

## Systems biology of innate immunity

Jesper Tegnér<sup>a,b,\*</sup>, Roland Nilsson<sup>b</sup>, Vladimir B. Bajic<sup>c</sup>,  
Johan Björkegren<sup>a</sup>, Timothy Ravasi<sup>d,e,f,\*</sup>

<sup>a</sup> Unit of Computational Medicine, King Gustaf V Research Institute, Department of Medicine, Karolinska Institute, SE-171 76 Stockholm, Sweden

<sup>b</sup> Computational Biology, Department of Physics, Linköping University, SE-581 53 Linköping, Sweden

<sup>c</sup> University of Western Cape, SANBI, Bellville 7535, South Africa

<sup>d</sup> Genome Exploration Research Group (Genome Network Project Core Group), RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

<sup>e</sup> Scripps NeuroAIDS Preclinical Studies Centre, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

<sup>f</sup> Department of Bioengineering, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Received 11 January 2007; accepted 31 January 2007  
Available online 11 April 2007

### Abstract

Systems Biology has emerged as an exciting research approach in molecular biology and functional genomics that involves a systematic use of genomic, proteomic, and metabolomic technologies for the construction of network-based models of biological processes. These endeavors, collectively referred to as systems biology establish a paradigm by which to systematically interrogate, model, and iteratively refine our knowledge of the regulatory events within a cell. Here, we present a new systems approach, integrating DNA and transcript expression information, specifically designed to identify transcriptional networks governing the macrophage immune response to lipopolysaccharide (LPS). Using this approach, we are not only able to infer a global macrophage transcriptional network, but also time-specific sub-networks that are dynamically active across the LPS response. We believe that our system biological approach could be useful for identifying other complex networks mediating immunological responses.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Systems biology; Networks; Transcriptional regulation; Regulatory circuits; Dynamics; Protein–protein interactions; Genome; Macrophages; Innate immunity

### 1. Introduction

Macrophages are without any doubts one of the most important cell type of the innate immune system, they play many roles spanning from development, wound healing and homeostasis; yet their destructive potential and secretory products are central to the pathology of acute and chronic inflammatory disease [1,2]. This functional diver-

sity reflects the complex interplay of regulatory networks regulating macrophages functions [3–5]. As part of the innate immune response, macrophages recognize and are activated by conserved components of microorganisms (pathogen-associated molecular patterns or PAMPs), and respond with a massive alteration in transcriptional output [5–8]. Expression profiling has revealed thousands of genes that are induced or repressed in macrophages in response to the classical activating agent lipopolysaccharide (LPS) and other PAMPs, reflecting the presence of complex regulatory mechanisms underlying this response.

Although, expression profiles alone are very informative they are limited to describing the complexity of the macrophage activation [4] as the underlying regulatory networks cannot be resolved. To advance the understanding beyond

\* Corresponding authors. Address: Department of Bioengineering, Jacobs School of Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA (T. Ravasi). Fax: +46 13 142355 (J. Tegnér), +1 858 822 4246 (T. Ravasi).

E-mail addresses: [jespert@ifm.liu.se](mailto:jespert@ifm.liu.se) (J. Tegnér), [travasi@bioeng.ucsd.edu](mailto:travasi@bioeng.ucsd.edu) (T. Ravasi).

associations between groups of genes either a systematic perturbation analysis has to be employed [9–12] or different types of data has to be integrated with the expression data [3,13]. Yet, up to date there are only few systems biology approaches were, expression measurements are integrated with other genome-wide datasets with the aim to gain deeper insights of regulatory mechanisms controlling the macrophage activation. Hence, in order to infer transcriptional networks regulating macrophage activation by LPS, we and others have collected and integrated in a systems approach several types of genome-wide measurements including, transcriptome changes, quantitative expression of transcriptional regulators and the promoter structure [3,13]. This approach required us to develop a new bioinformatics procedure on how to integrate these data into a network model representing the entire macrophage response to LPS. This enabled us to interrogate these time-dependent dynamic networks with specific biological questions.

## 2. Material and methods

### 2.1. Bone marrow-derived macrophages and LPS treatment

Generation of bone marrow-derived macrophages (BMM) was performed as described previously [14]. Macrophages were seeded at  $1 \times 10^7$  cells/ml and incubated with 10 ng/ml LPS.  $3 \times 10$  cm dishes were harvested for each time point—unstimulated (time 0, 2, 7, and 24 h). Total RNA from BMM were extracted by using Qiagen RNeasy Mini kits according to the manufacturer's protocols.

