Supporting Information

Self-Organizing Map Analysis of Toxicity-Related Cell Signaling Pathways for Metal and Metal Oxide Nanoparticles

Robert Rallo[a,b], Bryan France[a], Rong Liu[a], Sumitra Nair[a], Saji George[a,c], Robert Damoiseaux[b], Francesc Giralt[d,b], Andre Nel[a,c], Kenneth Bradley[a] and Yoram Cohen*[e,a]

[a] Center for the Environmental Implications of Nanotechnology. California Nanosystems Institute, University of California, Los Angeles, CA 90095. [b] Departament d'Enginyeria Informatica i Matematiques, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain. [c] Department of Medicine - Div. of NanoMedicine, University of California, Los Angeles, CA 90095. [d] Departament d'Enginyeria Quimica, Universitat Rovira i Virgili, Av. Paisos Catalans 26, 43007 Tarragona, Catalunya, Spain. [e] Chemical and Biomolecular Engineering Department. University of California, Los Angeles, Los Angeles, CA 90095.

Number of Pages (including the cover): 11

Number of Tables: 1

Number of Figures: 3

Generation of the stable luciferase reporter cell lines.

In order to obtain stable luciferase reporter cell lines, RAW 264.7 (ATCC #TIB-71) murine macrophage-like cells were transduced with Cignal Lenti Reporters (SABiosciences Corp., US) in Dulbecco's Modified Eagle Medium (DMEM; MediaTech, Inc., VA, US), supplemented with 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.292 mg/mL glutamine (growth medium), and 8 μg/mL polybrene. After 14 hours, the lentiviral-containing medium was removed and replaced with the growth medium. The RAW 264.7 cells were allowed to expand for approximately one week, at which point puromycin selection (3μg/mL) was applied for at least five days before experimentation. Ten cell lines were stably transduced with each of the ten pathway-specific reporters (Table S1). Each RAW 264.7 cell line contained a unique transcriptional response element (TRE) that drives the expression of the firefly luciferase gene. An additional control cell line was generated in which the TRE is absent (TRE^-), leaving only a basal promoter element upstream of the luciferase gene.

Each individual reporter cell line was seeded at $2x10^4$ cells per well in white 384 well plates (Thermo Fisher Sci., NH, US, Cat#4334). A predetermined volume of a nanoparticle (NP) dispersion (prepared from a stock suspension) was then added to plate wells containing the cells, to attain the desired nanoparticle concentration in each of the plate wells, and incubated at 37°C. The NP stock suspensions were prepared by dispersing the nanoparticles in DMEM following a previously established protocol (1, 2). NPs were transferred from these stock suspensions to achieve a two fold dilutions series ranging from 0.375 - 100 μg/mL in the plate wells. At specific times (i.e., t=3, 6, 12, 24 hours post nanomaterial addition), the nanomaterial-containing media were aspirated and replaced with Dulbecco's Phosphate-Buffered

Saline (DPBS; MediaTech, Inc., NH, US). Luciferase levels produced in response to each NP exposure were determined via Bright-Glo Luficerase Assay (Promega Corp., WI, US) with the relative luminescence values determined using a Victor 3V plate reader (Perkin Elmer Inc., MA, US).

References.

(1) Ji, Z.; Jin, X.; George, S.; Xia, T.; Meng, H.; Wang, X.; Suarez, E.; Zhang, H.; Hoek, E. M. V.; Godwin, H.; Nel, E. A.; Zink, J. Dispersion and Stability of TiO_2 Nanoparticles in Cell Culture Media. *Environ. Sci. Technol.* **2010** online; DOI 10.1021/es100417s.

(2) George, S.; Pokhrel, S.; Xia, T.; Gilbert, B.; Ji, Z.; Schowalter, M.; Rosenauer, A.; Damoiseaux, R.; Bradley, K. A.; Maedler, L.; Nel, E. A. Use of a Rapid Cytotoxicity Screening Approach To Engineer a Safer Zinc Oxide Nanoparticle through Iron Doping. *ACS Nano*, **2010**, 4, 1, 15-29.

