in reprogramming. Top: Venn diagram of Day 2 (D2) Myc and Gcn5 co-induced genes overlapped with D2 Myc and mESC Gcn5 co-bound genes. The percentage of co-induced genes bound is indicated along with the associated P value. Bottom: Table of enriched GO terms for the overlapping genes with associated P values.

(*G*) SAGA components regulate gene expression levels of RNA processing factors. The relative mRNA levels from a subset of RNA processing genes directly regulated

by Myc and Gcn5 (see Figure S5E) were measured in secondary MEFs transfected with Control, Myc, Gcn5, Trrap, siCcdc101, Taf12 or Pcaf siRNA after two days of reprogramming. Untreated 2^o MEFs are shown for comparison. The mRNA level of each gene is displayed relative to siControl. The average of three independent experiments is shown.

Figure S6. A subset of RNA processing factors regulated by Gcn5 and Myc are required for reprogramming.

(*A*) Reprogramming (dark blue) and mESC (light blue) RNAi screen results are displayed as relative DAPI area compared to siControl transfections in rank order. All control samples are displayed as black bars for the reprogramming screen and grey for the mESC screen. Error bars represent SD from three independent experiments.

(*B* and *C*) Alkaline phosphatase (AP) stained wells from a representative reprogramming (*B*) and mESC (*C*) RNAi screen.

Figure S7. Alternative splicing events regulated by the Myc / Gcn5 network during the initiation phase of reprogramming.

(*A*) Alternative splicing events regulated by Myc and Gcn5. The number of alternative splicing events that change (PSI values ≥ 15%) between Day 2 (D2) reprogramming cells versus MEFs are plotted as the total number of exon exclusion (grey) and exon inclusion (red) splicing events established during early

reprogramming (All). Of the events that change by ≥ 15% PSI, the number of exon exclusion and exon inclusion events regulated by Myc and Gcn5 are plotted.

(*B*) The overlap between Myc and Gcn5-regulated alternative splicing events (≥ 15% PSI change) after 2 days of reprogramming is shown as a Venn diagram.

(*C*) Knockdown of alternatively spliced (AS) genes during the initiation phase of reprogramming. Secondary MEFs were transfected one day prior to Dox exposure with siRNA pools targeting control (grey), AS genes co-regulated by Myc and Gcn5 (*), as well as co-regulated AS genes that fell just below the 15% PSI threshold (*Sdr39u1* and *Spag9*) and AS genes regulated primarily by Gcn5 (*Npnt* and *Rgs12*). Cells were fixed and stained for alkaline phosphatase (AP) activity after five days of Dox treatment. Results are displayed as relative AP area compared to siControl transfections in rank order. Error bars represent SD from three independent experiments.

(*D* and *E*) The mRNA splicing levels of *Slain2*, *Plod2*, *Fat1* and *Pcm1* in MEFs and Day 2 reprogramming cells (D2) or mouse embryonic stem cells (mESCs) (E) after knockdown of RNA processing factors directly regulated by Myc and Gcn5. Semi-quantitative percent spliced in (PSI) values are displayed. Presence of the red exon denotes exon inclusion. Representative images of three independent experiments are shown.

Supplemental Materials and methods

Cell culture

All mESCs were maintained in DMEM with high glucose (Gibco) supplemented with 15% FBS (Wisent), 2 mM L-glutamine (Gibco), 0.1 mM non-essential amino acids (Gibco), 1% penicillin/streptomycin (Gibco), 0.1 mM β-mercaptoethanol (Sigma) and 1000 U/mL of LIF / ESGRO (Chemicon). AB1 mESCs were cultured on 0.1% gelatin coated plates. R1 mESCs and *Gcn5*$^{flox/flox}$ mESCs were cultured on dishes with Mitomycin-C treated MEFs. Embryoid bodies were seeded as hanging drops of 400 mESCs in mES media lacking LIF. After 3 days the EBs were transferred to non-adherent bacterial grade dishes and cultured for a maximum of 6 addition days. 2^{o} 1B MEFs were maintained in DMEM with high glucose, 10% FBS, 2mM L-glutamine and 1% penicillin/streptomycin. 2^{o} MEFs were seeded for reprogramming at passage 5 in mES media and OKMS induced with 1.5 μg/mL doxycycline (Dox).

Generation of AB1-BirAV5-FLBio Gcn5 and AB1-BirAV5-FLBio stable mESCs

The pEF1αBirAV5 and pEF1αFLBio plasmids were kind gifts from Dr. Jianlong Wang (Mount Sinai Hospital, New York). Mouse Gcn5 cDNA was sub-cloned into the XbaI site of pEF1αFLBio vector to generate the pEF1αFLBioGcn5 plasmid. AB1 mESCs stably expressing *in vivo* biotinylated Gcn5 (BirAV5-FLBioGcn5) were generated as previously described (Kim et al. 2009). A control cell line (BirAV5-FLBio) that stably expressed BirA and FLBio alone was established in a similar manner. Stable expression of *in vivo* biotinylated Gcn5 was confirmed by anti-Gcn5

(Cell Signaling Technology), anti-Streptavidin (ThermoScientific), and anti-V5 (Invitrogen) immunoblots.

Derivation of the Gcn5$^{flox/flox}$ embryonic stem cell lines

Male and female *Gcn5$^{flox/flox}$* mice were bred and checked for plugs daily. At E3.5, uteri were removed and flushed with mES medium containing 0.02 mM HEPES. Blastocysts were cultured with mES medium in 60 mm culture dishes plated with gamma-irradiated STO feeders for 3 days. When blastocysts had adhered to the feeder layer, inner cell mass outgrowths were aspirated and plated individually in single wells of a 24-well plate. The next day, each well was split 1:1 in mES media using 0.25 % trypsin-EDTA. The cells were fed with mES cells medium daily and split 1:1 every 3 days until the ES cells colonies were visible. When several well-formed colonies were visible, the cells were spilt and plated on 60 mm culture dishes and expanded. For genotyping, feeders were removed by re-plating two or three times on non-gelatin coated dishes for 40-50 minutes, to remove the majority of feeder cells. The *Gcn5$^{flox/flox}$* genotype was verified by PCR with the primers WL20 5'-CACAGAGCTTCTTGGAGACC-3' and WL21-3 5'-GGCTTGATTCCTGTACCTCC-3'. To generate Gcn5$^{\Delta 3\text{-}18/\Delta 3\text{-}18}$ mESCs, *Gcn5 $^{flox/flox}$* cells were transiently transfected with the pBS598 EF1α-EGFPcre vector (Addgene) using Effectene Transfection Reagent (Qiagen). GFP positive and negative ES cells were sorted 72 hours after the transfection by FACS and subsequently re-seeded and expanded with mES media for nearly one week until individual colonies were picked from the GFP^{+} (4C2, 6C4) and GFP^{-} (4-1 and 6-1) populations. Cell were collected for genotyping from each line to

confirm that exons 3-18 of *Gcn5* had been excised by Cre recombinase in 4C2 and 6C4, but not 4-1 and 6-1 mESCs. All cells were genotyped by PCR with the primers WL20 and WL23 5'-ATAGTAGCGACTGCGCAACC-3'. qRT-PCR along with immunoblots verified that Gcn5 expression was not detectable in either 4C2 or 6C4 cells. G-banding cytogenetic analysis (Texas Children's Hospital and Cell Line Genetics) identified that all newly derived ES cell lines had a normal male karyotype.

