Cell Reports, Volume *22*

Supplemental Information

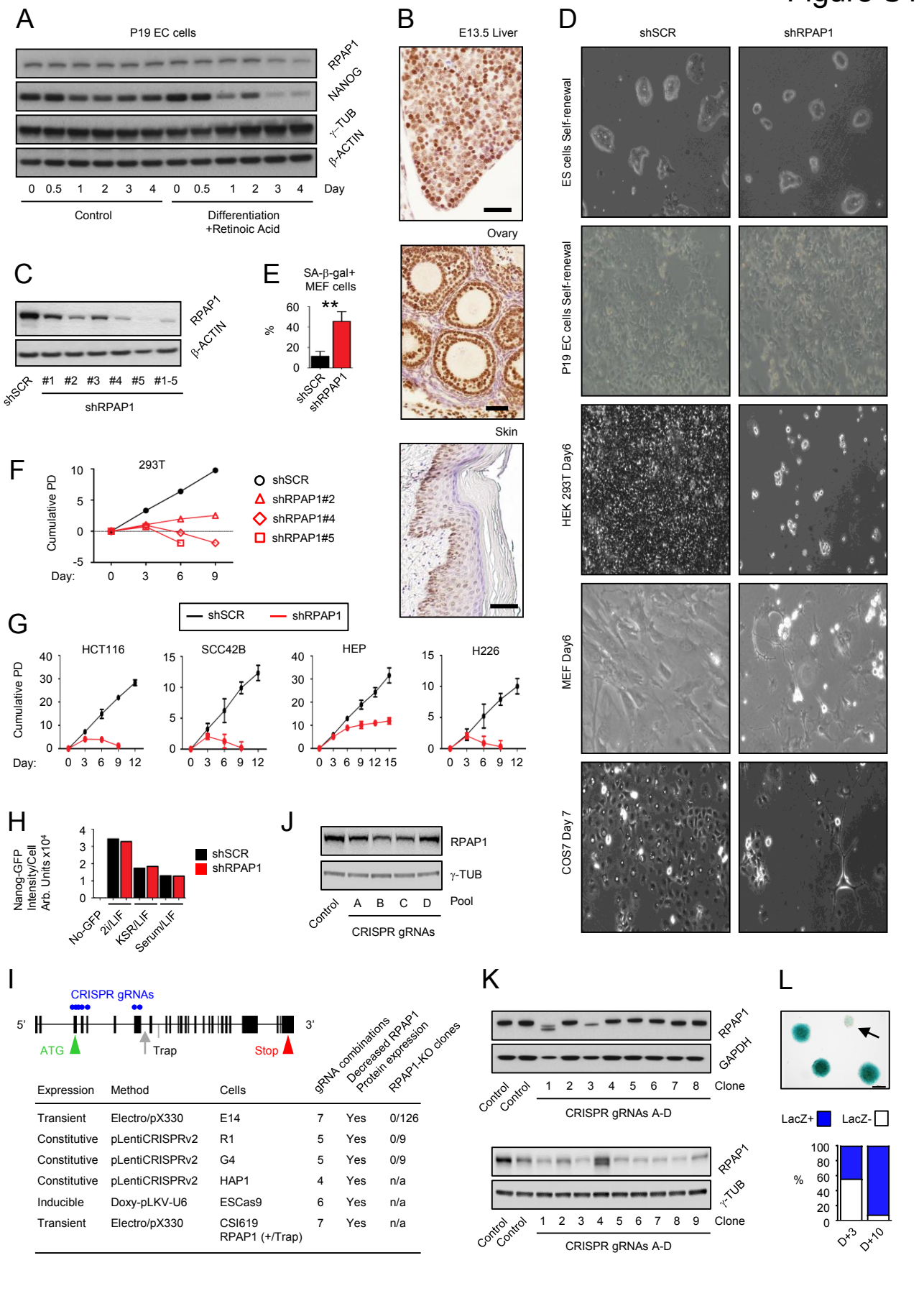
The RNA Polymerase II Factor RPAP1 Is Critical for Mediator-Driven Transcription and Cell Identity

Cian J. Lynch, Raquel Bernad, Isabel Calvo, Sandrina Nóbrega-Pereira, Sergio Ruiz, Nuria Ibarz, Ana Martinez-Val, Osvaldo Graña-Castro, Gonzalo Gómez-López, Eduardo Andrés-León, Vladimir Espinosa Angarica, Antonio del Sol, Sagrario Ortega, Oscar Fernandez-Capetillo, Enrique Rojo, Javier Munoz, and Manuel Serrano

Supplemental Information

- Supplemental Figures and Legends S1-S5
- Supplemental Experimental Procedures
- Supplemental References

Figure S1



Expression	Method	Cells			
Transient	Electro/pX330	E14	7	Yes	0/126
Constitutive	pLentiCRISPRv2	R1	5	Yes	0/9
Constitutive	pLentiCRISPRv2	G4	5	Yes	0/9
Constitutive	pLentiCRISPRv2	HAP1	4	Yes	n/a
Inducible	Doxy-pLKV-U6	ESCas9	6	Yes	n/a
Transient	Electro/pX330	CSI619 RPAP1 (+/Trap)	7	Yes	n/a

Figure S1. RPAP1 expression, localisation, and requirement for survival in stem cells versus differentiated cells, Related to Figure 1.

(A) Western blot of RPAP1 expression during a timecourse of P19EC cell differentiation by Retinoic Acid addition.

(B) Immunohistochemical staining for RPAP1 in mouse E13.5 liver, adult ovary, and adult skin. Scale bars represent 30 μm.

(C) Western blot of RPAP1 expression in mouse ES cells at day 6 following five separate lentiviral shRNA against RPAP1 (#1-#5).

(D) Photographs of the indicated cell lines at days 6-9 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1).

(E) Quantification of senescence-associated β-galactosidase staining in MEFs at day 9 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1). Mean+/-SD, n=3 replicates; **p<0.01.

(F,G) Proliferation curves (shown by cumulative population doubling) following lentiviral control (shSCR) or RPAP1 depletion (shRPAP1) in 293T cells using 3 different shRNAs (**F**), or in the indicated cell lines (**G**).

(H) FACS analysis of a Nanog-GFP reporter ES cell line (TNGA) cultured in three different media cocktails, at day 6 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1). Wild type non-GFP ES cells were used as negative control.

(I) Schematic of the 26-Exon mouse RPAP1 gene (Gene ID: 68925; NM_177294.5). Indicated are: the open reading frame ATG start (green, exon3) and TGA stop (red, exon26); the location of the gene trap in CSI619 ES cells (grey, intron8); and the location of CRISPR guide RNAs used in this study (blue, exons 3-7; see also: **Resource Tables** in **Supplemental Experimental Procedures**). Table (below) summarizes the effect of multiple CRISPR approaches on the expression of RPAP1. While RPAP1 protein levels were decreased in cell pools and in clonal lines, no RPAP1-null clones could be derived.

(J) Example of Western blot analyses of whole population from haploid HAP1 cells following CRISPR against RPAP1 using lentiviral constitutive CRISPR/Cas9 expression.

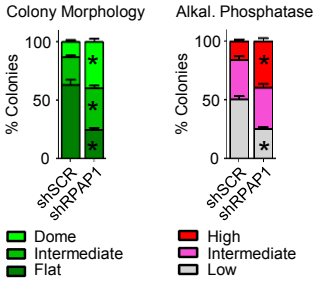
(K) Examples of Western blot analyses of ES clones following CRISPR against RPAP1 using CRISPR/Cas9 expression systems which were transient in wild-type E14 ES cells (pX330; above), or constitutive in G4 ES cells (lentiviral; below).

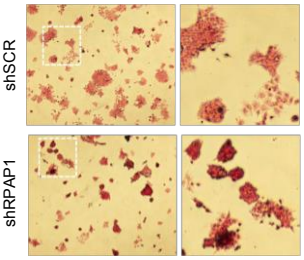
(L) CSI619 (RPAP1+/Trap) reporter ES cells stained for LacZ 3 days after CRISPR against RPAP1. Inset shows examples of ES colonies expanded from single cells and stained for LacZ (blue) expressed from the RPAP1 β-geo reporter allele. Arrow: a white/non-stained colony, indicating that CRISPR has successfully mutated at least one of the RPAP1 alleles, by knocking out the RPAP1 β-geo reporter allele. Bar chart below, shows the

percentage of non-staining colonies at Days +3 or +10 after CRISPR in the whole population. Significantly fewer non-staining colonies are observed at Day+10 (after a passage). This suggests that where CRISPR is active, and also knocks out the WT RPAP1 allele, the ES cells display a growth or survival phenotype within a few days.

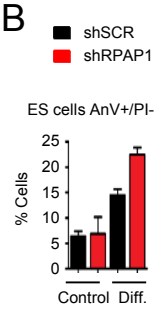
Figure S2

A

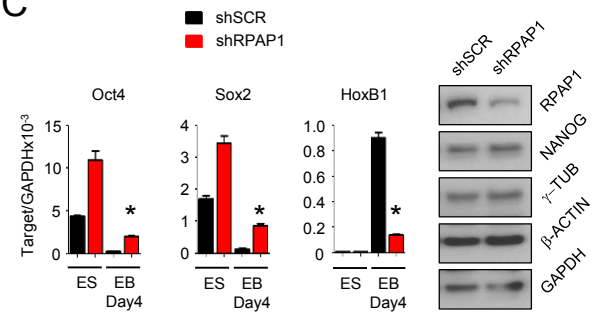




B



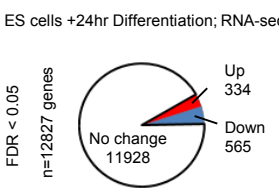
C



D

RPAP1(+/Trap) mouse generation		Germline transmission by chimera matings:	
Micro-injected embryos	254	Pups born	156
Pups born	21	Pups Agouti	7
Chimeras (Sex) % chimerism	10 (3M, 7F) 20-60%	Pups RPAP1+/Trap	0

E



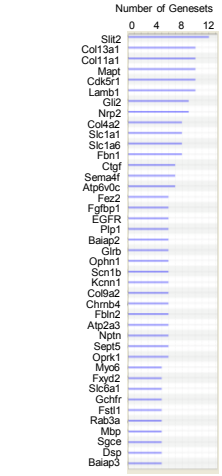
F

	GSEA Hallmark Pathways (from total of 50 genesets)	FDR q-value
UP	MYC TARGETS_V2	0.011
	MYC TARGETS_V1	0.029
	G2M CHECKPOINT	0.063
	E2F TARGETS	0.076
DOWN	MYOGENESIS	0.001
	EPITHELIAL MESENCHYMAL TRANSITION	0.002
	ANGIOGENESIS	0.009
	ESTROGEN RESPONSE LATE	0.013
	K-RAS SIGNALING_UP	0.013
	COAGULATION	0.025
	COMPLEMENT	0.027
	P53 PATHWAY	0.072
	ESTROGEN RESPONSE EARLY	0.086
	APOPTOSIS	0.090
	TNFA SIGNALING VIA NFKB	0.112
	UV RESPONSE DOWN	0.118
	APICAL JUNCTION	0.127
	IL2_STAT5 SIGNALING	0.153
	APICAL SURFACE	0.170
	SPERMATOGENESIS	0.178
	HYPOXIA	0.187
	INFLAMMATORY RESPONSE	0.294

G

	GSEA Gene Ontology Terms (from total of 1454 Genesets)	FDR q-value
UP	NUCLEOBASENUCLEOSIDE_AND_NUCLEOTIDE_METABOLIC_PROCESS	0.133
	NUCLEOTIDE_METABOLIC_PROCESS	0.141
	RNA_DEPENDENT_ATPASE_ACTIVITY	0.142
	ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CHROMATIN_ARCHITECTURE	0.146
	ATP_DEPENDENT_RNA_HELICASE_ACTIVITY	0.148
	CHROMATIN_REMODELING	0.168
DOWN	GLYCOSAMINOGLYCAN_BINDING	0.120
	ATPASE_ACTIVITY_COUPLED_TO_TRANSMEMBRANE_MOVEMENT_OF_IONS	0.122
	EXTRACELLULAR_MATRIX	0.122
	CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION	0.129
	POLYSACCHARIDE_BINDING	0.138
	PROTEINACEOUS_EXTRACELLULAR_MATRIX	0.142
	TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSINE_KINASE_ACTIVITY	0.143
	TRANSFERASE_ACTIVITY_TRANSFERRING_SULFUR_CONTAINING_GROUPS	0.151
	ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	0.152
	TRANSMEMBRANE_RECEPTOR_PROTEIN_KINASE_ACTIVITY	0.154
	SYSTEM_PROCESS	0.161
	ION_TRANSPORT	0.167
	NEURON_DEVELOPMENT	0.168
	AMINE_TRANSPORT	0.169
	PROTEIN_TYROSINE_KINASE_ACTIVITY	0.170
	SKELETAL_DEVELOPMENT	0.180

H

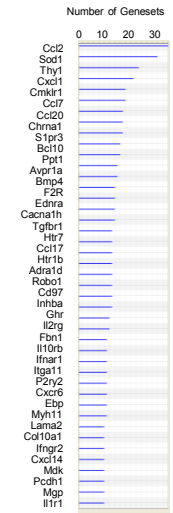


J

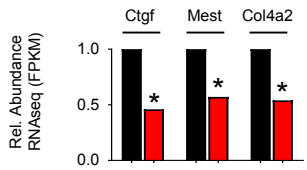
GSEA Hallmark Pathways down RPAP1-depletion, MEFs at day 3.

