Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Genomics 88 (2006) 133–142
GENOMICS
www.elsevier.com/locate/ygeno

Transcriptional network dynamics in macrophage activation

Roland Nilsson [a,b], Vladimir B. Bajic [c,d], Harukazu Suzuki [e], Diego di Bernardo [f], Johan Björkegren [a,b], Shintaro Katayama [e], James F. Reid [g], Matthew J. Sweet [h], Manuela Gariboldi [g], Piero Carninci [e,i], Yosihide Hayashizaki [e,i], David A. Hume [e,h,*], Jesper Tegner [a,b,*], Timothy Ravasi [e,j,*]

[a] *Center for Genomics and Bioinformatics, Karolinska Institutet, Stockholm, Sweden*
[b] *Computational Biology, IFM, Linkoping University, Linkoping, Sweden*
[c] *Knowledge Extraction Laboratory, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, 119613, Singapore*
[d] *University of Western Cape, Private Bag X17, SANBI, Bellville 7535, South Africa*
[e] *Genome Exploration Research Group (Genome Network Project Core Group), RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*
[f] *Téléthon Institute for Genetics and Medicine, Naples, Italy*
[g] *Department of Experimental Oncology, Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan, Italy*
[h] *ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane QLD 4072, Australia*
[i] *Genome Science Laboratory, Discovery Research Institute, RIKEN Wako Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*
[j] *Department of Bioengineering, Jacobs School of Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

Received 12 October 2005; accepted 25 March 2006
Available online 12 May 2006

Abstract

Transcriptional regulatory networks govern cell differentiation and the cellular response to external stimuli. However, mammalian model systems have not yet been accessible for network analysis. Here, we present a genome-wide network analysis of the transcriptional regulation underlying the mouse macrophage response to bacterial lipopolysaccharide (LPS). Key to uncovering the network structure is our combination of time-series cap analysis of gene expression with in silico prediction of transcription factor binding sites. By integrating microarray and qPCR time-series expression data with a promoter analysis, we find dynamic subnetworks that describe how signaling pathways change dynamically during the progress of the macrophage LPS response, thus defining regulatory modules characteristic of the inflammatory response. In particular, our integrative analysis enabled us to suggest novel roles for the transcription factors ATF-3 and NRF-2 during the inflammatory response. We believe that our system approach presented here is applicable to understanding cellular differentiation in higher eukaryotes.

Keywords: System biology; Regulatory networks; Network dynamics; Complex systems; Transcriptional regulation; Genome; Macrophages; Innate immunity; inflammation

Macrophages are the classical two-edged sword of the innate immune system. Their ability to recognize and destroy microorganisms is essential to host defense, and they have many roles in development, wound healing, and homeostasis; yet their destructive potential and secretory products are central to the pathology of acute and chronic inflammatory disease in mammals [1,2]. The destructive potential of macrophages is stringently controlled. Recognition of conserved nonself molecules expressed by microorganisms is mediated by so-called pattern recognition receptors, many of which belong to the Toll-like receptor family. The most studied of these receptors is Toll-like receptor 4 (TLR4), which mediates signals generated by lipopolysaccharide (LPS), a major component of the cell walls of gram-negative microorganisms. In response to

* Corresponding authors. T. Ravasi is to be contacted at Department of Bioengineering, Jacobs School of Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Fax: +1 858 822 4246.
E-mail address: travasi@bioeng.ucsd.edu (T. Ravasi).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.03.022

LPS, mouse macrophages undergo a major change in gene expression, in particular inducing the expression and release of numerous biologically active cytokines that orchestrate the inflammatory response. An extensive literature search revealed several hundred genes that were reported to be inducible in macrophages [3]. The LPS response in mouse macrophages, which is reflected in both morphology and gene expression patterns, has been analyzed on a number of different platforms [4,5]. Temporal profiling reveals a cascade of gene regulation, with many late-inducible genes responding to inducible transcription factors and/or inducible secreted regulators acting in an autocrine manner. This well-characterized, stereotypical response is ideal for identifying and understanding dynamic transcriptional networks [6,7]. Yet, the resolution of the networks that can be inferred directly from microarray data is limited, partly because transcription factors are often expressed below the detection limit of microarrays [8].

In this study, we have assembled a variety of data on the transcriptional response of murine bone marrow macrophages (BMMs) to LPS stimulation over time. BMMs actively proliferate in response to the lineage-specific growth factor macrophage colony-stimulating factor 1 (CSF-1). LPS down-regulates the CSF-1 receptor on the cells and causes growth arrest while at the same time promoting survival [9]. A substantial number of genes that are inducible by LPS in BMMs are actually induced by the inactivation of a repressive signal from the CSF-1 receptor. Hence, LPS signaling intersects with numerous fundamental biological events, such as proliferation, apoptosis, endocytosis, and secretion, common to most mammalian cells.

Data integration and perturbation are essential for recovering regulatory networks since, despite recent progress in identifying networks from genome-wide data in *Escherichia coli* [10,11] and yeast [6,12,13], it has not yet been possible to provide a reliable detailed map of the underlying regulatory transcriptional architecture. In addition, conventional clustering of coexpressed genes does not have sufficient resolution to detect regulatory interactions between genes. Several recent studies have demonstrated the advantage of integrating various data types obtained by high-throughput methods such as genome-wide expression profiling, genome-wide RNA interference, and chromatin immunoprecipitation complemented with promoter DNA microarray (ChIP-on-chip) [14]. In particular, Luscombe et al. [6] used transcription factor binding data from yeast to infer a "passive" network, whereas genome-wide expression data sampled from different states of the cell cycle defined the corresponding "active" subnetworks. However, their approach is limited to systems in which transcription factor binding experiments are feasible. This is not yet the case for mammalian systems, including the macrophage. Several studies have considered sequence-based promoter information, building on the belief that coexpressed genes are more likely to be coregulated by similar sets of transcription factors (TFs) [15,16]. However, genes in a coexpression cluster need not be coregulated by the same underlying mechanism. Two transcripts can have similar expression profiles and yet be regulated by different factors. An additional complication recognized by the FANTOM3 analysis of mouse promoters on a genome-wide scale is that the large majority of "genes" have more than one promoter with quite distinct regulation [17]. So, an arbitrary extraction of promoters based upon the sequence upstream of the longest known cDNA can combine distinct promoters with discordant regulation.

Here we instead design a method for discovering transcriptional networks active in a particular cell state, based on prediction of state-specific transcription factor binding sites (TFBSs), which are defined by experimentally validated transcription start sites. We present a novel algorithm for clustering transcripts based on similarity of promoter structure instead of coexpression. This analysis shows that genes within these clusters structures tend to be coexpressed and functionally related. Finally, we illustrate that our network inference method recovers many known features of the macrophage transcriptional response to LPS, and we discuss a number of novel findings. Although several studies attempted to link pairs of transcription factors to coregulation and coexpression in reconstruction of regulatory networks [12], we took the analysis one step further by determining coregulated genes based on similarities of promoter structures. Our approach can therefore account for complex combinatorial control by several transcription factors.