2º MEF transfections

Gene knockdown experiments were performed in 2º 1B MEFs. Cells were transfected with siRNA using Lipofectamine RNAiMAX (Invitrogen) upon seeding 2.25 x 10^4 cells / well of a 12 well plate. All siRNA pools (siGenome, Dharmacon) were used at 40 nM and individual siRNAs (siGenome, Dharmacon) were used at 10 nM concentrations. After 24 hrs, Dox was added to the cells in mES media. 2 days later cells were lysed for RNA isolation to assess knockdown efficiency, or remained in Dox mES media for a total of 5 days after which cells were fixed with 4% PFA and stained for alkaline phosphatase activity (Vector Red, Alkaline Phosphatase substrate kit; Vector Labs). Image analysis and quantification of AP positive area was determined using a Celigo Imaging Cell Cytometer (Cyntellect). For the rescue experiments, siRNA transfections were performed in an identical manner, but were transfected 4 hrs later with 0.5 μg of pCMV6, pCMVSport2 Flag-Gcn5 or pCMVSport2 Flag-Gcn5 E568A using Fugene HD (Roche) at a ratio of 2.8 μL to 1 μg of DNA. All siRNAs are listed in Table S7.

Somatic cell reprogramming of human cells

Human BJ foreskin fibroblasts (ATCC) were reprogrammed using a modified-mRNA based strategy (Mandal and Rossi 2013). BJ cells were seeded in feeder-free conditions, on plates coated with Matrigel (Corning) and transfected for ten consecutive days using RNAi-MAX (Life Technologies) with modified-mRNA transcripts of the reprogramming genes, *OCT4*, *SOX2*, *KLF4*, *c-MYC*, *LIN28A* and *eGFP* (TriLink). Cells were cultured in MEF-conditioned Pluriton (Stemgent) reprogramming media, supplemented with 200 ng/mL B18R recombinant protein (eBiosciences) for the duration of transfection period. At day 12, reprogramming colonies were fixed with 4% PFA and stained for AP activity.

Biotin chromatin immunoprecipitation (bioChIP) assay

Approximately 2.0 x 10^7 BirAV5-FLBioGcn5 and BirAV5-FLBio mESCs were crosslinked with 1% formaldehyde in the culture media for 10 min at room temperature, quenched with 125 mM glycine and washed three times with PBS containing protease inhibitors. The cells were lysed in SDS Lysis Buffer (1 % SDS, 10 mM EDTA and 50 mM Tris-HCl, pH 8.1) for 30 min on ice and sonicated with a Branson Sonifier 450 sonicator (output 1.5, duty 60 %) for 2 cycles, followed by 10 min in the Bioruptor (Diagenode) for 30 sec on/off at the high setting for 5 cycles. Sonicated samples were centrifuged to remove the insoluable debris, diluted 1:10 with ChIP Dilution Buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1 and 167 mM NaCl) and were precleared with Dynabeads Protein A (Invitrogen) for 1.5 hrs. Subsequently, the pre-cleared lysates were incubated with

Dynabeads MyOne Streptavidin T1 (Invitrogen) overnight at 4°C. The subsequent bioChIP steps were carried out as previously described (Kim et al. 2009).

Immunoprecipitation and immunoblot analysis

Immunoprecipitaions and immunoblots were performed as previously described using 5 μg of Myc (D84C12) antibody for co-immunoprecipiation (Cell Signaling Technology)(Beyer et al. 2013). Gcn5l2 (C26A10) (Cell Signaling Technology) and Max (C-17) (Santa Cruz) antibodies were used for immunoblots.

Viral production and MEF infection

pLEX 307 (Addgene) destination vectors for Oct4, Sox2, Klf4 and c-Myc were generated by Gateway cloning. To produce infectious particles, 293T cells were co-transfected with the lentiviral vector, psPAX2 (Addgene) and pMD2.G (Addgene) using Lipofectamine 2000 (Life Technologies). After 48 and 72 hours, viral media was collected, filtered and titered. MEFs were plated at a density of 8.0 x 10^4 cells per 12 well and infected with lentiviral media (MOI = 5) six hours later in the presence of 5 μg/mL hexadimethrine bromide.

ChIP assays in reprogramming cells

2° 1B MEFs (p5) were seeded as 1.2 x 10^6 cells on 0.1 % gelatin coated 15 cm plates and treated 24 hrs later with 1.5 μg/mL of Dox in mES media, while control cells were kept in mES media in the absence of Dox. The cells were collected after 2 days of Dox exposure and processed as previously described (Beyer et al. 2013), with the

exception of chromatin sonication, where 10 x 20 pulses (25 % amplitude) of the Sonicator 4000 (QSonica) were used, and 15 μg of chromatin was precipitated with 1 μg of c-Myc antibody (sc-764, Santa Cruz).

ChIP-seq

Libraries were prepared using a modified version of the Illumina TruSeq ChIP Sample Prep protocol. Briefly, ChIP DNA was end repaired (End-It DNA End-Repair Kit - Epicenter), 3' adenylated with Klenow 3'-5' exo (New England Biolabs) and ligated to Illumina Tru-Seq adapters (diluted 1:7) with DNA ligase (New England Biolabs). 5 x PCR amplification cycles were performed on the adapter ligated ChIP DNA, 250-300 bp products were purified from a 2 % agarose gel and amplified for 10 additional PCR cycles. Libraries concentrations were determined with a Library Quantification Kit (Kapa Biosystems). 10 pM of each library was sequenced on an Illumina HiSeq 2000 as outlined by Illumina.

qRT-PCR

Total RNA was isolated using the RNeasy Mini Kit (Qiagen). RNA was converted to cDNA with oligo (dT) primers, random hexamers and SuperScript II Reverse Transcriptase (Invitrogen) prior to qPCR using SYBR green PCR Master Mix (Roche), or 25 μg of total RNA was paired directly with the Power SYBR Green RNA-to-CT 1 Step Kit (Applied Biosystems). Primer sequences are detailed in Table S7. Expression levels were calculated by the ΔΔCt method, normalized to either *Actin* or *Ywhaz*.

RNA-seq

Total cellular RNA and genomic DNA were isolated from 1.0 x 10^6 cells of the 4-1, 4C2, 6-1 and 6C4 lines using the AllPrep DNA/RNA Mini Kit (Qiagen). The concentration of genomic DNA was monitored to ensure that the samples were nearly equivalent (within 7.5 % of each other). 1 μg of RNA was subsequently spiked with 2 μL of 1:100 ERCC Spike-In Mix #1 or #2 prior to RNA-seq library preparation. For RNA-seq performed on reprogramming cells, total RNA was isolated from 2^o 1B MEFs transfected with siControl, siGcn5 or siMyc that had been treated in the absence and presence of Dox for 2 days. RNA-seq libraries were prepared according to the Illumina TruSeq Strandard Total RNA Sample Preparation Kit (with Ribo-Zero Gold, RS-122-2301), 10 pM of each library was sequenced according to standard Illumina protocols on an Illumina HiSeq 2000.