	GSEA Hallmark Pathways (from total of 50 genesets)	FDR q-value
UP	E2F TARGETS	0.040
	DNA REPAIR	0.193
	G2M CHECKPOINT	0.230
DOWN	EPITHELIAL MESENCHYMAL TRANSITION	0.000
	INFLAMMATORY RESPONSE	0.000
	ANGIOGENESIS	0.004
	IL6_JAK STAT3 SIGNALING	0.010
	UV RESPONSE DN	0.011
	ALLOGRAFT REJECTION	0.013
	HYPOXIA	0.020
	INTERFERON_GAMMA RESPONSE	0.030
	MYOGENESIS	0.036
	APICAL JUNCTION	0.044
	MTORC1 SIGNALING	0.046
	COAGULATION	0.061
	TNFA SIGNALING VIA NFKB	0.062
	ANDROGEN RESPONSE	0.062
	K-RAS SIGNALING_UP	0.063
	APOPTOSIS	0.064
	GLYCOLYSIS	0.074
	ESTROGEN RESPONSE EARLY	0.098
	COMPLEMENT	0.110
	APICAL SURFACE	0.206
	ESTROGEN RESPONSE LATE	0.243
	PEROXISOME	0.251
	PI3K_AKT_MTOR SIGNALING	0.254
	NOTCH SIGNALING	0.260
	IL2_STAT5 SIGNALING	0.287

K



I



L

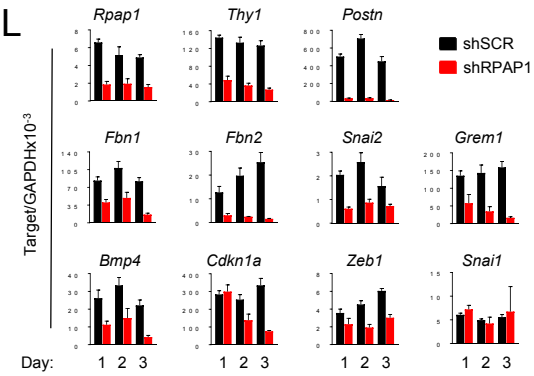


Figure S2. RPAP1 is required for the establishment and maintenance of cell identity, Related to Figure 2.

(A) Following shRNA-knockdown of RPAP1, ES cells were differentiated for 24 hours by LIF-removal, then fixed and scored per colony for morphology and Alkaline Phosphatase staining intensity. Photographs show examples of the delay in colony morphology changes and delay in attenuation of AP-staining intensity associated with RPAP1-depletion at +24 hours after LIF-removal. Mean+/-SEM, n=3 replicates; *p<0.05.

(B) FACS analyses for apoptosis levels by AnnexinV/Propidium Iodide double-staining in ES cells following induction of differentiation by LIF-removal for 24 hrs and then addition of retinoic-acid for 48hrs.

(C) qPCR analyses of pluripotency or cardiac development markers at the indicated time points from the EB differentiation assay in **Figure 2A**. Mean+/-SD, n=3 replicates; *p<0.05. Panel on right: Western blot confirming RPAP1 knockdown in ES cells during self-renewal.

(D) On left, table summarizing the generation of chimeric mice using CSI619, RPAP1(+/Trap) ES cells (1 wild-type and 1 null allele). A low percentage of chimeric pups survived to birth (21/254 micro-injected embryos), of which 10/21 pups displayed moderate chimerism based on coat colour 20-60% Agouti coat colour). On right, table summarizing the offspring generated by mating chimeric mice with eachother, or with wild-type mice, to look for germline transmission. Of 156 pups born from these matings, 7 pups had Agouti coat colour, indicating that the parental RPAP1(+/Trap) ES cells were viable. However, no pups carried the RPAP1(+/Trap) genotype, suggesting that a single RPAP1 allele was insufficient for germline transmission.

(E) Overview of differential gene expression in RNA-Seq transcriptome analysis of ES cells following RPAP1-knockdown then differentiation for 24 hours, as above, in **Figure S2A**. Proportional representation pie-chart indicates the proportion of mRNAs significantly up- or down-regulated with FDR q<0.05. See also **Table S1**.

(F,G) Table summarizing the most significantly up- or downregulated Hallmark genesets (**F**), or GO-term genesets (**G**), identified by GSEA analysis in RPAP1-depleted ES cells after 24 of differentiation, as above, in **Figure S2A** (FDR q<0.05; see also **Tables S1 and S2**). Genesets with FDR q<0.25 are significant.

(H) GSEA Leading Edge analysis of the most prevalent genes among those GO terms database genesets which were significantly downregulated in RPAP1-depleted ES cells after 24 hrs of differentiation, as above, in **Figure S2A** (FDR q<0.05; see also **Tables S1 and S2**).

(I) Normalized RNA-seq expression levels of mesenchymal, fibroblastic and development markers in RPAP1-depleted ES cells after 24 of

differentiation, as above, in **Figure S2A**. Data based on Mean FPKM values, n=3 replicates; * FDR q-value <0.05. See also **Table S1**.

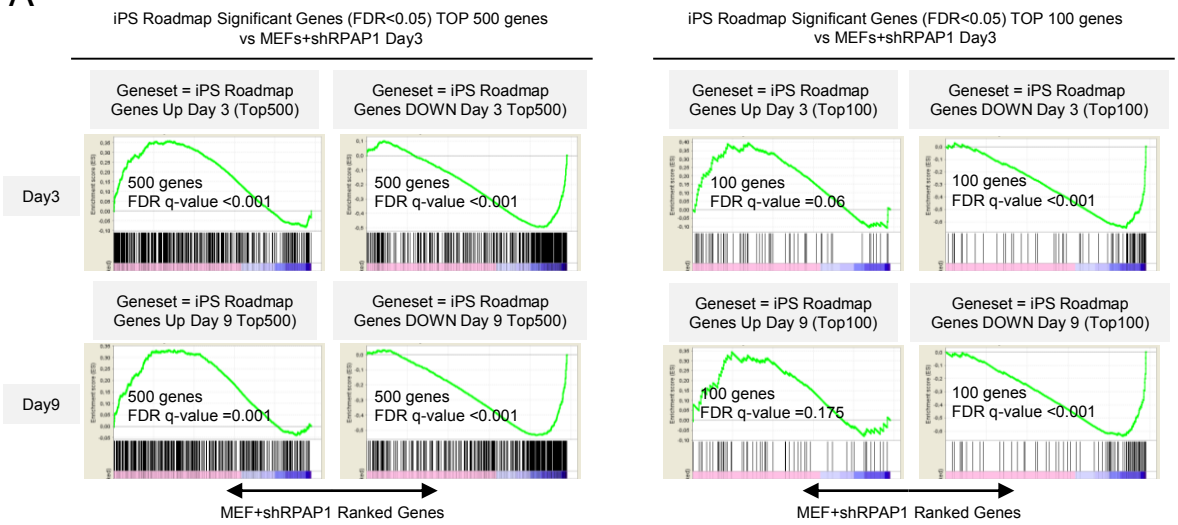
(J) Table summarizing the most significantly up- or down-regulated hallmark genesets identified by GSEA analysis of RNA-sea data at day 3 after RPAP1 depletion in MEFs (FDR q<0.05; see also **Tables S2 and S4**). Hallmark genesets with FDR q<0.25 are significant.

(K) Table summarizing GSEA Leading Edge analysis of the most prevalent genes among the genesets which were significantly downregulated at day 3 after RPAP1 depletion in MEFs (FDR q<0.05; see also Table S6) in a comparison versus the GSEA C5 GO terms database.

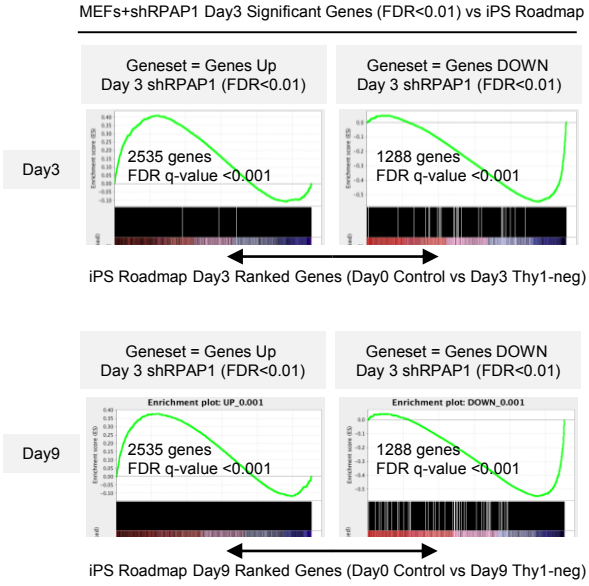
(L) qPCR analyses of fibroblastic, mesenchymal and development markers at the indicated early time points after RPAP1 shRNA depletion in MEFs. Mean+/-SD, n=3 replicates. Raw data from **Figure 2G** is displayed here relative to the housekeeper internal control *Gapdh* (whereas, in **Figure 2G**, the data is shown as fold-change, normalized to the shSCR non-targeting control).

Figure S3

A



B



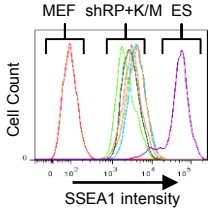
C

Control =KSR/LIF standard iPS media

-plus 12 growth-factor/inhibitor combinations

Growth Promoters +Inhibitors:	Growth Factor Combinations
FGF2 @10ng/mL	Control
EGF @20ng/mL	FGF2
Alki (ALK4/5/7, SB431542) @2/4uM	FGF2/EGF
DLPC @1-5uM	FGF2/Alki
SCF @10ng/mL	FGF2/DLPC
Forskolin @5uM	FGF2/EGF/Alki
5-AzadC @0.5-2uM	FGF2/EGF/DLPC
VPA @0.5mM	FGF2/EGF/Alki/DLPC
TSA @20nM	FGF2/EGF/Alki/Scf
BIX @100uM	FGF2/AZA
Kenpaullone @5uM	FGF2/VPA
Flavopiridol @50nM	FGF2/TSA
	FGF2/Kenpaullone
	FGF2/Forskolin
	FGF2/Flavopiridol

D



E

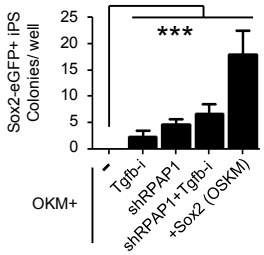


Figure S3. RPAP1-knockdown favors de-differentiation and reprogramming, Related to Figure 3.

(A) GSEA comparison of gene expression at day 3 after RPAP1 depletion in MEFs, versus, a published iPS roadmap gene expression profile (Polo et al., 2012). See Methods for assessment of the iPS Roadmap data from control MEFs versus Thy1-negative cells at day 3, or day 9, of reprogramming. The data here can be compared with **Figure 3A**. GSEA comparison of the published top 500 genes (on left), or Top 100 genes (on right (here, the reduced geneset size was used), up- or down-regulated at day 3, or day 9, of the iPS roadmap, versus, a ranked list of the gene expression profile at day 3 after RPAP1 depletion in MEFs in the current study (x-axis). FDR $q<0.25$ are significant.

(B) GSEA comparison of the significantly up- or down-regulated genes (FDR $q<0.01$) at day 3 after RPAP1 depletion in MEFs, versus, a ranked list of the published gene expression profile of the iPS roadmap at day 3, or day 9 (x-axis) as indicated. FDR $q<0.25$ are significant.

(C) Left: a list of growth factors and small molecule inhibitors, and the concentrations used, to test culture media supplementation in relation to the iPS reprogramming screen in **Figure 3E**. Right: the list of 12 combinations of the media supplements that were tested.

(D) Cells resembling putative reprogramming intermediates, which were generated by shRPAP1+*Klf4*/*cMyc* over-expression (see **Figure 3E**), were expanded to passage 4, then independent clones were analyzed for surface expression of SSEA1 by FACS. MEFS and ES cells were included as negative and positive controls for SSEA1 staining respectively.

(E) Sox2-eGFP-positive iPS colonies were counted per well at day14 following retroviral expression of the indicated combinations of Yamanaka factors, RPAP1 depletion, and/or TGFβ-signaling inhibition (see: Experimental Procedures). Mean +/-SD, 3 replicates; ***$p<0.001$, versus the control (lane 1).