Results

Experimental system

We used LPS to activate murine BMMs. Similar to previous small-scale studies [5], we collected gene expression data (described below) monitoring the LPS response over a time course of 0 (before LPS stimulation), 2, 7, and 24 h.

Macrophage transcriptome analysis

To monitor whole-genome expression during the LPS response, we used RIKEN cDNA arrays [17] containing over 60,000 probes. Triplicate array hybridizations were used at each time point. In addition, we performed quantitative real-time PCR (qPCR) of 1559 known and predicted transcripts coding for TFs and other putative nuclear proteins, since these are often expressed in amounts below the detection limit of microarrays [8]. Their putative roles in regulatory networks have therefore not yet been examined directly on a global scale.

The qPCR analysis revealed that 43% (673/1559) of TFs were significantly expressed and regulated (Supplementary Table 1). The diversity of TFs detected in macrophages is rather striking. We will not review the data herein, but the list contains all of the factors previously identified based upon a literature survey of known macrophage-expressed and/or LPS-inducible TFs [18]. It is also interesting to consider the TFs that were not detected, including most members of known TF families involved in lineage determination and patterning in embryonic development (e.g., Hox, Sox, GATA, Tbx, Neurog, Nkx, Lhx, and Fox). That is, any TF of unknown function that is not present in macrophages in any state of activation is likely to be

expressed in a defined window of embryonic development and control some aspect of cell lineage specification.

Of the 60,000 nonredundant array spots present on the RIKEN cDNA array chips 2219 (4%) were found to be above the detection limit and significantly changed over the time course after LPS stimulation (see Materials and methods for details). The data reported here are in accordance with previously published results on gene expression in macrophages in response to LPS stimulation [4,5]. In total, 2892 measured transcripts were used for the subsequent analysis after combining the qPCR and microarray datasets. These are listed in Supplementary Table 2.

To identify clusters of coexpressed genes, we performed hierarchical clustering on expression data from the selected set of transcripts (Fig. 1).

Four well-separated clusters of genes were identified: "down," down-regulated after LPS stimulation; "early," up-regulated at 2 h; "middle," up-regulated at 7 h; and finally, "late," up-regulated at 24 h. Other clusters resembling middle and late may be discerned at the bottom of Fig. 1 and above "late," respectively. However, we chose not to include these in the subsequent analysis since they deviated from middle and late in the expression distance metric, as can be seen by examining the cluster tree. Manual inspection of the corresponding gene lists revealed that the down cluster contains several cell-cycle-related genes, confirming previous evidence of the antiproliferative effect of the LPS [19]. In this cluster we also note a substantial enrichment of TFs (56% of the total set) (Fig. 2 and Supplementary Table 3). Most studies of transcriptional regulation in macrophages focus only on genes that are induced; but in many cases, the same outcome may be achieved, be dependent upon, or be complemented by the relief of repression by the numerous negative regulators of the LPS response. In the early cluster we found mainly genes involved in activation of the inflammatory response at different levels, like those of the MAP kinase signal cascade pathways. For instance, MAP3K8 is strongly up-regulated at this early stage. MAP3K8 has previously been linked to the posttranscriptional activation of tumor necrosis factor α (TNF-α, the archetype of proinflammatory cytokines) [20]. Candidate regulators of later gene responses in the transcriptional network can be seen in early up-regulation of TFs such as IRF-1, ETV-3, and Hivep3 [21,22] (Fig. 2 and Supplementary Table 4). Similarly, the middle cluster contains many TFs, reflecting a cascade effect on the network dynamics, but here we also detect production of proinflammatory cytokines and chemokines. We can also observe a change in the Toll-like receptor repertoire on the plasma membrane (down-regulation of TLR4 in the early cluster and up-regulation of TLR1 and TLR7), thus providing evidence of a staged response of the regulatory networks to complex pathogen-associated molecular patterns [23,24] (Fig. 2 and Supplementary Table 5). The late cluster contains more effectors genes of the innate immune response. In particular, we find an enrichment in lysosomal proteins such as members of the cathepsin family (cathepsins B, H, and Z), probably reflecting the preparation of the phagosome [25,26]. We also see enrichment in components of other intracellular compartment proteins associated with

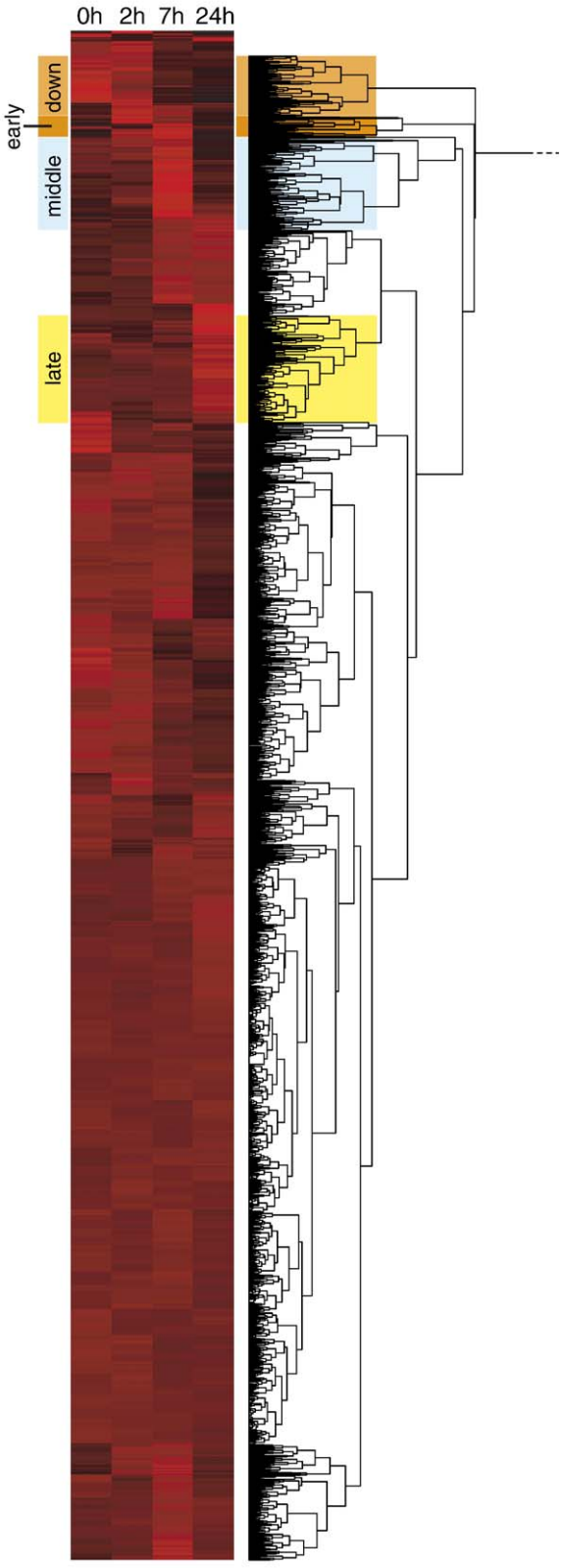


Fig. 1. Time-series expression clustering. Heat map of normalized expression data for 2892 genes (rows) across four time points (columns), together with the dendrogram obtained from hierarchical clustering of these data. Clusters named "down," "early," "middle," and "late" in the main text are highlighted accordingly. A set of outlier genes that cluster separately (topmost rows) is displayed without dendrogram.

