



# Application of biclustering of gene expression data and gene set enrichment analysis methods to identify potentially disease causing nanomaterials

Andrew Williams\* and Sabina Halappanavar

## Full Research Paper

Open Access

Address:  
Environmental Health Science and Research Bureau, Environmental  
and Radiation Health Sciences Directorate, Health Canada, Ottawa  
K1A 0K9, Canada

Email:  
Andrew Williams\* - andrew.williams@canada.ca

\* Corresponding author

Keywords:  
gene expression; risk assessment; toxicogenomics

*Beilstein J. Nanotechnol.* **2015**, *6*, 2438–2448.  
doi:10.3762/bjnano.6.252

Received: 19 August 2015  
Accepted: 30 November 2015  
Published: 21 December 2015

This article is part of the Thematic Series "Nanoinformatics for  
environmental health and biomedicine".

Guest Editor: Y. Cohen

© 2015 Williams and Halappanavar; licensee Beilstein-Institut.  
License and terms: see end of document.

## Abstract

**Background:** The presence of diverse types of nanomaterials (NMs) in commerce is growing at an exponential pace. As a result, human exposure to these materials in the environment is inevitable, necessitating the need for rapid and reliable toxicity testing methods to accurately assess the potential hazards associated with NMs. In this study, we applied biclustering and gene set enrichment analysis methods to derive essential features of altered lung transcriptome following exposure to NMs that are associated with lung-specific diseases. Several datasets from public microarray repositories describing pulmonary diseases in mouse models following exposure to a variety of substances were examined and functionally related biclusters of genes showing similar expression profiles were identified. The identified biclusters were then used to conduct a gene set enrichment analysis on pulmonary gene expression profiles derived from mice exposed to nano-titanium dioxide (nano-TiO<sub>2</sub>), carbon black (CB) or carbon nanotubes (CNTs) to determine the disease significance of these data-driven gene sets.

**Results:** Biclusters representing inflammation (chemokine activity), DNA binding, cell cycle, apoptosis, reactive oxygen species (ROS) and fibrosis processes were identified. All of the NM studies were significant with respect to the bicluster related to chemokine activity (DAVID; FDR p-value = 0.032). The bicluster related to pulmonary fibrosis was enriched in studies where toxicity induced by CNT and CB studies was investigated, suggesting the potential for these materials to induce lung fibrosis. The profibrogenic potential of CNTs is well established. Although CB has not been shown to induce fibrosis, it induces stronger inflammatory, oxidative stress and DNA damage responses than nano-TiO<sub>2</sub> particles.

**Conclusion:** The results of the analysis correctly identified all NMs to be inflammogenic and only CB and CNTs as potentially fibrogenic. In addition to identifying several previously defined, functionally relevant gene sets, the present study also identified

two novel genes sets: a gene set associated with pulmonary fibrosis and a gene set associated with ROS, underlining the advantage of using a data-driven approach to identify novel, functionally related gene sets. The results can be used in future gene set enrichment analysis studies involving NMs or as features for clustering and classifying NMs of diverse properties.

## Introduction

Metadata analysis that leverages genomics data has become increasingly popular as more experiments populate publicly available data repositories such as the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) and European Bioinformatics Institute (EBI; <https://www.ebi.ac.uk/arrayexpress/>). A systems biology approach through meta-analysis has the potential to reveal relationships and insight on resulting phenotypes that may not be possible to detect through the analysis of any individual experiment [1-12].

Conventional molecular approaches for the study of organismal response to toxicant exposures or diseases involve the study of one gene or a few genes at a time, whereas biological response is driven by a group of genes. Thus, when normal function of a specific biological process is perturbed, alterations and enrichment in the expression of a subset of co-functioning genes associated with that biological process are observed. Toxicogenomic tools such as gene expression profiling have become a widely used strategy for investigating the genome-wide changes relating to molecular mechanisms underlying many complex responses and diseases. The fact that genes interact with each other and are expressed in functionally relevant patterns implies that gene-expression data can be grouped into functionally meaningful gene sets across a subset of conditions [13-32]. The analysis of such predefined gene sets is a powerful alternative to individual gene analysis [13]. However, derivation of meaningful and relevant gene sets from the thousands of genes showing expression changes following exposure to toxicants is challenging.

Gene set data analysis, a computational technique which determines if a predefined set of genes exhibit statistically significant differential expression between two or more experimental conditions (time, dose, tissue, etc.), relies on the knowledge of annotated pathways relevant to the underlying physiology or biology being investigated. A survey conducted by Huang et al. [33] identified 68 different gene set enrichment tools. These methods are applied to manually and computationally curated [29] gene sets to identify enriched functional groupings of genes. These gene set enrichment tools include DAVID [21,22], EASE [34], GoMiner [35], MAPPFinder [36], Onto-express [37] and others, which consist of controlled descriptions of gene functions that are frequently used to define gene sets. Other tools, such as pathway databases including Gene Ontology [38], KEGG [39], BioCyc [40], TfactS [41], CTD [42], and BioCarta

(<http://www.biocarta.com>), have also been applied in gene set analysis. Despite the number of tools available, the effective identification of functional groups of genes relevant to the underlying physiology across several conditions still remains a challenge. As a result, these tools continue to be refined and improved.

Nanomaterials (NMs) are materials manufactured on the nanoscale (1–100 nm) and are the building blocks of nanotechnology. On the nanoscale, materials exhibit unique size-associated properties (optical, magnetic, mechanical, thermodynamic, electrical, etc.), which are harnessed for use in various commercial applications [43]. Current applications of NMs include therapeutic applications (e.g., nanomedicine, drug delivery, diagnostics), agriculture, manufacturing, electronics, cosmetics, textiles, and environmental remediation and protection. Although NMs are synthesized from their corresponding, known, bulk chemical substances, owing to their distinct size-associated properties, their biological or toxicological behavior are often different from their analogous bulk compound. Because of their smaller size and large surface area, NMs are known to have increased ability to interact with cellular membranes, they can easily cross cellular barriers and penetrate deeper regions of tissue (such as the highly vascularized alveolar regions of lungs), and they exhibit increased toxicological activity as compared to the corresponding bulk material or comparatively large particles [43]. A variety of conventional toxicology tools have been assessed using both in vitro and in vivo models for their suitability and applicability for toxicity testing of NMs. However, these tools are single-endpoint-based or targeted in nature, investigate only one type of response at a time, and lack detailed mechanistic information [44]. Given the rate at which nanotechnology is growing, and the limitations of the currently available toxicological testing tools, it is estimated that it will take several decades and millions of dollars to complete the assessment of NMs of various sizes, shapes and surface coatings that require immediate assessment [45]. Therefore, more efficient toxicity testing and prediction tools are needed to provide a comprehensive overview of the biological activities of NMs to rapidly screen the toxicological potential of NMs.

