Published online 19 August 2010 Nucleic Acids Research, 2010, Vol. 38, No. 22 8141–8148
doi:10.1093/nar/gkq729

# Building promoter aware transcriptional regulatory networks using siRNA perturbation and deepCAGE

**Morana Vitezic[1,2,*], Timo Lassmann[1,*], Alistair R. R. Forrest[1], Masanori Suzuki[1], Yasuhiro Tomaru[1], Jun Kawai[1], Piero Carninci[1], Harukazu Suzuki[1], Yoshihide Hayashizaki[1] and Carsten O. Daub[1]**

[1]Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045 Japan and [2]Department of Cell and Molecular Biology (CMB), Karolinska Institute, SE-171 77, Stockholm, Sweden

Received February 1, 2010; Revised and Accepted August 2, 2010

## ABSTRACT

**Perturbation and time-course data sets, in combination with computational approaches, can be used to infer transcriptional regulatory networks which ultimately govern the developmental pathways and responses of cells. Here, we individually knocked down the four transcription factors PU.1, IRF8, MYB and SP1 in the human monocyte leukemia THP-1 cell line and profiled the genome-wide transcriptional response of individual transcription starting sites using deep sequencing based Cap Analysis of Gene Expression. From the proximal promoter regions of the responding transcription starting sites, we derived *de novo* binding-site motifs, characterized their biological function and constructed a network. We found a previously described composite motif for PU.1 and IRF8 that explains the overlapping set of transcriptional responses upon knockdown of either factor.**

## INTRODUCTION

The human genome project (1) and the subsequent annotation efforts (2,3) provided us a catalog of genes present in our genome. These efforts quickly gave rise to system approaches aiming at understanding the interactions between genes that ultimately govern phenotype and disease pathology (4). The complex interactions among transcription factors derived from such networks point to diverse regulatory programs responsible for cell differentiation during development and cellular responses to outside stimuli.

A powerful technique to understand gene regulatory networks is the perturbation of individual transcription factors in concert with high-throughput expression profiling of all genes (5). Commonly, microarrays are used to measure the changes in gene expression (6–8). In addition to defining regulatory interactions, transcription factor binding site (TFBS) motifs can be extracted from promoter regions of affected genes. Searching the genome sequence *in silico* with such motifs can reveal putative downstream targets of the transcription factors. However, these predictions are fraught with difficulties summarized by the 'futility theorem' (9). In brief, most predicted binding sites will have no functional role in general and, despite binding *in vitro*, may not be functional in the cellular model studied or may only be functional in presence of additional factors (co-regulation). Therefore, it is desirable to couple computational approaches with experimental techniques to identify actively used TFBS.

Chromatin immunoprecipitation (ChIP) in conjunction with tiling microarrays or sequencing is able to tell us the possible binding sites of transcription factors. To be able to perform experiments for specific transcription factors, however, specific antibodies are needed whose production is both difficult and, for many of the transcription factors, not yet available (10). Additional specific experimental optimizations are required.

Here, we describe the use of deep sequencing based Cap Analysis of Gene Expression (deepCAGE) (11) to study the effects of transcription factor (TF) perturbations on target gene expression at the promoter level. Previously, deepCAGE was used to accurately define and compare the transcriptional start sites (TSS) of genes in various tissues (7), determine the distance of the

*To whom correspondence should be addressed. Tel: +81 45 503 9220; Fax: +81 45 503 9216; Email: mvitezic@gsc.riken.jp
Correspondence may also be addressed to Timo Lassmann. Tel: +81 45 503 9220; Fax: +81 45 503 9216; Email: lassmann@gsc.riken.jp

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

TATA-box from the TSS (12), as well as during cell differentiation (3). Restricting TFBS analysis to the accurately mapped TSSs discards many false-positive predictions in intergenic regions and thus improves the accuracy of transcriptional regulatory networks (3). In contrast to previous approaches, this allows for the construction of transcriptional regulatory gene networks at the resolution of individual promoters.

In this study, we combined our deepCAGE (3,13) technology with knockdown (KD) perturbation experiments of four key transcription factors (PU.1, IRF8, MYB and SP1) expressed in the human monoblastic leukemia cell line THP-1 (14). Previously, we demonstrated by using siRNA-mediated gene knockdown and microarray profiling that these four factors regulate large numbers of genes important to monocyte biology. In particular, MYB knockdown promotes monocytic differentiation of THP-1 cells, indicating a central role in maintaining the undifferentiated monoblast state (3).

DeepCAGE profiles were generated for each of the samples and compared to cells treated with a scrambled negative control oligo. This approach allowed us to identify the most strongly affected TSSs for each TF knockdown and their corresponding promoter regions. We then attempted to derive *de novo* TFBS motifs from the promoter regions and compared our results to the known binding-site models in the TRANSFAC database. Finally, these data were used to draw a basic regulatory network based on the direct regulatory interactions we identified.

## MATERIALS AND METHODS

### Cell culture and knockdown experiments

We used RNA extracted from the same knockdown human leukemia THP-1 cell batches used in the recent FANTOM4 project (3,8). In brief, transfection was performed using stealth siRNA (Invitrogen) and RNA was harvested after 48 h. TF gene-expression levels in THP-1 cells treated with gene-specific siRNAs (SP1, PU.1, IRF8 and MYB) or the calibrator negative control (NC) siRNA were estimated by qRT-PCR in triplicate [see Supplementary material of Suzuki *et al.* (3)].

