SUPPLEMENTAL METHODS

RNAi screen development

For the siRNA screen, dual reporter ESCs were plated in gelatin-coated 96-well plates at 1000 cells/well in FBS+LIF+2i medium (day -1). The next day (day 0), LIF and 2i were removed to induce differentiation, and ESCs in each well were transfected with 25nM SMARTpool siGENOME siRNAs (pool of 4 distinct siRNAs) to 356 separate genes using Dharmafect 1 transfection reagent (GE/Dharmacon). The siRNA to NF110 was custom designed (sense: GCAGAAAGGCUAUGGCCAUUU, antisense: AUGGCCAUAGCCUUUCGUU [GE/Dharmacon]). On day 3, plates were fixed and stained with Hoechst 33258 (ThermoFisher Scientific) and imaged on the INCell Analyzer 2000 (GE). The area and intensity of miR-302-eGFP (GFP) and miR-290-mCherry (RFP) fluorescence were measured using the INCell Developer Toolbox software (GE). ESCs maintained in FBS+LIF+2i (“l2i”), ESCs allowed to differentiate in the absence of siRNA and transfection reagent (“untreated”), ESCs treated with transfection reagent only (“dfect”), and ESCs treated with control non-targeting siRNA (“siCtrl”) served as differentiation and transfection toxicity controls. ESCs treated with siRNAs against GFP (“siGFP”) and Oct4 (“siOct4”) were used as positive controls for siRNA efficiency (knockdown of Oct4 resulted in cell death). Specifically, each plate contained sets of siRNAs against 32 test genes, as well as 4 replicates of every control condition described above, resulting in a total of 68 conditions evaluated on each plate. Each plate design was repeated in triplicate to ensure consistency in the response of any test gene.

For screen hits, cells were analyzed by flow cytometry on an LSRII flow cytometer (BD Biosciences) to evaluate levels of eGFP and mCherry fluorescence. Cells were also analyzed by qRT-PCR to confirm knockdown of the targeted gene as well as changes in miR-290 and miR-302 expression.

RNAi screen analyses

Two levels of normalization were used to control for sample size during testing and to ensure that values across replicates and plate designs were comparable. (1) A variable of interest (DAPI [Hoechst] area, GFP area, or RFP area) was first normalized across all the replicates of each plate design. This was done by shifting the data by a location estimator (Maronna and Zamar 2002) and then scaling the data by the estimator from the same source (the implementation uses the scaleTau2 function defined in R’s robustbase package). This procedure assumes that the distribution of the variable is the same across replicates. However, this is not guaranteed to be a valid assumption across different plates because we cannot ensure that all plates follow an identical distribution for a given variable. (2) To normalize across plates, we made the assumption that the normalized distribution of all the controls within each plate (not including the test genes) should be identical across plates. Therefore, the distribution of the control siRNAs for each plate was used to calculate a robust measurement of location and scale by which to shift and divide that plate’s data. This approach re-scales the data to describe each observation in terms of robust control deviations away from the control center.

This 2-step normalization was first used to transform the DAPI area values to be comparable across all replicates and plate designs. The mean and standard deviation of DAPI area were then calculated across siOct4 wells to estimate the parameters of the background distribution of cell
death. We then calculated the z-score for all other genes and denote any gene that is above the 95th FDR-adjusted quantile as live cells. In this way, we identified 49 genes whose knockdown induced cell death (Supplementary Table S1, “Viability testing”). To ensure that we test for changes in differentiation upon knockdown, we removed any target genes whose effect might be poorly estimated. We used stricter exclusion criteria and removed any genes that had a large percentage of cell death in at least one of its wells, which we defined as having a DAPI area value below the maximum value in the siOct4 samples after normalizing the DAPI area of siOct4 across all replicates and plate designs. For the analysis, we only tested for changes in the RFP and GFP area in the 146 genes whose 3 replicates have DAPI areas that are all above maximum siOct4 DAPI area levels, indicating that the knockdown of that gene does not induce cell death.

To identify the genes whose knockdown significantly changed GFP or RFP signal (Supplemental Table S1, “Differentiation testing”), we first took the difference of the log fluorescence area and log DAPI area to account for the total amount of cells present. Next, we normalized the log fluorescence data across all replicates and applied the t-test with equal variance for each gene siRNA against the combined siCtrl samples within each plate design. We then controlled for multiple hypothesis testing by calculating the q-value of each test using the qvalue package in bioconductor. To identify the siRNAs that significantly alter fluorescence signal, it is not enough to only use a cutoff on the q-value, since that ignores other controls that could be useful in assessing an RBP’s importance. For example, the “untreated” controls were often marked as being “significantly” different from the siCtrl samples. Therefore, we considered RBPs to be “significant” only if they are below a q-value cutoff (q < 0.001) and have a q-value less than the “untreated” control.

**Generation of NF90+NF110 and NF110 KO ESCs by CRISPR**

CRISPR was performed as previously described (Ran et al. 2013). In brief, guide RNAs (sequences obtained from GeCKO [http://genome-engineering.org/gecko/] or designed using crispr.mit.edu) were inserted into the PX458 construct or a modified PX458 in which eGFP was replaced with BFP. The oligos used to create the NF90+NF110 KO allele target exon 3 of NF90/NF110 (CACC-GAATGATGATCGCCACGTGA, AAAC-TCACGTGGCGATCATCATT). The oligos used to create the NF110 KO allele target the intron preceding the exons private to NF110 (CACC-G-TGCTCTGAGCAGTCGGCTGC, AAAC-GCAGCCGACTGCTCAGAGCA-C) and the 3’UTR of NF110 (CACC-G-TGCCCTGGTGTGAGTTCCATC, AAAC-GATGGAACTCACACCAGGCA-C); the intron-targeting guide was cloned into PX458 and the 3’UTR-targeting guide was cloned into PX458-BFP. The constructs (3-10ug) were introduced into V6.5 mouse ESCs (3 million) using the Amaxa Mouse ES Cell Nucleofector Kit (Lonza). GFP+ (for NF90+NF110 KO) or GFP+/BFP+ (for NF110 KO) ESCs were purified by fluorescence-activated cell sorting (FACS) on a FACSAria II UV machine (BD Biosciences) 2 days after nucleofection, and individual colonies were picked for analysis 6 days after sorting. ESCs that had successfully undergone genetic modifications of interest were identified by PCR. For NF90+NF110 KO ESCs, a PCR product different from 387bp when using the primer pair CGTAGAGTGCCCCATTCCTG and CAGCCCCTTTCTTGGCTCTCA indicates a mutation in NF90/NF110; for NF110 KO ESCs, the absence of a product when using the primer pair TTGGCTCAGGCCCTATAATCAC and CCTGTGCACCTTAAAGCCCT, as well as the presence of an product smaller than 2447bp when using the primer pair CATCTGCAAAGCTGCTGTCG and
AGCGGAATTCAAATGTACTGTCT indicate successful removal of the exons private to NF110. PCR products were TOPO-cloned (ThermoFisher Scientific) and sequenced to determine the nature of the mutations. NF90/NF110 protein expression status was confirmed by Western blot.