Figure S4

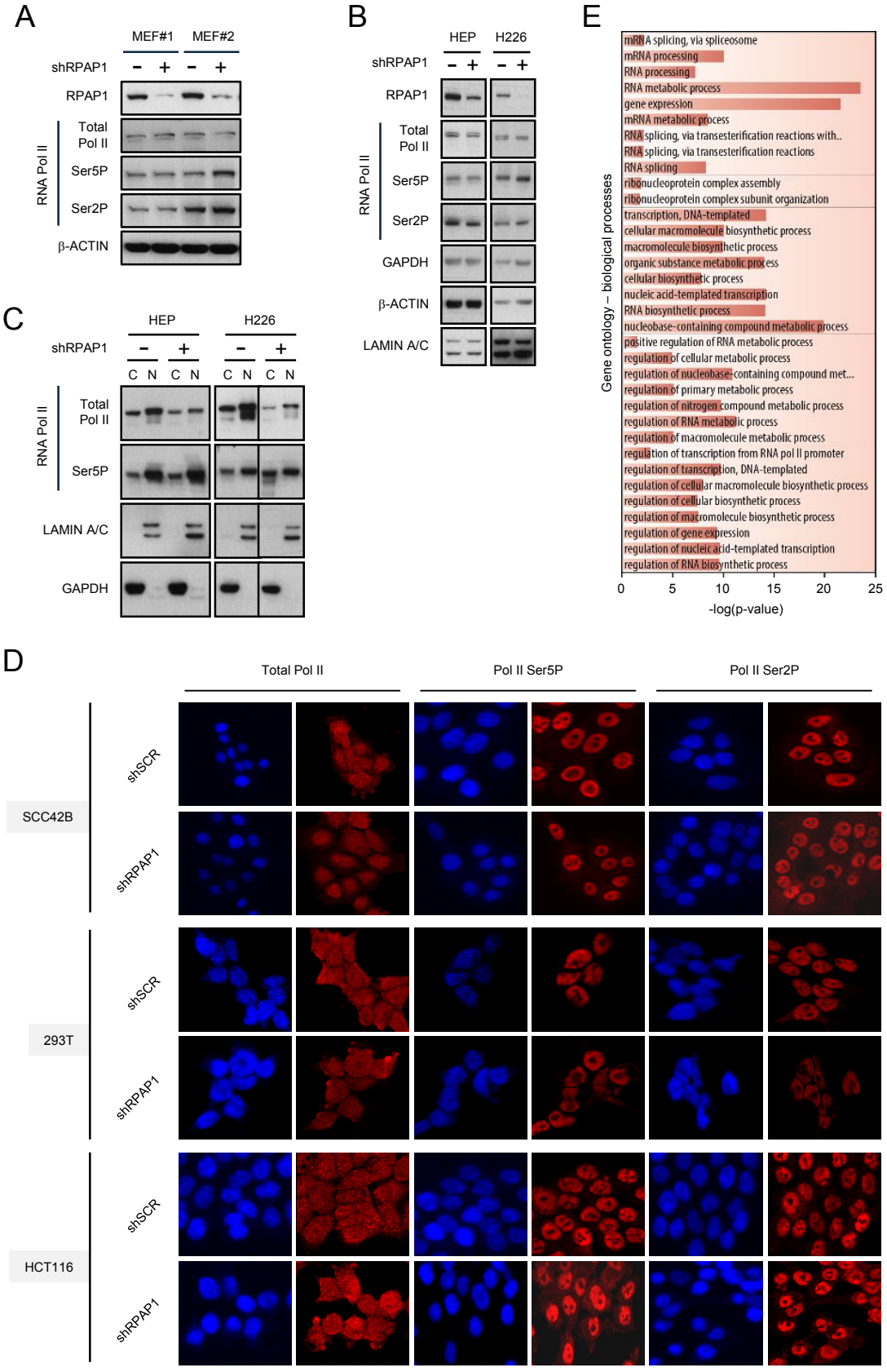

A
MEF#1
MEF#2
shRPAP1
RPAP1
RNA Pol II
Total Pol II
Ser5P
Ser2P
β-ACTIN
B
HEP
H226
shRPAP1
RPAP1
RNA Pol II
Total Pol II
Ser5P
Ser2P
GAPDH
β-ACTIN
LAMIN A/C
C
HEP
H226
shRPAP1
C N C N C N C N
RNA Pol II
Total Pol II
Ser5P
LAMIN A/C
GAPDH
E
mRNA splicing, via spliceosome
mRNA processing
RNA processing
RNA metabolic process
gene expression
mRNA metabolic process
RNA splicing, via transesterification reactions with..
RNA splicing, via transesterification reactions
RNA splicing
ribonucleoprotein complex assembly
ribonucleoprotein complex subunit organization
transcription, DNA-templated
cellular macromolecule biosynthetic process
macromolecule biosynthetic process
organic substance metabolic process
cellular biosynthetic process
nucleic acid-templated transcription
RNA biosynthetic process
nucleobase-containing compound metabolic process
positive regulation of RNA metabolic process
regulation of cellular metabolic process
regulation of nucleobase-containing compound met...
regulation of primary metabolic process
regulation of nitrogen compound metabolic process
regulation of RNA metabolic process
regulation of macromolecule metabolic process
regulation of transcription from RNA pol II promoter
regulation of transcription, DNA-templated
regulation of cellular macromolecule biosynthetic process
regulation of cellular biosynthetic process
regulation of macromolecule biosynthetic process
regulation of gene expression
regulation of nucleic acid-templated transcription
regulation of RNA biosynthetic process
Gene ontology – biological processes
0
5
10
15
20
25
-log(p-value)
D
Total Pol II
Pol II Ser5P
Pol II Ser2P
SCC42B
293T
HCT116
shSCR
shRPAP1

Figure S4. RPAP1 regulates the Pol II interactome, not its expression or localization, Related to Figure 4.

(A,B) Western blots of RPAP1, Pol II Total (RPB1), Ser5P, or Ser2P expression in whole cell lysates from two independent MEF lines at day 3 (**A**), or from HEP and H226 cell lines (**B**), at day 3 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1). GAPDH, β-ACTIN, and LAMIN A/C used as internal loading controls.

(C) Western blots of Pol II Total (RPB1), Ser5P, or Ser2P expression in Nuclear/Cytoplasmic fractions from HEP and H226 cell lines at day 3 after RPAP1 depletion. GAPDH and LAMIN A/C used as indicators of fraction separation. N, nuclear fraction. C, cytoplasmic fraction.

(D) Immunofluorescence of Pol II Total (RPB1), Ser5P or Ser2P in a range of human cell lines at day 3 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1). Nuclei stained with DAPI.

(E) Gene ontology analysis for the enrichment of biological processes among the Pol II interactors lost following RPAP1-depletion, with p-value corrected for multiple testing (Bonferroni).

Figure S5

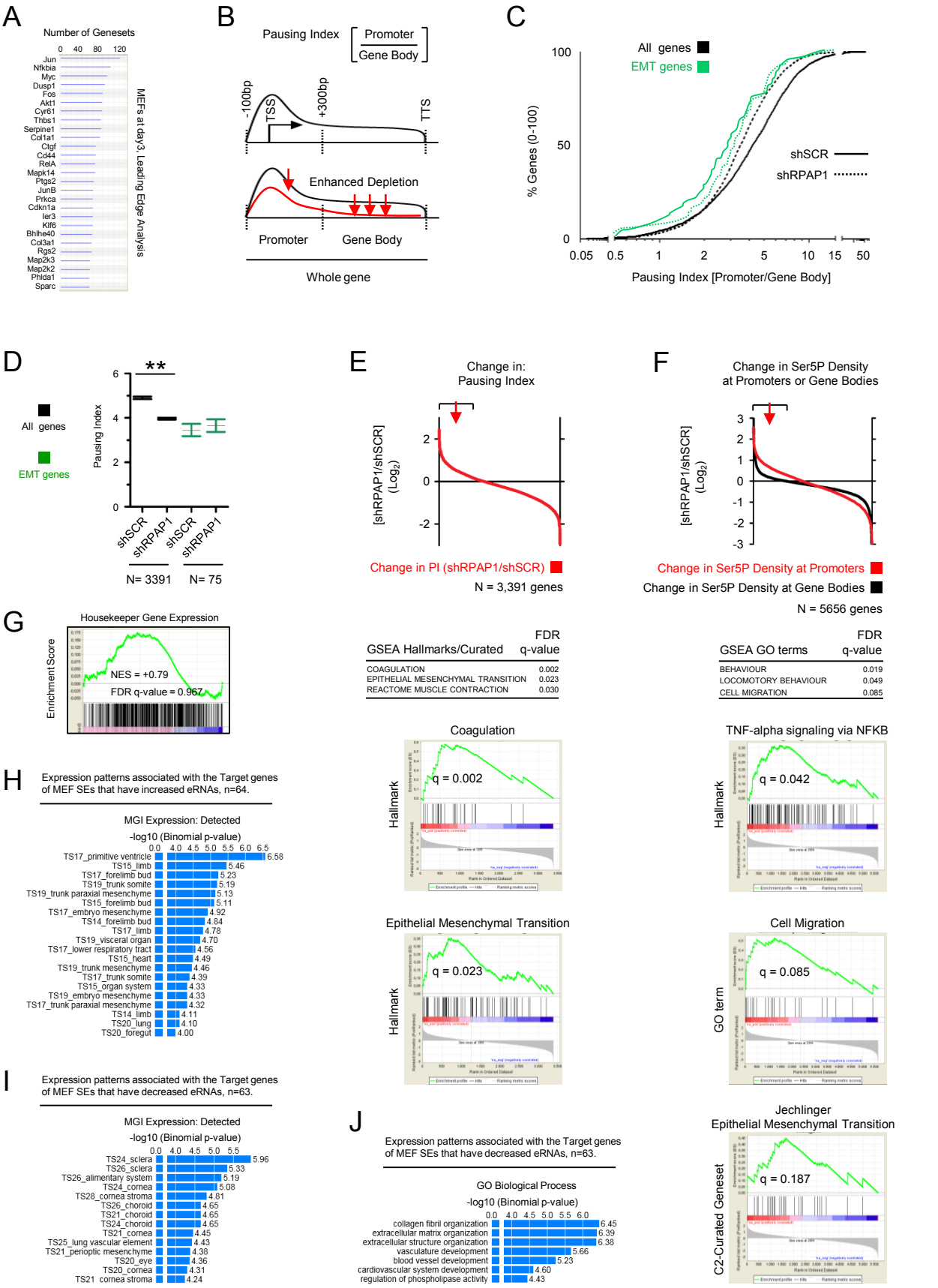

A
Number of Genesets
0 40 80 120
Jun
Nfkbia
Myc
Dusp1
Fos
Akt1
Cyr61
Thbs1
Serpine1
Col1a1
Ctgf
Cd44
RelA
Mapk14
Ptgs2
JunB
Prkca
Cdkn1a
Ier3
Klf6
Bhlhe40
Col3a1
Rgs2
Map2k3
Map2k2
Phlda1
Sparc
MEFs at day3, Leading Edge Analysis
B
Pausing Index
Promoter
Gene Body
-100bp
TSS
+300bp
TTS
Enhanced Depletion
Promoter
Gene Body
Whole gene
C
All genes
EMT genes
shSCR
shRPAP1
% Genes (0-100)
100
50
0
0.05 0.5 1 2 5 10 15 50
Pausing Index [Promoter/Gene Body]
D
All genes
EMT genes
Pausing Index
**
shSCR
shRPAP1
shSCR
shRPAP1
N= 3391
N= 75
E
Change in:
Pausing Index
[shRPAP1/shSCR] (Log2)
Change in PI (shRPAP1/shSCR)
N = 3,391 genes
F
Change in Ser5P Density
at Promoters or Gene Bodies
[shRPAP1/shSCR] (Log2)
Change in Ser5P Density at Promoters
Change in Ser5P Density at Gene Bodies
N = 5656 genes
G
Housekeeper Gene Expression
Enrichment Score
NES = +0.79
FDR q-value = 0.967
GSEA Hallmarks/Curated
FDR q-value
COAGULATION 0.002
EPITHELIAL MESENCHYMAL TRANSITION 0.023
REACTOME MUSCLE CONTRACTION 0.030
GSEA GO terms
FDR q-value
BEHAVIOUR 0.019
LOCOMOTORY BEHAVIOUR 0.049
CELL MIGRATION 0.085
Coagulation
Hallmark
q = 0.002
TNF-alpha signaling via NFKB
Hallmark
q = 0.042
Epithelial Mesenchymal Transition
Hallmark
q = 0.023
Cell Migration
GO term
q = 0.085
H
Expression patterns associated with the Target genes of MEF SEs that have increased eRNAs, n=64.
MGI Expression: Detected
-log10 (Binomial p-value)
TS17_primitive ventricle 6.58
TS15_limb 5.46
TS17_forelimb bud 5.23
TS19_trunk somite 5.19
TS19_trunk paraxial mesenchyme 5.13
TS15_forelimb bud 5.11
TS17_embryo mesenchyme 4.92
TS14_forelimb bud 4.84
TS17_limb 4.78
TS19_visceral organ 4.70
TS17_lower respiratory tract 4.56
TS15_heart 4.49
TS19_trunk mesenchyme 4.46
TS17_trunk somite 4.39
TS15_organ system 4.33
TS19_embryo mesenchyme 4.33
TS17_trunk paraxial mesenchyme 4.32
TS14_limb 4.11
TS20_lung 4.10
TS20_foregut 4.00
I
Expression patterns associated with the Target genes of MEF SEs that have decreased eRNAs, n=63.
MGI Expression: Detected
-log10 (Binomial p-value)
TS24_sclera 5.96
TS26_sclera 5.33
TS26_alimentary system 5.19
TS24_cornea 5.08
TS28_cornea stroma 4.81
TS26_choroid 4.65
TS21_choroid 4.65
TS24_choroid 4.65
TS21_cornea 4.45
TS25_lung vascular element 4.43
TS21_perioptic mesenchyme 4.38
TS20_eye 4.36
TS20_cornea 4.31
TS21_cornea stroma 4.24
J
Expression patterns associated with the Target genes of MEF SEs that have decreased eRNAs, n=63.
GO Biological Process
-log10 (Binomial p-value)
collagen fibril organization 6.45
extracellular matrix organization 6.39
extracellular structure organization 6.38
vasculature development 5.66
blood vessel development 5.23
cardiovascular system development 4.60
regulation of phospholipase activity 4.43
Jechlinger
Epithelial Mesenchymal Transition
C2-Curated Geneset
q = 0.187

Figure S5. RPAP1 is required for Pol II transcription in MEFs, particularly on developmental and mesenchymal genes, Related to Figure 5.

