

Published in final edited form as:

Science. 2014 June 6; 344(6188): 1156–1160. doi:10.1126/science.1248882.

Defining an Essential Transcription Factor Program for Naïve Pluripotency

S-J Dunn^{#1}, G Martello^{#2,*‡}, B Yordanov^{#1}, S Emmott¹, and AG Smith^{2,3,*}

¹Computational Science Laboratory, Microsoft Research, Cambridge, CB1 2FB, UK

²Wellcome Trust - Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, CB2 1QR, UK

³Department of Biochemistry, University of Cambridge, Cambridge, UK

These authors contributed equally to this work.

Abstract

The gene regulatory circuitry through which pluripotent embryonic stem (ES) cells choose between self-renewal and differentiation appears vast and has yet to be distilled into an executive molecular program. Here we developed a data-constrained, computational approach to reduce complexity and derive a set of functionally validated components and interaction combinations sufficient to explain observed ES cell behavior. This minimal set, the simplest version of which comprises only 16 interactions, 12 components and 3 inputs, satisfies all prior specifications for self-renewal and furthermore predicts unknown and non-intuitive responses to compound genetic perturbations with an overall accuracy of 70%. We propose that propagation of ES cell identity is not determined by a vast interactome, but rather can be explained by a relatively simple process of molecular computation.

Mouse embryonic stem (ES) cells exhibit the capacity to self-renew indefinitely, the plasticity to generate all somatic lineages and germ cells, and the ability to re-enter embryogenesis through blastocyst injection. Collectively these properties are described as ground state pluripotency (1). The pluripotent state of mouse ES cells has been considered as a vast network of genetic interactions (2-6). However, only a limited number of Transcription Factors (TFs) have been validated rigorously. Two have been found to be irreplaceable: Oct4 and Sox2. In contrast, factors such as Nanog, Klf4, and Esrrb, are individually dispensable, yet their overexpression can support self-renewal (7-9).

Culture environments for efficiently sustaining mouse ES cells have been progressively refined. Currently optimized conditions comprise the cytokine leukemia inhibitory factor (LIF), together with two selective inhibitors (2i) of GSK3 (Chiron99021, CH) and Mek (PD0325901, PD) (10). A number of direct transcriptional targets downstream of LIF, CH and PD have been identified (Fig. S1A) (8, 9, 11-14). Although there is evidence of extensive cross-regulation between the TFs, it is not understood how the environmental

*Correspondence to: graziano.martello@unipd.it, austin.smith@cscr.cam.ac.uk.

‡Current address: Dept. of Molecular Medicine, University of Padua, Padua, 35131 Italy

signals are processed to stabilize the network, and to ensure self-renewal. Characterizing this circuitry should allow us to explain and predict the self-renewal behavior of ES cells.

ES cells can be stably cultured as a substantially homogeneous population in 4 different environments: 2i+LIF, 2i, LIF+CH, LIF+PD (Fig. 1A and S2A-C). Under these conditions, cells show no signs of differentiation, self-renew at clonal density, homogeneously express pluripotency markers, and form germline chimeras after blastocyst injection (Fig. S2E). This system allows us to study gene expression of pluripotency factors at steady state in different culture conditions with similar ability to maintain ES cells. Furthermore, cells can be efficiently transitioned from one culture condition to another without affecting pluripotency or cell survival (Fig. S2D), allowing adjustments in gene expression circuitry to be measured over time.

We analyzed gene expression for 17 factors implicated in ES cell maintenance and inferred to be cross-regulatory (Fig. 1B) (1). To identify plausible TF interactions, we examined expression correlation between all possible TF pairs across the different culture conditions. Steady state data were obtained from ES cells grown in each condition for 10 days, and timecourse data were obtained from cells cultured for 10 days in each of 2i, LIF+CH and LIF+PD, before adding the third component to reconstitute 2i+LIF (Fig. 1B). By way of example, the strong negative correlation observed between *Esrrb* and *Tcf3* (Fig. 1C) suggests a repressive interaction.

