nature genetics

The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line

The FANTOM Consortium and the Riken Omics Science Center[1]

Using deep sequencing (deepCAGE), the FANTOM4 study measured the genome-wide dynamics of transcription-start-site usage in the human monocytic cell line THP-1 throughout a time course of growth arrest and differentiation. Modeling the expression dynamics in terms of predicted *cis*-regulatory sites, we identified the key transcription regulators, their time-dependent activities and target genes. Systematic siRNA knockdown of 52 transcription factors confirmed the roles of individual factors in the regulatory network. Our results indicate that cellular states are constrained by complex networks involving both positive and negative regulatory interactions among substantial numbers of transcription factors and that no single transcription factor is both necessary and sufficient to drive the differentiation process.

Development, organogenesis and homeostasis in multicellular systems involve the proliferation of precursor cells, followed by growth arrest and the acquisition of a differentiated cellular phenotype. Upon stimulation with phorbol myristate acetate (PMA), human THP-1 myelomonocytic leukemia cells cease proliferation, become adherent and differentiate into a mature monocyte- and macrophage-like phenotype[1,2]. This study aimed to understand the transcriptional network underlying growth arrest and differentiation in mammalian cells using THP-1 cells as a model system.

Most existing methods for regulatory network reconstruction collect genes into coexpressed clusters and associate these clusters with regulatory motifs or pathways (for example, see refs. 3–5). Alternatively, one can model the expression patterns of all genes explicitly in terms of predicted regulatory sites in promoters and the post-translational activities of their cognate transcription factors (TFs)[6–8]. Although this approach is challenging in complex eukaryotic genomes owing to large noncoding regions, ChIP-chip data[9] indicates that the highest density of regulatory sites is found near transcription start sites (TSSs) and regulatory regions originally thought to be distal may often be alternative promoters[10,11]. Precise identification of TSS locations is thus likely to be a crucial factor for accurate modeling of transcription regulatory dynamics in mammals.

In this study, we extend our previous observations of genome-wide TSS usage by Cap Analysis of Gene Expression (CAGE)[12] and using deep sequencing to identify promoters active during a time course of differentiation and quantify their expression dynamics. DeepCAGE data are used in combination with cDNA microarrays, other genome-scale approaches, novel computational methods and large-scale siRNA validation to provide a comprehensive analysis of growth arrest and differentiation in the THP-1 cell model.

RESULTS

Outline of the analysis strategy

In most cell line models, only a subset of cells undergoes growth arrest and differentiation. To maximize the sensitivity in this study, we identified a subclone of THP-1 cells in which the large majority of cells became adherent in response to PMA (**Supplementary Fig. 1** online). Our strategy began with deepCAGE, which identified active TSSs at single-base-pair resolution, and simultaneously measured their time-dependent expression (using normalized tag frequency) as cells differentiated in response to PMA. The same RNA was subjected to cDNA microarray analysis on an Illumina platform. The differentiation of the cells was evident from the large increase in expression of macrophage-specific genes such as *CD14* and *CSF1R* detected by both deepCAGE and microarray in all replicates (**Supplementary Fig. 2** online).

Figure 1 summarizes our Motif Activity Response Analysis (MARA) strategy. Promoters were defined as local clusters of coexpressed TSSs and promoter regions as their immediate flanking sequences (**Fig. 1a**,**b**). To reconstruct transcription regulatory dynamics we refined earlier computational methods[6–8] by incorporating comparative genomic information and each TF's positional preferences relative to the TSS in the prediction of regulatory sites. Binding sites for a comprehensive and unbiased collection of mammalian regulatory motifs were predicted in all proximal promoter regions (**Fig. 1c**) and the observed promoter expression profiles (**Fig. 1d**) were combined with the predicted site-counts (**Fig. 1e**) to infer time-dependent activity profiles of regulatory motifs (**Fig. 1f**). We inferred individual regulatory interactions (edges) between motifs and promoters by comparing the promoter expression and motif activity profiles (**Fig. 1g**). Rigorous Bayesian probabilistic methods were developed for all steps of the computational analysis. Finally, a core network was

[1]A full list of authors and affiliations is provided at the end of this paper.

Received 16 July 2008; accepted 25 March 2009; published online 19 April 2009; doi:10.1038/ng.375

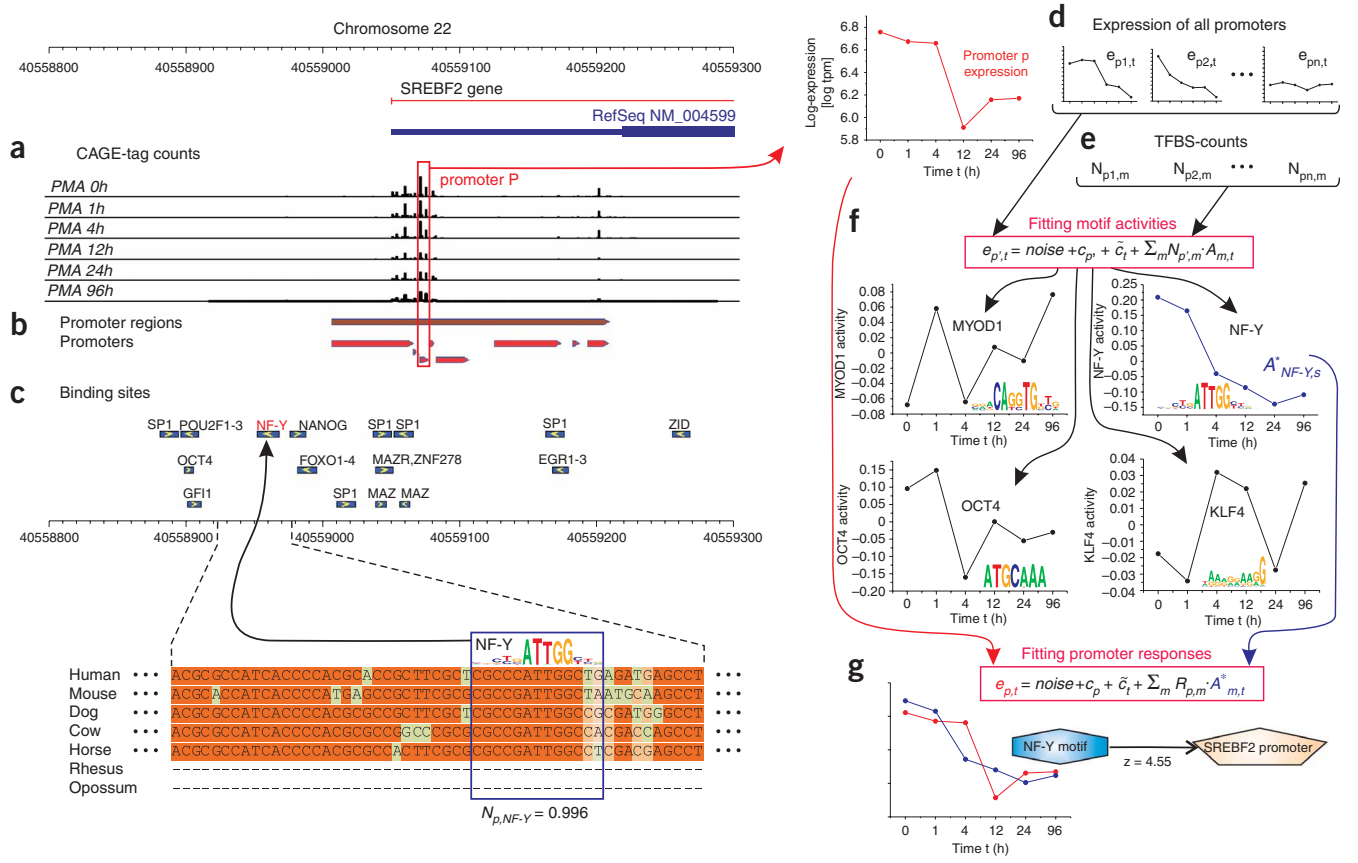

Figure 1 Motif Activity Response Analysis (MARA). (**a**) CAGE tags are mapped to the human genome and their expression is normalized; vertical lines represent TSS positions, and their height is proportional to the normalized expression. (**b**) Mapped tags are clustered into promoters on the basis of their relative expression, and neighboring promoters are joined into promoter regions. (**c**) A window of −300 to +100 flanking each promoter region is extracted, multiply aligned and the MotEvo algorithm is used to predict binding sites for known motifs. (**d–f**) Observed expression of all promoters (**d**) and predicted site-counts (**e**) are used to infer motif activities (**f**). (**g**) The statistical significance of the regulatory edge from motif to promoter is calculated based on correlation of the promoter expression and motif activity profiles.

constructed by selecting the motifs that explained the greatest proportion of the expression variance, obtaining all predicted regulatory edges between TFs corresponding to these motifs and selecting those regulatory edges that had independent experimental support. Using this approach, we reconstructed the transcriptional regulatory dynamics associated with cellular differentiation in human THP-1 cells, and validated a subset of predicted regulatory interactions.