**Colony formation/clonogenicity assay**
Cells were plated in 96-well plates at 500 cells/well and stained with Vector Red alkaline phosphatase (AP) 4 days later (Vector Laboratories). The ratio of AP+ colonies formed to the number of cells plated was defined as the colony formation potential.

**Population doubling time assay**
Cells were plated in 6-well plates at 100,000 cells/well and counted 24, 48, and 72hrs later. Population doubling time was calculated as previously described (Wang et al. 2007) using the formula, \( Y_{end} = Y_{start} \times 2^{(t/T)} \), where \( T \) is the cell population doubling time, \( Y_{start} \) is the starting number of ESCs (e.g., number of ESCs counted at 24hrs), and \( Y_{end} \) is the ending number of ESCs (e.g., number of ESCs counted at 72hrs) after growth for a period of time \( t \) (e.g., 48hrs).

**Embryoid body (EB) generation**
Embryoid bodies were generated using Aggrewell 800 plates (STEMCELL Technologies) according to the manufacturer’s protocol. In brief, 1.5 million ESCs were seeded per Aggrewell (~5000 cells per EB) in differentiation medium (DMEM, 15% FBS, non-essential amino acids, L-glutamine, penicillin/streptomycin, 55uM beta-mercaptoethanol). After 3 days, EBs were transferred to ultra-low attachment plates (Corning) and harvested every 3 days for qRT-PCR analysis.

**qRT-PCR primer sequences**

| Primer               | Sequence                           |
|----------------------|------------------------------------|
| NF45 qPCR F          | AAGCCAGCACCTGTGATGAC               |
| NF45 qPCR R          | TTCTGTGGGCAAATCAATTCAAG            |
| NF90+NF110 qPCR F    | TGTGGAGGTAGACGCGAAGTA              |
| NF90+NF110 qPCR R    | CTGGAGTCTCTGCGGCCCAGC             |
| NF110 qPCR F         | GGGCTCTGACTACAGCTACG              |
| NF110 qPCR R         | CCTCCATGTGAGCCTGTGTT              |
| NF90 qPCR F          | ACATGAATGCTGGTGCTGGA              |
| NF90 qPCR R          | CGCTCTAGGAAGCCCAAAAA              |
| Rpl7-qPCR F          | GAACCAAGCTGGCCTTTGTCACTC          |
| Rpl7-qPCR R          | CAATGTATGGGCTCCAATCCGCA           |
| T/Brachyury qPCR F   | CTGGGAGCTCTGCTTCTTCA              |
| T/Brachyury qPCR R   | GAGGACGTGGCAGCTGAGA               |
| Gata6 qPCR F         | TGACTCTCTACTCCTCTTCTTC            |
| Gata6 qPCR R         | TACTTGGAGCTGACTGCTTC              |
| Fgf5 qPCR F          | CTTTGCACGCCAGGAGCTTA              |
| Fgf5 qPCR R          | CGGTCGTGGTACCTTGTGAGG             |
| Klf4 qPCR F          | TGTGGAAACACCTATACCAAGAG           |
| Klf4 qPCR R          | CACAGCGTCCCAGTCAC                 |
| Rex1 qPCR F          | GAAAGTGGAGATTAGCCCCGAG            |
Identifying binding targets from RIP-seq data

NF90/NF110 RIP-seq was performed in NF110 KO, NF90+NF110 KO, and NF90+NF110 WT ESCs. Each experiment contains paired input (total RNA) and IP samples, and experiments were repeated in biological triplicate. We consider a theoretical RIP-seq experiment as consisting of 4 conditions: WT input, WT IP, KO input, and KO IP. In this theoretical RIP-seq experiment, genes enriched in KO IP compared to KO input contain nonspecific targets of the antibody and experimental noise, while genes enriched in WT IP compared to WT input contain true binding targets of the protein of interest as well as nonspecific targets of the antibody and experimental noise. Thus, searching for transcripts enriched in the WT IP/WT input comparison but not in the KO IP/KO input comparison yields candidate true binding targets. Our NF90/NF110 RIP-seq can be regarded as three realizations of the theoretical experiment by replacing the 4 conditions in the theoretical experiment with the conditions of the real experiment, as summarized in the following table (also see Fig. 6B).

| Experiment | Conditions | Results |
|------------|------------|---------|
| Theoretical experiment | WT input | WT IP | KO input | KO IP | True binding targets of protein of interest |
| 1 | WT input | WT IP | NF110 KO input | NF110 KO IP | NF110 only targets |
| 2 | WT input | WT IP | NF90+NF110 KO input | NF90+NF110 KO IP | NF90+NF110 targets |
| 3 | NF110 KO input | NF110 KO IP | NF90+NF110 KO input | NF90+NF110 KO IP | NF90 targets (including targets redundantly bound by NF110) |

We discuss in the following sections how to identify binding targets from a theoretical experiment. The methods can then be implemented separately in the three realizations of the
experimental setup. We developed two complementary approaches for identifying binding targets from a theoretical experiment: Poisson ratio test and log ratio test. The Poisson ratio test utilizes the stochastic nature of count data, treating input and IP samples as independent samples. In contrast, the log ratio test compares input and IP samples in pairs, but does not make full use of the distributive properties of count data. Thus, the two methods provide insights from different angles. Combining the two methods in calling binding targets is expected to yield optimal results. In both methods, we restricted to the 14705 genes with total library-size normalized read count (calculated by DESeq2) summed over all RIP-seq samples ≥ 50 (Supplemental Table S5).