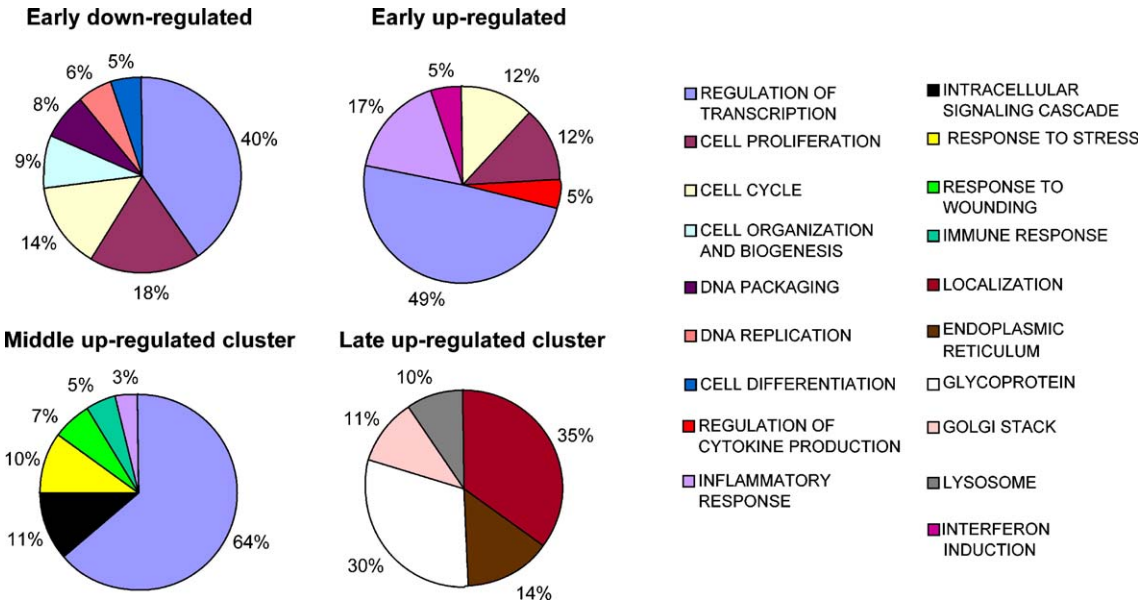


Fig. 2. Gene Ontology annotation of the four main regulated gene expression clusters. Gene Ontology analysis of the four expression clusters defined in Fig. 1 is presented. The annotations are based on the primary accession numbers as shown in the supplementary material (Tables 1, 2, 3, and 4). We used the Database for Annotation Visualization and Integration Discovery version 2.1 (DAVID 2.1, http://apps1.niaid.nih.gov/david/).

secretion, such as Golgi apparatus and endoplasmic reticulum and vesicle trafficking components (Fig. 2 and Supplementary Table 6). This part of the response has been examined in detail elsewhere [27]. From those studies, we identified syntaxin 6 and vti1b as components of a novel SNARE complex that is up-regulated in activated macrophages to facilitate exocytosis of TNF-α [27]. Taken together, these results validate our expression data and cluster analysis.

The TFBS analysis

To set the stage for a network analysis, we identify promoters by designing macrophage-specific cap analysis of gene expression (CAGE) libraries [28,29] for each time point after LPS stimulation. A detailed analysis of these CAGE libraries is presented elsewhere [17]. For our purposes, the CAGE may be viewed solely as a means to locate precisely the promoter sequence of each transcript. Of the 2892 transcripts considered, 1784 were detected as present by CAGE. For these, we extracted the corresponding promoter sequences and performed prediction of TFBSs based on Transfac matrix models (see Materials and methods). With this method, we identified 298 distinct TFs with at least one associated TFBS in at least one promoter. In total, over 30,000 TFBSs were identified by this procedure, with the most prevalent TFBSs being present in more than half of the transcripts. However, a large fraction of these are likely to be false positives from the prediction software or nonspecific interactions that are not particularly important to the transcriptional regulation program of the LPS response. We therefore filtered these data by removing predicted binding sites that were frequently found also in a background set of 40,000 random mouse promoters (see Materials and methods). After filtering, we retained 3824 TF–promoter interactions. This leaves us with a set of interactions that are highly specific to the LPS response.

To validate our TFBS predictions, we first asked whether transcripts with similar promoter content are also coexpressed. To investigate this issue, we defined a measure d of dissimilarity between any two promoters, inversely related to the number of TFBSs common to the two promoters (see Materials and methods). We found that promoters with small d have significantly higher expression correlation than would be expected by chance ($p < 10^{-3}$, permutation test, see Materials and methods).

Using the dissimilarity measure d, we performed hierarchical clustering of the 1784 promoters (Fig. 3). This creates a partition of genes, *promoter clusters*, based on actual promoter structure rather than expression patterns. To test the reliability of this method, we investigated a number of functional gene groups that are known to be coregulated. For example, the ribosomal genes were found to be significantly close together in the promoter clustering tree ($p = 0.001$; see Supplementary Table 7 for remaining groups). In this sense the promoter clustering outperformed even conventional (expression-based) clustering in some cases, for example, the TAF transcription factors, which are significantly close according to the TFBS metric ($p = 0.002$) but not based on the expression metric ($p = 0.42$).

We manually selected 17 promoter clusters that were well separated and of reasonable size. Of these, clusters 1, 6, 8, 9, and 10 were deemed to be of great interest for the LPS response, shown in Fig. 3. An important advantage of the clustering method based on promoter content compared to expression clustering is that each cluster centroid provides insight into the specific regulation mechanism of that cluster (Fig. 3). For example, we readily find that cluster 1 is strongly regulated by TFs from the ETS family and contains several transcripts that are highly enriched in macrophages, confirming the important role of the ETS family of transcription

