Published in final edited form as:
Nature. 2009 November 19; 462(7271): 358–362. doi:10.1038/nature08575.

Systems-level dynamic analyses of fate change in murine embryonic stem cells

Rong Lu[1,†], **Florian Markowetz**[2,†,*], **Richard D. Unwin**[3,*], **Jeffrey T. Leek**[4,†], **Edoardo M. Airoldi**[2,†], **Ben D. MacArthur**[4,5], **Alexander Lachmann**[5], **Roye Rozov**[4,†], **Avi Ma'ayan**[5], **Laurie A. Boyer**[6], **Olga G. Troyanskaya**[2], **Anthony D. Whetton**[3], and **Ihor R. Lemischka**[1,4]

[1]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

[2]Lewis-Sigler Institute for Integrative Genomics and Department of Computer Science, Princeton University, Princeton, New Jersey 08544, USA

[3]Stem Cell and Leukaemia Proteomics Laboratory, School of Cancer and Imaging Sciences, Manchester Academic Health Science Centre, University of Manchester, Wolfson Molecular Imaging Centre, Manchester M20 4QL, UK

[4]Department of Gene and Cell Medicine and The Black Family Stem Cell Institute, Mount Sinai School of Medicine, New York, New York 10029, USA

[5]Department of Pharmacology and System Therapeutics and Systems Biology Center New York (SBCNY), Mount Sinai School of Medicine, New York, New York 10029, USA

[6]Massachusetts Institute of Technology, Department of Biology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

Abstract

Molecular regulation of embryonic stem cell (ESC) fate involves a coordinated interaction between epigenetic[1–4], transcriptional[5–10] and translational[11,12] mechanisms. It is unclear how these different molecular regulatory mechanisms interact to regulate changes in stem cell fate. Here we present a dynamic systems-level study of cell fate change in murine ESCs following a well-defined perturbation. Global changes in histone acetylation, chromatin-bound RNA polymerase II, messenger RNA (mRNA), and nuclear protein levels were measured over 5 days after downregulation of Nanog, a key pluripotency regulator[13–15]. Our data demonstrate how a single genetic perturbation leads to progressive widespread changes in several molecular

Correspondence and requests for materials should be addressed to I.R.L. (ihor.lemischka@mssm.edu) or R.L. (rlu@stanford.edu).
[†]Present addresses: Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Beckman Center B261, 279 Campus Drive, Stanford, California 94305, USA (R.L.); Cancer Research UK, Cambridge Research Institute, Cambridge CB2 0RE, UK (F.M.); Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics 615 North Wolfe Street, Baltimore, Maryland 21205, USA (J.T.L.); Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, Massachusetts 02128, USA (E.M.A.); Blavatnik School of Computer Science, Tel Aviv University, 69978 Tel Aviv, Israel (R.R.).
[*]These authors contributed equally to this work.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions R.L. and I.R.L. designed the experiments. R.L. prepared the cell samples for all the experiments, performed the RNA polymerase II ChIP-chip, the mRNA microarray, and verification experiments such as western blot, ChIP and quantitative PCR. R.D.U. and A.D.W. performed the proteomic experiments and primary analysis on proteomic data. L.A.B. performed the histone acetylation ChIP-chip experiments. R.L., F.M., E.M.A., R.R. and O.G.T. performed general data processing and statistical analyses. R.L. and F.M. plotted Figs 1–3. A.L., B.D.M. and A.M. developed and plotted interactive Fig. 4a. A.L. and A.M. developed and plotted interactive Fig. 4b. R.L., J.L., F.M. and I.R.L performed network analysis shown in Fig. 3. R.L. and I.R.L. wrote the paper.

regulatory layers, and provide a dynamic view of information flow in the epigenome, transcriptome and proteome. We observe that a large proportion of changes in nuclear protein levels are not accompanied by concordant changes in the expression of corresponding mRNAs, indicating important roles for translational and post-translational regulation of ESC fate. Gene-ontology analysis across different molecular layers indicates that although chromatin reconfiguration is important for altering cell fate, it is preceded by transcription-factor-mediated regulatory events. The temporal order of gene expression alterations shows the order of the regulatory network reconfiguration and offers further insight into the gene regulatory network. Our studies extend the conventional systems biology approach to include many molecular species, regulatory layers and temporal series, and underscore the complexity of the multilayer regulatory mechanisms responsible for changes in protein expression that determine stem cell fate.