DeepCAGE quantification of dynamic TSS usage

CAGE tags generated from mRNA harvested at each time point were mapped to the human genome. Promoters were defined as clusters of nearby TSSs that showed identical expression profiles (within measurement noise) and were substantially expressed in at least one time point (**Fig. 1a**,**b**). Using these criteria we identified 29,857 promoters expressed in THP-1 cells containing 381,145 unique TSS positions (which is a subset of the nearly 2 million TSSs detected at least once in THP-1). These promoters were contained within 14,607 promoter regions (separated by at least 400 bp; Methods and **Supplementary Fig. 3** online). The deepCAGE data was validated using genome tiling-array ChIP for markers of active transcription. Of the promoters identified, 79% and 78% were associated with H3K9Ac and RNA polymerase II, respectively (both markers of active transcription[13,14]), compared to 18% and 27% for inactive promoters (**Supplementary Note** online).

Among the identified promoters 84% (24,984) were within 1 kb of the starts of known transcripts and 81% (24,327) could be associated with 9,452 Entrez genes. Approximately half of the remaining promoters were more than 1 kb away from the loci of known genes (**Supplementary Fig. 4** online). These newly identified promoters are conserved across mammals, suggesting that they are true transcription starts of currently unknown transcripts (**Supplementary Fig. 5** online). The association of 24,327 promoters with 9,452 Entrez genes extends previous evidence of alternative promoter usage[11]—in this case even within a single cell type (**Supplementary Table 1** online)—and demonstrates that promoter regions frequently contain multiple promoters with distinguishable expression profiles (**Supplementary Table 2** online). In addition, for genes with known multiple promoters deepCAGE frequently identified only one promoter to be active in the THP-1 samples (**Supplementary Fig. 6** online). Hence, deepCAGE samples a distinct aspect of transcriptional activity that can and does vary independently of mRNA abundances as measured by hybridization to representative microarray probes.

Promoter expression

Using the normalized tags per million (tpm) counts assigned to the promoters, we tested reproducibility among the three biological replicates and compared the outcome to the Illumina array from the same samples (**Supplementary Fig. 7** online). DeepCAGE

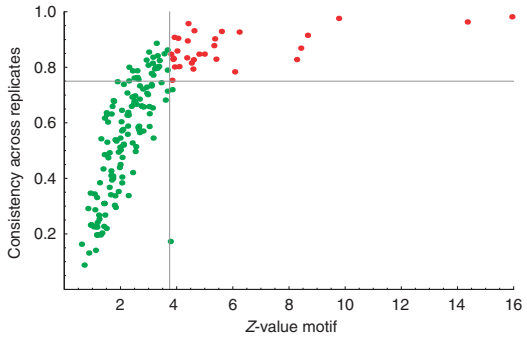

Figure 2 Statistical significance and consistency across replicates of the inferred motif activity profiles. Each dot corresponds to a motif. The significance of each motif in explaining the observed expression variation is quantified by the *z* value of its activity profile (horizontal axis, see Methods). The consistency of the inferred activity profile of each motif is quantified by the fraction of the variance (FOV) in the activity profile across all six replicates (three biological replicates for both CAGE and Illumina), which is reproduced in each replicate (vertical axis, see Methods).

expression measurements were comparatively noisy (**Supplementary Fig. 7a**). Nevertheless, the median Pearson correlation between the replicate-averaged expression profiles of CAGE and microarray was around 0.72 (**Supplementary Fig. 7b**), which is comparable to that observed with other deep transcriptome sequencing datasets[15]. As predicted, the correlation is lower for genes with multiple promoter regions (**Supplementary Fig. 7b** and discussed further in **Supplementary Note**).

Comprehensive regulatory site prediction

Known binding sites from the JASPAR and TRANSFAC databases[16,17] were used to construct a set of 201 regulatory motifs (position-specific weight matrices, WMs), which represent the DNA binding specificities of 342 human TFs. We predicted transcription factor binding sites (TFBSs) for all motifs within the proximal promoter regions (−300 to +100 bps) of all CAGE-defined promoters. Extending the proximal promoter regions beyond the −300 to +100 window decreased the quality of the fitted model described below (data not shown). In contrast to previous approaches that used simple WM scanning[6], we incorporated information from orthologous sequences in six other mammals and used a Bayesian regulatory-site prediction algorithm that uses explicit models for the evolution of regulatory sites[18,19] (**Fig. 1c** and Methods). Notably, different motifs had distinct and highly specific positional preferences with respect to TSS (**Supplementary Fig. 8** online), extending a previous genome-scale analysis[20]. Positional preferences were incorporated in the TFBS prediction by assigning each site a probability that it is under selection and correctly positioned. This analysis generated approximately 245,000 predicted TFBSs for the 201 motifs genome-wide. For each promoter–motif combination, the TFBS prediction was summarized by a count N_{pm}, which represents the estimated total number of functional TFBSs for motif m in promoter p. The TFBS predictions were compared with published high-throughput protein–DNA interaction datasets (ChIP-chip) and predicted target genes were significantly (P values ranged from 0.02 for *ETS1* to 6.60E–263 for *GABPA*) enriched among genes for which binding was observed (**Supplementary Table 3** online).

Inferring key TFs and their time-dependent activities

The details of our Motif Activity Response Analysis (MARA) are described in Methods. Briefly, for each motif m and each time point t, there is an (unknown) motif activity A_{mt}, which represents the time-dependent nuclear activity of positive and negative regulatory factors that bind to the sites of the motif (for example, the E2F activity will depend on nuclear E2F1-8, and DP1-2 levels, as well as RB1 phosphorylation status). As in previous work[6–8,21], motif activities were inferred by assuming that the expression e_{pt} of promoter p at time t is a linear function of the activities A_{mt} of those motifs that have predicted sites in p. Additionally, the effect of motif m on the expression of promoter p is assumed to be proportional to the predicted number of functional sites N_{pm}. Assuming that the deviations of the predicted expression levels $e_{pt}^{\text{theo}} = \text{constant} + \sum_m N_{pm}A_{mt}$ from the observed levels e_{pt} are Gaussian distributed, and using a Gaussian prior on the activities, we determine fitted activities A_{mt}^{*} that have maximal posterior probability (Methods).

The inferred motif activities were validated using a number of internal tests. First, our Bayesian procedure quantifies both the significance of each motif in explaining the observed expression variation as well as the reproducibility of its activity across replicates (**Fig. 2** and **Supplementary Table 4** online). The activity profiles of the top motifs are extremely reproducible across replicates and different measurement technologies (**Figs. 2** and **3a** and **Supplementary Fig. 9** online). It should be stressed that, although motif activities are inferred by fitting the expression profiles of all promoters, the model cannot be expected to predict expression profiles of individual genes from the predicted TFBS in proximal promoters alone. The effects of chromatin structure, distal regulatory sites, nonlinear interactions between regulatory sites, and the contribution of the large numbers of human TFs for which no motif is known, are not considered. Furthermore, especially for genes that are dynamically regulated, mature mRNA abundance can be dynamically regulated independently of transcription initiation and promoter activity through selective mRNA elongation, processing and degradation. Our aim is not to predict expression profiles of individual genes but rather to predict the key regulators and their time-dependent activities, which can be inferred from integration of global expression information in a system undergoing dynamic change. We validated the significance of the inferred activity profiles by comparing the fraction of the 'expression signal' (expression variance minus replicate noise) that is explained by the model, compared to randomized versions, and under a tenfold cross-validation test (**Supplementary Fig. 10** online). The explained expression signal is highly significant and this significance is maintained under tenfold cross-validation (Methods). In addition, the highly peaked positional profiles of TFBSs (**Supplementary Fig. 8**) suggest that knowing the exact TSS is important for accurate TFBS prediction. Indeed, the predicted TFBSs from CAGE promoters explain substantially more of the expression signal in microarrays than predicted TFBSs of the associated RefSeq promoters (**Supplementary Fig. 10**). We observe that the model better predicts the expression profiles of those promoters that are more strongly expressed, more reproducible across replicates, and have higher expression variance (**Supplementary Fig. 11** online). Similarly, samples at the start and end of the differentiation time course are better predicted than those at intermediate time points (**Supplementary Fig. 12** online), possibly because individual cells differentiate at different rates and leave the cell populations less homogeneous at intermediate time points.