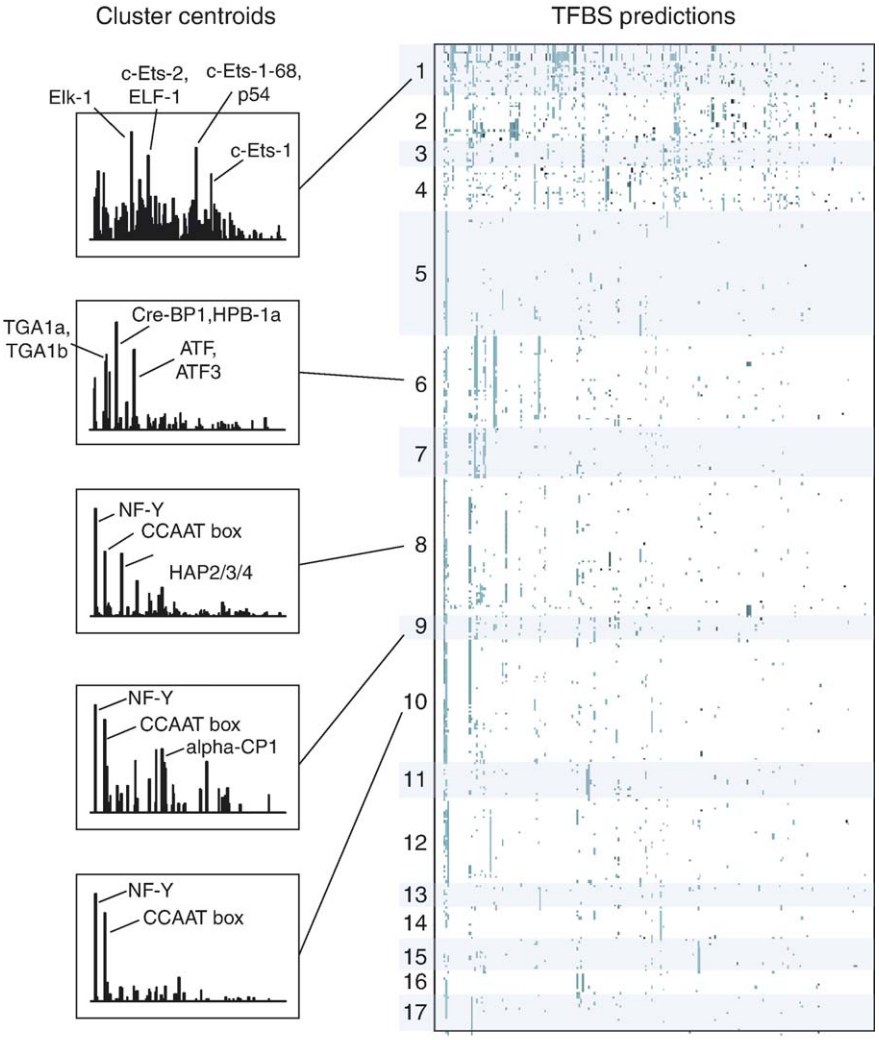


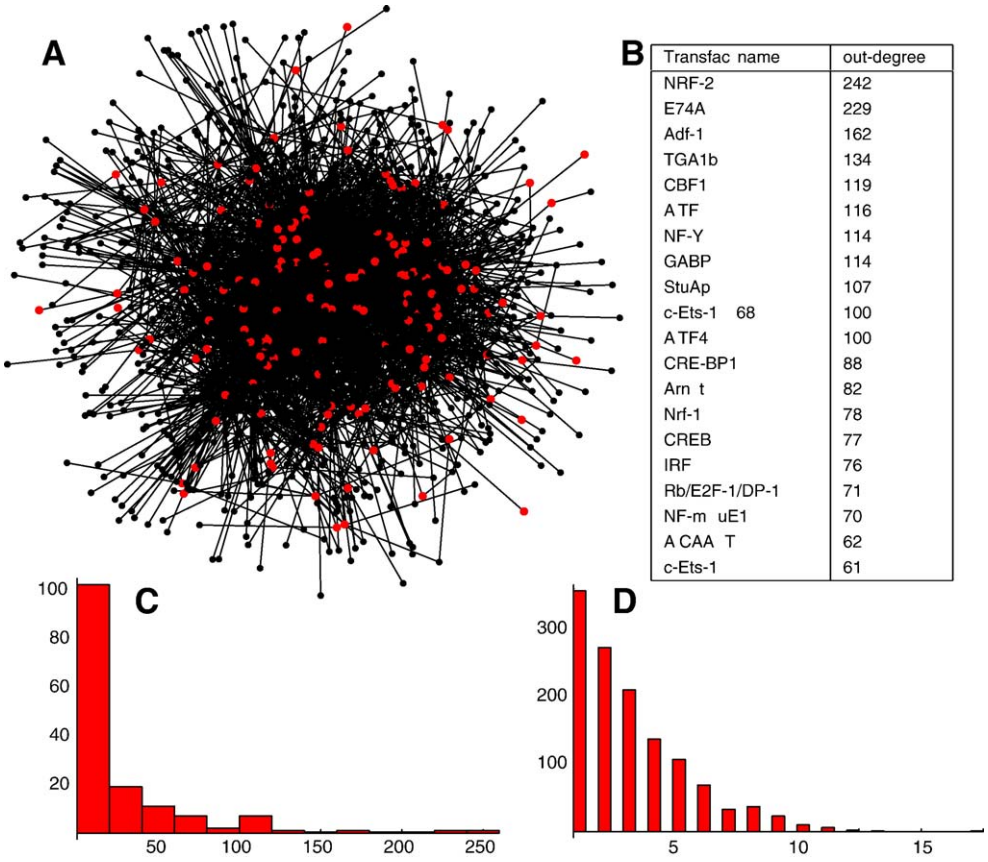
Fig. 3. Promoter clusters. Right: A heat map showing the predicted transcription factor binding sites for each transcript (rows) and each transcription factor (columns). Dark colors mean highly specific sites. A selection of 17 clusters obtained by clustering (see Materials and methods) is shown here. Left: Cluster centroids for clusters 1, 6, 8, 9, and 1, shown as bar charts, each bar representing the mean of the corresponding heat map rows. Some important transcription factors regulating cluster 1 are highlighted.

factors in macrophage differentiation and regulation [30,31], while cluster 6 is regulated by members of the ATF family, an important player during the inflammatory response [32]. Taken together, this is strong evidence that our approach can identify functional and expression-related clusters based on promoter structure.

LPS network

To infer a gene regulatory network from the TFBS predictions, we connected each TF to each transcript that has a predicted binding site for that TF (Fig. 4A). Due to our selection of differentially expressed transcripts and the filtering procedure described above, we obtain a specific *LPS network*, enriched for connections that are active predominantly during the LPS response. This is important, as it is known from previous studies on yeast that different *subnetworks* of the total regulatory network will be active in different cellular states [6]. Our method is able to extract such subnetworks from any given experimental condition. In the LPS network, ubiquitous parts of the regulatory machinery, such as TF motifs required for basal transcription, are suppressed. For example, the CCAAT box, which appeared in approx 20% of all promoters in the background set, was not found to be sufficiently overrepresented in the target set and thus did not appear in the LPS network.

The global characteristics of the LPS network are shown in Fig. 4. In agreement with previous studies [7], Fig. 4C shows a broad-tailed distribution of out-degrees (number of transcripts each TF regulates). The most regulating TF, NRF-2, is predicted to regulate 242 transcripts (Fig. 4B). This TF is known to coordinate protection against oxidative stress, an essential part of the macrophage LPS response as macrophages themselves must produce oxidative agents to combat pathogens [33,34]. The fact that NRF-2 has very high out-degree in the LPS network suggests that protection against oxidative stress is a major task of the LPS response. This prediction was recently



Transfac name	out-degree
NRF-2	242
E74A	229
Adf-1	162
TGA1b	134
CBF1	119
A TF	116
NF-Y	114
GABP	114
StuAp	107
c-Ets-1 68	100
A TF4	100
CRE-BP1	88
Arn t	82
Nrf-1	78
CREB	77
IRF	76
Rb/E2F-1/DP-1	71
NF-m uE1	70
A CAA T	62
c-Ets-1	61

Fig. 4. LPS network properties. (A) Global view of the LPS network reveals transcription factors (red) as central nodes. (B) The 20 most regulating transcription factors. (C) Out-degree (number of regulated transcripts) distribution for the LPS network. (D) In-degree (number of regulating transcription factors) distribution for the LPS network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

independently validated by the groups of Ishii and Iizuka [35,36] using NRF-2 knockout mice.