Hierarchical clustering and Heat Map Analysis of HTS data. Cluster heat map representation of the complete HTS assay (Figure S2), after data preprocessing and normalization, were also developed for comparison with the clustering obtained using the SOM approach (Figure 3). Hierarchical heat map clustering was developed using the same Euclidean distance similarity metric (1, 2) between signaling pathway response (SPR) vectors as utilized for the SOM unit construction.

In the heat map shown in Figure S2, black and white cells correspond, respectively, to statistically significant up-regulation (i.e., SSMD ≥ 3) and down-regulation (i.e., SSMD ≤-3) of the corresponding cell signaling pathways. The hierarchical clustering followed a bottom-up approach where the hierarchy was developed from the individual signaling-pathway response vectors by progressively merging clusters based on their similarity (i.e., agglomerative clustering). The similarity between two individual SPR vectors or two clusters of SPR vectors *i* and *j* is defined as $sim(i,j) = 1 - d(i,j) / d_{\max}$ where $d(i,j)$ is the Euclidean distance between the two objects and $d_{\max}$corresponds to the maximum distance between SPR vectors in the dataset. A similarity level of unity indicates that of all the individual SPR vectors are at the leaves level of the hierarchical tree. Conversely, a similarity of zero indicates merging of all the SPR vectors into a single cluster.

In order to compare the hierarchical clustering with the SOM clustering (Figure 3b) which yielded five distinct clusters, a similarity level of 0.54 was required in the hierarchical clustering approach to also achieve five clusters as in the SOM analysis. At this similarity level, four clusters labeled A, B, C and E in Figure S2 corresponding to unique SPR vectors (i.e., nanoparticles at specific concentrations inducing dissimilar responses merge at the upper levels in the hierarchical tree) were identified from the hierarchical tree by long stems. Cluster E in Figure S2, which

merges with the remaining eNMs at the highest level of the hierarchical tree, identifies the effects of ZnO nanoparticles at high concentration (i.e., 50 and 100 μg/mL) as the most distinct signaling pathway response vectors within the current HTS dataset. Cluster A (Figure S2), also containing ZnO nanoparticles at a moderate concentration of 12.5 and 1.6 μg/mL, merges with cluster E at the top of the hierarchy. Clusters B and C (Figure S2) included only Pt nanoparticles in the concentration range of 12.5 μg/mL to 100 μg/mL. The above clustering analysis indicated that the exposure to ZnO and Pt nanoparticles at moderate-high concentrations triggers a distinct response in RAW 264.7 cells. Finally, cluster D, the largest cluster in Figure S2, groups the majority of the SPR vectors; thus, the response vectors of the eNMs in cluster D cannot be discriminated to a greater level of detail by this hierarchical clustering. Greater level of differentiation among the SPR vectors can be achieved in the heat map representation through the selection of a greater similarity measure. However, it is important to note that the optimal similarity level that would identify valid (or "real") clusters is not known *a priori* and thus an appropriate optimization approach must be utilized such as the consensus clustering method employed in the present work. Since the proximity of nanoparticles in the heat map display is not necessarily related to its proximity along the hierarchical tree, it is difficult to visually extract relationships between the different SPR vectors. In other words, hierarchical clustering does not preserve the topological relationships present in the original HTS dataset. For example, clusters B and E in Figure S2 (corresponding to the response vectors of ZnO and Pt at high concentrations) are located on opposite sides of the hierarchical tree (Figure S2).

References.

(1) Wilkinson, L.; Friendly, M. The history of the cluster heat map. The American Statistician, 2009, 63 (2), 179-184.

(2) Kohonen, T. Physiological interpretation of the self-organizing map algorithm. *Neural Networks* **1993**, 6, 895-905.

Table S1. Description of the 10 toxicity-related cell-signaling pathways.

