An integrative computational approach for a prioritization of key transcription regulators associated with nanomaterial-induced toxicity

Vadim Zhernovkov,[*] Tapesh Santra,[*] Hilary Cassidy,[*] Oleksii Rukhlenko,[*] David Matallanas,[*,†] Aleksandar Krstic,[*] Walter Kolch,[*,†,‡] Vladimir Lobaskin,[§] and Boris N. Kholodenko[*,†,‡¶,1]

[*]Systems Biology Ireland, University College Dublin, Dublin 4, Ireland

[†]School of Medicine and Medical Science, University College Dublin, Belfield, Dublin 4, Ireland

[‡]Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Ireland

[§]School of Physics, University College Dublin, Dublin 4, Ireland

[¶]Department of Pharmacology, Yale University School of Medicine, New Haven CT 06520, USA

[1]Corresponding author

E-mail addresses:

VZ: vadim.zhernovkov@ucd.ie

TS: tapesh.santra@gmail.com

HC: hilary.cassidy@ucd.ie

OR: oleksii.rukhlenko@ucd.ie

DM: david.gomez@ucd.ie

AK: aleksandar.krstic@ucd.ie

WK: walter.kolch@ucd.ie

VL: vladimir.lobaskin@ucd.ie

BNK: boris.kholodenko@ucd.ie

Abstract

A rapid increase of new nanomaterial products poses new challenges for their risk assessment. Current traditional methods for estimating potential adverse health effect of nanomaterials (NMs) are complex, time consuming and expensive. In order to develop new prediction tests for nanotoxicity evaluation, a systems biology approach and data from high-throughput omics experiments can be used. We present a computational approach that combines reverse engineering techniques, network analysis and pathway enrichment analysis for inferring the transcriptional regulation landscape and its functional interpretation. To illustrate this approach, we used published transcriptomic data derived from mice lung tissue exposed to carbon nanotubes (NM-401 and NRCWE-26). Because fibrosis is the most common adverse effect of these NMs, we included in our analysis the data for bleomycin (BLM) treatment, which is a well-known fibrosis inducer. We inferred gene regulatory networks for each NM and BLM to capture functional hierarchical regulatory structures between genes and their regulators. Despite the different nature of the lung injury caused by nanoparticles and BLM, we identified several conserved core regulators for all agents. We reason that these regulators can be considered as early predictors of toxic responses after NMs exposure. This integrative approach, which refines traditional methods of transcriptomic analysis, can be useful for prioritization of potential core regulators and generation of new hypothesis about mechanisms of nanoparticles toxicity.

Keywords: Nanoparticles, Carbon nanotubes, Bleomycin, Gene regulatory network, Gene regulation, Fibrosis

Introduction

Nanomaterials (NMs) have gained increased attention in the last 30 years due to their unique properties, exploited in various industrial and commercial products. NMs are widely used in electronics, cosmetics and biomedical applications, such as drug delivery (De Volder *et al.*, 2013). However, inhalation of certain NMs causes adverse health effects, such as chronic inflammation, pulmonary fibrosis, carcinogenesis and other undesirable effects (Dong and Ma, 2015). Among these NMs, carbon nanotubes (CNT) were extensively investigated in the health-related context in recent years. Toxicological effects linked to CNT exposure include inflammation, fibrosis, DNA genotoxicity and tumorigenesis (Dong and Ma, 2015). The biological response to CNT exposure has many similarities with the responses observed for bleomycin (BLM) treatment, which is widely used as a classic model of inducing lung fibrosis (Peng *et al.*, 2013).

Although many studies have investigated mechanisms and signaling pathways of CNTs and BLM toxicity (Nikota *et al.*, 2016; Peng *et al.*, 2013; Poulsen *et al.*, 2015), the global transcription regulation program altered in response to these agents is mainly unknown. This lack of knowledge hinders the development of accurate prediction tests for nanotoxicity evaluation. Currently, about a hundred of transcriptomics data sets for NMs have been included in GEO database. These *omics* data sets present expression values for thousands of transcripts altered in response to NM treatment. Current approaches for the analysis of gene expression responses to NMs include identification of differentially expressed genes (DEGs), subsequent enrichment analysis of signaling pathways and the use of GO terms for functional interpretation of these findings. This type of analysis can provide information about distinct genes and enriched functional groups, but a collective, systems-level response to NMs remains hidden. Novel and complementary approaches can move forward the nanotoxicology field by capturing mutual variation in gene expression at the systems

level. Previous studies have demonstrated the power of reverse engineering techniques to achieve this aim. Such techniques were used to identify molecular biomarkers and drug targets in cancer and other diseases (Chen *et al.*, 2014), the deduction of adverse outcome pathways (AOPs) for chemicals from high-throughput transcriptomic data sets (Perkins *et al.*, 2011; Villeneuve *et al.*, 2014), the functional interpretation of responsive modules from gene expression data sets (Marwah *et al.*, 2018), and the characterization of macrophage responses to pathogen stimuli (McDermott *et al.*, 2011).

Another knowledge gap in the nanotoxicology field is related to functional roles of transcription factors (TFs), which shape the global transcription response. Far too little attention in *omics* analysis has been paid to this type of genes, although TFs determine the landscape of toxicological response (Andersen *et al.*, 2013; Chepelev *et al.*, 2015; Souza *et al.*, 2017). Furthermore, some TFs are considered as early indicators of stress-induced changes (Jennings *et al.*, 2013).

The aim of this paper is to provide a computational approach for inferring the transcriptional regulation landscape and its functional interpretation, using network maps of gene regulation by TFs known as gene regulatory networks (GRNs). GRNs provide important information about critical TFs that guide gene expression and their potential targets. In this study we inferred GRNs from transcriptomic data derived from mice lung tissue exposed to long and short carbon nanotubes (NM-401and NRCWE-26, respectively) and BLM. Based on these GRNs, we identified common potential core regulators of fibrosis-associated processes for all agents. We present a computational approach for the functional characterization of stress-response modules in these networks and prioritization of key mediators associated with specific biological processes or disease progression.

Materials and Methods

Datasets

Two types of multi-walled CNT were included in the analysis, NM-401 (4048±366 nm in length) and NRCWE-26 (847±102 nm in length). Both CNTs induced transcriptional and histological pulmonary fibrotic changes, but NM-401 triggered it with more pronounced effects compared to NRCWE-26 (Poulsen *et al.*, 2015). For comparison, we included data for bleomycin (BLM) treatment, that is widely used as a classic model of exogenously induced lung fibrosis (Peng *et al.*, 2013).

NM-401 is a long and rigid multi-walled CNT, NRCWE-26 is a short and entangled multi-walled CNT. Both CNTs induce strong pulmonary acute phase and inflammatory response that reaches the highest point at day 3 and is maintained at day 28. Both CNTs enhance immune cell infiltration in bronchoalveolar lavage fluid (Poulsen *et al.*, 2015). Histological analysis shows an interstitial pneumonia pattern with granulomas in lung tissue from mice exposed to high doses of NM-401 and NRCWE-26 CNTs at day 28, but more severe outcomes are found following NM-401 exposure, which includes granulomas, alveolar septal lymphoid infiltration and fibrosis (Poulsen *et al.*, 2015). Transmission Electron Microscopy (TEM) showed that curled and agglomerated NRCWE-26 nanoparticles were identified in the cytoplasm, inside vesicles on day 3. For NM-401 CNTs, TEM imaging also showed their localization in cytoplasm vesicles on day 3. But NM-401 appeared to induce some damage effects, including piercing of these vesicles and visible damage of cells, which was not observed with NRCWE-26 (Poulsen *et al.*, 2015). The toxic effects of these CNTs are related to reactive oxygen species (ROS) production and oxidative stress, but the mechanisms are unclear.

The biological response to CNT exposure has many similarities with the responses

observed for bleomycin (BLM) treatment (Peng *et al.*, 2013). BLM is a chemotherapeutic drug that causes DNA strand breaks *via* oxidative mechanisms and can induce lung fibrosis in animals and in human patients as a severe side effect. A single BLM dose initiates an acute inflammatory phase in lung tissue characterized by infiltration of immune cells and release of pro-inflammatory cytokines. The fibrotic phase is initiated seven days after BLM instillation with increased expression of pro-fibrotic cytokines and increased fibroblast proliferation and collagen accumulation (Williamson *et al.*, 2015). However, there are differences in endpoint effects of BLM and CNTs. For CNT exposure, genotoxic and pro-fibrotic responses together with immunomodulation components prevail, leading to chronic inflammation, fibrosis and possibly cancer in a long-term period (Rahman *et al.*, 2017). In contrast, single dose BLM models generally show attenuation of fibrotic features after 28 days post-exposure (Dong and Ma, 2016).

The gene expression profiles were obtained from Gene Expression Omnibus Database (https://www.ncbi.nlm.nih.gov/geo/) using the R/Bioconductor package GEOquery (Davis and Meltzer, 2007). Table 1 shows the summary characteristics of these data sets. Data for all CNTs and BLM were generated using Agilent microarrays and Affymetrix microarrays, respectively, in *in vivo* mice experiments. Three different doses of CNTs (18, 54, 162 μg) and three timepoints (1, 3, 28 days) following a single intratracheal instillation were used. BLM was administered in one dose, 2U/kg body weight, using a single intratracheal instillation method, and the lung tissue was harvested at 7 post-instillation timepoints (1, 2, 7, 14, 21, 28, 35 days). All experiments were conducted with vehicle controls for each timepoint. The total number of analyzed samples in each data sets ranged from 67 to 111, which included controls, different doses, times, and replicas.