We used the Pearson coefficient as a metric to quantify the correlation between any two TFs (Fig. S3A). Accordingly, we constructed a ‘meta-model’ of the pluripotency network (Fig. 1D), defined as a set of possible undirected interactions, each suggested by the Pearson correlation data (Section S1J). The meta-model therefore consists of a set of models, each a unique instantiation of possible interactions that might explain ES cell behavior. Inferred interactions are assumed to be functional, but not necessarily direct.

The possible combinations of interactions between the TFs number into billions of billions. Distillation to only those models that capture observed behavior, and can explain ES cell information processing and decision-making, necessitated development of a new computational approach.

We adopted Boolean network formalism to abstract gene expression levels as low or high within the endogenous expression range. Computational analyses were performed using a bespoke software tool designed to encode possible genetic interactions and behavioral constraints (Section S1I). The tool implements formal verification procedures (15) to synthesize those interaction combinations and gene regulation functions that *provably* satisfy experimental data, thereby constraining the set of possible models. Moreover, the tool enables interrogation of the entire constrained meta-model to generate predictions of behavior under genetic and environmental perturbations.

We defined a set of 23 experimental constraints, comprising conversion between culture conditions, loss of signals, and gain/loss of gene function experiments (Fig. 2A, Section S1K). Each constraint consists of an initial and final gene expression pattern that must be

reached in a finite number of steps. Starting from the meta-model, we identified a subset of models that can reproduce all 23 constraints simultaneously.

To reduce model complexity, and to maximize the predictive capability of the meta-model (Section S1M), we sought only those components and interactions that are essential. A high Pearson correlation threshold is desirable to select significant possible interactions. The maximum Pearson correlation threshold that identified the smallest set of interactions needed to satisfy all experimental constraints simultaneously is 0.792. This gave a set of 70 possible interactions. We found that in addition to known signal targets, 11 of the possible interactions were present in all of these models. We conclude that these are essential interactions for the pluripotency network (Fig. 2B). We further found that 5 components are not required and can be eliminated: Rex1, Nr0b1, Mbd3, Mi2b and Klf5 (Fig. 2B, in grey). We thus derived a first set of minimal models to explain the known behavior of ES cells.

The set of possible interactions were compared to those inferred from open-source microarray data following genetic perturbations, and also with ChIP-Seq data. We found our set of possible interactions are supported by these independent data, but that a set of interactions derived only from perturbation data was insufficient to satisfy experimental observations (Section S2A, Figs. S4, S5). Secondly, we compared our computational approach with Bayesian network inference (16). The latter yielded fewer interactions and only one model that was inconsistent with experimental observations (Fig. S5B).

ES cells cultured in 2i+LIF show a remarkable degree of robustness because they can tolerate the singular loss of key TFs such as Nanog or Esrrb (8, 9). However, the response of ES cells to compound perturbations is unknown, difficult to predict, and time-consuming to investigate by unsupervised experimental tests. We used the complete set of data-constrained models to predict the response of ES cells to single and double knockdowns in 2i+LIF.

We determined whether a given knockdown was predicted to result in maintenance or collapse of self-renewal (Section S1L) (17-19). We accepted predictions only when all of the candidate models were in agreement. This led to 11 predictions from the 24 possible double knockdowns, and 1 single knockdown prediction (Fig. 3A, left panel, prediction column).

To test the predictions we employed siRNA transfection followed by clonal assay, which provides a quantitative measure of self-renewal capacity (Fig. 3A, left panel, exp. validation column). A subset of predictions was also tested using knockout ES cells (Fig. S6B-C). The experimental results confirmed all predictions, with the sole exception of the *Tbx3/Klf2* double knockdown. Several predictions were non-intuitive: for example, the effect of *Klf2/Sall4* double knockdown could not be inferred from single knockdown results (Fig. 3B).

No single model within the meta-model is capable of producing the observed behavior after *Tbx3/Klf2* double knockdown. We therefore used this incorrect prediction as a new experimental constraint to derive a more accurate set of models by incrementally lowering the Pearson coefficient threshold to identify the most supported subset of interactions

(Section S1M). We found that including an optional positive interaction between Oct4 and Sox2 was sufficient to introduce the necessary flexibility to the network.

With the refined meta-model we sought predictions of the response to genetic perturbations in different conditions (Fig. 3A, right panel, Fig. S6F-G). In 2i, LIF+CH and LIF+PD conditions, we found 17 of 28 tested predictions were correct (Fig. 3A and S6D).