We applied a single well-defined perturbation to murine ESCs by downregulating Nanog, a key pluripotency factor[13–15]. A lentiviral-based complementation system was introduced into mouse ESCs in which short hairpin RNA (shRNA) depletes endogenous *Nanog* mRNA, and normal levels of Nanog expression are restored in a doxycycline-dependent manner from an shRNA 'immune' version[7] (Fig. 1b). Previously, we showed that this engineered ESC clone is fully pluripotent *in vitro* and *in vivo* when maintained in the presence of doxycycline[7]. After doxycycline removal, Nanog mRNA and protein levels rapidly decline (Fig. 1c), and both pluripotency and self-renewal capacities of ESCs diminish with time. We collected data from four molecular layers. Specifically, we performed: (1) chromatin-immunoprecipitation microarray (ChIP-chip) analysis of histone H3 lysine 9 and 14 acetylation (acH3K9/14) at gene promoter regions to assess chromatin modification (designated as HIS); (2) ChIP-chip analysis of RNA polymerase II localization at 3′ exons of gene coding regions to reveal active transcription (designated as POL); (3) gene expression microarrays to quantify mRNA abundance (designated as RNA); and (4) protein mass spectrometry to measure nuclear protein abundance (designated as PRO) (Fig. 1a). Fold changes were calculated for each gene by comparing the expression levels of a molecular layer on days 1, 3 and 5 (doxycycline absent, Nanog depleted) to day 0 (doxycycline present, Nanog expressing), allowing for comparisons across the different experimental platforms (Supplementary Fig. 1). To estimate experimental noise, a significance threshold in each experiment was determined based on the experimental replicates of all measured genes at a false discovery rate (FDR) of 5% (Fig. 1d and Supplementary Fig. 2).

Although changes between different gene expression steps are generally correlated (Supplementary Fig. 3), both concordances and discordances exist on the individual gene level. The discordances show regulatory events that alter gene expression. We performed a supervised gene/protein classification to identify the key regulatory step that is most responsible for changes in protein levels, which directly determine cellular phenotype. We anchored our analysis on observed changes in protein levels and assessed the concordance of changes in the other three layers by comparing PRO to RNA, then RNA to POL, and finally POL to HIS (Fig. 2a). Proteins with significant changes were assigned to one of four categories at each time-point: category 1 proteins exhibit discordant PRO and RNA changes in expression, which is indicative of translational and posttranslational regulation; category 2 proteins exhibit concordant PRO and RNA changes in expression, but discordant RNA and POL changes in expression, which is indicative of post-transcriptional regulation; category 3 proteins exhibit concordant PRO, RNA and POL changes in expression, but discordant POL and HIS changes in expression, which is indicative of transcriptional regulation; and category 4 proteins exhibit concordant changes in expression across all four layers, which is indicative of regulation through chromatin modification. Proteins tend to stay in the same category over time (Supplementary Fig. 4). Category 1 constitutes 43–52% of all the genes with significant changes in protein levels, indicating that translational and

post-translational regulatory mechanisms have important roles in ESC fate decisions[11,12,16,17]. However, it is unclear whether this is specific to stem cells or whether it is characteristic of other biological systems.

In addition to providing a genome-wide perspective of ESC fate change, our concordance analysis also provides useful information on the level of individual genes (Fig. 2b). For example, the ESC transcriptional regulator Esrrb[7] falls into the category 2 concordance pattern at all time points. This indicates that ultimate levels of Esrrb protein are primarily regulated post-transcriptionally, at least under our experimental conditions, and not by direct Nanog regulation at the transcriptional level. It has been proposed that Esrrb and Nanog mutually regulate each other by a positive feedback circuit[6,18]. Our concordance pattern analysis of Esrrb indicates that at least one other component is likely to be involved in this circuit, which is responsible for the post-transcriptional regulation of Esrrb, possibly a microRNA[19,20].

Gene-ontology analyses across the four molecular layers suggest a complex interaction between different molecular regulatory mechanisms in cell fate regulation (Fig. 2c and Supplementary Fig. 5). For example, differentiation- and development-related genes are over-represented among the genes that only show changes in acH3K9/14 levels, but not on the other three layers (Fig. 2c). Furthermore, chromatin- and nucleosome-assembly-related genes are overrepresented among the genes upregulated on the RNA polymerase II binding layer but not on any of the other three layers (Fig. 2c), suggesting that the chromatin modifiers are primarily regulated at the transcription step. Therefore, reconfiguration of chromatin structure, although an important factor in ESC fate alteration, may have a secondary role to primary regulation by transcription factors[5,6,8,21–23].