Gene regulatory network (GRN) inference and analysis

GRNs were inferred individually for each agent. In order to generate lists of predicted interaction scores for TF-gene pairs, we applied an ensemble approach that combined results from three different algorithms: (*i*) the Bayesian variable selection algorithm (BVS) that is based on a linear regression model of gene regulation (Santra, 2014), (*ii*) the mutual information algorithm ARACNe-AP (Lachmann *et al.*, 2016), and (*iii*) the tree-based regression algorithm GENIE3 (Huynh-Thu *et al.*, 2010). Each method has benefits and disadvantages in their ability to capture mutual variation in gene expression. Moreover, there is currently no gold standard for such methods applied to mammals. Importantly, the prediction accuracy can be significantly improved by a combination of results from multiple statistical methods, especially using diverse type of methodologies (Marbach *et al.*, 2012). Therefore, in our analysis, we have chosen algorithms from different categories, which showed better prediction performance in Marbach's study, namely regression, mutual information and tree-based approaches. The steps of our pipeline were applied separately to each data set, as described below.

1. First, the expression values for multiple probes representing the same gene were averaged. The final data set was stored as a gene expression matrix where rows represented genes and columns represented samples (different conditions).
2. Next, DEGs were identified. The analysis was performed using the limma package in R/Bioconductor (Ritchie *et al.*, 2015). The list of genes was considered as significantly differentially expressed if the absolute values of the expression changes were equal to or greater than certain cut-off, namely 1.5-fold for CNT or BLM and 1.3-fold for CB treatments compared to non-treated controls for each experimental condition. To decrease the risk of a Type I error (false positive results) during multiple comparison tests, the Benjamini Hochberg (BH) correction procedure was

applied. Genes were considered as differentially expressed if BH-adjusted p-values were less than or equal to 0.05.

3. Non-differentially expressed genes were excluded from the gene expression matrix. The list of gene regulators (TFs) was obtained from the AnimalTFDB 3.0 database (Hu *et al.*, 2019) and mapped to the list of genes in the expression matrix.
4. The above set is then used to infer a list of predicted interaction score for TF - gene pairs. Bayesian variable selection and GENIE3 algorithms were run with recommended default parameters (Huynh-Thu *et al.*, 2010; Santra, 2014). The ARACNe-AP algorithm was run with three key steps: MI threshold estimation, bootstrapping/MI network reconstruction, building consensus network (only significant interactions are filtered, $p \leq 0.05$, BH corrected).
5. Next, for improving prediction accuracy, we integrated the results of all three algorithms by unweighted Borda count ranking, as described by Marbach et al (Marbach *et al.*, 2012) in the supplementary materials to their paper. Briefly, interaction scores predicted by each algorithm were ranked for all gene pairs. Subsequently, for each gene pair the integrated rank was calculated as the average of the individual interaction scores over three algorithms.
6. To identify an optimal threshold cutoff for selecting ranked gene pairs with high mutual dependencies and discard unlikely interactions, we used a gene ontology (GO) semantic similarity approach that estimates the function similarity for gene pairs based on GO graph structure (Yu *et al.*, 2010). The list of TFs and their potential targets were sorted by a descending order of interaction scores, and a GO similarity score was calculated for each gene pair using the GOSemSim R/Bioconductor package (Yu *et al.*, 2010). The semantic similarity of two GO terms was measured using the Wang method (Yu *et al.*, 2010), using biological process

ontologies for this analysis. Then, the list containing the calculated GO score for each gene pair was subdivided into bins with the bin width equal to 100 rows, and the median score values were calculated. Next, these values were fitted by a 4^{th}-order polynomial dependence on this median score. Then, the median of all median score values and their standard deviation were calculated. Taking this median score value plus the standard deviation divided by 2, the intersection of the fitted 4^{th}-order polynomial curve was found that defined the optimal threshold cutoff (Borate *et al.*, 2009). We calculated the threshold cutoff for each ranked list of TFs and their potential targets and chose the minimum value. Subsequently, based on the identified common minimum threshold, the top TF-gene pairs were selected in each list for subsequent analysis.

7. These gene pairs formed a network, where nodes represent genes and edges denote interactions between them. Network visualization and analysis were performed using the open source software platform Cytoscape version 3.4.0 (Shannon *et al.*, 2003). Topological parameters of the network were estimated using the NetworkAnalyzer plugin available within the Cytoscape software. To identify network modules, we applied the Girvan-Newman algorithm, named as Glay clustering method in the clusterMaker Cytoscape plugin (Girvan and Newman, 2002; Su *et al.*, 2010). The Polychrome R package (https://cran.r-project.org/web/packages/Polychrome/index.html) was used for generating a 36 colors palette, which was used for color visualization of different network modules.
8. The identified gene modules of inferred networks were then annotated based on the KEGG database (Kanehisa *et al.*, 2017), using the over-representation method and the gProfileR R package (Reimand *et al.*, 2016). The analysis was performed for each condition (time points; doses) separately. A fold change cut-off of 1.5 for DEGs

was used in this analysis. A threshold for the minimum number of genes per module was 50. The heatmaps representing fold-changes of DEGs for the network modules were generated using the Pheatmap R package (https://cran.r-project.org/web/packages/pheatmap/index.html). The fold-change values of DEGs used in the heatmaps were transformed using the floor and ceiling transformation. Namely, the values of the matrix elements that were above “2” were replaced with “2”, all values in the matrix below “-2” were replaced with “- 2”.

9. Finally, in order to prioritize gene regulators which are associated with fibrosis, we used the inferred network and identified TFs with the largest numbers of connections with fibrosis markers. We considered as fibrosis markers the genes associated with pulmonary fibrosis in the Comparative Toxicogenomics Database (Davis *et al.*, 2017). Then, we found TFs that are directly connected with these genes. Each TF from this list was characterized using the number of connections with fibrosis markers for each condition (time points; doses), provided that these markers were differentially expressed.

Data processing and statistical analysis were performed with R version 3.5.2 (https://www.r-project.org/) and RStudio version 1.1.383 (https://www.rstudio.com). For inferring GRNs using the ARACNE-AP algorithm, we used the JAVA executable command-line tool (https://github.com/califano-lab/ARACNe-AP). The Bayesian variable selection algorithm was run in a Matlab environment Version 6 (Mathworks).

Results

Inferring gene regulatory networks

Gene regulatory networks (GRNs) provide useful insights into transcriptional regulatory

mechanisms. GRNs have hierarchical structures where a few highly interconnected genes, usually TFs, are the hubs that account for most interactions. GRNs inferred from the expression data can suggest which TFs are responsible for the changes in gene expression observed following the exposure to CNTs or other agents. To infer gene regulatory networks, we used gene expression matrices for different experimental conditions and reverse engineering algorithms. These algorithms included a linear regression model (BVS), the mutual information algorithm (ARACNe-AP), and the tree-based regression algorithm (GENIE3). As a result, we obtained lists containing predicted interactions score for TF - gene pairs. To improve prediction accuracy, we applied an ensemble approach based on the Borda count ranking (see Methods), that combines results from these three algorithms. A term 'wisdom of crowds' was coined for this approach (Marbach *et al.*, 2012). The GRNs were reconstructed for each agent separately. An overview schematic of our analysis pipeline is presented in Figure 1 and technical characteristics for each step are shown in Table 2.

The TF – gene interaction score results can help us establish the relationships in GRNs. However, to select the most critical, top relationships, we have to estimate a cut-off threshold. We used the approach based on gene ontology (GO) semantic similarity measurement (Pesquita, 2017), which calculates the closeness between two genes based on the graph structure of gene ontology database terms (see Methods section for more details). Using this approach, we calculated optimal thresholds for each list of TFs and their potential targets (Figure 2 and Table 2). The identified thresholds were 7700 TF – gene pairs for the NM-401 list, 5700 for the NRCWE-26 list and 6500 for the BLM list. The 5700 value was chosen as a common optimal threshold for all interaction lists. Using this value, top TF-gene pairs were selected for subsequent analysis. We then built a network for these gene pairs where nodes represented genes and edges denoted interactions between them.

The main characteristics of these networks are presented in Table 2. The network diameter (maximum distance between nodes) and the characteristic path length in the BLM network was greater than in the CNT networks. Nonetheless, the clustering coefficients and average number of neighbors were generally similar for CNTs and BLM, indicating that all GRNs are well-connected, small world networks (Table 2). The visualization of these networks is shown in Figure 3, and the Cytoscape session file can be found in the Supplementary file 1. Next, in order to prioritize gene regulators for each network, we ranked TFs based on their connection numbers (Table 3). TFs with the largest numbers of connections are referred to as TFs hubs. These topological features are widely used in the analysis of GRNs, and TF hubs are deemed important in the cellular regulatory programs (Basso *et al.*, 2005; Langfelder *et al.*, 2013). Table 3 shows the TFs with the highest connectivity in all networks, such as *E2f8, Litaf, Foxm1, Mxd3, Myc* and *Irf7.*

Analysis of biological functions of genes in GRN modules

A feature of GRNs is colocalization of the genes, which are involved in similar biological processes, in the same network modules (van Dam *et al.*, 2018). We used this feature for a subsequent analysis of biological processes controlled by TFs. To identify network modules, we used the Girvan-Newman algorithm (Girvan and Newman, 2002). Considering modules containing 50 or more genes, this method revealed 12, 13 and 7 modules in the NM-401, NRCWE-26 and BLM networks, respectively. Figure 3 shows these modular structures for the inferred networks. To capture the dynamic complexity of transcriptional responses, heatmaps for upregulated and downregulated DEGs were plotted for each network module (Figure 3).

