LETTERS

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] & Ihor R. Lemischka[1,4]

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 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 multi-layer 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

[1]Department of Molecular Biology, [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, [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. †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.

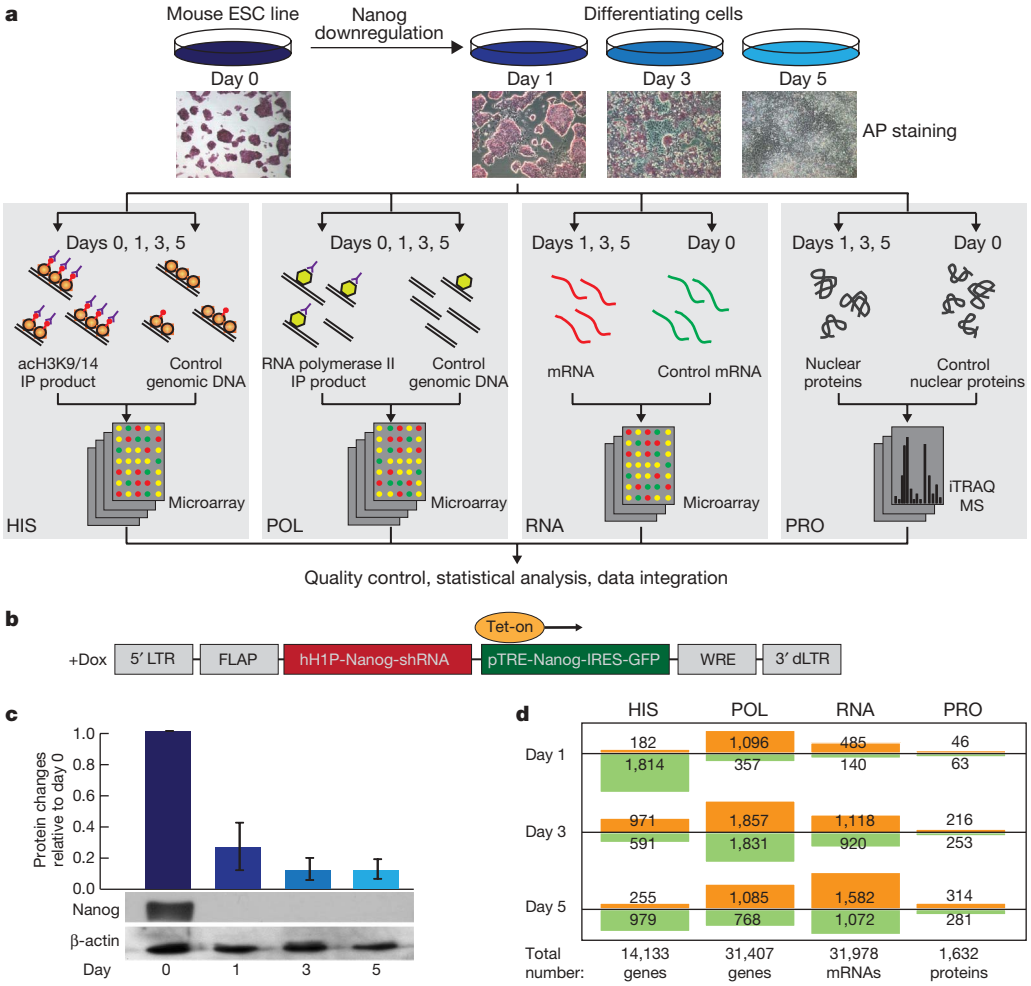

Figure 1 | Measuring changes in the epigenome, the transcriptome and the nuclear proteome after Nanog downregulation. **a**, Experimental design. AP, alkaline phosphatase; IP, immunoprecipitation; iTRAQ, isobaric tag for relative and absolute quantification; MS, mass spectrometry. **b**, The lentiviral vector construct to conditionally regulate Nanog expression levels[7]. dLTR, deleted long-terminal repeat; FLAP, nucleotide segment that improves transduction efficiency; Tet-on, 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.