Motif activities that were independently inferred from all 11,995 expressed microarray probes were combined with the inferred motif activities from all CAGE and microarray replicates into a final set of time-dependent motif activities (Methods). From these, we selected 30 'core' motifs that contribute most to explaining the expression

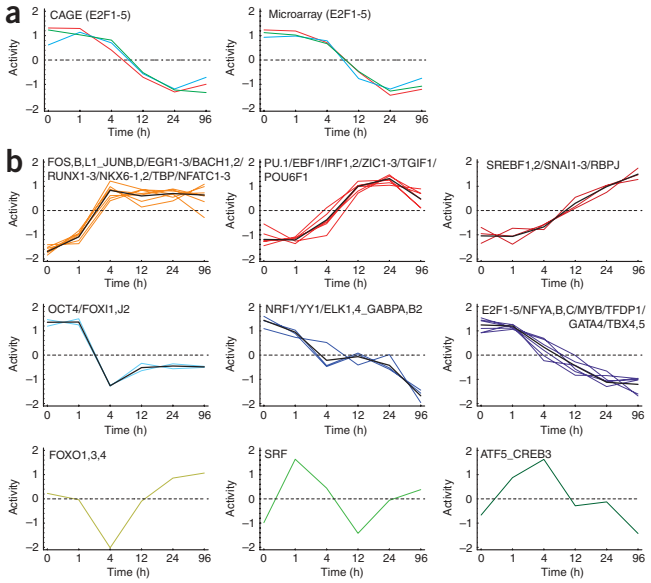

Figure 3 Inferred time-dependent activities of the key regulatory motifs. (**a**) The time-dependent activity profile of the E2F1-5 regulatory motif as inferred from CAGE (left) and microarray (right) data. The three biological replicates are shown in red, blue and green. (**b**) The 30 most significant motifs with consistent activity profiles across all replicates (CAGE and microarray) were clustered into nine sets of motifs with similar dynamics. Each panel shows the activity of the members of the cluster (colored curves), the names of motifs contributing and the cluster average activity profile (black).

variation (red dots in **Fig. 2**) and segregated their activity profiles using a Bayesian procedure into nine clusters (**Fig. 3b** and Methods), including three clusters of upregulated motifs, three clusters of downregulated motifs and three clusters containing single motifs with profiles involving different transient dynamics. The genome-wide set of target promoters for each of the motifs was determined as described in Methods. The significance of each regulatory 'edge' from a motif to a putative target promoter (containing a predicted TFBS) was quantified by the z value of the correlation between the motif's activity profile and the promoter's expression profile (**Fig. 1e**).

Core transcriptional regulatory network

The final aim in reconstructing transcriptional regulatory networks is to infer not only the key regulators and their target gene sets, but also the way in which the actions of these key regulators are coordinated. For this purpose, we collected all 199 predicted regulatory edges (z value ≥ 1.5) between the 30 core motifs. Recognizing that the prediction of individual regulatory edges is still prone to error, we constructed a core regulatory network (**Fig. 4**) of 55 highly trusted edges by filtering the predicted edges according to experimental validation, either within our data or in existing literature (**Supplementary Table 5** online). In addition, for each core motif we extracted the set of predicted target genes (z value ≥ 1.5) and checked for enrichment of gene ontology terms. A selection of significantly enriched terms is shown as oval nodes in **Figure 4** (full set of GO enrichments are available as **Supplementary Table 6** online).

Whereas our method infers the key regulators *ab initio*, the majority of factors within this core network are known to be important in the monocyte-macrophage lineage, thereby validating the method. In addition the predicted targets of these motifs are enriched for biological processes known to be involved in differentiation of the monocytic lineage.

The gene ontology enrichments can broadly be divided into four groups. Downregulated motifs E2F1-5, NFYA,B,C and MYB are associated with cell cycle–related terms, consistent with the growth arrest observed during PMA-induced differentiation and the specific downregulation of numerous genes required for DNA synthesis and cell cycle progression within 24 h of PMA addition. Notably, MYB targets are also enriched specifically for microtubule-cytoskeleton–associated genes. Conversely, targets of upregulated motifs are associated with the terms immune response, cell adhesion, plasma membrane, vacuole and lysosome, all of which are consistent with differentiation into an adherent monocyte-like cell. The targeting of lysosomal genes by cholesterol-regulated SREBFs (sterol regulatory element-binding transcription factors) is of note, as lipid homeostasis is important in the macrophage in atherosclerosis and lysosomal storage diseases[22]. We also saw enrichment of signal transduction genes among targets of the early induced motifs EGR1-3 and TBP. Finally, there is a set of motifs whose targets are enriched in TFs. These motifs correspond to the transiently induced/repressed motifs, ATF5_CREB3, FOXO1,3,4 and SRF, and the repressed pair of OCT4 and FOXI1,J2 motifs.

Validation of edge predictions

THP-1 cells, even in an 'undifferentiated' state, are clearly a myeloid cell line. In seeking to validate the transcriptional network, we noted that there was a large set of TF genes expressed constitutively in the cells that were rapidly downregulated in response to PMA, of which *MYB* is an example, and another set that was expressed but further upregulated during differentiation. It is technically difficult to apply siRNA knockdown to genes that are only expressed later in the differentiation. To validate predicted edges empirically, we therefore chose to carry out siRNA knockdowns in undifferentiated THP-1 cells for genes encoding 28 TFs that are expressed in the undifferentiated state and for which we have associated motifs. To assess whether siRNA knockdown carried out in the undifferentiated state is appropriate to address factors that increase expression during the time course, we carried out the technically more difficult experiment of siRNA knockdown combined with PMA treatment for *SPI1* (more commonly known in the literature as *PU.1*). All knockdowns were carried out in biological triplicate and qRT-PCR was used to confirm RNA-level knockdown, which in most cases was greater than 80% (**Supplementary Table 7** online; in addition, protein-level knockdown was confirmed by protein blot for 14 siRNAs, see **Supplementary Fig. 13** online). Changes in gene expression caused by TF knockdown were measured by Illumina microarrays. For each knocked-down TF gene, we obtained the list of predicted regulatory targets for the associated motif and divided the microarray probes into predicted targets and nontargets for a range of z-value thresholds. Higher-confidence targets in general show greater expression changes upon knockdown (**Fig. 5a** shows the example TF genes *MYB*, *SNAI3*, *EGR1* and *RUNX1*; additional examples are shown in **Supplementary Fig. 14** online). For *SPI1*, even in the absence of PMA treatment siRNA knockdown caused significant downregulation of predicted SPI1 targets, but the effects were much stronger when knockdown was combined with 1 h or 24 h of PMA treatment (**Fig. 5b**), confirming that PMA causes upregulation of *SPI1* activity. A good correlation between target confidence (z-value cut-off) and average log expression ratio was observed for the large majority of experiments (**Fig. 5c**). For an intermediate cut-off of $z = 1.5$ we quantified the difference in log expression ratio of predicted targets and nontargets (**Fig. 5d**) and

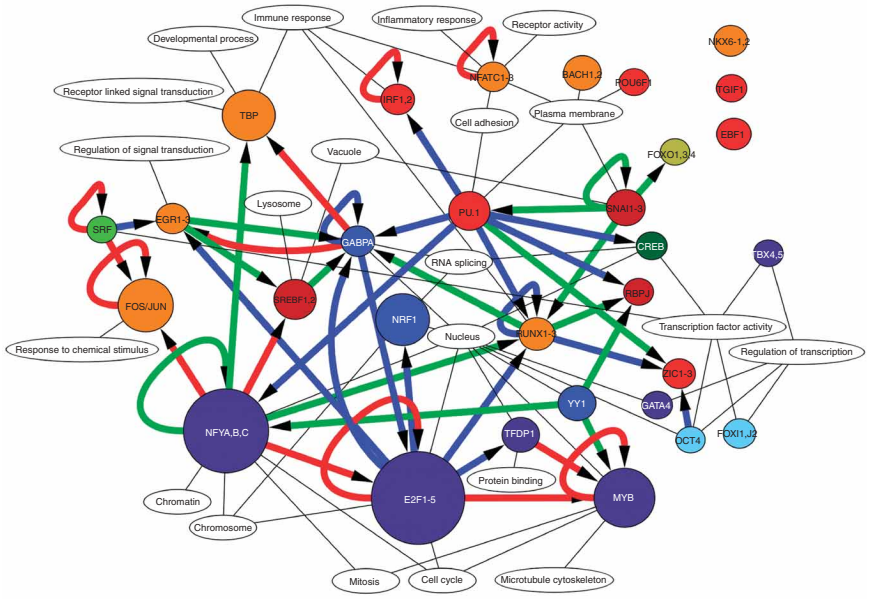

Figure 4 Predicted core regulatory network of the 30 core motifs. An edge X→Y is drawn whenever the promoter of at least one of the TFs associated with motif Y has a predicted regulatory edge for motif X (z value ≥ 1.5) and the edge has independent experimental support. The color of each node reflects its cluster membership and the size of the node reflects the significance of the motif. Edges confirmed in the literature, by ChIP or by siRNA are shown in red, blue and green, respectively. In cases where there are multiple lines of support only one evidence type is shown. **Supplementary Table 5** shows all predicted edges and their experimental support. GO terms significantly enriched among target genes are shown as white nodes with black edges. FOS/JUN (FOS,B,L1_JUNB,D), CREB (ATF5_CREB3), GABPA (ELK1,4_GABPA,B2).

found significant changes (z-value larger than 2) for 23 of 33 cases with *SPI1* knockdown combined with 24 h of PMA treatment and *MYB* knockdown being the most significant (**Supplementary Fig. 15** online shows the entire distribution of log expression ratios of targets and nontargets for eight example TFs). Notably, for the TF genes *LMO2*, *MXI1* and *SP1*, the knockdown led to a significant upregulation of their targets, suggesting that the three encoded TFs act primarily as repressors in undifferentiated THP-1 cells (**Fig. 5d**, also see **Supplementary Fig. 14a**). Together these results provide compelling experimental validation of our predicted regulatory edges.

