Cell Systems

Supplemental Information

Systems Analyses Reveal Shared and Diverse Attributes of Oct4 Regulation in Pluripotent Cells

Li Ding, Maciej Paszkowski-Rogacz, Maria Winzi, Debojyoti Chakraborty, Mirko Theis, Sukhdeep Singh, Giovanni Ciotta, Ina Poser, Assen Roguev, Wai Kit Chu, Chunaram Choudhary, Matthias Mann, A. Francis Stewart, Nevan Krogan, and Frank Buchholz

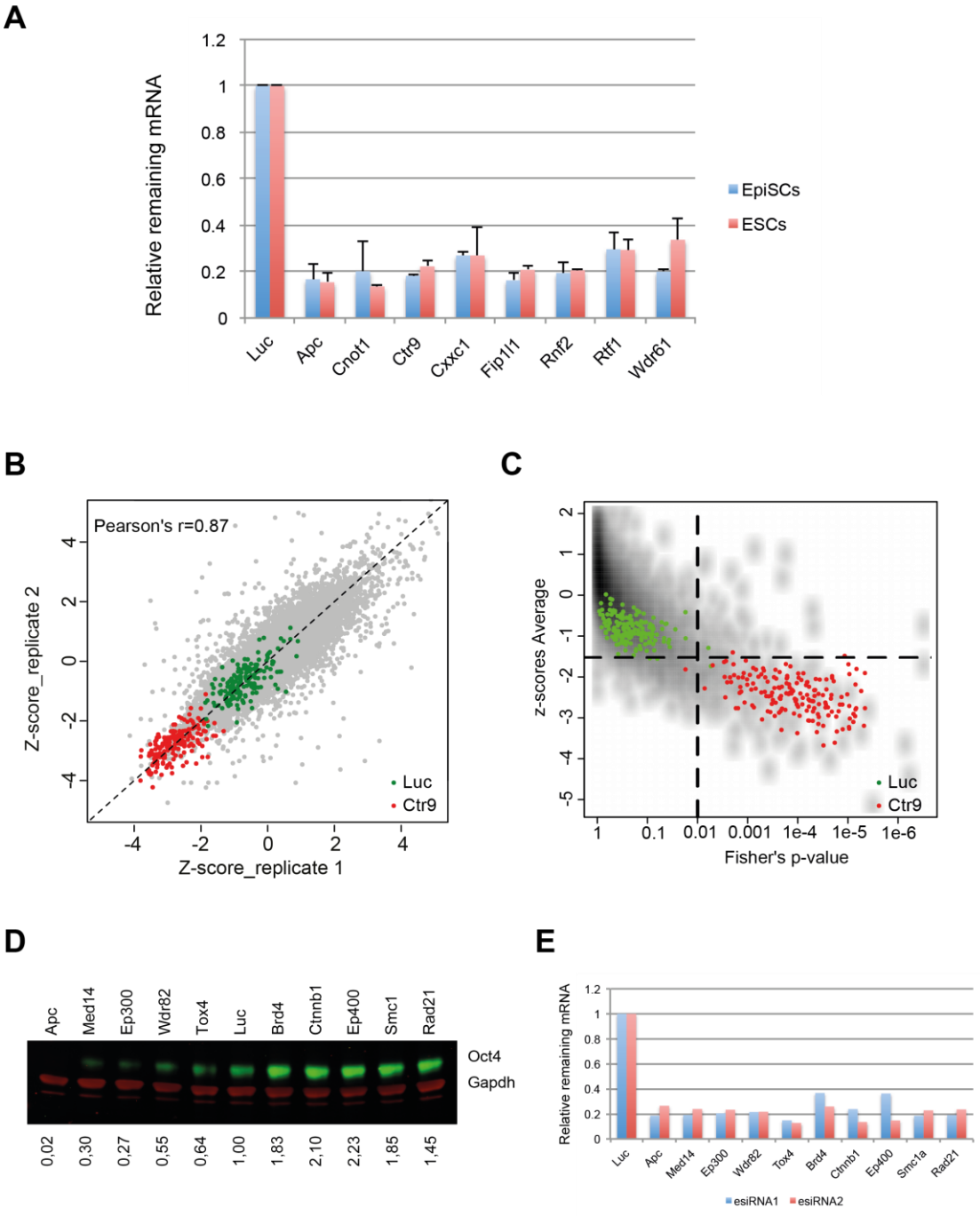


Figure S1. RNAi in EpiSCs. Related to Figure 1. (A) EsiRNA transfection achieves comparable knockdown efficacy in ESCs and EpiSCs. Total RNAs were prepared from ESCs and EpiSCs 24 hours post esiRNA transfection. Knockdown efficiencies for indicated esiRNAs were determined by qRT-PCR,

using primers listed in Table S8. Values are means ± SD from triplicate samples. **(B)** Scatter plot of Z score replicates of the primary screen. Each esiRNA is marked as a grey dot. Negative controls (non-targeting Luc esiRNA) are marked in green. Positive controls (Ctr9 esiRNA) are marked in red. The dotted line marks the diagonal. The Pearson correlation for the replicate experiment is indicated. **(C)** Fisher's combined probability test of the primary screen. Each esiRNA is marked as a grey dot. Negative controls (non-targeting Luc esiRNA) are marked in green. Positive controls (Ctr9 esiRNA) are marked in red. The dotted lines mark the thresholds used to define primary hits. **(D)** Quantitative western bot analysis of Oct4 after knockdown. Oct4 protein levels (green) were quantified and normalized to Gapdh (red) after treatment with indicated esiRNAs. The non-targeting Luc esiRNA was used as control. Numbers represent the Oct4 fold change in comparison to Luc control transfected cells. **(E)** The primary and secondary esiRNA transfection achieves comparable knockdown efficacy in EpiSCs. Total RNAs were prepared 24 hours post esiRNA transfection. Knockdown efficiencies for indicated esiRNAs were determined by qRT-PCR, using primers listed in Table S8. Values are means ± SD from triplicate samples.