ChIP-seq analysis

The raw reads in ChIP-seq datasets were mapped to the NCBI build 37 (UCSC mm9) using Bowtie (version 0.12.8) (Langmead et al. 2009) with following parameters: -n 1, -m 1, -S, --best, --strata, and --chunkmbs 320. The mapped reads were used to identify enriched regions and enriched regions were normalized over input, using peak calling algorithm MACS (version 2.0.10) (Zhang et al. 2008) at a cut-off q-value ≤ 0.05. In two replicates of Gcn5 ChIP-seq data, we obtained a similar number of uniquely mapped reads, 17,319,169 and 22,898,752. Each replicate of ChIP-seq data was normalized to their respective input. We subtracted the Biotin-enriched peaks from Gcn5-bound peaks in two replicate experiments to eliminate the background

signal. Then, the remaining Gcn5-bound regions (peaks) in two replicates were intersected with BEDTools (version 2.13.3) (Quinlan and Hall 2010) and called as overlapped if they were within a distance of 3 kb of each other.

Generation of density files for browser data visualization

Alignment files were processed to obtain normalized density files in bedgraph format using the makeUCSCfile program in Homer (Hypergeometric Optimization of Motif Enrichment) software (Heinz et al. 2010). First, each tag was extended to 200 bp. Then, ChIP-fragment density was defined by the total number of overlapping ChIP-fragments at each genomic position and reported at each 25 bp bins (normalized to the total number of 10 million tags). For visualization, normalized density files in bedgraph format were uploaded to the UCSC genome browser (http://genome.ucsc.edu) and IGV (Integrative Genomics Browser) (Robinson et al. 2011; Thorvaldsdottir et al. 2013).

Heatmap generation

ChIP-seq signal densities were computed at defined equally sized bins within ± 5 kb of enriched regions and plotted as a heatmap using seqMiner (Ye et al. 2011) for Figure 2B, E and Supplemental Fig. S3A. The correlation heatmap shows the Pearson correlation coefficient between each pair of factors in Figure 3B. The target gene of a certain transcription factor (TF), histone modifier or histone modification is defined as a gene that has binding of this factor at its promoter region ± 2 kb. Pearson correlation coefficient between each pair of factors based on their target gene

similarity was computed, and graphically displayed using corrplot package in R. Analysis utilized unpublished H3K9K14ac ChIP-seq data from R. Young, (http://younglab.wi.mit.edu).

Distance plot calculations

The distance of each binding site to the transcription start site (TSS) of the closest gene was calculated and plotted as a pie chart in Figure 2C.

Average gene profile calculations

The enrichment of ChIP-seq signal density over its corresponding input was determined. For each ChIP-seq data set, ChIP-seq signal density was calculated by seqMiner (Ye et al. 2011) along the average gene profile across all datasets. Each gene in RefSeq gene model regardless of its size was partitioned into 100 equally sized bins. Its 5' end and 3' end flanking regions were partitioned into 20 equally sized bins. The average ChIP-seq signal density in each bin and its enrichment score over input were computed. The log2 fold ratio of enrichment scores were plotted along all bins.

Motif analysis

Transcription factor Affinity Prediction (TRAP) motif analysis tool (http://trap.molgen.mpg.de/cgi-bin/trap_multi_seq_form.cgi) was used to detect enriched motif sequences for the top 1000 enriched regions (± 300 bp of peak summits) with default P value threshold of 10^{-3}. To evaluate these motif results

against a background model, a control dataset of random sequences were generated of the same length and similar GC content, using RSAT (Regulatory Sequence Analysis Tool) (http://rsat.ulb.ac.be/) (Turatsinze et al. 2008; Thomas-Chollier et al. 2011). The corresponding motifs and p-values are represented in Supplemental Table S2.

Gene ontology analysis

DAVID functional annotation tool (Huang da et al. 2009) was used to find the enriched Gene Ontology Biological Process terms and their associated P values for gene lists extracted from ChIP-seq and/or RNA-seq analysis.

RNA-seq analysis

RNA-seq reads were aligned to NCBI build 37 (UCSC mm9), using TopHat 2.0.9 (Trapnell et al. 2009). For RNA-seq data in mESCs, three steps of normalization were performed. The first step of normalization was performed using the ERCC synthetic spike-in RNA sequences (http://tools.invitrogen.com/downloads/ERCC92.fa) and aligning them along with all other RNA-seq reads. Second, expression values were computed as fragments per kilobase of exon per million fragments mapped (FPKM) for each gene and spike-in sequence using NGSUtils (Breese and Liu 2013). Lastly, FPKM values for each gene were normalized to spike-in values using loess normalization through loess.normalize function in affy package of R (Gautier et al. 2004). The $\log_2$ fold change of normalized FPKM values were used to detect differentially expressed

genes. For reprogramming RNA-seq data, the number of reads aligning to the exons of each gene was computed using HTSeq (Anders and Huber 2010). EdgeR was applied to determine differentially expressed genes at a false discovery rate (FDR) adjusted P value (q-value) of 1 % (Robinson et al. 2010). Integrative Genomics Browser (IGV) (Robinson et al. 2011; Thorvaldsdottir et al. 2013) was used to visualize RNA-seq reads along *Gcn5* gene, using the NCBI build 37 (UCSC mm9).

Alternative splicing analysis

Transcriptome-wide AS profiling was performed using the recently described pipeline (*vast-tools*; (Irimia et al. 2014)). *vast-tools* uses reads mapping to exon-exon (or exon-intron) junctions (EEJ or EIJ) to accurately detect and quantify different types of AS events. Percent spliced in (PSI) prediction was performed as previously described (Barbosa-Morais et al. 2012). Exon skipping events were selected if coverage and balance scores were satisfied across all samples and if at least a 15% difference in PSI values were present between MEFs and Day 2 siControl cells. Some events were excluded from further analysis if PSI values differed beyond 15% between the duplicate samples. In total, 59 selected events were plotted as Z score based on PSI level as a heatmap.

Binomial test

P value calculation for significance of overlapping regions in Venn diagrams was performed using Binomial test.

Supplemental References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* **11**: R106.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587-1593.

Beyer TA, Weiss A, Khomchuk Y, Huang K, Ogunjimi AA, Varelas X, Wrana JL. 2013. Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. *Cell reports* **5**: 1611-1624.

Breese MR, Liu Y. 2013. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* **29**: 494-496.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307-315.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**: 576-589.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**: 44-57.

Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallieres M, Tapial J, Raj B, O'Hanlon D et al. 2014. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**: 1511-1523.

Kim J, Cantor AB, Orkin SH, Wang J. 2009. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nature protocols* **4**: 506-517.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Mandal PK, Rossi DJ. 2013. Reprogramming human fibroblasts to pluripotency using modified mRNA. *Nature protocols* **8**: 568-582.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Thomas-Chollier M, Hufton A, Heinig M, O'Keeffe S, Masri NE, Roider HG, Manke T, Vingron M. 2011. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nature protocols* **6**: 1860-1869.