Interestingly, the outcomes of some double knockdowns, including *Esrrb/Nanog*, were predicted to be strictly dependent on the culture environment: impeding self-renewal in 2i but not in 2i+LIF (Fig. 3B). Thus our modelling approach can accurately recapitulate the behavior of the network in response to input signals.

The meta-model predictions suggest that the pluripotency network is less resistant to genetic perturbation in 2i than in 2i+LIF (compare panels, Fig. 3A). We tested whether a similar difference in robustness could be observed upon environmental perturbation. We transferred ES cells from either 2i+LIF or 2i only into unsupplemented medium, then after different periods performed a clonogenicity assay by replating in 2i+LIF. The number of self-renewable cells declined more rapidly for 2i than 2i+LIF cultures (Fig. S6H-I). Rex1GFP was also downregulated faster from 2i (Fig. 3C). These results suggest that the pluripotent state collapses more readily in 2i ES cells, consistent with the network being less stable than in 2i+LIF.

To identify the simplest among the set of equally valid candidate models constrained using the additional experimental data from knockdowns in 2i+LIF, we explicitly sought those with the fewest interactions. Only 16 interactions between 12 components may be required because *Tbx3* is dispensable as a regulator of any other component (Section S1N). There is only one valid model of this size, representing the minimal possible transcriptional program governing ES cell self-renewal (Fig. 3D). Interestingly, a repressive effect of *Esrrb* on *Oct4* is required in all of the constrained models.

Our results reveal that ES cell decision-making is not necessarily dependent on a vast gene network, as widely considered, but instead can be explained by a program that, in its simplest version, consists of just 16 interactions, 12 components and 3 inputs. This example model from the meta-model identifies the most parsimonious network that could maintain naïve pluripotency under four different culture conditions. Environmental signals are processed via interactions between pluripotency factors to stabilize the gene expression pattern that defines a self-renewing, pluripotent cell. The program itself is unchanging under the action of different inputs, but computes the appropriate stable state as a consequence of these inputs. This conclusion is consistent with the proposition that individual cell states are attractors (20).

It will be instructive to learn how the current meta-model can be refined further to explain how cells exit from the pluripotent state for differentiation, and how this state is generated in the developing embryo or by reprogramming. Furthermore, the general design of our computational approach means that it can be applied to understand biological computation in other cell systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank T. Kalkan, H. Kugler and S. Dupont for stimulating discussions, and comments on this manuscript. We also thank B. Simons for reading the manuscript. GM was a recipient of a Human Frontier Science Program Fellowship. AS is a Medical Research Council Professor and is also funded by The Wellcome Trust.

References and Notes

1. Nichols J, Smith A. *Cold Spring Harb. Perspect. Biol.* 2012; 4:a008128. [PubMed: 22855723]
2. Chen X, et al. *Cell.* 2008; 133:1106–17. [PubMed: 18555785]
3. Lu R, et al. *Nature.* 2009; 462:358–62. [PubMed: 19924215]
4. van denBerg DLC, et al. *Cell Stem Cell.* 2010; 6:369–81. [PubMed: 20362541]
5. Som A, et al. *PLoS One.* 2010; 5:e15165. [PubMed: 21179244]
6. Yeo J-C, Ng H-H. *Cell Res.* 2013; 23:20–32. [PubMed: 23229513]
7. Chambers I, et al. *Nature.* 2007; 450:1230–4. [PubMed: 18097409]
8. Martello G, et al. *Cell Stem Cell.* 2012; 11:491–504. [PubMed: 23040478]
9. Martello G, Bertone P, Smith A. *EMBO J.* 2013:1–14. [PubMed: 23211745]
10. Ying Q-L, et al. *Nature.* 2008; 453:519–23. [PubMed: 18497825]
11. Silva J, et al. *Cell.* 2009; 138:722–37. [PubMed: 19703398]
12. Hall J, et al. *Cell Stem Cell.* 2009; 5:597–609. [PubMed: 19951688]
13. Tai C-I, Ying Q-L. *J. Cell Sci.* 2013; 126:1093–8. [PubMed: 23345404]
14. Lanner F, et al. *Stem Cells.* 2010; 28:191–200. [PubMed: 19937756]
15. Yordanov B, Wintersteiger CM, Hamadi Y, Kugler H. *NASA Formal Methods Symposium.* 2013:78–92.
16. Wilczy ski B, Dojer N. *Bioinformatics.* 2009; 25:286–7. [PubMed: 18826957]
17. Masui S, et al. *Nat. Cell Biol.* 2007; 9:625–35. [PubMed: 17515932]
18. Ivanova N, et al. *Nature.* 2006; 442:533–538. [PubMed: 16767105]
19. Niwa H, Miyazaki J, Smith AG. *Nat. Genet.* 2000; 24:372–6. [PubMed: 10742100]
20. Enver T, Pera M, Peterson C, Andrews PW. *Cell Stem Cell.* 2009; 4:387–97. [PubMed: 19427289]