To gain further insight into systems-level regulatory control of changes in ESC fate, we combined our data with that of previous stem cell regulatory network studies to form a new synthesis (Fig. 3)[6,8,24]. A core protein–protein interaction network was previously identified in murine ESCs involving 26 proteins centred around Nanog[24]. We found that this interactome is enriched in proteins that decreased in expression after downregulation of Nanog (Supplementary Fig. 6). On day 5, 8 out of the 26 interactome proteins are at significantly reduced levels (Supplementary Fig. 7). These are: Sall4, Rnf2, Oct4 (also known as Pou5f1), Ilf2, Nanog, Mybbp1a, Sall1 and Esrrb. Of these eight proteins only one (Rnf2) does not directly interact with Nanog (Fig. 3a). This suggests interdependence between the Nanog interactome and the network of genes under Nanog transcriptional control.

Nanog protein binds to thousands of genomic locations in undifferentiated ESCs[5,6]. Our data show that approximately 20% of the previously identified Nanog-binding genes change their transcription levels (POL) during the first 5 days after Nanog downregulation. Of those that changed, approximately 50% also exhibit changes in protein levels (PRO) (Fig. 3b and Supplementary Fig. 7). To determine how the changes in expression develop after the downregulation of Nanog, we analysed the temporal alterations of mRNAs in the context of an extended transcriptional regulatory network proposed previously[8] (Fig. 3c). Our data show that most genes in this network are downregulated after the removal of Nanog. In particular, downregulation of Oct4 and Sox2 (protein levels shown in Supplementary Fig. 7) occurred later than downregulation of Klf4 or Rex1. This suggests that decreases in Oct4 and Sox2 expression are not responsible for decreases in Klf4 and Rex1 expression under our experimental conditions. The temporal sequence of changes in gene expression is loosely correlated with the chromatin-binding data[6,8]. These two sources provide independent and complementary information about the ESC gene regulatory network. Using the same principle that later molecular events cannot regulate earlier events, we can extract

new sets of useful information concerning the gene regulatory relations from the temporal order of the network reconfiguration (Fig. 4 and Supplementary Fig. 8).

To facilitate comparisons and visualization of the multilayered time series, we generated interactive movies to display our data (Fig. 4 and Supplementary Fig. 8; http://amp.pharm.mssm.edu/ronglu). Expression changes for 400 genes with the most significant changes in protein levels on day 5 were projected onto two-dimensional hexagonal arrays (Fig. 4a). Individual hexagons representing specific genes are dynamically coloured according to the fold changes in each of the four molecular layers. This approach facilitates genome-wide and temporal comparisons among the different molecular layers, and allows clustering of genes with similar dynamics on multiple gene expression regulatory layers. We have also generated interactive scatter plot movies to help visualize concurrent changes across the different molecular layers (Fig. 4b). In these movies, individual genes can be selected to illustrate the concurrent changes between pairs of molecular layers. For instance, Fig. 4b demonstrates that changes in Esrrb mRNA and protein expression are monotonically related, whereas Sall1 and Oct4 both show increased mRNA levels without any corresponding increase in protein levels during the early stage of ESC differentiation. Similar dynamics are also exhibited by several other previously identified essential ESC factors[25] (shown as red dots in Fig. 4b). These genes are regulated on different regulatory layer(s) compared to Esrrb, and suggest that the transcription layer undergoes an early cell fate reconfiguration without significant accompanying changes in protein production. Recent studies proposed that fluctuating levels of Nanog may discriminate between alternative pluripotent states of ESCs, in which high or low Nanog levels render ESCs resistant or susceptible to differentiation inducing stimuli, respectively[15,26–29]. In our system, the early time point of Nanog downregulation is comparable to the ‘low’ Nanog state from these studies. Thus, the absence of changes in protein levels during the mRNA layer reconfigurations could reflect the nature of these distinct pluripotent states. Collectively, the variety of the multilayered expression patterns underscores the complexity of the molecular regulation of ESC fate and suggests an intricate regulatory network involving several molecular regulatory layers.