(A) GSEA Leading Edge analysis. The genesets which were significantly depleted in Pol II abundance were identified in MEFs at day 3 after RPAP1 depletion (see **Table S2; and Figure 5E**). The table lists the most prevalent genes among the GO term genesets.

(B) Definition and analysis of Pol II loading ratio on Promoter-Body (or Pausing Index, PI). Schematics outline the parameters used to define the whole gene, promoter, gene body, and Pausing Index (PI) in this study, (see also **Supplemental Experimental Procedures and Table S5**). An example of preferential depletion of Pol II from the gene body is shown, in lower panel.

(C,D) Plots showing the PI ratio for all genes, or all regulators of the Epithelial-Mesenchymal Transition (defined by the GSEA Hallmark geneset #M5930, MySigDB, Broad Institute). In (**D**), data are Mean +/- SEM of "n" genes as indicated; **p<0.01. In (**C**) and (**D**), at day 3 after RPAP1 depletion in MEFs, the Mean PI significantly decreases for many genes (ΔPI <1.0), however, the Mean PI increases for EMT-regulatory genes (ΔPI >1.0). This is consistent with preferential depletion of Pol II from the gene body, as depicted in lower panel of (**B**), above.

(E) Ratio of shRPAP1/shSCR for the change in PI for each gene (ΔPI), at day 3 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1) in MEFs. Arrow highlights the region containing genes with increased PI at their promoters. Table shows the top three GSEA results which identify that genesets and genes with increased pausing index (region highlighted in plot) are enriched for MEF cell identity and developmental regulators (FDR q<0.25 is significant). Below: examples of GSEA plots for the most significantly enriched genesets with increased PI (see also **Table S5** for PI calculations per gene, and **Table S2** for full GSEA results).

(F) Graph of the change in Ser5P density comparing shRPAP1/shSCR, at the promoter (red), or in the gene body, (black). Data from day 3 after lentiviral control (shSCR) or RPAP1 depletion (shRPAP1) in MEFs. Arrow highlights the region containing genes with increased Pol II Ser5P density at their promoters. GSEA analyses were performed on the entire ranked lists for promoters (red line) or the gene bodies (black line), however, significant enrichment of genesets was only observed for those genes with increased Pol II Ser5P density at their promoters (Red plot line, region as indicated by arrow). Table below shows the top three GSEA results which identify that genes with increased promoter Ser5P density (region highlighted in red plot line by arrow, above) are enriched for cell identity and developmental regulators (FDR q<0.25 is significant). Below: examples of GSEA plots for the most significantly enriched genesets with

increased Ser5P density at their promoter (see also **Table S5** for Ser5P density calculations per gene, and **Table S2** for full GSEA results).

(G) GSEA to assess mRNA expression levels of housekeeper genes, (as defined: see **Supplemental Experimental Procedures**) in primary MEFs at day3 after RPAP1 knockdown. No significant change in housekeeper geneset expression was observed.

(H) Developmental stages significantly associated with the super-enhancer target genes ($P < 10^{-4}$) where the enhancers display increased eRNA levels in MEFs at day 3 after RPAP1 knockdown.

(I,J) Developmental stages (**I**) and GO Biological Processes (**J**) significantly associated with the super-enhancer target genes ($P < 10^{-4}$) where the enhancers display decreased eRNA levels in MEFs at day 3 after RPAP1 knockdown.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Content list:

Data and Software availability

Accession numbers: Three datasets (two RNA-seq and one ChIP-seq experiment) are available from the GEO database, Accession: GSE78795. The mass spectrometry proteomics data are available from the ProteomeXchange Consortium/PRIDE repository with the dataset identifier PXD007114.

RESOURCE TABLE

Primers, Antibodies, shRNAs, gRNAs and Cas9 expression systems used in this study are listed in the following Resource Tables:

PRIMERS

qRT-PCR primers used in this study		
Target mouse genes	**Forward**	**Reverse**
β-Actin	GGCACCACACCTTCTACAATG	GTGGTGGTGAAGCTGTAGCC
Gapdh	TTCACCACCATGGAGAAGGC	CCCTTTTGGCTCCACCCT
Pou5f1/Oct4	TCTTTCCACCAGGCCCCCGGCTC	TGCGGGCGGACATGGGGAGATCC
Sox2	TAGAGCTAGACTCCGGGCGATGA	TTGCCTTAAACAAGACCACGAAA
Klf4	GCGAACTCACACAGGCGAGAAACC	TCGCTTCCTCTTCCTCCGACACA
Nanog	CAGGTGTTTGAGGGTAGCTC	CGGTTCATCATGGTACAGTC
RPAP1	CACCCTTCTCTGCCTGGGCC	TAGCAGCTGCGGATGCTGGG
E-cadherin /Cdh1	TTTTCGGAAGACTCCCGATTCA	AGCTTGTGGAGCTTTAGATGC
N-cadherin /Cdh2	CTGATAGCCCGGTTTCACTTG	CAGGCTTTGATCCCTCTGGA
Zscan4c	GAGATTCATGGAGAGTCTGACTGATGAGTG	GCTGTTGTTTCAAAAGCTTGATGACTTC
BMP4	CGCTTCTGCAGGAACCAATGGAGC	CCGGTCTCAGGTATCAAACTAGC
Cytokeratin14	GACCATCGAGGACCTGAAGAGCAAGAT	GCCTCCACGCTCATGCGCAGGCTC

Cardiac α-Actinin	CTGGTATTGCCGATCGTATG	CTTGCTGATCCACATTTGCT
Atrial Natriuretic Peptide	ACTAGGCTGCAACAGCTTCC	TGACACACCACAAGGGCTTA
Snai1	CACACGCTGCCTTGTGTCT	GGTCAGCAAAAGCACGGTT
Snai2	TGGTCAAGAAACATTTCAACGCC	GGTGAGGATCTCTGGTTTTGGTA
Zeb1	GCTGGCAAGACAACGTGAAAG	GCCTCAGGATAAATGACGGC
Zeb2	CAGGCTCGGAGACAGATGAAG	CTTGCAGAATCTCGCCACTG
Twist1	GGACAAGCTGAGCAAGATTCA	CGGAGAAGGCGTAGCTGAG
Cdkn1a/p21	GTGGGTCTGACTCCAGCCC	CCTTCTCGTGAGACGCTTAC
Cdkn2a/p16	CGTACCCCGATTCAGGTGAT	TTGAGCAGAAGAGCTGCTACGT
Thy1	TTACCCTAGCCAACTTCACCACCA	AAATGAAGTCCAGGGCTTGGAGGA
Fbn1	TCGAGTCCTACACGAGCCATGG	ACCAGGTAAGGTTTTCCATCCAGG
Fbn2	CTGACGAAGGGTGGTCAGAC	GCCAAGAGCGCACAGAAGGAG
Fn1	AGAGGCAGGCTCAGCAAAT	TGCTTCCCATTGTCAAAACA
Grem1	ACTAGGTGCGCCCTTCGCAGAC	GGTCCTCAGGTTCTTCGCTGTGG
Postn	ACAACAATCTGGGGCTTTTT	AATCTGGTTCCCATGGATGA
Hoxb1	CCCTTCCAACTCAGTTCAGTGCCT	TTGGTGGCGATTGGGCTCACACTC
Ccnd1	TAGGCCCTCAGCCTCACTC	CCACCCCTGGGATAAAGCAC
Meox1	ATTGCATGGTACTTGGGACGATCG	ATATCTCCGGAGCCGGGTCAGG TAG
Meox2	GAGAACTAGAGGCAGAATTTGCCC	CCTGACAGCTCTGACGGAAGAAG
Nup210	TTTATAAAGCTGCAGACAAACAGG	CCATCAAGGACACGGTAGC
Ccl2	TCCACGTGTTGGCTCAGCCAGATG	CAGCTTCTTTGGGACACCTGCTGC
Ccnb1	TGGCCTCACAAAGCACATGA	GCTGTGCCAGCGTGCTAATC
Lce1f	CACTGATCTTGTGCTGTCCACA	CAGCATCCTCCAGAGCTAC

	GTCT	AGCAG
Lce1h	CTGGCTGACTGAGATACCCAC AGATC	CCAAGCTACAGCAGGAAGA CACAG
E2F2	AATTGTGCGATGTGCACCCGCA GG	AGCACCTCGGCTGCCCAGT TCAG
Rasgef1b	GTACTACCATGACAACAACCTC C	TCATCTGGTTCTTATCGCCG TCC
Sprr1a	ACAGAGAACCTGCTCTTCTCTG AG	CTTGGGGTTGCAGGGCTCA GGAAC
Fibrillarin /Fbl	CCGAGGTGGGGGCTTCCAGTCT G	GGATAGCTGCTGCCAGCTT GGAGC
Foxc2/Mfh1	AGAACAGCATCCGCCACAAC	GCACTTTCACGAAGCACTC ATT
Desmoplaki n/Dsp	ACCGTCAACGACCAGAACTC	TTTGCAGCATTTCTTGGATG

ChIP-qPCR primers (locus indicated in brackets)		
Target (mouse genes) (position relative to TSS, size)	**Forward**	**Reverse**
Mapk1 (Erk2) 1F-1R (Promoter: +19 to +160, 122bp)	TCCCACTCGTAGCCCGCCCGTC	CGCAGGAACCGCGCTGC
Map2k1 (Mek1) 2F-2R (Promoter: +25 to +149, 125bp)	GCGGCGTCTCGGAGCGCCGGAGC	AACTCTCGCCTCAGCACACCGGTTC
Map2k2 (Mek2) 1F-1R (Promoter: +32 to +123, 92bp)	TGCGCTGCAGCGTCAGCTTCACTC	TAGGCCGGGCAGAAGGTGGAAGG
Eif1a 1F-1R (Promoter: +12 to +111, 100bp)	TGGCCGGCCGTTGCCTAGGAAG	AACTTGGTACTCACAGTGACC

ANTIBODIES

Antibodies used in this study		
Target	**Company**	**Code**
RPAP1	Proteintech	15138-1-AP
RPAP1	Cosmo Bio	MK14030910
RPAP1	Abcam	ab21827
Nanog	Chemicon/Millipore	#AB5731
Total RNA Pol II (RPB1)	Santa Cruz	sc-899x (N-20)
RNA Pol II Ser-5P	Abcam	ab5131
RNA Pol II Ser-2P	Abcam	ab5095
Gapdh	Sigma	G8795
β-Actin	Sigma	A5441
γ-Tubulin	Sigma	#T6557, CLONE GTU-88 ascites fluid
Lamin A/C	Santa Cruz	sc-6215 (N-18)

shRNAs

shRNAs used in this study			
From: Open Biosytems (#RMM4534-NM_177294; TRC Mission Library) with a pLKO.1 lentiviral backbone.			
From these 5 shRPAP1 shRNA clones we identified that the best knockdown of RPAP1 expression was achieved using clone TRCN0000173186, hereafter "shRPAP1#5".			
shRNA clone	**Clone**	**Target Sequence 21bp (mature antisense)**	**Locus on target transcript**
shRPAP1#1	TRCN0000176144	AAAGGCTCAAACAATGTGCTC	At +4581; in the 3'-UTR
shRPAP1#2	TRCN0000174973	TATCATGATGTAATACGAGTC	At +4616; in the 3'-UTR.
shRPAP1#3	TRCN0000173273	TTGGGAAGCTTATAGTGGAGG	At +4354; in the 3' of Coding region
shRPAP1#4	TRCN0000173170	ATGATACCAAGAGAAGGTGCG	At +1745; Central to Coding region
shRPAP1#5	TRCN0000173186	TTGAGGTTTGGCAAGACTTGG	At +3871; in the 3' of Coding region
shSCR non-targeting	scramble shRNA was acquired from Addgene (plasmid 1864)		Non-targeting