TRE	Pathway	Pathway Description
NF-κB	NF-κB	Nuclear factor-kappa B plays a key role in inflammation, immune response, cell proliferation and protection against apoptosis. NF-κB is widely used by eukaryotic cells as a regulator of genes that control cell proliferation and cell survival. NF-κB controls genes involved in inflammation and is found to be chronically active in many inflammatory diseases.
CRE	cAMP/PKA	The cAMP response element (CRE) is a point of convergence for many extracellular and intracellular signals, including cAMP, calcium, G-protein coupled receptors (GPCR) and neurotrophins. cAMP activates the cAMP-dependent protein kinase, protein kinase A (PKA). PKA plays very diverse roles in the cell. It participates in regulation of cell cycle and proliferation, metabolism, transmission of nerve impulses, cytoskeleton remodeling, muscle contraction, cell survival and other cell processes.
HIF1A	Hypoxia	The hypoxia-inducible factor-1 (HIF-1) protein is a key regulator of oxygen homeostasis and plays significant roles in cancer progression as well as in cardiovascular diseases. In addition to its other roles in adaptation to hypoxia, HIF-1 has been shown to play a role in inflammation. High p53 expression destructs HIF-1α protein and thereby eliminates HIF-1 reporter activity.
AP1	MAPK/JNK	Activator protein-1 (AP1) transcription factor is a hetero- or homo-dimeric complex that comprises members of the proto-oncogene Jun protein family (c-Jun, JunB and JunD) and Fos protein family (c-Fos, Fos B, Fra-1 and Fra-2). The stress-activated protein kinase/Jun N-terminal kinase (SAPK/JNK) signal transduction pathway is responsible for the phosphorylation and activation of Jun, which in turn activates AP1.
NFAT	PKC/Ca^{++}	The NFAT family of transcription factors plays a pivotal role in the transcriptional regulation of cytokine genes and other genes critical for the immune response. Several pathways are associated with activation of the NFAT enhancer element, including calcineurin and protein kinase C. There is considerable evidence that a number of toxic environmental chemicals target these Ca2+ signaling processes, alter them, and induce cell death by apoptosis.
SMAD	TGF-β	The transforming growth factor β (TGFβ) signaling pathway is involved in many cellular processes, including cell cycle arrest, differentiation, homeostasis, and immunosuppression. TGFβ signaling induces phosphorylation and activation of the SMAD2 and

		SMAD3 proteins, which then form complexes with the mediator SMAD4. These SMAD complexes then translocate to the nucleus, where they activate the expression of TGFβ-responsive genes.
Elk-1/SRF	MAPK/ERK	The ternary complex factors, TCR and Elk-1, form a complex with the SRF over the serum response element (SRE), and activate gene expression. The Elk-1 protein is phosphorylated by mitogen-activated protein kinase (MAPK), causing increased DNA binding, ternary complex formation, and transcriptional activation of target genes.
E2F	Cell Cycle	The E2F family of transcription factors is a key regulator of cell-cycle checkpoints in mammalian cells. The E2F protein is a major target of the retinoblastoma gene product (Rb) and the activity of E2F/pRb is intimately connected with the G1-S transition of the cell cycle.
Myc/Max	c-myc	The c-Myc protein is a transcription factor that heterodimerizes with an obligatory partner, Max, and regulates the transcription of genes important for cell proliferation, differentiation, and apoptosis.
p53	DNA Damage	p53 signaling plays an important role in DNA repair, cell cycle arrest, and apoptosis.

