An ancestral molecular response to nanomaterial particulates

del Giudice, G.; Serra, A.; Saarimäki, L. A.; Kotsis, K.; Rouse, I.; Colibaba, S. A.; Jagiello, K.; Mikolajczyk, A.; Fratello, M.; Papadiamantis, A. G.; Sanabria, N.; Annala, M. E.; Morikka, J.; Kinaret, P.A.S.; Voyiatzis, E.; Melagraki, G.; Afantitis, A.; Tämm, K.; Puzyn, T.; Gulumian, M.

DOI:
10.1038/s41565-023-01393-4

nature nanotechnology

Analysis

https://doi.org/10.1038/s41565-023-01393-4

An ancestral molecular response to nanomaterial particulates

Received: 11 April 2022

Accepted: 31 March 2023

Published online: 08 May 2023

Check for updates

G. del Giudice[1], A. Serra[1,2], L. A. Saarimäki[1], K. Kotsis[3], I. Rouse[3], S. A. Colibaba[3], K. Jagiello[4,5], A. Mikolajczyk[4,5], M. Fratello[1], A. G. Papadiamantis[6,7], N. Sanabria[8,9], M. E. Annala[1], J. Morikka[1], P. A. S. Kinaret[1,10], E. Voyiatzis[7], G. Melagraki[11], A. Afantitis[7], K. Tämm[12], T. Puzyn[4,5], M. Gulumian[8,13,14], V. Lobaskin[3], I. Lynch[6], A. Federico[1,2] & D. Greco[1,10,15] ✉

The varied transcriptomic response to nanoparticles has hampered the understanding of the mechanism of action. Here, by performing a meta-analysis of a large collection of transcriptomics data from various engineered nanoparticle exposure studies, we identify common patterns of gene regulation that impact the transcriptomic response. Analysis identifies deregulation of immune functions as a prominent response across different exposure studies. Looking at the promoter regions of these genes, a set of binding sites for zinc finger transcription factors C_2H_2, involved in cell stress responses, protein misfolding and chromatin remodelling and immunomodulation, is identified. The model can be used to explain the outcomes of mechanism of action and is observed across a range of species indicating this is a conserved part of the innate immune system.

Nano-toxicogenomics aims at unravelling the potential toxicity of engineered nanomaterials (ENMs). To date, a plethora of transcriptomics data has been generated for this purpose[1].

Finding commonalities among environmental exposures allows grouping of ENM by mechanism of action (MOA), which would streamline their safety assessment[2].

However, because of the high complexity and variability of systemic responses to ENMs, transcriptomic profiles are heterogeneous and typically biased by small-scale datasets (low numbers of ENM, limited doses and time points). This results in a myriad of toxicogenomic signatures, with low similarity to each other (Supplementary Fig. 1).

The heterogeneity of the ENM transcriptomic signatures hampers the possibility to highlight commonalities between in vitro responses and real-life exposure scenarios.

Improving in vitro–in vivo extrapolation requires the definition of models able to transpose the mechanisms of toxicity from shorter observations in vitro (hours or days) to long timescales (weeks or months) in vivo[3].

In contrast to the specificity of the transcriptional response to environmental signals, the regulation of gene expression is usually well conserved across species[4].

The environment modulates the epigenome[5], and several studies suggested that epigenetic signals can stably maintain modulation well

[1]FHAIVE, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [2]Tampere Institute for Advanced Study, Tampere, Finland. [3]School of Physics, University College Dublin, Dublin, Ireland. [4]Group of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland. [5]QSAR Lab Ltd, Gdańsk, Poland. [6]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK. [7]Novamechanics Ltd, Nicosia, Cyprus. [8]National Institute for Occupational Health, National Health Laboratory Services, Johannesburg, South Africa. [9]School of Health Systems and Public Health, University of Pretoria, Pretoria, South Africa. [10]Institute of Biotechnology, Helsinki Institute of Life Sciences (HiLife), University of Helsinki, Helsinki, Finland. [11]Division of Physical Sciences and Applications, Hellenic Military Academy, Vari, Greece. [12]Institute of Chemistry, University of Tartu, Tartu, Estonia. [13]Haematology and Molecular Medicine Department, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa. [14]Water Research Group, Unit for Environmental Sciences and Management, North West University, Potchefstroom, South Africa. [15]Division of Pharmaceutical Biosciences, Faculty of Pharmacy, University of Helsinki, Helsinki, Finland. ✉e-mail: dario.greco@tuni.fi

https://doi.org/10.1038/s41565-023-01393-4

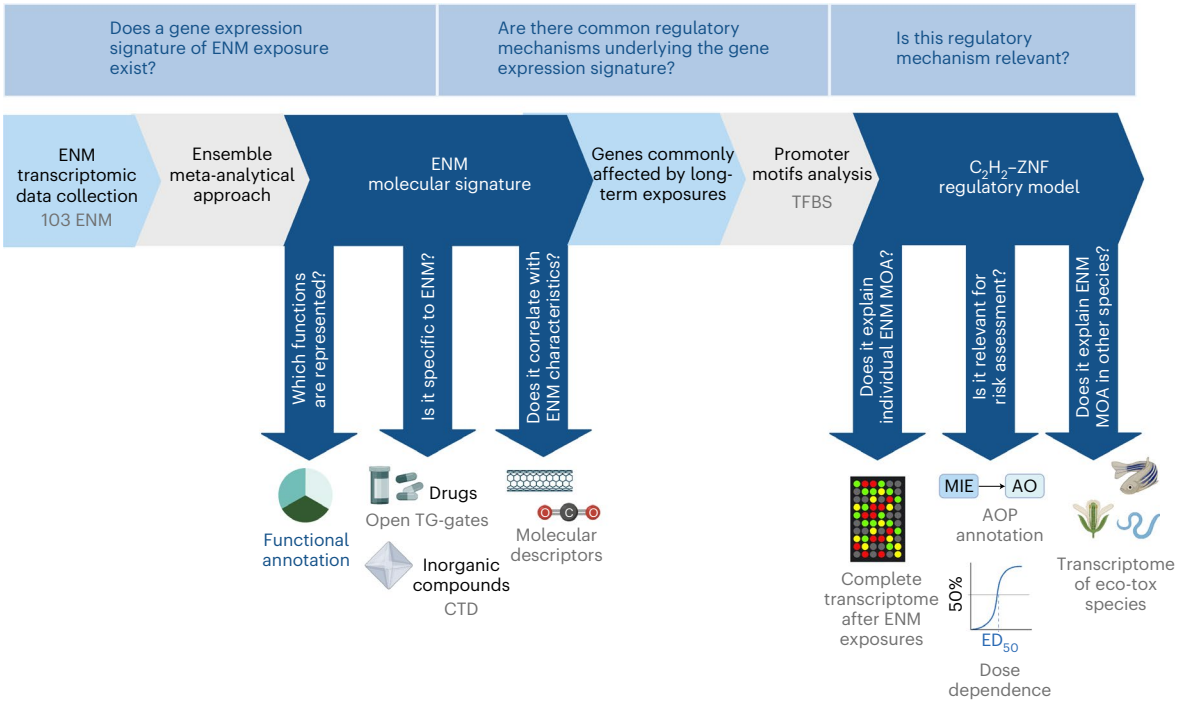


Fig. 1 | Study workflow. Study workflow, including the data used, their sources (grey text) and methodological steps (light grey arrow boxes and blue text). The light blue boxes at the top of the figure indicate the hypothesis underpinning the study. AO, adverse outcome. ED_{50}, mediant effective dose.

beyond the duration of the stimulus (adaptive plasticity)[6]. Transcriptional changes usually follow an 'impulse-like' kinetic, while regulation of transcription is achieved by multiple layers of sustained epigenetic signals[7]. For this, gene regulation is associated with more stable physiopathological changes and can detect exposure-induced alteration more reliably than transcript changes alone.