Single TF knockdowns affect multiple motif activities

Besides validating predicted targets, the siRNA knockdowns can also be used to assess the effects of the knockdown of one TF gene on the motif activities of other TFs. In addition to the 28 TFs perturbed above, we included a further 24 TFs that lacked motifs but were naturally repressed during PMA differentiation, or had been reported to have a role in myeloid differentiation or leukemia (**Supplementary Table 8** online).

The motif activity inference method was used to determine the changes in activities of all motifs upon knockdown of each TF gene. To assess the role of each TF in differentiation, we defined the differentiative overlap between a TF gene knockdown and the PMA time course as the fraction of all motifs that significantly changed their activity in the same direction upon TF gene knockdown as in the PMA differentiation (Methods). By far the largest differentiative overlap (69%) was observed for the *MYB* knockdown, which not only affected MYB motif activity, but also the activity of most motifs in the core network, with the most significant activity changes all in the same direction as in the PMA time course (**Fig. 6a**). Knockdown of 13 other TF genes generated an overlap greater than the negative control (**Supplementary Table 9** online), and **Figure 6** shows three further examples (*E2F1*, *HOXA9* and *CEBPG*).

As for *MYB*, *E2F1* knockdown reproduced some of the downregulation of MYB and E2F activity observed upon PMA stimulation, but it failed to reproduce the upregulation of SREBF1,2, PU.1, NFATC1-3 and FOS,B,L1_JUNB,D activity (**Fig. 6b**). Similarly, the activity changes that *HOXA9* knockdown induced were mostly in the same direction as in the PMA differentiation; however, the SNAI1-3 and IRF1,2 motif activities failed to be induced and the GATA4 and TBX4,5 motif activities failed to be downregulated (**Fig. 6c**). Notably, knockdown of *CEBPG*, encoding one of the PMA-downregulated factors, for which we do not have a motif, also generated activity changes that significantly overlapped those observed in response to PMA (**Fig. 6d**). Finally, instead of comparing the motif activity changes that different knockdowns induced, we can also directly compare the expression changes of all genes with the expression changes observed in the PMA time course. We found that *MYB*, *HOXA9*, *CEBPG*, *GFI1*, *CEBPA*, *FLI1* and *MLLT3* knockdowns all generated changes in gene expression that reiterated some of those observed with PMA treatment (**Supplementary Table 8**). *MYB* knockdown was exceptional, as it induced 35% (340/967) and repressed 19% (172/916) of the genes upregulated and downregulated with PMA, respectively. In addition the cells became adherent (**Supplementary Fig. 16** online) and began to express the monocytic markers CD11B (*ITGAM*), CD54 (*ICAM1*), *CD14*, *APOE* and *CSF1R* (**Supplementary Fig. 2**), three of which we confirmed by flow cytometry (**Supplementary Table 10** online). This development of adherence could be linked to the GO enrichment for cytoskeleton-associated genes among *MYB* targets noted above. Given these observations one might wonder whether MYB is a master regulator of the differentiation process and whether stronger and longer knockdown would have reproduced the complete differentiation observed under PMA treatment. Several observations argue strongly against this. First, the gene sets perturbed by MYB and by the other pro-differentiative TFs overlap only partially (**Supplementary Table 11** online). Second, of the six other pro-differentiative TF genes only two (*CEBPG* and *GFI1*) are affected by *MYB* knockdown. Both these facts indicate that the other pro-differentiative TF genes are not simply downstream of *MYB*. Third, *MYB* downregulation does not occur until after the second hour of the PMA time course (**Fig. 3b**), which is at odds with the idea of *MYB* sitting at the top of the regulatory hierarchy. It is also worth noting that THP-1 cells harbor a leukemogenic fusion[23] between *MLL* (mixed-lineage leukemia) and *MLLT3* (MLL translocation partner 3) and that the *MLLT3* siRNA targets this leukemogenic fusion (note that full-length *MLLT3* does not seem to be expressed in THP-1 as there is no CAGE 5′ signal for this gene). Our data indicate that this fusion interferes with differentiation and that neither PMA treatment nor *MYB* knockdown affects *MLL-MLLT3* levels, suggesting these stimuli can bypass the differentiative block. Conversely, *MLLT3* knockdown had no effect on *MYB* levels. These

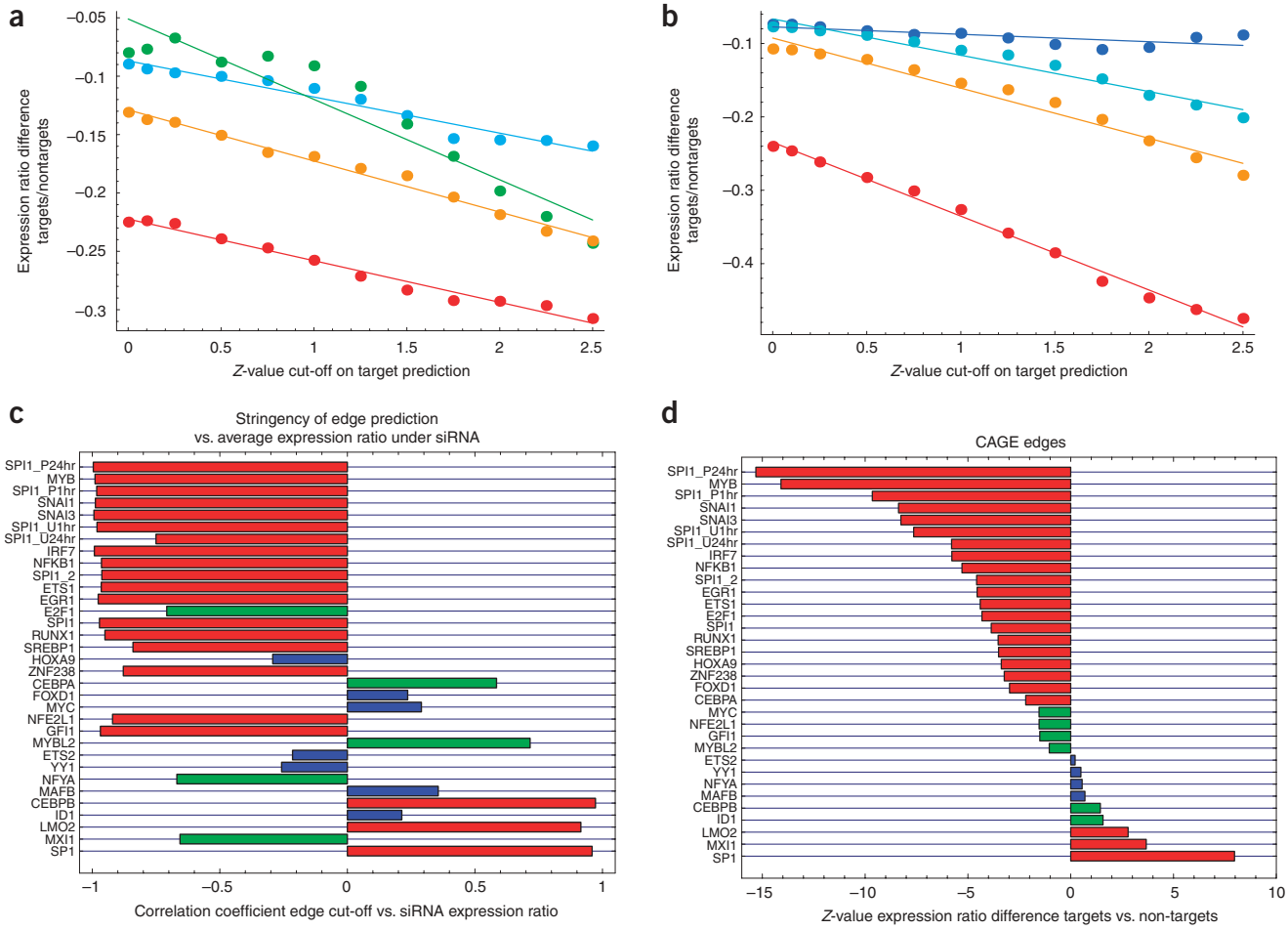

Figure 5 Validation of predicted target promoter sets using siRNA knockdowns. (**a**) Difference in the average log expression ratio upon knockdown between predicted target promoters and predicted nontargets (vertical axis) as a function of the *z*-value cut-off on target prediction (horizontal axis, more stringent cut-offs are on the right) for knockdown of the TF genes *MYB* (red), *SNAI3* (orange), *RUNX1* (green) and *EGR1* (light blue). (**b**) As in **a** but now for knockdown of *SPI1* followed by 1 h without treatment (light blue), 24 h without treatment (dark blue), 1 h of PMA treatment (orange) and 24 h of PMA treatment (red). All straight lines are linear regression fits. (**c**) Pearson correlation coefficients between the average log expression ratio difference of targets and nontargets and the cut-off on target predictions (horizontal axis). Red bars indicate correlation coefficients larger than 0.75 in absolute value; green bars, absolute values between 0.5 and 0.75; and blue bars, less than 0.5. (**d**) Significance (*z* value) of the difference in log expression ratio between predicted targets and nontargets (cut-off $z = 1.5$) for all 28 TFs associated with a motif, measured as a *z* value (number of standard errors). Red bars correspond to significant changes, that is, greater than two standard errors; green bars, changes between 1 and 2 standard errors; and blue bars, changes less than 1 standard error. siRNA knockdowns were carried out in biological triplicate and knockdown was assessed by qRT-PCR (**Supplementary Table 7**).