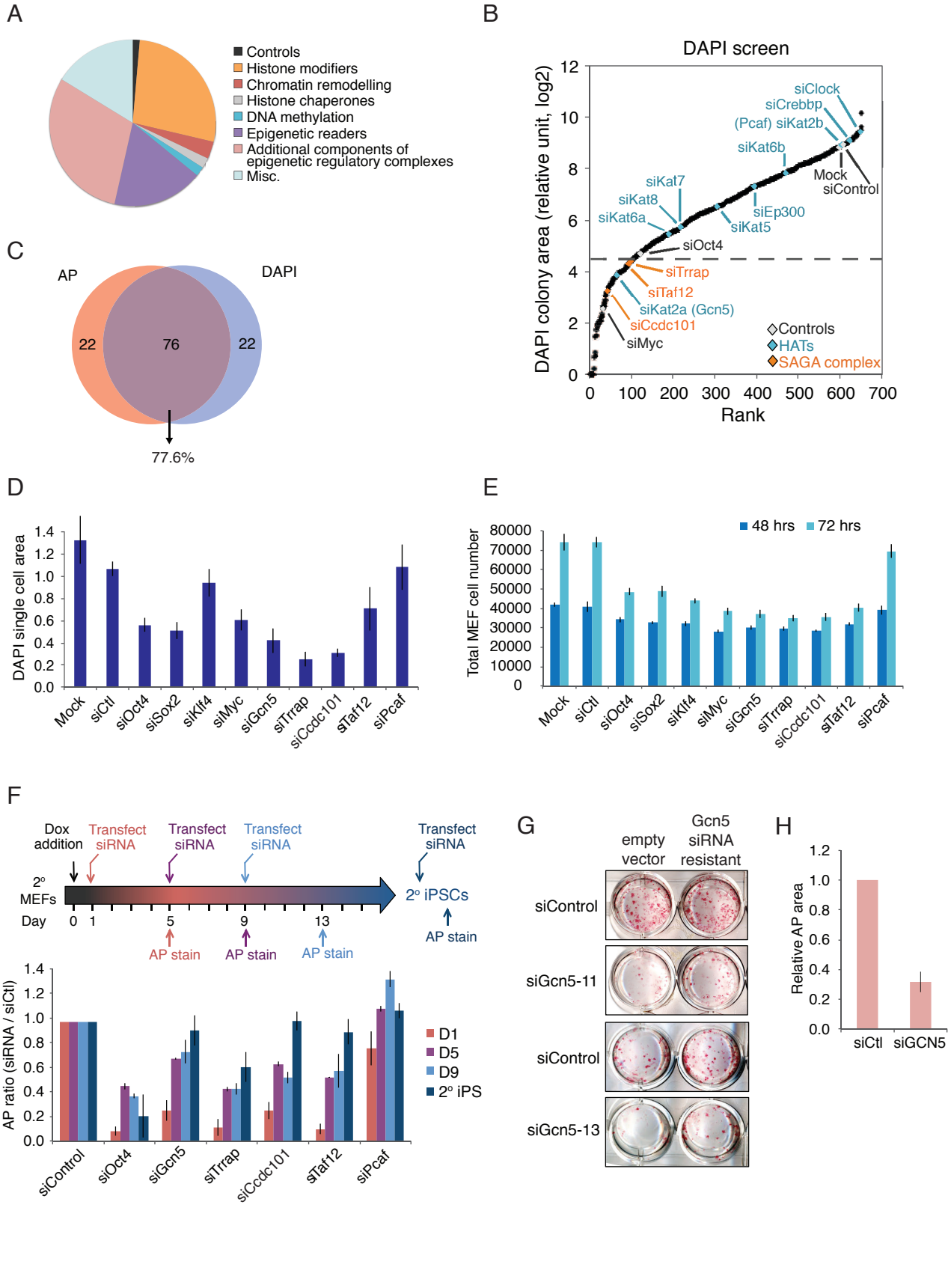
Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**: 178-192.

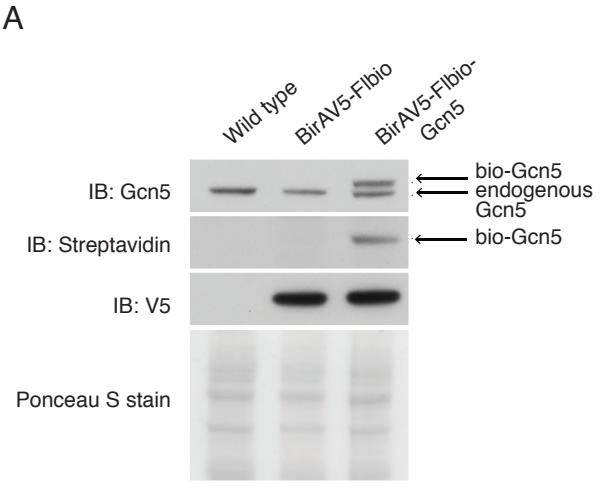
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

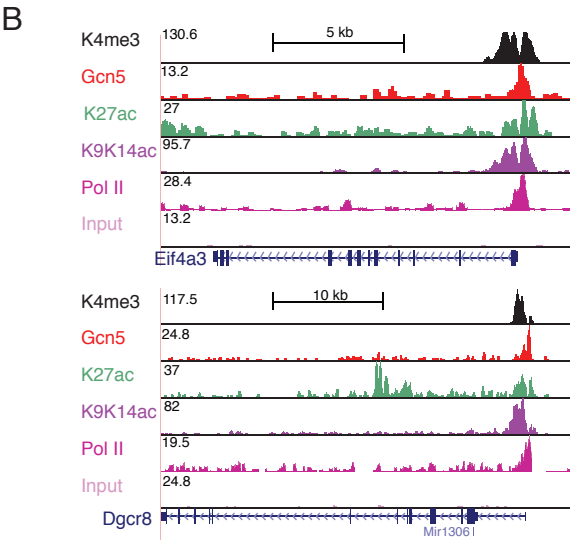
Turatsinze JV, Thomas-Chollier M, Defrance M, van Helden J. 2008. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. *Nature protocols* **3**: 1578-1588.

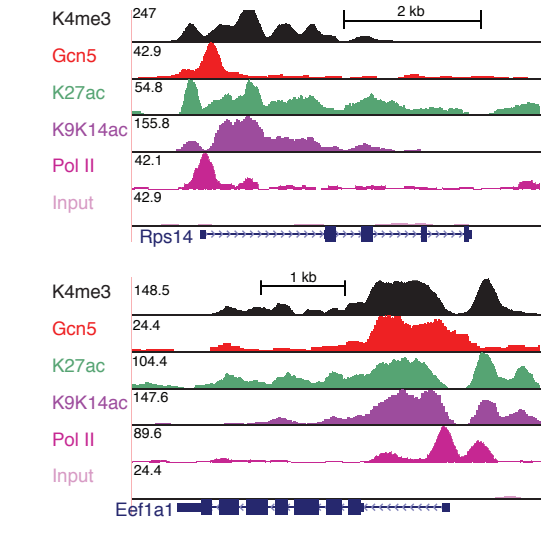
Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic acids research* **39**: e35.