In this Analysis, we hypothesized that common patterns of gene regulation underlie the response of multiple biological systems exposed to a variety of ENMs (Fig. 1). We analysed the most comprehensive transcriptomics data collection for ENMs to date, in which the expression of 3,676 genes is measured across 584 experimental conditions[8–10]. This collection includes multiple human and mouse cell types and tissues, both in vitro and in vivo, exposed to 103 ENMs varying in chemistry, geometry and size (Supplementary Fig. 2a).

An ENM exposure-specific molecular signature

In toxicogenomics, each significant variation observed between exposed and unexposed samples is attributed to the test compound and is commonly defined as its MOA. In the absence of cell death, the biological systems reach a new homeostasis, which can differ from the pre-exposure state. Hence, we first performed differential expression analysis for each dataset in the ENM collection and computed their pairwise similarity. Our results indicate that transcriptomic signatures are substantially dissimilar from each other (Supplementary Fig. 1), corroborating the hypothesis that transcriptional MOAs do not reveal common patterns of response to exposures[11]. These results are not surprising when considering the diversity of the experimental conditions represented in our collection.

Despite the intrinsic complexity of the individual datasets, we hypothesized that a signature of molecular alteration across ENM exposures exists and can be retrieved through meta-analysis. For this, we utilized an ensemble meta-analytical approach, prioritizing different aspects of the gene expression alteration, to highlight robust patterns of molecular deregulation detected in multiple ENM exposure systems in vivo and in vitro. Initially, we investigated the effect of including the nominal fold change derived from each study. Given the high similarity of the final ENM molecular signature (correlation > 0.75), we performed a canonical meta-analysis based solely on gene expression statistics (Supplementary Fig. 3) to rank 3,676 genes across the datasets (Supplementary Table 1).

The ENM signature has enriched biological pathways belonging to six main categories (cell stress response, innate immunity, cell death, cell cycle perturbation, neurological diseases and adaptive immunity) (Fig. 2a and Supplementary Table 2). Oxidative stress is a prevalent mechanism of ENM toxicity, as their reactive surface can induce hydroxyl radicals, which in turn results in lipid peroxidation, protein interference and DNA damage[12–14]. Cytotoxicity is modulated by ENM intrinsic characteristics[15]. Moreover, many ENMs have been previously associated with neuronal toxicity[2], protein unfolding and fibrillation[16], supporting our results.

Pathways related to both innate and adaptive immunity were strongly enriched (Supplementary Table 2). Interferons are known to mediate early innate and ancestral defence mechanisms, especially upon viral infection[17,18], and have also been reported as being activated by exposure to some ENMs[19–22]. It is noteworthy that, in response to ENMs with diverse intrinsic properties, immune-related pathways are frequently altered across multiple cells and tissues, even those with no primary immune function. Toxicogenomics also concerns molecular response differences across exposures; therefore, we reported the ENM responses with specific physico-chemical characteristics (Supplementary Fig. 4). We next investigated the possible mechanisms of toxicity associated with this gene rank and tested whether ion release can explain the response of different biological systems to ENMs. For this, we progressively removed ion-releasing materials from the initial collection and compared the effect on the ENM signature (Fig. 3a). Our results indicate that ion release cannot alone describe the toxicity mechanism associated with the complete set of ENMs (Fig. 3b–e). Furthermore, we assessed whether removing ion-releasing materials affects the functional profile of the ENMs signature. As expected, the main functional categories initially identified are preserved (Fig. 2a and Supplementary Fig. 5). Altogether, these results may be relevant in future studies focusing on ENM ion release.

https://doi.org/10.1038/s41565-023-01393-4

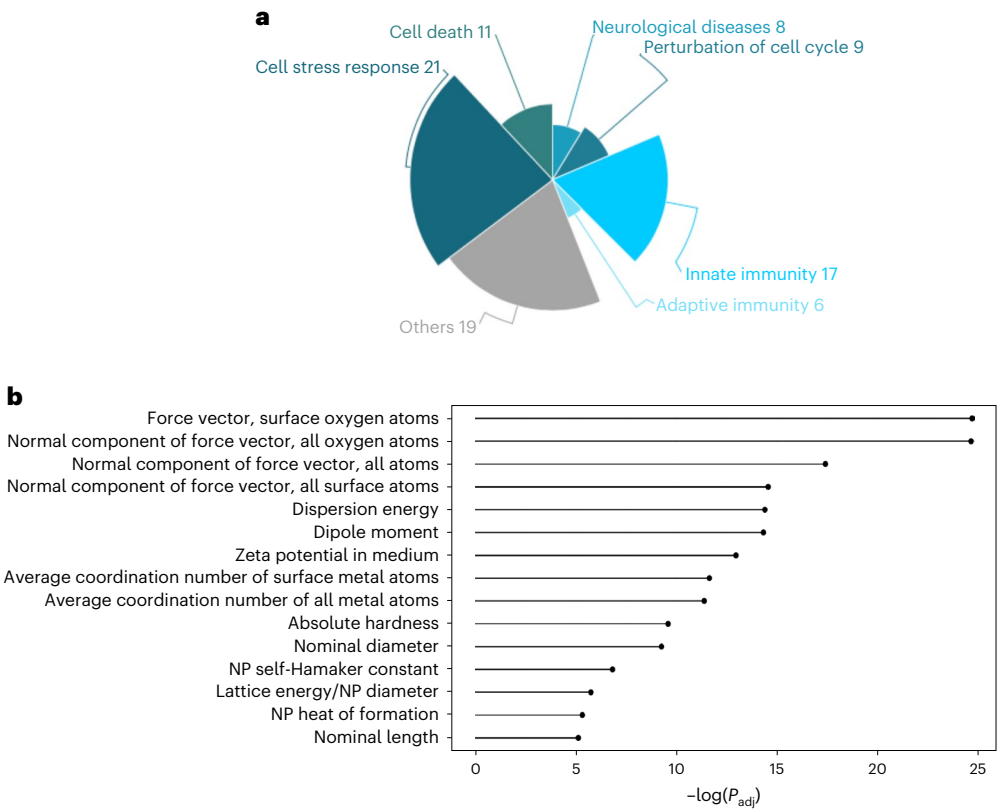


Fig. 2 | Characterization of the ENM molecular signature. a, Pie chart representing the functions associated with the meta-analysis rankings, grouped into six main biological categories (otherwise listed as 'Others'). The number of pathways falling within each category determines their size. The complete list of pathways is shown in Supplementary Table 2. **b**, Molecular descriptors correlated with genes at the top of the ENM ranking. For each molecular descriptor, the top 10% of genes ($n = 367$ for each bar) in the dataspace, whose expression is correlated, are selected. A GSEA approach is used to highlight the descriptors whose correlated genes are enriched at the top of the ENM rank. The P value from the GSEA was corrected for multiple comparisons with the FDR method ($P < 0.01$). For more details, see Methods. NP, nanoparticle; P_{adj}, adjusted P value.