### 2.2. Microarray data analysis and identification of clones differentially expressed over time

In order to identify statistically significant differentially expressed clones over time, following a pretreatment of the data [13], we performed paired *t*-tests by comparing expression data sampled from the following time points: 0–2, 0–7, 0–24, 2–7, 2–24, and 7–24 h. A single list representing all the transcripts differentially expressed at least in one of these comparisons was obtained by taking the union of the lists.

A stringent significance threshold for differential expression ( $p < 0.001$ ) was used to limit the number of false positive findings together with a multivariate permutation test [15] which provided a 75% confidence estimation that the number of false-positive genes did not exceed 10. Expression data was normalized so that the average of each gene across time points equals 1. Hierarchical clustering was performed with Mathematica 5.1 (Wolfram Research, Inc., [www.wolfram.com](http://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.

### 2.3. 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](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. A detail protocol for the qPCR and the list of the primers and accession numbers can be found elsewhere [13].

### 2.4. TFBS analysis

Our analysis covered the [–500, +200] regions relative to the transcription start site (TSS). 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 TRANSFAC Professional (Version. 7.4) database [16] and the analysis was performed by using the command-line version of the MATCH program and the matrix models were mapped to the extracted promoter sequences. We used minSUM profiles [17] 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 specificity of our target set of macrophage promoters, we calculated an over-representation index (ORI) [18] 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 over-representation of the motif in the target promoter group as compared to the background; a larger ORI value indicates larger over-representation. We filtered out all TFBSs with ORI below 2.8. The remaining TFBSs were used to annotate the target promoters, a detail protocol including equations used in this analysis are available at [13].

### 2.5. Networks visualization and analysis

Networks presented in Fig. 1 have been visualized using Cytoscape version 2.2, a freely available network visualization and analysis software (<http://www.cytoscape.org>) [19]. Network statistics was calculated by using in-house computer scripts running on the Mathematica 5.1 platform (Wolfram Research, Inc., [www.wolfram.com](http://www.wolfram.com)).

## 3. Results

### 3.1. Definition of the node and edge libraries for the construction of the macrophage network

Our analysis follows a three step strategy. First, the nodes in the network are defined by the genes which are