### deepCAGE library generation, mapping and clustering of deepCAGE tags

deepCAGE libraries were prepared for the five knockdown experiments according to the deepCAGE protocol (3,13) and sequenced using the Roche 454 sequencer. In total, 6 187 981 deepCAGE tags were mapped to the human reference genome sequence (hg18) using Nexalign (Lassmann,T., http://genome.gsc.riken.jp/osc/english/dataresource/) allowing up to one mismatch or one indel. Tags with TSS falling into windows of 20 bp were grouped into 396 118 tag clusters (TCs). For all further analyses, we focused on a filtered set of 3332 robustly detected TCs with a minimum average deepCAGE expression across the five (four KD and control) libraries of 30 tags per million (TPM).

### Comparison of deepCAGE and microarray expression

For comparing the perturbation of deepCAGE expression profiles with microarray expression, we first mapped the 3332 robustly detected TCs to Entrez gene models, requiring that the tags originated within the boundaries of known transcripts for the locus or up to 1 kb upstream. The 3332 TCs mapped to 3114 Entrez genes using this approach, with 84 genes possessing more than one robustly detected TC. Fold change for the deepCAGE data was then calculated by dividing the gene expression in TF KD by the expression in the negative control experiment. Microarray probe mapping to Entrez gene and expression fold changes were obtained as described in Suzuki *et al.* (3). This then allowed direct comparison of fold changes measured by deepCAGE with the corresponding measurement by microarray.

### *De novo* motif prediction, TFBS prediction and ChIP-chip data

Proximal promoter regions of TSSs were defined as previously described (3) and include 300 bp upstream and 100 bp downstream of the deepCAGE-defined TSS. We extracted the corresponding active deepCAGE promoter regions from the human genome (hg18) and applied the motif-finding program MEME (15). We applied MEME to regions which are at least 1.5-fold up- or downregulated in both microarray and deepCAGE measurement. The selection was further restricted to the top 50 of such regions based on recommendations found in Bailey *et al.* (15). We hypothesize that this selection enriches for promoters that are direct targets of the transcription factor. In the case of IRF8, SP1 and PU.1, fewer than 50 TCs were upregulated by at least 1.5-fold (20, 22 and 38, respectively); therefore, smaller training sets were used for these classes.

MEME can report multiple motifs for each set of the proximal promoter regions. In such cases, we only selected the motif with the most significant *E*-value for further analysis. We did not attempt to merge similar motifs.

To assess whether the obtained motifs are biologically relevant, we searched the remaining TCs (3332 TCs, excluding the training sets) using the program Fimo from the Meta-MEME package (16). For comparison, we used the TRANSFAC database and the accompanying Match program (17,18) to scan our sequences for the presence of TRANSFAC defined motifs. Furthermore, we overlaid our TCs with previously published ChIP-chip data (3) for PU.1 and SP1 (detailed Methods available in the Supplementary Data).

We used UCSC browser Vertebrate Multiz Alignment & PhastCons trac to look for conservation of our motifs. A base position in the motif was deemed to be conserved if the conservation was at least 80%.

### Accession codes

DNA Data Bank of Japan (DDBJ) Read Archive: DRX000341 (CAGE library I05).

## RESULTS

### deepCAGE and microarray profiling of siRNA knockdowns identifies overlapping sets of perturbed genes

To evaluate deepCAGE as a platform for measuring gene-expression perturbation, we used the same batches of RNAs for both TF suppression and negative control samples as were used in the microarray analysis for the FANTOM4 main paper (3). For these samples, the efficient knockdown was already confirmed by qRT-PCR and western blotting. We observed an overall positive correlation for all four TF knockdown samples across both platforms (Figure 1). In general, deepCAGE fold changes were greater than those measured by microarrays, as has been previously noted (19).

### *De novo* motif prediction using knockdown deepCAGE identifies known core motifs, extended motifs and a composite motif for PU.1 and IRF8

Knockdown of SP1, IRF8, PU.1 and MYB led to induction of 267, 347, 189 and 307 genes and repression of 428, 527, 260 and 1160 genes by 1.5-fold up- or downregulation, respectively. Eight sets of proximal promoter regions were extracted corresponding to the top 50 most upregulated and most downregulated TCs for each knockdown experiment (see 'Methods' section). The *de novo* motif-finding algorithm MEME (15) was used to identify motifs enriched in the perturbed promoters. We identified motifs for all four downregulated promoter sets and also identified motifs in the promoters of the upregulated sets for MYB and PU.1.

Enrichment in the upregulated set of promoters suggests the TF works as a repressor,whereas enrichment in the downregulated set of promoters suggests the TF works as an activator. As an example, we find that knockdown of IRF8, a known activator (20), results in downregulation in both the deepCAGE and microarray experiments of XAF1, a gene which we predict to contain our novel motif (Figure 2). The observation that MYB knockdown yielded motifs for both up- and downregulated sets is consistent with its known role as both a transcriptional activator and repressor (21).