Over the last few years, genome-wide expression analysis tools have been used as an alternative approach to comprehensively investigate the toxicological response induced by various

classes of NMs and to identify the properties of NMs that are responsible for eliciting adverse effects. We have previously used transcriptomics profiling tools to investigate the underlying mechanisms of toxicity induced by nanoparticles of titanium dioxide (nano-TiO<sub>2</sub>) [46–48] and carbon nanotubes (CNTs) [49,50] of various sizes and properties. This work identified the properties of nano-TiO<sub>2</sub> that influence their inflammatory potential [51]. These studies have generated a large repository of gene expression data that reflect the diversity of NM-induced biological response across a variety of experimental conditions. However, the challenge lies in the effective use of these data to discern individual or networks of genes

conferring adverse outcomes of regulatory importance or disease phenotypes.

In the present study, we used a meta-analysis approach like that described by Turcan et al. [20] to identify functionally related biclusters of genes showing similar expression profiles, derived from publicly available gene expression data sets describing specific lung diseases (Table 1). One advantage of biclustering is that genes in the same cluster do not have to behave similarly over all experimental conditions. Unlike classical clustering techniques, biclusters can overlap with each other. This is ideal for mining functionally related gene sets as genes can be asso-

**Table 1:** Publically available datasets.

	GEO accession; reference	Platform	Disease model/nanomaterial
Lung disease models	GSE4231 [57]	UCSF 10Mm Mouse v.2 Oligo Array (GPL1089); UCSF GS Operon Mouse v.2 Oligo Array (GPL3330); UCSF 11Mm Mouse v.2 Oligo Array (GPL3331); UCSF 7Mm Mouse v.2 Oligo Array (GPL3359)	Lung inflammation models
	GSE6116 [58]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Biomarkers to predict female mouse lung tumors
	GSE6858 [59]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Model of experimental asthma
	GSE8790 [60]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Cigarette smoke-induced emphysema
	GSE11037 [11]	Agilent-011978 Mouse Microarray G4121A (GPL891)	Emphysema
	GSE18534 [61]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Mouse small cell lung cancer model
	GSE19605 [62]	Illumina MouseRef-8 v2.0 expression beadchip (GPL6885)	Lung carcinogenesis
	GSE25640 [63]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Pulmonary fibrosis
	GSE31013 [64]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Spontaneous lung tumors
	GSE40151 [65]	Affymetrix Mouse Genome 430 2.0 Array (GPL1261)	Idiopathic pulmonary fibrosis
	GSE42233 [66]	Illumina Mouse WG-6 v2.0 expression beadchip (GPL6887)	Lung cancer
	GSE52509 [67]	Illumina MouseRef-8 v2.0 expression beadchip (GPL6885)	COPD
NM studies	GSE29042 [68]	GPL4134 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	CNT: MWCNT-7
	GSE35193 [48]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	CB: Printex 90
	GSE41041 [47]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	TiO <sub>2</sub> : UV-Titan L181
	GSE47000 [49]	GPL10787 Agilent-028005 SurePrint G3 Mouse GE 8x60K Microarray	CNT: Mitsui7
	GSE60801 [51]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	TiO <sub>2</sub> : NRCWE-025, NRCWE-030
	GSE60801 [51]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	TiO <sub>2</sub> Sanding dust: Indoor-R, Indoornano TiO <sub>2</sub>
	GSE60801 [51]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	TiO <sub>2</sub> : Sanding dust NRCWE-032, sanding dust NRCWE-033
	GSE60801 [51]	GPL7202 Agilent-014868 Whole Mouse Genome Microarray 4x44K G4122F	TiO <sub>2</sub> : NRCWE 001 (no charge), NRCWE 002 (positively charged)
	GSE61366 [50]	GPL10787 Agilent-028005 SurePrint G3 Mouse GE 8x60K Microarray	CNT: NRCWE-26, NM-401

ciated with more than one biological process. Several studies [3,52–55] have shown that biclustering is a useful methodology to uncover processes that are active only over some but not all experimental conditions [56].

In this study, experiments investigating lung diseases (including lung inflammation, emphysema, chronic obstructive pulmonary disease (COPD) or lung cancer) in mice using the whole genome gene expression tools were obtained from GEO. For each study, raw data were downloaded from GEO and normalized as described in the methods below. Biological replicates for each of the experimental conditions were averaged. All studies were merged together and biclustering was employed. Through this analysis, ten biclusters representing ten functional gene sets were identified. Using DAVID [21,22], the biological functions associated with these biclusters were identified. Next, we applied these candidate gene sets/biclusters to nine, publicly available, toxicogenomic gene expression studies (Table 1, published studies from our laboratory) to examine the toxicity induced by a variety of NMs (nano-TiO<sub>2</sub>, CB and CNTs) to determine the disease significance of the altered gene expression profiles following exposure to NMs. The analysis was restricted to lung disease models since pulmonary response following NM exposure is well characterized.

## Results and Discussion

### Identification of biclusters of genes from lung disease models

To develop a data-driven view of the mouse lung response following exposure to NMs, publicly available genomic data from GEO that describe characteristic features of select lung diseases were leveraged. Eleven studies encompassing 52 experimental conditions with 8752 common gene symbols were assembled and specific gene sets were extracted using the repeated Bimax [69] biclustering method. A total of ten distinct biclusters were identified. The results of the biclustering are visually summarized in Figure 1.

Bicluster-1 consisted of studies investigating small cell lung carcinoma, spontaneous lung tumor, asthma and pulmonary fibrosis. This bicluster consisted of 19 gene symbols (C1qa, C3ar1, Cd68, Clec4n, Ctsk, Ect2, Fcgr3, Gp2, Igf1, Mmp12, Ms4a6d, Ms4a7, Pbk, Prc1, Saa3, Shcbp1, Spp1, Timp1 and Ube2c). Submitting these gene symbols into the DAVID functional annotation analysis tool (<http://david.abcc.ncifcrf.gov>) resulted in no significant gene ontology (GO). The top three ranked GO terms based on unadjusted p-values were acute inflammatory response (p-value = 0.0023), extracellular region (p-value = 0.0067) and extracellular region part (p-value = 0.0083). The lung disease models that comprised this bicluster were the model for human small cell lung carcinoma