In this study we have provided a dynamic multimolecular layer view of a murine ESC fate change in response to the downregulation of Nanog. In our experimental system the transcription of *Nanog* is regulated by exogenous manipulation and not by the endogenous regulatory circuit. This disrupts the balance of mutually regulated ESC molecular circuits[15,26–29], and allows for rapid and synchronous cell fate changes within the population. However, our results nonetheless represent the average of a large cell population, as we have shown previously that removing Nanog results in a complex mixture of cell lineages[7]. In this work, our primary goal was to analyse the molecular dynamics that are associated with the transition away from the pluripotent state as it occurs in most of the cells. *In vivo*, cell fate alteration is probably triggered by several perturbations and inputs dynamically. The single gene perturbation that we have used does not reflect the natural signals that pluripotent cells are subjected to *in vivo*. However, it is a powerful tool to dissect the complex regulatory networks that underpin ESC fate changes and offers an initial window into the dynamic complexity of ESC fate regulation across multiple molecular levels.

METHODS SUMMARY

AcH3K9/14 levels were assayed using ChIP-chip. Acetylated regions in a 1-kilobase window around the transcription initiation position were identified to generate acetylation profiles (Supplementary Figs 9 and 10). ChIP-chip was also used to measure RNA polymerase II localization on 3′ exons to directly assess transcriptional activity (elongation).

Changes in mRNA levels were monitored using Agilent two-colour microarrays. Nuclear protein levels were measured using peptide isobaric tagging followed by two-dimensional liquid chromatography mass spectrometry (LC-MS/MS)[16]. We chose to measure nuclear protein levels because cell fate determination is largely controlled in the nucleus. For technical reasons, attempts to measure the entire proteome would have significantly decreased the sensitivity of the nuclear protein measurements, as these only constitute approximately 20% of all proteins in ESCs. All experiments were conducted in triplicate except for the acH3K9/14 measurements, which were performed in duplicate. Reliability of all data sets was verified using independent experimental assays, including conventional chromatin immunoprecipitation (ChIP), quantitative PCR (qPCR), and western blot assays for key pluripotency regulator genes (Supplementary Figs 11 and 12). Experimental reproducibility was also verified using a linear analysis of variance (ANOVA) model[30]. After data pre-processing and normalization, we were able to validate 1,627 nuclear proteins and 12,488 genes (HIS/POL/RNA) with high confidence. For 1,212 nuclear proteins, we were able to jointly obtain high-quality data across all four layers (HIS/POL/RNA/PRO). Supplementary Fig. 1 provides an overview of the entire data processing pipeline and the results of the quality-control procedures (ANOVA analysis). The significance of change is determined at a FDR of 5% using an empirical Bayes' model with Benjamini–Hochberg correction on the basis of experimental replicates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank D. Storton for technical support, and E. Wieschaus, Y. Shi, S. Tavazoie and N. Slavov for constructive discussions. We also acknowledge the laboratories of the following people for providing antibodies for western blot: A. Okuda, J. Flint and Y. Kang. This work was supported by the NIH, and in part supported by the BBSRC and Leukaemia Research UK. O.G.T., F.M. and E.M.A. were partially supported by the NIH and US National Science Foundation.

References

1. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]
2. Boyer LA, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 2006; 441:349–353. [PubMed: 16625203]
3. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–560. [PubMed: 17603471]
4. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–770. [PubMed: 18600261]
5. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. [PubMed: 16153702]
6. Loh YH, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nature Genet. 2006; 38:431–440. [PubMed: 16518401]
7. Ivanova N, et al. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006; 442:533–538. [PubMed: 16767105]
8. Kim J, Chu J, Shen X, Wang J, Orkin SH. An extended transcriptional network for pluripotency of embryonic stem cells. Cell. 2008; 132:1049–1061. [PubMed: 18358816]
9. Chickarmane V, Peterson C. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. PLoS One. 2008; 3:e3478. [PubMed: 18941526]
10. Ying QL, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453:519–523. [PubMed: 18497825]