To identify the signaling pathways and functional processes which were altered in each network module following instillation of CNTs and BLM, we applied functional annotation

analysis using the KEGG database (Supplementary file 2). This analysis was performed for each condition (the time point and dose) separately. To capture major altered signaling pathways, we selected enriched pathways using the BH corrected p-value (Figure 4). As these Tables demonstrate, instillation of CNTs and BLM affected various physiological and pathological processes, such as (*i*) the immunomodulatory response (innate and adaptive), in particular, pathways activated by cytosolic pattern recognition receptors (PRRs), (*ii*) response to DNA damage/integrity, and (*iii*) cell death and senescence pathways, suggesting the involvement of these processes in the adverse effects observed upon CNTs and BLM instillation.

The inflammatory immune response is strongly stimulated by CNTs and BLM instillation. Several network modules identified above include the pathways involved in this response. The GRNs corresponding to each agent show that DEGs participating in the inflammatory response belong to modules of two different types: (1) early immune response, and (2) activation of cytosolic pattern recognition receptors (PRRs). Genes that comprised the first type of module were altered on day 1 and maintained altered expression on day 3 in response to NM-401 (module 2 and 11) and NRCWE-26 (module 2), while in the BLM network, these genes were induced on day 1 and showed maximal expression on days 7-14 (module 2 in the BLM network). For all modules of the first type, the effects increased with the CNT doses. These modules contain genes from inflammatory-related pathways, including IL-17 signaling and cytokine-cytokine receptor interaction (in the NM-401 and BLM networks), TNF signaling (in the NM-401 network), and NF-kappa B signaling (in the NRCWE-26 network).

The second type of modules for inflammatory immune response contained the genes, which were associated with anti-viral activity or activation of cytosolic pattern recognition receptors (modules 14, 11 and 19 in the NM-401, NRCWE-26 and BLM networks,

respectively). The temporal changes of gene expression varied in different networks. The genes from module 14 in the NM-401 network were mainly altered on day 3 in response to middle and high doses of NM-401. Module 11 in the NRCWE-26 network contained genes, the expression of which was changed on day 3 and at a later time point (day 28) in response to low and middle doses of NRCWE-26, while the high dose of the CNT triggered expression changes only on day 3. Noteworthy, the response to NRCWE-26 CNT in this module was dose dependent, but in contrast with the first type of modules, the effect was more pronounced at the low and middle doses than at the high dose of this CNT. Module 19 in the BLM network included genes, the expression of which was altered on day 1 and reached the highest point on day 2 (Figure 3C, 4C).

The immune response was accompanied by initiation of DNA damage processes (modules 6, 9 and 3 in the NM-401, NRCWE-26 and BLM networks, respectively). DEGs in these modules were mainly altered on day 3 (CNT networks) and on days 2-21 (BLM network). These DEGs contained genes from cell cycle, homologous recombination and DNA replication. The necroptosis pathway was identified only in the CNT network module. DNA damage associated effects were dose dependent, and the response increased with increasing dose of agents, while components of the necroptosis signaling pathway were triggered mainly by low doses of CNTs.

At late time points, expression of many genes did not return to normal levels. Module 9 in the NM-401 network consisted of sufficiently high numbers of up- and down-regulated DEGs (see Figure 3A). Among the enriched pathways in this module, we identified the lysosome and phagosome signaling pathways, which were altered on day 3 and were active up to the day 28. BLM also altered the lysosome pathway, but mainly on day 7 (module 5) with subsequent attenuation of the effect. In addition, at late time points, the BLM network included two modules (modules 6 and 1) with a high number of inhibited

genes (see heatmap in Figure 3C). KEGG pathway enrichment analysis of DEGs from module 6 identified the natural killer cytotoxicity pathway, while module 1 included DEGs from the pathways, which are associated with the regulation and function of muscle cells (Figure 4C). Actin and myosin family of genes were mapped to these pathways (see Supplementary file 2).

Prioritization of gene regulators of fibrosis

In order to prioritize gene regulators (TFs) associated with activation of fibrotic processes, we next focused on a specific part of each GRN that included pulmonary fibrosis markers and directly linked TFs. As fibrosis markers, we used 94 pulmonary fibrosis-associated genes derived from the Comparative Toxicogenomics Database (Davis *et al.*, 2017), which contains manually curated gene–chemicals/nanoparticles–disease associations. This gene set includes several matrix metallopeptidases and their inhibitors, interleukins, chemokines, ECM regulators, and other genes which are involved in parenchymal injury of the lung (this list of genes is provided in Supplementary file 3). We considered these 94 genes as targets and found their regulators using the inferred GRNs for NM-401, NRCWE-26 and BLM. Each TF was characterized using the number of direct connections (connectivity) with fibrosis markers for each condition (time points/doses), provided that the target gene (fibrosis marker) was differentially expressed. A list of identified TFs with their network modules and connectivity values for different conditions are presented in the Supplementary file 5. Hierarchical clustering of TFs based on these connectivity values was used to generate a heatmap (see Figure 5). The heatmap columns can be separated into three main groups: (1) BLM treatment responses, (2) responses to the high dose of NM-401 on day 3, and (3) other responses to NM-401 treatment and responses to NRCWE-26. TFs that form heatmap rows clustered into two main groups, namely the BLM cluster and the CNT cluster.

The BLM cluster contains the TFs *Litaf, Mafb* and *Batf3*. *Litaf* regulates the expression of cytokines, pro-inflammatory and pro-fibrogenic genes (Ceccarelli *et al.*, 2015; Tang *et al.*, 2005, 2011). Transcription of *Litaf* can be induced by the tumor suppressor p53 (Myokai *et al.*, 1999) and Toll-Like receptors (TLR 2/4) (Tang *et al.*, 2006). In the CNT networks, *Litaf* is directly connected with *Cd14* and *Myd88* genes (see Supplementary file 6), which encode toll-like receptor interacting proteins. *Mafb* has connections with fibrosis markers in all networks, but connectivity intensities were more prominent in the NM-401 and BLM GRNs (Figure 5). In all GRNs, *Mafb* is directly connected with Fcgamma receptor (FCGR) coding genes, including *Fcgr3* and *Fcgr2b* genes (see Supplementary file 6). Published data show, that *Mafb* can enhance phagocytic activity of macrophages by stimulating Fcgr3 (Nemoto *et al.*, 2017) and has a key role in the activation of anti-inflammatory macrophage profile by inducing M1/M2 macrophage polarization, which is important for fibrosis development (Cuevas *et al.*, 2017; Kim, 2017). The other TF, *Batf3*, which has a high number of connections with fibrosis markers in the NM-401 and BLM networks, is involved in the formation of $CD103^{+}$ and $CD8^{+}$ dendritic cells that may facilitate lung fibrosis (Bantsimba-Malanda *et al.*, 2010; Kopf *et al.*, 2015). In line with this result, liver fibrosis is attenuated in *Batf3* $^{-/-}$ knockout mice (Chen *et al.*, 2017).

The CNT cluster is partitioned into two subclusters (see Figure 5), which denote TFs with high connectivity for (1) both CNTs and (2) specific to NM-401. The TFs, *Arid5a, Myc* and Nme2 expose a high number of connections with fibrosis markers in the CNT networks. *Arid5a* also has a high number of connections with fibrosis markers in the CNT networks (see Figure 5). *Arid5a* controls IL6 mRNA stability and protects IL6 mRNA from regnase-1-mediated degradation (Masuda *et al.*, 2013). Importantly, *Arid5a* is regulated by the NF-κB and MAPK signaling pathways, which in turn are activated by Toll-like receptor 4 (Nyati *et al.*, 2017). In the NM-401 GRN, *Arid5a* and *Myc* colocalized *in the same module. Myc* was

identified in the top 10 TF hubs with the highest number of connections in all CNT networks (Table 3), indicating that Myc plays a critical role in responses to CNT treatment. In contrast, in the BLM network Myc was not linked with fibrosis markers. *Myc* activity is essential for cell cycle progression, apoptosis and other biological processes. Myc has cross-regulatory interactions with cytokines, including IL1, IL2, IL4, IL6, IL8, IL10, and TNF-α (Liu *et al.*, 2015). Interestingly, another TF from this cluster is *Nme2,* which controls *Myc* transcription (Yao *et al.*, 2014).

The TFs *Srebf2, Thrb, Atf3, Hif1a, Thyn1 and Mxd1,* are presented in the NM-401-specific subcluster. The connectivity of *Srebf2*, also known as *Srebp2*, rose as the NM-401 dose increased. *Srebf2* induces the expression of genes that are involved in cholesterol and fatty acid synthesis, cholesterol transport (Fessler, 2017; Madison, 2016), and the formation of lipid-laden macrophages (foam cells) (Li *et al.*, 2013), which are associated with lung fibrosis (Romero *et al.*, 2015). *Thrb* regulates biological functions of thyroid hormone, which can be involved in induction of oxidative stress and inflammation (Mancini *et al.*, 2016). *Atf3* is involved in TLR2/4 downstream signaling and plays essential roles in the regulation of stress-induced responses by inhibiting IL6, TNF-α and other cytokines expressions, in apoptosis regulation (Kondo *et al.*, 2012; Thompson *et al.*, 2009). Hif1a is an oxidative stress sensitive TF that may initiate fibrosis by facilitating epithelial- mesenchymal transition (Cummins *et al.*, 2016; Xiong and Liu, 2017). *Thyn1*, also known as Thy28, is an apoptosis-associated gene (Toyota *et al.*, 2012).