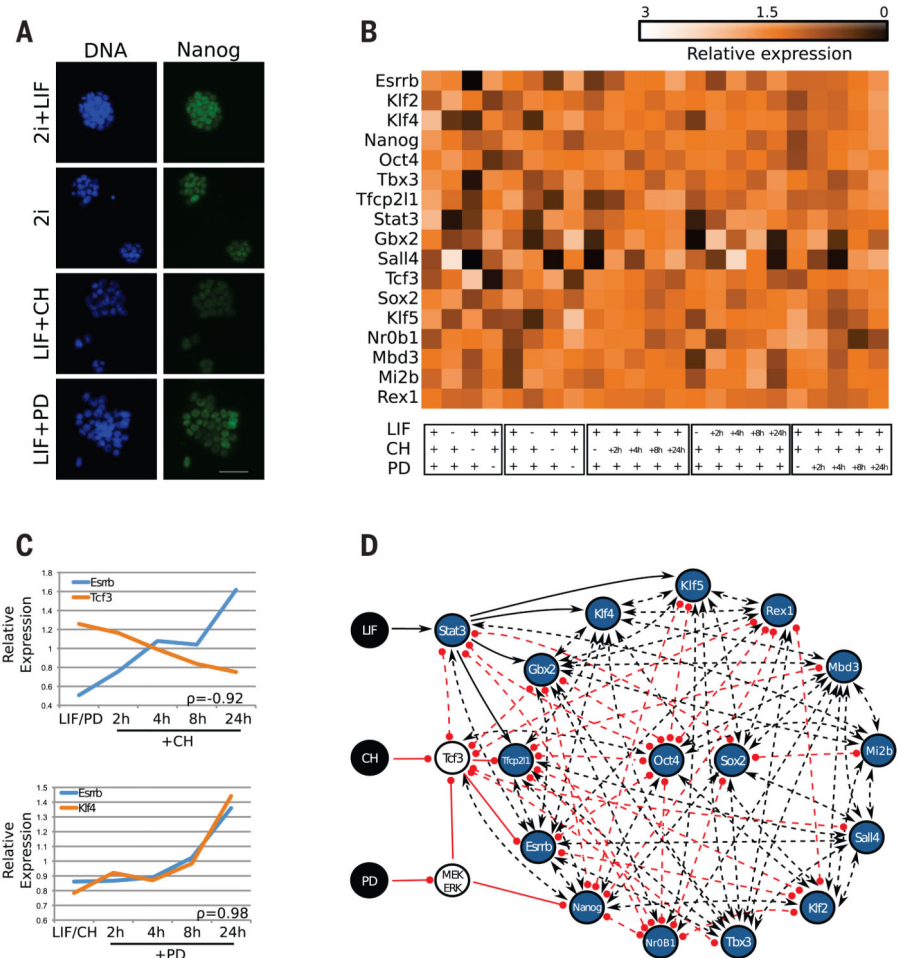


Fig. 1. Flexible culture conditions for mouse ES cells allow deduction of possible interactions between pluripotency factors