11. Sampath P, et al. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. Cell Stem Cell. 2008; 2:448–460. [PubMed: 18462695]
12. Chang WY, Stanford WL. Translational control: a new dimension in embryonic stem cell network analysis. Cell Stem Cell. 2008; 2:410–412. [PubMed: 18462690]
13. Chambers I, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 2003; 113:643–655. [PubMed: 12787505]
14. Mitsui K, et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell. 2003; 113:631–642. [PubMed: 12787504]
15. Chambers I, et al. Nanog safeguards pluripotency and mediates germline development. Nature. 2007; 450:1230–1234. [PubMed: 18097409]
16. Unwin RD, et al. Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. Blood. 2006; 107:4687–4694. [PubMed: 16507774]
17. Spooncer E, et al. Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting. Mol Cell Proteomics. 2008; 7:573–581. [PubMed: 18083999]
18. van den Berg DL, et al. Estrogen-related receptor β interacts with Oct4 to positively regulate *Nanog* gene expression. Mol Cell Biol. 2008; 28:5986–5995. [PubMed: 18662995]
19. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to *Nanog*, *Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation. Nature. 2008; 455:1124–1128. [PubMed: 18806776]
20. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521–533. [PubMed: 18692474]
21. Bonifer C, Hoogenkamp M, Krysinska H, Tagoh H. How transcription factors program chromatin—lessons from studies of the regulation of myeloid-specific genes. Semin Immunol. 2008; 20:257–263. [PubMed: 18579409]
22. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]
23. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. Nature. 2007; 448:313–317. [PubMed: 17554338]
24. Wang J, et al. A protein interaction network for pluripotency of embryonic stem cells. Nature. 2006; 444:364–368. [PubMed: 17093407]
25. Macarthur BD, Ma'ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. Nature Rev Mol Cell Biol. 2009; 10:672–681. [PubMed: 19738627]
26. Graf T, Stadtfeld M. Heterogeneity of embryonic and adult stem cells. Cell Stem Cell. 2008; 3:480–483. [PubMed: 18983963]
27. Dietrich JE, Hiiragi T. Stochastic patterning in the mouse pre-implantation embryo. Development. 2007; 134:4219–4231. [PubMed: 17978007]
28. Singh AM, Hamazaki T, Hankowski KE, Terada N. A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells. 2007; 25:2534–2542. [PubMed: 17615266]
29. Kalmar T, et al. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol. 2009; 7:e1000149. [PubMed: 19582141]
30. Airoldi EM. Getting started in probabilistic graphical models. PLOS Comput Biol. 2007; 3:e252. [PubMed: 18069887]

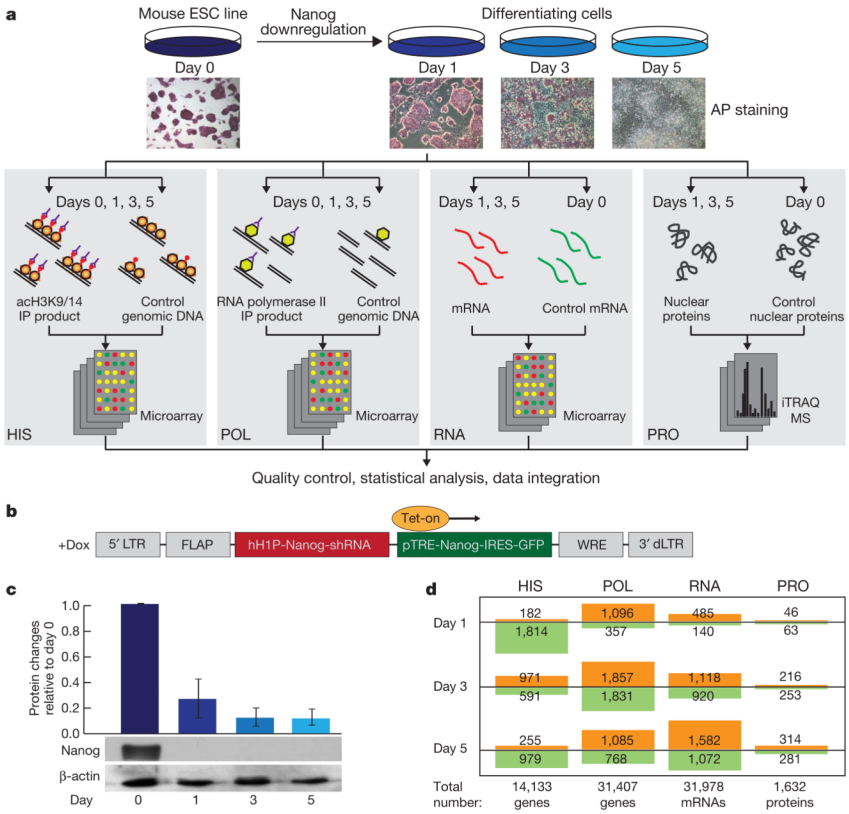

Figure 1. Measuring changes in the epigenome, the transcriptome and the nuclear proteome after Nanog downregulation