Figure S2

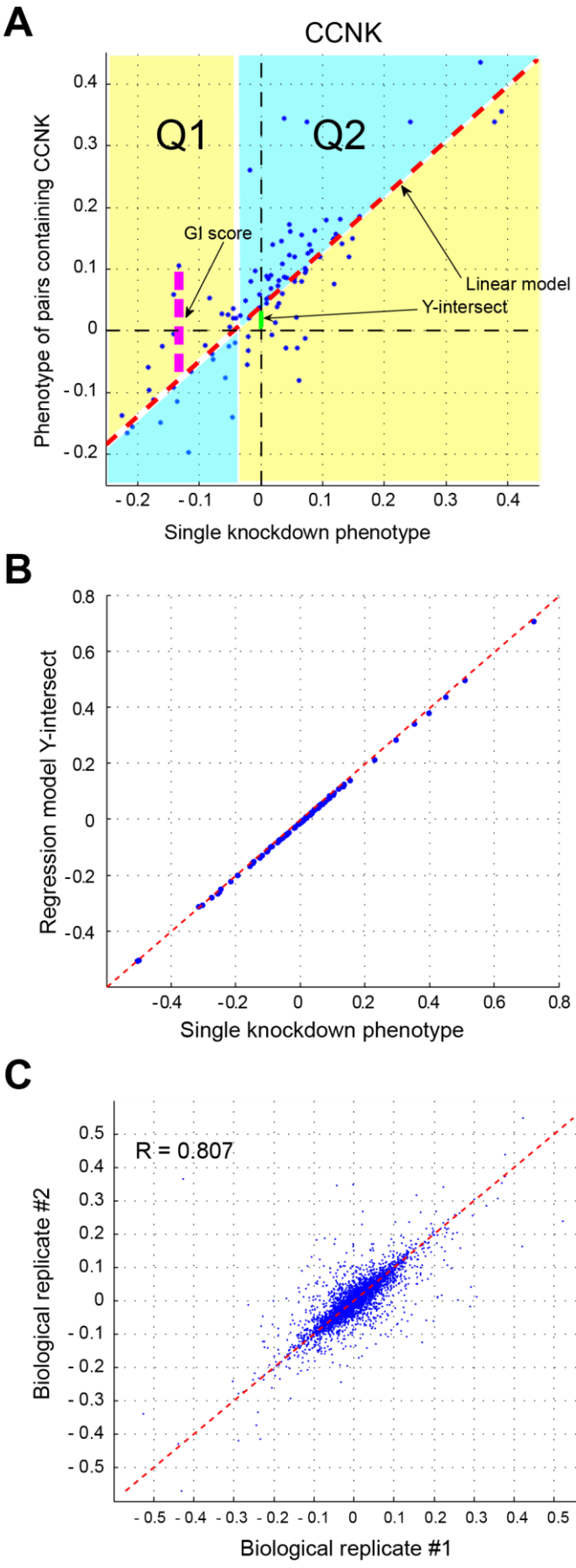