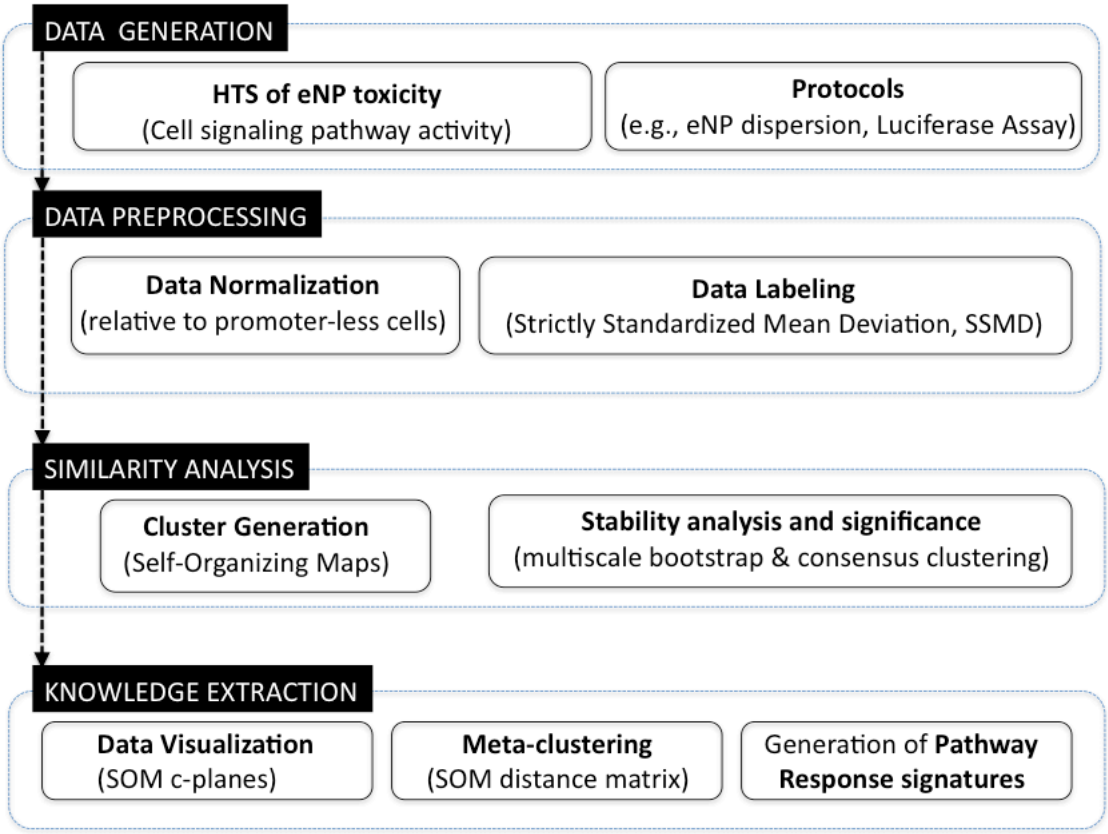

Figure S1. Workflow for data mining and knowledge extraction from HTS of eNMs toxicity-pathway data sets.