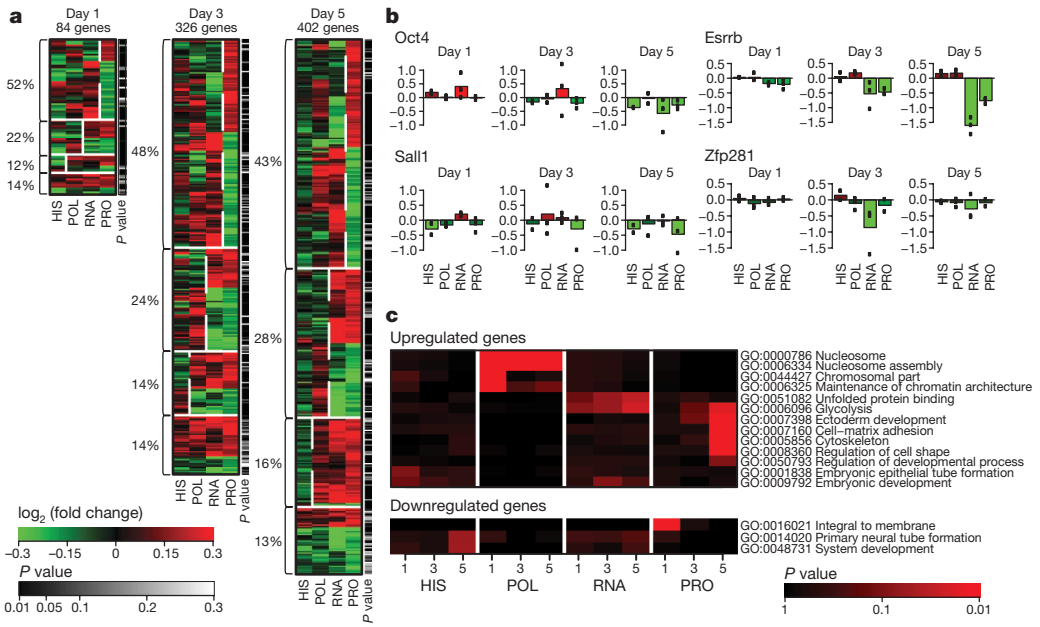

Figure 2 | Comparisons across different molecular regulatory layers. **a**, Proteins with significant changes on each day are assigned to one of four categories on the basis of concordance between expression steps (Methods). The percentages on the left are calculated according to the number of proteins 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.

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 overrepresented 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.