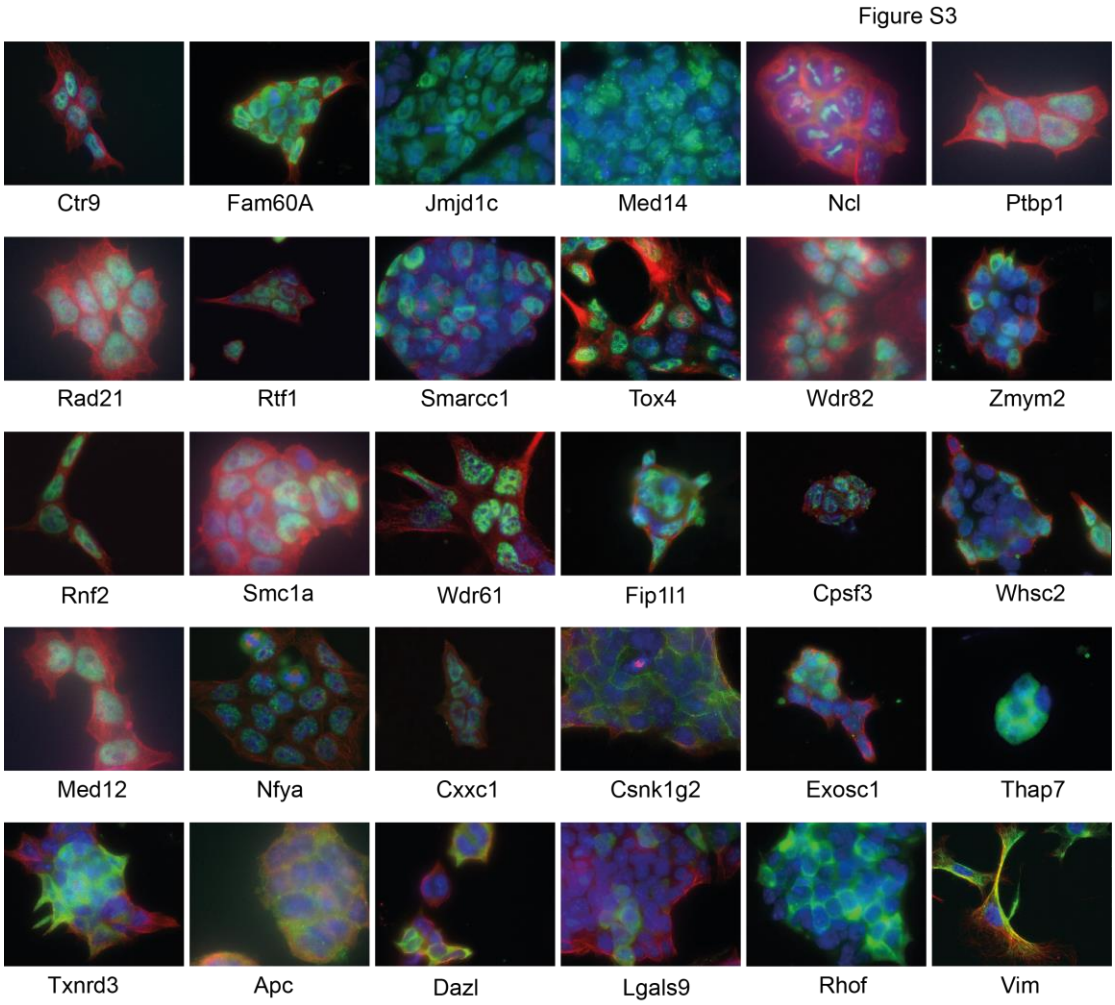