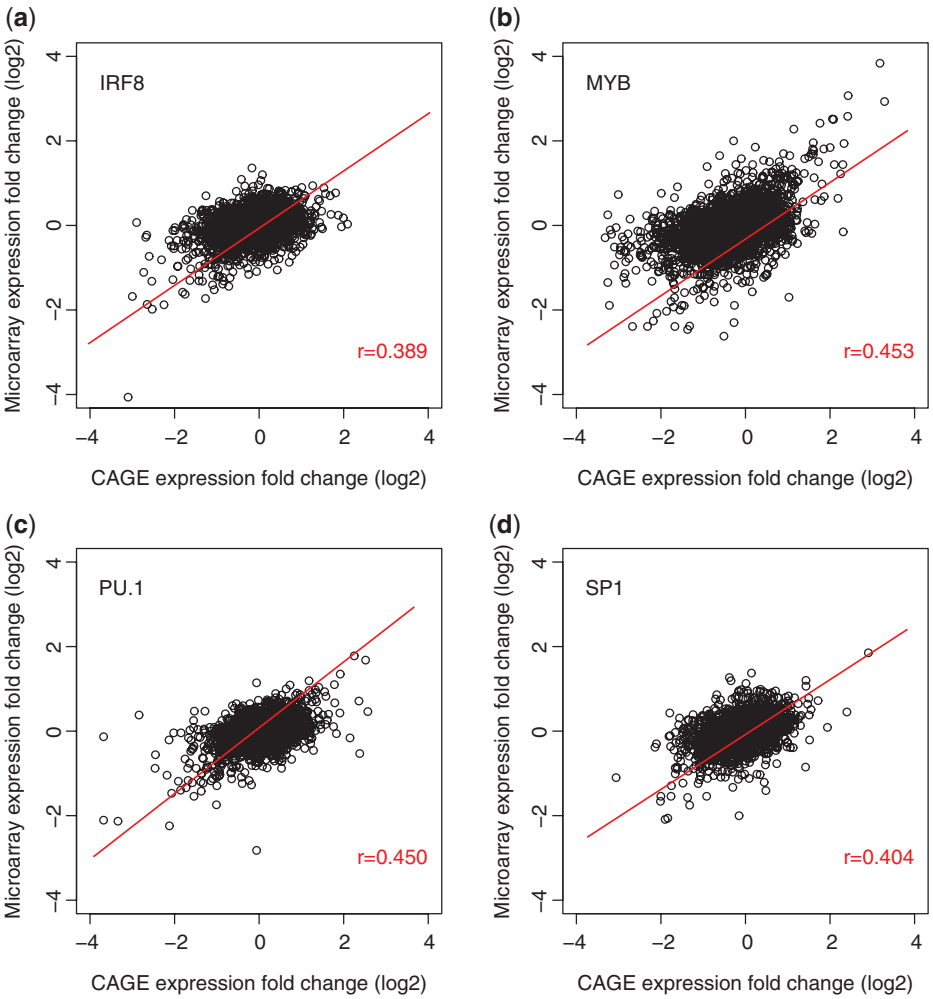


**Figure 1.** DeepCAGE and microarrays detect overall similar expression changes. The transcriptome-profiling technologies deepCAGE and microarrays showed overall similar transcriptional response (log2 expression fold-change) comparing before and after siRNA-based knockdown of the transcription factors IRF8, MYB, PU.1 and SP1. The Pearson correlation values for these two platforms are: (**a**) 0.389 ($P = 1.3e^{-12}$) for IRF8, (**b**) 0.453 ($P = 2.2e^{-16}$) for MYB, (**c**) 0.450 ($P = 1.2e^{-11}$) for PU.1 and (**d**) 0.404 ($P = 6.7e^{-10}$) for SP1.