a, Experimental design. AP, alkaline phosphatase; IP, immunoprecipitation; iTRAQ, isobaric tag for relative and absolute quantification; MS, mass spectrometry. **b**, The lentiviral vector construct to conditionally regulate Nanog expression levels[7]. dLTR, deleted long-terminal repeat; FLAP, nucleotide segment that improves transduction efficiency; Teton, tetracycline transactivator; WRE, woodchuck hepatitis virus post-transcriptional regulatory element. **c**, Efficacy of Nanog protein downregulation as measured by mass spectrometry (bar chart) and western blot (image, bottom). Error bars denote the s.d. of duplicate measurements. **d**, Summary of the numbers of genes with significant changes at different molecular layers on each day. Increased and decreased levels are shown in orange and green, respectively.

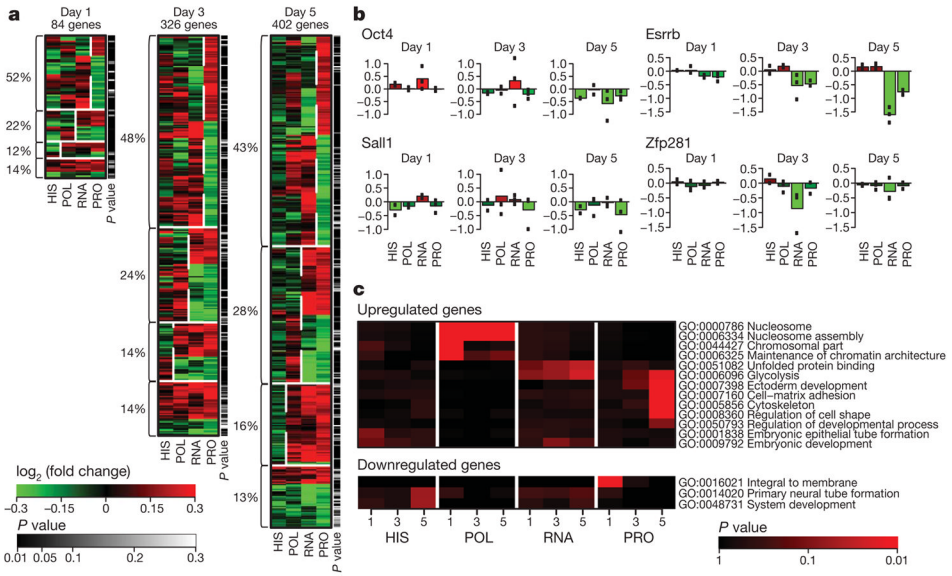

Figure 2. Comparisons across different molecular regulatory layers

a, Proteins with significant changes on each day are assigned to one of four categories on the basis of concordance between expression steps (Methods). The percentages on the left are calculated according to the number of proteins in each category. The *P*-value bar on the right gives the inclusion significance level. **b**, Examples of proteins from each of the four categories. Black dots represent the exact values for each experimental replicate. **c**, Selected gene-ontology (GO) categories that are overrepresented at each gene expression step. The complete panel is shown in Supplementary Fig. 5.