We then investigated which ENM characteristics are associated with our signature. We curated a set of 159 molecular descriptors and 46 experimental labels and computed their correlation with gene expression (Supplementary Table 3).

The top correlated descriptors suggest that nanoparticle thermodynamic stability and ion release has a role in defining toxicity (Fig. 2b). However, our results indicate that the nanoparticle-induced ion release is not overlapping with bulk ionic compounds. Interestingly, Gupta et al. hypothesized that the difference in the MOA of copper ionophores and nano-copper may relate to massive internalization of nanoparticles and their ability to trigger protein aggregation and proteasomal inhibition in lysosomes[23]. Indeed, we cannot exclude that intracellular localization and other unique properties of ENMs may induce different responses and, ultimately, differences in nucleus translocation. Enriched descriptors relevant for ENM–bio interactions inform on the tendency of the nanoparticles to aggregate, form protein corona and interact with cell membranes, eventually modulating the toxicological potential of nanomaterials. It is noteworthy that nanoparticle characteristics highlighted so far refer to their chemical composition and do not directly depend on particle size. Our analysis, however, also indicates the importance of geometrical descriptors. This suggests that nanoscale properties are relevant for toxicity in addition to the intrinsic size-independent representation of ENMs.

Finally, we investigated whether our gene signature is ENM specific or related more generally to xenobiotic exposure. To this end, we tested its similarity with other gene signatures associated with small molecules and inorganic compounds in their bulk form. We applied the same meta-analysis pipeline to the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System dataset, where the rat liver transcriptome is measured after multi-dose treatment with 158 small molecules. Our analysis showed substantial differences between small molecules and ENMs (Supplementary Fig. 6), suggesting that the identified transcriptional signature is indeed specific to ENMs. As the ENM molecular signature may still be shared with other types of compound, we tested whether it shares similarities with ionic and/or covalent compounds of the same elemental composition. We retrieved the gene signatures of 142 inorganic substances from the Comparative Toxicogenomics Database (CTD) and demonstrated that no bulk compound elicits significantly similar molecular responses to those observed upon ENM exposure (Fig. 3f and Supplementary Table 4).

This suggests that the unique properties of nanoforms influence the response of biological systems. Indeed, ENMs differ from drugs and inorganic compounds because of their unique physicochemical properties, requiring more information to be described than just their structure[24,25].

Identification of a regulatory model for ENM exposure

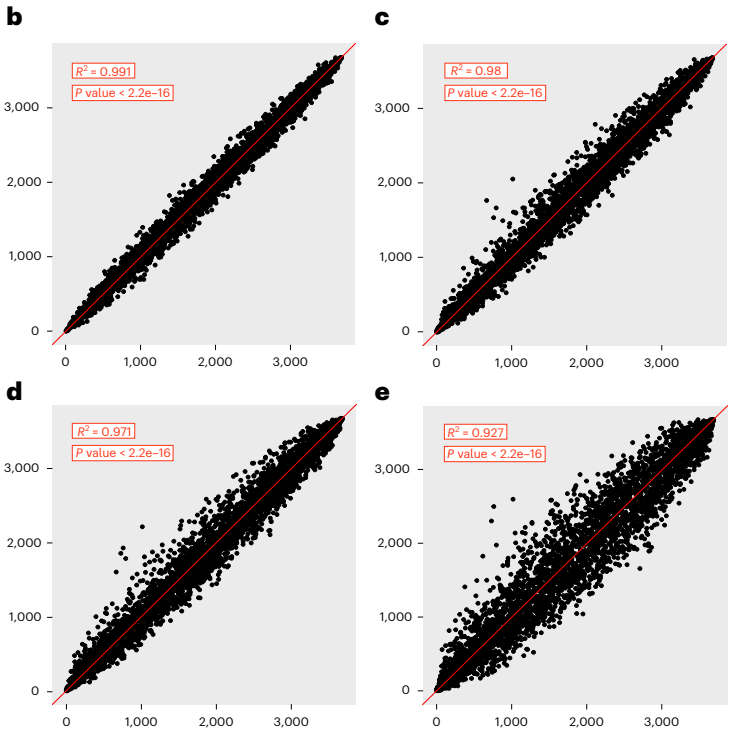
We hypothesized that the transcriptomics meta-analysis allows the identification of specific patterns of alterations common to long-term ENM exposure, in vitro and in vivo. We focused on the top-ranked 1,872 genes with the most significant functional enrichment (Methods).

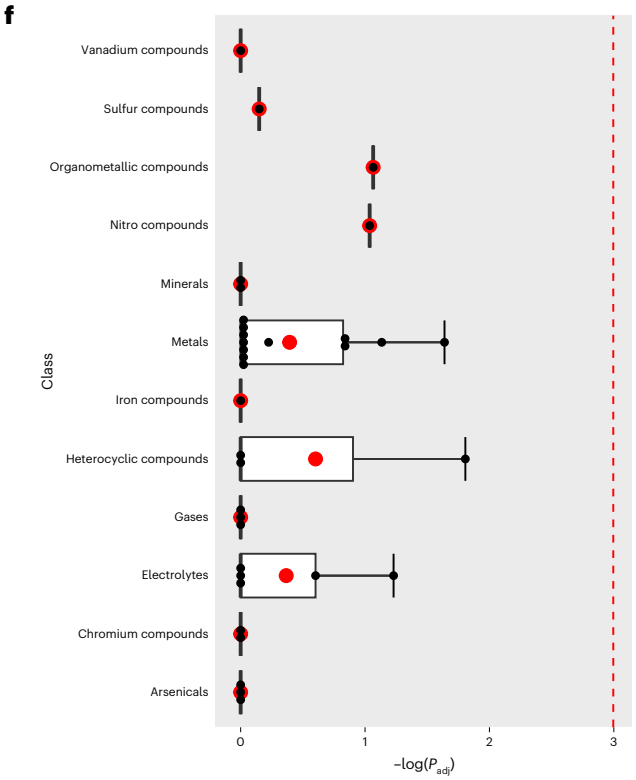
When toxic doses are used, both in vitro and in vivo assays succeed in capturing ENM acute toxicity. However, these effects are less evident at sub-toxic doses, especially long-term. About 81% of the studies in our collection screened sub-toxic exposures; hence, we investigated

https://doi.org/10.1038/s41565-023-01393-4

a

	Complete dataset	Complete dataset – ions	Complete dataset – ions – silver	Complete dataset – ions – silver – metals	Complete dataset – ions – silver – metals – functionalized
Complete dataset	1.0000000	0.9421595*	0.9140515*	0.8975029*	0.8327685*
Complete dataset – ions		1.0000000	0.9163835*	0.8996916*	0.8339855*
Complete dataset – ions – silver			1.0000000	0.9209102*	0.8484531*
Complete dataset – ions – silver – metals				1.0000000	0.8792296*
Complete dataset – ions – silver – metals – functionalized					1.0000000