A
CCNK
Q1
Q2
GI score
Linear model
Y-intersect
Phenotype of pairs containing CCNK
Single knockdown phenotype
B
Regression model Y-intersect
Single knockdown phenotype
C
R = 0.807
Biological replicate #2
Biological replicate #1

Figure S2. Gene interaction scores. Related to Figure 3. (A) Graphical representation of the GI score – for each gene (CCNK in this case) a robust linear model of the normalized double knockdown phenotypes as a function of the single mutant phenotypes (blue dots) is computed (dashed red line) and represents the expected phenotype for each esiRNA pair. GI scores are the residuals of the linear model (dashed purple line). The Y-intersect (light green) is a measure of the single knockdown phenotype for the gene of interest. The GI is determined by the spatial domain, where the data falls with respect to the linear model, with negative sign assigned to synergistic GIs and positive sign assigned to suppressive GIs. The point where the regression line crosses the x-axis is important as it determines how interactions are interpreted since expected phenotypes below this point have negative and above it positive values respectively. For example, in the top left yellow quadrant (Q1) the expected phenotype of the double esiRNA knockdown has a negative value e.g. combinations of esiRNAs in this space have a tendency to cause lower GFP expression. Points above the regression line (like the one used for annotation here) have a weaker than expected phenotype (a stronger than expected phenotype here will be even lower GFP expression e.g. a more extreme negative value). Therefore, data in this quadrant represents suppressive interactions. Conversely for a point in the top right blue quadrant (Q2) the expected phenotype has a positive value e.g. double knockdowns in this space have a tendency to cause higher GFP expression. Data points in this space have stronger than expected phenotype and are therefore classified as synergistic GIs. **(B)** The calculated single knockdown phenotypes are strongly correlated with the measurements derived from the

Y-intersects in the linear models. The dotted red line runs across the diagonal of the graph and represents perfect correlation. **(C)** A scatter plot of independent replicate measurements. GI scores are highly reproducible, as the whole combinatorial space has been saturated in this experiment. For each esiRNA pair there are two independent replicate measurements. R - pearson correlation coefficient=0.807.



Protein	Localization
Ctr9	Nucleus
Fam60A	Nucleus
Jmjd1c	Nucleus
Med14	Nucleus
Ncl	Nucleus
Ptbp1	Nucleus
Rad21	Nucleus
Rtf1	Nucleus
Smarcc1	Nucleus
Tox4	Nucleus
Wdr82	Nucleus
Zmym2	Nucleus
Rnf2	Nucleus
Smc1a	Nucleus
Wdr61	Nucleus
Fip1l1	Nucleus
Cpsf3	Nucleus
Whsc2	Nucleus
Med12	Nucleus
Nfya	Nucleus
Cxxc1	Nucleus

Protein	Localization
Csnk1g2	Cytoplasm
Exosc1	Cytoplasm
Thap7	Cytoplasm
Txnrd3	Cytoplasm
Apc	Cytoplasm
Dazl	Cytoplasm
Lgals9	Cytoplasm
Rhof	Cytoplasm