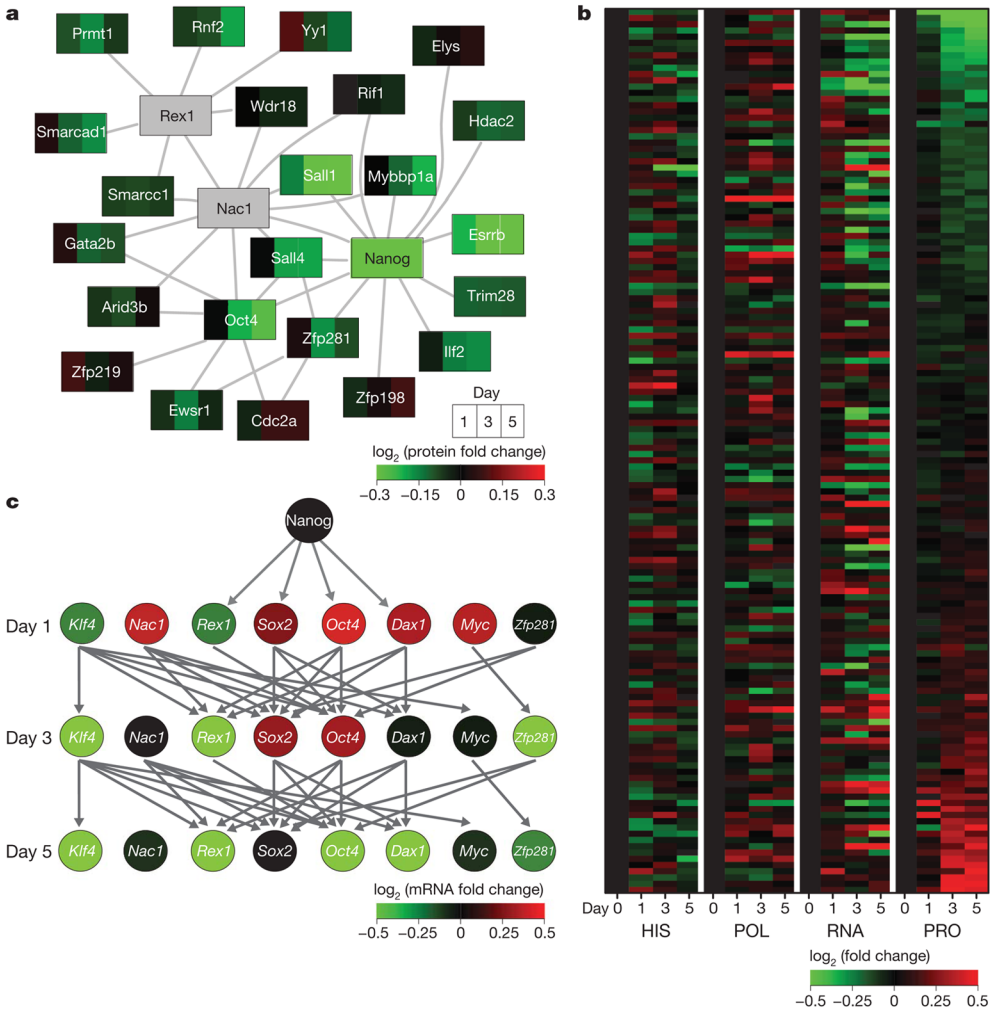

Figure 3. Dynamic changes in ESC networks

a, The core ESC protein–protein interaction network[24] (connections) overlaid with dynamic protein changes observed in our data (rectangles are divided into three segments representing changes on days 1, 3 and 5 compared to day 0). **b**, Heat map of multimolecular layer gene expression changes for Nanog-binding genes[6]. Shown are the genes whose data were obtained with high confidence on all four molecular layers. Genes are ranked on the basis of changes in protein levels. **c**, The pluripotency transcriptional regulatory network[8] (arrows) overlaid with mRNA fold changes (colours) from our data.

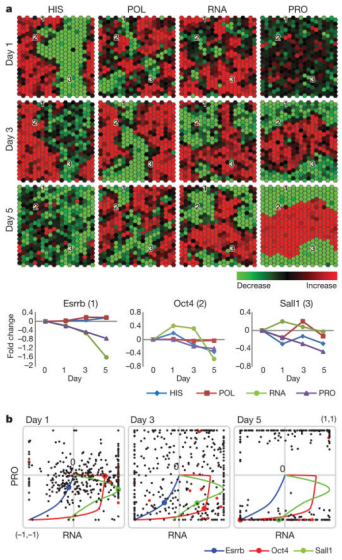

Figure 4. Interactive visualization of the multilayer dynamic data

a, Snapshots from heat map movies showing 400 genes with the most significant changes in protein levels on day 5. The position (pixel) of each gene locus is the same in all 12 heat maps. **b**, Snapshots from dynamic scatter plots illustrating concurrent changes in mRNAs and proteins. Red dots represent genes that have been identified to have important roles in ESCs[25]. Supplementary Fig. 8 and the website http://amp.pharm.mssm.edu/ronglu are interactive and each gene can be displayed as a line plot as exemplified by *Esrrb*, *Oct4* and *Sall1*.