g

Class	Mean	s.d.	Min	Max	Wish_low	Wish_high
Arsenicals	0.00	0.00	0.00	0.00	0.00	0.00
Chromium compounds	0.00	0.00	0.00	0.00	0.00	0.00
Electrolytes	0.36	0.55	0.00	1.23	0.00	0.60
Gases	0.00	0.00	0.00	0.00	0.00	0.00
Heterocyclic compounds	0.60	1.04	0.00	1.81	0.00	0.90
Iron compounds	0.00	NA	0.00	0.00	0.00	0.00
Metals	0.39	0.57	0.00	1.64	0.00	0.82
Minerals	0.00	0.00	0.00	0.00	0.00	0.00
Nitro compounds	1.04	NA	1.04	1.04	1.04	1.04
Organo-metallic compounds	1.07	NA	1.07	1.07	1.07	1.07
Sulphur compounds	1.15	NA	1.15	1.15	1.15	1.15
Vanadium compounds	0.00	NA	0.00	0.00	0.00	0.00

Fig. 3 | The ENM molecular signature is not shared with inorganic compounds. a, Correlation values between the meta-analysis rank obtained from the complete dataset and the rank obtained from partial datasets, progressively removing ion-releasing ENMs. Statistical significance between ranks has been determined with Kendall's τ correlation test (marked with a red asterisk; $P < 1 \times 10^{-16}$). **b**–**e**, Comparison between the positions of the rank obtained by including the complete dataset and progressively removing ENMs potentially releasing ions. X-axis represents the gene positions in the original rank. The four plots include the following individual comparisons: complete set minus ions versus complete set (y-axis represents the gene positions in the rank minus ions) (**b**); complete set minus ions and silver versus complete set (y-axis represents the gene positions in the rank minus ions and silver) (**c**); complete set minus ions, silver and pristine metals versus complete set (y-axis represents the gene positions in the rank minus ions, silver and pristine metals) (**d**); and complete set minus ions, silver, pristine metals and functionalized metals versus complete set (y-axis represents the gene positions in the rank minus ions, silver, pristine metals and functionalized metals) (**e**). The fit is reported with a red line. In the top left boxes, the R^2 and the P value of the limma fit are included. The P value was determined with a moderated t-test. **f**, Box plot of the gene set enrichment P values of the gene signatures of inorganic compounds derived from the CTD and the ENM molecular signature. The complete results are reported in Supplementary Table 2. The results are presented per compound category, as described by the CTD annotation. The significance threshold is indicated with a red dashed line. As the P value is expressed as minus logarithm, the significance threshold of 0.05 has been drawn at 2.995732. Individual data points are reported as black dots. The data are presented as mean values (red dots) ± standard error of independent experiments (n = 1, 1, 1, 1, 2, 12, 1, 3, 3, 5, 2, 3 for each bar, respectively). **g**, Box plot in **f** has been defined in terms of minima, maxima, bounds of box, whiskers and percentile.

the longer experimental time points and hypothesized that they could inform on shared responses of biological systems.

When comparing patterns of gene expression alteration in long-term post-exposure monitoring in vitro and in vivo (compare with Methods; Fig. 4a,b), we identified two distinct clusters of genes deregulated in at least 40% of instances (Fig. 4c and Supplementary Tables 5 and 6). In vivo and in vitro clusters contained 319 and 273 genes, respectively, sharing 56 genes (Supplementary Table 7). Functional