The in-degree (number of TFs each transcript is regulated by) distribution is more short-tailed (Fig. 4D). The out- and in-degree parameters reflect a scale-free property of the macrophage transcriptional network, a recurrent theme in many biological networks [7,37].

Among the top 200 most regulated transcripts, 43 (22%) themselves code for TFs, suggesting a complex cascade of regulatory interactions in the LPS response.

Dynamics of the LPS network

Combining the TFBS analysis with expression dynamics, we can discern parts of dynamics of the LPS network. Fig. 5 shows a subset of the LPS network containing the transcripts in the four expression clusters displayed in Fig. 1. As the transcriptional program progresses over time, we can discern how each individual TF assumes different regulatory roles, participating in the various stages of the LPS response in different combinations. This observation, once more, suggests an important combinatorial role for collections of TFs in complex biological responses.

To validate the dynamic network, we selected three well-known TFs from the network in Fig. 5A and analyzed their local networks in detail. Activating transcription factor-3 (ATF-3) is a transcriptional repressor that has previously been shown to be important for the regulation of several cytokines [38]. In the dynamic network, ATF-3 is one of the most connected TFs, interacting with several transcripts involved in cell cycle progression and DNA replication (Mcm2, Mcm5, Cfk2, and AQR) (Fig. 5B). ATF-3 is transiently up-regulated in the LPS response at 2 h after LPS stimulation, with peak activity at 7 h, whereas the expression of the above-mentioned interacting transcripts is down-regulated at 7 h. Combined with this information on transcriptional activity, the ATF-3 subnetwork in Fig. 5B leads us to infer a novel role for ATF-3 as a key regulator of the antiproliferative effects of macrophage LPS stimulation.

In a similar fashion, analysis of the ETS-2 subnetwork confirms the active role of ETS-2 in regulating the inflammatory response in macrophages. Of particular relevance here is the positive activation of inflammatory mediators such Cias-1, Lcp, Ifi204, and Ltb4dh (Fig. 5B).

NRF-2 is known to be involved in the regulation of the oxidative stress response [36,39]. Our local network analysis confirms this, but also suggests a more general role for NRF-2 in the inflammatory response, since NRF-2 is found to regulate not only oxidative stress-related genes, but also other genes that have previously been implicated in the inflammatory response (Fig. 5B).

Keap1 is a known negative regulator of NRF-2 [40]. Interestingly, in our dynamic network Keap1 is itself predicted

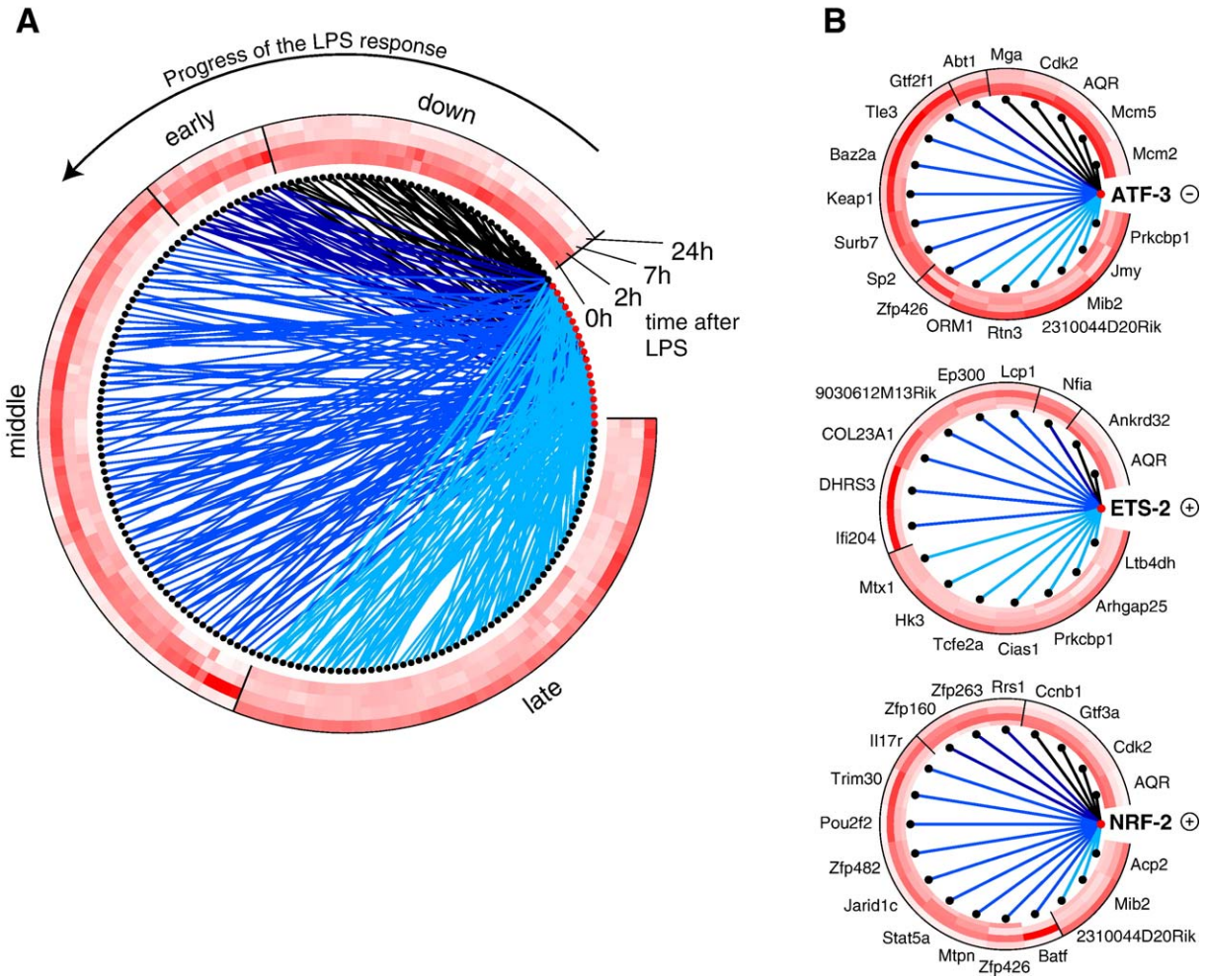


Fig. 5. LPS network dynamics. (A) Subset of the LPS network for the four expression clusters described in Fig. 1. Center, network plot with black dots representing transcripts and red dots transcription factors; lines show predicted connections, color-coded according to the cluster regulated (down, early, middle, late, as defined in the text). Clusters are arranged to show the LPS response progress counterclockwise (upper arrow). Surrounding the network plot is an expression heat map for the microarray time-course data for each transcript for 0, 2, 7, and 24 h, radially from center and outward. (B) Network plots similar to (A), showing the inferred gene regulation for the transcriptions factors ATF-3, ETS-2, and NRF-2. A plus sign indicates a positive regulation or transcriptional activation, whereas a minus sign indicates either negative regulation or transcriptional repression. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to be repressed by ATF-3. Again, the expression data corroborate this prediction, as Keap1 is repressed during the early stage of LPS activation (when ATF-3 is active) and subsequently up-regulated at 7 h. This leads to the hypothesis of a possible interplay between the ATF-3 and the NRF-2 regulatory circuits shown in Fig. 5B. In this way, we are able to elucidate how the inflammatory response is orchestrated by dynamic and stage-specific transcriptional networks over time.