Poisson ratio test. We denote the four conditions, WT input, WT IP, KO input, and KO IP, in the theoretical experiment as 0, 0’, 1, 1’, respectively. We further denote the library-size normalized read count (calculated by DESeq2 and rounded to the closest integer) of gene $i$ from sample $j$ as $K_{ij}$. The library-size normalized read count was modeled as a Poisson distribution:

$$K_{ij} \sim \text{Po}(\mu_{i,\rho(j)})$$

where $\mu_{i,\rho(j)}$ is the Poisson mean, and $\rho(j) = 0, 0', 1, 1'$ denotes the condition of sample $j$. We are interested in testing whether the following null hypothesis is true:

$$H_0: \frac{\mu_{i,0'}}{\mu_{i,0}} = \frac{\mu_{i,1'}}{\mu_{i,1}}$$

Rejecting the null hypothesis means that gene $i$ is significantly differentially enriched in WT and KO when comparing IP to input. As test statistics, we define

$$K_{i,\rho} = \sum_{j: \rho(j) = \rho} K_{ij}.$$  

We further define the joint probability of observing $K_{i,0}, K_{i,0'}, K_{i,1}, K_{i,1'}$ being equal to $a, b, c, d$, respectively, as $P_l(a, b, c, d)$. In the spirit of DESeq (Anders and Huber 2010), we define the $p$-value for the proposed null hypothesis as:

$$p_l = \frac{\sum_{a+b+c+d = k_{i,0}+k_{i,0'}+k_{i,1}+k_{i,1'}} P_l(a, b, c, d)}{\sum_{a+b+c+d = k_{i,0}+k_{i,0'}+k_{i,1}+k_{i,1'}} P_l(a, b, c, d)},$$

where $k_{i,\rho}$ are the observed values of $K_{i,\rho}$. We next show that given any $k_{i,\rho}$, the $p$-value defined above can be efficiently estimated by a sequential importance sampling scheme. Under the null hypothesis, the probability $P_l(a, b, c, d)$ can be factorized as $P_l(a, b, c, d) = P_{l,0}(a)P_{l,0'}(b)P_{l,1}(c)P_{l,1'}(d)$, where $P_{l,\rho}(x)$ is the probability of $K_{i,\rho} = x$ according to a Poisson distribution with mean $\lambda_{i,\rho}$ constrained by the null hypothesis. The mean values $\lambda_{i,\rho}$ can be estimated as follows. We first take the maximal likelihood estimates of $\mu_{i,\rho}$:

$$\hat{\mu}_{i,\rho} = \frac{1}{m} \sum_{j: \rho(j) = \rho} k_{ij},$$

where $m$ is the number of samples under any specific condition $\rho$ (here, $m$ is assumed to be the same for all conditions; in our data $m = 3$), and $k_{ij}$ is the observed value of $K_{ij}$. $\lambda_{i,\rho}$ can then be estimated under the null hypothesis as follows:
In order to combine the results from Poisson ratio test and log ratio test, we first obtained a pseudo-log ratios, we simply perform Welch’s t-test between the log ratios of WT and KO to obtain the p-value for each gene i, given the estimated \( \hat{\lambda}_{l,p} \) as calculated above:

for \( l \) from 1 to \( N \):

1) Draw \( a \sim Unif(0, n_i) \), where \( Unif(0, n_i) \) denotes a uniform distribution over integers from 0 to \( n_i \) and \( n_i = k_{i,0} + k_{i,0'} + k_{i,1} + k_{i,1'} \) is the total read count of gene \( i \) in all samples.

2) Draw \( b \sim Unif(0, n_i - a) \).

3) Draw \( c \sim Unif(0, n_i - a - b) \).

4) Calculate \( d = n_i - a - b - c \).

5) Calculate the importance weight:

\[
w_l = \frac{P\left(\frac{k_{i,0}}{k_{i,0}'}(\frac{1}{n_i+1})(\frac{1}{n_i-a+1})(\frac{1}{n_i-a-b+1})\right)}{P\left(\frac{k_{i,0}}{k_{i,0}'}(\frac{1}{n_i+1})(\frac{1}{n_i-a+1})(\frac{1}{n_i-a-b+1})\right)}
\]

where \( P\left(\frac{k_{i,0}}{k_{i,0}'}(\frac{1}{n_i+1})(\frac{1}{n_i-a+1})(\frac{1}{n_i-a-b+1})\right) \) is the Poisson probability mass function with mean \( \hat{\lambda}_{l,p} \).

6) Calculate the indicator function:

\[
I_l = 1\left(P\left(\frac{k_{i,0}}{k_{i,0}'}(\frac{1}{n_i+1})(\frac{1}{n_i-a+1})(\frac{1}{n_i-a-b+1})\right)\leq P\left(\frac{k_{i,0}}{k_{i,0}'}(\frac{1}{n_i+1})(\frac{1}{n_i-a+1})(\frac{1}{n_i-a-b+1})\right)\right)
\]

The estimated p-value is given by:

\[
p_l = \frac{\sum_{l=1}^{N} w_l I_l}{\sum_{l=1}^{N} w_l}
\]

We generated \( N = 10^6 \) samples for each gene to estimate the p-value (Supplemental Table S5).

Log ratio test. We follow the same notations as in Poisson ratio test. In the log ratio test, we consider the log ratios of library-size normalized read counts between IP and input: \( \xi_j = log_2\left(\frac{k_{i,j}}{k_{i,j}}\right) \) for WT, where \( \rho(j) = 0 \) and \( j' \) is the paired IP sample of \( j \); and \( \xi_j = log_2\left(\frac{k_{i,j}}{k_{i,j}}\right) \) for KO, where \( \rho(j) = 1 \) and \( j' \) is the paired IP sample of \( j \). To test the difference between these two log ratios, we simply perform Welch’s t-test between the log ratios of WT and KO to obtain the p-value (Supplementary Table S5). To avoid zeros in the denominator of the log ratios, we used a pseudo-count of 1 in both the numerator and the denominator.