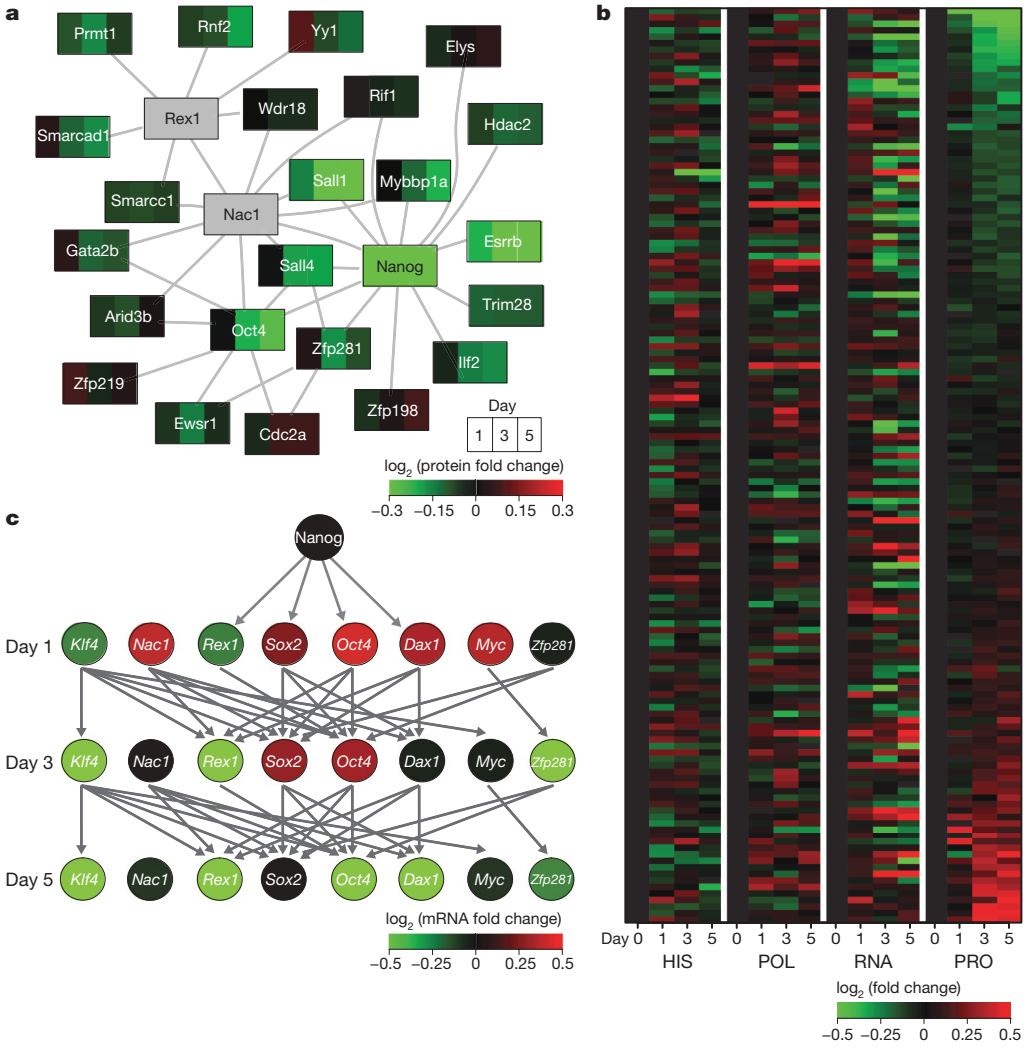

Figure 3 | Dynamic changes in ESC networks. **a**, The core ESC protein–protein interaction network[24] (connections) overlaid with dynamic protein changes observed in our data (rectangles are divided into three segments representing changes on days 1, 3 and 5 compared to day 0). **b**, Heat map of multimolecular 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.

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.