Discussion

Our analysis of the global properties of the LPS network adds to the existing body of evidence [7] for universal organization principles of complex systems (Fig. 4). The out-degree distribution of the LPS network shows a characteristic long tail, resembling the power-law distribution that is evident in all sorts of complex systems, including protein–protein interaction maps, social networks, and power grids [41,42]. This characteristic of the macrophage network allows us to define the most connected TFs (the "hubs") of the network (Fig. 4B). Among these we found macrophage-specific TFs such as the ETS family and several TFs with central roles in inflammation, such as the interferon regulatory factors. Each of these hubs is predicted to be central to some aspect of the LPS response and suggests candidates for experimental validation through knock-outs or overexpression.

Although we did find a significant correlation between promoter structure and expression, this dependence was comparatively weak. This is not surprising, as coexpression can have causes other than coregulation. There are a number of caveats to our analysis and the available data. First, both promoter-based and expression-based clustering are constrained by limits on detection that are quite distinct. Therefore, genes expressed at relatively low levels may be detected to different extents by the two approaches. Second, we have chosen an arbitrary window of 700 bp (−500, +200 relative to

transcription start sites) to analyze for TFBS. In many cases, this window may contain more than one functional promoter. For example, the 500-bp region upstream of the macrophage promoter in the Csf1r gene also contains a completely independent trophoblast promoter with a distinct set of trophoblast-specific TFBSs [43]. Third, the TFBS analysis used is inherently limited to binding sites described by existing Transfac matrix models, rendering the network incomplete and biased toward more well-known TFs. An extension of this work would be to test *ab initio* detection of shared sequence motifs in the macrophage-expressed gene sets. Furthermore, Transfac matrices commonly contain binding sites for multiple members of gene families, and one cannot be certain as to which member is able to bind with highest affinity nor whether the site is available to be occupied in native chromatin structures. Clearly, ChIP-on-chip data for each of the candidate TFs identified as hubs in our analysis would greatly increase the confidence of the dataset and corresponding network analysis. In addition, our dissimilarity metric may be refined to take into account the position and orientation of sequence motifs; this has previously been shown to increase correlation with expression [44].

Yet, our combined approach convincingly demonstrates the "first principle" in transcription initiation, since transcripts clustered based on promoter structure are also significantly coexpressed. We find clear examples of coexpressed modules in macrophage activation; modules enriched in functional categories (Fig. 2) and composed of transcripts previously associated with the inflammatory response. By association, we also infer the function of a set of novel transcripts. At the same time, for each such module we have generated a hypothetical regulatory program in terms of the TFs that participate in activation of the involved promoters, as detailed in Figs. 2 and 5. We were able to predict a novel function for the transcriptional repressor ATF-3 as a key regulator of the antiproliferative effects of macrophage LPS stimulation and also a pivotal role for NRF-2 and ETS-2 in orchestrating the LPS response.

In summary, we have performed the first dynamic analysis of the macrophage activation network. We have demonstrated that by combining several genome-wide datasets with novel bioinformatics approaches we can readily identify cell-state-specific gene regulatory networks in mammalian systems. In particular, in the macrophage, this approach can be used to study the differences in regulatory networks underlying activation by different pathogens products, developmental stages, and macrophage populations.

Materials and methods

Bone marrow-derived macrophages and LPS treatment

Generation of BMMs was performed as described previously [45]. Briefly, BMMs were cultured in complete RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% FCS (JRH Biosciences, Inc., Lenexa, KS, USA), 20 U/ml penicillin (Invitrogen), 20 μg/ml streptomycin (Invitrogen), and 2 mM L-glutamine (Invitrogen). The primary cells were maintained in a 37°C incubator containing 5% CO_2. LPS from *Salmonella minnesota* (Sigma–Aldrich, St. Louis, MO, USA) was sonicated in 0.1% triethylamine and then further diluted in PBS.

Macrophages were seeded at 1×10^7 cells/ml and incubated with 10 ng/ml LPS. Three 10-cm dishes were harvested for each time point—unstimulated (time 0, 2, 7, and 24 h). Total RNA from BMMs was extracted using Qiagen RNeasy Mini kits according to the manufacturer's protocols. In all RNA samples, contaminating genomic DNAs were removed by digestion of the DNA in the RNA samples with an RNase-free DNase I (DNA*free* kit from Ambion) performed according to Ambion's methods.

Microarray data generation and analysis

The experimental design, using the 17.5-dpc C57BL/6J embryo as a common reference, has been described previously [46]. Briefly, RNA was labeled indirectly with aminoallyl-conjugated Cy3 (time course) or Cy5 (embryo) and hybridized overnight to RIKEN 60,000 full-length cDNA microarrays [46]. Each hybridization was done in triplicate and the data are presented as average values. Slides were washed and scanned on a ScanArray 5000 confocal laser scanner. Molecularware (Digital Genome) was use to process the images, and data were corrected for local background and confidence status was flagged for empty spots, signal/noise ratio, spot CV ratio, and spot morphology.

The average ratio of the three triplicate experimental signals to the control signal for each spot was calculated; intensity-dependent normalization was also applied, by which the ratio was reduced to the residual of the Lowess fit of the intensity versus ratio curve. The dataset was restricted to those spots passing confidence status (as defined above) on the control channel of every hybridization. This matrix was imported to the software tool BRB-ArrayTools version 3.1 (developed by Dr. Richard Simon and Amy Peng, Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute; http://linus.nci.nih.gov/BRB-ArrayTools.html).

Filtering

No missing values were allowed and all control clones were removed.

Identification of clones differentially expressed over time

To identify statistically significant differentially expressed clones over time we performed paired *t* tests by comparing the following time steps: 0–2, 0–7, 0–24, 2–7, 2–24, and 7–24 h. A single list representing all the transcripts differentially expressed in at least one of these comparisons was obtained by taking the union of the lists.

Genes were considered statistically significant if their p value was less than 0.001. A stringent significance threshold was used to limit the number of false-positive findings; we used the multivariate permutation test [47] to provide 75% confidence that the number of false-positive genes did not exceed 10.