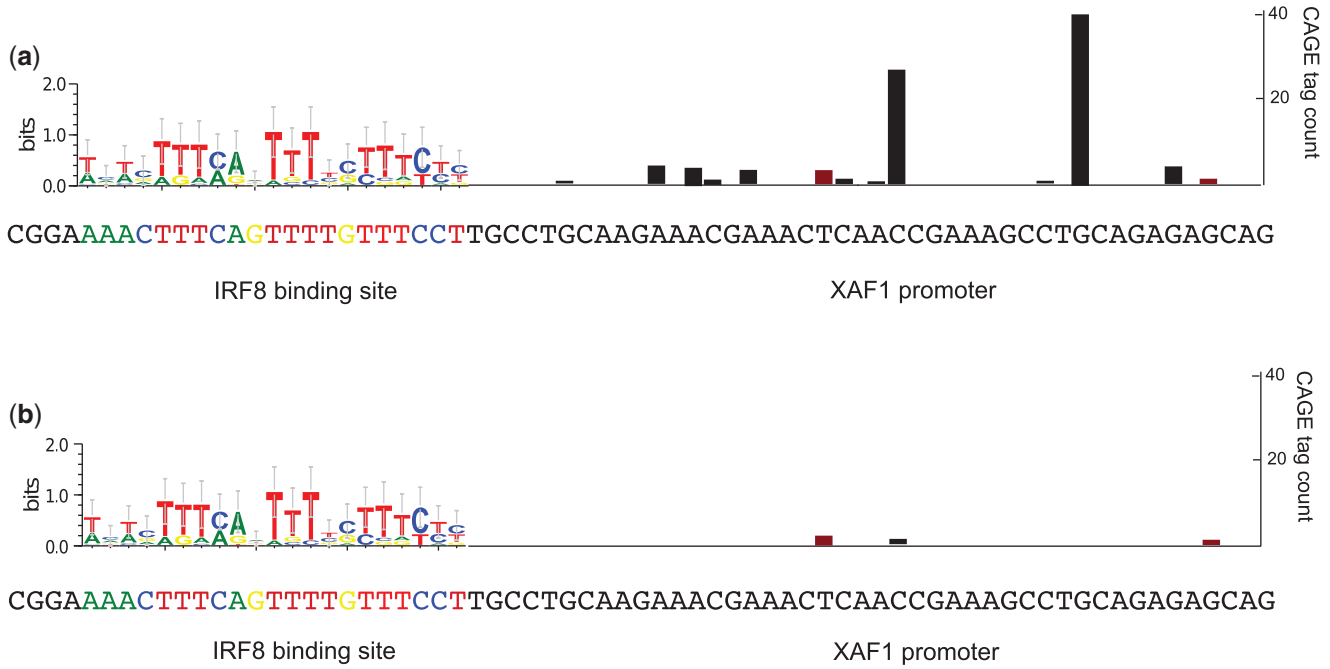


**Figure 2.** DeepCAGE identified individual transcription starting sites responding to transcription factor knockdown. DeepCAGE profiling of the transcriptome quantitatively measures individual transcription starting sites (TSS) of capped mRNA indicated by the vertical bars (**a**) before and (**b**) after the knockdown of the IRF8 transcription factor. Red bars indicate CAGE tags that do not change upon knockdown while the black bars represent tags showing significant change upon knockdown. One transcript cluster (TC) is shown in the promoter region of the XAF1 gene on chromosome 17 (positions 6 600 047–6 600 115, hg18) together with the defining TSSs.

Despite this, the motifs found in either set appear to be different, which may suggest different modes or different co-factors for binding repressive and activating sites.

To assess whether our *de novo* motifs identify functional sites, we examined the expression fold-changes of TCs containing the predicted motifs compared to all other TCs. The TCs used to derive the motifs in the first place were excluded. Instead of using CAGE data based on a single experiment, we used microarray expression data based on three biological replicas from the same RNA batch since we deemed it to be more reliable (Figure 3). However, when using CAGE expression data instead, there are no discernible differences (Supplementary Figure S1).

Of the motifs tested, those for MYB up, MYB down, PU.1 up and SP1 down data sets (Supplementary Figures S2 and S3) did not show significant expression differences between the sequences containing the motifs and those that do not contain the motifs. Hence, we did not further analyze these motifs.

However, promoters containing PU.1 down or IRF8 down motifs were expressed at significantly lower levels than promoters lacking the motif. Moreover, when the same test was carried out using the published TRANSFAC (17,18) motifs for PU.1 and IRF8, or using ChIP data for PU.1 to identify PU.1-bound promoters, neither outperformed the novel motif (Figure 3). Furthermore, comparison to UCSC's vertebrate-conservation track revealed that 32.8 and 35.5% of the novel PU.1 and IRF8 base positions, respectively, are strictly conserved, while 11 out of 47 and 7 out of 20 PU.1 and IRF8 motifs are completely conserved. This compares with 3–8% average overall conservation and 11–24% conservation in coding regions.

In a parallel effort, we used the program CLOVER (22) to detect enriched motifs in the top 50 downregulated IRF8 and PU.1 CAGE clusters. As expected, we found enrichment for the corresponding known motifs in both data sets (for details see Supplementary Data and Figures S3 and S4 and Tables S2 and S3). However, the enriched motifs are only weakly overrepresented when considering all downregulated clusters. Therefore, the *de novo* derived motifs describe the transcriptional response to TF knockdown better than using known motifs or the present ChIP-chip data.

The motifs obtained for PU.1 and IRF8 were longer than the corresponding motifs in the TRANSFAC database (Figure 4a). Manual alignment of our matrices to each other and to the TRANSFAC motifs revealed that both of our motifs contain regions similar to the TRANSFAC PU.1 and the IRF8 motifs. Furthermore, we observed 44 promoters that were downregulated in both IRF8 and PU.1 knockdown (Supplementary Figure S5 and Table S4). Our IRF8 motif contains three triple-T (TTT) regions. To understand their significance, we truncated our IRF8 motif by removing the triple-T sub-motif from either end. The expression differences in the test set became less pronounced (Figure 4b), indicating that all three triple-Ts are important for the specificity. Similar examples of combinatorial regulation were previously described for IRF8 and other IRF family members and for the PU.1 transcription factor (20,23).