In order to combine the results from Poisson ratio test and log ratio test, we first obtained candidate targets by imposing the following criteria. For Poisson ratio test, we require:

1) Library-size normalized read count ratio WT IP/WT input > KO IP/KO input: \( \hat{\mu}_{i,0}/\hat{\mu}_{i,0} > \hat{\mu}_{i,1'}/\hat{\mu}_{i,1} \).
2) Library-size normalized read count in WT IP > that in WT input: \( \hat{\mu}_{i,0'} > \hat{\mu}_{i,0} \).
3) The RNA abundance is called significantly different by DESeq2 between WT IP and WT input (adjusted \( p \)-value \( \leq 0.05 \)), with log₂ fold change (estimated by DESeq2) log₂(WT IP/WT input) > 0.

4) Mean FPKM in WT input \( \geq 10 \).

For log ratio test, we require:

1) Mean log ratio in WT > that in KO: \( \frac{1}{m} \sum_{j; \rho(j)=0} \xi_j > \frac{1}{m} \sum_{j; \rho(j)=1} \xi_j \).

2) Mean log ratio in WT > 0: \( \frac{1}{m} \sum_{j; \rho(j)=0} \xi_j > 0 \).

3) The RNA abundance is called significantly different by DESeq2 between WT IP and WT input (adjusted \( p \)-value \( \leq 0.05 \)), with log₂ fold change (estimated by DESeq2) log₂(WT IP/WT input) > 0.

4) Mean FPKM in WT input \( \geq 10 \).

A gene satisfying either the criteria for Poisson ratio test or those for log ratio test is included in the candidate target set. We then employed a multi-objective optimization algorithm based on the concept of Pareto dominance (Algorithm 1 in (Diaz et al., 2015)) to provide a combined rank for the genes in the candidate target set, where -log₁₀ \( p \)-values from the Poisson ratio test and the log ratio test were used for the ranking. The final set of binding targets was identified by including the top \( n \) genes such that the \( n + 1 \) ranked gene is the first gene that has at least one of the two \( p \)-values > 0.05.

In Supplemental Figure S6A, -log₁₀ \( p \)-value of the Poisson ratio test was plotted on the y-axis, and log₂(\( \frac{\hat{\mu}_{L, j} / \hat{\mu}_{L, 0}}{\hat{\mu}_{L, s} / \hat{\mu}_{L, 1}} \)) was plotted on the x-axis. Identified targets are indicated in red. In Supplemental Figure S6B, -log₁₀ \( p \)-value of the log ratio test was plotted on the y-axis, and log₂(\( \frac{\sum_{j; \rho(j)=0} \xi_j}{\sum_{j; \rho(j)=1} \xi_j} \)) was plotted on the x-axis. Identified targets are also indicated in red. In Supplemental Figure S6C, the candidate target set was plotted, with -log₁₀ \( p \)-values of the Poisson ratio test and log ratio test on the x- and y-axes, respectively. Again, the final identified targets are indicated in red.

**Combinatorial gene set expression analysis**

In this section, we strive to infer the first order approximation to the functional interactions between NF45, NF90, and NF110 using the RNA-seq data in our genetic knockout models. In particular, various modes may exist in the regulatory network of these three proteins. Specifically, NF45 and NF90/NF110 can function both cooperatively as a complex and independently by themselves. Furthermore, when functioning independently of NF45, NF90 and NF110 may regulate the same genes redundantly or distinct sets of genes. The same possibilities also exist for NF45-NF90 and NF45-NF110 complexes. Employing a combinatorial gene set expression analysis framework, we were able to rule out certain regulatory modes inconsistent with our transcriptomic data, while confirming the existence of other possible modes. The results provide novel insights about the gene regulatory network associated with NF45 and NF90/NF110 in ESCs (Fig. 7B).

In order to carry out the combinatorial gene set expression analysis, we adopt the following notations.
Known gene sets:
The following gene sets were identified from transcriptomic data (Materials and Methods, Differential expression).
S\(_1\): 554 differentially expressed (DE) genes in NF110 KO vs. NF90/NF110 WT;
S\(_2\): 655 DE genes in NF90+NF110 KO vs. NF90/NF110 WT;
S\(_3\): 140 DE genes in NF90+NF110 KO vs. NF110 KO;
S\(_4\): 675 DE genes in NF45 KO vs. NF45 WT.

Hypothesized gene sets:
The following genes sets are hypothesized based on all possible regulatory modes associated with NF45 and NF90/NF110.
A: Genes regulated only by NF45, independently of NF90/NF110;
B: Genes regulated only by NF90, independently of NF45 and NF110;
C: Genes regulated redundantly by either NF90 or NF110, independently of NF45;
D: Genes regulated only by NF110, independently of NF45 and NF90;
E: Genes regulated only by NF45-NF90 complex;
F: Genes regulated redundantly by either NF45-NF90 or NF45-NF110 complex;
G: Genes regulated only by NF45-NF110 complex.

Vector of log\(_2\) fold changes in FPKM:
The following vectors are of dimension equal to the total number of genes, where each entry is defined as the log\(_2\) fold change in the mean FPKM value of one gene between two corresponding conditions (numerator vs. denominator).
x\(_1\): NF110 KO vs. NF90/NF110 WT;
x\(_2\): NF90+NF110 KO vs. NF90/NF110 WT;
x\(_3\): NF90+NF110 KO vs. NF110 KO;
x\(_4\): NF45 KO vs. NF45 WT.

Other notations:
0 denotes empty gene set;
A+B denotes the union of gene set A and B;
A-B denotes the gene set consisting of genes in set A but not in set B;
A\(\cap\)B denotes the intersection of gene set A and B;
r(x\(_1\), x\(_2\)|S\(_1\)) denotes the Spearman correlation coefficient between vectors x\(_1\) and x\(_2\), restricted to gene set S\(_1\).