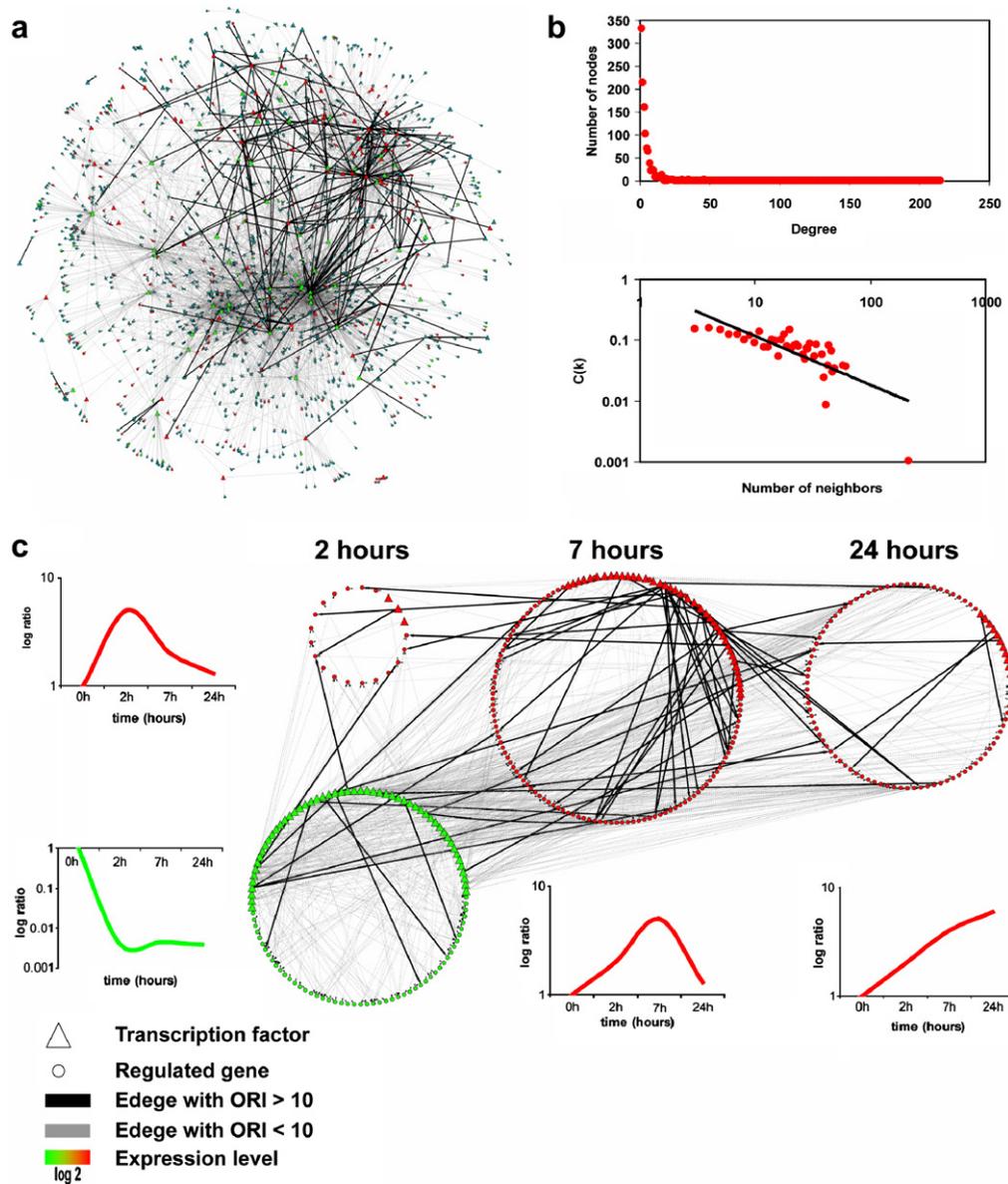


Fig. 1. (a) Global view of the macrophage LPS transcriptional network. Nodes represent transcription factors (triangle) or regulated genes (circle); inferred transcriptional control events are represented by an edge between a transcription factor and the regulated gene (solid line). This figure clearly illustrates the presence of highly connected transcription factors (hubs of the network). (b) Topological analysis of the global network. The macrophage transcriptional network is scale free, the degree distribution of the interactions follows a power-law relationship (top plot). In other words there are only few nodes (hubs) with high probability of interaction  $P(k)$ . Also the distribution of the clustering coefficients  $C(k)$ , follows a power-law relationship (bottom plot). These are global characteristic common to many “real” networks [25]. (c) Networks dynamics during the macrophage LPS response. Our approach enables us to define time-specific sub-networks, based on the dynamic profiles of TFs and their regulated genes; we have identified sub-networks that are particularly active at specific time-points during the macrophage response to LPS. Interestingly there are evidences of high connectivity between the sub-networks (edges across the four sub-network). The plots show the representative expression profile of the majority of the TFs and genes for each sub-network. The thickness of the edge indicates the score calculated from the TFBSs over-representation index (ORI) as described in Section 2.

“active” in the gene expression data. Next, we perform a promoter analysis to collect an edge library for the active genes. Finally, we define time-dependent sub-networks by combining the edge library with the genes that are active in each time-point of the LPS response.

**Nodes.** In order to define the nodes in the macrophage transcriptional network we integrated two types of genome-wide measurements from murine bone marrow-derived macrophage (BMM) at four time-points after

LPS stimulation (unstimulated; 2, 7 and 24 h post-stimulation).

First, we used RIKEN 60,000 cDNA array as a platform to monitor the genome-wide changes in macrophage genes expression over the time-course. 2219 (4%) non-redundant array spots out of the 60,000 were above the detection limit and significantly changed over the time-course after LPS stimulation. Our results were in agreement with previously published analysis [3–5] showing

groups of genes with precise time-specific dynamic expression where the main classification categories were: “early” up-regulated (0–2 h), “middle” up-regulated (2–7 h), “late” up-regulated (7–24 h), “early” down-regulated (0–2 h) and transcripts expressed continuously over the time-course.

As a second data-source for defining the nodes we performed quantitative real-time PCR (qPCR) of 1559 known and predicted transcripts coding for transcription factors and other putative nuclear proteins (TFs). The rationale is that we are focusing on the transcriptional networks and because transcriptional regulators are often expressed below the detection limit of microarrays [20]. From the qPCR data, we identified 43% (673/1559) of the TFs monitored to be significantly expressed or differentially regulated over the time-course with a dynamical response resembling the transcriptome changes detected by RIKEN cDNA microarray described previously [5,13].