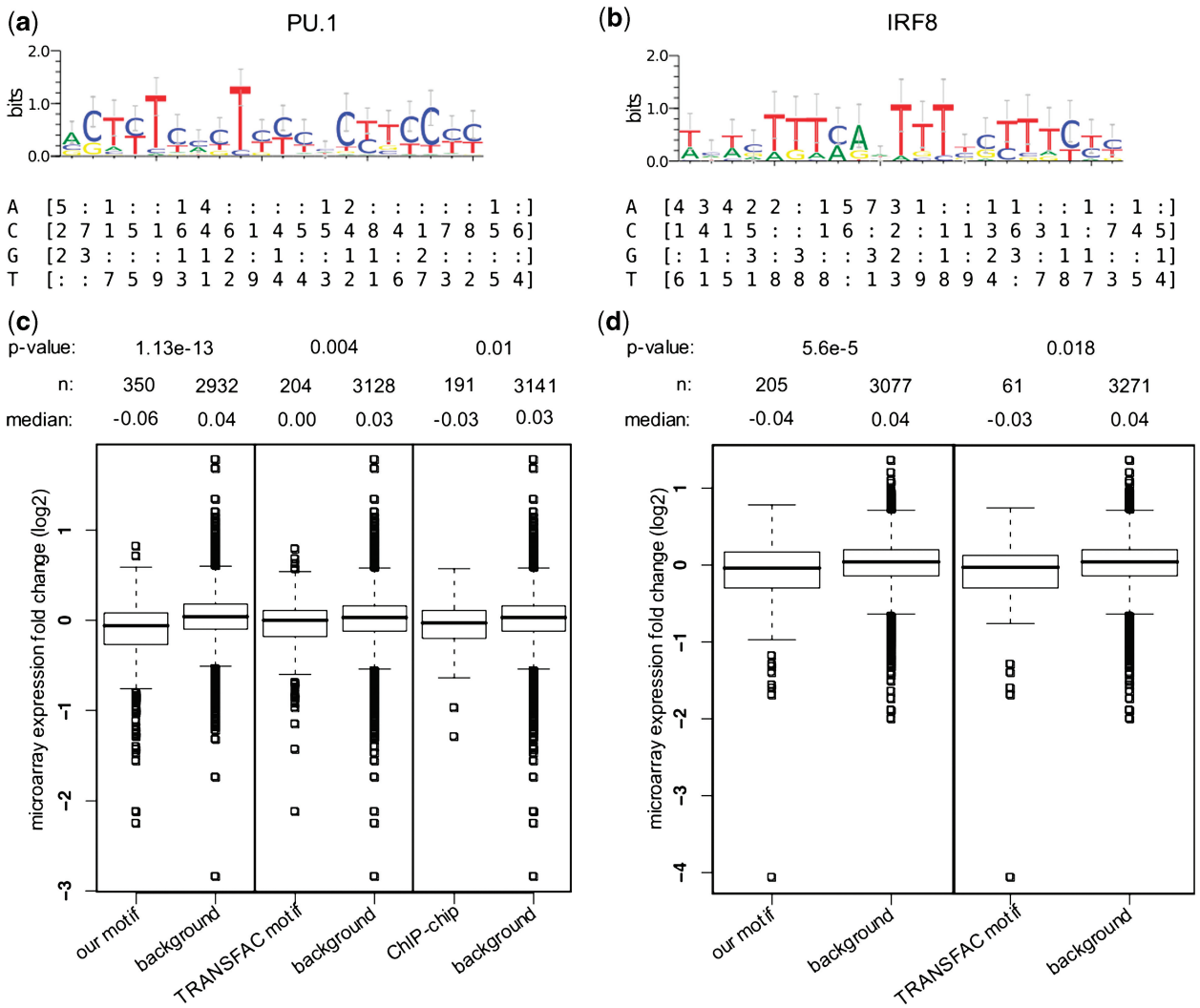


**Figure 3.** TFBS motifs derived for PU.1 and IRF8 as activators. The 50 strongest downregulated TCs after knockdown of each of the two TFs PU.1 and IRF8 and their corresponding promoter regions were used as training data set to identify binding-site motifs and their respective PWMs (**a** and **b**). The PU.1 motif was present in 47 out of 50 sequences with an *E*-value of $4.6e^{-23}$ and is 20 nucleotides wide while the IRF8 motif was present in 20 out of 50 sequences with an *E*-value of $2.2e^{-9}$ and is 21 nts wide. The expression levels of deepCAGE TSSs containing the motif in their promoter sequences excluding the training data were contrasted to all other TSSs (**c** and **d**). The same comparisons were performed on promoter regions containing the TRANSFAC motif as well as for regions where the TFs bound to DNA according to ChIP-chip measurements. *P*-values were calculated using Student's *t*-test on microarrays values.

## A promoter-based gene regulatory network

Above we have demonstrated that KD followed by deepCAGE expression profiling (KD-CAGE) can be effectively used to identify promoters regulated by a given transcription factor. Moreover, highly downregulated promoters in the PU.1 and IRF8 KDs were shown to contain PU.1 and IRF8 motifs indicating they are direct targets of these factors. The approach can thus be used to directly generate a transcriptional network model (24). For illustration purposes, we generated a small sub-network based on genes co-perturbed by the knockdown of at least two of the four factors (Figure 5). Edges upregulated upon knockdown are shown in red and those downregulated are shown in blue. Genes co-regulated by PU.1 and IRF8 were predominantly co-downregulated upon knockdown. Interestingly, there is an antagonistic relationship for genes co-regulated by PU.1 and MYB, with the majority downregulated upon PU.1 KD but upregulated upon MYB KD. The network predicts 47 genes as targets of our novel PU.1 motif. Eight of these (CD74, HCLS1, NRGN, TNFSF13B, IFI6, MLC1, MARCH3 and CHI3L1) are supported by ChIP signal for PU.1 (Supplementary Table S1). Most of these are known to be important in hematopoietic lineages and IFI6 is known to be an interferon-inducible gene. CHI3L1 has been previously reported as a PU.1 target (25). However, this is the first report that TNFSF13B, a myeloid-associated marker gene, is regulated by both PU.1 and IRF8.

These directed edges reflect the regulation of individual TSSs rather than responses at the gene level and represents a powerful new approach to building alternative promoter-aware networks in the near future.