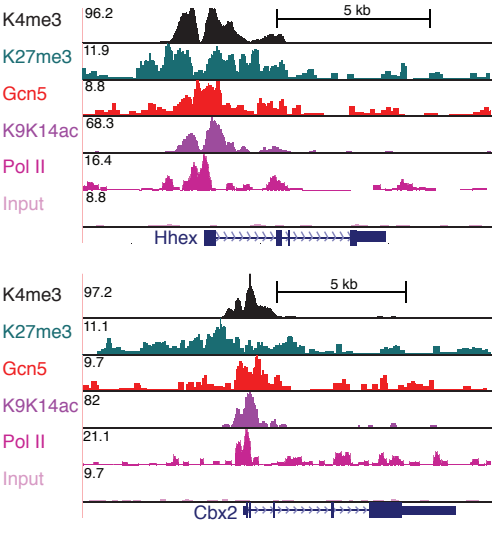
Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**: R137.


A
Controls
Histone modifiers
Chromatin remodelling
Histone chaperones
DNA methylation
Epigenetic readers
Additional components of epigenetic regulatory complexes
Misc.
B
DAPI screen
DAPI colony area (relative unit, log2)
Rank
siClock
siCrebbp
(Pcaf) siKat2b
siKat6b
Mock
siControl
siKat7
siKat8
siKat6a
siEp300
siKat5
siOct4
siTrrap
siTaf12
siKat2a (Gcn5)
siCcdc101
siMyc
Controls
HATs
SAGA complex
C
AP
DAPI
22
76
22
77.6%
D
DAPI single cell area
Mock
siCtl
siOct4
siSox2
siKlf4
siMyc
siGcn5
siTrrap
siCcdc101
siTaf12
siPcaf
E
48 hrs
72 hrs
Total MEF cell number
F
Dox addition
Transfect siRNA
2° MEFs
Day
2° iPSCs
AP stain
AP ratio (siRNA / siCtl)
D1
D5
D9
2° iPS
siControl
G
empty vector
Gcn5 siRNA resistant
siControl
siGcn5-11
siControl
siGcn5-13
H
Relative AP area
siCtl
siGCN5

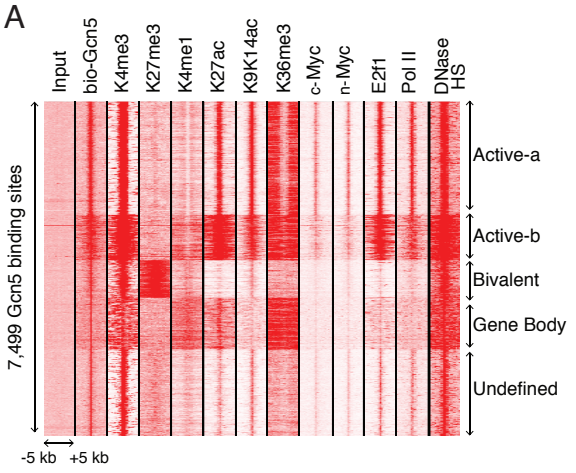

A
Wild type
BirAV5-Flbio
BirAV5-Flbio-Gcn5
IB: Gcn5
IB: Streptavidin
IB: V5
Ponceau S stain
bio-Gcn5
endogenous Gcn5
bio-Gcn5


B
K4me3
Gcn5
K27ac
K9K14ac
Pol II
Input
130.6
13.2
27
95.7
28.4
13.2
5 kb
Eif4a3
117.5
24.8
37
82
19.5
24.8
10 kb
Dgcr8
Mir1306

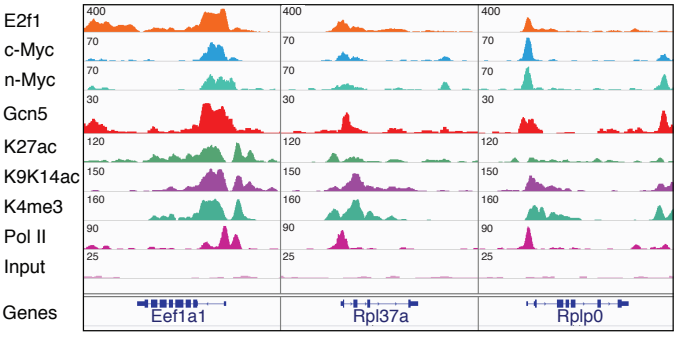

C
K4me3
Gcn5
K27ac
K9K14ac
Pol II
Input
247
42.9
54.8
155.8
42.1
42.9
2 kb
Rps14
148.5
24.4
104.4
147.6
89.6
24.4
1 kb
Eef1a1


D
K4me3
K27me3
Gcn5
K9K14ac
Pol II
Input
96.2
11.9
8.8
68.3
16.4
8.8
5 kb
Hhex
97.2
11.1
9.7
82
21.1
9.7
5 kb
Cbx2

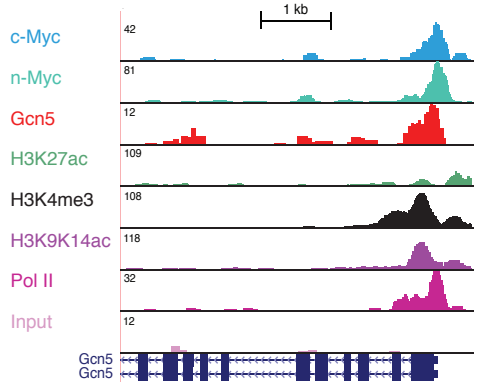
A



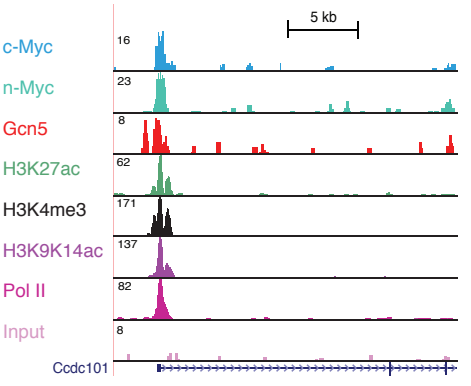
B



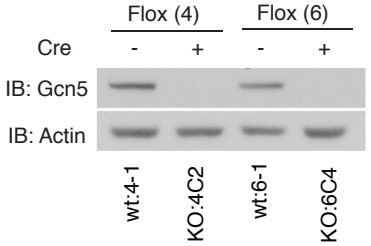
C



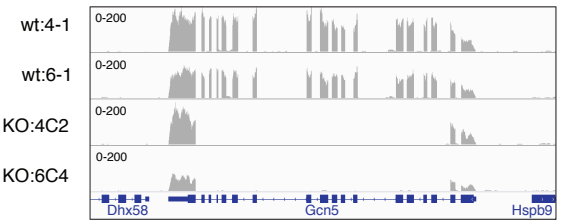
D



E



F



G

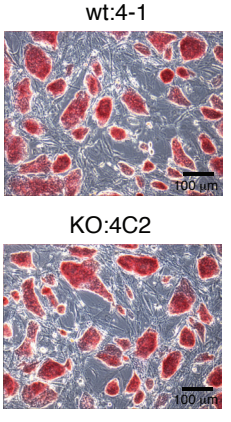
Gcn5-induced target genes bound by c-Myc/n-Myc

Biological Function	P value
Cell cycle process	3.06×10^{-5}
Translation	6.91×10^{-5}
Cell cycle	3.09×10^{-4}