Another TF with high number of connections in the NM-401 network was *Mxd1,* also known as *Mad*. This TF is closely involved in the cell regulation by MYC (Lüscher and Vervoorts, 2012). Heterodimerization of MYC and MAX is necessary for activation of MYC target genes. The protein MAD, which is encoded by the *Mxd1* gene, competes with MYC for binding to MAX and thereby inhibits MYC activity (Lüscher and Vervoorts, 2012). Analysis

of the log fold change (logFC) values for *Myc, Max, Mxd1* and cytokines showed that the expression changes of the most upregulated cytokines (*IL6, IL1b, Cxcl10* and *Ccl12)* were accompanied by the concurrent upregulation of *Myc* and *Mxd1* genes on day 3 in response to high NM-401 doses (Supplementary file 4).

Discussion

Approaches that are widely used in nanotoxicology for the analysis of gene expression data commonly include the identification of DEGs and subsequent enrichment analysis of signaling pathways and GO terms for functional interpretation of these changes. However, these methods do not provide information about key regulators, which orchestrate transcriptional stress responses to NMs exposure. Here, we used an integrative computational approach based on reverse engineering algorithms and microarray mRNA data to analyze biological responses associated with NM-induced toxicity. We applied two criteria for selecting data sets from GEO data base for our analysis: (1) NMs that induce fibrosis; (2) that the number of samples was as large as possible. The second criterion is required to enhance performance in predicting the mutual dependencies between genes. In addition, to improve prediction accuracy, we restricted candidate regulators to TFs. The prediction accuracy can also be improved by a combination of results from multiple statistical algorithms, especially using diverse type of methodologies (Marbach *et al.*, 2012). Therefore, our approach includes a step with integration of results from different algorithms from diverse categories, which showed the best prediction performance in Marbach's study, namely regression, mutual information and tree-based approaches.

We applied this ensemble approach to NM-401, NRCWE-26 and BLM experimental data sets, inferred GRNs and performed a two-stage analysis: (1) we examined biological

functions of genes in GRN modules; and (2) we prioritized TFs associated with the activation of fibrotic processes.

Functional characterization of GRNs identified several altered gene modules that belong to stress-responses, such as early immune response, response to DNA damage/integrity, and apoptotic pathways (necroptosis, cellular senescence). Interestingly, the virus immune response was also among these stress-responses. In particular, the activation of cytosolic DNA-sensing receptor and Influenza A pathways were strongly stimulated by CNTs and BLM instillation. This effect might be related to CNT- and BLM-induced mitochondrial damage. Oxidative stress induced by CNTs or BLM damages the mtDNA in cells (Szczesny *et al.*, 2018). Damaged mtDNA serves releases into the cytoplasm, which is sensed by cytosolic DNA-sensing receptors and triggers the interferon signaling pathway. Moreover, mtDNA fragments are excreted by exosomes into the extracellular space and promote an inflammatory response in other cells (Boyapati *et al.*, 2017).

Clustering analysis of TFs associated with activation of fibrotic processes showed co-localization of BLM samples and NM-401 samples for the 3rd day of NM-401 exposure (Figure 5). We conclude that the 3rd day of NM-401 exposure determines the early switch to the fibrosis state. Since NM-401 induced much stronger fibrotic changes compared to NRCWE-26, we consider the following list of TFs as early predictors of toxic responses after CNT exposure: *Mafb, Litaf, Batf3, Nme2-Myc-Mxd1, Thyn1, Hif1a, Stat5b, Egr2, Plagl1, Atf3, Etv5, Thrb, Srebf2.* We included into this list TFs, which have high connectivity values primarily in NM-401 GRN (Figure 5).

In conclusion, we have developed an integrative computational approach to prioritize key transcription regulators, their associated biological processes and signaling pathways, which were altered in response to toxic compounds. We applied omics-based tools using a systems biology approach which can have a pivotal role in moving toxicity testing away from *in vivo* to

in vitro and *in silico* models (Nymark *et al.*, 2018). Our method uses transcriptomics data, generates interaction networks that are specific to each toxic agent, and is independent from bias in the reference databases for pathway mapping as it infers connections and pathways *de novo* purely based on the data. Inferred networks can be used as a basis for further integration with proteomics data that might enhance the power of NMs toxicity prediction models (Riebeling *et al.*, 2017). Moreover, this approach and inferred interaction networks can be useful for generating toxicity pathways and adverse outcome pathway schemes for toxic NMs.

List of abbreviations

BLM: bleomycin

CNTs: Carbon nanotubes

DEGs: differentially expressed genes

ECM: extracellular matrix

FCGR: Fcgamma receptor

GRN: gene regulatory network

PRRs: pattern recognition receptors

TF: transcription factor

Availability of data and material

All experimental data is available at the Gene Expression Omnibus database repository (accession number GSE55286, GSE40151). Other data supporting the conclusions of this

article are included within the article and its supplementary files (available from the Dryad repository at doi: 10.5061/dryad.3b025hh).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the EU H2020 grants SmartNanoTox (Grant no. 686098) and NanoCommons (Grant no. 731032).

References

Andersen, M.E., McMullen, P.D., and Bhattacharya, S. (2013). Toxicogenomics for transcription factor-governed molecular pathways: moving on to roles beyond classification and prediction. *Arch Toxicol*. **87**, 7–11.

Bantsimba-Malanda, C., Marchal-Sommé, J., Goven, D., Freynet, O., Michel, L., Crestani, B., and Soler, P. (2010). A role for dendritic cells in bleomycin-induced pulmonary fibrosis in mice? *Am J Respir Crit Care Med*. **182**, 385–395.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat Genet*. **37**, 382–390.

Borate, B.R., Chesler, E.J., Langston, M.A., Saxton, A.M., and Voy, B.H. (2009). Comparison of threshold selection methods for microarray gene co-expression matrices. *BMC Res Notes*. **2**, 240.

Boyapati, R.K., Tamborska, A., Dorward, D.A., and Ho, G.-T. (2017). Advances in the understanding of mitochondrial DNA as a pathogenic factor in inflammatory diseases. [version 1; peer review: 3 approved]. *F1000Res*. **6**, 169.

Ceccarelli, S., Panera, N., Mina, M., Gnani, D., Stefanis, C. De., Crudele, A., Rychlicki, C., Petrini, S., Bruscalupi, G., Agostinelli, L., *et al.* (2015). LPS-induced TNF-α factor mediates pro-inflammatory and pro-fibrogenic pattern in non-alcoholic fatty liver disease. *Oncotarget*. **6**, 41434–41452.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., *et al.* (2014). Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell*. **159**, 402–414.

Chen, L., Zhang, D., Zhang, W., Zhu, Y., Hou, M., Yang, B., Xu, Z., Ji, M., and Wu, G. (2017). Absence of Batf3 results in reduced liver pathology in mice infected with Schistosoma japonicum. *Parasit Vectors*. **10**, 306.

Chepelev, N.L., Moffat, I.D., Labib, S., Bourdon-Lacombe, J., Kuo, B., Buick, J.K., Lemieux, F., Malik, A.I., Halappanavar, S., Williams, A., *et al.* (2015). Integrating toxicogenomics into human health risk assessment: lessons learned from the benzo[a]pyrene case study. *Crit Rev Toxicol*. **45**, 44–52.

Cuevas, V.D., Anta, L., Samaniego, R., Orta-Zavalza, E., Vladimir de la Rosa, J., Baujat, G., Domínguez-Soto, Á., Sánchez-Mateos, P., Escribese, M.M., Castrillo, A., *et al.* (2017). MAFB Determines Human Macrophage Anti-Inflammatory Polarization: Relevance for the Pathogenic Mechanisms Operating in Multicentric Carpotarsal Osteolysis. *J Immunol*. **198**, 2070–2081.

Cummins, E.P., Keogh, C.E., Crean, D., and Taylor, C.T. (2016). The role of HIF in immunity and inflammation. *Mol Aspects Med*. **47-48**, 24–34.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., King, B.L., McMorran, R., Wiegers, J., Wiegers, T.C., and Mattingly, C.J. (2017). The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res*. **45**, D972–D978.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. **23**, 1846–1847.

Dong, J., and Ma, Q. (2015). Advances in mechanisms and signaling pathways of carbon nanotube toxicity. *Nanotoxicology*. **9**, 658–676.

Dong, J., and Ma, Q. (2016). Myofibroblasts and lung fibrosis induced by carbon nanotube exposure. *Part Fibre Toxicol*. **13**, 60.

Fessler, M.B. (2017). A new frontier in immunometabolism. cholesterol in lung health and disease. *Annals of the American Thoracic Society*. **14**, S399–S405.

Girvan, M., and Newman, M.E.J. (2002). Community structure in social and biological networks. *Proc Natl Acad Sci USA*. **99**, 7821–7826.

Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q., and Guo, A.-Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res*. **47**, D33–D38.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. **5**,.

Jennings, P., Limonciel, A., Felice, L., and Leonard, M.O. (2013). An overview of transcriptional regulation in response to toxicological insult. *Arch Toxicol*. **87**, 49–72.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. **45**, D353–D361.

Kim, H. (2017). The transcription factor MafB promotes anti-inflammatory M2 polarization and cholesterol efflux in macrophages. *Sci Rep*. **7**, 7591.

Kondo, T., Kawai, T., and Akira, S. (2012). Dissecting negative regulation of Toll-like receptor signaling. *Trends Immunol*. **33**, 449–458.

Kopf, M., Schneider, C., and Nobs, S.P. (2015). The development and function of lung-resident macrophages and dendritic cells. *Nat Immunol*. **16**, 36–44.

Lachmann, A., Giorgi, F.M., Lopez, G., and Califano, A. (2016). ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*. **32**, 2233–2235.