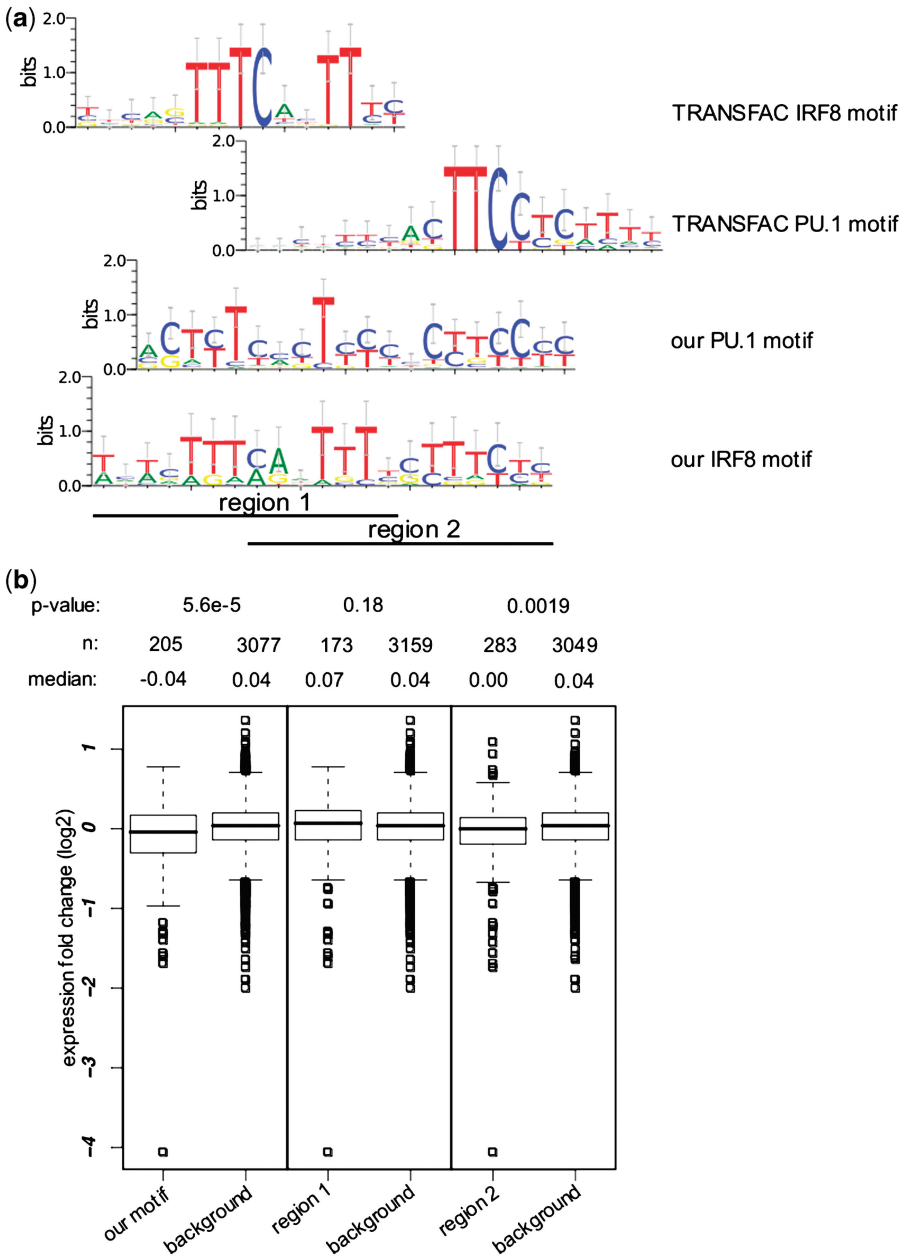


**Figure 4.** Overlapping motifs for PU.1 and IRF8 transcription factors. (**a**) The binding-site motifs we found for IRF8 and PU.1 were longer than the TRANSFAC motifs and both our motifs contained each of the TRANSFAC motifs as sub-motifs. Our motif for IRF8 was longer than the motifs of other IRF family members (data not shown). (**b**) Trimming the characteristic TTT sub-motif from either side of the IRF8 motif reduced the ability of the motif to explain the changes in expression levels. *P*-values were calculated using Student's *t*-test.

## DISCUSSION

We have demonstrated for the first time that deepCAGE technology is a feasible alternative to microarrays for measuring RNAi-mediated perturbations and generating perturbation networks. As the technique is a direct measure of promoter expression, it allows focusing on the actual promoters used in a given cellular context, rather than ambiguous mapping of microarray expression to the 5′-ends of known transcripts. Furthermore, we have shown that our approach can be used to *de novo* identify regulatory motifs with a clear demonstration of functional motifs for PU.1 and IRF8 with similarity to the published TRANSFAC motifs. The motifs described by us perform better at describing the response to the KD than TRANSFAC and ChIP-chip data.