Transcription factors quantitative real-time PCR

Gene-specific primer pairs were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), with an optimal primer size of 20 bases, amplification size of 140 bp, and annealing temperature of 60°C for a set of 1559 proteins designated as transcription factors by either Gene Ontology or conserved domain architecture (the list of accession numbers can be found in Supplementary Table 1).

Reverse transcription was performed using a 17-mer oligo(dT) and the Superscript III RNase H Reverse Transcriptase Kit or Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen), according to the manufacturer's instructions. Negative control samples (no first-strand synthesis) were prepared by performing reverse transcription reactions in the absence of reverse transcriptase. Quantitative real-time PCR was carried out with 12.5 ng of total RNA per test well using the LightCycler DNA Master SYBR Green I kit (Roche) following the manufacturer's instructions. The PCRs were performed using an ABI Prism machine (Applied Biosystems) using the following cycling protocols: 1 min hot start at 94°C, followed by 45 cycles of 1 s at 94°C, 10 s at 60°C, and 15 s at 72°C. The threshold cycle (C_t) value was calculated from amplification plots, in which the fluorescence signal detected was plotted against the PCR cycle. The threshold was empirically set as 4% of the top 10 to 50% of fluorescence signals in the plateau phase.

The amplification efficiency was calculated from the slope of the standard curve by the following algorithm: $E = 10^{-1/\text{slope}}$. $E^{\text{Csf3r}} = 1.81$ and $E_{\text{Arg1}} = 1.84$. cDNA levels during the linear phase of amplification were normalized against Gapdh controls and the data of three independent reactions for each sample were averaged.

TFBS analysis

Our analysis covered the (−500, +200) regions relative to the transcription start site. Promoters that contained 5% or more ambiguous nucleotides were excluded from consideration, leaving 1784 promoters in our target set. We used all available matrix models of TFBSs contained in the Transfac Professional (version 7.4) database [48]. This was done using the command-line version of the MATCH program and we mapped these to the extracted promoter sequences. We used *minSUM* profiles [49] for thresholding of the matrix models: these profiles contain threshold values for the core and matrix scores, optimized to minimize the sum of false-positive and false-negative TFBS predictions. To determine TFBSs specific to our target set of macrophage promoters, we calculated an overrepresentation index (ORI) [50] using a background set of 40,101 randomly chosen unique mouse promoters; this calculation was done by in-house computer scripts. Readers interested in applying the same filtering process for annotation of promoters based on ORI should contact the authors.

An ORI value of 1 means no overrepresentation of the motif in the target promoter group compared to the background; a larger ORI value means greater overrepresentation. We filtered out all TFBSs with an ORI below 2.8. The remaining TFBSs were used to annotate the target promoters.

Expression clustering

Expression data were normalized so that the mean of each gene across time points was equal to 1. Hierarchical clustering was performed with Mathematica 5.1 (Wolfram Research, Inc.; www.wolfram.com). For robustness against outliers, we chose the L1 (Manhattan) distance metric and average linkage. Clusters were delimited manually from the resulting dendrogram.

TFBS dissimilarity metric and clustering

By viewing each promoter as a set of TFBSs, dissimilarity metric between any two promoters was defined as

$$d(T_1, T_2) = \frac{|T_1 \cup T_2|}{|T_1 \cap T_2|} - 1. \quad (1)$$

This gives $d = 0$ for between-transcript pairs with identical sets of TFBSs and $d > 0$ otherwise. If the two transcripts have no TFBSs in common, we define d = infinity. This metric was used to cluster promoters using hierarchical clustering by Mathematica 5.1 (Wolfram Research, Inc.). Cluster centroids were computed by treating TFBS binding data as a vector and computing all such vectors in a promoter cluster. This was done by in-house computer scripts.

Calculation of ORI

The ORI of any given promoter element in the target group relative to the background was calculated as

$$ORI(PEi) = \frac{DensityT(PEi)}{DensityB(PEi)} \times \frac{PROPORTION\_t}{PROPORTION\_b} = \frac{\frac{PEi\_tar}{TotalLengthT} \times \frac{Ntar}{N\_TAR}}{\frac{PEi\_back}{TotalLengthR} \times \frac{Nback}{N\_BACK}} \quad (2)$$

where DensityT/B is the density at which promoter element PEi is found in the target/background set, PROPORTION_t/b is the proportion of target/background sequences in which promoter element PEi is found, PEi_tar/back is the number of promoter elements PEi in the target/background sequence set, TotalLengthT/B is the total length of the target/background sequence set, Ntar/back is the number of target/background sequences in which PEi is found, and N_TAR/BACK is the number of target/background sequences.

Permutation tests

We used standard permutation tests to determine statistical significance for observed expression distances and cluster tree distances (defined as the number of splits between transcripts in the observed cluster tree). In each case, the observed value was compared to an empirical distribution obtained by computing that same statistics for data sampled at random from the full dataset, and the resulting percentile is reported as a p value. Throughout, 1000 permutations were used.

Acknowledgments

This research has been supported by a research grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science, and Technology of the Japanese Government to Y.H.; a research grant for the Advanced and Innovational Research Program in Life Science to Y.H.; a grant for the Genome Network Project from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to Y.H.; a grant for the Strategic Programs for R&D of RIKEN to Y.H.; and grants from the Australian National Health and Medical Research Council to D.A.H. and T.R. The authors also thank Dr. Hiroaki Kitano for valuable comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.03.022.

References

[1] C.A. Wells, T. Ravasi, D.A. Hume, Inflammation suppressor genes: please switch out all the lights, J. Leukocyte Biol. 78 (2005) 9–13.
[2] C. Nathan, Points of control in inflammation, Nature 420 (2002) 846–852.
[3] D.A. Hume, et al., The mononuclear phagocyte system revisited, J. Leukocyte Biol. 72 (2002) 621–627.
[4] T. Ravasi, et al., Generation of diversity in the innate immune system: macrophage heterogeneity arises from gene-autonomous transcriptional probability of individual inducible genes, J. Immunol. 168 (2002) 44–50.
[5] C.A. Wells, et al., Genetic control of the innate immune response, BMC Immunol. 4 (2003) 5 (Electronic publication).
[6] N.M. Luscombe, et al., Genomic analysis of regulatory network dynamics reveals large topological changes, Nature 431 (2004) 308–312.
[7] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev., Genet. 5 (2004) 101–113.
[8] M.J. Holland, Transcript abundance in yeast varies over six orders of magnitude, J. Biol. Chem. 277 (2002) 14363–14366.
[9] D.P. Sester, et al., Bacterial/CpG DNA down-modulates colony stimulating factor-1 receptor surface expression on murine bone marrow-derived macrophages with concomitant growth arrest and factor-independent survival, J. Immunol. 163 (1999) 6541–6550.
[10] J.L. Reed, B.O. Palsson, Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states, Genome Res. 14 (2004) 1797–1805.
[11] S. Ott, A. Hansen, S.Y. Kim, S. Miyano, Superiority of network motifs over optimal networks and an application to the revelation of gene network evolution, Bioinformatics 21 (2005) 227–238.