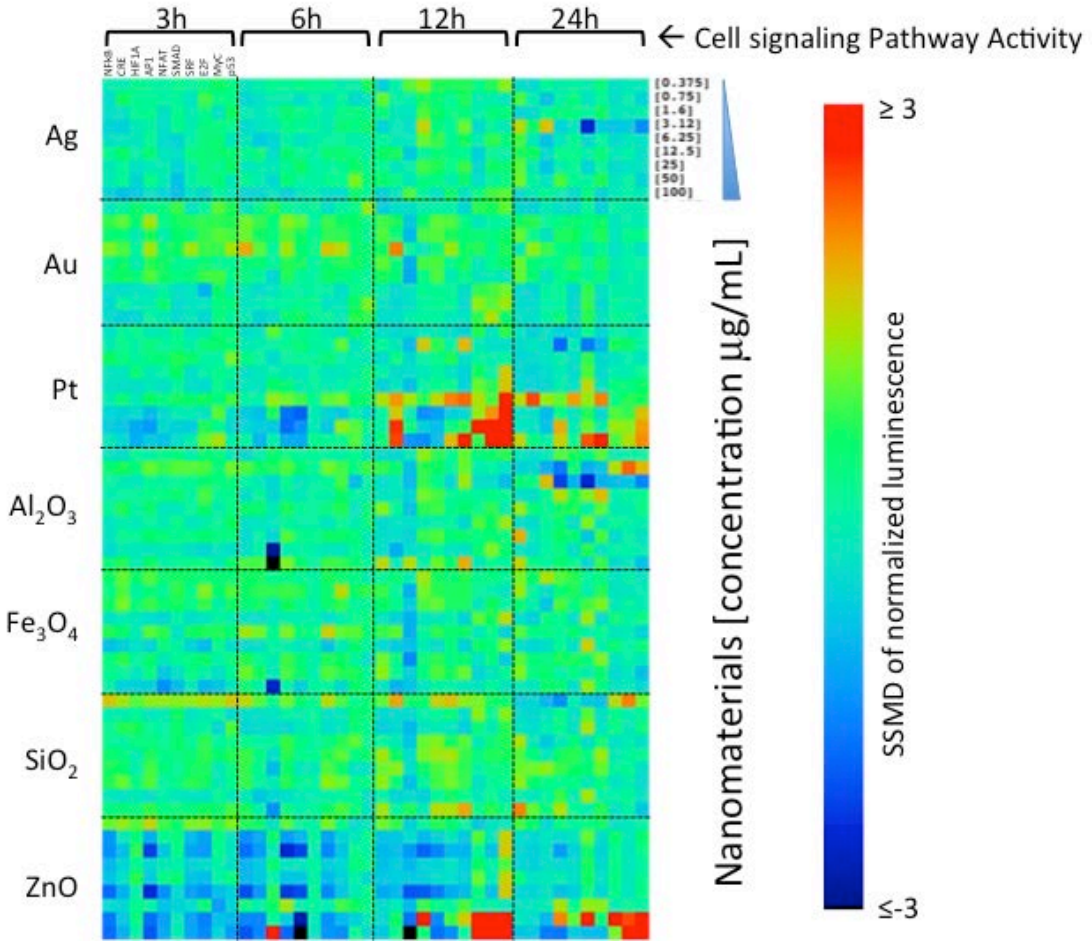

Figure S2. Heat Map of the HTS cell signaling pathway data set after normalization and preprocessing. Dose-response behaviors in specific signaling pathways (e.g., p53, MyC, E2F) can be clearly identified for diverse nanoparticles such as Pt and ZnO after 6 h of exposure time.