results agree with previous RNAi studies that conclude that down-regulation of *MLL* leukemogenic fusion proteins can promote growth arrest but is not required for terminal differentiation[24,25]. Thus, individual TF gene knockdowns affect the activities of multiple motifs and elicit different, but overlapping, subsets of the regulatory changes observed in the PMA time course. Taken together, the data indicate that the independent perturbation of expression of multiple TFs in response to PMA is both necessary and sufficient to initiate partial differentiation.

Many TFs are involved in the differentiation process

The network predictions and the siRNA results above suggest that upregulation and downregulation of the activities of multiple co-operating TFs is required for differentiation. Of a curated list[26] of 1,322 human TFs, 610 were detected by both CAGE and microarray in at least one time point (**Supplementary Table 12** online); however, only 155 of these are covered by weight matrices, suggesting that other factors may well be important in these cells. Of the 610 expressed TFs 64 were most highly expressed in the undifferentiated and 34 in the differentiated state. In addition, 101 TFs were transiently induced or repressed during differentiation. To elucidate the connection of these TFs to the inferred network, we compared the predicted regulatory inputs of co-regulated subsets of TFs with the predicted regulatory inputs of the set of all 610 expressed TFs.

Whereas no motifs are overrepresented among inputs of statically expressed TFs, inputs of dynamically expressed TFs showed enrichment for a subset of motifs. TFs downregulated from 0 to 96 h PMA were most enriched for three downregulated motifs of the core network: OCT4 (3.4×), GATA4 (3.3×) and NFYA,B,C (2.2×) (**Supplementary Table 13a** online). Similarly, TFs upregulated from 0 to 96 h were most enriched for core network motifs that increase activity during differentiation: SNAI1-3 (4.6×) and TBP (5.2×) (**Supplementary Table 13b**). Finally, transiently regulated TFs were enriched for the SRF (3.5×) and NHLH1,2 (3×) motifs (**Supplementary Table 13c**).

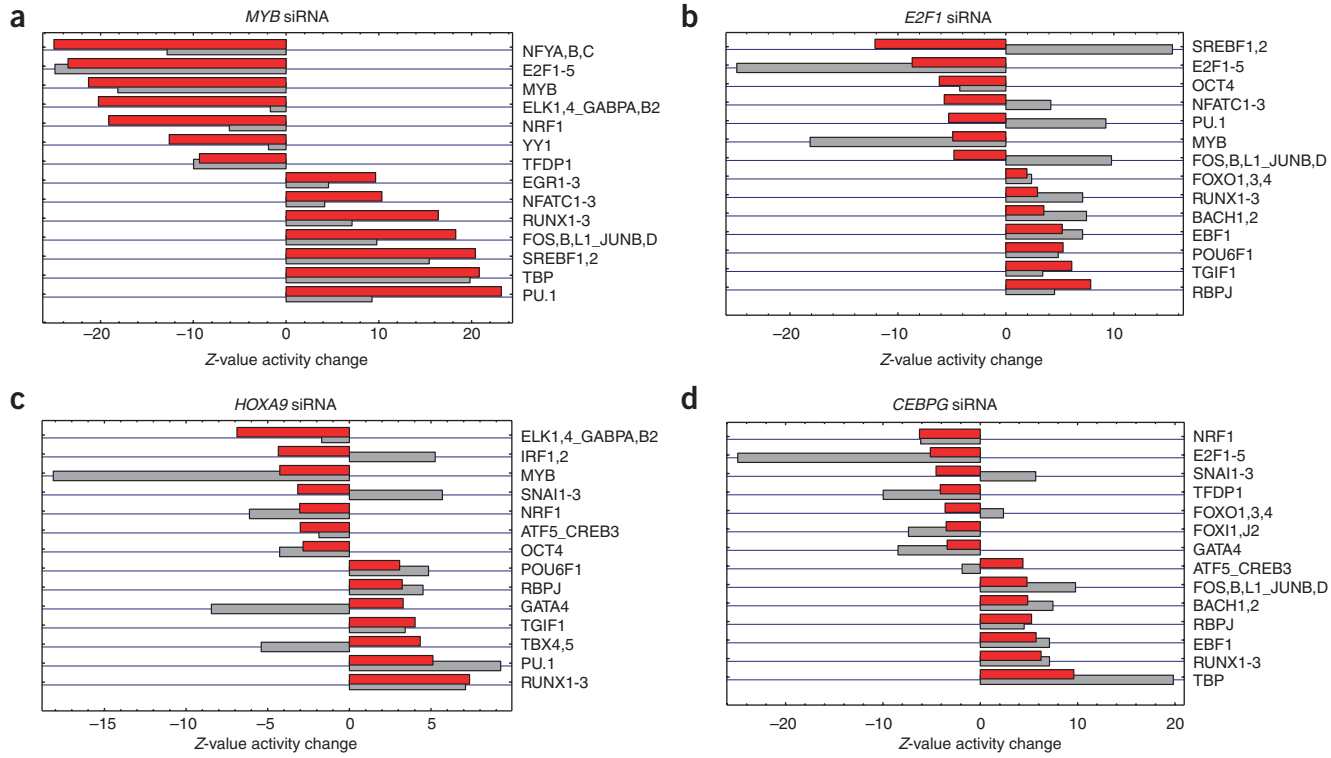

Figure 6 Most significant motif activity changes (as measured by *z* value, red bars) for four TF gene knockdowns that induce motif activity changes that have a differentiative overlap with the PMA time course of more than 50%. The corresponding motif activity changes observed in the PMA time course are shown as gray bars.

Notably, TFs that are predicted targets of SRF are mostly induced in the first hour of PMA-induced differentiation. During this first hour 55 of the 57 genes whose expression was perturbed are induced and 30% encode TFs (**Supplementary Fig. 17a** online). The regulatory inputs of these early-induced TFs are enriched for the motifs SRF, TBP and FOSL2 (**Supplementary Table 13d**), which all correspond to known PMA-responsive TFs[27–30]. Among the early-induced TFs, five correspond to upregulated core network motifs themselves (FOSB, EGR1-3 and SNAI1) and two (MAFB and EGR1) are known to induce pro-differentiative changes[31,32]. It is also worth noting that significant downregulation did not occur until the second hour, and this may require both early induction of transcriptional repressors and the RNA degradation proteins BTG2 and ZFP36 (tristetraprolin)[33,34] (**Supplementary Fig. 17b**). Together, these results suggest that induction of SRF target genes in the first hour is critical to establishing the differentiative program and is required before factors maintaining the undifferentiated state are downregulated (**Supplementary Fig. 17b,c**).

Web interface to data and analysis results

To facilitate the use of the data and analysis of results amassed here, we provide an online tool, EdgeExpressDB, as part of the FANTOM4 web resource, which allows users to explore our annotations of the structure, expression and regulation of promoters genome-wide. It also integrates published TF–promoter interactions, the siRNA perturbations and genome-wide chromatin immunoprecipitation experiments. Our complete set of regulatory-interaction predictions provides a large collection of hypotheses that can be targeted for validation, for example, through chromatin immunoprecipitation, gel shift assays or reporter assays. The value of this resource is illustrated by detailed examination of individual loci. For example, the osteopontin gene (*SPP1*) is massively induced from 12 h of differentiation (**Supplementary Note**). Our predictions confirm RUNX and PU.1 as regulators and support a previous analysis in mouse implicating the TGIF1 factor. In addition our analysis identifies NFAT, STAT, NKX6.2 and LIM domain and homeobox proteins as candidates for further testing.

Finally, our set of human promoters, TF motifs, genome-wide annotation of TF-binding sites and their predicted effects on the expression of the target promoters are available through the SwissRegulon website. A web interface, allowing researchers to automatically perform Motif Activity Response Analysis (MARA) of their own expression data in terms of our genome-wide predictions of TFBSs, is also available at SwissRegulon.