Protein	Localization
Vim	Cytoskeletal

Protein	Localization
Lpar4	Cell membrane
Tyw3	Cell membrane
Camta2	Cell membrane

Protein	Localization
Syce2	Nucleus/Cytoplasm
Cnot2	Nucleus/Cytoplasm
Cnot3	Nucleus/Cytoplasm

Figure S3. Assessment of Localization and Affinity Purification (LAP) - tagged proteins. Related to Figure 4. GFP fusion proteins are shown in

green; α-tubulin is shown in red; and DAPI staining is presented in blue. The classification of the individual proteins is shown in the table to the right.

Figure S4

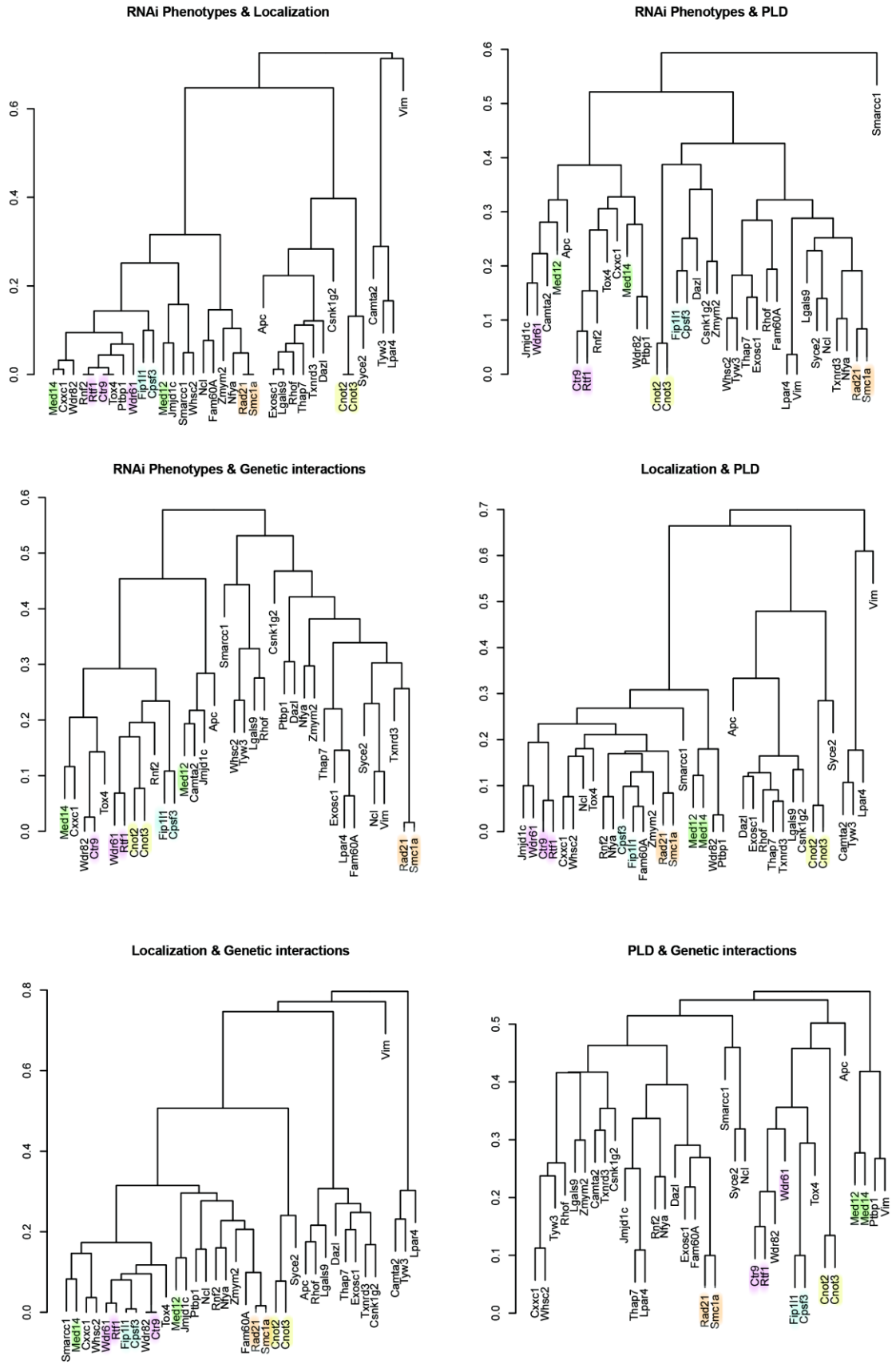

RNAi Phenotypes & Localization
RNAi Phenotypes & PLD
RNAi Phenotypes & Genetic interactions
Localization & PLD
Localization & Genetic interactions
PLD & Genetic interactions

Figure S4. Graphical presentation of hierarchical cluster analysis of indicated Omics datasets in an all-against-all manner. Related to Figure 5. A binary distance metric was used for the localization data and an Euclidean distance metric was employed for all other data sets. Components of known protein complexes are highlighted with the same color.

Supplemental Tables:

Table S1. Related to Figure 1. Primary RNAi screen in EpiSCs (sheet1), primary hits that down-regulated Oct4 expression (sheet2), and primary hits that up-regulated Oct4 expression (sheet3).

Table S2. **Related to Figure 1.** Validation of EpiSC screen primary hits using independent esiRNAs.

Table S3. **Related to Figure 2.** Categorization of validated hits.

Table S4. **Related to Figure 3.** Genetic interaction mapping of Oct4 regulators in EpiSCs (sheet1), raw GFP readout of double knockdown (sheet2), and chart illustration (chart1). Note: raw GFP values of the double knockdowns (one esiRNA with combination of all other esiRNAs) can be illustrated in the chart1 panel by typing gene names in the Cell “I2” or “J2” of sheet2.

Table S5. **Related to Figure 4.** Z-scores of protein level dependencies.