(A) Under different combinations of LIF, CH, and PD, ES cells show homogeneous expression of Rex1, Oct4, Esrrb, and Nanog. See also Figure S2A-C. Scale bar: 50 μ m. (B) Gene expression for the set of 17 pluripotency-associated TFs from 5 experiments under different culture conditions. Columns 1-4 and 5-8 are biological replicates in the indicated steady state conditions, while columns 9-23 illustrate the change in gene expression over time from cells in LIF+PD, 2i or LIF+CH switched to 2i+LIF. Gene expression was measured by qRT-PCR and normalized to mean expression level. β -actin was used as an internal control. (C) Examples showing negative correlation between Esrrb and Tcf3, and positive correlation between expression of Esrrb and Klf4 under time series experiments. Pearson correlation coefficient indicated on each plot. (D) The meta-model derived from a Pearson coefficient threshold of 0.7, with the expected gene expression under 2i+LIF conditions. Gene expression is discretized to high (blue) or low (white). Positive regulation indicated by a black arrow, negative regulation indicated by a red circle-headed line. Dashed lines indicate optional interactions, solid lines indicate definite interactions.

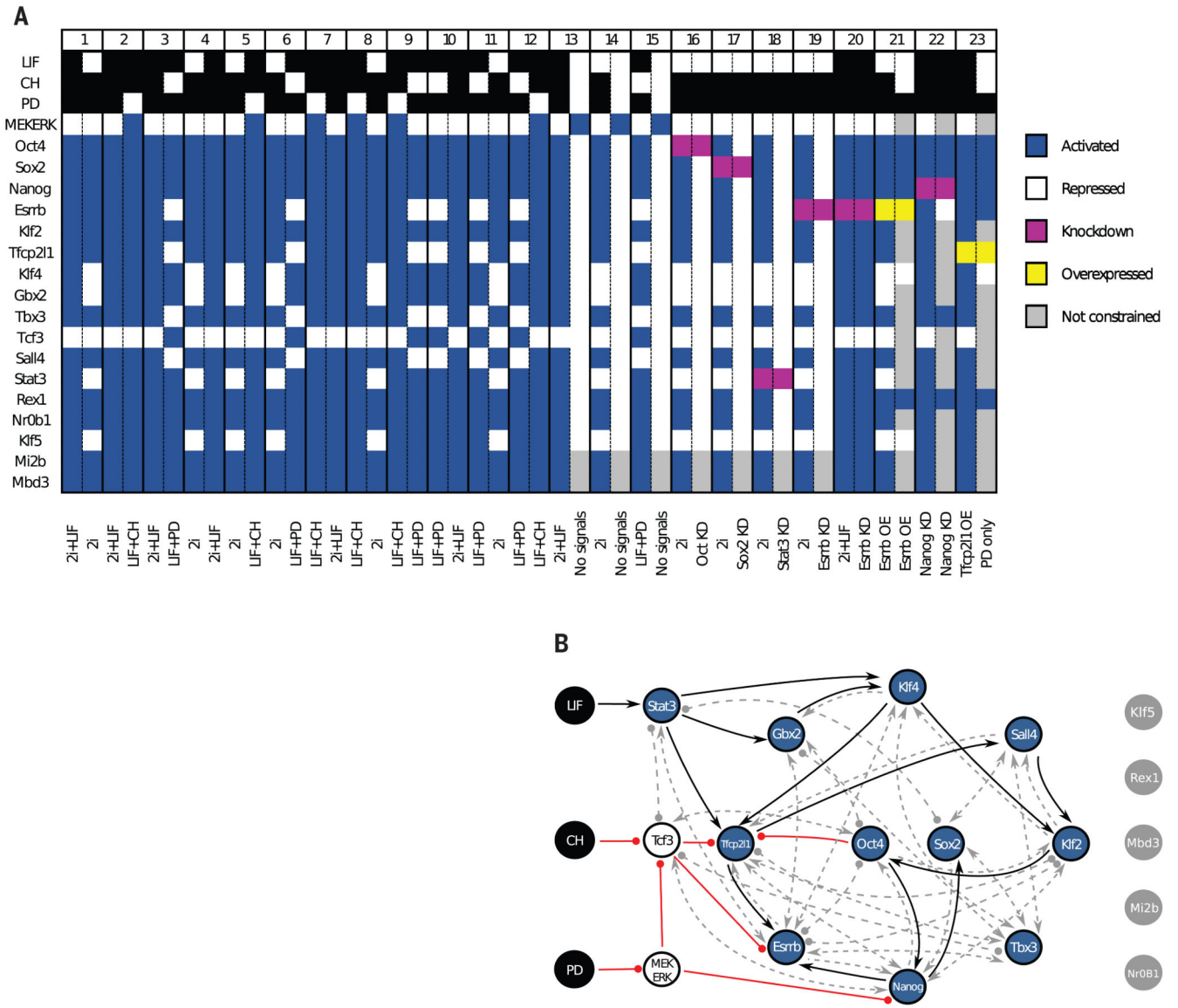


Fig. 2. Iteration with experimental observations reduces model complexity