DISCUSSION

We have devised a new integrated approach that combines genome-wide identification of TSSs and their time-dependent expression with computational modeling to reconstruct the transcriptional regulatory dynamics of a differentiating human cell line. The CAGE tag sequencing used here is tenfold deeper than in previous studies[11], and this is the first study to our knowledge to quantitatively monitor dynamic expression changes of individual TSSs genome-wide. Using this data we developed a new computational method in which promoter expression profiles were modeled directly in terms of the TFBSs occurring in their proximal promoter regions. This method allowed us to infer which regulatory motifs are most predictive of expression changes and the time-dependent activities of the corresponding TFs *ab initio*. We identified more than two dozen different regulatory motifs that significantly change their activity during PMA-induced

differentiation and a complex network of regulatory interactions between them that have independent experimental support. Notably, although the modeling considers only TFBSs in proximal promoter sequences, the core network in **Figure 4** contains most of the known regulators of macrophage differentiation. Furthermore, siRNA perturbation of these TFs confirmed many of their predicted targets, and by analyzing changes in motif activity we found that each knockdown led to a distinct transcriptional state that was associated with changes in the activities of multiple motifs.

The changes in motif activity that we observed during THP-1 macrophage differentiation do not necessarily imply that the factor(s) that act upon a motif are themselves transcriptionally regulated. For example, PU.1 (SPI1) activity increases significantly in response to PMA and we have confirmed that, besides a moderate increase in mRNA expression, the SPI1 protein is also activated by phosphorylation[35] and nuclear translocation[36] (data not shown). For other motifs such as E2F, multiple redundant factors can bind to the same sites[37]. Motif activity analysis is conducted without any assumptions about the TFs that act through these regulatory elements. That is, because motif activity is inferred directly from expression changes of predicted targets, the most active motifs can be identified before ascertaining the responsible TF(s) and their mode of regulation. Thus, motif activity analysis is a powerful approach compared to analysis of TF mRNA expression alone.

What do our results teach us about the general structure of regulatory networks in cellular differentiation? An often evoked picture is that differentiation pathways consist of well-defined cascades of regulatory events which are initiated by master regulators that sit at the top of fixed regulatory hierarchies. A prime candidate for such a master regulator in our system would be *MYB*, as its siRNA-mediated knockdown reconstituted a significant fraction of the expression and phenotypic changes observed under PMA-induced differentiation. Indeed, this observation is consistent with earlier reports that *MYB* antisense treatment of myeloid leukemia lines causes differentiative growth arrest[38] and that *MYB* is a repressor of expression of mature macrophage-expressed genes such as *CSF1R*[39]. Our data indicate that *MYB* probably acts on such genes indirectly, by a transcriptional program that represses upregulation of *SPI1* activity and downregulation of proliferation.

However, several observations argue against *MYB* as a master regulator: *MYB* downregulation is not among the first events in the PMA time course, *MYB* knockdown far from completely mimics the PMA-induced differentiation and there are several other TFs, which are not downstream of *MYB*, whose knockdown reconstituted different subsets of the PMA-induced expression changes (**Supplementary Tables 8**, **9** and **11**). Moreover, it is known that additional factors can also drive differentiation, for example, enforced expression of *SPI1* and *CEBPA* in mouse fibroblasts is sufficient to drive acquisition of a macrophage-like phenotype[40], and overexpression of *EGR1* and *MAFB* also drives differentiation, as we noted above. Yet, evidence from mouse knockouts indicates that the whole EGR family is dispensable for macrophage proliferation, differentiation and function[41].

Rather than a fixed hierarchy with one or very few master regulators at the top, the picture that emerges is that of a recurrent network in which multiple TFs mutually coordinate their activity changes to implement the differentiation. In addition, whereas different partial differentiation pathways can be initiated by multiple independent perturbations, it appears that complete differentiation requires the coordinated downregulation of multiple factors that maintain the undifferentiated state. This observation draws some similarities to the TF network that both maintains proliferation and prevents differentiation in embryonic stem cells[42]. Enforced expression of four stem cell transcription factors (MYC, OCT4, KLF4, SOX2) is sufficient to dedifferentiate committed adult cells into a stem cell–like state[43]. Maintenance of an undifferentiated proliferative state is important in cancer, and it is worth noting that 10 of the 64 downregulated TFs (16%) have Entrez gene annotations containing the term 'myeloid leukemia' (compared to 50 of the remaining 1,258 TFs (4%); **Supplementary Table 14** online). In addition we have demonstrated that knockdown of the *MLL-MLLT3* leukemogenic fusion found in THP-1 also partially promotes differentiation.

From our time-course analysis, we see distinct phases of early, middle and late, induction and repression. Our modeling predicts, and the literature supports, *SRF* as the major effector of transcriptional activation of immediate early genes (IEG)[44]. However, SRF activation and IEG responses are not restricted to the PMA stimulus, the monocytic lineage or differentiation[28,45–47], suggesting that this response has a more general function. We speculate that a generalized immediate early response may be used to put the cell into a transient receptive state, which permits downregulation of the multiple TFs that maintain the undifferentiated state. This fits with the concept of stable cellular states as attractors of the regulatory network dynamics. The associated attractor basins[48,49] of cellular states are analogous to local minima in energy landscapes surrounded by slopes, and homeostatic interactions between the TFs can be considered as providing a kind of inertia to maintain this state. We suggest that the immediate early response may help overcome this inertia, that is, by moving the system out of its attractor basin.

METHODS

URLs. FANTOM4 web resource, http://fantom.gsc.riken.jp/4/; SwissRegulon, http://www.swissregulon.unibas.ch. All methods are described in **Supplementary Methods**.

Accession codes. Accession numbers of the data sets, deposited in public databases, are shown in **Supplementary Table 15** online.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank A. Ambesi, H. Atsui, M. Bansal, V. Belcastro, H. Daub, D. di Bernardo, M. Furuya, A. Hasegawa, K. Hayashida, A. Hirakiyama, F. Hori, K. Koseki, S. Kuhara, N. Miyamoto, S. Miyano, M. Nishikawa, C. Ohinata, M. Persson, S. Saihara, C. Sakaba, H. Sano, E. Shibazaki, T. Takagi, K. Toyoda, Y. Tsujimura and M. Yamamoto for discussion, encouragement and technical assistance. We thank M. Muramatsu, T. Ogawa, Y. Sakaki and A. Wada for support and encouragement. This work was mainly supported by grants for the Genome Network Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan (Y.H.), Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government (Y.H.) and the RIKEN Frontier Research System, Functional RNA research program (Y.H.). A.R.R.F. is supported by a CJ Martin Fellowship from the Australian National Health and Medical Research Council (ID 428261). E.v.N. acknowledges support from SNF grant SNF #3100A0-118318.

AUTHOR CONTRIBUTIONS

A.A., A.M.C., A.D., A. Kruger, A. Krogh, A.R., A.R.R.F., A.S., A.S.S., A.W., B.L., C.A.M., C.A.S., C.A.W., C.O.D., C.M., C. Simons, C. Schönbach, C.W., D.B., E.A., E.V., E.v.N., G.J.F., H. Kawaji, H. Kitano, H. Matsuda, J.L.F., J.G., J.M., J.Q., J.S., J.S.M., J.T., K. Ikeo, K.T., K.W., K.Y., L.H., M.d.H., M.E., M.G., M. Hörnquist, M. Kaur, M. Lizio, M. Maqungo, M.P., M. Sera, M.S.T., M.T., M.Z., N.B., N.C., O.H., O.W., P.J.B., P.G.E., R.I., R.J.T., R.S., R.D.T., S.F., S. Kondo, S. Katayama, S. Kimura, S. Meier, S.S., S. Teichmann, T.B., T.G., T.H., T.I., T. Konno, T.L., T.O., T.R., V.B.B., W.H., Y. Kimura, Y.N. and Y. Takenaka were involved in bioinformatic aspects of the project. A.G.B., A.J., A. Kaiho, A. Kubosaki, A. Kumar,