https://doi.org/10.1038/s41565-023-01393-4

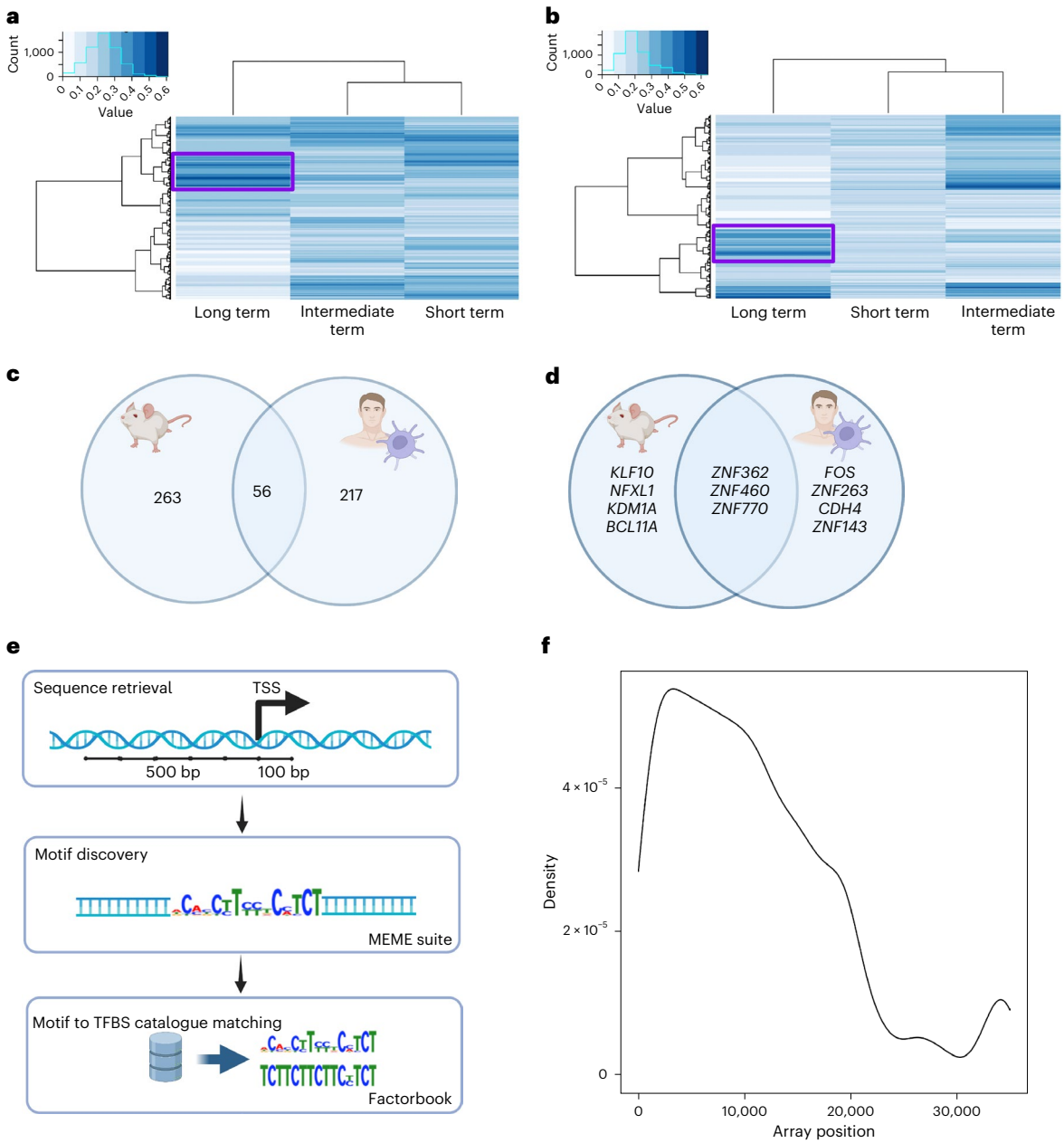


Fig. 4 | Identification of a gene regulatory model for ENMs exposures. **a**,**b**, Genes prioritized through functional analysis in vitro (**a**) and in vivo (**b**) are clustered according to the deregulation frequency following short, intermediate and long-term post-exposure monitoring. The cluster containing genes altered in at least 40% of the long-term monitoring samples (highlighted by the purple boxes) are selected for further investigation. **c**, Venn diagram showing the overlap between the two clusters of genes (in vitro and in vivo). **d**, Venn diagram showing the overlap of the regulatory motifs whose expression was statistically significant in the promoter regions of the two clusters of genes. **e**, Conceptual pipeline of the promoter analysis performed in this study. First, the DNA sequence of the [−500, +100] region around the TSS is retrieved. Motif discovery is performed through the MEME suite, finding all the DNA motifs between 6 and 15 base pairs that would satisfy a *P* value threshold of 0.05. MEME estimates the significance using an approximation to the *E* value of the information content of the motif. Finally, all the motifs are matched to the closest TFBS returned by the Factorbook database. **f**, Density plot of the position of genes regulated by the C_2H_2–ZNF family of transcription factors in the original complete ENMs dataset. On the *x* axis, the positions of genes ordered by the *P* value of their alteration is indicated. The *P* value of the gene expression is determined with a moderated *t*-test in the limma package R. On the *y* axis, the frequency of C_2H_2–ZNF targets for each position.

annotation highlighted unfolded protein response (UPR), apoptosis, and alteration of cellular metabolism and cell membranes. UPR is induced after accumulation of misfolded proteins in the endoplasmic reticulum, activating processes to ease reticulum stress and ultimately inducing apoptosis if it cannot be reverted[26]. ENMs induce endoplasmic reticulum stress through various mechanisms, including reactive oxygen species, leading to UPR[27,28]. ENMs are also known to cause protein unfolding and fibrillation, potentially disrupting cellular proteostasis[16]. Complement proteins, acute phase and tissue leakage proteins form ENM coronas acting as 'thorns' inducing proinflammatory response[29]. All these responses are expected to occur in sub-toxic long-term exposures (Supplementary Table 8).

Long-term physiopathological changes are well captured by epigenetic modifications, while transient molecular responses are usually associated with transcriptional alteration[30]. Therefore, we investigated whether the long-term altered genes were commonly regulated.