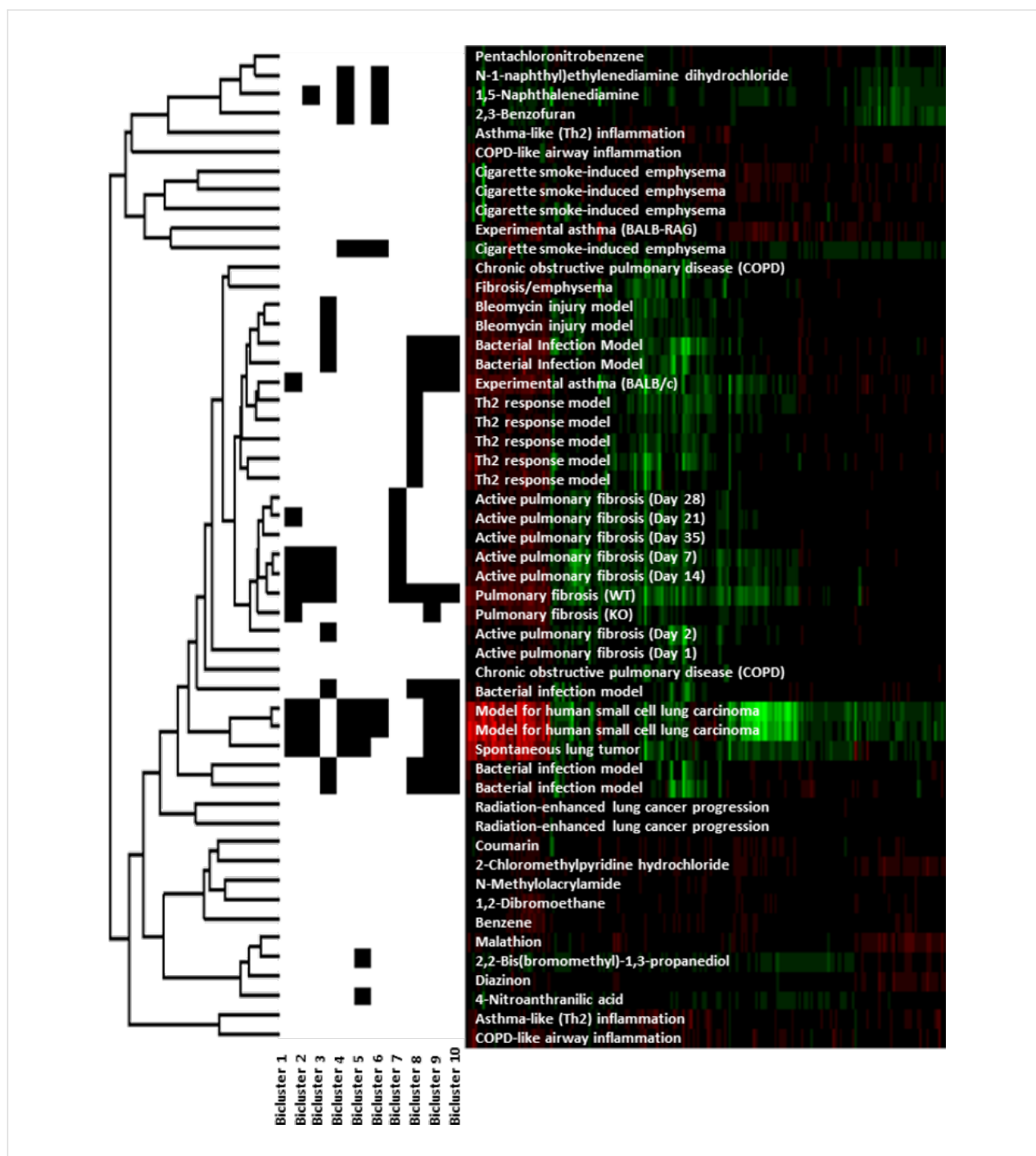
(GSE18534), spontaneous lung tumor (GSE31013), experimental asthma (GSE6858), active pulmonary fibrosis days 7, 14, and 21 (GSE40151) and pulmonary fibrosis (GSE25640).

The second bicluster consisted of twenty gene symbols (4632434I11Rik, Ccna2, Ccnb1, Ccnb2, Cdc20, Cdca8, Cldn4, Hells, Kif22, Mad2l1, Megf10, Melk, Msr1, Mx1, Plk4, Psat1, Rad51, Rrm2, Sprr1a and Uhrf1) with lung disease models such as a model for human small cell lung carcinoma, spontaneous lung tumor, chemical-induced lung carcinogenesis model from GSE6116 (1,5-naphthalenediamine; NAPD) and pulmonary fibrosis. Using DAVID, many GOs and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were found significant (FDR p-value < 0.05). Ten of the twenty gene symbols from this bicluster were elements of the cell cycle GO (FDR p-value =  $4.9 \times 10^{-6}$ ) and five were part of the KEGG pathway (FDR p-value =  $2.5 \times 10^{-4}$ ).

The bleomycin injury and the bacterial infection models (GSE4231), as well as lung disease models related to pulmonary fibrosis, constituted the third bicluster. This bicluster contained 17 gene symbols (Aif1, Ccl2, Ccl9, Ccr5, Cdkn1a, Chl1, Cxcl9, Cyp7b1, Ereg, Fcgr1, Mt2, Retnla, Sfn, Sfrp1, Slc26a4, Socs3, and Tnc). Nine of the seventeen gene symbols are part of the extracellular region GO (FDR p-value = 0.0056). Other significant GO terms included chemokine receptor binding (FDR p-value = 0.017), extracellular region part (FDR p-value = 0.021) and chemokine activity (FDR p-value = 0.032).

The fourth bicluster contained gene symbols associated with chromatin binding (Arid4b, Atrx, Cnot6, Ezh2, Glmn, Hif1a, Ncl, Npm1, Ofd1, Sdccag1, Ssb, Tfrc, Tpp2, Ttc3, Zfp386) (FDR p-value = 0.019). This bicluster contained lung disease models associated with chemical exposure to known lung carcinogens (NAPD, *N*-1-naphthyl)ethylenediamine dihydrochloride (NEDD), 2,3-benzofuran (BFUR)) (GSE6116), a model for human small cell lung carcinoma, spontaneous lung tumor and cigarette smoke-induced emphysema (GSE8790). Many of the gene symbols found in this bicluster are transcription factors involved in the gene expression regulation and are associated with one form of cancer or another.

The fifth bicluster consisted of 35 gene symbols (1700019G17Rik, Ap1m2, Arg1, Atic, Cdc6, Ckmt1, Cldn7, Ddit4, Fetub, Galnt2, Gatm, Grb7, H1f0, Hdac11, Ildr1, Mapk13, Mcm2, Mcm5, Mcm6, Mrps15, Nup50, Pgls, Plek2, Psm8, Rbp4, Rfc4, Rgl3, Rrs1, Serpine1, Sh3yl1, Slc25a13, Slc39a11, Spata5, Tk1, and Tmprss4). The lung disease models that formed this bicluster included the model for human small



**Figure 1:** Heatmap of the gene symbols obtained from the bicluster data analysis. The distance metric used for the cluster analysis was 1-correlation estimated using Spearman correlation with average linkage.

cell lung carcinoma, spontaneous lung tumor, cigarette smoke-induced emphysema and two chemical exposures, 2,2-bis(bromomethyl)-1,3-propanediol (BBMP; lung carcinogen) and 4-nitroanthranilic acid (NAAC; which resulted in no observed tumors). DNA replication for the GO term (FDR p-value =  $4.1 \times 10^{-3}$ ) and KEGG pathway (FDR p-value =  $4.1 \times 10^{-3}$ ) were significant. The only other

significant GO term was DNA replication initiation (FDR p-value = 0.028). A few genes showed association with matrix degradation, inflammation and energy metabolism.