[12] T.S. Gardner, D. di Bernardo, D. Lorenz, J.J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling, Science 301 (2003) 102–105.
[13] D. di Bernardo, et al., Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks, Nat. Biotechnol. 23 (2005) 377–383.
[14] I. Lee, S.V. Date, A.T. Adai, E.M. Marcotte, A probabilistic functional network of yeast genes, Science 306 (2004) 1555–1558.
[15] S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, G.M. Church, Systematic determination of genetic network architecture, Nat. Genet. 22 (1999) 281–285.
[16] E. Segal, et al., Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data, Nat. Genet. 34 (2003) 166–176.
[17] P. Carninci, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.
[18] S.R. Himes, S. Cronau, C. Mulford, D.A. Hume, The Runx1 transcription factor controls CSF-1-dependent and -independent growth and survival of macrophages, Oncogene 24 (2005) 5278–5286.
[19] T.Y. Chin, Y.S. Lin, S.H. Chueh, Antiproliferative effect of nitric oxide on rat glomerular mesangial cells via inhibition of mitogen-activated protein kinase, Eur. J. Biochem. 268 (2001) 6358–6368.
[20] C.D. Dumitru, et al., TNF-alpha induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-dependent pathway, Cell 103 (2000) 1071–1083.
[21] J. Liu, X. Guan, X. Ma, Interferon regulatory factor 1 is an essential and direct transcriptional activator for interferon {gamma}-induced RANTES/CCl5 expression in macrophages, J. Biol. Chem. 280 (2005) 24347–24355.
[22] J.I. Kollet, T.M. Petro, IRF-1 and NF-kappaB p50/cRel bind to distinct regions of the proximal murine IL-12 p35 promoter during costimulation with IFN-gamma and LPS, Mol. Immunol. 43 (6) (2005) 623–633.
[23] C. Pasare, R. Medzhitov, Toll-like receptors: linking innate and adaptive immunity, Adv. Exp. Med. Biol. 560 (2005) 11–18.
[24] M. Muzio, N. Polentarutti, D. Bosisio, P.P. Manoj Kumar, A. Mantovani, Toll-like receptor family and signalling pathway, Biochem. Soc. Trans. 28 (2000) 563–566.
[25] M. Santic, M. Molmeret, Y. Abu Kwaik, Modulation of biogenesis of the Francisella tularensis subsp. novicida-containing phagosome in quiescent human macrophages and its maturation into a phagolysosome upon activation by IFN-gamma, Cell. Microbiol. 7 (2005) 957–967.
[26] A. Toyohara, K. Inaba, Transport of phagosomes in mouse peritoneal macrophages, J. Cell Sci. 94 (Pt 1) (1989) 143–153.
[27] R.Z. Murray, F.G. Wylie, T. Khromykh, D.A. Hume, J.L. Stow, Syntaxin 6 and Vti1b form a novel SNARE complex, which is up-regulated in activated macrophages to facilitate exocytosis of tumor necrosis factor-alpha, J. Biol. Chem. 280 (2005) 10478–10483.
[28] M. Nakamura, P. Carninci, [Cap analysis gene expression: CAGE], Tanpakushitsu Kakusan Koso 49 (2004) 2688–2693.
[29] R. Kodzius, et al., Absolute expression values for mouse transcripts: re-annotation of the READ expression database by the use of CAGE and EST sequence tags, FEBS Lett. 559 (2004) 22–26.
[30] R. Hines, B.R. Sorensen, M.A. Shea, W. Maury, PU.1 binding to ets motifs within the equine infectious anemia virus long terminal repeat (LTR) enhancer: regulation of LTR activity and virus replication in macrophages, J. Virol. 78 (2004) 3407–3418.
[31] H.G. Kim, et al., The ETS family transcription factor PU.1 is necessary for the maintenance of fetal liver hematopoietic stem cells, Blood 104 (2004) 3894–3900.
[32] J. Proffitt, et al., An ATF/CREB-binding site is essential for cell-specific and inducible transcription of the murine MIP-1 beta cytokine gene, Gene 152 (1995) 173–179.
[33] S.A. Rushworth, X.L. Chen, N. Mackman, R.M. Ogborne, M.A. O'Connell, Lipopolysaccharide-induced heme oxygenase-1 expression in human monocytic cells is mediated via Nrf2 and protein kinase C, J. Immunol. 175 (2005) 4408–4415.
[34] K. Srisook, Y.N. Cha, Super-induction of HO-1 in macrophages stimulated with lipopolysaccharide by prior depletion of glutathione decreases iNOS expression and NO production, Nitric Oxide 12 (2005) 70–79.
[35] K. Itoh, et al., Transcription factor Nrf2 regulates inflammation by mediating the effect of 15-deoxy-Delta(12,14)-prostaglandin j(2), Mol. Cell. Biol. 24 (2004) 36–45.
[36] T. Iizuka, et al., Nrf2-deficient mice are highly susceptible to cigarette smoke-induced emphysema, Genes Cells 10 (2005) 1113–1125.
[37] T.I. Lee, et al., Transcriptional regulatory networks in Saccharomyces cerevisiae, Science 298 (2002) 799–804.
[38] S. Boehlk, et al., ATF and Jun transcription factors, acting through an Ets/CRE promoter module, mediate lipopolysaccharide inducibility of the chemokine RANTES in monocytic Mono Mac 6 cells, Eur. J. Immunol. 30 (2000) 1102–1112.
[39] T. Ishii, et al., Role of Nrf2 in the regulation of CD36 and stress protein expression in murine macrophages: activation by oxidatively modified LDL and 4-hydroxynonenal, Circ. Res. 94 (2004) 609–616.
[40] H. Motohashi, M. Yamamoto, Nrf2–Keap1 defines a physiologically important stress response mechanism, Trends Mol. Med. 10 (2004) 549–557.
[41] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[42] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[43] R.T. Sasmono, et al., A macrophage colony-stimulating factor receptor–green fluorescent protein transgene is expressed throughout the mononuclear phagocyte system of the mouse, Blood 101 (2003) 1155–1163.
[44] M.A. Beer, S. Tavazoie, Predicting gene expression from sequence, Cell 117 (2004) 185–198.
[45] M.J. Sweet, D.A. Hume, CSF-1 as a regulator of macrophage activation and immune responses, Arch. Immunol. Ther. Exp. (Warsz) 51 (2003) 169–177.
[46] R. Miki, et al., Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays, Proc. Natl. Acad. Sci. USA 98 (2001) 2199–2204.
[47] Y. Xiao, R. Frisina, A. Gordon, L. Klebanov, A. Yakovlev, Multivariate search for differentially expressed gene combinations, BMC Bioinform. 5 (2004) 164 (Electronic publication).
[48] V. Matys, et al., TRANSFAC: transcriptional regulation, from patterns to profiles, Nucleic Acids Res. 31 (2003) 374–378.
[49] A.E. Kel, et al., MATCH: a tool for searching transcription factor binding sites in DNA sequences, Nucleic Acids Res. 31 (2003) 3576–3579.
[50] V.B. Bajic, V. Choudhary, C.K. Hock, Content analysis of the core promoter region of human genes, In Silico Biol. 4 (2004) 109–125.