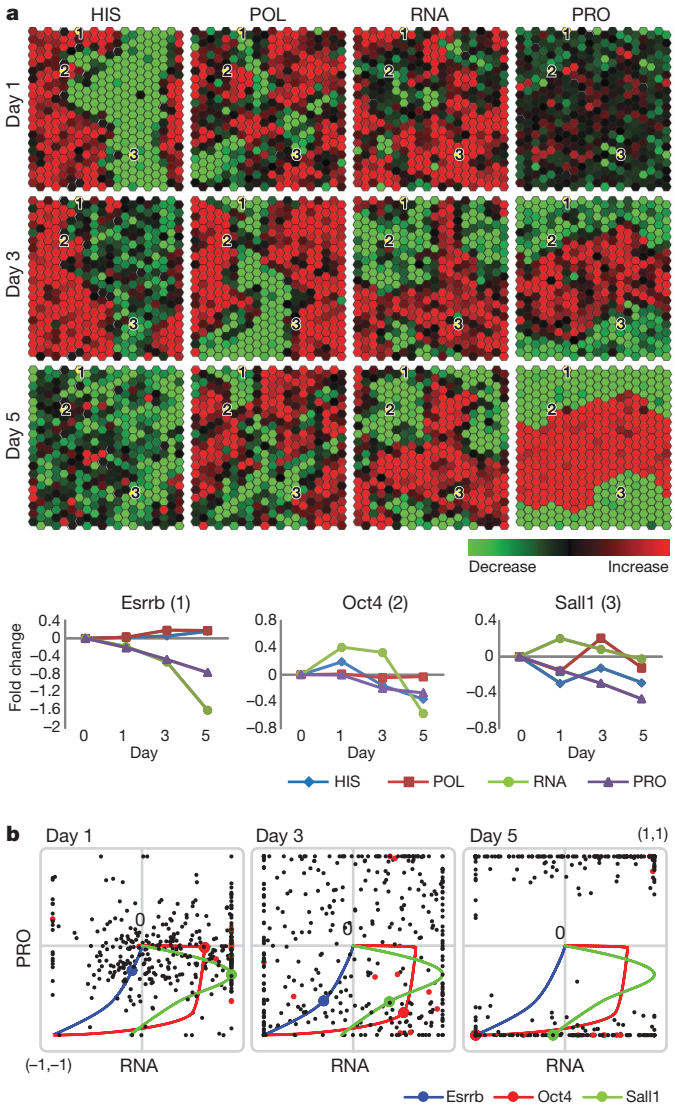

Figure 4 | Interactive visualization of the multilayer dynamic data. **a**, Snapshots from heat map movies showing 400 genes with the most significant changes in protein levels on day 5. The position (pixel) of each gene locus is the same in all 12 heat maps. **b**, Snapshots from dynamic scatter plots 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*.

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.

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

Received 12 February; accepted 9 October 2009.

1. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
2. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).
3. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
4. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
5. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
6. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38,** 431–440 (2006).
7. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442,** 533–538 (2006).
8. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
9. Chickarmane, V. & Peterson, C. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS One* **3,** e3478 (2008).
10. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
11. Sampath, P. *et al.* A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **2,** 448–460 (2008).
12. Chang, W. Y. & Stanford, W. L. Translational control: a new dimension in embryonic stem cell network analysis. *Cell Stem Cell* **2,** 410–412 (2008).
13. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643–655 (2003).
14. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
15. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
16. Unwin, R. D. *et al.* Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood* **107,** 4687–4694 (2006).
17. Spooncer, E. *et al.* Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting. *Mol. Cell. Proteomics* **7,** 573–581 (2008).
18. van den Berg, D. L. *et al.* Estrogen-related receptor β interacts with Oct4 to positively regulate *Nanog* gene expression. *Mol. Cell. Biol.* **28,** 5986–5995 (2008).
19. Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. MicroRNAs to *Nanog, Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation. *Nature* **455,** 1124–1128 (2008).
20. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).
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.* **20,** 257–263 (2008).
22. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
23. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
24. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444,** 364–368 (2006).
25. Macarthur, B. D., Ma'ayan, A. & Lemischka, I. R. Systems biology of stem cell fate and cellular reprogramming. *Nature Rev. Mol. Cell Biol.* **10,** 672–681 (2009).
26. Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3,** 480–483 (2008).
27. Dietrich, J. E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development* **134,** 4219–4231 (2007).
28. Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25,** 2534–2542 (2007).
29. Kalmar, T. *et al.* Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7,** e1000149 (2009).
30. Airoldi, E. M. Getting started in probabilistic graphical models. *PLOS Comput. Biol.* **3,** e252 (2007).

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

Acknowledgements 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.

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.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.R.L. (ihor.lemischka@mssm.edu) or R.L. (rlu@stanford.edu).

METHODS

Cell culture. A murine ESC line with controllable Nanog expression was constructed and characterized previously[7], and was cultured as described. ESCs were cultured without feeder cells (primary mouse embryonic fibroblasts) for all experiments. To induce differentiation, we withdrew doxycycline ($1\,\mu g\,ml^{-1}$) from the media, but still maintained all of the routine ESC nutrients (DMEM with 15% FBS (Hyclone), 100 mM MEM non-essential amino acids, 0.1 mM 2-mercaptoethanol, 1 mM L-glutamine (Invitrogen) and $10^3\,U\,ml^{-1}$ LIF (Chemicon)).

Days 1, 3 and 5 were selected because: (1) our previous studies[7], which investigated the differentiation process using microarrays and quantitative PCR analysis over the course of 12 days, suggested that days 1, 3 and 5 are sufficiently early such that no major differentiation events have yet occurred in the population, but are also sufficiently late and temporally spaced to study transitions from pluripotency and temporal differences. (2) Our preliminary proteomic experiment had shown that a reasonable number of proteins had changed during this time frame. In other words, on days 1, 3 and 5 the numbers of proteins that had changed were large enough to analyse using mass spectrometry, and were also small and distinct enough from each other to study the initial dynamic changes (Fig. 1d).