*Edges.* an essential step toward the construction of networks is the definition of edge libraries, namely interactions among the networks nodes (i.e. protein; genes; promoters etc.). In order to computationally define edges between the 2892 (2219 from the microarray + 673 from the qPCR) nodes expressed/regulated in our system we used the promoter sequences as defined by the Genome Network Project (GNP) by using the Cap Analysis of Gene Expression (CAGE) [6,21]. This new technique enables a systematic mapping of transcription starting sites (TSS) in the genome and it is therefore a more accurate definition of proximal promoters of virtually any gene expressed in the system under investigation [22]. The GNP generated CAGE libraries for each of the four time-points of the LPS time-course and the corresponding results of the entire CAGE dataset analysis for the macrophage system has been presented elsewhere [6,21]. For the current analysis we have defined the active 800 kilobases promoter sequences of the 2892 nodes in CAGE libraries corresponding to the four time-points during the LPS time-course and performed predictions of transcription factor binding sites (TFBSs) based on the Transfac matrix models. With this method, we identified 298 distinct transcription factors (TFs) from our node library with at least one associated TFBS in at least one of the 2892 promoters, filtered by removing the predicted binding sites that were frequently found also in a background set. In summary, we obtained 3824 TF-promoter interactions.

### 3.2. The global macrophage network

Integrating these datasets we have been able to infer the first macrophage transcriptional network, where the topological representation of the entire network is displayed in Fig. 1a. To further validate our approach we characterized the topological properties of the macrophage network. Previous studies, mainly by Barabasi and colleagues [23–25], revealed that complex systems including biological systems exhibit scale-free properties, where some nodes act as “highly connected hubs” (having many more connections

than the average), although most nodes have few connections (Fig. 1b) [23,24,26]. A similar analysis demonstrated that some of the most connected TFs in the macrophages network are: Nrf-2 (out-degree = 242); Atf-3 (out-degree = 116), Ets-1 (out-degree = 100) and Irf-1 (out-degree = 76); TFs with central roles in controlling the LPS response in macrophage have been extensively verified (Ets-1 and Irf-1) [27,28] or recently linked to this biology as for the case of Nrf-2 and Atf-3 [3,13,29].

### 3.3. Network dynamics

Because of the dynamical nature of the measurements we used for the network construction we were able to analyze the entire network in order to identify sub-networks which were particularly active (up-regulated or down regulated) during specific time-points of the LPS time-course. Here we defined a 20 sub-networks active at discrete time-points defining dynamic regulatory circuits likely to be necessary for the innate immunity program in macrophages (Fig. 1c).

## 4. Discussion

Here, we have presented a scalable systems biology approach to construct a regulatory network in LPS activated macrophages by integrating expression measurement and promoter structure. Our approach turned out to be very powerful and unbiased in defining not only time-specific sub-networks but also high connected TFs (hubs of the network) that are likely to be play fundamental roles during the innate immune program.

There are several noise sources in our analysis which could potentially contaminate the resulting network. For example, the data quality of the high-throughput datasets which define the nodes in the network, the edges of the network are not defined from physical interactions (i.e. protein-protein interaction and TF location analysis) but on the promoter occurrence inferred from computationally predicted TFBS. Despite these limitations we believe that our system approach is applicable in understanding virtually any mammalian cell type. The definition of nodes will be improved with increasing quality and amount of expression data generated from array technologies. The edge library can also be improved by adding other data-sources such as physical binding information [30] and algorithms for identifying gene-to-gene interactions [12]. Therefore, we will be able to further refine our predictions and infer a more robust macrophage network that can be interrogated by other researchers with specific biological questions related to immunity.

### Acknowledgments

This research has been supported by Research Grant for the Scripps NeuroAIDS Preclinical Studies Center (SNAPS) from the National Institute of Mental Health

(NIMH), Grant No. 2P30MH062261-07 to T.R. and the Ph.D. programme in medical bioinformatics (R.N.), Swedish Research Council (J.B.), Clinical Gene Networks AB and Linköping University (J.T.).