Table S6. Related to Figure 5. Gene ontology analysis of genes regulated upon Tox4 knockdown.

Table S7. Related to Figure 6. Mass spectrometry results of Tox4 and Ctr9 Co-IPs.

Table S8. Related to Figures 1, S1, 4, and 5. Primer sequences used in this study.

Supplemental Experimental Procedures

Cell culture and high throughput esiRNA screen

ESCs (Oct4-Gip and BAC-transgenic ESC lines) were cultured on gelatin-coated plates in Glasgow Minimum Essential Medium (Sigma) supplemented with 10% FBS (Pan biotech), 2.2 mM L-glutamine (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 50 mM 2-mercaptoethanol (Invitrogen), NEAA (Invitrogen), and 5 µg/ml LIF (generated in house) as previously described (Ding *et al*., 2009). ESCs were trypsinized and split every 2 days, and the medium was changed daily.

The Oct4-GFP EpiSC line OE7 was cultured on Fibronectin (Millipore)-coated plates in N2B27 medium, supplemented with 20 ng/ml Activin A (generated in house) and 12 ng/ml Fgf2 (generated in house) as previously described (Guo *et al*., 2009). EpiSCs were fed daily, and split every 2 days. For the genome-scale screen, reverse transfections were performed using mixtures of 50 ng esiRNA and 0.075 µl Lipofectamine 2000 (Invitrogen) in 10 µl OptiMEM medium (Invitrogen). EpiSCs were plated in Fibronectin-coated 384-well plates with a density of 3000 cells per well in 80 µl N2B27/Activin A/Fgf2 medium. On each plate, four negative controls (Renilla Luciferase esiRNA), and four positive controls (Ctr9 esiRNA) were placed to monitor the transfection efficiency and phenotypic readout. GFP fluorescence and cell numbers were measured 72 hours post transfection using a FACS Calibur (BD biosciences) equipped with an HTS loader for high throughput analysis.

esiRNA double knockdowns and Epistasis analysis

Pair-wise esiRNA matrix (25 ng esiRNA(A) and 25 ng esiRNA(B) in 10 µl OptiMEM medium) were pipetted onto Fibronectin coated 384-well plates.