We assert that the following two claims are true.

Claim 1: D\(\neq\)0.
Proof: By exhausting all possibilities, we can identify S\(_4\)=A+E+F+G(\(+D\)), where (\(+D\)) indicates the indirect effect of reduced NF110 upon NF45 KO. Note that only those genes whose regulation depends on NF110 and not redundantly on NF90 are affected in NF110 KO. Hence, S\(_1\)=G+D, so that D=0 implies that S\(_1\)=G. Thus, if D were empty, then NF110 KO and NF45 KO should have the same effect on the gene set S\(_1\), i.e., r(x\(_1\), x\(_4\)|S\(_1\)) should be close to 1. However, we find that r(x\(_1\), x\(_4\)|S\(_1\)) is only about 0.15 (Supplemental Fig. S8A). We therefore conclude that
D \neq 0. When D \neq 0, S_1 = D + G; and NF45 KO affects only G, while NF110 KO affects both D and G, which explains the small \( r(x_1, x_4|S_1) \).

**Claim 2:** G is small compared to D.

**Proof:** As before, we have \( S_1 = D + G \). Note that NF45 KO and NF110 KO have the same effect on the set G. Thus, if G were not sufficiently small, then the correlation \( r(x_1, x_4|S_1) \) would be dominated by the contribution from G and be large, contradicting the observed small value of 0.15 (Supplemental Fig. S8A). We therefore conclude that to first order approximation, \( G = 0 \).

The small but statistically significant correlation \( r(x_1, x_4|S_1) \) reflects the indirect effect of reduced NF110 upon NF45 KO.

In order to further validate Claim 1 and 2, we can check the following corollary against data.

**Corollary 1:**

1. \( r(x_2, x_3|S_3) \) should be close to 1;
2. \( r(x_2, x_3|S_2) < r(x_2, x_3|S_3) \);
3. \( r(x_2, x_3|S_2 - S_1) \) should be close to 1 and \( r(x_2, x_3|S_2 - S_1) > r(x_2, x_3|S_2) \);
4. \( r(x_1, x_2|S_1) \) should be close to 1;
5. \( r(x_1, x_4|S_1) \) should be statistically significant.

**Proof:** According to Claim 1 and 2, D \neq 0 and G = 0. We can thus identify \( S_1 = D \), \( S_2 = B + C + D + E + F \), \( S_3 = B + C + E + F \), and \( S_4 = A + E + F + (D) \), where \((D)\) still indicates the indirect effect of reduced NF110 upon NF45 KO. The corollaries above can be easily proved by these identities. Take (1) for example. Restricting to \( S_3 \), NF90+NF110 KO vs. NF90/NF110 WT (\( x_2 \)) should be similar to NF90+NF110 KO vs. NF110 KO (\( x_3 \)).

Corollary 1 is empirically checked in Supplemental Figure S8B, in which each panel (i-v) supports the respective claim (1-5) in Corollary 1.

We proceed by proving the following claim.

**Claim 3:** B+C \neq 0.

**Proof:** If B+C=0, we can identify \( S_1 = D \), \( S_2 = D + E + F \), \( S_3 = E + F \), and \( S_4 = A + E + F + (D) \); as a result,  
1. \( r(x_3, x_4|S_3) \) would be close to 1; 
2. \( r(x_2, x_4|S_2 - S_1) \) would be close to 1. 

However, \( r(x_3, x_4|S_3) \) is only 0.58 (Supplemental Fig. S8C panel (i)), and \( r(x_2, x_4|S_2 - S_1) \) is only 0.56 (Supplemental Fig. S8C panel (ii)). We therefore conclude that B+C \neq 0.

At this point, we have shown that the following relations hold: \( S_1 = D \), \( S_2 = B + C + D + E + F \), \( S_3 = B + C + E + F \), and \( S_4 = A + E + F + (D) \). Notice that our experimental design does not allow us to distinguish B from C, or E from F, since they always appear together in these equations. However, we are able to further simplify the regulatory modes by employing the following assumptions based on the structure of NF45 and NF90/NF110. First, since all nucleic acid binding motifs present in NF90 are also present in NF110 (Guan et al. 2008), it is reasonable to assume that NF90 does not regulate any additional genes on top of those also redundantly regulated by NF110, i.e., B=0. Second, since the interaction site (DZF motif) between NF45 and NF90/NF110 is common to NF90 and NF110 (Guan et al. 2008; Wolkowicz et al. 2012), it is
also reasonable to assume that NF45-NF90 complex does not regulate any additional genes besides those also redundantly regulated by NF45-NF110 complex, i.e., $E=0$. Therefore, our final model of the regulation network consists of only 4 dominant regulatory modes: A, C, D, and F (Fig. 7B), suggesting the following relations: $S_1=D$, $S_2=C+D+F$, $S_3=C+F$, and $S_4=A+F(+D)$.

Given the relations $S_1=D$, $S_2=C+D+F$, $S_3=C+F$, and $S_4=A+F(+D)$, we can infer A, C, D and F (Supplemental Table S7) from $S_1$, $S_2$, $S_3$, and $S_4$ as follows:

1. $D=S_1$ (554 genes);
2. $F=S_1 \cap S_4$ or $F=S_2 \cap (S_4-D)$. We took the union of genes obtained by the two formulas and got 76 genes for $F$.
3. $C=S_2-D-F$ or $C=S_3-F$. We took the union of genes obtained by the two formulas and got 473 genes for C.
4. $A=S_4-D-F$ (588 genes).

We have empirically checked the following corollary against our data for a final validation.

Corollary 2:

1. $r(x_3, x_4|F)$ should be close to 1, and $r(x_3, x_4|F) > r(x_3, x_4|S_3)$.
2. $r(x_2, x_4|F)$ should be close to 1, and $r(x_2, x_4|F) > r(x_2, x_4|S_2-S_1)$.
3. $r(x_2, x_3|C)$ should be close to 1.
4. $r(x_2, x_3|F)$ should be close to 1.
5. $r(x_2, x_4|D)$ should be statistically significant.