Langfelder, P., Mischel, P.S., and Horvath, S. (2013). When is hub gene selection better than standard meta-analysis? *PLoS One*. **8**, e61505.

Li, L.-C., Varghese, Z., Moorhead, J.F., Lee, C.-T., Chen, J.-B., and Ruan, X.Z. (2013). Cross-talk between TLR4-MyD88-NF-κB and SCAP-SREBP2 pathways mediates macrophage foam cell formation. *Am J Physiol Heart Circ Physiol*. **304**, H874–84.

Liu, T., Zhou, Y., Ko, K.S., and Yang, H. (2015). Interactions between Myc and Mediators of Inflammation in Chronic Liver Diseases. *Mediators Inflamm*. **2015**, 276850.

Lüscher, B., and Vervoorts, J. (2012). Regulation of gene transcription by the oncoprotein MYC. *Gene*. **494**, 145–160.

Madison, B.B. (2016). Srebp2: A master regulator of sterol and fatty acid synthesis. *J Lipid Res*. **57**, 333–335.

Mancini, A., Segni, C. Di., Raimondo, S., Olivieri, G., Silvestrini, A., Meucci, E., and Currò, D. (2016). Thyroid hormones, oxidative stress, and inflammation. *Mediators Inflamm*. **2016**, 6757154.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., DREAM5 Consortium., Kellis, M., Collins, J.J., *et al.* (2012). Wisdom of crowds for robust gene network inference. *Nat Methods*. **9**, 796–804.

Marwah, V.S., Kinaret, P.A.S., Serra, A., Scala, G., Lauerma, A., Fortino, V., and Greco, D. (2018). Inform: inference of network response modules. *Bioinformatics*. **34**, 2136–2138.

Masuda, K., Ripley, B., Nishimura, R., Mino, T., Takeuchi, O., Shioi, G., Kiyonari, H., and Kishimoto, T. (2013). Arid5a controls IL-6 mRNA stability, which contributes to elevation of IL-6 level in vivo. *Proc Natl Acad Sci USA*. **110**, 9409–9414.

McDermott, J.E., Archuleta, M., Thrall, B.D., Adkins, J.N., and Waters, K.M. (2011). Controlling the response: predictive modeling of a highly central, pathogen-targeted core response module in macrophage activation. *PLoS One*. **6**, e14673.

Myokai, F., Takashiba, S., Lebo, R., and Amar, S. (1999). A novel lipopolysaccharide-induced transcription factor regulating tumor necrosis factor alpha gene expression: molecular cloning, sequencing, characterization, and chromosomal assignment. *Proc Natl Acad Sci USA*. **96**, 4518–4523.

Nemoto, T., Shibata, Y., Inoue, S., Igarashi, A., Tokairin, Y., Yamauchi, K., Kimura, T., Sato, M., Sato, K., Nakano, H., *et al.* (2017). MafB enhances the phagocytic activity of RAW264.7 macrophages by promoting Fcgr3 expression. *Biochem Biophys Res Commun*. **482**, 375–381.

Nikota, J., Williams, A., Yauk, C.L., Wallin, H., Vogel, U., and Halappanavar, S. (2016). Meta-analysis of transcriptomic responses as a means to identify pulmonary disease outcomes for engineered nanomaterials. *Part Fibre Toxicol*. **13**, 25.

Nyati, K.K., Masuda, K., Zaman, M.M.-U., Dubey, P.K., Millrine, D., Chalise, J.P., Higa, M., Li, S., Standley, D.M., Saito, K., *et al.* (2017). TLR4-induced NF-κB and MAPK signaling regulate the IL-6 mRNA stabilizing protein Arid5a. *Nucleic Acids Res*. **45**, 2687–2703.

Nymark, P., Kohonen, P., Hongisto, V., and Grafström, R.C. (2018). Toxic and genomic influences of inhaled nanomaterials as a basis for predicting adverse outcome. *Annals of the American Thoracic Society*. **15**, S91–S97.

Peng, R., Sridhar, S., Tyagi, G., Phillips, J.E., Garrido, R., Harris, P., Burns, L., Renteria, L., Woods, J., Chen, L., *et al.* (2013). Bleomycin induces molecular changes directly relevant to idiopathic pulmonary fibrosis: a model for “active” disease. *PLoS One*. **8**, e59348.

Perkins, E.J., Chipman, J.K., Edwards, S., Habib, T., Falciani, F., Taylor, R., Aggelen, G. Van., Vulpe, C., Antczak, P., and Loguinov, A. (2011). Reverse engineering adverse outcome pathways. *Environ Toxicol Chem*. **30**, 22–38.

Pesquita, C. (2017). Semantic similarity in the gene ontology. *Methods Mol Biol*. **1446**, 161–173.

Poulsen, S.S., Saber, A.T., Williams, A., Andersen, O., Køhler, C., Atluri, R., Pozzebon, M.E., Mucelli, S.P., Simion, M., Rickerby, D., *et al.* (2015). MWCNTs of different physicochemical properties cause similar inflammatory responses, but differences in transcriptional and histological markers of fibrosis in mouse lungs. *Toxicol Appl Pharmacol*. **284**, 16–32.

Rahman, L., Jacobsen, N.R., Aziz, S.A., Wu, D., Williams, A., Yauk, C.L., White, P., Wallin, H., Vogel, U., and Halappanavar, S. (2017). Multi-walled carbon nanotube-induced genotoxic, inflammatory and pro-fibrotic responses in mice: Investigating the mechanisms of pulmonary carcinogenesis. *Mutat Res*. **823**, 28–44.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*. **44**, W83–9.

Riebeling, C., Jungnickel, H., Luch, A., and Haase, A. (2017). Systems biology to support nanomaterial grouping. *Adv Exp Med Biol*. **947**, 143–171.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**, e47.

Romero, F., Shah, D., Duong, M., Penn, R.B., Fessler, M.B., Madenspacher, J., Stafstrom, W., Kavuru, M., Lu, B., Kallen, C.B., *et al.* (2015). A pneumocyte-macrophage paracrine lipid axis drives the lung toward fibrosis. *Am J Respir Cell Mol Biol*. **53**, 74–86.

Santra, T. (2014). A bayesian framework that integrates heterogeneous data for inferring gene regulatory networks. *Front Bioeng Biotechnol*. **2**, 13.

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.

Souza, T.M., van den Beucken, T., Kleinjans, J.C.S., and Jennen, D.G.J. (2017). Inferring transcription factor activity from microarray data reveals novel targets for toxicological investigations. *Toxicology*. **389**, 101–107.

Su, G., Kuchinsky, A., Morris, J.H., States, D.J., and Meng, F. (2010). GLay: community structure analysis of biological networks. *Bioinformatics*. **26**, 3135–3137.

Szczesny, B., Marcatti, M., Ahmad, A., Montalbano, M., Brunyánszki, A., Bibli, S.-I., Papapetropoulos, A., and Szabo, C. (2018). Mitochondrial DNA damage and subsequent activation of Z-DNA binding protein 1 links oxidative stress to inflammation in epithelial cells. *Sci Rep*. **8**, 914.

Tang, X., Marciano, D.L., Leeman, S.E., and Amar, S. (2005). LPS induces the interaction of a transcription factor, LPS-induced TNF-alpha factor, and STAT6(B) with effects on multiple cytokines. *Proc Natl Acad Sci USA*. **102**, 5132–5137.

Tang, X., Metzger, D., Leeman, S., and Amar, S. (2006). LPS-induced TNF-alpha factor (LITAF)-deficient mice express reduced LPS-induced cytokine: Evidence for LITAF-dependent LPS signaling pathways. *Proc Natl Acad Sci USA*. **103**, 13777–13782.

Tang, X., Yang, Y., and Amar, S. (2011). Novel regulation of CCL2 gene expression by murine LITAF and STAT6B. *PLoS One*. **6**, e25083.

Thompson, M.R., Xu, D., and Williams, B.R.G. (2009). ATF3 transcription factor and its emerging roles in immunity and cancer. *J Mol Med*. **87**, 1053–1060.

Toyota, H., Jiang, X.-Z., Asakura, H., and Mizuguchi, J. (2012). Thy28 partially prevents apoptosis induction following engagement of membrane immunoglobulin in WEHI-231 B lymphoma cells. *Cell Mol Biol Lett*. **17**, 36–48.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J.P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinformatics*. **19**, 575–592.

Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S., Nepelska, M., *et al.* (2014). Adverse outcome pathway (AOP) development I: strategies and principles. *Toxicol Sci*. **142**, 312–320.

Volder, M.F.L. De., Tawfick, S.H., Baughman, R.H., and Hart, A.J. (2013). Carbon nanotubes: present and future commercial applications. *Science*. **339**, 535–539.

Williamson, J.D., Sadofsky, L.R., and Hart, S.P. (2015). The pathogenesis of bleomycin-induced lung injury in animals and its applicability to human idiopathic pulmonary fibrosis. *Exp Lung Res*. **41**, 57–73.

Xiong, A., and Liu, Y. (2017). Targeting Hypoxia Inducible Factors-1α As a Novel Therapy in Fibrosis. *Front Pharmacol*. **8**, 326.

Yao, Y., Li, C., Zhou, X., Zhang, Y., Lu, Y., Chen, J., Zheng, X., Tao, D., Liu, Y., and Ma, Y. (2014). PIWIL2 induces c-Myc expression by interacting with NME2 and regulates c-Myc-mediated tumor cell proliferation. *Oncotarget*. **5**, 8466–8477.

Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. **26**, 976–978.

Tables

Table 1. Data sets from Gene Expression Omnibus Database used in the paper.