In the case of PU.1 and IRF8, many of the same promoters responded to either knockdown and a longer composite motif was identified. While the known IRF8 TFBS contains two copies of a triple-T motif, ours contains three copies. This longer motif, however, is functionally relevant as truncating the motif by removing the first or third triple-T reduced our ability to explain the transcriptional response to IRF8 knockdown. These observations are supported by the previously reported

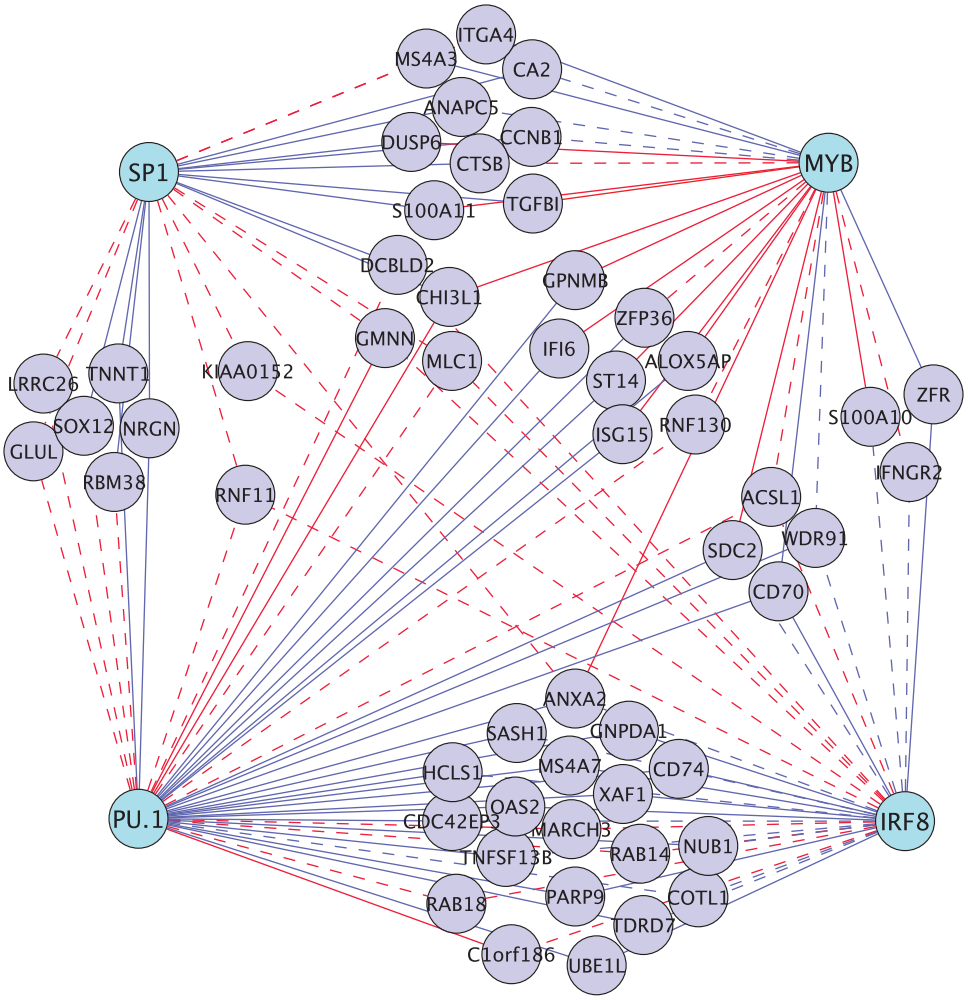


**Figure 5.** Network inferred from deepCAGE knockdown data. Our data can be transferred into network view using Cytoscape (24). The transcription factors represent the nodes and the promoters associated to their genes are the edges. Edges drawn in red indicate upregulation after TF knockdown while edges drawn in blue indicate downregulation. The dotted lines present edges that are detected by CAGE, while solid lines represent the edges that have a motif found by our method. For easier viewing, we have only shown those nodes from the training set that are influenced by more than one transcription factor.

cooperative binding of both factors (20,23). As the significant motifs were identified in the promoters of downregulated genes, we conclude that PU.1 and IRF8 in combination act primarily as activators as previously reported (22), while the motifs observed for MYB suggest it can act both as a repressor or an activator (Supplementary Figure 2A and B).

This pilot experiment paves the way for building regulatory networks and identifying regulatory motifs for the majority of transcription factors. Genome-wide ChIP of TFs is an alternative approach to identify transcriptional regulatory regions (26), which is extensively being used in the ENCODE project (4). However, to date only 160 ChIP grade antibodies are available for the estimated 882 DNA-binding transcription factors in mammals (27). KD-CAGE is not restricted by such reagents, and in the light of constantly reducing costs of DNA sequencing (28) it is possible to test a large collection of all DNA-binding proteins to characterize their function. In addition to the 330 regulatory interactions, we reported in our four knockdown experiments (Supplementary Table S1), only 3 were supported by current ChIP-chip experiments. This highlights that there are sites where the TF is bound but is functionally inactive, as noted by Wasserman and Sandelin (9). However, in spite of this, a combined approach would potentially be a very powerful method to discriminate indirect targets from direct targets bound by factors at both proximal and distal sites including enhancers and insulators.

Finally, we have previously described the application of motif activity response analysis (MARA) in a developmental time course to predict the regulation by TFs on individual promoters (3). However, this approach depends on known TFBS motifs. The approach described here can be used to identify TFBS motifs *de novo*. In the future, we will aim to extend the set of known motifs using this approach and extend our network analyses to encompass the function and targets of uncharacterized DNA-binding proteins and to provide a network of interactions among such proteins.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

Mr Akira Hasegawa assisted in the alignment of the deepCAGE tags. P.C. developed the deepCAGE technology. T.L. conceived perturbation deepCAGE. M.V. and T.L. designed the experiments and carried out the motif analyses and network building. ARRF carried out the microarray analysis and Entrez gene mapping. Y.T. and M.S. carried out the knockdowns. T.L., M.V., A.R.R.F. and C.D. wrote the manuscript. H.S., Y.H. and C.D. advised on the experimental design. All authors read and approved the final manuscript.