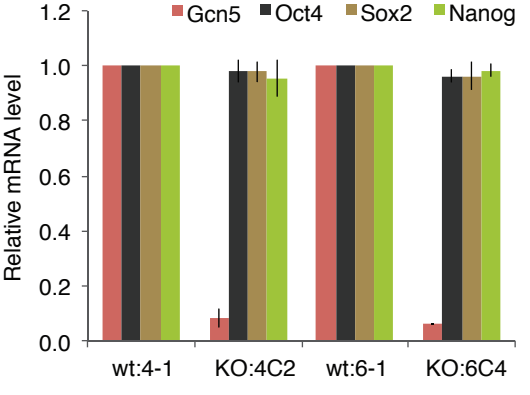
Gcn5-induced target genes bound by E2f1

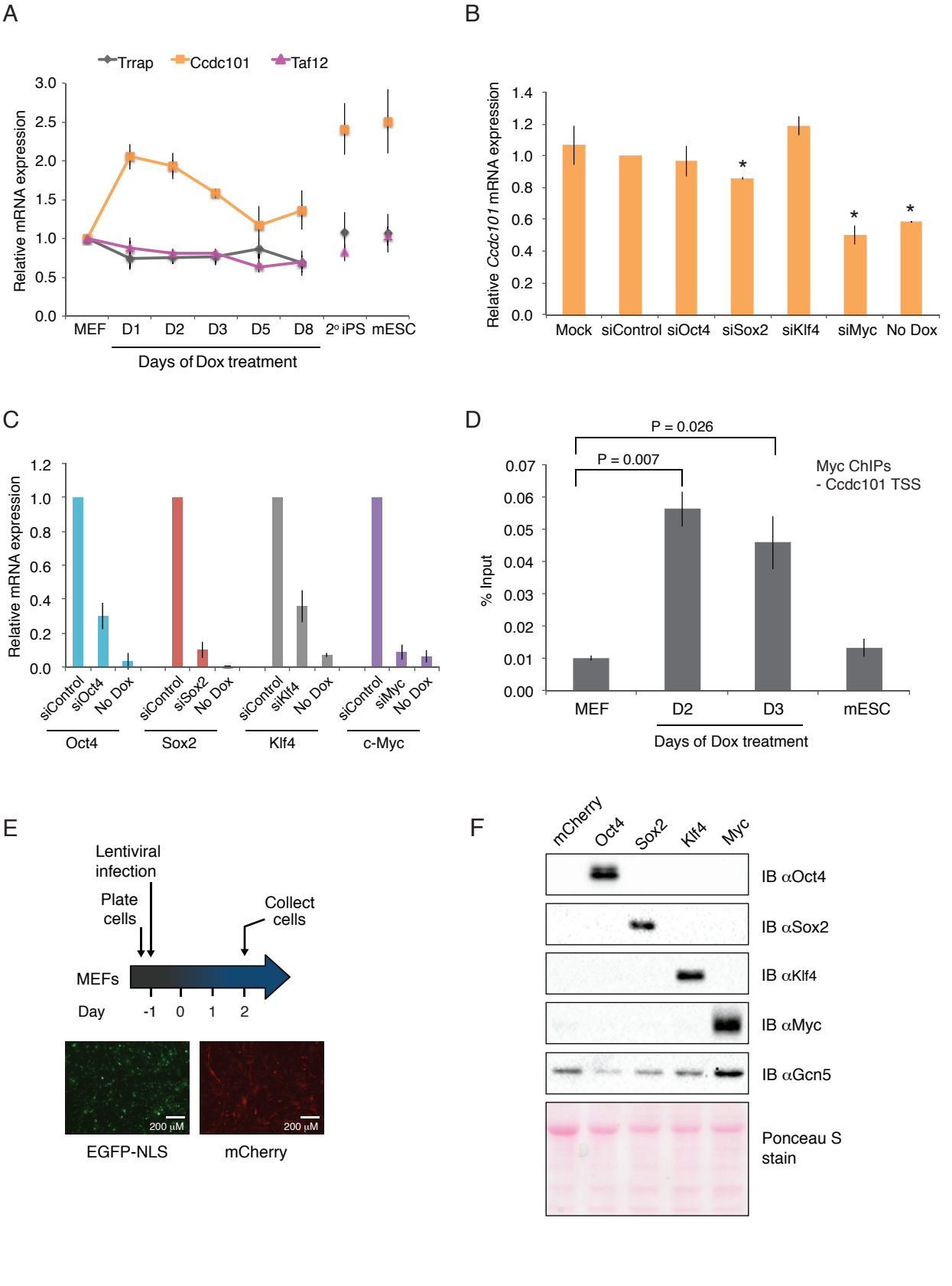
Biological Function	P value
Cell cycle	4.13×10^{-5}
Response to DNA damage	1.91×10^{-4}
Cell cycle process	3.96×10^{-4}

H

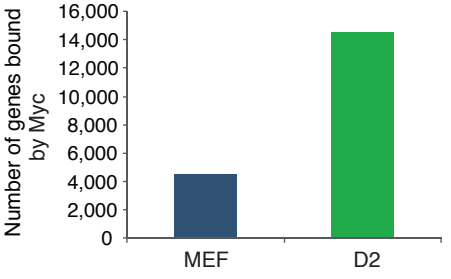


I

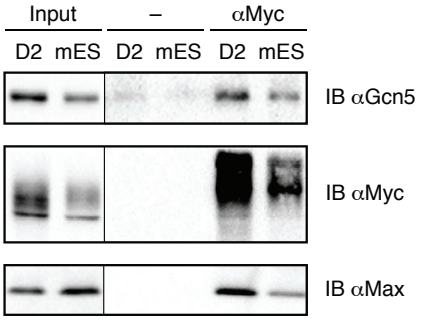



A
Trrap
Ccdc101
Taf12
Relative mRNA expression
3.0
2.5
2.0
1.5
1.0
0.5
0.0
MEF
D1
D2
D3
D5
D8
2° iPS
mESC
Days of Dox treatment
B
Relative Ccdc101 mRNA expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Mock
siControl
siOct4
siSox2
siKlf4
siMyc
No Dox
C
Relative mRNA expression
1.2
1.0
0.8
0.6
0.4
0.2
0.0
siControl
siOct4
No Dox
siControl
siSox2
No Dox
siControl
siKlf4
No Dox
siControl
siMyc
No Dox
Oct4
Sox2
Klf4
c-Myc
D
P = 0.026
P = 0.007
Myc ChIPs
- Ccdc101 TSS
% Input
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
MEF
D2
D3
mESC
Days of Dox treatment
E
Lentiviral
infection
Plate
cells
Collect
cells
MEFs
Day
-1
0
1
2
200 μM
200 μM
EGFP-NLS
mCherry
F
mCherry
Oct4
Sox2
Klf4
Myc
IB αOct4
IB αSox2
IB αKlf4
IB αMyc
IB αGcn5
Ponceau S
stain