The sixth bicluster consisted of models for human small cell lung carcinoma, cigarette smoke-induced emphysema and chemical exposures BFUR, NAPD and NEDD. DAVID annota-

tion analysis of the 23 gene symbols (Atm, Baz1b, Belaf1, Ccar1, Dek, Dhx9, Epb4.1l3, F5, Hgf, Kif5b, Mier1, Pgm2l1, Plcb4, Ppil4, Rabep1, Smc1a, Stk3, Syncrip, Tcerg1, Ugcg, Usp9x, Zfml, and Zfp292) showed that this group of genes was primarily involved in the acetylation process (FDR p-value = 0.0056). Many GO terms related to the regulation of apoptosis were present in the results obtained by DAVID analysis. However, these results were not statistically significant after the FDR adjustment.

The seventh bicluster contained lung disease models related to pulmonary fibrosis only. DAVID analysis of the gene symbols included in this bicluster (Ccl3, Cd200r1, Chodl, Clec5a, Col24a1, Cxcl10, Emr1, Fxyd4, Gpnmb, Havcr2, Igj, Il1rn, Mmp10, Slc37a2, Syt12, Tgm1, Tlr8, Trem2, Wfdc12, and Zranb3) showed association with pulmonary fibrosis but no significant gene sets were derived. This bicluster can potentially serve as a candidate gene set for pulmonary fibrosis.

The eighth bicluster consisted of models for bacterial infection, Th2 response (GSE4231), asthma (GSE6858) and pulmonary fibrosis (GSE25640) with sixteen gene symbols (C1qb, Ch25h, Clec4a2, Ctss, F7, Fcgr2b, Itgam, Itgb2, Lgmn, Lpxn, Ly86, S100a4, Serpina3g, Serpina3n, Slc7a2, and Tbxas1). These gene symbols resulted in three significant GOs: response to wounding (FDR p-value = 0.0037), defense response (FDR p-value = 0.0063) and inflammatory response (FDR p-value = 0.0045).

The ninth bicluster consisted of the down-regulated gene symbols (Actc1, Cfd, Ckm, Ckmt2, Cox7a1, Cox8b, Csrp3, Eno3, Fmo3, Myh6, Myl1, Myl7, Pln, Pon1, Smpx, Sult1d1, Tnnc1, and Tnni3) and included a bacterial infection model, a model for human small cell lung carcinoma, spontaneous lung tumor, an asthma model and a pulmonary fibrosis model. These gene symbols were significantly associated with KEGG pathway cardiac muscle contraction (FDR p-value < 0.0001) and GO terms such as myosin complex (FDR p-value = 0.02) and regulation of system process (FDR p-value = 0.0015).

The tenth bicluster resulting from the analysis of the genes that were 2-fold down-regulated consisted of lung inflammation and disease models such as the bacterial infection model, a model for human small cell lung carcinoma, the study on spontaneous lung tumor, an asthma model and pulmonary fibrosis. This bicluster consisted of seventeen gene symbols (Aldh3a1, Bmp6, Cyp1a1, Cyp4b1, Eng, Fmo1, Fmo2, Gpr155, Igfbp6, Mapt, Ndr2, Omd, Pcolce2, Pgam2, Scube2, Slc7a10, and Tnxb). These genes were associated with a variety of functions including fatty acid metabolism; however, DAVID functional annotation analysis of these gene symbols resulted in no

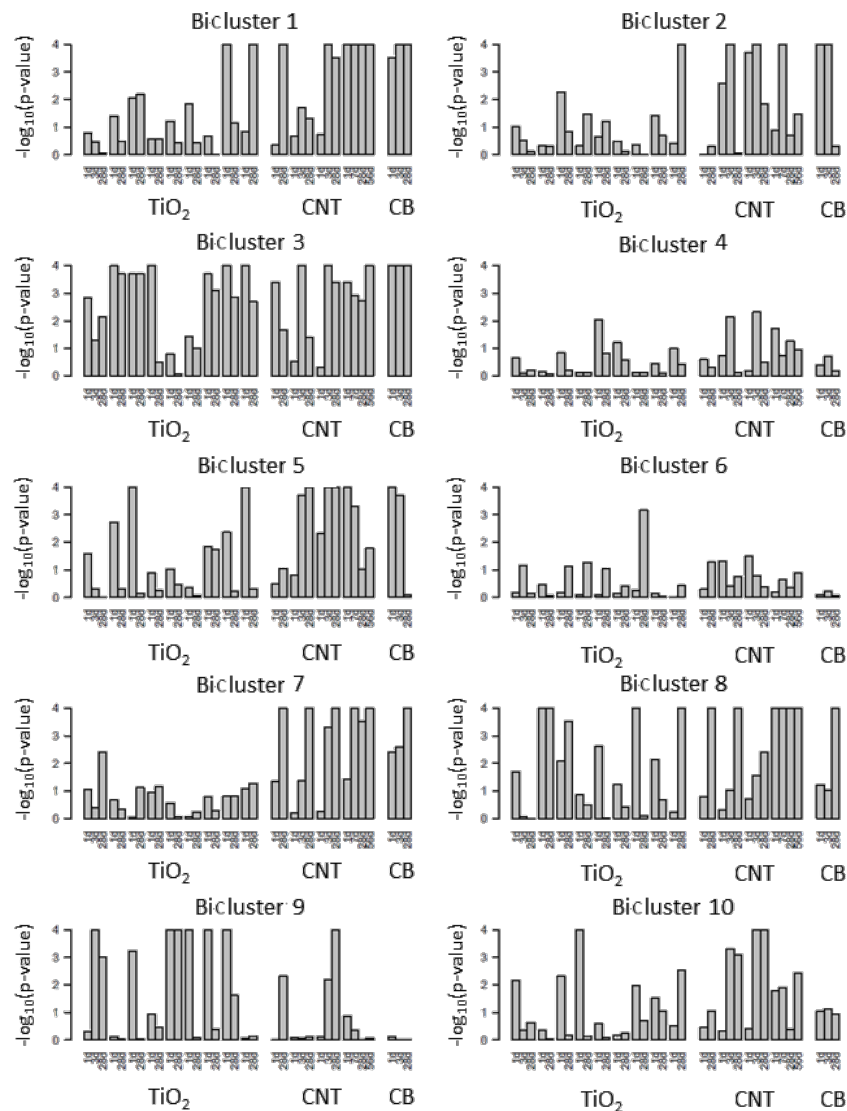
statistically significant results to known annotated gene sets. However, several of these genes are associated with reactive oxygen species (ROS), which may not be a well-established gene set.

## Application of biclusters to classify NM-induced lung response

Next, gene set enrichment analysis (GSEA) [29] using the bicluster-method-derived genes sets was conducted on the nine publically available studies [47-51,68] that examined NM-induced pulmonary toxicity. These results are presented in Figure 2. Bicluster-3 (genes associated with chemokine activity reflecting pulmonary inflammation) was enriched for most of the NMs. These results are in alignment with other studies in the literature that have shown pulmonary inflammation to be the predominant response following exposure to a variety of NMs. Bicluster-7 was the other significant cluster that was enriched in most of the experiments related to CNTs and CB. This cluster consisted of gene symbols showing strong association with pulmonary fibrosis. CNTs are well known to induce pulmonary fibrosis [50]. Although exposure to CB was not shown to cause lung fibrosis at the tested doses [48], studies have shown that CB exposure enhances bleomycin-induced lung fibrosis [70]. These results suggest that both carbon-based NMs may perturb similar biological processes and functions and factors in addition to the altered expression of a few genes in the gene set may contribute to the initiation of lung fibrosis.