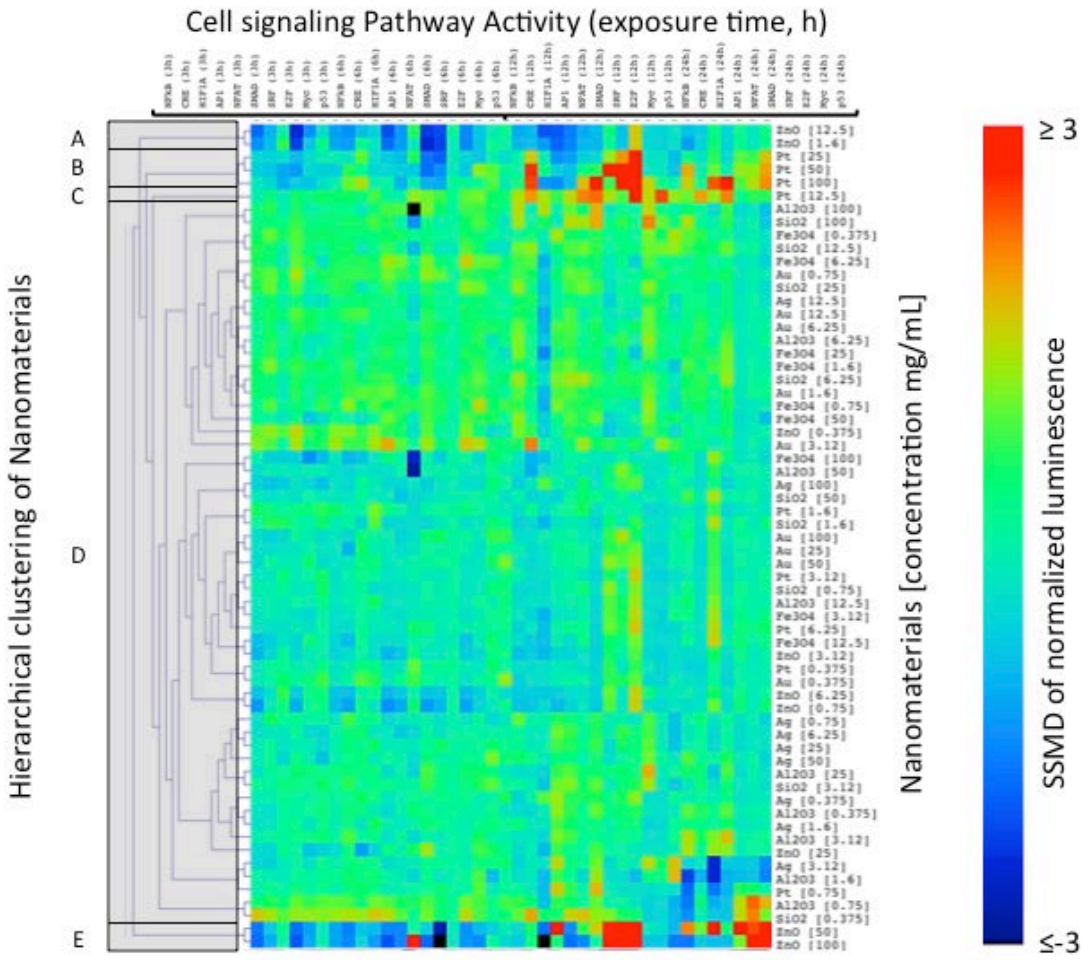

Figure S3. Cluster heat map of the HTS cell signaling pathway data set. The set of five clusters (labeled A to E) were identified from the hierarchical nanoparticle clustering using a similarity threshold of 0.54.

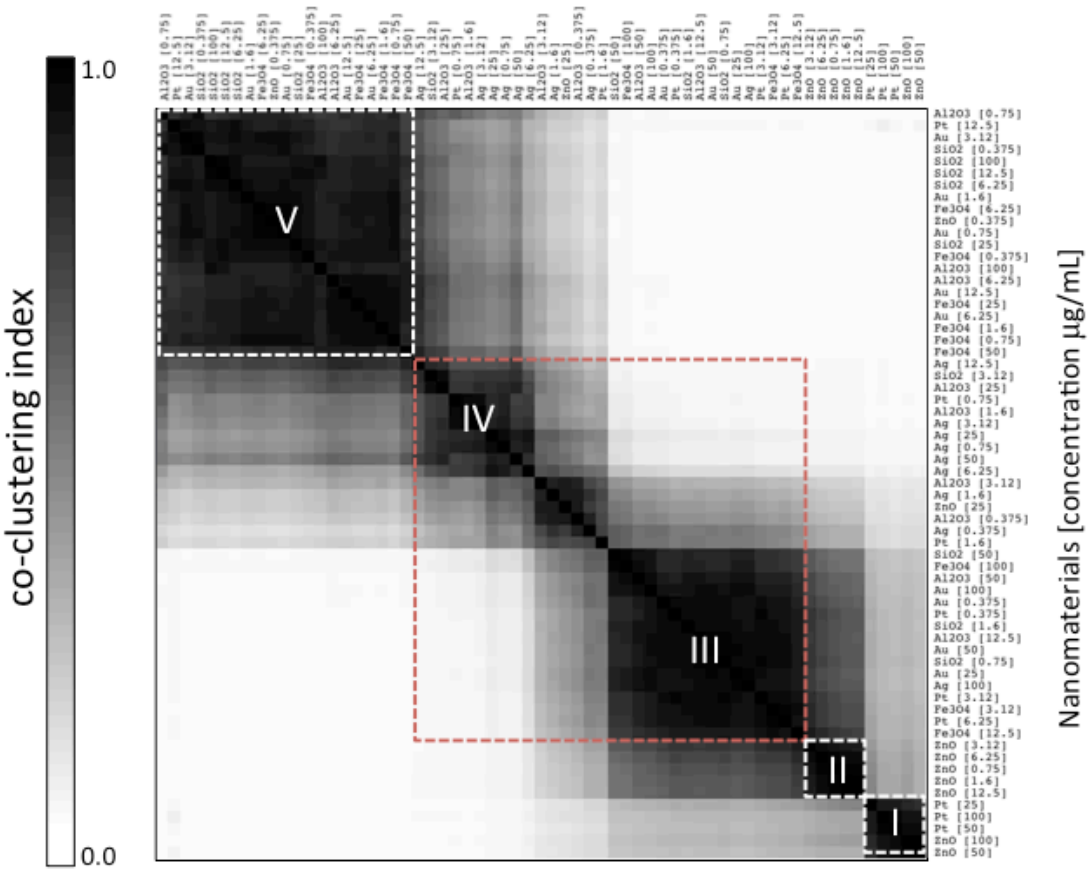

Figure S4. Visualization of the symmetric consensus matrix obtained from the consensus clustering analysis of SOM. The dashed boxes identify the five main structures identified from the clustering of SOM units. The red box shows the region that corresponds to the two SOM clusters (III and IV) with lowest consensus clustering index.