## FUNDING

Research Grant for RIKEN Omics Science Center from Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to Y.H.); International Program Associate stipend from RIKEN (to M.V.). Funding for open access charge: Research Grant for RIKEN Omics Science Center from Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to Y.H.).

*Conflict of interest statement*. None declared.

## REFERENCES

1. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing, analysis of the human genome. *Nature*, **409**, 860–921.
2. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
3. Suzuki,H., Forrest,A.R., van Nimwegen,E., Daub,C.O., Balwierz,P.J., Irvine,K.M., Lassmann,T., Ravasi,T., Hasegawa,Y., de Hoon,M.J. *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
4. Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigo,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
5. Quackenbush,J. (2007) Extracting biology from high-dimensional biological data. *J. Exp. Biol.*, **210(Pt 9)**, 1507–1517.
6. Segal,E., Shapira,M., Regev,A., Pe'er,D., Botstein,D., Koller,D. and Friedman,N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, **34**, 166–176.
7. Carninci,P., Sandelin,A., Lenhard,B., Katayama,S., Shimokawa,K., Ponjavic,J., Semple,C.A., Taylor,M.S., Engström,P.G. and Frith,M.C. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
8. Tomaru,Y., Simon,C., Forrest,A.R., Miura,H., Kubosaki,A., Hayashizaki,Y. and Suzuki,M. (2009) Regulatory interdependence of myeloid transcription factors revealed by Matrix RNAi analysis. *Genome Biol.*, **10**, R121.
9. Wasserman,W.W. and Sandelin,A. (2004) Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.*, **5**, 276–287.
10. Sikder,D. and Kodadek,T. (2005) Genomic studies of transcription factor-DNA interactions. *Curr. Opin. Chem. Biol.*, **9**, 38–45.
11. Kodzius,R., Kojima,M., Nishiyori,H., Nakamura,M., Fukuda,S., Tagami,M., Sasaki,D., Imamura,K., Kai,C., Harbers,M. *et al.* (2006) CAGE: Cap analysis of gene expression. *Nat. Methods.*, **3**, 211–222.
12. Ponjavic,J., Lenhard,B., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. and Sandelin,A. (2006) Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol.*, **7**, R78.
13. Valen,E., Pascarella,G., Chalk,A., Maeda,N., Kojima,M., Kawazu,C., Murata,M., Nishiyori,H., Lazarevic,D., Motti,D. *et al.* (2009) Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.*, **19**, 255–265.
14. Tsuchiya,S., Yamabe,M., Yamaguchi,Y., Kobayashi,Y., Konno,T. and Tada,K. (1980) Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int. J. Cancer.*, **26**, 171–176.
15. Bailey,T.L., Williams,N., Misleh,C. and Li,W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, **34(Web Server issue)**, W369–W373.
16. Grundy,W.N., Bailey,T.L., Elkan,C.P. and Baker,M.E. (1997) Meta-MEME: Motif-based Hidden Markov Models of Biological Sequences. *Comput. Appl. Biosci.*, **13**, 397–406.
17. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I.L., Barre-Dirrie,S., Reuter,A., Chekmenev,I., Krull,D., Hornischer,M.K. *et al.* (2006) TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34(Database issue)**, D108–D110.
18. Kel,A.E., Gößling,E., Reuter,I., Cheremushkin,E., Kel-Margoulis,O.V. and Wingender,E. (2003) MATCHTM: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.*, **31**, 3576–3579.
19. de Hoon,M. and Hayashizaki,Y. (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques*, **44**, 627–628, 630, 632.
20. Meraro,D., Gleit-Kielmanowicz,M., Hauser,H. and Levi,B.Z. (2002) IFN-stimulated gene 15 is synergistically activated through interactions between the myelocyte/lymphocyte-specific transcription factors, PU.1, IFN regulatory factor-8/IFN consensus sequence binding protein, and IFN regulatory factor-4: characterization of a new subtype of IFN-stimulated response element. *J. Immunol.*, **168**, 6224–6231.
21. Lüscher,B. and Eisenman,R.N. (1990) New light on Myc and Myb. Part II. Myb. *Genes Dev.*, **4**, 2235–2241.
22. Frith,M.C., Fu,Y., Yu,L., Chen,J.F., Hansen,U. and Weng,Z. (2004) Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.*, **32**, 1372–1381.
23. Marecki,S. and Fenton,M.J. (2000) PU.1/Interferon Regulatory Factor interactions: mechanisms of transcriptional regulation. *Cell Biochem. Biophys.*, **33**, 127–148.
24. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
25. Rehli,M., Niller,H.H., Ammon,C., Langmann,S., Schwarzfischer,L., Andreesen,R. and Krause,S.W. (2003) Transcriptional regulation of CHI3L1, a marker gene for late stages of macrophage differentiation. *J. Biol. Chem.*, **278**, 44058–44067.
26. Pillai,S. and Chellappan,S.P. (2009) ChIP on chip assays: genome-wide analysis of transcription factor binding and histone modifications. *Methods Mol. Biol.*, **523**, 341–366.
27. Fulton,D.L., Sundararajan,S., Badis,G., Hughes,T.R., Wasserman,W.W., Roach,J.C. and Sladek,R. (2009) TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.*, **10**, R29.
28. Service,R.F. (2006) GENE SEQUENCING: the race for the $1000 Genome. *Science*, **311**, 1544–1546.