## Conclusion

In this study, we examined the applicability of a data-driven approach to identify gene sets from the comprehensive gene expression data using a biclustering method. The results showed that the lung response to NM exposure predominantly reflects responses observed following bacterial infections and bleomycin injury models that involve acute inflammation. The combined biclustering and gene set enrichment analysis also identified CNT and CB as potentially fibrogenic NMs. Although several genes sets associated with acute DNA binding, cell cycle, apoptosis, and ROS response that were specific to different disease models were also observed to be perturbed following exposure to NMs, the implication of such perturbation was not clear from this analysis. In addition the identification of several previously defined, functionally relevant gene sets, the present study also identified two novel genes sets: Bicluster-7 (consisting of genes associated with pulmonary fibrosis) and Bicluster-10 (consisting of genes associated with ROS), underlining the advantage of using a data-driven approach to identify novel, functionally related gene sets. The results can be used in future gene set enrichment analysis studies involving NMs or as features for clustering and classifying NMs of diverse properties.



**Figure 2:** Gene set enrichment results of the NM datasets. Barplots of the  $-\log_{10}(\text{p-value})$  from the GSEA are presented for each of the NM studies. The studies are ordered in the barplots as follows:  $\text{TiO}_2$ : UV-Titan L181, NRCWE-025, NRCWE-030, Sanding Dust Indoor-R, Sanding Dust Indoor-nano, Sanding dust NRCWE-032, Sanding dust NRCWE-03, NRCWE 001 (No charge), NRCWE 002 (positively charged); CNT: Mitsui7, NRCWE-26, NM-401, MWCNT-7; CB: Printex 90.

While powerful, data-driven meta-analysis approaches have several limitations. One important limitation is that the analysis is conditional on the subset of studies selected from the public data repositories such as GEO and EBI. Also, the experiments available in these repositories may not be representative of the population. For example, there are other mouse models of lung diseases that were not included in the present study due to lack of publicly available data or failure to meet the criteria set by the present study (time points, mouse strain, microarray platforms used).

The analysis is also limited to the gene symbols that were consistently investigated across the various microarray plat-

forms from the different studies included in the analyses. Furthermore, the bicluster analysis is conditional to the two-fold change cut-off employed to create the binary matrix for the Bimax algorithm and the choice of the Bimax parameters. Modifying the fold cut-off to 1.75- and 1.5-fold, an additional 28 (23 up and 5 down) and 100 (89 up and 11 down) biclusters were identified. However, the interpretations were derived from the 2-fold cut-off as it provides the most conservative approach. The biclusters were stable when varying the minimum number of rows and when varying the minimum number of columns. Here, additional clusters were identified when these parameters were reduced and clusters were eliminated when these parameters were increased. Changes to any of the above

could impact the final results and therefore the interpretation of the data.

## Experimental Lung disease models

The data were obtained from the GEO. The accession numbers for the studies [11,57–67] used in the exploration of novel gene sets are presented in Table 1. These data sets cover a variety of lung diseases and lung injury outcomes, including different lung inflammation models, emphysema, chronic obstructive pulmonary disease and experiments studying lung cancer and lung tumors. Several different microarray platforms including the Illumina expression beadchip were used in these studies. The analysis was restricted to lung disease models since pulmonary responses following exposure to NMs are well characterized.

## Data processing and normalization

The  $\log_2$  transformation was applied to all signal intensity measurements. For the two color microarray studies, the LOWESS normalization method [71] using the R statistical software environment [72] was applied. For studies using the Affymetrix GeneChips<sup>®</sup>, the RMA normalization was applied using the justRMA function in the affy [73] R package. Quantile normalization was applied for studies that utilized the Illumina beadchip. This was done using the lumiN function in the lumi [74] R package.

Probes with technical replicates were then averaged using the median. The data for each study was then merged to its appropriate annotation file to obtain the gene symbol. Probes with the same gene symbol were then averaged using the median. The experimental conditions with biological replicates were averaged using the median. The median was used as it is a robust estimate of the central tendency.

For each experimental condition, the data was further normalized by centering to the matched control. The control samples were then removed from the data set. The remaining data is presented relative to the control, equivalently the  $\log_2$  of the fold change (estimated using medians) for all the studies. The data were then merged across studies using the gene symbol. The mining the  $\log_2$  of the fold changes was done in an attempt to minimize the cross-platform differences. However, platform differences may exist through compression of the fold-change values [75].

## Biclustering

The biclustering data analysis was conducted in R using the biclust [69] package. The repeated Bimax [56] method was selected for this analysis. Bimax uses a simple data model that

assumes two possible states for each expression level, no change and change with respect to a control experiment. For this analysis, two binary matrices were constructed: one matrix, consisting of zeros and ones, where the ones indicated genes that were 2-fold up-regulated and a second matrix, where the ones identify genes that were 2-fold down-regulated.

The option for the minimum number of rows for the Bimax method was set at 15. The minimum number of columns (which represent the experimental conditions) was set as 5 and the maximum number of columns was set as 15. This resulted in 8 biclusters from the binary matrix representing the up-regulated genes and 2 biclusters were identified for the matrix representing the down-regulated genes.

## NM-induced lung response data sets

The data sets examining differential gene expression in mouse lung exposed to CB, nano-TiO<sub>2</sub> or CNTs were compiled from GEO. Since this is a proof-of-concept study, the investigation was limited to those NMs for which lung toxicological response is well characterized. Also, the genomics datasets with multiple doses and post-exposure time points were considered in the analysis. The GEO accession numbers for these studies are presented in Table 1. These studies utilized the two color Agilent microarray reference design [76]. The data were LOWESS normalized and probes with technical replicates were averaged. The annotation file containing the gene symbol was merged with the expression data and probes with multiple gene symbols were averaged using the median expression.

## Gene set enrichment

As the NM-induced lung response data sets contained multiple doses, the test statistic from the Attract [19] approach was used. Using this method, the overall F-statistic for the dose effect was estimated for each gene. The F-statistics were then  $\log_2$ -transformed. A two sample t-test (assuming unequal variances) was then conducted, comparing the mean of the  $\log_2$  F-statistics within the bicluster to the mean of the  $\log_2$  F-statistics for all genes. The observed t-statistics and p-values are reported in Figure 2.