ChIP-chip. ChIPs were performed as described[5]. Fifty-million to five-hundred-million cells were fixed in a formaldehyde solution and sonicated into chromatin fragments containing 500–1,000 base pairs of DNA. ChIP was performed using 100 µl of a protein G magnetic bead suspension from Dynal coated by 10 µg of antibody (anti-acH3K9/14 (06-599) from Upstate; RNA polymerase II antibody (MMS-126R) from Covance). After reversal of the cross-links, the isolated DNA and non-ChIP-enriched control DNA were tailed with polyA by terminal transferase (TdT)[31]. T7 $(dT)_{24}$ primer was used to incorporate the T7 promoter during the second-strand synthesis reaction. The DNA fragments were then linearly amplified and labelled with Cy3 and Cy5 during the *in vitro* transcription, following the protocol provided by Agilent for dye incorporation and array hybridization (Agilent low RNA input fluorescent linear amplification kit; Agilent 60-mer oligo microarray processing protocol version 2.1). The histone acetylation ChIP-chip was performed by L. Boyer. The amplification step is slightly different[5].

Microarrays. The histone acetylation ChIP-chip used the mouse promoter array from Agilent, custom-designed by the R. Young laboratory. An Agilent whole mouse genome oligonucleotide microarray that covered 41,000 well-characterized mouse genes and transcripts was used for the mRNA assays and RNA polymerase II ChIP-chip experiments.

Nuclear proteome. Nuclear protein samples were prepared with the Nuclear/Cytosol fractionation kit (BioVision). Proteomic measurements were performed according to published protocols[16]. Samples from four different time points (day 0, and days 1, 3 and 5 after doxycycline removal) were labelled using four-channel isobaric tagging reagents (iTRAQ, Applied Biosystems) and analysed by strong cation-exchange fractionation followed by reverse-phase liquid chromatography on line to a QStar XL quadropole time-of-flight mass spectrometer. We used ProQUANT (Applied Biosystems) and ProGROUP (Applied Biosystems) to identify and quantify proteins. We checked our proteomic data with the proteomic data from a previous study[32]. Only 3.1% of the proteins that we considered to be well-reproduced nuclear proteins were not identified as nuclear proteins in their study.

Data confirmation (qPCR and western blot). ChIP-chip results for RNA polymerase II localization and histone acetylation were verified using a commercial ChIP kit (Upstate), followed by qPCR. RNA microarray data were verified by qPCR. The qPCR kit was obtained from Stratagene (Brilliant SYBR Green QPCR Master Mix). Proteomic data were confirmed using western blot. The verification experimental results are shown in Supplementary Figs 11 and 12. Antibodies used to perform western blot were: Oct4 antibody from BD; Nanog antibody from Cosmo Bio; Dnmt3b antibody from Abgent; p53 antibody from J. Flint; β-actin antibody from Santa Cruz; HSP 90 antibody from Upstate; Histone H1.0 antibody from Abcam; Utf1 antibody from A. Okuda.

Processing microarray data. Background correction was performed using a Normal+Exponential convolution model[33] that adjusts the foreground to the background and yields strictly positive intensities. Furthermore, we used an offset to dampen the variation of the log-ratios for very low intensities near 0. This stabilized our estimated fold changes. Arrays were normalized using a global loess, which is a well-tested general-purpose normalization method using local regressions to straighten the 'banana-shape' seen in raw measurements[34]. To confirm data quality, microarrays with remaining spatial (and other) artefacts were discarded and the experiments repeated.