CRISPR-Cas9 guide RNAs and vector systems

CRISPR gRNAs used in this study (related to Figure S1)						
Guide RNA name	Target sequence		Location in Rpap1	Predicted %AA's deleted	Mouse or human RPAP1	Next best hit
Transient system/ pX330	Forward (*additional bases for cloning*)	Reverse (*additional bases for cloning*)				
mRPAP1 CRISPR_A	*CACCG*CACAGACCAAA TCTAGTCAC	*AAAC*GTGACTAGATT TGGTCTGTG*C*	Mid Exon 7	~80%	mouse	16/20
mRPAP1 CRISPR_B	*CACCG*CCCTTTCCTGT GACTAGATT	*AAAC*AATCTAGTCAC AGGAAAGGG*C*	Mid Exon 7	~80%	mouse	18/20 hit in intergenic region
mRPAP1 CRISPR_C	*CACC*GCCTGTGTTCCA TCGCTCTC	*AAAC*GAGAGCGATGG AACACAGGC	Mid Exon 5	~85%	mouse	15/20
mRPAP1 CRISPR_D	*CACCG*TCGTGAGGCAG CGGGTGACC	*AAAC*GGTCACCCGCT GCCTCACGA*C*	Mid Exon 4	~90%	mouse	14/20
Constitutive system/ pLentiCRISPRv2	Forward (*additional bases for cloning*)	Reverse (*additional bases for cloning*)				
mRPAP1 CRISPR_A	*CACCG*CACAGACCAAA TCTAGTCAC	*AAAC*GTGACTAGATT TGGTCTGTG*C*	Mid Exon 7	~80%	mouse	16/20
mRPAP1 CRISPR_B	*CACCG*CCCTTTCCTGT GACTAGATT	*AAAC*AATCTAGTCAC AGGAAAGGG*C*	Mid Exon 7	~80%	mouse	18/20 hit in intergenic region
mRPAP1 CRISPR_C	*CACC*GCCTGTGTTCCA TCGCTCTC	*AAAC*GAGAGCGATGG AACACAGGC	Mid Exon 5	~85%	mouse	15/20
mRPAP1 CRISPR_D	*CACCG*TCGTGAGGCAG CGGGTGACC	*AAAC*GGTCACCCGCT GCCTCACGA*C*	Mid Exon 4	~90%	mouse	14/20
hRPAP1-1	*CACCG*ATGCTGTCGAG ACCGAAGCC	*AAAC*GGCTTCGGTCT CGACAGCAT*C*	Mid Exon 2 at ATG start	100%	human	13/20
hRPAP1-2	*CACCG*CAGAGTCAGTT TCTCGCAGC	*AAAC*GCTGCGAGAAA CTGACTCTG*C*	Mid Exon 2	~99%	human	17/20
hRPAP1-3	*CACCG*TCACCACATCC CGATGGTCC	*AAAC*GGACCATCGGG ATGTGGTGA*C*	Mid Exon 2	~99%	human	13/20
hRPAP1-4	*CACCG*TGCTGTGTTCC TTCGCTCGC	*AAAC*GCGAGCGAAG GAACACAGCA*C*	Mid Exon 4	~97%	human	17/20
Inducible system/ pLKV-U6 in ESCas9 cells	Forward (*additional bases for cloning*)	Reverse (*additional bases for cloning*)				
mRPAP1 CRISPR_A	*CACCG*CACAGACCAAA TCTAGTCAC*GT*	*TAAAAC*GTGACTAGA TTTGGTCTGTG*C*	Mid Exon 7	~80%	mouse	16/20
mRPAP1 CRISPR_C	*CACC*GCCTGTGTTCCA TCGCTCTC*GT*	*TAAAAC*GAGAGCGAT GGAACACAGGC	Mid Exon 5	~85%	mouse	15/20
mRPAP1 CRISPR_D	*CACCG*TCGTGAGGCAG CGGGTGACC*GT*	*TAAAAC*GGTCACCCG CTGCCTCACGA*C*	Mid Exon 4	~90%	mouse	14/20
mRPAP1 CRISPR_E	*CACCG*CATCCCGATGG TCCTGCGGT*GT*	*TAAAAC*ACCGCAGGA CCATCGGGATG*C*	3' end of Exon3	~96%	mouse	14/20
mRPAP1 CRISPR_F	*CACCG*ATGCTGTCCAG ACCGAAGCC*GT*	*TAAAAC*GGCTTCGGT CTGGACAGCAT*C*	Exon 3, in ATG start	100%	mouse	15/20

			codon			

CRISPR-Cas9 Expression System and guide RNA combinations used

Expression System	Delivery Method	Plasmid	guide RNA combinations used (Related to **Figure S1**)
Transient	Electroporation	pX330 (Addgene #42230)	A, B, C, D, AC, BD, A-D
Constitutive (in mouse)	Lentivirus	pLentiCRISPRv2 (Addgene #52961)	A, B, C, D, A-D
Constitutive (in human)	Lentivirus	pLentiCRISPRv2 (Addgene #52961)	1, 2, 3, 4
Inducible	Doxycyclin	pLKV-U6 (Addgene #50946)	A, C, D, E, F, A-F

CONTACT FOR REAGENT AND RESOURCE SHARING

Please contact Manuel Serrano. Manuel.serrano@irbbarcelona.org

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Animal experimentation at the CNIO, Madrid, was performed according to protocols approved by the CNIO-ISCIII Ethics Committee for Research and Animal Welfare (CEIyBA).

Cells and culture conditions

Primary mouse embryo fibroblasts (wild-type, MEFs, passage 2) were obtained at E13.5 from pure inbred C57BL6 background mice, as described previously (Palmero et al., 2001). Immortalized primary mouse hepatocytes HEP cells have been previously described (Lopez-Guadamillas et al., 2016). Mouse P19EC cells, monkey COS7 cells, and the human cell lines 293T, HCT116, SCC42B and H226 were from ATCC. All the above-mentioned cells were maintained in DMEM medium with 10% FBS (Gibco) with antibiotics (penicillin/streptomycin 100 U/ml). The mouse ES cells E14Tg2a.4 (wild-type parental) and CSI619 RPAP1 (+/Trap) mouse ES cells containing a pGT0Lxf genetrap with *LacZ* reporter in Intron8 were from BayGenomics/MMRRC genetrap resource, University of California. Nanog-GFP knockin mouse ES cells (TNGA) were previously described (Chambers et al., 2007) and were shared by the laboratory of Austin Smith. The mouse ES cells R1, G4, doxy-inducible ESCas9 as described (Ruiz et al., 2016). HAP1 cells (a kind gift from T Brummelkamp) were grown in IMDM (Invitrogen) and 15%FBS. The Sox2-eGFP MEFS (Sox2-Promoter/GFP transgenic) were as described (D'Amour and Gage, 2003). Mouse ES cells and iPS cells, were routinely cultured on gelatin-coated plates in either "Serum/LIF" (15% FBS), or Knockout Serum Replacement (KSR, Invitrogen) "KSR/LIF" (15% KSR), in DMEM (high glucose) basal media, with LIF (1000 Units/mL), non-essential amino acids, glutamax and beta-mercaptoethanol plus antibiotics. Where used, the "2i" drug cocktail comprised 1 μM Mek-inhibitor (PD0325901, Axon Medchem, #1408) plus 3 μM GSK3b-inhibitor (CHIR 99021, Axon Medchem #1386) as described (Ying et al., 2008). Reprogrammed iPS cells were initially derived and expanded on mitomycin-C inactivated feeder cells on gelatin-coated plates, before transfer to gelatin-only. Cultures were routinely tested for mycoplasma and were always negative. C57BL/6 ES cells were derived at the Transgenic Mice Unit of the Spanish National Cancer Research Center (commonly abbreviated as CNIO, from the name in Spanish: Centro Nacional de Investigaciones Oncológicas) from E4.5 C57BL6 blastocysts, or mixed background C57BL6/129 blastocysts. ES cell self-renewal and pluripotency was scored, by cytometry (Nanog-GFP heterogeneity and overall intensity), by immunofluorescence (see below), by colony morphology (see **Figures S1D and S2A**), by alkaline phosphatase staining of fixed cells (Promega #S3771), and by qPCR for stemness markers Nanog, Oct4 and Sox2 (See: **Figures 2B and S2C**), in addition to their differentiation capacity in retinoic acid or embryoid body cardiac centre development (see below). To inhibit CRM1-dependent nuclear export, cells were treated for 3 hrs

with 10 nM Leptomycin B (Sigma #L2913). For proliferation curves, cells were counted and serially passaged every 3 days to monitor the cumulative doubling rate. Senescence-associated β-gal staining was performed as described (Munoz et al., 2013). Staining for *Lac*Z expression in the CSI619 ES cells RPAP1(+/Trap), where the genetrap contains a β-geo reporter, was performed as described (Munoz et al., 2013).

METHOD DETAILS

CRISPR/Cas9-based gene editing

To target human or mouse RPAP1 sequences, we used the MIT CRISPR design tool (http://tools.genome-engineering.org) to design the sgRNAs as described (Ran et al., 2013). Six mouse sgRNAs were used targeting mouse/human RPAP1 Exons 4-7 (see **Figure S1I**) or 4 sgRNAs targeting human RPAP1 Exons 2-4, either individually to generate indels, or in combinations to generate deleted regions (see: **Resource Tables,** in **Supplemental Experimental Procedures**, for sgRNA sequences, plasmid details and gRNA combinations used). RPAP1-knockout was assessed by Western blot of entire cellular pools, or derivation and expansion of individual clones.

Briefly, three CRISPR strategies were pursued. Transient CRISPR/Cas9 expression was by electroporation of mouse ES cells (Neon Transfection System; 1200V, 20 msec, 2 pulses) using the pX330 plasmid (Addgene #42230). Constitutive CRISPR/Cas9 expression was by pLentiCRISPRv2 (Addgene: #52961) as described (Ruiz et al., 2010). For the human HAP1 cell line, human specific CRISPR-sgRNAs oligos (**Resource Tables,** in **Supplemental Experimental Procedures**) were cloned into the pLenti-CRISPRV2 (Addgene plasmid #52961). For doxycyclin-inducible CRISPR/Cas9: CRISPR-sgRNAs oligos cloned into the pKLV-U6-gRNA (BbsI)-PGKpuro2ABFP (Addgene #50946) to generate doxy-inducible ESCas9 cells as described (Ruiz et al., 2016). Individual lentiviral vectors pKLV-U6gRNA-PGKpuro2ABFP (Addgene #50946) or pLentiCRISPR v2 (Addgene plasmid #52961) were co-transfected with third generation packaging vectors in 293T cells using Lipofectamine 2000 (Invitrogen) in order to generate viral supernatants as described (Ruiz et al., 2010). A total of 10^5 ES cells were infected in suspension with 500 μls of viral supernatant for 1 hour at 37ºC and plated on a layer of fresh feeder cells. Two days after infection, G4 and R1 ES cells were selected with Puromycin 1μg/ml and maintained for a week in culture in order to allow efficient gene editing. For the doxycycline-inducible ESCas9 cell line (Ruiz et al, 2016), two days after infection, cells were split into media with or without 1 μg/ml doxycycline and maintained for an additional week in culture in order to allow efficient gene editing. In the case of Hap1 cells, spinfection was used to infect as follows: a total of 10^5 HAP1 cells in one 6-well were incubated with 1.5 ml of viral supernatant and centrifuged at 1850 rpm for 1 hour. Two days after infection, cells were selected with Puromycin 1μg/ml and maintained for a week in culture in order to allow efficient gene editing.

Production of retrovirus and lentivirus, and infection of recipient cells

Briefly, retroviral and lentiviral supernatants were produced in HEK-293T cells ($5x10^6$ cells per 100 mm diameter dish). Vector transfections were performed using Fugene-6 transfection reagent (Roche) according to the manufacturer's protocol. Two days later, viral supernatants (10 ml) were collected serially during the subsequent 48 hours, at 12-hour intervals, each time adding fresh medium to the cells (10 ml). The recipient cells were seeded the previous day ($1.5x10^5$ cells per well in a 6-well plate) and each well received 1.0 ml of the corresponding retroviral and/or lentiviral supernatants as indicated in each Figure. This procedure was repeated every 12 hours for 2 days (a total of 4 additions).