NM/ chemical	Data set	Publication	Mouse strain	Number of samples	Post-exposure time point, days	Doses	Type of microarray
NM-401	GSE55286	(Poulsen et al. 2015)	C57BL/6	70	1, 3, 28	18, 54, 162 µg, vehicle control	Agilent SurePrint G3 Mouse GE 8x60K Microarray
NRCWE-26	GSE55286	(Poulsen et al. 2015)	C57BL/6	69	1, 3, 28	18, 54, 162 µg, vehicle control	Agilent SurePrint G3 Mouse GE 8x60K Microarray
Bleomycin	GSE40151	(Peng et al. 2013)	C57BL/6	111	1, 2, 7, 14, 21, 28, 35	2U/kg, vehicle control	Affymetrix Mouse 430 2.0 arrays

Table 2. General features of the inferred networks. *The analysis of network features was performed using the NetworkAnalyzer plugin available within the Cytoscape software*.

	NM-401	**NRCWE-26**	**BLM**
Initial gene expression matrix			
Number of genes x sample size	24163 x 70	24163 x 69	17358 x 111
Inferring GRNs			
Number of genes which were used for inferring networks	4768	5063	4771
Number of predicted TF-gene pairs in common network	84202	87012	81280
Identified cut-off based on GO similarity, number of top rows	7700	5700	6500
Network characteristics for final GRNs			
Nodes x edges	3076 x 5587	3045 x 5584	2929 x 5623
Transcription factors	238	244	227
Clustering coefficients	0.073	0.086	0.088
Network diameter	12	11	21
Shortest paths	35601	36576	57598
Characteristic path length	3.641	4.168	6.143
Avg. number of neighbors	3.633	3.668	3.840

Table 3. Prioritized list of TFs with high number of connections in the networks. *The outdegree columns show the number of connections for listed TF in each network, the rank columns denote the ranks of these TFs based on the number of connections (a rank increases with the decrease in the connection number). The last column depicts the average rank of the TF outdegree for all networks. TFs are shown according to increasing the average rank values.*

TFs	Outdegree			Rank			Average rank
	NM-401	NRCWE-26	BLM	NM-401	NRCWE-26	BLM	
E2f8	168	177	138	2	4	10	5.33
Litaf	165	100	217	3	11	2	5.33
Foxm1	116	125	156	7	7	6.5	6.83
Mxd3	93	102	86	12	10	16	12.67
Myc	107	192	54	9	2	30	13.67
Irf7	98	187	44	11	3	39	17.67
Mis18bp1	47	75	167	31.5	16.5	5	17.67
Thyn1	153	36	156	4	43	6.5	17.83
E2f7	36	116	113	40.5	8	12	20.17
Mafb	65	29	207	21.5	55.5	3	26.67
Plek	104	13	88	10	108.5	15	44.5
Batf3	39	7	233	36.5	144	1	60.5
Nme2	208	72	0	1	20	222.5	81.17
Srebf2	141	67	0	5	22	222.5	83.17
Thrb	133	32	0	6	51.5	222.5	93.33
Hes2	111	28	0	8	58.5	222.5	96.33
Hoxb5	35	0	183	42.5	245	4	97.17
Irf8	0	143	27	242.5	6	55	101.17
Mecp2	11	3	147	124	176	8	102.67
Gatad1	20	156	0	81	5	222.5	102.83
Sp110	11	193	0	124	1	222.5	115.83
Maff	12	104	0	118	9	222.5	116.5
Arid3b	0	0	142	242.5	245	9	165.5

Figures

Figure 1. Workflow schematics for this study. (A) *To infer a GRN from transcriptomics data, reverse engineering algorithms were applied. We used three different algorithms for GRN inferring (BVS, GENIE3 and ARACNE-AP) and then integrated the results of all three algorithms by Borda count ranking. Optimal threshold cutoff for selecting ranked gene pairs was calculated using a gene ontology (GO) semantic similarity approach.* ***(B)*** *To identify network modules, the GLay community clustering algorithm was applied. Heatmaps for DEGs were plotted for each network module. KEGG pathway enrichment analysis was performed to identify the biological functions of the genes in the modules.* ***(C)*** *To prioritize gene regulators which are associated with fibrosis, we identified TFs directly connected with fibrosis markers in the inferred networks. Each TF from this list was characterized using the number of connections with fibrosis markers for each condition (time points/doses), provided that these markers were differentially expressed. The visualization of this scheme is provided in the panel C, where green diamonds represent TFs and red circles denote differentially expressed markers. Each identified TF was characterized as a set of connectivity values for each condition. To compare the connectivity patterns of alternative treatments and different TFs, we applied hierarchical clustering.*

Figure 2. GO similarity scores. A) The NM-401 GRN, B) The NRCWE-26 GRN, and C) The BLM GRN. *Black dots show the median values of 100 row slots. The blue line is the 4th order polynomial regression of GO similarity score values. The bigger value means the higher level of GO similarity. A dashed line represents the calculated optimal threshold for selecting ranked gene pairs with high mutual dependencies. The threshold was defined using the fitted curve as a row number in which a GO similarity score exceeded the median value plus standard deviation divided by 2 (see Materials and Methods).*

Figure 3. Inferred gene regulatory networks for CNTs and BLM. *A) NM-401, B) NRCWE-26, C) BLM treatments. Colored nodes represent genes. Groups of nodes with the same color denote distinct network modules. TFs are shown in white font and the size of the font is proportional to the number of their external connections (outdegree). The KEGG database was used for the functional*

annotation of altered genes in the modules, the top enriched pathways are shown. The heatmaps represent fold-changes of DEGs for the network modules.

Figure 4. KEGG pathway enrichment analysis of DEGs in A) NM-401, B) NRCWE-26 and C) BLM network modules. *The values in the tables indicate -log10(BH corrected p-value), the red color scale denotes the magnitude of the values. The analysis was performed for each condition (time point; dose) separately. The following requirements were applied for including the enriched KEGG pathways in the tables: (i) if the BH corrected p-value was <0.0005 in any conditions, all enriched KEGG pathways were included; (ii) if the BH corrected p-value was <0.005, only the top 5 enriched pathways were included.*

Figure 5. **Hierarchical clustering of TFs.** Heatmap intensities show the number of connections between a TF and its direct gene target in a network. Columns in the heatmap represent samples and different conditions, rows denote TFs. A hierarchical clustering algorithm was applied using the Euclidian distance measure.

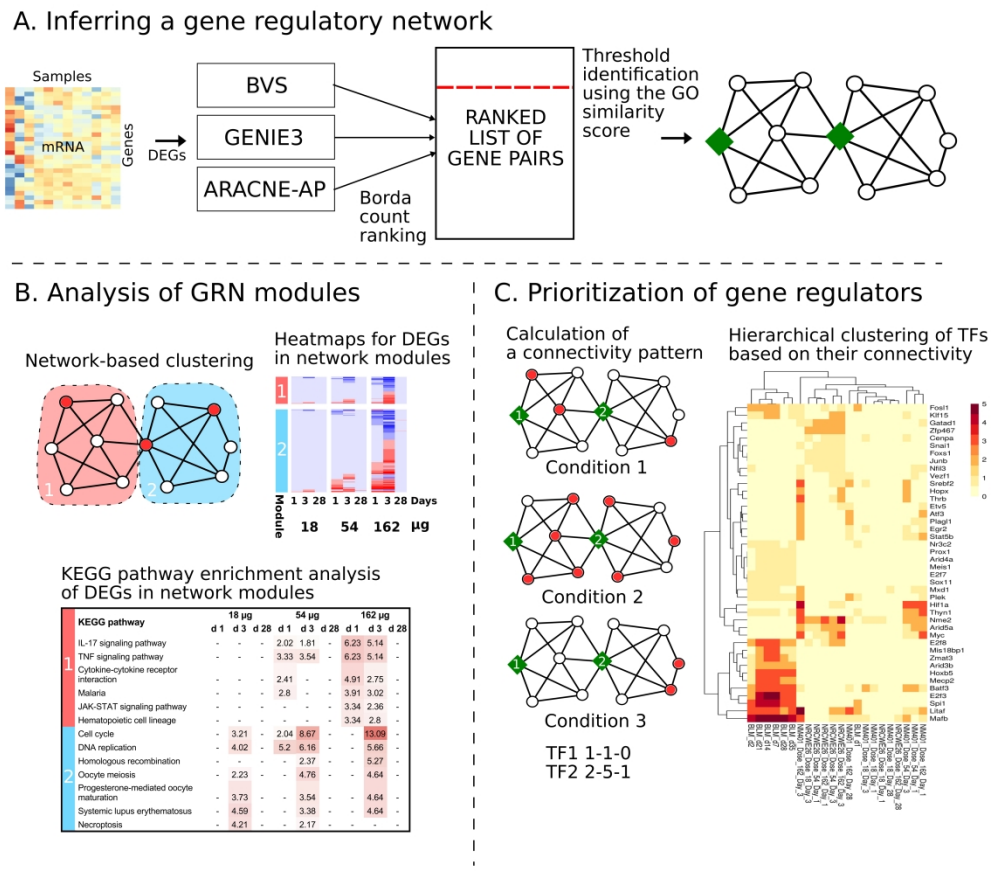


Figure 1. Workflow schematics for this study.