Processing proteomic data. We used ProQUANT and ProGROUP software (Applied Biosystems) to analyse the mass spectrometric data, giving confidence values for the relative quantification analysis. Our proteomic analysis was based only on proteins that were identified with more than 95% confidence. We further filtered proteins based on two filters: (1) filter criteria based on raw data: the error factor of the measurement must be smaller than 2 and the protein must have been detected in at least two of the three runs. (2) Assessing reproducibility of protein measurements: we fitted a linear model (two-way ANOVA) to obtain temporal and replicate effects. If a significant replicate effect existed, we deemed the protein to be 'non-reproducible' and discarded it from further analysis.

Identification of histone acetylation regions. We compared the measurement for each probe on the promoter array against the distribution of measurements for all the negative control probes (null distribution), and calculated a *P* value for every probe (Supplementary Fig. 9). We use a FDR cut-off of 0.1 on the *P* value distribution to define which probes were acetylated and which were not. Supplementary Fig. 10 shows example acetylation profiles that indicate the acetylated regions and illustrate the main changes that occurred there.

Assessing experimental reproducibility and merging data. (1) Assessing reproducibility of probes: for every microarray probe, we fitted a linear model[30,35] (two-way ANOVA) to extract temporal and replicate effects. If a probe had a significant replicate effect, we deemed it to be non-reproducible and discarded it from further analysis. (2) Averaging probes that represent the same gene: for RNA polymerase II and mRNA expression data, we performed a three-way ANOVA with temporal, replicate and probe effects. Only genes with non-significant probe effects were used for further analysis (that is, those for which all probes behave coherently). For the histone acetylation data set, we averaged acetylated probes in a 1-kilobase window around the transcription start position (red lines in Supplementary Fig. 10 mark this region). (3) Combining gene isoforms: data from different molecular layers were merged based on our ID matching strategy (details later). For genes with more than one isoform, we applied a three-way ANOVA to determine temporal, replicate and gene effects. If the gene effect was significant (showing non-coherent behaviour), we discarded the data. The data for each gene in each data set at each time point were averaged if coherent behaviour was observed on both probe and gene levels.

ID matching. We matched protein IDs, microarray IDs, and MGI symbols (for GoMiner) using Ensembl BioMart (http://www.ensembl.org/Multi/martview), supplemented with protein information from the following databases: http://www.ebi.uniprot.org/uniprot-srv/uniProtEntryListSearch.do; http://www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi?db=Nucleotide; and http://www.pir.uniprot.org/search/idmapping.shtml. Histone acetylation ChIP-chip data were matched to the RNA microarray data using the UCSC database and the Ensembl database. The ID match file is included in the Supplementary Information.

Determining significance thresholds. For each of the four data sets, we computed the standard deviations of each gene using values from the replicate experiments. We then used the median value of the entire set of standard deviations in each data set as an estimate of the experimental error. For each of the four data sets, *P* values were independently computed using a Gaussian model for the measurements of each gene, under the null hypothesis given by setting the mean at zero, and the standard deviation at the experimental error estimate. Corrected *P* value was then obtained using the FDR correction[36]. Up- and downregulated genes were considered to be significant at a confidence level of $\alpha = 0.05$. An overview of the results is given in Supplementary Fig. 2, which shows the number of up- and downregulated genes in all data sets and for genes with protein data.

Methods for Fig. 2a. The method we used to generate Fig. 2a is not a clustering per se. Conventional clustering method is only applied at the very last step for visualization, but does not determine the categories. Our method is basically an iterative gene selection procedure, starting on the PRO level and working from there 'backwards' to RNA, POL and finally to HIS. The step-by-step description is as follows: (1) for each day, select all genes with significant protein changes. Genes without significant protein changes are discarded. (2) Select all genes that show a direction of change on the PRO level that is opposite to that on the RNA level. These genes form category 1. (3) Select all genes that show the same direction of change on PRO and RNA, but the opposite direction on POL. These genes form category 2. (4) Select all genes that show the same direction of change in PRO, RNA and POL, but the opposite direction in HIS. These genes form category 3. (5) All remaining genes show the same direction of change in all layers—PRO, RNA, POL and HIS. These genes form category 4. (6) Within each category, we cluster the genes with standard hierarchical clustering ('hclust' function in R) using complete linkages and a Euclidean distance. This clustering does not influence the definition of the four categories. It only improves the 'readability' of the resulting heatmap. Data are normalized within each column (molecular layer).