Proof: The corollaries above follow immediately from the relations $S_1=D$, $S_2=C+D+F$, $S_3=C+F$, and $S_4=A+F(+D)$.

Corollary 2 is checked against the data as follows. (1) can be checked by Supplemental Figure S8D panel (i) and comparing with Supplemental Figure S8C panel (i); (2) can be checked by Supplemental Figure S8D panel (ii) and by comparing with Supplemental Figure S8C panel (ii). (3) to (5) can be checked by Supplemental Figure S8D panels (iii) to (v).

In summary, our data imply that (1) there exists a set of genes that are regulated only by NF110, independently of both NF45 and NF90 (Claim 1); (2) genes regulated by NF45-NF110 complex can also be redundantly regulated by NF45-NF90 complex, and vice versa (Claim 2 and the assumption that $E=0$); (3) there exists a set of genes that are regulated by either NF90 or NF110, independent of NF45 (Claim 3 and the assumption that $B=0$). We can therefore narrow down the regulatory network dictated by NF45 and NF90/NF110 to four distinct regulatory modes (Fig. 7B): modes A, C, D, and F. NF45, NF90, and NF110 may regulate ESC proliferation, pluripotency, and differentiation through these four distinct modes (Supplemental Fig. S8E). The genes in each mode were also identified in our analysis (Supplemental Table S7), and may serve as a reference for further studies.
SUPPLEMENTAL REFERENCES

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

Diaz AA, Qin H, Ramalho-Santos M, Song JS. 2015. HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res.* **43**: e16-e16.

Guan D, Altan-Bonnet N, Parrott AM, Arrigo CJ, Li Q, Khaleduzzaman M, Li H, Lee CG, Pe’ery T, Mathews MB. 2008. Nuclear factor 45 (NF45) is a regulatory subunit of complexes of NF90/NF110 involved in mitotic control. *Mol. Cell. Biol.* **28**: 4629-4641.

Kuwano Y, Pullmann R Jr., Marasa BS, Abdelmohsen K, Lee EK, Yang X, Martindale, J.L., Zhan, M., and Gorospe, M. 2010. NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res.* **38**: 225-238.

Maronna RA, Zamar RH. 2002. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* **44**: 307-317.

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

Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat. Genet.* **39**: 380-385.

Wolkowicz UM, Cook AG. 2012. NF45 dimerizes with NF90, Zfr and SPNR via a conserved domain that has a nucleotidyltransferase fold. *Nucleic Acids Res.* **40**: 9356-9368.
SUPPLEMENTAL FIGURES

Supplemental Figure S1 (to go with Figure 1)
Expression analysis by qRT-PCR of (A) NF45, (B) NF90 and NF110, and (C) the miRNAs of the miR-290 family in ESCs or ESCs treated with the indicated siRNAs and differentiated for 3 days. Error bars represent SD of 3 biological replicates. Numbers above the graph indicate the p-values of the comparisons marked.

Supplemental Figure S2 (to go with Figure 2)
Transcript levels by qRT-PCR of (A, E) Klf4, (B, F) Rex1, (C, G) Fgf5, and (D, H) T/Brachyury in NF45 WT, NF45 KO, NF90+NF110 WT, NF90+NF110 KO, and NF110 KO ESCs. Error bars represent SD of 3-4 biological replicates. Numbers above graphs indicate the p-values of the comparisons marked.

Supplemental Figure S3 (to go with Figure 3)
Transcript levels by qRT-PCR of (A, D) Klf4, (B, E) Fgf5, and (C, F) miR-302 family miRNAs in NF45 WT, NF45 KO, NF90+NF110 WT, NF90+NF110 KO, and NF110 KO ESCs and 2D differentiated EpiCs. Error bars represent SD of 3-4 biological replicates. Numbers above graphs indicate the p-values of the comparisons marked.

Supplemental Figure S4 (to go with Figure 4)
(A) Subcellular localization of NF45, NF90, and NF110 proteins in WT ESCs and EpiCs (3 days of -LIF, -2i diff). Nanog was used as a nuclear marker as well as an indicator for differentiation progression. (B) Subcellular localization of NF45, NF90, and NF110 in NF45 WT and NF45 KO ESCs. Tubulin and Oct4B were used as cytoplasmic markers; Oct4A was used as a nuclear marker. For A and B, equal amounts of protein were loaded for each lane. (C, D) Transcript levels by qRT-PCR of NF45, NF90, and NF110 in NF45 WT, NF45 KO, NF90+NF110 WT, NF90+NF110 KO, and NF110 KO ESCs. Error bars represent SD of 3-4 biological replicates.

Supplementary Figure S5 (to go with Figure 5)
FPKM values by RNA-seq of (A, E) Klf4, (B, F) Rex1, (C, G) Fgf5, and (D, H) T/Brachyury in NF45 WT, NF45 KO, NF90/NF110 WT, NF90+NF110 KO, and NF110 KO ESCs. Error bars represent standard deviation. Compare with Supplemental Fig. S2 A-H. (I) GO analysis of Module II and VIII genes as defined in Figure 5B. (J) and (K) Correlation analysis of NF45 KO/NF45 WT vs. NF110 KO/NF90+NF110 WT or NF90+NF110 KO/NF90+NF110 WT transcriptomes with respect to (J) molecular pathways or (K) GO terms (Materials and Methods).

Supplementary Figure S6 (to go with Figure 6)
(A) Poisson ratio test identifies binding targets from RIP-seq data. The three volcano plots represent the three realizations of the theoretical experiment for identifying NF110 only, NF90, and NF90+NF110 targets, respectively (Supplemental Methods). In each plot, \(-\log_{10} p\)-value of the Poisson ratio test was plotted on the y-axis, and \(\log_{2}(\frac{\tilde{\pi}_{l,0}}{\tilde{\pi}_{l,t}}/\tilde{\pi}_{l,t})\) was plotted on the x-axis (Supplemental Methods). Identified targets are indicated in red. (B) Log ratio test identifies binding targets from RIP-seq data. The three volcano plots represent the three realizations of the theoretical experiment for identifying NF110 only, NF90, and NF90+NF110 targets, respectively.
In each plot, $-\log_{10} p$-value of the log ratio test was plotted on the y-axis, and $\log_2\left(\frac{\sum_{f \in \mathcal{F}(j)} k_{ij}^f}{\sum_{f \in \mathcal{F}(j)} k_{ij}^{\phi_f}}\right)$ was plotted on the x-axis (Supplemental Methods). Identified targets are indicated in red. (C) A multi-objective optimization algorithm combined the results of the Poisson ratio test and the log ratio test. In each plot, the candidate target set was plotted, with $-\log_{10} p$-values of the Poisson ratio test and log ratio test on the x- and y-axes, respectively (Supplemental Methods). The final identified targets are indicated in red.