A.L., A.R.R.F., C.A.W., C. Kai, C. Kawazu, C.O., C.P., C. Simon, C.W., D.A.H., E.B., E.M.-S., F.B., G.S.L., H. Koga, H. Miura, H.N., H.O.-Y., H.S., H.Y., J.B., J.C., J.K., J.O., J.S.M., J.Y., K.F., K. Imamura, K.M., K.M.I., K.N., K. Schroder, K. Shirahige, L.W., M.A., M.C.K., M.F., M. Hashimoto, M. Hatakeyama, M.J.S., M.K.-K., M. Kojima, M. Murata, M.N., M.R., M. Suzuki, M.T., N.A.M., N.I., N.N., N.P., R.K., R.D.T., S.M.G., S.H., S.I., S. Miyamoto, S. Noma, S. Nygaard, S. Takeda, T.A., T. Kawashima, T. Kojima, T. Sano, T. Suzuki, V.O., Y.A., Y. Hasegawa, Y.I., Y. Kitazume, Y.N., Y.O., Y. Takahashi and Y. Tomaru were involved in biological aspects of the project. A.M.C., A.R.R.F., A.S., B.L., C.O.D., D.F., E.A., E.v.N., G.J.F., H.A., H.S., J.D., J.M., J.Q., J.S.M., K.W., M. Lindow, M.Z., N.C., N.M., O.H., P.J.B., P.C., R.J.T., R.S., S.M.G., S. Kondo, T.L., T.R. and V.O. were involved in the genome-wide and RNA analyses. E.v.N. and P.J.B. designed and carried out the motif activity response analysis. A.R.R.F., E.v.N., Y. Tomaru and M.K.-K. carried out the siRNA analysis. A.R.R.F., C.O.D., D.A.H., E.v.N., H.S., J.K., P.C. and Y. Hayashizaki oversaw the project. H.S., A.R.R.F., E.v.N., and D.A.H. wrote the manuscript with assistance from T.R., T.L., M.J.S., Y. Hasegawa, M.d.H., K.M.I., K.Schloder, P.J.C., P.J.B., E.A., N.P., M.R., S.M.G., C.A.W., J.Q., W.H., A. Kubosaki, Y. Tomaru, V.B.B., M. Suzuki and Y. Hayashizaki.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Tsuchiya, S. *et al.* Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. *Cancer Res.* **42**, 1530–1536 (1982).
2. Abrink, M., Gobl, A.E., Huang, R., Nilsson, K. & Hellman, L. Human cell lines U-937, THP-1 and Mono Mac 6 represent relatively immature cells of the monocyte-macrophage cell lineage. *Leukemia* **8**, 1579–1584 (1994).
3. Beer, M.A. & Tavazoie, S. Predicting gene expression from sequence. *Cell* **117**, 185–198 (2004).
4. Ramsey, S.A. *et al.* Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* **4**, e1000021 (2008).
5. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**, 166–176 (2003).
6. Das, D., Nahle, Z. & Zhang, M.Q. Adaptively inferring human transcriptional subnetworks. *Mol. Syst. Biol.* **2**, 2006.0029 (2006).
7. Gao, F., Foat, B.C. & Bussemaker, H.J. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics* **5**, 31 (2004).
8. Nguyen, D.H. & D'Haeseleer, P. Deciphering principles of transcription regulation in eukaryotic genomes. *Mol. Syst. Biol.* **2**, 2006.0012 (2006).
9. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
10. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
11. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).
12. Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. USA* **100**, 15776–15781 (2003).
13. Roh, T.Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).
14. Sandoval, J. *et al.* RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nucleic Acids Res.* **32**, e88 (2004).
15. Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **5**, 613–619 (2008).
16. Vlieghe, D. *et al.* A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* **34**, D95–D97 (2006).
17. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **24**, 238–241 (1996).
18. Moses, A.M., Chiang, D.Y., Pollard, D.A., Iyer, V.N. & Eisen, M.B. MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model. *Genome Biol.* **5**, R98 (2004).
19. van Nimwegen, E. Finding regulatory elements and regulatory motifs: a general probabilistic framework. *BMC Bioinformatics* **8** (Suppl. 6), S4 (2007).
20. Frith, M.C. *et al.* A code for transcription initiation in mammalian genomes. *Genome Res.* **18**, 1–12 (2008).
21. Bussemaker, H.J., Foat, B.C. & Ward, L.D. Predictive modeling of genome-wide mRNA expression: from modules to molecules. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 329–347 (2007).
22. Schmitz, G. & Grandl, M. Lipid homeostasis in macrophages—implications for atherosclerosis. *Rev. Physiol. Biochem. Pharmacol.* **160**, 93–126 (2008).
23. Odero, M.D., Zeleznik-Le, N.J., Chinwalla, V. & Rowley, J.D. Cytogenetic and molecular analysis of the acute monocytic leukemia cell line THP-1 with an MLL-AF9 translocation. *Genes Chromosom. Cancer* **29**, 333–338 (2000).
24. Martino, V. *et al.* Down-regulation of MLL-AF9, MLL and MYC expression is not obligatory for monocyte-macrophage maturation in AML-M5 cell lines carrying t(9;11)(p22;q23). *Oncol. Rep.* **15**, 207–211 (2006).
25. Pession, A. *et al.* MLL-AF9 oncogene expression affects cell growth but not terminal differentiation and is downregulated during monocyte-macrophage maturation in AML-M5 THP-1 cells. *Oncogene* **22**, 8671–8676 (2003).
26. Roach, J.C. *et al.* Transcription factor expression in lipopolysaccharide-activated peripheral-blood-derived mononuclear cells. *Proc. Natl. Acad. Sci. USA* **104**, 16245–16250 (2007).
27. Biggs, J.R., Ahn, N.G. & Kraft, A.S. Activation of the mitogen-activated protein kinase pathway in U937 leukemic cells induces phosphorylation of the amino terminus of the TATA-binding protein. *Cell Growth Differ.* **9**, 667–676 (1998).
28. Iyer, D. *et al.* Serum response factor MADS box serine-162 phosphorylation switches proliferation and myogenic gene programs. *Proc. Natl. Acad. Sci. USA* **103**, 4516–4521 (2006).
29. Morton, S., Davis, R.J. & Cohen, P. Signalling pathways involved in multisite phosphorylation of the transcription factor ATF-2. *FEBS Lett.* **572**, 177–183 (2004).
30. Trejo, J. *et al.* A direct role for protein kinase C and the transcription factor Jun/AP-1 in the regulation of the Alzheimer's beta-amyloid precursor protein gene. *J. Biol. Chem.* **269**, 21682–21690 (1994).
31. Kelly, L.M., Englmeier, U., Lafon, I., Sieweke, M.H. & Graf, T. MafB is an inducer of monocytic differentiation. *EMBO J.* **19**, 1987–1997 (2000).
32. Krishnaraju, K., Hoffman, B. & Liebermann, D.A. The zinc finger transcription factor Egr-1 activates macrophage differentiation in M1 myeloblastic leukemia cells. *Blood* **92**, 1957–1966 (1998).
33. Mauxion, F., Faux, C. & Seraphin, B. The BTG2 protein is a general activator of mRNA deadenylation. *EMBO J.* **27**, 1039–1048 (2008).
34. Blackshear, P.J. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem. Soc. Trans.* **30**, 945–952 (2002).
35. Carey, J.O., Posekany, K.J., deVente, J.E., Pettit, G.R. & Ways, D.K. Phorbol ester-stimulated phosphorylation of PU.1: association with leukemic cell growth inhibition. *Blood* **87**, 4316–4324 (1996).
36. Foster, N., Lea, S.R., Preshaw, P.M. & Taylor, J.J. Pivotal advance: vasoactive intestinal peptide inhibits up-regulation of human monocyte TLR2 and TLR4 by LPS and differentiation of monocytes to macrophages. *J. Leukoc. Biol.* **81**, 893–903 (2007).
37. Xu, X. *et al.* A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res.* **17**, 1550–1561 (2007).
38. Anfossi, G., Gewirtz, A.M. & Calabretta, B. An oligomer complementary to c-myb-encoded mRNA inhibits proliferation of human myeloid leukemia cell lines. *Proc. Natl. Acad. Sci. USA* **86**, 3379–3383 (1989).
39. Reddy, M.A. *et al.* Opposing actions of c-ets/PU.1 and c-myb protooncogene products in regulating the macrophage-specific promoters of the human and mouse colony-stimulating factor-1 receptor (c-fms) genes. *J. Exp. Med.* **180**, 2309–2319 (1994).
40. Feng, R. *et al.* PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proc. Natl. Acad. Sci. USA* **105**, 6057–6062 (2008).
41. Carter, J.H. & Tourtellotte, W.G. Early growth response transcriptional regulators are dispensable for macrophage differentiation. *J. Immunol.* **178**, 3038–3047 (2007).
42. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
43. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
44. Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U. & Nordheim, A. Serum response factor is essential for mesoderm formation during mouse embryogenesis. *EMBO J.* **17**, 6289–6299 (1998).
45. Cooper, S.J., Trinklein, N.D., Nguyen, L. & Myers, R.M. Serum response factor binding sites differ in three human cell types. *Genome Res.* **17**, 136–144 (2007).
46. Fleige, A. *et al.* Serum response factor contributes selectively to lymphocyte development. *J. Biol. Chem.* **282**, 24320–24328 (2007).
47. Poser, S., Impey, S., Trinh, K., Xia, Z. & Storm, D.R. SRF-dependent gene expression is required for PI3-kinase-regulated cell proliferation. *EMBO J.* **19**, 4955–4966 (2000).
48. Huang, S. & Ingber, D.E. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp. Cell Res.* **261**, 91–103 (2000).
49. Kauffman, S. *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).