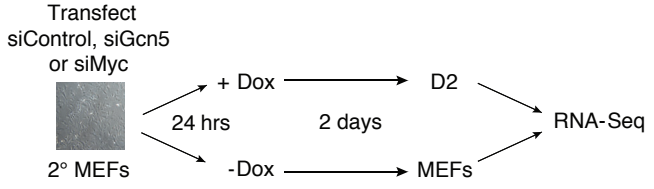
A



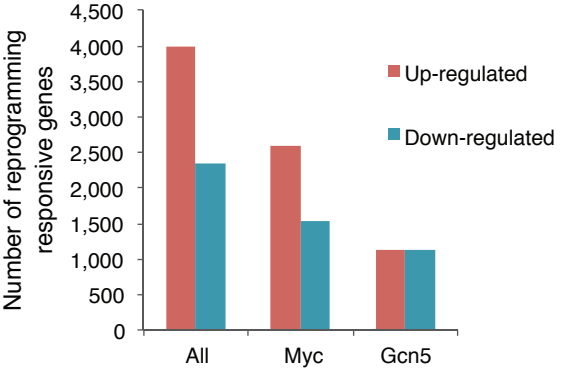
B



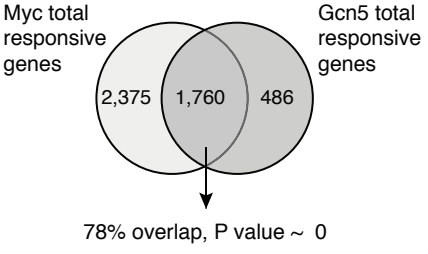
C



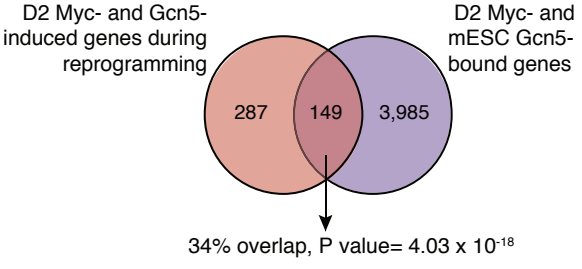
D



E

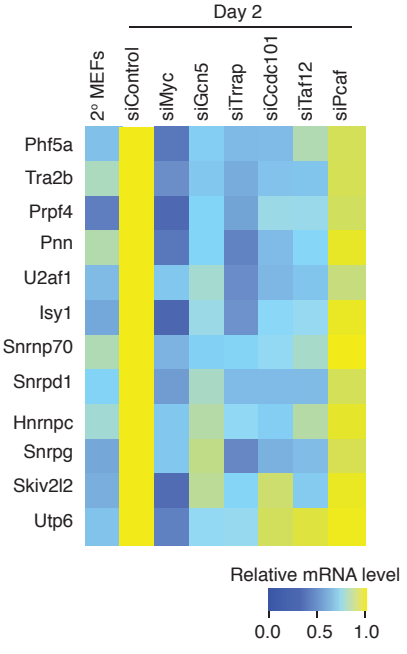


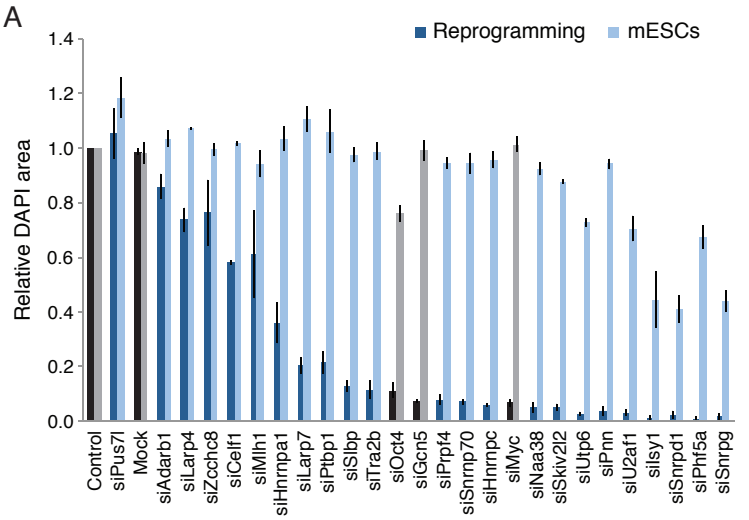
F

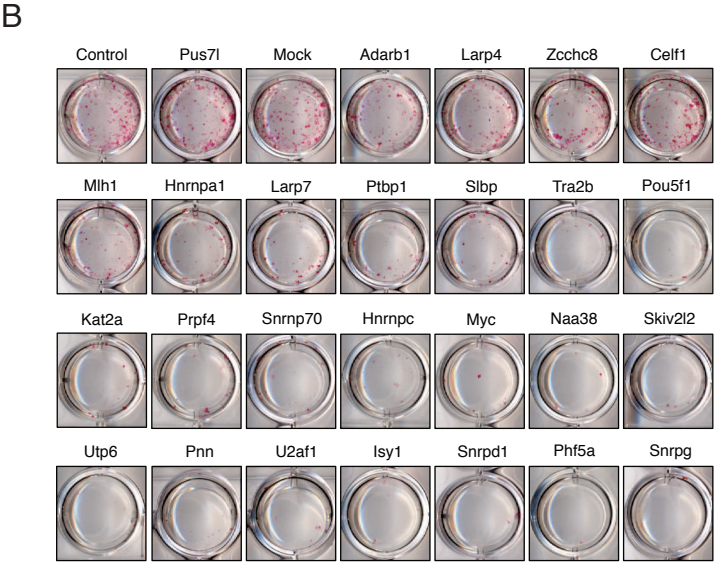


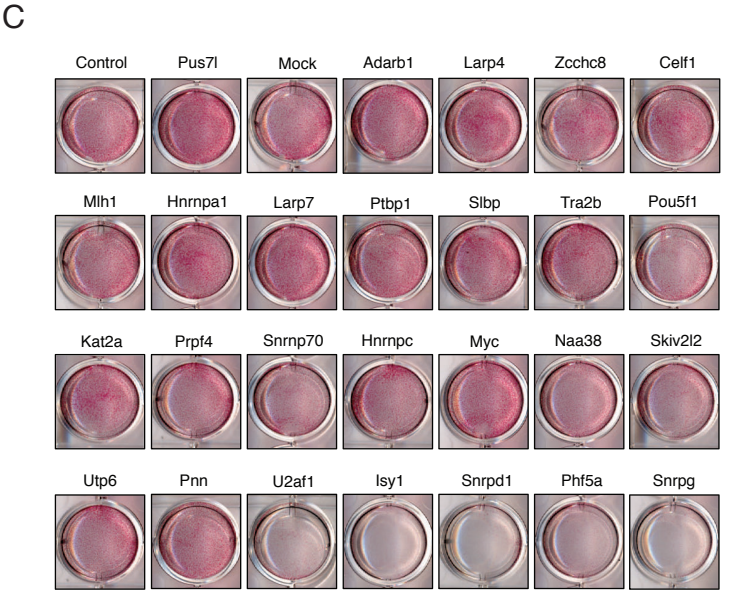
Biological Function	P value
mRNA processing	2.78×10^{-10}
RNA processing	1.64×10^{-9}
mRNA metabolic process	2.21×10^{-9}
RNA splicing	7.42×10^{-9}
DNA metabolic process	2.31×10^{-7}