(A) 23 experimental constraints are defined, each with initial (left column) and final (right column) conditions (Section 1K). (B) The meta-model derived from a Pearson correlation threshold of 0.792, constrained to satisfy the expected experimental behavior (panel A). 11 of the possible interactions are present in all possible models, and the remaining possible interactions are shown by dashed grey lines. All of the constraints can be satisfied without including five components of the network (grey).

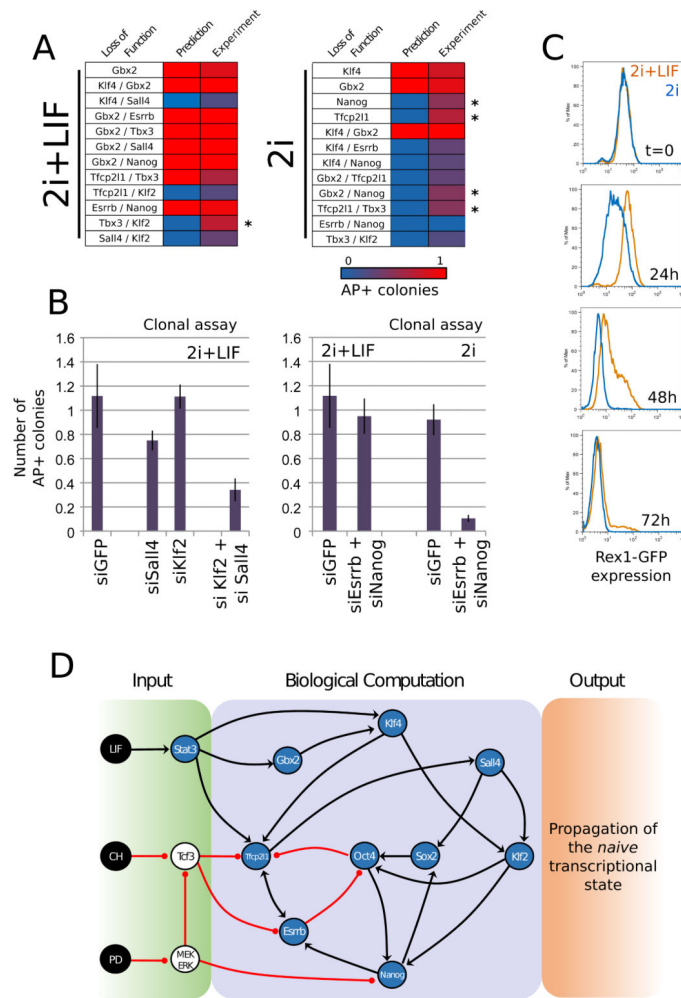


Fig. 3. Experimental validation of model predictions

(A) Summary of model predictions and experimental validation. Incorrect predictions highlighted with an asterisk. Each box in the “Exp. Validation” column indicates the number of AP+ colonies obtained in a clonal assay. See also Fig. S5. (B) Clonal assay of ES cells in the indicated culture conditions. For each sample the number of AP+ colonies, relative to non-targeting siRNA transfected cells, is indicated. Each bar represents the mean and SEM of at least 4 independent experiments. (C) Rex1GFPd2 flow profile of cells withdrawn from either 2i+LIF or 2i conditions for the indicated time. Cells taken from 2i downregulate the reporter more rapidly. (D) The simplest model of 16 interactions and 12 components required to satisfy all experimental constraints. The complete set of constrained models is shown in Fig. S7.