https://doi.org/10.1038/s41565-023-01393-4

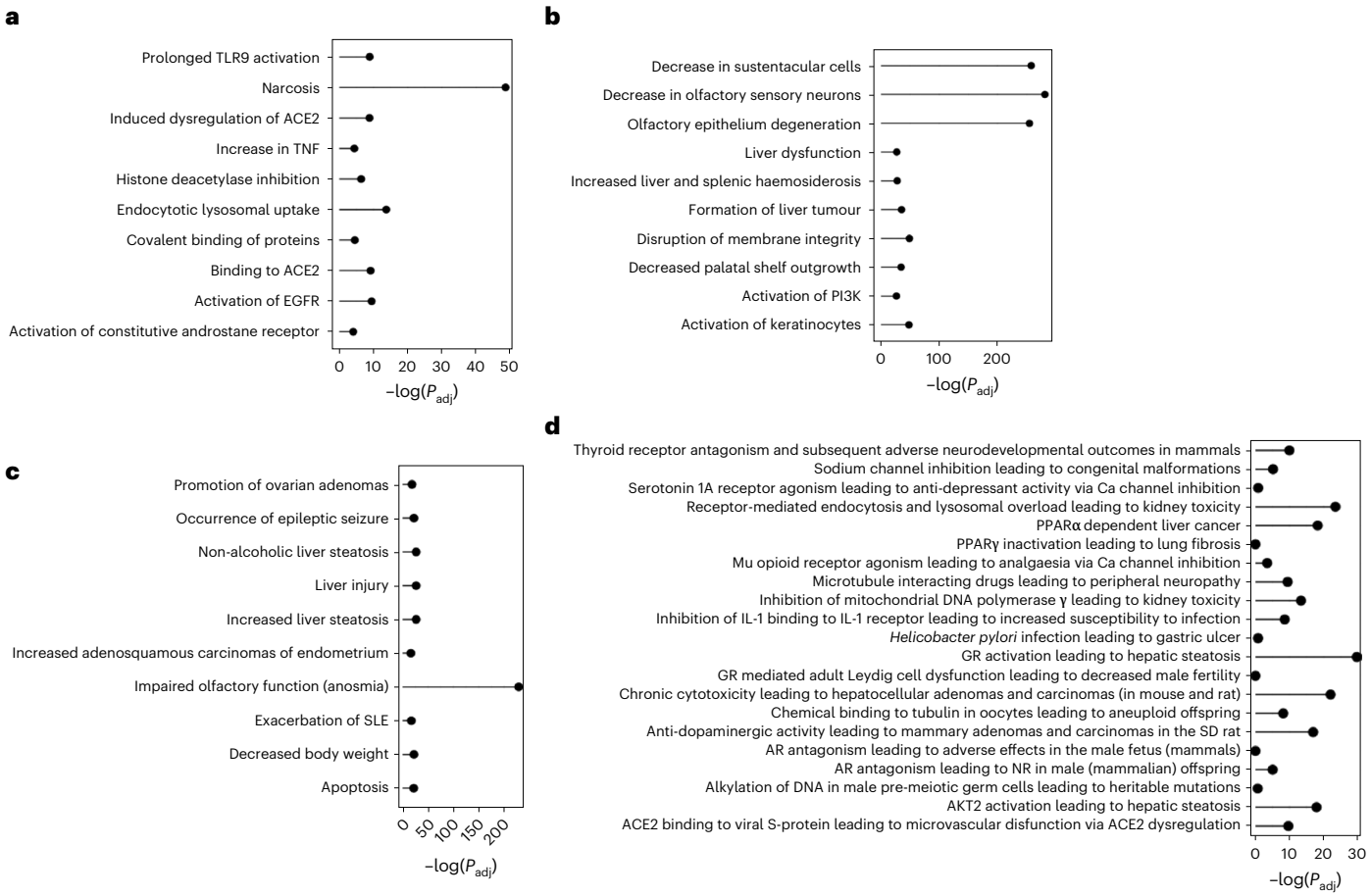


Fig. 5 | C_2H_2 family members enrich relevant AOPs. a–**c**, Top ten enriched MIEs, key events and adverse outcomes, with their respective adjusted *P* value (expressed as minus the logarithm of the value). TLR, toll-like receptor; ACE2, angiotensin-converting enzyme 2; TNF, tumour necrosis factor; EGFR, epidermal growth factor receptor; SLE, systemic lupus erythematosus; PPAR, peroxisome proliferator-activated receptor; IL-1, interleukin-1; GR, glucocorticoid receptor; SD, Sprague–Dawley; AR, androgen receptor; NR, nipple retention.; AP-1, activator protein 1; PIN-1, peptidylprolyl cis/trans isomerase NIMA-interacting 1; EMT epithelial mesenchymal transition. **d**, Top 20 enriched AOPs; JNK, jun N-terminal kinase; TGF, transforming growth factor. Besides the classical enrichment, we filtered pathways according to the proportion of individual key events enriched and discarded the ones in which less than one-third of the pathway is covered. **e**, Example of the 'endocytic lysosomal uptake leading to liver fibrosis' AOP, enriched by the regulatory model we identified. MIEs and adverse outcomes are represented in blue and green, respectively. Events reported as transparent do not pass the significance threshold ($P < 0.05$). **f**, Example of the 'Toll-like receptor 4 activation and peroxisome proliferator-activated receptor gamma inactivation leading to pulmonary fibrosis' AOP, enriched by the regulatory model we identified. Graphical annotation is the same as in **e**. The statistical significance in **a**–**e** was determined with a Fisher test, and multiple comparisons adjustment method was performed with the Bonferroni correction.

We retrieved the promoter sequences of each gene both in vitro and in vivo and searched for conserved motifs around the transcription starting site (TSS; Fig. 4e). We identified ten DNA motifs in vivo and eight in vitro ($P < 0.05$) (Fig. 4d,e and Supplementary Table 9). It is noteworthy that, when the regulatory layer of the transcriptome is considered, the similarity between in vitro and in vivo increases from 20% gene similarity (Fig. 4c) to more than 40% regulatory motif similarity (Fig. 4d). The discovered motifs mainly bind members of the

https://doi.org/10.1038/s41565-023-01393-4

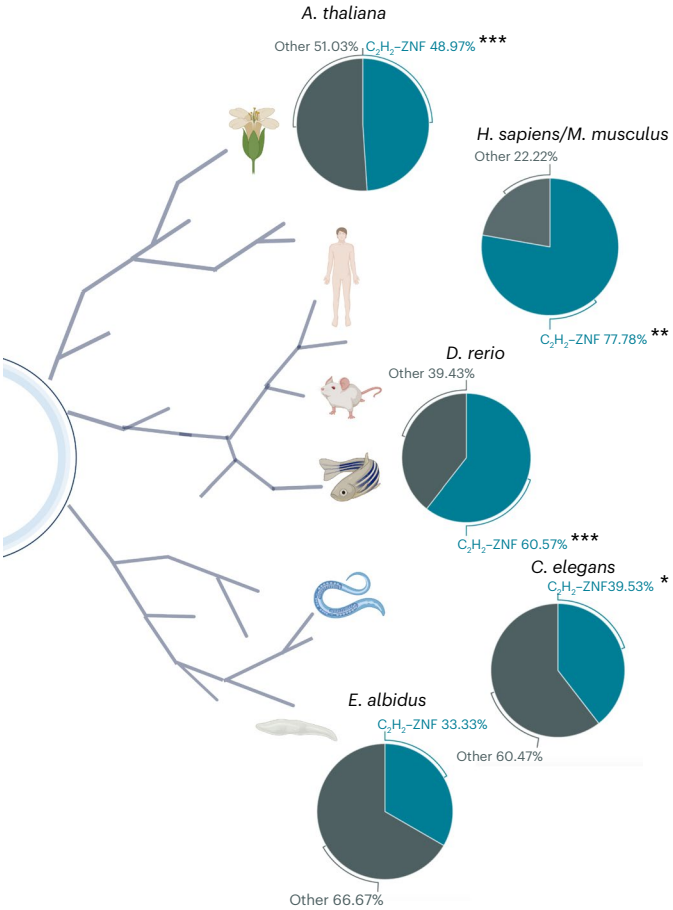


Fig. 6 | The C_2H_2 model is conserved in other non-mammal species. Overview of the species included in our study and relative proportion of C_2H_2–ZNF TFBS in the promoter region of the genes involved in their adaptation response. The statistical significance was estimated with the Fisher test and is indicated by the number of asterisks. *$P < 0.005$, **$P = 1.4 \times 10^{-8}$ and ***$P < 2.2 \times 10^{-16}$.

C_2H_2 zinc finger subfamily (C_2H_2–ZNF), eluding to their central role in the conserved response to ENMs[31].

C_2H_2–ZNF are a large family of transcription factors in eukaryotes with key roles in development and differentiation[32], the majority of which conserve binding sites and effectors throughout evolution[33]. They bind to repeated and contiguous motifs on target sites, often associated with transposable elements[33]. More than 700 C_2H_2–ZNF genes exist in humans, accounting for more than 2% of human genes[34]. Some members have also been linked to immunomodulation and inflammation[35–37]. Disruption of ZNF has been associated with metal-ion toxicity, a secondary effect of metal concentration changes, altering gene expression and DNA repair[38]. Importantly, while they have already been studied in plants as abiotic stress regulators[39], their role in human toxicological responses remains unclear.

C_2H_2–ZNF transcription factors play a role in chromatin plasticity and recruitment of repressor complexes[40]. Modulation of chromatin structure is a commonly observed response to exogenous stimuli, for example, reduction of chromatin accessibility to avoid viral genomic insertion[41]. However, epigenetic mechanisms are involved in generic stress responses and can alter chromatin structure to produce transient translocation and nuclear reorganization[42,43]. This provides epigenetic memory of the environmental stimulus[44]. Notably, this mechanism has been observed in plants and simpler eukaryotic organisms[42,43]. Furthermore, topological associated domain borders are enriched in the C_2H_2–ZNF member CTCF (CCCTC-binding factor), whose binding determines chromatin domains[45]. Our results suggest that the epigenetic mechanisms found here, although still largely unexplored, can substantially aid the reconstruction of the ENM MOA. To validate our results, we selected ZNF362 (Fig. 3d), as it is expressed in all the main lung-derived cell lines (Supplementary Fig. 7a) and its epigenetic modifications have been associated with lung function in previous studies[46]. We performed dual luciferase assay to determine the activity of ZNF362 promoter in BEAS-2B cell line upon exposure to two different concentrations of carbon materials (MWCNT NM401 and carbon black) after 24 and 48 h, respectively (Supplementary Fig. 7c,d). At 24 h, carbon black induces ZNF362 especially at higher concentration (Supplementary Fig. 7c). At 48 h, ZNF362 is induced at high concentrations for NM401, although not significantly (Supplementary Fig. 7d). This is not surprising given the transient nature of the reporter vector transfections, which may be sub-optimal after 48 h. These findings provide an initial proof of concept of our model.

C_2H_2–ZNF regulate toxicologically relevant genes

We hypothesized that the C_2H_2–ZNF model can be used to interpret a larger proportion of ENM transcriptomic responses, not limited to the 3,676 genes common across the studies analysed here. We investigated whether the most altered genes in response to individual ENMs are regulated by C_2H_2–ZNF in 84 transcriptomic datasets, including 19 not present in the initial meta-analysis (Supplementary Table 8). We show that the C_2H_2–ZNF regulatory model explains the most relevant transcriptional alterations in all individual transcriptomes (Fig. 4f and Supplementary Fig. 8).