For lentiviral shRNA production, per dish, 293T cells were transfected with 3 plasmids: (i) the ecotropic lentiviral envelope packaging plasmid pMD2.G (0.3 μg; Addgene, plasmid #12259; containing the VsVg gene); (ii) the lentiviral packaging plasmid pCMV-dR8.91 (3.0 μg); (from: Harvard Medical School, plasmid #516); (iii) plus either one of the following 6 lentiviral shRNA constructs (3.0 μg) expressing mouse shRNAs against RPAP1 (shRPAP1#1-5, respectively), or the corresponding non-targeting control (Scramble, shSCR) vector. After lentiviral infection was completed, lentiviral RPAP1-knockdown shRNA recipient cells were selected with puromycin (1 μg/ml). A panel of five lentiviral shRNA against RPAP1 were from Open Biosystems (#RMM4534-NM_177294; TRC Mission Library) with a pLKO.1 lentiviral backbone. From these 5 clones we identified that the best knockdown of RPAP1 expression was achieved using clone TRCN0000173186, hereafter "shRPAP1#5". See shRNA clone details in **Resource Tables,** in **Supplemental Experimental Procedures**.

For retrovirus, per dish, 293T cells were transfected with the ecotropic packaging plasmid pCL-Eco (4 μg) together with one of the following retroviral constructs (4 μg): pMXs-Oct4, pMXs-Sox2, pMXs-Klf4, pMXs-cMyc, or pMXs-Nanog (obtained from Addgene and previously described (Takahashi and Yamanaka, 2006) -the backbone is pMXs plasmid in all cases and the expression of the coding sequences of the reprogramming factors are driven by the MMLV LTR promoter.

Generation of iPS cells from primary MEFs or i4F-MEFs

For retroviral-mediated iPS reprogramming of primary (passage 2-4) mouse embryo fibroblasts was performed by a previous protocol (Li et al., 2009a). Briefly, after infection of primary MEFs with retrovirus expressing the four Yamanaka transcription factors (OSKM), as outlined above, MEF media was replaced by KSR/LIF medium (see above). Cultures were maintained in the absence of drug selection with medium changes every 48 hrs (Li et al., 2009a).

For reprogramming of the secondary-system doxycyclin-inducible 4-Factor (i4F) MEFs which inducibly-express the four Yamanaka factors Oct4, Sox2, Klf4, and cMyc (OSKM) was performed as previously described (Abad et al., 2013). Briefly, i4F-MEFs were treated with doxycyclin (1 μg/mL) continuously to induce expression of

the OSKM transcription factors in the presence of the KSR/LIF iPS medium described above, which was replaced every 48 hrs..

After 7-10 days, iPS colonies with ES-like morphology were counted as they became visible and were subsequently scored by Alkaline Phosphatase staining according to manufacturer's protocol (AP detection kit, Chemicon International, or, Promega #S3771). Colonies of iPS cells were picked after 2 weeks and expanded on feeder fibroblasts using standard procedures. Sox2-eGFP MEFs (D'Amour and Gage, 2003) were used in iPS reprogramming experiments since they become Sox2-GFP-positve (reflecting activation of the endogenous pluripotency network) only in the final stages of iPS reprogramming (see: **Figures 3E, 3F and S3E**).

Growth factors and small molecules to improve iPS reprogramming

The media supplements to improve iPS reprogramming, at the indicated concentrations shown in Figure S4C, are as follows: FGF2 (R+D Systems #233-FB/CF); EGF (Sigma # E9644); Alki (SB431542; ALK4/5/7 inhibitor; Sigma# #S4317); Forskolin (Sigma # F6886); SCF (R+D Systems #455-MC/CF); DLPC (Lrh1 agonist; Stratech # 850335P); 5-Aza-Deoxycitidine (Sigma # A3656-5MG); VPA (Calbiochem # 676380); TSA (Trichostatin A; Sigma; T8552); BIX (BIX 01294; Tocris #3364); Kenpaullone (Tocris #1398); Flavopiridol (Santa Cruz # CAS 146426-40-6).

Differentiation with retinoic acid

Differentiation of ES cells with retinoic acid (RA) was performed essentially as described (Savatier et al., 1996). LIF was first removed for 24 hrs by culture in LIF-free Differentiation medium (that is DMEM (high glucose) supplemented with serum 15%, non-essential amino acids, glutamax and beta-mercaptoethanol; hereinafter referred as "differentiation medium"). Next, LIF-free differentiation media was supplemented with Retinoic Acid at 10 μM from +24 to +72 hrs, followed by LIF-free differentiation medium alone from +72 to +96 hrs. P19EC cell differentiation was by Retinoic Acid addition at 10 μM.

EB hanging-drop differentiation

This was performed essentially as described (Marikawa et al., 2009). ES cells were transferred to Differentiation medium (that is DMEM (high glucose) supplemented with serum 15%, non-essential amino acids, glutamax and beta-mercaptoethanol; hereinafter referred as "differentiation medium"), and suspended in hanging drop culture at a cell density of 5000 cells/20 μLs. ES cells were allowed to form spherical aggregates known as Embryoid Bodies (EBs) for 48 hours in the hanging drops before transfer to suspension culture in low-adherence petri-dishes. In suspension culture, fresh Differentiation medium was added every 3 days, and the percent of EBs was scored daily for the development of beating cells in cardiac centres.

Wound healing scratch assay

Three MEF clones were assessed for their ability to migrate and close a scratched region at day 3 +/- RPAP1 depletion. Scratch wounds (12 per experimental condition)

were made in shSCR and shRPAP1 cultures and photographed at both +0 and +24 hrs in order to quantify the percent area of the original damage which remained at +24 hrs, using ImageJ software analyses of the photographs.

Cytometry

FACS was performed as described (Li et al 2009a). Briefly, for SSEA1 analysis, cells were collected by scraping and pipetting to unicellularize, before resuspension in 500 μLs 1xPBS and incubation with anti-SSEA1 antibody conjugated to allophycocyanin (R+D Systems, #FAB2155A) for 15 mins at room temperature. For AnnexinV analysis of apoptosis, the cells were collected by trypsinization before re-suspension in 1x binding buffer and incubation with anti-AnnexinV antibody conjugated to FITC (BD Pharmingen, # 556570). Data were analyzed with FlowJo 9.6.2 software. The percent of cells in S-phase was quantified using the Click-iT EdU staining kit (Invitrogen #C35002). Briefly, cells were exposed to EdU in culture for 45 minutes followed by fixation and staining according to the manufacturer's protocol.

Cell lysis and Western blot

Whole cell extracts were prepared using 50 mM TrisHCl pH8; 1 mM EDTA; 150 mM NaCl; 1% NP40; 0.5% Triton X-100; 1.0% SDS, with freshly added protease inhibitors (Roche #11873580001). A total protein of 10 μg was loaded per lane and resolved on NuPAGE 4-12% gradient Bis-Tris gels, transferred to nitrocellulose and hybridized using antibodies as described in **Resource Tables,** in **Supplemental Experimental Procedures**. Nuclear/Cytosolic Fractionation was performed by using the NE-PER Nuclear and Cytoplasmic Extraction Kit by Thermo Scientific, following the manufacturer's instructions.

Histopathology and immunohistochemistry

Mouse tissues were fixed in formalin at 4ºC, embedded in paraffin block, and sectioned at a thickness of 5 μm. Sections were stained with hematoxylin and eosin for pathological examination or processed for immunohistochemical analysis with antibodies against mouse RPAP1 (for a list of the antibodies used, see **Resource Tables,** in **Supplemental Experimental Procedures**). E3.0 morulae and E4.0 blastocyst embryos were collected in KSOM media (Chemicon #3699) and gently resuspended in 10% Formalin at 4C overnight to fix. Next day, embryos were resuspended in 100-200 μls of sterile 5% gelatin/dH_2O pre-warmed at 37C, then placed at 4C to allow gelatin solidification, followed by equilibration of the solid gelatin pellet in cold 10% formalin before embedding in paraffin block, and sectioning as above.

Immunofluorescence

Cells were grown on chamber slides using the same protocols as for the rest of the experiments. Briefly, at day 3 after RPAP1 depletion, cells were fixed with 4% paraformaldehyde for 2 minutes at room temperature, washed with PBS and permeabilized with PBS containing 0.02% Tween-20 for 20 minutes. Cells were blocked in PBS with 50% Australian FBS for 1 h and incubated with antibodies against

RPAP1 or Pol II (for a list of the antibodies used, see **Resource Tables,** in **Supplemental Experimental Procedures**) at 1:200 to 1:1000 in PBS-4%BSA, for 3 hrs, washed with PBS and further incubated with secondary anti-rabbit antibodies conjugated with Alexa-488, Alex-555 and/or Alexa-647 (1:500 in PBS-4%BSA). Nuclei were counter-stained with DAPI. Confocal immunofluorescence cell images were captured using a Leica SP5, equipped with white light laser and hybrid detection.

RNA Pol II interactome analysis and LC/LC mass spectrometry

RNA Pol II immunoprecipitation was performed on Day+2 after lentiviral shRNA knockdown of RPAP1 in primary MEFs. Cells were washed x2 with ice-cold 1xPBS, then scrape-harvested in ice-cold 1xPBS. Lysates were prepared from two replicate experiments, sonicated, and clarified by centrifugation at 10C, at 10,000g, for 10 minutes. The supernatants were pre-cleared by exposure to Protein A/G beads (Santa Cruz #sc-2003). The Pol II complex was immunoprecipitated using a cocktail of three antibodies against RPB1/Polr2a, the largest and core catalytic subunit of Pol II, in order to immunoprecipitate Pol II throughout all the stages of transcription. The antibodies targeted the N-terminus of Pol II (Santa Cruz, sc-899x), the Serine5-phosphorylated C-terminal domain (Abcam #5131), and the Serine-2-phosphorylated C-terminal domain (Abcam #5095). The immunoprecipitate fraction was eluted, specific Pol II protein interactors were determined by Mass Spectrometry, and the Pol II-interactome was analysed, as described below.

Immunoprecipitate sample preparation for mass spectrometry.

Proteins were eluted from the agarose beads in two consecutive steps by shaking for 10 min at 1250 rpm in an Eppendorf Thermomixer in 2 bead volumes (~100 µl) of elution buffer (UT: 8M Urea, 100mM Tris-HCl pH=8.0). The supernatant obtained was digested by means of standard FASP (Filter Aided Sample Preparation) protocol (Wiśniewski et al., 2009). Proteins were then reduced with 10 mM DTT, alkylated using 50 mM IAA for 20 min in the dark. Proteins were digested with Lys-C (Wako, Neuss, Germany) for 6 hours (1:50). Finally, samples were diluted in 50 mM ammonium bicarbonate to reduce the urea concentration to less than 1M, and were subsequently digested with Trypsin (Promega, Madison, WI; 1:100 sample concentration, overnight at 37 °C). Resulting peptides were desalted using a Sep-Pak C18 cartridge for SPE (Waters Corp., Milford, MA). Eluted peptides were vacuum-dried. To comprehensively identify the Pol II interactome, peptides were further pre-fractionated into five fractions using high pH reverse phase micro-columns (Batth et al., 2014), packing three discs (16 g diameter) of 3M Empore C18 at the bottom of a conventional 200 µl micropipette tip. After conditioning the tip, peptides were dissolved in 50 µl of Buffer A (20mM NH_3, pH $\geq$ 10). Using an adapter, the tip was mounted on a 1.5 ml tube and fit in a benchtop centrifuge. During each fractionation step, centrifuge was operated at 1500 g for 2 minutes until all the volume passed through the C18 membrane. Peptides were subsequently eluted increasing the percentage of Buffer B (20mM NH3 in CH3CN) (i.e. 4, 8, 12, 80%) of Buffer B. All

the five fractions and the flow through were dried by speed-vacuum and resuspended in 22 µl 0.5% FA.

LC−MS/MS analysis

The five fractions of the eight different samples were analyzed by RP chromatography using a nanoLC Ultra system (Eksigent, Dublin, CA), directly coupled with a LTQ-Orbitrap Velos instrument (Thermo) via nanoESI (Proxeon Biosystems, Waltham, MA). Peptides were loaded onto a Reprosil-Pur C18 column (3 µm, 400x0.075 mm; Dr. Maisch, Ammerbuch-Entringen Germany), with a trapping column (Prot Trap Column 0.3 x 10 mm, ReproSil C18-AQ, 5 µm), for 10 minutes with a flow rate of 2.5 L/min of loading buffer (0.1% FA). Elution was performed with a 120 minute linear gradient (buffer A: 2% ACN, 0.1%FA; buffer B: 100% ACN, 0.1%FA) at 300 nl/min. Peptides were directly electrosprayed into the mass spectrometer using a PicoTip emitter (360/20 OD/ID µm tip ID 10 µm, New Objective) at 1.4 kV spray voltage with a heated capillary temperature of 325°C and S-Lens of 60%. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS scans using a top 10 method. MS spectra were acquired with a resolution of 60,000 (FWHM) at 400 m/z in the Orbitrap, scanning a mass range between 350 and 1500 m/z (AGC = 1e6, Max IT = 500 ms). Peptide fragmentation was performed using collision-induced dissociation (CID) with read out in the ion trap (AGC = 5e3, Max IT = 100 ms) and a normalized collision energy of 35%.