OE7 EpiSCs were reverse transfected using 0.075 µl Lipofectamine 2000 (Invitrogen) in 10 µl OptiMEM medium (Invitrogen) with a density of 3000 cells per well in 80 µl N2B27/Activin A/Fgf2 medium with triplicates for each combination. Oct4 expression was quantified 72 hours post transfection using a FACS Calibur (BD biosciences) equipped with an HTS loader for high throughput analysis.

Data analysis relies on the assumption that strong genetic interactions are rare and therefore most RNAi combinations will display weak or no effect. Raw data was normalized to the median value of the GFP signal and the resulting distribution was centered over zero. Phenotypes emerging from individual RNAi knockdowns were defined for each gene as the median of the normalized GFP signal distribution for all gene pairs containing the gene of interest. For each set of double RNAi combinations containing a particular gene, a linear model describing the relationship between the single RNAi phenotypes and the phenotypes observed upon double RNAi knockdown (e.g. D = a*S + b, where D is the phenotype upon double RNAi, S is the computed single RNAi effect, a is the slope of the fit and b is the y-intercept) was computed using a robust linear fitting algorithm minimizing the effect of the outliers. The y-intercepts of these models represent the phenotypes from single RNAi knockdown for each gene and are in very good agreement with the computed single RNAi phenotypes. Genetic interaction scores for each RNAi pairwise combination are defined as the residuals of that fit.

Protein level dependency

LAP-tagged EpiSC lines were derived from ESC stably expressing the BAC-tagged transgene by *in vitro* differentiation according to the protocols

established by Guo *et al.* (Guo *et al*., 2009). In brief, ESCs were seeded onto 15 μg/ml fibronectin (Roche, Milan, Italy, http://www.roche.it) coated six-well plates at 1 × 10^5 cells per well in ESC medium or in 2i/LIF. After 24 hours, the medium was switched to EpiSC medium. When cells reached 80%–90% confluence, they were passaged at high density (typically at 5 × 10^5 per six-well). The differentiating cultures were collected at passages 4 and 6 for quantification of the expression of the EpiSCs marker Fgf5. Once established, cells were passaged for an additional 1 weeks in the presence of 0.6μM JAK inhibitor I (Calbiochem, 420099) in order to ensure the complete elimination of any residual ESCs. Each LAP-tagged EpiSC line was transfected with the pre-pipetted esiRNA array in triplicates using the same conditions as for OE7 transfection. The expression of GFP tagged proteins were quantified using a FACS Calibur (BD biosciences) equipped with an HTS loader. The expression of each GFP-tagged protein after esiRNA knockdown was calculated as z scores of GFP fluorescence.

Quantitative Western hybridization

1.5x10^5 EpiSCs cells were reverse transfected with 1000 ng esiRNAs and 2 μl lipofectamine 2000 in fibronectin-coated 6-well plates. 72 hours post transfection, EpiSC cells were harvested and lysed in Laemmli sample buffer. 10 μg of protein extracts were separated on NuPAGE 4-12% Bis-tris protein gels (Invitrogen) and blotted to nitrocellulose membrane (Millipore). The membranes were probed with the primary antibodies against Oct4 (Santa Cruz, sc-8628), Ctr9 (Abcam, ab84487), Tox4 (Sigma-Aldrich, HPA017880), and Gapdh (Novus Biologicals, NB300-221), and corresponding secondary antibodies (RDye® 680RD/800CW anti-mouse IgG, anti-rabbit IgG, and anti-

goat IgG). The membranes were scanned on an Odyssey Infrared Imager, and the proteins were quantified using the software Image Studio.