Next, we evaluated whether the C_2H_2–ZNF model is relevant for ENM risk assessment.

A main theme in modern toxicology is the ability to generate exposure mechanistic models. Adverse outcome pathways (AOPs) emerged as robust multi-scale models linking chemical exposures to adverse outcomes. The annotation of toxicogenomics data into AOPs is currently evaluated by the Organisation for Economic Cooperation and Development to include toxicogenomic evidence in regulatory safety assessment[47].

Here, we investigated molecular events regulated by C_2H_2–ZNF by enriching a recently curated gene annotation of AOPs[48]. Given the large number of genes targeted by any C_2H_2–ZNF, we restricted the analysis to specific C_2H_2–ZNFs identified in this study (Fig. 4d).

Events in AOPs are defined as molecular initiating events (MIEs), key events and adverse outcomes, based on their relationship with a specific endpoint. For this, we investigated the statistically significant over-representation of C_2H_2–ZNF targets against AOP-, MIEs-, key-events-, and adverse-outcomes-specific gene sets.

The enrichment of narcosis ($P < 7.82 \times 10^{-22}$), interaction with membrane receptors ($P < 1 \times 10^{-4}$) and lysosomal dysfunction ($P < 1.09 \times 10^{-6}$) MIEs suggest that the C_2H_2–ZNF regulatory model captures the initial events of the ENM–host interaction and encompasses cellular homeostasis disruption (Fig. 5a).

Furthermore, activation of the phosphoinositide 3-kinase (PI3K) signalling pathway ($P < 3.58 \times 10^{-12}$) induces inflammation and reactive oxygen species production ($P < 6.62 \times 10^{-11}$; Fig. 5b). Among PI3K effectors, NFkB has a central role in ENM response (Supplementary Table 10). Finally, the previously discussed effect on membrane homeostasis and immune system is supported by adverse outcome enrichment (Fig. 5c).

We also evaluated the enrichment of AOPs, restricting cases in which less than one-third of key events associated with an AOP are significant. We report the 20 most significant enrichments and graphically describe two AOPs, showcasing that the C_2H_2–ZNF model is directly linked to ENM MOA (Fig. 5d–f).

Focusing on dose-dependent molecular alterations is also relevant for regulatory purposes[49]. Therefore, we evaluated the portion of dose-dependent genes that are regulated by C_2H_2–ZNFs in 62 studies[50]. We demonstrate that C_2H_2–ZNF regulates on average 55.3% of all

dose-dependent genes across a wide range of experimental conditions (Supplementary Fig. 9).

Taken together, these results show that C_2H_2–ZNFs can explain molecular responses to ENMs also in a regulatory relevant perspective.

The C_2H_2–ZNF model is conserved in eco-toxicological species

Finally, we hypothesized that C_2H_2–ZNFs mediate ENM transcriptomic responses in other species of eco-toxicological interest. We analysed 17 datasets recently curated, including ENM exposures to *Danio rerio*, *Caenorhabditis elegans*, *Enchytraeus albidus* and *Arabidopsis thaliana*[11,51] (Supplementary Table 11). Our results indicate that in non-mammal organisms the ENM response is also regulated by C_2H_2–ZNF (Supplementary Table 12). Interestingly, the relative proportion of C_2H_2–ZNF transcription factor binding sites (TFBS) decreases across the phylogenetic tree, suggesting a possible association with organismal complexity[34] (Fig. 6). The organisms considered here have variable levels of organization of adaptive and innate immunity. Plants show host immunity controlled by polymorphic host genes, where resistance protein-mediated defence is based on 'altered-self' mechanisms of recognition[52]. These mechanisms commonly use epigenetic modification and chromatin remodelling to establish infection memory, achieving immunity even in the absence of specialized immune cells[53]. Recently, ref. 54 described immune-deregulation and stress response as a shared feature to lithium cobalt oxide exposure across taxonomic groups.

Conclusions

Response to environmental stimuli is a primitive function of living organisms that influences their evolution through epigenetic mechanisms that minimize harmful events, while increasing their fitness. Although ENMs originated during industrialization, all organisms were exposed to natural nanoparticles across geological eras[55–57]. This suggests that responses and mechanisms of adaptation result from a long evolutionary exposure history to nano-sized matter. Our results suggest that a very ancestral regulatory mechanism may be conserved across the tree of life and explain the response to ENMs. This proposed model captures commonalities of molecular response beyond the divergence of individual studies and points to shared immunomodulation of ENMs, even in non-immune cells.

The stability of the regulatory layer across species paves the way to new toxicological tests that can bridge ecotoxicology and human toxicology[58], possibly facilitating the development of more effective in vitro–in vivo extrapolation.

Exposome impacts on the epigenome, contributing to disease susceptibility[59]. Hence, association between DNA modifications (for example, DNA methylation) and exposures to environmental factors have been widely investigated[60,61]. Moreover, the epigenome regulates chronic responses to drugs[62]. Our results further suggest the role of additional layers of (epi)genomic regulation on long-term responses to ENMs.

However, efforts such as the one described here are challenged by limited homogeneity in reporting nanosafety experiments[63,64]. Nonetheless, our model is statistically robust and validated in the largest collections of curated nanotoxicogenomics data. Moreover, it is potentially relevant for risk assessment in multiple species and hence throughout the ENMs life cycle.

In conclusion, this study suggests that a wide range of biological systems respond to ENMs through a set of commonly regulated genes. We demonstrate that this signature is specific to ENMs and interlinked to their nano-properties. This study sheds light on an unexplored layer of ENM MOA and proposes a solution to the 'one-chemical-one-signature problem', currently limiting the use of toxicogenomics in chemical safety assessment.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41565-023-01393-4.