**Supplementary Figure S7 (to go with Figure 6)**

(A) Heatmap of $\log_2$(IP/input) RIP-seq expression values of previously reported NF90/NF110 targets (Kuwano et al. 2010) in WT, NF110 KO, and NF90+NF110 KO ESC RIP-seq. Note that NF90 targets are expected to have higher IP to input enrichment in NF110 KO ESCs when compared with NF90+NF110 KO ESCs. (B) Intersection of expression Module III (Fig. 5B, upregulated in all KO ESCs) and the union of NF110 only, NF90, and NF90+NF110 targets. Genes in red are related to cell development and differentiation. (C) Intersection of expression Module IX (Fig. 5B, downregulated in all KO ESCs) and the union of NF110 only, NF90, and NF90+NF110 targets. Genes in blue are related to cell division.

**Supplementary Figure S8 (to go with Figure 7)**

Combinatorial gene set analysis (Supplemental Methods) used to model the functional interactions among NF45, NF90, and NF110. Correlation plots supporting (A) Claim 1 and 2, (B) Corollary 1, (C) Claim 3, and (D) Corollary 2. (E) GO analysis of genes in identified regulatory modes through the combinatorial gene set analysis.
Supplementary Figure S1 (to go with Figure 1)

A
Expression relative to siCtrl
Normalized to Rpl7

B
Expression relative to siCtrl
Normalized to Rpl7

C
Expression relative to ESC
Normalized to sno202

Legend:
- siCtrl
- siNF90+NF110
- siNF110

Gene expression levels:
- NF45
- miR-293
- miR-294
- miR-295
Supplementary Figure S2 (to go with Figure 2)

A  Klf4

Expression normalized to Rpl7

NF45  NF45
WT   KO

B  Rex1

Expression normalized to Rpl7

NF45  NF45
WT   KO

C  Fgf5

Expression normalized to Rpl7

NF45  NF45
WT   KO

D  T

Expression normalized to Rpl7

NF45  NF45
WT   KO

E  Klf4

Expression normalized to Rpl7

NF90/ NF90+ NF110
NF110 NF110 KO

F  Rex1

Expression normalized to Rpl7

NF90/ NF90+ NF110
NF110 NF110 KO

G  Fgf5

Expression normalized to Rpl7

NF90/ NF90+ NF110
NF110 NF110 KO

H  T

Expression normalized to Rpl7

NF90/ NF90+ NF110
NF110 NF110 KO
Supplementary Figure S4 (to go with Figure 4)

A

| Total cell lysate | Cytoplasm | Nucleus |
|-------------------|-----------|---------|
| ESC               | EpiC      | ESC     | EpiC    |
| NF45              |           |         |         |
| NF110             |           |         |         |
| NF90              |           |         |         |
| Tubulin           |           |         |         |
| Nanog             |           |         |         |

B

| Cytoplasm | Nucleus |
|-----------|---------|
| NF45      |         |
| NF110     |         |
| NF90      |         |
| Tubulin   |         |
| Oct4A     |         |
| Oct4B     |         |

C

Expression relative to Rev NF90-NF110 levels
Normalized to Rpl7

D

NF45 expression relative to WT ESC
Normalized to Rpl7
Supplementary Figure S5 (to go with Figure 5)

(A) Klf4
(B) Rex1
(C) Fgf5
(D) T

(E) Klf4
(F) Rex1
(G) Fgf5
(H) T

(I) Module II
(II) Module VIII

- apoptotic process
- oxidation-reduction process
- negative regulation of cell growth
- negative regulation of cell proliferation
- negative regulation of fibroblast proliferation
- fructose 1,6-bisphosphate metabolic process
- positive regulation of astrocyte differentiation
- response to hypoxia
- response to toxic substance
- positive regulation of apoptotic process

- transcription, DNA-templated
- covalent chromatin modification
- regulation of transcription, DNA-templated
- positive regulation of transcription from RNA polymerase II promoter
- transforming growth factor beta receptor signaling pathway
- methylation
- mitotic nuclear division
- negative regulation of transcription from RNA polymerase II promoter
- placenta development
- negative regulation of cytokinesis

- Spearman $p = 0.30, p = 2.2 \times 10^{-7}$
- Spearman $p = -0.08, p = 0.16$
- Spearman $p = 0.24, p = 3.5 \times 10^{-47}$
- Spearman $p = -0.11, p = 1.1 \times 10^{-11}$
Supplementary Figure S6 (to go with Figure 6)

A

NF10 only targets

NF90 targets

NF90+NF110 targets

B

NF10 only targets

NF90 targets

NF90+NF110 targets

C

NF10 only targets

Spearmann r = 0.62

NF90 targets

Spearmann r = 0.74

NF90+NF110 targets

Spearmann r = 0.79
**Supplementary Figure S7** (to go with Figure 6)