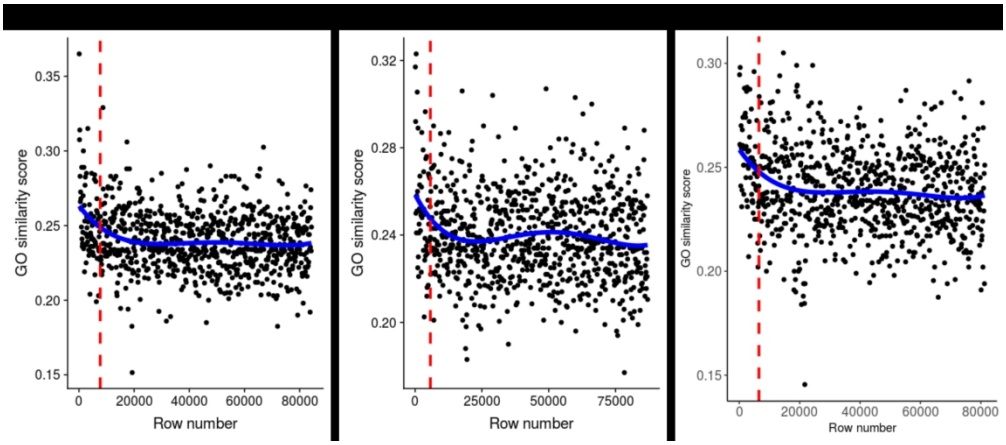


Figure 2. GO similarity scores.

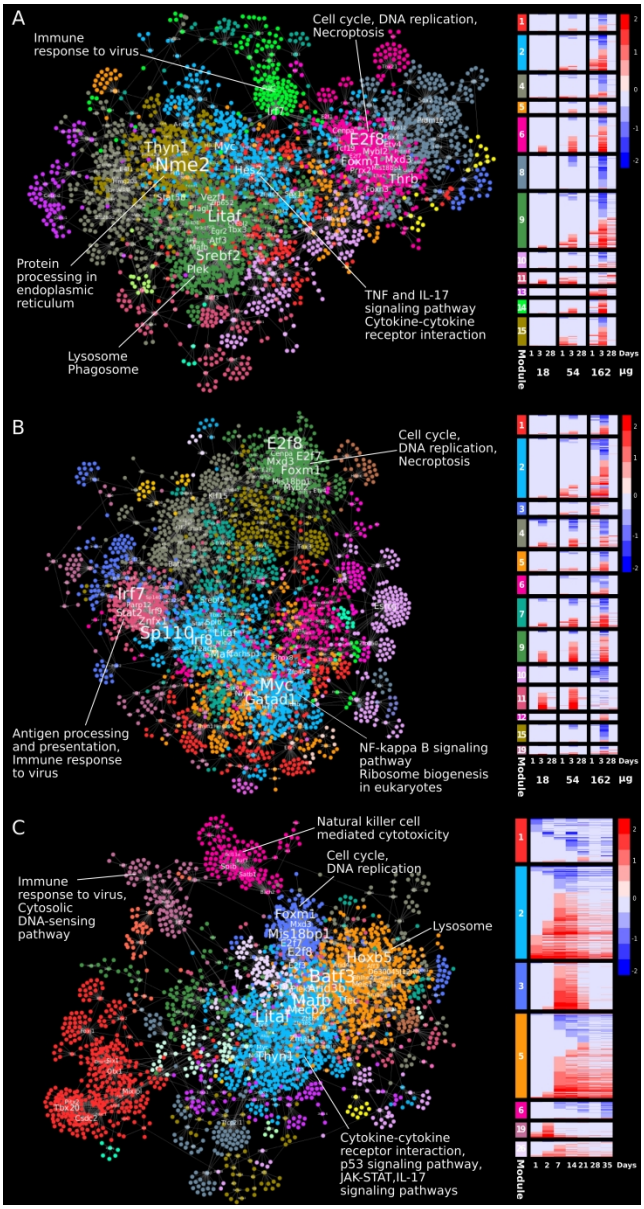
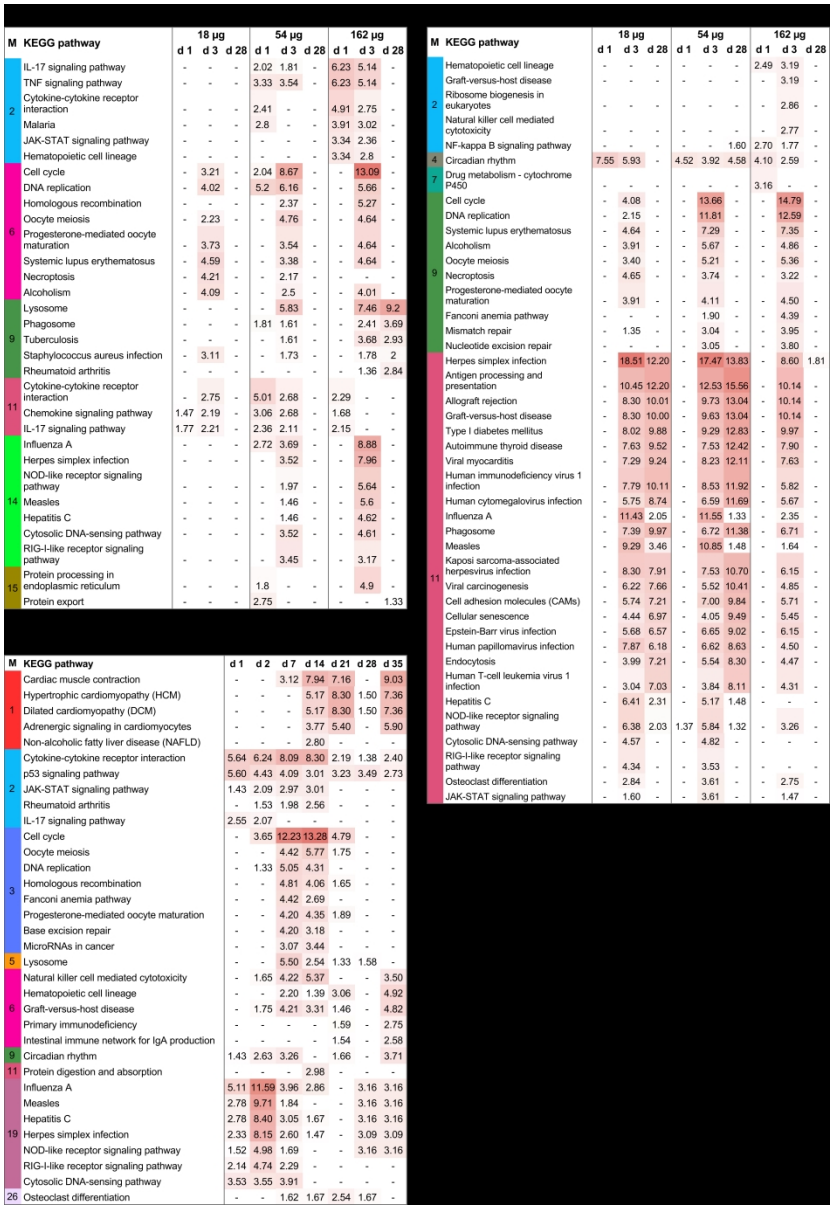


Figure 3. Inferred gene regulatory networks for CNTs and BLM.



M	KEGG pathway	18 µg			54 µg			162 µg		
		d 1	d 3	d 28	d 1	d 3	d 28	d 1	d 3	d 28
	IL-17 signaling pathway	-	-	-	2.02	1.81	-	6.23	5.14	-
	TNF signaling pathway	-	-	-	3.33	3.54	-	6.23	5.14	-
2	Cytokine-cytokine receptor interaction	-	-	-	2.41	-	-	4.91	2.75	-
	Malaria	-	-	-	2.8	-	-	3.91	3.02	-
	JAK-STAT signaling pathway	-	-	-	-	-	-	3.34	2.36	-
	Hematopoietic cell lineage	-	-	-	-	-	-	3.34	2.8	-
	Cell cycle	-	3.21	-	2.04	8.67	-	-	13.09	-
	DNA replication	-	4.02	-	5.2	6.16	-	-	5.66	-
	Homologous recombination	-	-	-	-	2.37	-	-	5.27	-
6	Oocyte meiosis	-	2.23	-	-	4.76	-	-	4.64	-
	Progesterone-mediated oocyte maturation	-	3.73	-	-	3.54	-	-	4.64	-
	Systemic lupus erythematosus	-	4.59	-	-	3.38	-	-	4.64	-
	Necroptosis	-	4.21	-	-	2.17	-	-	-	-
	Alcoholism	-	4.09	-	-	2.5	-	-	4.01	-
	Lysosome	-	-	-	-	5.83	-	-	7.46	9.2
	Phagosome	-	-	-	1.81	1.61	-	-	2.41	3.69
9	Tuberculosis	-	-	-	-	1.61	-	-	3.68	2.93
	Staphylococcus aureus infection	-	3.11	-	-	1.73	-	-	1.78	2
	Rheumatoid arthritis	-	-	-	-	-	-	-	1.36	2.84
11	Cytokine-cytokine receptor interaction	-	2.75	-	5.01	2.68	-	2.29	-	-
	Chemokine signaling pathway	1.47	2.19	-	3.06	2.68	-	1.68	-	-
	IL-17 signaling pathway	1.77	2.21	-	2.36	2.11	-	2.15	-	-
	Influenza A	-	-	-	2.72	3.69	-	-	8.88	-
	Herpes simplex infection	-	-	-	-	3.52	-	-	7.96	-
	NOD-like receptor signaling pathway	-	-	-	-	1.97	-	-	5.64	-
14	Measles	-	-	-	-	1.46	-	-	5.6	-
	Hepatitis C	-	-	-	-	1.46	-	-	4.62	-
	Cytosolic DNA-sensing pathway	-	-	-	-	3.52	-	-	4.61	-
	RIG-I-like receptor signaling pathway	-	-	-	-	3.45	-	-	3.17	-
15	Protein processing in endoplasmic reticulum	-	-	-	1.8	-	-	-	4.9	-
	Protein export	-	-	-	2.75	-	-	-	-	1.33