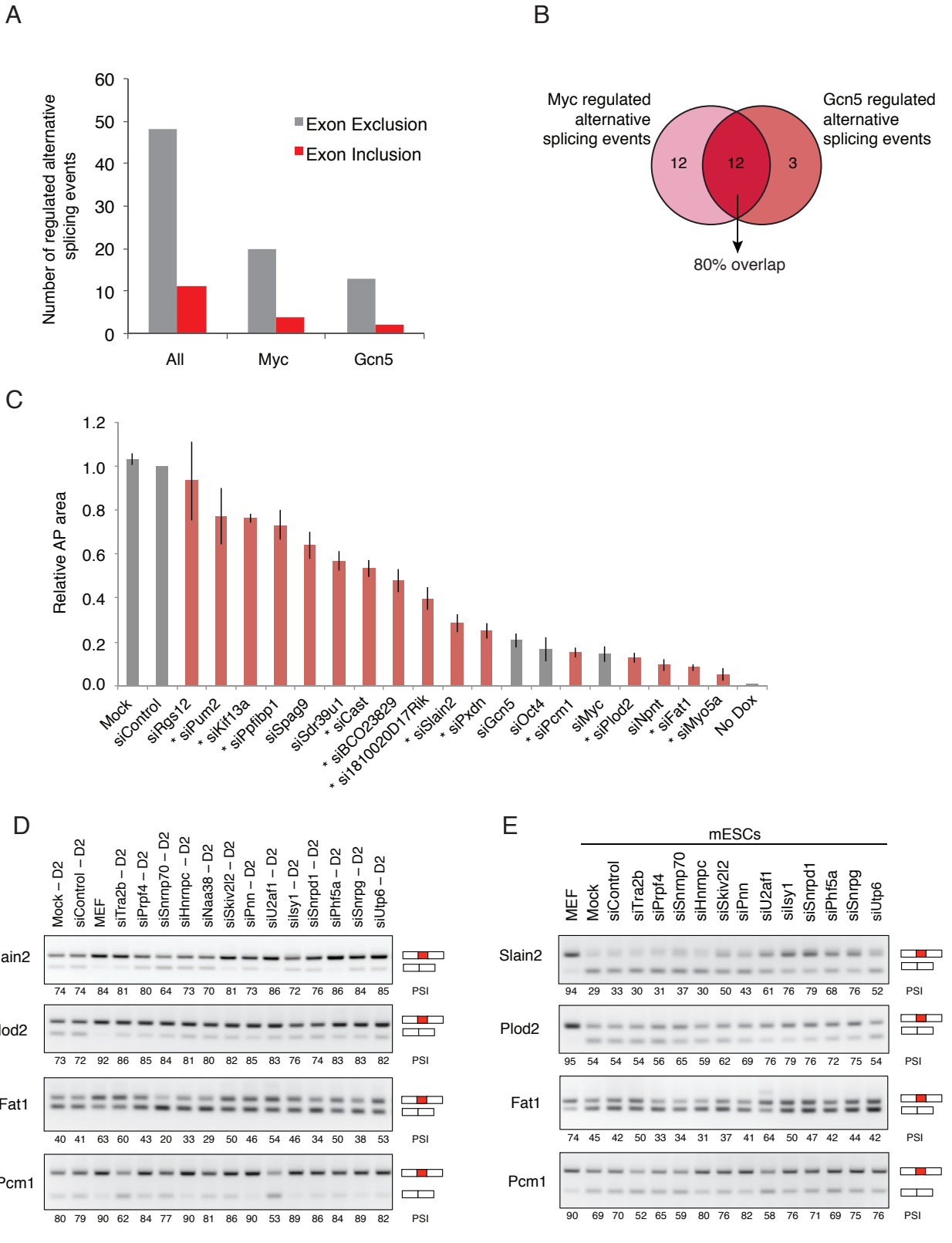
G




A
Reprogramming
mESCs
Relative DAPI area
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Control
siPus7l
Mock
siAdarb1
siLarp4
siZcchc8
siCelf1
siMlh1
siHnrnpa1
siLarp7
siPtbp1
siSlbp
siTra2b
siOct4
siGcn5
siPrpf4
siSnrnp70
siHnrnpc
siMyc
siNaa38
siSkiv2l2
siUtp6
siPnn
siU2af1
siIsy1
siSnrpd1
siPhf5a
siSnrpg


B
Control
Pus7l
Mock
Adarb1
Larp4
Zcchc8
Celf1
Mlh1
Hnrnpa1
Larp7
Ptbp1
Slbp
Tra2b
Pou5f1
Kat2a
Prpf4
Snrnp70
Hnrnpc
Myc
Naa38
Skiv2l2
Utp6
Pnn
U2af1
Isy1
Snrpd1
Phf5a
Snrpg


C
Control
Pus7l
Mock
Adarb1
Larp4
Zcchc8
Celf1
Mlh1
Hnrnpa1
Larp7
Ptbp1
Slbp
Tra2b
Pou5f1
Kat2a
Prpf4
Snrnp70
Hnrnpc
Myc
Naa38
Skiv2l2
Utp6
Pnn
U2af1
Isy1
Snrpd1
Phf5a
Snrpg


A
Number of regulated alternative splicing events
60
50
40
30
20
10
0
Exon Exclusion
Exon Inclusion
All
Myc
Gcn5
B
Myc regulated alternative splicing events
Gcn5 regulated alternative splicing events
12
12
3
80% overlap
C
Relative AP area
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Mock
siControl
siRgs12
* siPum2
* siKif13a
* siPpfibp1
siSpag9
siSdr39u1
* siCast
* siBCO23829
* si1810020D17Rik
* siSlain2
* siPxdn
siGcn5
siOct4
* siPcm1
siMyc
* siPlod2
siNpnt
* siFat1
* siMyo5a
No Dox
D
Mock – D2
siControl – D2
MEF
siTra2b – D2
siPrpf4 – D2
siSnrnp70 – D2
siHnrnpc – D2
siNaa38 – D2
siSkiv2l2 – D2
siPnn – D2
siU2af1 – D2
siIsy1 – D2
siSnrpd1 – D2
siPhf5a – D2
siSnrpg – D2
siUtp6 – D2
Slain2
74 74 84 81 80 64 73 70 81 73 86 72 76 86 84 85
PSI
Plod2
73 72 92 86 85 84 81 80 82 85 83 76 74 83 83 82
PSI
Fat1
40 41 63 60 43 20 33 29 50 46 54 46 34 50 38 53
PSI
Pcm1
80 79 90 62 84 77 90 81 86 90 53 89 86 84 89 82
PSI
E
mESCs
MEF
Mock
siControl
siTra2b
siPrpf4
siSnrnp70
siHnrnpc
siSkiv2l2
siPnn
siU2af1
siIsy1
siSnrpd1
siPhf5a
siSnrpg
siUtp6
Slain2
94 29 33 30 31 37 30 50 43 61 76 79 68 76 52
PSI
Plod2
95 54 54 54 56 65 59 62 69 76 79 76 72 75 54
PSI
Fat1
74 45 42 50 33 34 31 37 41 64 50 47 42 44 42
PSI
Pcm1
90 69 70 52 65 59 80 76 82 58 76 71 69 75 76
PSI