Protein Pol II-interactome data collection and analysis

Forty raw files (i.e. two experiments "SCR-Pol II vs SCR-IgG" and "shRPAP1-Pol II vs shRPAP1-IgG" with two biological replicates each and fractionated into five fractions), were analyzed using MaxQuant 1.5.3.30 (Cox and Mann, 2008) with Andromeda (Cox et al., 2011) as the search engine against a *Mus musculus* database (UniProtKB/Swiss-Prot, 43,539 sequences). Carbamidomethylation of cysteine was included as fixed modification and oxidation of methionine, acetylation of protein N-terminal were included as variable modifications. Precursor mass tolerance was 20 ppm for the first search, and 4.5 ppm for the main search. Fragment mass tolerance was set to 0.5 Da. Minimal peptide length was set to 6 amino acids and a maximum of two missed-cleavages were allowed. Peptides were filtered at 1% FDR. For protein assessment (FDR <1%) in MaxQuant, at least one unique peptide was required for both identification and quantification. Other parameters were set as default. A total of 4,384 proteins were identified. Afterwards, the "protein-group" file was loaded in Perseus (v1.5.1.6) (Tyanova et al., 2016). After removing proteins annotated as contaminants, only identified by site and/or reversed a total of 3,944 proteins were quantified. Missing values in the IgG runs were replaced by the minimum LFQ value (i.e. 10) detected in the whole experiment. Using the LFQ values, all four possible pairwise comparisons between the two biological replicates of "SCR-Pol II vs SCR-IgG" were calculated. The same four comparisons were calculated for the "shRPAP1-Pol II vs shRPAP1-IgG" experiments. A protein was declared as specific interactor when the log2 enrichment ratio against its IgG was larger than 2.5 in three out of the four

comparisons in at least one of the two IP experiments. In total, 294 proteins were found as specific interactors (see **Table S4**). Among them, we identified all the subunits of the RNA pol II complex (12 proteins) and 28 out of 30 subunits of the Mediator complex. To identify interactors affected upon RPAP1 depletion, the data were normalized using the RPB1/Polr2a bait protein levels. Then, all four possible pairwise comparisons between "shRPAP-Pol II vs SCR-Pol II" experiments were calculated, and proteins were declared to be decreasing in the shRPAP1 if the log2 ratio was smaller than -1.5 in three out of the four comparisons. Proteins were declared to be increasing in the shRPAP-Pol II if the log2 ratio was larger than 1.5 in three out of the four comparisons. The RPB1/Polr2a interactome in cells treated with the shRPAP1 showed alterations, specifically 104 interactors were absent or significantly reduced, while 5 new interactors were found (see **Table S4**). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD007114.

Protein Pol II-interactome functional analysis

The interactors found to be affected in the RPAP1-depleted cells were functionally categorized using Panther database (http://pantherdb.org) by GO molecular function, GO biological process and GO cellular component. Statistical over-representation of GO terms (mouse genome was used as the background data set) was determined with a Binomial test and used the Bonferroni correction for multiple testing. P-values were then –log10 transformed for better graphical representation. These analyses revealed that the affected interactors in shRPAP1 were enriched in processes related to transcription and splicing ($p<0.00001$) (**Figure S4E**).

To find out whether these affected pol II-interactor proteins belong to specific complexes, we mapped our interactome data to the Corum database (Comprehensive resource of mammalian protein complexes) (http://mips.helmholtz-muenchen.de/corum/) that contains more than 3000 manually curated mammalian protein complexes. The number of subunits identified in the interactome data (specific interactors) for each known complex was retrieved. The same mapping was done with the list of interactors found to be affected in the RPAP1-depleted cells. Corum complexes with less than 6 subunits were not considered and redundant complexes (those sharing identical subsets of proteins) were also removed. Several well-known complexes were represented in our dataset of Pol II-interactors which were affected by RPAP1 depletion (**Table S4**). Among them, the Mediator complex (which is formed by 30 subunits) was ranked the highest (**Figure 4E**) with eleven subunits affected following RPAP1 depletion (MED27, MED28, MED9, MED13, MED25, MED22, MED29, MED10, MED31, CDK8, MED14) indicating an important alteration in the functions controlled by this complex.

RNA isolation and quantitative real-time PCR (qPCR)

Total RNA was extracted from cells on column by RNeasy kit with DNA digestion following provider's recommendations (Qiagen #74104) and retro-transcribed into cDNA following manufacturer´s protocol with Superscript Reverse Transcriptase (Life

Technologies). Quantitative real time-PCR was performed using Syber Green Power PCR Master Mix (Applied Biosystems) in an ABI PRISM 7700 thermocycler (Applied Biosystem). Input normalization of all the qRT-PCR data was by the $2^{-\Delta\Delta Ct}$ method (Yuan et al., 2006) using the housekeeping genes β-Actin or Gapdh as indicated in each Figure, and as described (Ortega-Molina et al., 2015). Primers used are in **Resource Tables,** in **Supplemental Experimental Procedures**.

RNA-seq transcriptomic analyses

For RNA-seq, samples of 1 μg of total RNA, with RIN numbers in the range 9.8 to 10 (Agilent 2100 Bioanalyzer), was used. PolyA+ fractions were processed using TruSeq Stranded mRNA Sample Preparation Kit (Agilent). Adapter-ligated library was completed by PCR with Illumina PE primers (8 cycles). The resulting directional cDNA libraries were sequenced for 40 bases in a single-read format (Genome Analyzer IIx, Illumina). The complete set of reads has been deposited in GEO (GSE78795). Sequencing quality for RNA-seq samples was analyzed with FastQC. Reads were aligned to the mouse genome (GRCm38/mm10) with TopHat-2.0.4 (Trapnell et al., 2012) (using Bowtie 0.12.7 (Langmead et al., 2009) and Samtools 0.1.16 (Li et al., 2009c), allowing two mismatches and five multi-hits. Transcripts assembly, estimation of their abundance, and differential expression, were calculated with Cufflinks 1.3.0 (Trapnell et al., 2012), using the mouse genome annotation data set GRCm38/mm10 from the UCSC Genome Browser (Rosenbloom et al., 2015).

Functional analyses of differential gene expression

For differential gene expression lists (see data in **Table S1**: ES cells +24hr differentiation; or **Table S3**: MEFs at day 3 after RPAP1 depletion). Genes were ranked using the FDR q-value statistic to identify significant genes (FDR<0.05 or FDR<0.01, as indicated in the Figures), then by fold change in expression. Selected differentially-expressed genes identified in the RNA-seq were validated by qPCR. Venn diagrams were generated by JVenn (Bardou et al., 2014) and hypergeometric testing was performed to assess any significant overlaps. Pathway analyses were by Ingenuity Pathway Analysis software (www.ingenuity.com). Gene Set Enrichment Analysis (GSEA; Subramanian et al., 2005). GSEAPre-ranked was used to perform a gene set enrichment analysis of annotations from the MsigDB Hallmarks, C5-Gene Ontology (GO) terms, C2-Curated, KEGG, Reactome and NCI databases, with standard GSEA and Leading Edge analysis settings. We used the RNA-seq gene list ranked by statistic, setting 'gene set' as the permutation method and ran it with 1000 permutations for Kolmogorov-Smirnoff correction for multiple testing. We considered only those gene sets with significant enrichment levels (FDR q-value <0.25) (Subramanian et al., 2005) (see: **Table S2**). GSEA Enrichment data were obtained and ranked according to their FDR q-value (see: **Table S2**). Heatmaps of GSEA data (**Figures 2H and 5F**) or qPCR data (**Figure 3G**) were generated using Gene Pattern (Reich et al., 2006).

Supervised network analysis

Investigation of differential gene expression for dominant gene-ontologies or functions was performed by supervised network analyses. Briefly, network analyses were performed starting from the list of differentially expressed genes induced by RPAP1 depletion followed by 24hrs of ES differentiation, or separately, RPAP1 depletion for 3 days in MEFs. Next these lists were used to find gene interaction information in the Metacore™ database, including manually curated experimentally validated interaction data. The interaction datasets generated (including information of the interaction direction –i.e. source and target genes, and interaction effect –i.e. inhibition or activation) were contextualized for obtaining the gene regulatory networks of the RPAP1-depletion and control phenotypes, using an algorithm developed in-house (Crespo et al., 2013; Zickenrott S et al., 2016). Finally, the phenotype-specific networks were compared to identify the pathway enrichment in genes in the "up-regulated" or "down-regulated" lists. In this comparison, we estimate the statistical significance (i.e. enrichment) of the interactions among genes in each category, which constitute an indication of the differences in the regulatory mechanisms underlying the phenotypical changes caused by RPAP1 depletion. See **Table S1, sheets#6-11**; and **Table S3, sheets#6-10**.

Comparison of differential gene expression with the iPS roadmap

Gene expression changes have been comprehensively characterized in the subset of successfully-progressing cells during iPS reprogramming by overexpression of the OSKM Yamanaka factors (Polo et al., 2012; Hansson et al., 2012). We first identified the gene expression changes which occur between day 0 and day +3, or between day 0 and day +9, of successful iPS reprogramming by comparing RNA-seq data in the parental MEFs (day 0) versus day +3 (or day +9) Thy1-negative cells in the published datasets. Next, we used GSEA to compare these iPS roadmap genesets of top 100 or top 500 up- or down-regulated mRNAs versus the complete ranked list of differential gene expression in MEFs at day 3 after RPAP1 depletion. We also performed the analysis in reverse, comparing the genesets of significantly differentially expressed mRNAs up- or down-regulated in MEFs at day 3 after RPAP1 depletion, versus, the complete ranked list of differential gene expression at day 3 (or day 9) of the iPS roadmap. GSEA results are shown in **Figure 3A and S3A and S3B**. Data with $P<0.05$ and $FDR<0.25$ are considered significant.

Conversion of plant gene expression data to mammalian homologs

The effect of RPAP1-mutation on mRNA expression levels was previously published in Arabidopsis (Sanmartin et al., 2011). We converted the published data from plant (31,200 genes; see: **Table S2, sheet#4**) to mammal (mouse) via protein sequence similarity (**Table S2, sheets #5 and #6**), filtering the data by three thresholds: (i) "100% coverage", that is, the whole plant protein is included in the alignment against whole mouse proteins; (ii) the best "% Amino Acid Identity" possible, always greater than 20% (% of amino acids than are totally conserved in both sequences); (iii) the best "% Positive Amino Acids" as possible (this takes into account synonymous amino acids (that is based on similarity in terms of size and charge). We filtered out: 250 genes that

did not map (neither in Arabidopsis TAIR 10 database nor in EnsEMBL), and a further 1933 Arabidopsis genes were without homology/orthology in mouse, however, the majority of these were transposons (**Table S2, sheet#7**). We ran GSEA using the MSigDB Hallmark, C5-GO terms, and C2-Curated databases against the entire remaining ranked list of homologous plant proteins/genes (for the ranked list of genes converted to mouse, see **Table S2, sheets #5 and #6**) to identify significant genesets up or down-regulated by RPAP1-mutation in plants which have a homolog in mouse (see summary of results in **Table S2, sheet #7**). In **Figure 5F**, the heatmap compares the GSEA hallmark database analysis from the plant-mammal conversion above (**Table S2, sheet #7**), versus, the GSEA hallmark database analyses results for three other experiments: (i) GSEA on the ranked list of differential mRNA expression in MEFs at day 3 +/- RPAP1-depletion; (ii) GSEA on the ranked list of differential mRNA expression in ES cells at +24hrs after inducing differentiation, +/- RPAP1-depetion; (iii) GSEA on the ranked list of differential RNA Pol II abundance at the promoter at day 3 +/- RPAP1 depletion.

Chromatin immunoprecipitation (ChIP) and deep-sequencing

ChIP-qPCR was performed as described (Li et al., 2012) with primers listed in **Resource Tables,** in **Supplemental Experimental Procedures** and antibodies for Total Pol II (Santa Cruz N20, sc-899x) and RPAP1 (Cosmo Bio MK14030910). ChIP-seq for Pol II was performed as described (Rahl et al., 2010). Briefly, cells were fixed using 1% formaldehyde, scrape-harvested, resuspended in ChIP lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) and sonicated using Covaris water bath sonicator to generate fragments of 150 to 500 bp. Soluble chromatin was diluted 10 fold in ChIP Dilution buffer (1% Triton X-100, 2 mM EDTA pH 8.0, 150 mM NaCl) precleared with Agarose Protein A/G beads (Santa Cruz), and then incubated with antibody specific for total RNA Pol II (N-20, sc-899x, Santa Cruz) or specific for the RNA Pol II Ser5P-phoshorylated form (Abcam #ab5131). After incubation, immunocomplexes were collected with Agarose Protein A/G beads (Santa Cruz). Next, the immunocomplexes were washed sequentially with Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 150mM NaCl), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 500mM NaCl), LiCl Wash Buffer (0.25M LiCl, 1% NP40, 1% deoxycholate-Na, 1mM EDTA, 10mM Tris-HCl, pH 8.1) and washed twice with TE (10 mM Tris-HCl pH7.5, 1mM EDTA). Immunocomplexes were eluted in ChIP elution buffer (1%SDS, 0.1M NaHCO3) and the crosslinking was reverted by incubation at 65 ºC for 8 hrs with 200 mM NaCl. Samples were treated with Proteinase K and RNase A, and DNA was extracted using Phenol-Chloroform. DNA precipitation was in 100% ethanol with 0.1 M NaAcetate ph5.2 and 2 uLs glycogen (Roche). The DNA pellet was washed with 70% ethanol, and resuspended in ddH2O. Purified chromatin was used for library construction.