## References

- Hughes, T. R.; Marton, M. J.; Jones, A. R.; Roberts, C. J.; Stoughton, R.; Armour, C. D.; Bennett, H. A.; Coffey, E.; Dai, H.; He, Y. D.; Kidd, M. J.; King, A. M.; Meyer, M. R.; Slade, D.; Lum, P. Y.; Stepaniants, S. B.; Shoemaker, D. D.; Gachotte, D.; Chakraburty, K.; Simon, J.; Bard, M.; Friend, S. H. *Cell* **2000**, *102*, 109–126. doi:10.1016/S0092-8674(00)00015-5
- Rhodes, D. R.; Barrette, T. R.; Rubin, M. A.; Ghosh, D.; Chinnaiyan, A. M. *Cancer Res.* **2002**, *62*, 4427–4433.
- Ihmels, J.; Friedlander, G.; Bergmann, S.; Sarig, O.; Ziv, Y.; Barkai, N. *Nat. Genet.* **2002**, *31*, 370–377. doi:10.1038/ng941



4. Ramaswamy, S.; Ross, K. N.; Lander, E. S.; Golub, T. R. *Nat. Genet.* **2002**, *33*, 49–54. doi:10.1038/ng1060
5. Xu, L.; Tan, A. C.; Naiman, D. Q.; Geman, D.; Winslow, R. L. *Bioinformatics* **2005**, *21*, 3905–3911. doi:10.1093/bioinformatics/bti647
6. Lamb, J.; Crawford, E. D.; Peck, D.; Modell, J. W.; Blat, I. C.; Wrobel, M. J.; Lerner, J.; Brunet, J. P.; Subramanian, A.; Ross, K. N.; Reich, M.; Hieronymus, H.; Wei, G.; Armstrong, S. A.; Haggarty, S. J.; Clemons, P. A.; Wei, R.; Carr, S. A.; Lander, E. S.; Golub, T. R. *Science* **2006**, *313*, 1929–1935. doi:10.1126/science.1132939
7. Hibbs, M. A.; Hess, D. C.; Myers, C. L.; Huttenhower, C.; Li, K.; Troyanskaya, O. G. *Bioinformatics* **2007**, *23*, 2692–2699. doi:10.1093/bioinformatics/btm403
8. Hassane, D. C.; Guzman, M. L.; Corbett, C.; Li, X.; Abboud, R.; Young, F.; Liesveld, J. L.; Carroll, M.; Jordan, C. T. *Blood* **2008**, *111*, 5654–5662. doi:10.1182/blood-2007-11-126003
9. Dudley, J. T.; Tibshirani, R.; Deshpande, T.; Butte, A. J. *Mol. Syst. Biol.* **2009**, *5*, 307. doi:10.1038/msb.2009.66
10. Daigle, B. J., Jr.; Deng, A.; McLaughlin, T.; Cushman, S. W.; Cam, M. C.; Reaven, G.; Tsao, P. S.; Altman, R. B. *PLoS Comput. Biol.* **2010**, *6*, e1000718. doi:10.1371/journal.pcbi.1000718
11. Thomson, E. M.; Williams, A.; Yauk, C. L.; Vincent, R. *Am. J. Pathol.* **2012**, *180*, 1413–1430. doi:10.1016/j.ajpath.2011.12.020
12. Jackson, A. F.; Williams, A.; Recio, L.; Waters, M. D.; Lambert, I. B.; Yauk, C. L. *Toxicol. Appl. Pharmacol.* **2014**, *274*, 63–77. doi:10.1016/j.taap.2013.10.019
13. Parfett, C.; Williams, A.; Zheng, J. L.; Zhou, G. *Regul. Toxicol. Pharmacol.* **2013**, *67*, 63–74. doi:10.1016/j.yrtph.2013.06.005
14. Barry, W. T.; Nobel, A. B.; Wright, F. A. *Bioinformatics* **2005**, *21*, 1943–1949. doi:10.1093/bioinformatics/bti260
15. Leung, Y. F.; Cavalieri, D. *Trends Genet.* **2003**, *19*, 649–659. doi:10.1016/j.tig.2003.09.015
16. Lu, Y.; Liu, P.-Y.; Xiao, P.; Deng, H.-W. *Bioinformatics* **2005**, *21*, 3105–3113. doi:10.1093/bioinformatics/bti496
17. Qin, H.; Feng, T.; Harding, S. A.; Tsai, C.-J.; Zhang, S. *Bioinformatics* **2008**, *24*, 1583–1589. doi:10.1093/bioinformatics/btn215
18. Tuglus, C.; van der Laan, M. J. *Stat. Appl. Genet. Mol. Biol.* **2009**, *8*, 1–15. doi:10.2202/1544-6115.1397
19. Mar, J. C.; Matigian, N. A.; Quackenbush, J.; Wells, C. A. *PLoS One* **2011**, *6*, e25445. doi:10.1371/journal.pone.0025445
20. Turcan, S.; Vetter, D. E.; Maron, J. L.; Wei, X.; Slonim, D. K. In *Proceedings of the Pacific Symposium, Kohala Coast, Hawaii, Jan 3–7, 2011*; World Scientific Publishing Co Pte Ltd: Singapore, 2011; pp 50–61. doi:10.1142/9789814335058\_0006
21. Huang, D. W.; Sherman, B. T.; Tan, Q.; Collins, J. R.; Alvord, W. G.; Roayaei, J.; Stephens, R.; Baseler, M. W.; Lane, H. C.; Lempicki, R. A. *Genome Biol.* **2007**, *8*, R183. doi:10.1186/gb-2007-8-9-r183
22. Huang, D. W.; Sherman, B. T.; Tan, Q.; Kir, J.; Liu, D.; Bryant, D.; Guo, Y.; Stephens, R.; Baseler, M. W.; Lane, H. C.; Lempicki, R. A. *Nucleic Acids Res.* **2007**, *35*, W169–W175. doi:10.1093/nar/gkm415