A  
NF90 targets in HeLa cells

| Gene       | Expression Level |
|------------|------------------|
| Pfkb       | WT > WT KO       |
| Eftf3m     | WT > WT KO       |
| Cttna1     | NF90 KO > WT KO  |
| Effl       | NF90 KO > WT KO  |
| Cnpl       | NF90 KO > WT KO  |
| Mrps30     | NF90 KO > WT KO  |
| C1qb       | NF90 KO > WT KO  |
| Mapre1     | NF90 KO > WT KO  |
| Xbp1       | NF90 KO > WT KO  |
| Psma1      | NF90 KO > WT KO  |
| Slc25a3    | NF90 KO > WT KO  |
| Atnal1     | NF90 KO > WT KO  |
| Mps10      | NF90 KO > WT KO  |
| Mad2l1     | NF90 KO > WT KO  |
| Psma1      | NF90 KO > WT KO  |
| Mat2b      | NF90 KO > WT KO  |
| Cnih1      | NF90 KO > WT KO  |
| Prdx3      | NF90 KO > WT KO  |
| Cse1l      | NF90 KO > WT KO  |
| Ostc       | NF90 KO > WT KO  |
| C1qbp      | NF90 KO > WT KO  |
| Mat2b      | NF90 KO > WT KO  |
| Pex3       | NF90 KO > WT KO  |
| Ckap5      | NF90 KO > WT KO  |
| Gng10      | NF90 KO > WT KO  |
| Cld        | NF90 KO > WT KO  |
| C1d        | NF90 KO > WT KO  |
| Nus1       | NF90 KO > WT KO  |
| Cse1l      | NF90 KO > WT KO  |
| Cnpl       | NF90 KO > WT KO  |
| Cnih1      | NF90 KO > WT KO  |
| Selt       | NF90 KO > WT KO  |

B  
Intersection between expression module III (Fig. 5B, up-regulated in all KO ESCs) and the union of NF110 only, NF90, and NF90+NF110 targets

**Red:** related to cell development and differentiation.

- **Ints1**: Subunit of the Integrator complex.
- **Slc25a3**: Involved in the development and differentiation of digestive smooth muscle cells.
- **Lama5**: This gene encodes one of the vertebrate laminin alpha chains.
- **Sess2**: The encoded protein may function in the regulation of cell growth and survival. This protein may be involved in cellular response to different stress conditions.
- **Ckap5**: Receptor for netrin required for axon guidance. Mediates axon repulsion of neuronal growth cones in the developing nervous system upon ligand binding.
- **Calccol1**: Functions as a coactivator for aryl hydrocarbon and nuclear receptors.
- **Alpi**: This gene encodes a member of the alkaline phosphatase family of proteins.
- **Chf**: This gene is thought to play a role in tumor progression and left-right patterning.
- **Posk6**: The protein encoded by this gene is a single-pass type I transmembrane protein of unknown function.
- **Celari1**: The protein encoded by this gene is a member of the flamingo subfamily. This particular member is a developmentally regulated, neural-specific gene which plays an unspecified role in early embryogenesis.
- **Ccm1**: This gene encodes a member of the cyclin family. Acting as an inhibitor of centrosome duplication.
- **Npdc1**: Neural Proliferation Differentiation And Control Protein 1. Suppresses oncogenic transformation in neural and non-neural cells and down-regulates neural cell proliferation.
- **Chf**: Chondroitin Polymersing Factor.
- **Btbd2**: BTB Domain Containing 2.
- **Lamb2**: Laminin Subunit Beta 2.
- **Tjp1**: Tight Junction Associated Protein 1.
- **Nomcl**: This gene may antagonize Nodal signaling.

C  
Intersection between expression module IX (Fig. 5B, down-regulated in all KO ESCs) and the union of NF110 only, NF90, and NF90+NF110 targets

**Blue:** related to cell development and differentiation.

- **Syncr**: Synaptotagmin Binding Cytoplasmic RNA Interacting Protein.
- **Snx6**: Sorting Nexin 6.
- **Rnpg2**: ribosome-associated protein complex subunit of the ORC complex.
- **Orc2**: Component of the origin recognition complex (ORC) that binds origins of replication. Studies in yeast demonstrated that ORC binds specifically to origins of replication and serves as a platform for the assembly of additional initiation factors such as Cdc6 and Mcm proteins. The protein encoded by this gene is a subunit of the ORC complex.
- **Smc4**: Subunit of the condensin complex, a complex required for conversion of interphase chromatin into mitotic-like condense chromosomes.
- **Top1**: Topoisomerase (DNA) 1.
- **Rdx**: Radixin is a cytoskeletal protein that may be important in linking actin to the plasma membrane.
- **Farsb**: This gene encodes a highly conserved enzyme that belongs to the aminoacyl-tRNA synthetase class IIc subfamily.
- **Dek**: This gene encodes a protein with one SAP domain. This protein binds to cruciform and superhelical DNA and induces positive supercoils into closed circular DNA, and is also involved in splice site selection during RNA processing.
- **Gmps**: Involved in the de novo synthesis of guanine nucleotides which are not only essential for DNA and RNA synthesis, but also provide GTP, which is involved in a number of cellular processes important for cell division.
- **Ckap5**: Binds to the plus end of microtubules and regulates microtubule dynamics and microtubule organization. Promotes cytoplasmic microtubule nucleation and elongation. Plays a major role in organizing spindle poles.
Supplementary Figure S8 (to go with Figure 7) continued

E

- Fructose 1,6-bisphosphate metabolic process
- Positive regulation of macrophage cytokine production
- Positive regulation of apoptotic signaling pathway
- Positive regulation of neuron fate commitment
- Response to hypoxia
- Multicellular organism development
- Amino acid transport
- Cell differentiation
- Neuron fate commitment
- Positive regulation of cell proliferation

Mode A

- Axon guidance
- Transcription, DNA-templated
- Lung alveolus development
- Neuron projection development
- Negative regulation of microtubule depolymerization
- Negative regulation of cell migration
- Covalent chromatin modification
- Mitotic cytokinesis
- Actin cytoskeleton organization
- Adipose tissue development

Mode C

- Cell cycle
- Mitotic nuclear division
- Cell division
- Regulation of transcription, DNA-templated
- Chromosome organization
- Ubiquitin-dependent protein catabolic process
- Mitotic chromosome condensation
- Transcription, DNA-templated
- mRNA transport
- Chromosome segregation

Mode D

- Regulation of defense response to virus by host
- Oxidation-reduction process
- Protein homotetramerization
- Telomere maintenance via telomere lengthening
- Positive regulation of astrocyte differentiation
- Negative regulation of protein processing
- Heart trabecula formation
- Retinal cone cell development

Mode F