qRT-PCR

$1.5x10^5$ EpiSCs cells were reverse transfected with 1000 ng esiRNAs and 2 µl lipofectamine 2000 in fibronectin-coated 6-well plates. 72 hours post transfection, total RNA was isolated by using the RNeasy Mini kit (Qiagen) and 1 µg RNA was reverse transcribed with SuperScript III Reverse transcriptase (Invitrogen) utilizing an oligo(dT)18 primer. qPCRs were performed with the Sybr green qPCR kit (Abgene) on a C1000 touch Thermal Cycler (Biorad). Measured transcript levels were normalized to Gapdh. Samples were run in triplicates. Primers used are listed in Table S8.

Co-IP and Mass spec

GFP-tagged protein complexes were isolated by immunoaffinity chromatography using a fully automated liquid-handling platform as described (Hubner *et al*., 2010). LC-MS/MS analysis was performed by using a mass spectrometer (Q Exactive; Thermo Scientific). The Q-Exactive was operated using Xcalibur 2.2 in the data-dependent mode to automatically switch between MS and MS/MS acquisition as described (Kelstrup *et al.*, 2012; Michalski *et al*., 2011). Raw data files were processed using the MaxQuant software (v1.2.6.20; http://www.maxquant.org) as described previously (Cox *et al*., 2011). Parent ion (MS) and fragment (MS2) spectra were searched against the UniProt species-specific fasta files from the February 2012 release. Label-free quantification was performed with the Perseus software (http://www.perseus-framework.org) and the MaxQuant-based program QUBICvalidator as described previously (Hubner *et al*., 2010). Proteins with

more than 2 unique peptides in all 3 triplicate samples were further analyzed. Results were then plotted by using the open source statistical software R (https://www.Rproject.org).

Integration of Omics data

Combined clustering analysis of the experimental results obtained from the RNAi screen, localization and genetic interaction studies, as well as the protein level dependency analyses was performed with the R software package, version 3.0.2.

In order to group genes into clusters, we employed agglomerative hierarchical clustering with complete linkage. Dissimilarity values between individual genes were calculated as quasi geometric means of dissimilarity values obtained from individual experiments (Gower, 1971).

For the RNAi data, z-scores obtained from the screen were converted into ranks and dissimilarities between genes were calculated using euclidean distance metric. Obtained values in an all-against-all matrix were scaled to be in a range between 0 and 1. The same approach was employed for the protein level dependency data, where ranks were calculated individually for each BAC cell line. In case of genetic interaction scores, dissimilarities were calculated using Pearson's correlation coefficient, and the final dissimilarity matrix was also scaled to be in a range from 0 to 1. Protein localization data was converted into 0, if two genes share the same localization pattern, and 1 if their localization is different.

Since each obtained dissimilarity matrix had values in a range between 0 and 1, prior to calculating their geometric mean all values were incremented by 1.

After calculating the geometric mean, a value of 1 was subtracted from obtained combined dissimilarity values.

Supplemental References

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

Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., Slabicki, M.M., Heninger, A.-K., de Vries, I., Kittler, R., Junqueira, M., Shevchenko, A., Schulz, H., et al. (2009). A Genome-Scale RNAi Screen for Oct4 Modulators Defines a Role of the Paf1 Complex for Embryonic Stem Cell Identity. Cell Stem Cell *4*, 403–415.

Gower, J.C. (1971). A General Coefficient of Similarity and Some of Its Properties. Biometrics *27*, 857–871.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of Cell Biology *189*, 739–754.

Kelstrup, C.D., Young, C., Lavallee, R., Nielsen, M.L., and Olsen, J.V. (2012). Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. J. Proteome Res. *11*, 3487–3497.

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell Proteomics *10*, M111.011015.