M	KEGG pathway	18 µg			54 µg			162 µg		
		d 1	d 3	d 28	d 1	d 3	d 28	d 1	d 3	d 28
	Hematopoietic cell lineage	-	-	-	-	-	-	2.49	3.19	-
	Graft-versus-host disease	-	-	-	-	-	-	-	3.19	-
2	Ribosome biogenesis in eukaryotes	-	-	-	-	-	-	-	2.86	-
	Natural killer cell mediated cytotoxicity	-	-	-	-	-	-	-	2.77	-
	NF-kappa B signaling pathway	-	-	-	-	-	1.60	2.70	1.77	-
4	Circadian rhythm	7.55	5.93	-	4.52	3.92	4.58	4.10	2.59	-
7	Drug metabolism - cytochrome P450	-	-	-	-	-	-	3.16	-	-
	Cell cycle	-	4.08	-	-	13.66	-	-	14.79	-
	DNA replication	-	2.15	-	-	11.81	-	-	12.59	-
	Systemic lupus erythematosus	-	4.64	-	-	7.29	-	-	7.35	-
	Alcoholism	-	3.91	-	-	5.67	-	-	4.86	-
	Oocyte meiosis	-	3.40	-	-	5.21	-	-	5.36	-
9	Necroptosis	-	4.65	-	-	3.74	-	-	3.22	-
	Progesterone-mediated oocyte maturation	-	3.91	-	-	4.11	-	-	4.50	-
	Fanconi anemia pathway	-	-	-	-	1.90	-	-	4.39	-
	Mismatch repair	-	1.35	-	-	3.04	-	-	3.95	-
	Nucleotide excision repair	-	-	-	-	3.05	-	-	3.80	-
	Herpes simplex infection	-	18.51	12.20	-	17.47	13.83	-	8.60	1.81
	Antigen processing and presentation	-	10.45	12.20	-	12.53	15.56	-	10.14	-
	Allograft rejection	-	8.30	10.01	-	9.73	13.04	-	10.14	-
	Graft-versus-host disease	-	8.30	10.00	-	9.63	13.04	-	10.14	-
	Type I diabetes mellitus	-	8.02	9.88	-	9.29	12.83	-	9.97	-
	Autoimmune thyroid disease	-	7.63	9.52	-	7.53	12.42	-	7.90	-
	Viral myocarditis	-	7.29	9.24	-	8.23	12.11	-	7.63	-
	Human immunodeficiency virus 1 infection	-	7.79	10.11	-	8.53	11.92	-	5.82	-
	Human cytomegalovirus infection	-	5.75	8.74	-	6.59	11.69	-	5.67	-
	Influenza A	-	11.43	2.05	-	11.55	1.33	-	2.35	-
	Phagosome	-	7.39	9.97	-	6.72	11.38	-	6.71	-
	Measles	-	9.29	3.46	-	10.85	1.48	-	1.64	-
	Kaposi sarcoma-associated herpesvirus infection	-	8.30	7.91	-	7.53	10.70	-	6.15	-
11	Viral carcinogenesis	-	6.22	7.66	-	5.52	10.41	-	4.85	-
	Cell adhesion molecules (CAMs)	-	5.74	7.21	-	7.00	9.84	-	5.71	-
	Cellular senescence	-	4.44	6.97	-	4.05	9.49	-	5.45	-
	Epstein-Barr virus infection	-	5.68	6.57	-	6.65	9.02	-	6.15	-
	Human papillomavirus infection	-	7.87	6.18	-	6.62	8.63	-	4.50	-
	Endocytosis	-	3.99	7.21	-	5.54	8.30	-	4.47	-
	Human T-cell leukemia virus 1 infection	-	3.04	7.03	-	3.84	8.11	-	4.31	-
	Hepatitis C	-	6.41	2.31	-	5.17	1.48	-	-	-
	NOD-like receptor signaling pathway	-	6.38	2.03	1.37	5.84	1.32	-	3.26	-
	Cytosolic DNA-sensing pathway	-	4.57	-	-	4.82	-	-	-	-
	RIG-I-like receptor signaling pathway	-	4.34	-	-	3.53	-	-	-	-
	Osteoclast differentiation	-	2.84	-	-	3.61	-	-	2.75	-
	JAK-STAT signaling pathway	-	1.60	-	-	3.61	-	-	1.47	-

M	KEGG pathway	d 1	d 2	d 7	d 14	d 21	d 28	d 35
	Cardiac muscle contraction	-	-	3.12	7.94	7.16	-	9.03
	Hypertrophic cardiomyopathy (HCM)	-	-	-	5.17	8.30	1.50	7.36
1	Dilated cardiomyopathy (DCM)	-	-	-	5.17	8.30	1.50	7.36
	Adrenergic signaling in cardiomyocytes	-	-	-	3.77	5.40	-	5.90
	Non-alcoholic fatty liver disease (NAFLD)	-	-	-	2.80	-	-	-
	Cytokine-cytokine receptor interaction	5.64	6.24	8.09	8.30	2.19	1.38	2.40
	p53 signaling pathway	5.60	4.43	4.09	3.01	3.23	3.49	2.73
2	JAK-STAT signaling pathway	1.43	2.09	2.97	3.01	-	-	-
	Rheumatoid arthritis	-	1.53	1.98	2.56	-	-	-
	IL-17 signaling pathway	2.55	2.07	-	-	-	-	-
	Cell cycle	-	3.65	12.23	13.28	4.79	-	-
	Oocyte meiosis	-	-	4.42	5.77	1.75	-	-
	DNA replication	-	1.33	5.05	4.31	-	-	-
3	Homologous recombination	-	-	4.81	4.06	1.65	-	-
	Fanconi anemia pathway	-	-	4.42	2.69	-	-	-
	Progesterone-mediated oocyte maturation	-	-	4.20	4.35	1.89	-	-
	Base excision repair	-	-	4.20	3.18	-	-	-
	MicroRNAs in cancer	-	-	3.07	3.44	-	-	-
5	Lysosome	-	-	5.50	2.54	1.33	1.58	-
	Natural killer cell mediated cytotoxicity	-	1.65	4.22	5.37	-	-	3.50
	Hematopoietic cell lineage	-	-	2.20	1.39	3.06	-	4.92
6	Graft-versus-host disease	-	1.75	4.21	3.31	1.46	-	4.82
	Primary immunodeficiency	-	-	-	-	1.59	-	2.75
	Intestinal immune network for IgA production	-	-	-	-	1.54	-	2.58
9	Circadian rhythm	1.43	2.63	3.26	-	1.66	-	3.71
11	Protein digestion and absorption	-	-	-	2.98	-	-	-
	Influenza A	5.11	11.59	3.96	2.86	-	3.16	3.16
	Measles	2.78	9.71	1.84	-	-	3.16	3.16
	Hepatitis C	2.78	8.40	3.05	1.67	-	3.16	3.16
19	Herpes simplex infection	2.33	8.15	2.60	1.47	-	3.09	3.09
	NOD-like receptor signaling pathway	1.52	4.98	1.69	-	-	3.16	3.16
	RIG-I-like receptor signaling pathway	2.14	4.74	2.29	-	-	-	-
	Cytosolic DNA-sensing pathway	3.53	3.55	3.91	-	-	-	-
26	Osteoclast differentiation	-	-	1.62	1.67	2.54	1.67	-

Figure 4. KEGG pathway enrichment analysis of DEGs in A) NM-401, B) NRCWE-26 and C) BLM network modules.

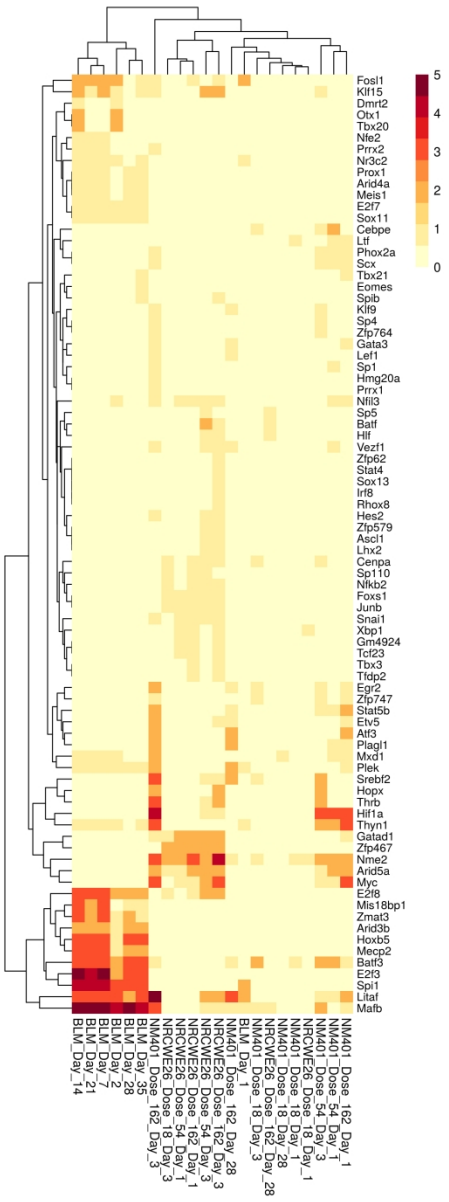


Figure 5. Hierarchical clustering of TFs.