23. Engreitz, J. M.; Daigle, B. J., Jr.; Marshall, J. J.; Altman, R. B. *J. Biomed. Inf.* **2010**, *43*, 932–944. doi:10.1016/j.jbi.2010.07.001
24. Huang, Y.; Li, H.; Hu, H.; Yan, X.; Waterman, M. S.; Huang, H.; Zhou, X. J. *Bioinformatics* **2007**, *23*, i222–i229. doi:10.1093/bioinformatics/btm222
25. Eisen, M. B.; Spellman, P. T.; Brown, P. O.; Botstein, D. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14863–14868. doi:10.1073/pnas.95.25.14863
26. Spellman, P. T.; Sherlock, G.; Zhang, M. Q.; Iyer, V. R.; Anders, K.; Eisen, M. B.; Brown, P. O.; Botstein, D.; Futcher, B. *Mol. Biol. Cell* **1998**, *9*, 3273–3297. doi:10.1091/mbc.9.12.3273
27. Segal, E.; Shapira, M.; Regev, A.; Pe'er, D.; Botstein, D.; Koller, D.; Friedman, N. *Nat. Genet.* **2003**, *34*, 166–176. doi:10.1038/ng1165
28. Bar-Joseph, Z.; Gerber, G. K.; Lee, T. I.; Rinaldi, N. J.; Yoo, J. Y.; Robert, F.; Gordon, D. B.; Fraenkel, E.; Jaakkola, T. S.; Young, R. A.; Gifford, D. K. *Nat. Biotechnol.* **2003**, *21*, 1337–1342. doi:10.1038/nbt890
29. Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15545–15550. doi:10.1073/pnas.0506580102
30. Ihmels, J.; Bergmann, S.; Barkai, N. *Bioinformatics* **2004**, *20*, 1993–2003. doi:10.1093/bioinformatics/bth166
31. Wang, X.; Dalkic, E.; Wu, M.; Chan, C. *Curr. Opin. Biotechnol.* **2008**, *19*, 482–491. doi:10.1016/j.copbio.2008.07.011
32. Fehrmann, R. S. N.; de Jonge, H. J. M.; ter Elst, A.; de Vries, A.; Crijns, A. G. P.; Weidenaar, A. C.; Gerbens, F.; de Jong, S.; van der Zee, A. G. J.; de Vries, E. G. E.; Kamps, W. A.; Hofstra, R. M. W.; te Meerman, G. J.; de Bont, E. S. J. M. *PLoS One* **2008**, *3*, e1656. doi:10.1371/journal.pone.0001656
33. Huang, D. W.; Sherman, B. T.; Lempicki, R. A. *Nucleic Acids Res.* **2009**, *37*, 1–13. doi:10.1093/nar/gkn923
34. Hosack, D. A.; Dennis, G.; Sherman, B. T.; Lane, H. C.; Lempicki, R. A. *Genome Biol.* **2003**, *4*, R70. doi:10.1186/gb-2003-4-10-r70
35. Zeeberg, B. R.; Feng, W.; Wang, G.; Wang, M. D.; Fojo, A. T.; Sunshine, M.; Narasimhan, S.; Kane, D. W.; Reinhold, W. C.; Lababidi, S.; Bussey, K. J.; Riss, J.; Barrett, J. C.; Weinstein, J. N. *Genome Biol.* **2003**, *4*, R28. doi:10.1186/gb-2003-4-4-r28
36. Doniger, S. W.; Salomonis, N.; Dahlquist, K. D.; Vranizan, K.; Lawlor, S. C.; Conklin, B. R. *Genome Biol.* **2003**, *4*, R7. doi:10.1186/gb-2003-4-1-r7
37. Draghici, S.; Khatri, P.; Bhavsar, P.; Shah, A.; Krawetz, S. A.; Tainsky, M. A. *Nucleic Acids Res.* **2003**, *31*, 3775–3781. doi:10.1093/nar/gkg624
38. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. *Nat. Genet.* **2000**, *25*, 25–29. doi:10.1038/75556
39. Kanehisa, M.; Goto, S. *Nucleic Acids Res.* **2000**, *28*, 27–30. doi:10.1093/nar/28.1.27
40. Karp, P. D.; Ouzounis, C. A.; Moore-Kochlacs, C.; Goldovsky, L.; Kaipa, P.; Ahrén, D.; Tsoka, S.; Darzentas, N.; Kunin, V.; López-Bigas, N. *Nucleic Acids Res.* **2005**, *33*, 6083–6089. doi:10.1093/nar/gki892
41. Essaghir, A.; Toffalini, F.; Knoop, L.; Kallin, A.; van Helden, J.; Demoulin, J.-B. *Nucleic Acids Res.* **2010**, *38*, e120. doi:10.1093/nar/gkq149
42. Gohlke, J. M.; Thomas, R.; Zhang, Y.; Rosenstein, M. C.; Davis, A. P.; Murphy, C.; Becker, K. G.; Mattingly, C. J.; Portier, C. J. *BMC Syst. Biol.* **2009**, *3*, 46. doi:10.1186/1752-0509-3-46
43. Maynard, A. D.; Warheit, D. B.; Philibert, M. A. *J. Toxicol. Sci.* **2011**, *120*, S109–S129. doi:10.1093/toxsci/kfq372
44. Geraci, C. L.; Castranova, V. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **2010**, *2*, 569–577. doi:10.1002/wnan.108
45. Lai, D. Y. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **2012**, *4*, 1–15. doi:10.1002/wnan.162