The full list of authors and affiliations is as follows:
The FANTOM Consortium:

Harukazu Suzuki[2,50–52], Alistair R R Forrest[2,3,50–52], Erik van Nimwegen[4,50–52], Carsten O Daub[2,51,52], Piotr J Balwierz[4,51], Katharine M Irvine[5,51,52], Timo Lassmann[2,51,52], Timothy Ravasi[6,51,52], Yuki Hasegawa[2,51], Michiel J L de Hoon[2,51], Shintaro Katayama[2,51], Kate Schroder[5,51,52], Piero Carninci[2,51], Yasuhiro Tomaru[2,51], Mutsumi Kanamori-Katayama[2,51], Atsutaka Kubosaki[2,51], Altuna Akalin[7], Yoshinari Ando[2], Erik Arner[2], Maki Asada[8], Hiroshi Asahara[8], Timothy Bailey[5],

Vladimir B Bajic[9,51], Denis Bauer[5], Anthony G Beckhouse[3], Nicolas Bertin[2], Johan Björkegren[10], Frank Brombacher[11], Erika Bulger[2], Alistair M Chalk[3], Joe Chiba[12], Nicole Cloonan[13], Adam Dawe[9], Josee Dostie[14], Pär G Engström[7], Magbubah Essack[9], Geoffrey J Faulkner[13], J Lynn Fink[15], David Fredman[7], Ko Fujimori[16], Masaaki Furuno[2], Takashi Gojobori[17,51], Julian Gough[18], Sean M Grimmond[13,51], Mika Gustafsson[19], Megumi Hashimoto[8], Takehiro Hashimoto[2], Mariko Hatakeyama[20], Susanne Heinzel[21], Winston Hide[9,22,51], Oliver Hofmann[9,22], Michael Hörnquist[19], Lukasz Huminiecki[23], Kazuho Ikeo[17], Naoko Imamoto[24], Satoshi Inoue[25], Yusuke Inoue[26], Ryoko Ishihara[2], Takao Iwayanagi[27], Anders Jacobsen[28], Mandeep Kaur[9], Hideya Kawaji[2], Markus C Kerr[15], Ryuichiro Kimura[12], Syuhei Kimura[29], Yasumasa Kimura[2], Hiroaki Kitano[30], Hisashi Koga[31], Toshio Kojima[20], Shinji Kondo[2], Takeshi Konno[17], Anders Krogh[28], Adele Kruger[9], Ajit Kumar[32], Boris Lenhard[7,51], Andreas Lennartsson[2], Morten Lindow[28], Marina Lizio[2], Cameron MacPherson[9], Norihiro Maeda[2], Christopher A Maher[9], Monique Maqungo[9], Jessica Mar[33], Nicholas A Matigian[3], Hideo Matsuda[34], John S Mattick[13], Stuart Meier[9], Sei Miyamoto[17], Etsuko Miyamoto-Sato[35], Kazuhiko Nakabayashi[12], Yutaka Nakachi[36], Mika Nakano[2], Sanne Nygaard[28], Toshitsugu Okayama[17], Yasushi Okazaki[36], Haruka Okuda-Yabukami[2], Valerio Orlando[37], Jun Otomo[38], Mikhail Pachkov[4], Nikolai Petrovsky[21], Charles Plessy[2], John Quackenbush[33,51], Aleksandar Radovanovic[9], Michael Rehli[39], Rintaro Saito[40], Albin Sandelin[28], Sebastian Schmeier[9], Christian Schönbach[41], Ariel S Schwartz[6], Colin A Semple[42], Miho Sera[17], Jessica Severin[2], Katsuhiko Shirahige[43], Cas Simons[13], George St. Laurent[32], Masanori Suzuki[2], Takahiro Suzuki[2], Matthew J Sweet[44], Ryan J Taft[13], Shizu Takeda[38], Yoichi Takenaka[34], Kai Tan[6], Martin S Taylor[45], Rohan D Teasdale[15], Jesper Tegnér[10,46,51], Sarah Teichmann[47], Eivind Valen[28], Claes Wahlestedt[48], Kazunori Waki[2], Andrew Waterhouse[2], Christine A Wells[3,51], Ole Winther[28], Linda Wu[21], Kazumi Yamaguchi[2], Hiroshi Yanagawa[35], Jun Yasuda[2], Mihaela Zavolan[4] & David A Hume[49,51,52]

Riken Omics Science Center:
Takahiro Arakawa[2], Shiro Fukuda[2], Kengo Imamura[2], Chikatoshi Kai[2], Ai Kaiho[2], Tsugumi Kawashima[2], Chika Kawazu[2], Yayoi Kitazume[2], Miki Kojima[2], Hisashi Miura[2], Kayoko Murakami[2], Mitsuyoshi Murata[2], Noriko Ninomiya[2], Hiromi Nishiyori[2], Shohei Noma[2], Chihiro Ogawa[2], Takuma Sano[2], Christophe Simon[2], Michihira Tagami[2], Yukari Takahashi[2] & Jun Kawai[2,51]

General Organizer:
Yoshihide Hayashizaki[2,51,52]

[2]RIKEN Omics Science Center, RIKEN Yokohama Institute, Kanagawa, Japan. [3]The Eskitis Institute for Cell and Molecular Therapies, Griffith University, Australia. [4]Biozentrum, University of Basel, and Swiss Institute of Bioinformatics, Basel, Switzerland. [5]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. [6]Department of Bioengineering, Jacobs School of Engineering, University of California, San Diego, La Jolla, California, USA. [7]Bergen Center for Computational Science, Bergen, Norway. [8]National Research Institute for Child Health and Development, Tokyo, Japan. [9]South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa. [10]Computational Medicine Group, Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden. [11]Institute of Infectious Disease and Molecular Medicine (IIDMM), Wolfson Pavilion Level 2, Faculty of Health Sciences, University of Cape Town, Observatory, South Africa. [12]Department of Biological Science and Technology, Tokyo University of Science, Japan. [13]Australian Research Council (ARC) Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Australia. [14]Department of Biochemistry, McGill University, Montreal, Quebec, Canada. [15]Australian Research Council (ARC) Centre of Excellence in Bioinformatics, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, Australia. [16]Laboratory of Biodefense and Regulation, Osaka University of Pharmaceutical Sciences, Osaka, Japan. [17]Research Organization of Information and Systems, Center for Information Biology and DNA Data Bank of Japan (DDBJ), National Institute of Genetics, Shizuoka, Japan. [18]Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, UK. [19]Department of Science and Technology, Linköping University, Norrköping, Sweden. [20]Computational and Experimental Systems Biology Group, RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, Kanagawa Japan. [21]Department of Diabetes and Endcrinology, Flinders University and Medical Centre, Bedford Park, Adelaide, Australia. [22]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [23]Department of Cell and Molecular Biology (CMB), Karolinska Institutet, Stockholm, Sweden. [24]Cellular Dynamics Laboratory, Discovery and Research Institute, RIKEN Wako Institute, Saitama, Japan. [25]Graduate School of Medicine and Faculty of Medicine, the University of Tokyo, Tokyo, Japan. [26]Department of Biological and Chemical Engineering, Gunma University Faculty of Engineering, Gunma, Japan. [27]R&D Solution Center, Research & Development Group, Hitachi Ltd., Tokyo, Japan. [28]The Bioinformatics Centre, Department of Biology and Biotech Research & Innovation Centre, University of Copenhagen, Copenhagen, Denmark. [29]Department of Information and Knowledge Engineering, Faculty of Engineering, Tottori University, Tottori, Japan. [30]The Systems Biology Institute, Shibuya, Tokyo, Japan. [31]Department of Human Gene Research, Kazusa DNA Research Institute, Chiba, Japan. [32]Department of Biochemistry and Molecular Biology, the George Washington University Medical Center, Washington, D.C., USA. [33]Department of Biostatistics, Harvard School of Public Health, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [34]Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. [35]Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, Yokohama, Japan. [36]Division of Functional Genomics and Systems Medicine, Research Center for Genomic Medicine, Saitama Medical School, Saitama, Japan. [37]Dulbecco Telethon Institute, IRCCS Fondazione Santa Lucia at EBRI, Rome and IGB CNR, Naples, Italy. [38]Central Research Laboratory, Hitachi Ltd., Tokyo, Japan. [39]Department of Hematology and Oncology, University of Regensburg, Hospital, Regensburg, Germany. [40]Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan. [41]School of Biological Science, Division of Genomics and Genetics, Nanyang Technological University, Singapore. [42]MRC Human Genetics Unit, Western, General Hospital, Crewe Road, Edinburgh, UK. [43]Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan. [44]Institute for Molecular Bioscience, School of Molecular and Microbial Sciences, CRC for Chronic Inflammatory Diseases, The University of Queensland, St. Lucia, Australia. [45]EMBL European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [46]Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden. [47]Structural Studies Division MRC Laboratory of Molecular Biology, Hills Rd., Cambridge, UK. [48]Biochemistry/Neuroscience, the Scripps Research Institute, Jupiter, Florida, USA. [49]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Roslin, UK. [50]These authors contributed equally to this work. [51]These authors are the core writing group. [52]These authors are affiliated with the FANTOM 4 headquarters. Correspondence should be addressed to D.H. (david.hume@roslin.ed.ac.uk) or Y.H. (yosihide@gsc.riken.jp).