## References

- [1] C. Nathan, Points of control in inflammation, *Nature* 420 (2002) 846–852.
- [2] C.A. Wells, T. Ravasi, D.A. Hume, Inflammation suppressor genes: please switch out all the lights, *J. Leukoc. Biol.* 78 (2005) 9–13.
- [3] M. Gilchrist, V. Thorsson, B. Li, A.G. Rust, M. Korb, K. Kennedy, T. Hai, H. Bolouri, A. Aderem, Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4, *Nature* 441 (2006) 173–178.
- [4] T. Ravasi, C. Wells, A. Forest, D.M. Underhill, B.J. Wainwright, A. Aderem, S. Grimmond, D.A. Hume, 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, T. Ravasi, G.J. Faulkner, P. Carninci, Y. Okazaki, Y. Hayashizaki, M. Sweet, B.J. Wainwright, D.A. Hume, Genetic control of the innate immune response, *BMC Immunol.* 4 (2003) 5.
- [6] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, et al., The transcriptional landscape of the mammalian genome, *Science* 309 (2005) 1559–1563.
- [7] S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, et al., Antisense transcription in the mammalian transcriptome, *Science* 309 (2005) 1564–1566.
- [8] C.A. Wells, T. Ravasi, R. Sultana, K. Yagi, P. Carninci, et al., Continued discovery of transcriptional units expressed in cells of the mouse mononuclear phagocyte lineage, *Genome Res.* 13 (2003) 1360–1365.
- [9] M. Bansal, G.D. Gatta, D. di Bernardo, Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, *Bioinformatics* 22 (2006) 815–822.
- [10] D. di Bernardo, M.J. Thompson, T.S. Gardner, S.E. Chobot, E.L. Eastwood, A.P. Wojtovich, S.J. Elliott, S.E. Schaus, J.J. Collins, Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks, *Nat. Biotechnol.* 23 (2005) 377–383.
- [11] 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.
- [12] J. Tegnér, J. Björkegren, Perturbations to uncover gene networks, *Trends Genet.* 23 (2007) 34–41.
- [13] R. Nilsson, V.B. Bajic, H. Suzuki, D. di Bernardo, J. Björkegren, et al., Transcriptional network dynamics in macrophage activation, *Genomics* 88 (2006) 133–142.
- [14] 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.
- [15] Y. Xiao, R. Frisina, A. Gordon, L. Klebanov, A. Yakovlev, Multivariate search for differentially expressed gene combinations, *BMC Bioinformatics* 5 (2004) 164.
- [16] V. Matys, E. Fricke, R. Geffers, E. Gossling, M. Haubrock, et al., TRANSFAC: transcriptional regulation, from patterns to profiles, *Nucleic Acids Res.* 31 (2003) 374–378.
- [17] A.E. Kel, E. Gossling, I. Reuter, E. Cheremushkin, O.V. Kel-Margoulis, E. Wingender, MATCH: a tool for searching transcription factor binding sites in DNA sequences, *Nucleic Acids Res.* 31 (2003) 3576–3579.
- [18] 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.
- [19] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504.
- [20] M.J. Holland, Transcript abundance in yeast varies over six orders of magnitude, *J. Biol. Chem.* 277 (2002) 14363–14366.
- [21] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, et al., Genome-wide analysis of mammalian promoter architecture and evolution, *Nat. Genet.* 38 (2006) 626–635.
- [22] V.B. Bajic, S.L. Tan, A. Christoffels, C. Schonbach, L. Lipovich, et al., Mice and men: their promoter properties, *PLoS Genet.* 2 (2006) e54.
- [23] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, *Science* 286 (1999) 509–512.
- [24] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* 5 (2004) 101–113.
- [25] E. van Nimwegen, Scaling laws in the functional content of genomes, *Trends Genet.* 19 (2003) 479–484.
- [26] N.M. Luscombe, M.M. Babu, H. Yu, M. Snyder, S.A. Teichmann, M. Gerstein, Genomic analysis of regulatory network dynamics reveals large topological changes, *Nature* 431 (2004) 308–312.
- [27] 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.* (2005).
- [28] T. Roger, I. Miconnet, A.L. Schiesser, H. Kai, K. Miyake, T. Calandra, Critical role for Ets, AP-1 and GATA-like transcription factors in regulating mouse Toll-like receptor 4 (Tlr4) gene expression, *Biochem. J.* 387 (2005) 355–365.
- [29] S.A. Rushworth, X.L. Chen, N. Mackman, R.M. Ogborne, 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.
- [30] C.T. Workman, H.C. Mak, S. McCuine, J.B. Tagne, M. Agarwal, O. Ozier, T.J. Begley, L.D. Samson, T. Ideker, A systems approach to mapping DNA damage response pathways, *Science* 312 (2006) 1054–1059.