Methods for Fig. 4 and online movies. We selected the 400 genes with the most significant changes in protein expression on day 5. Because there are four time points (days 0, 1, 3 and 5), the data from each molecular layer is a 400×4 matrix. To consider correlations across layers, we first concatenated the time series from

all four layers into a 400 × 16 data matrix D. To visualize systems-level regulatory dynamics we then projected this data matrix onto a regular hexagonal array H by assigning each row of the data matrix to a unique hexagon h in H. A hexagonal array was chosen because it presents the data in a form that is easy to visualize. To provide a continuous geometric object with no boundaries we associated the left- and right-hand sides of the array with each other, and the top and the bottom of the array with each other (to make the surface of a torus). These conditions ensure that there are no special places on the array and all molecular species are treated equally.

Not all arrangements of the data on the array will capture the system-level regulatory dynamics equally well: most arrangements will not capture the collective dynamics because molecular species with similar expression patterns will not be close to each other on the array. To construct an arrangement that best captures collective dynamics we assigned to each arrangement a fitness

$$\text{Fit} = \frac{1}{2{,}400}\sum_{i=1}^{400}\sum_{j\in N_j} C_{ij},$$

in which C_{ij} is the Pearson's correlation coefficient between the time series i and j, and N_j are the six neighbours of the hexagon h_j. Fit measures how well a given arrangement captures the collective dynamics of the system in general: arrangements with low fitness do not capture system-level dynamics, whereas arrangements with high fitness capture system-level dynamics well. To find the arrangement of the time series on the array with the maximal fitness we used a simulated annealing algorithm, and ran the annealing algorithm overnight (12 h) to ensure as close to an optimal arrangement as possible.

Movies of systems-level dynamics were then generated by dynamically assigning colours to each of the hexagons in the array based upon the expression fold changes of the gene to which it is assigned. To create a movie that interpolates smoothly between time-points, each time series was normalized such that all expression series range from 0 to 1 and a piecewise cubic Hermite interpolation was implemented before visualization. Similar movies can be created using GATE (http://amp.pharm.mssm.edu/maayan-lab/gate.htm), a system we developed for this purpose.

We note here that the clustering technique we have used is similar to a self-organizing map (SOM), and the movies we create are similar to those created by the Gene Expression Dynamics Inspector (GEDI)[37] using SOMs. Given a set of time-series data describing expression changes in a large number of genes, the GEDI uses SOMs to project the expression time series onto a two-dimensional rectangular array, and colours rectangles according to the genes to which they are associated. However, because the GEDI uses a SOM, individual rectangles are associated with a cluster of genes that share similar expression patterns. In our study, we are concerned with the gene expression at different molecular layers. Thus it was important to track the molecular regulation of individual (rather than clusters of) genes. For this reason, we used the above custom-written algorithm that assigns molecular species to hexagons in a strictly one-to-one manner.

Code and software. Data pre-processing, data normalization and large parts of the analysis were performed in the computing languages Python and R (http://www.r-project.org/) using packages available from the Bioconductor website (http://www.bioconductor.org/). In particular, we relied on the limma package (http://bioinf.wehi.edu.au/limma/) including the Norm-Exp model for background correction as described previously[33]. To create Fig. 4, we used GATE (http://amp.pharm.mssm.edu/maayan-lab/gate.htm) and AS3/Flash. Our pre-processing and analysis pipeline is available from the authors on request.

31. Liu, C. L., Schreiber, S. L. & Bernstein, B. E. Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics* **4,** 19 (2003).
32. Kislinger, T. *et al.* Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125,** 173–186 (2006).
33. Ritchie, M. E. *et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23,** 2700–2707 (2007).
34. Yang, Y. H. *et al.* Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30,** e15 (2002).
35. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** Article3 (2004).
36. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125,** 279–284 (2001).
37. Eichler, G. S., Huang, S. & Ingber, D. E. Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* **19,** 2321–2322 (2003).