For ChIP-seq the amount of DNA used was ~5 ng from each sample (as quantitated by fluorometry). Samples were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters as in Illumina's "TruSeq

DNA Sample Preparation Guide" (part # 15005180 Rev. C). Adapter-ligated libraries were completed by limited-cycle PCR with Q5 High-Fidelity DNA Polymerase (NEB) and Illumina PE primers (15 cycles), and further purified with a double-sided SPRI size selection to obtain a size distribution in the range of 230-500bp. Libraries were applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v5) and sequenced on the Genome Analyzer IIx with SBS TruSeq v5 reagents by following manufacturer's protocols, to 20-25 million reads per sample.

Pol II ChIP-seq data analyses

Definition of promoter and gene body regions (See: **Figure S5B**) and the calculation of Pol II total and Ser5P abundance along genes was based on methods of Young and colleagues (Rahl el al., 2010) (see **Table S5**). Sequencing quality for ChIP-seq samples was analyzed with FastQC (Andrews, 2011). Reads were aligned with Bwa 0.7.5a (Li and Durbin, 2009) to the mouse reference genome (GRCm38/mm10) using the default seed length (32) and allowing 1 mismatch in the seed. SAMtools 0.1.16 (Li et al., 2009b) was used to convert the output alignment SAM files to the BAM file format, sort the alignments and eliminate duplicated reads. BEDTools 2.23.0 (Quinlan, 2014) was used to convert the resulting files to the BED format. All ChIP and input samples were randomly normalized to the same number of reads. Peak calling was performed with MACS 2.0.10.20130712 (Feng et al., 2012) using the input sample as control for each one of the ChIP samples. BigWig files were obtained with bedGraphToBigWig (Kent et al., 2010) from the BedGraph files generated with MACS. Resulting peaks were annotated with PeakAnalyzer 1.4 (Salmon-Divon et al., 2010), and the distribution of peaks was plotted with SeqMiner 1.3.3e (Ye et al., 2014) with color-scaled intensities are in units of reads per million mapped reads (rpm). Transcription Start Sites (TSS) and Transcription Termination Sites (TTS) were identified using the Database of Transcriptional Start Sites (http://dbtss.hgc.jp). Metagenes were aligned +/- 5 Kb around the TSS. The Pausing Index (PI) for gene promoters versus gene bodies was calculated as described (Rahl et al., 2010; see also **Figure S5B**). First, the number of reads per nucleotide was computed with BEDTools 'genomecov'; second, to extend this number to the number of reads per gene promoter or gene body, BEDTools 'map' was used; and third, the Promoter/Body ratio, or Pausing Index (PI) was calculated for each gene promoter or gene body as PI = ((number of reads in region / region size)*scaling factor)*10^5. Scaling factor = (total number of reads in sample/genome length).

Definition of MEF super-enhancers, their target genes, and their eRNA levels

For **Figure 1I**, H3K27ac enrichment before/after ES cell differentiation was defined by H3K27ac ChIP-seq as previously described (Creyghton et al., 2010) using datasets with the following Accession numbers: GSE24164, GSM594578, GSM594585. For **Figure 5G**: MEF super-enhancers were defined by H3K27ac ChIP-seq signal and ranking by ROSE, as previously described (Whyte et al., 2013; Shen et al., 2012; Khan and Zhang, 2016; dbSUPER, http://bioinfo.au.tsinghua.edu.cn/dbsuper/). To identify the single-nearest target gene to each MEF super-enhancer, GREAT analysis was performed as described (GREAT v3.0.0; McLean et al., 2010). In **Figure 5G**, this geneset of MEF

super-enhancer target genes was used in GSEA analysis of the mRNA expression levels of these genes at day 3 after RPAP1 knockdown in primary MEFs in our data.

For **Figure 5H**: the same MEF super-enhancer regions were assessed for enhancer-RNA (eRNA) abundance which has been reported to be proportional to enhancer activity (Andersson et al., 2014). MEF super-enhancers grouped by K-means clustering according to changes in their eRNA levels at day 3 after RPAP1 knockdown into 3 groups: increased (~10%, n=64), decreased (~10%, n=63), no-change/not-detected (~80%). In **Figure 5H**, average RNA abundance on the super-enhancers with increased or decreased eRNA levels was visualized in 50 bp bins from start to end of feature using SeqMINER (Ye et al., 2014).

For **Figure S5G**, GSEA was performed as described above for **Figure 5G**, except here, in order to assess any changes in housekeeper mRNA expression levels. No significant change in housekeeper gene expression was detected, despite performing GSEA using the following housekeeper genesets: (i) a full set of 3384 housekeeper genes (defined in Eisenberg and Levanon, 2013); (ii) 10 sets of 500 genes, each randomly selected from the full 3384 housekeepers, performed so that the individual genes could be visualized in the GSEA enrichment plot. The data shown in **Figure S5G** is representative of one of the random selections of 500 housekeeper genes from the above list of 3384 genes where no significant change in housekeeper mRNA was detected.

For **Figures S5H, S5I, and S5J**: GREAT analysis (GREAT v3.0.0; McLean et al., 2010) was performed on the following two groups of super-enhancers, defined above, to identify the single-nearest target gene of each super-enhancer: (i) super-enhancers with increased eRNA levels (**Figure S5H**); (ii) super-enhancers with decreased eRNA levels (**Figures S5I and S5J**). Next, these two enhancer-target-gene groups were assessed separately for any enrichments in their functions (Gene Ontology Biological process) or the developmental stage associated with their expression (MGI Expression-Detected; Theiler Stage of embryo development). The data is presented in **Figures S5H to S5J**. In **Figure 5I**, the Theiler Stage of embryo development associated with these two enhancer-target-gene groups (the enhancers with increased or decreased eRNA/activity levels) is shown, together with the approximate embryo day-post-coitus (dpc) (emouseatlas.org; Bard et al., 1998). Super-enhancers with decreased eRNA levels (and thus putatively decreased activity) were associated with target-genes expressed during the period embryo dpc E11.5-E17. Conversely, super-enhancers with increased eRNA levels (and thus putatively increased activity) were associated with target-genes expressed during the period embryo dpc E8-E13. Since primary MEFs derive from E13.5, the decrease in activity of enhancers associated with E11.5-E17, coupled with the increase in activity of enhancers associated with E8-E13 mirrors the gene expression analysis in **Figure 2**, where MEFs at Day3 after RPAP1 knockdown appear to have de-differentiated, and in **Figure 3**, where this pattern of de-differentiation correlates significantly with the first 3 days of iPS reprogramming when MEF cell identity is erased (Polo et al., 2012).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless otherwise specified quantitative data are presented as mean +/- SD and significance was assessed by the two-tailed Student's t test; *p<0.05, **p<0.01, ***p<0.001.

DATA AND SOFTWARE AVAILABILITY

Data Resources. Accession Numbers: Three datasets (two RNA-seq and one ChIP-seq experiment) are available from the GEO database: GSE78795. The mass spectrometry proteomics data are available from the ProteomeXchange Consortium/PRIDE repository with the dataset identifier PXD007114.

SUPPLEMENTAL REFERENCES

Abad M, Mosteiro L, Pantoja C, Cañamero M, Rayon T, Ors I, Graña O, Megías D, Domínguez O, Martínez D, et al. (2013). Reprogramming in vivo produces teratomas and iPS cells with totipotency features. Nature *502,* 340-345.

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Andrews, S (2011). FastQC, a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bard JBL, Kaufman MH, Dubreuil C, Brune RM, Burger A, Baldock RA, and Davidson DR. (1998). An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. Mech. Develop. *74*, 111-120.

Bardou P, Mariette J, Escudié F, Djemiel C, and Klopp C. (2014). Jvenn: an interactive Venn diagram viewer. BMC Bioinformatics *15*, 293.

Batth TS, Francavilla C, and Olsen JV. (2014). Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. J Proteome Res. *13*, 6176-6186.

Cox J, and Mann M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotech. *26*, 1367-1372.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, and Mann M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794-1805.

Crespo I, Krishna A, Le Béchec A, and Del Sol A. (2013). Predicting missing expression values in gene regulatory networks using a discrete logic modeling optimization guided by network stable states. Nucleic Acids Res. *41*, e8.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.

D'Amour KA, and Gage FH. (2003). Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells. PNAS *100*, 11866-11872.

Eisenberg E, and Levanon EY. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569-574.

Feng J, Liu T, Qin B, Zhang Y, and Liu XS. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728-1740.

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

Kent WJ, Zweig AS, Barber G, Hinrichs AS, and Karolchik D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

Khan A, and Zhang X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acid Res. *44*, D164-171.

Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, Blasco MA, and Serrano M. (2009a). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature *460*, 1136-1139.

Li H, Collado M, Villasante A, Matheu A, Lynch CJ, Cañamero M, Rizzoti K, Carneiro C, Martínez G, Vidal A et al. (2012). p27(Kip1) directly represses Sox2 during embryonic stem cell differentiation. Cell Stem Cell *11*, 845-852.

Li H, and Durbin R. (2009b). Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics *25*, 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genome Project Data Processing Subgroup. (2009c). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lopez-Guadamillas E, Fernandez-Marcos PJ, Pantoja C, Muñoz-Martin M, Martínez D, Gómez-López G, Campos-Olivas R, Valverde AM, and Serrano M. (2016). $p21^{Cip1}$ plays a critical role in the physiological adaptation to fasting through activation of PPARα. Sci. Rep. *6*, 34542.

Marikawa Y, Tamashiro DA, Fujita TC, and Alarcón VB. (2009). Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis. Genesis *47*, 93-106.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495-501.

Munoz P, Blanco R, Flores JM, and Blasco MA. (2005). XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer. Nat. Genet. *37*, 1063-1071.

Muñoz-Espín D, Cañamero M, Maraver A, Gómez-López G, Contreras J, Murillo-Cuesta S, Rodríguez-Baeza A, Varela-Nieto I, Ruberte J, Collado M, et al. (2013). Programmed cell senescence during mammalian embryonic development. Cell *155*, 1104-1118.

Palmero, I and Serrano, M. (2001). Induction of senescence by oncogenic Ras. Methods Enzymol. *333*, 247-256.

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617-1632.

Quinlan AR. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1-11.12.34.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, and Zhang F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281-2308.

Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, and Mesirov JP. (2006). GenePattern 2.0. Nat. Genet. *38*, 500-501.

Rosenbloom, KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670-681.

Ruiz S, Panopoulos AD, Herrerías A, Bissig KD, Lutz M, Berggren WT, Verma I, and Izpisua Belmonte JC. (2010). A high proliferation rate is required for somatic cell reprogramming and maintenance of human embryonic stem cell identity. Curr. Biol. *21*, 45-52.

Ruiz S, Mayor-Ruiz C, Lafarga V, Murga M, Vega-Sendino M, Ortega S, and Fernández-Capetillo O. (2016). A genomewide CRISPR screen identifies CDC25A as a determinant of sensitivity to ATR inhibitors. Mol. Cell *62*, 307-313.

Salmon-Divon M, Dvinge H, Tammoja K, and Bertone P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics *11*, 415.

Savatier P, Lapillonne H, van Grunsven LA, Rudkin BB, and Samarut J. (1996). Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. Oncogene *12*, 309-322.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545-15550.

Takahashi K, and Yamanaka S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, and Pachter L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562-578.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein M, Geiger T, Mann M, and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13,* 731–740.

Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, et al. (2016). 2016 update of the PRIDE database and related tools. Nucleic Acids Res. *44*, D447-D456.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, and Young RA. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wiśniewski JR, Zougman A, Nagaraj N, and Mann M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359-362.

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, and Smith A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Ye T, Ravens S, Krebs AR, and Tora L. (2014). Interpreting and visualizing ChIP-seq data with the seqMINER software. Methods Mol. Biol. *1150*, 141-152.

Yuan JS, Reed A, Chen F, and Stewart CN. (2006). Statistical analysis of real-time PCR data. BMC Bioinformatics *7*, 85.

Zickenrott S, Angarica VE, Upadhyaya BB, and Del Sol A. (2016). Prediction of disease-gene-drug relationships following a differential network analysis. Cell Death Dis. *7*, e2040.