46. Halappanavar, S.; Jackson, P.; Williams, A.; Jensen, K. A.; Hougaard, K. S.; Vogel, U.; Yauk, C. L.; Wallin, H. *Environ. Mol. Mutagen.* **2011**, *52*, 425–439. doi:10.1002/em.20639
47. Husain, M.; Saber, A. T.; Guo, C.; Jacobsen, N. R.; Jensen, K. A.; Yauk, C. L.; Williams, A.; Vogel, U.; Wallin, H.; Halappanavar, S. *Toxicol. Appl. Pharmacol.* **2013**, *269*, 250–262. doi:10.1016/j.taap.2013.03.018
48. Bourdon, J. A.; Halappanavar, S.; Saber, A. T.; Jacobsen, N. R.; Williams, A.; Wallin, H.; Vogel, U.; Yauk, C. L. *Toxicol. Sci.* **2012**, *127*, 474–484. doi:10.1093/toxsci/kfs119
49. Poulsen, S. S.; Jacobsen, N. R.; Labib, S.; Wu, D.; Husain, M.; Williams, A.; Bøgelund, J. P.; Andersen, O.; Købler, C.; Mølhave, K.; Kyjovska, Z. O.; Saber, A. T.; Wallin, H.; Yauk, C. L.; Vogel, U.; Halappanavar, S. *PLoS One* **2013**, *8*, e80452. doi:10.1371/journal.pone.0080452
50. Poulsen, S. S.; Saber, A. T.; Williams, A.; Andersen, O.; Købler, C.; Atluri, R.; Pozzebon, M. E.; Mucelli, S. P.; Simion, M.; Rickerby, D.; Mortensen, A.; Jackson, P.; Kyjovska, Z. O.; Mølhave, K.; Jacobsen, N. R.; Jensen, K. A.; Yauk, C. L.; Wallin, H.; Halappanavar, S.; Vogel, U. *Toxicol. Appl. Pharmacol.* **2015**, *284*, 16–32. doi:10.1016/j.taap.2014.12.011
51. Halappanavar, S.; Saber, A. T.; Decan, N.; Jensen, K. A.; Wu, D.; Jacobsen, N. R.; Guo, C.; Rogowski, J.; Koponen, I. K.; Levin, M.; Madsen, A. M.; Atluri, R.; Snitka, V.; Birkedal, R. K.; Rickerby, D.; Williams, A.; Wallin, H.; Yauk, C. L.; Vogel, U. *Environ. Mol. Mutagen.* **2015**, *56*, 245–264. doi:10.1002/em.21936
52. Cheng, Y.; Church, G. Biclustering of expression data. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, San Diego, CA, U.S.A., Aug 19–23, 2000; 2000; pp 93–103.
53. Ben-Dor, A.; Chor, B.; Karp, R.; Yakhini, Z. In *Proceedings of the 6th Annual International Conference on Computational Biology*, Washington, DC, April 18–21, 2002; ACM Press: New York, NY, U.S.A., 2002; pp 49–57. doi:10.1145/565196.565203
54. Tanay, A.; Sharan, R.; Kupiec, M.; Shamir, R. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2981–2986. doi:10.1073/pnas.0308661100
55. Murali, T. M.; Kasif, S. Extracting conserved gene expression motifs from gene expression data. In *Pacific Symposium on Biocomputing*, Lihue, Hawaii, Jan 3–7, 2003; World Scientific Publishing Co Pte Ltd: Singapore, 2003; pp 77–88.
56. Prelić, A.; Bleuler, S.; Zimmermann, P.; Wille, A.; Bühlmann, P.; Gruissem, W.; Hennig, L.; Thiele, L.; Zitzler, E. *Bioinformatics* **2006**, *22*, 1122–1129. doi:10.1093/bioinformatics/btl060
57. Lewis, C. C.; Yang, J. Y. H.; Huang, X.; Banerjee, S. K.; Blackburn, M. R.; Baluk, P.; McDonald, D. M.; Blackwell, T. S.; Nagabhushanam, V.; Peters, W.; Voehringer, D.; Erle, D. J. *Am. J. Respir. Crit. Care Med.* **2008**, *177*, 376–387. doi:10.1164/rccm.200702-333OC
58. Thomas, R. S.; Pluta, L.; Yang, L.; Halsey, T. A. *Toxicol. Sci.* **2007**, *97*, 55–64. doi:10.1093/toxsci/kfm023
59. Lu, X.; Jain, V. V.; Finn, P. W.; Perkins, D. L. *Mol. Syst. Biol.* **2007**, *3*, 98. doi:10.1038/msb4100138
60. Rangasamy, T.; Misra, V.; Zhen, L.; Tankersley, C. G.; Tuder, R. M.; Biswal, S. *Am. J. Physiol.: Lung Cell. Mol. Physiol.* **2009**, *296*, L888–L900. doi:10.1152/ajplung.90369.2008
61. Schaffer, B. E.; Park, K.-S.; Yiu, G.; Conklin, J. F.; Lin, C.; Burkhart, D. L.; Karnezis, A. N.; Sweet-Cordero, E. A.; Sage, J. *Cancer Res.* **2010**, *70*, 3877–3883. doi:10.1158/0008-5472.CAN-09-4228
62. Ochoa, C. E.; Mirabolfathinejad, S. G.; Ruiz, V. A.; Evans, S. E.; Gagea, M.; Evans, C. M.; Dickey, B. F.; Moghaddam, S. J. *Cancer Prev. Res.* **2011**, *4*, 51–64. doi:10.1158/1940-6207.CAPR-10-0180
63. Liu, T.; Baek, H. A.; Yu, H.; Lee, H. J.; Park, B.-H.; Ullenbruch, M.; Liu, J.; Nakashima, T.; Choi, Y. Y.; Wu, G. D.; Chung, M. J.; Phan, S. H. *J. Immunol.* **2011**, *187*, 450–461. doi:10.4049/jimmunol.1000964
64. Pandiri, A. R.; Sils, R. C.; Ziglioli, V.; Ton, T.-V. T.; Hong, H.-H. L.; Lahousse, S. A.; Gerrish, K. E.; Auerbach, S. S.; Shockley, K. R.; Bushel, P. R.; Peddada, S. D.; Hoenerhoff, M. J. *Toxicol. Pathol.* **2012**, *40*, 1141–1159. doi:10.1177/0192623312447543
65. Peng, R.; Sridhar, S.; Tyagi, G.; Phillips, J. E.; Garrido, R.; Harris, P.; Burns, L.; Renteria, L.; Woods, J.; Chen, L.; Allard, J.; Ravindran, P.; Bitter, H.; Liang, Z.; Hogaboam, C. M.; Kitson, C.; Budd, D. C.; Fine, J. S.; Bauer, C. M.; Stevenson, C. S. *PLoS One* **2013**, *8*, No. e59348. doi:10.1371/journal.pone.0059348
66. Delgado, O.; Batten, K. G.; Richardson, J. A.; Xie, X.-J.; Gazdar, A. F.; Kaisani, A. A.; Girard, L.; Behrens, C.; Suraokar, M.; Fasciani, G.; Wright, W. E.; Story, M. D.; Wistuba, I. I.; Minna, J. D.; Shay, J. W. *Clin. Cancer Res.* **2014**, *20*, 1610–1622. doi:10.1158/1078-0432.CCR-13-2589
67. John-Schuster, G.; Hager, K.; Conlon, T. M.; Irmeler, M.; Beckers, J.; Eickelberg, O.; Yildirim, A. Ö. *Am. J. Physiol.: Lung Cell. Mol. Physiol.* **2014**, *307*, L692–L706. doi:10.1152/ajplung.00092.2014
68. Guo, N. L.; Wan, Y.-W.; Denvir, J.; Porter, D. W.; Pacurari, M.; Wolfarth, M. G.; Castranova, V.; Qian, Y. *J. Toxicol. Environ. Health, Part A* **2012**, *75*, 1129–1153. doi:10.1080/15287394.2012.699852
69. Kaiser, S.; Santamaria, R.; Khamiakova, T.; Sill, M.; Theron, R.; Quintales, L.; Leisch, F.; De Troyer, E. biclust: BiCluster Algorithms, R package, Version 1.2.0. 2013; <http://CRAN.R-project.org/package=biclust> (accessed Dec 8, 2015).
70. Kamata, H.; Tasaka, S.; Inoue, K.-i.; Miyamoto, K.; Nakano, Y.; Shinoda, H.; Kimizuka, Y.; Fujiwara, H.; Ishii, M.; Hasegawa, N.; Takamiya, R.; Fujishima, S.; Takano, H.; Ishizaka, A. *Exp. Biol. Med.* **2011**, *236*, 315–324. doi:10.1258/ebm.2011.010180
71. Yang, Y. H.; Dudoit, S.; Luu, P.; Lin, D. M.; Peng, V.; Ngai, J.; Speed, T. P. *Nucleic Acids Res.* **2002**, *30*, e15. doi:10.1093/nar/30.4.e15
72. R Core Team. *R Foundation for Statistical Computing*, Vienna, Austria, 2014.
73. Gautier, L.; Cope, L.; Bolstad, B. M.; Irizarry, R. A. *Bioinformatics* **2004**, *20*, 307–315. doi:10.1093/bioinformatics/btg405
74. Du, P.; Kibbe, W. A.; Lin, S. M. *Bioinformatics* **2008**, *24*, 1547–1548. doi:10.1093/bioinformatics/btn224
75. Wang, Y.; Barbacioru, C.; Hyland, F.; Xiao, W.; Hunkapiller, K. L.; Blake, J.; Chan, F.; Gonzalez, C.; Zhang, L.; Samaha, R. R. *BMC Genomics* **2006**, *7*, 59. doi:10.1186/1471-2164-7-59
76. Kerr, M. K.; Churchill, G. A. *Genet. Res.* **2001**, *77*, 123–128. doi:10.1017/S0016672301005055

## License and Terms

This is an Open Access article under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The license is subject to the *Beilstein Journal of Nanotechnology* terms and conditions: (<http://www.beilstein-journals.org/bjnano>)

The definitive version of this article is the electronic one which can be found at:  
[doi:10.3762/bjnano.6.252](https://doi.org/10.3762/bjnano.6.252)