References

1. Kinaret, P. A. S. et al. Transcriptomics in toxicogenomics, part I: experimental design, technologies, publicly available data, and regulatory aspects. *Nanomaterials* **10**, 750 (2020).
2. Serra, A. et al. INSIdE NANO: a systems biology framework to contextualize the mechanism-of-action of engineered nanomaterials. *Sci. Rep.* **9**, 179 (2019).
3. Krewski, D. et al. Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health B Crit. Rev.* **13**, 51–138 (2010).
4. ENCODE Project Consortium. The ENCODE (encyclopedia of DNA elements) project. *Science* **306**, 636–640 (2004).
5. Herman, J. J., Spencer, H. G., Donohue, K. & Sultan, S. E. How stable "should" epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution* **68**, 632–643 (2014).
6. D'Urso, A. & Brickner, J. H. Mechanisms of epigenetic memory. *Trends Genet.* **30**, 230–236 (2014).
7. Yosef, N. & Regev, A. Impulse control: temporal dynamics in gene transcription. *Cell* **144**, 886–896 (2011).
8. Saarimäki, L. A. et al. Manually curated transcriptomics data collection for toxicogenomic assessment of engineered nanomaterials. *Sci. Data* **8**, 49 (2021).
9. Gallud, A. et al. Multiparametric profiling of engineered nanomaterials: unmasking the surface coating effect. *Adv. Sci. (Weinh.)* **7**, 2002221 (2020).
10. Kinaret, P. A. S. et al. Toxicogenomic profiling of 28 nanomaterials in mouse airways. *Adv. Sci.* **8**, 2004588 (2021).
11. Burkard, M., Betz, A., Schirmer, K. & Zupanic, A. Common gene expression patterns in environmental model organisms exposed to engineered nanomaterials: a meta-analysis. *Environ. Sci. Technol.* **54**, 335–344 (2020).
12. Manke, A., Wang, L. & Rojanasakul, Y. Mechanisms of nanoparticle-induced oxidative stress and toxicity. *BioMed. Res. Int.* **2013**, 942916 (2013).
13. Zhang, H. et al. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano* **6**, 4349–4368 (2012).
14. Papadiamantis, A. G. Predicting cytotoxicity of metal oxide nanoparticles using Isalos Analytics Platform. *Nanomaterials* **10**, 2017 (2020).
15. Sun, H., Jiang, C., Wu, L., Bai, X. & Zhai, S. Cytotoxicity-related bioeffects induced by nanoparticles: the role of surface chemistry. *Front. Bioeng. Biotechnol.* **7**, 414 (2019).
16. Mahmoudi, M., Kalhor, H. R., Laurent, S. & Lynch, I. Protein fibrillation and nanoparticle interactions: opportunities and challenges. *Nanoscale* **5**, 2570–2588 (2013).
17. Sadler, A. J. & Williams, B. R. G. Interferon-inducible antiviral effectors. *Nat. Rev. Immunol.* **8**, 559–568 (2008).
18. Shaw, A. E. et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol.* **15**, e2004086 (2017).
19. Chen, H. et al. In vivo study of spherical gold nanoparticles: inflammatory effects and distribution in mice. *PLoS ONE* **8**, e58208 (2013).
20. Khatri, M. et al. Evaluation of cytotoxic, genotoxic and inflammatory responses of nanoparticles from photocopiers in three human cell lines. *Part. Fibre Toxicol.* **10**, 42 (2013).

21. Perkins, T. N. et al. Differences in gene expression and cytokine production by crystalline vs. amorphous silica in human lung epithelial cells. *Part. Fibre Toxicol.* **9**, 6 (2012).
22. Ye, S.-F., Wu, Y.-H., Hou, Z.-Q. & Zhang, Q.-Q. ROS and NF-kappaB are involved in upregulation of IL-8 in A549 cells exposed to multi-walled carbon nanotubes. *Biochem. Biophys. Res. Commun.* **379**, 643–648 (2009).
23. Gupta, G. & Fadeel, B. Toxpoint: copper is the new showstopper. *Toxicol. Sci.* **189**, 3–4 (2022).
24. Papadiamantis, A. G. et al. Metadata stewardship in nanosafety research: community-driven organisation of metadata schemas to support FAIR nanoscience data. *Nanomaterials* **10**, 2033 (2020).
25. Walczyk, D., Bombelli, F. B., Monopoli, M. P., Lynch, I. & Dawson, K. A. What the cell "sees" in bionanoscience. *J. Am. Chem. Soc.* **132**, 5761–5768 (2010).
26. Adams, C. J., Kopp, M. C., Larburu, N., Nowak, P. R. & Ali, M. M. U. Structure and molecular mechanism of ER stress signaling by the unfolded protein response signal activator IRE1. *Front. Mol. Biosci.* **6**, 11 (2019).
27. Simard, J.-C., Durocher, I. & Girard, D. Silver nanoparticles induce irremediable endoplasmic reticulum stress leading to unfolded protein response dependent apoptosis in breast cancer cells. *Apoptosis* **21**, 1279–1290 (2016).
28. Christen, V., Camenzind, M. & Fent, K. Silica nanoparticles induce endoplasmic reticulum stress response, oxidative stress and activate the mitogen-activated protein kinase (MAPK) signaling pathway. *Toxicol. Rep.* **1**, 1143–1151 (2014).
29. Cai, R. et al. Corona of thorns: the surface chemistry-mediated protein corona perturbs the recognition and immune response of macrophages. *ACS Appl. Mater. Interfaces* **12**, 1997–2008 (2020).
30. Scala, G. et al. Multi-omics analysis of ten carbon nanomaterials effects highlights cell type specific patterns of molecular regulation and adaptation. *NanoImpact* **11**, 99–108 (2018).
31. Kurowska, E. et al. The C2H2 zinc finger transcription factors are likely targets for Ni(II) toxicity. *Metallomics* **3**, 1227–1231 (2011).
32. Cassandri, M. et al. Zinc-finger proteins in health and disease. *Cell Death Discov.* **3**, 17071 (2017).
33. Najafabadi, H. S. et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat. Biotechnol.* **33**, 555–562 (2015).
34. Seetharam, A. & Stuart, G. W. A study on the distribution of 37 well conserved families of C2H2 zinc finger genes in eukaryotes. *BMC Genomics* **14**, 420 (2013).
35. Stik, G. et al. CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nat. Genet.* **52**, 655–661 (2020).
36. Wu, L.-C. ZAS: C2H2 zinc finger proteins involved in growth and development. *Gene Expr.* **10**, 137–152 (2002).
37. Gounari, F. & Kee, B. L. Fingerprinting ikaros. *Nat. Immunol.* **14**, 1034–1035 (2013).
38. Hartwig, A. Zinc finger proteins as potential targets for toxic metal ions: differential effects on structure and function. *Antioxid. Redox Signal.* **3**, 625–634 (2001).
39. Han, G. et al. C2H2 zinc finger proteins: master regulators of abiotic stress responses in plants. *Front. Plant Sci.* **11**, 115 (2020).
40. Schmitges, F. W. et al. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res.* **26**, 1742–1752 (2016).
41. Liu, X. et al. Human virus transcriptional regulators. *Cell* **182**, 24–37 (2020).
42. Pascual-Ahuir, A. et al. Dose dependent gene expression is dynamically modulated by the history, physiology and age of yeast cells. *Biochim. Biophys. Acta, Gene Regul. Mech.* **1862**, 457–471 (2019).
43. Ahmed, S. et al. DNA zip codes control an ancient mechanism for gene targeting to the nuclear periphery. *Nat. Cell Biol.* **12**, 111–118 (2010).
44. McLeod, D. V., Wild, G. & Úbeda, F. Epigenetic memories and the evolution of infectious diseases. *Nat. Commun.* **12**, 4273 (2021).
45. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
46. Nicodemus-Johnson, J. et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI Insight* **1**, e90151 (2016).
47. *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)* (OECD, 2017); https://doi.org/10.1787/44bb06c1-en
48. Saarimäki, L. A. et al. Toxicogenomics data for chemical safety assessment and development of new approach methodologies: an adverse outcome pathway-based approach. *Adv. Sci.* e2203984, https://doi.org/10.1002/advs.202203984 (2022).
49. Serra, A. et al. Characterization of ENM dynamic dose-dependent MOA in lung with respect to immune cells infiltration. *Nanomaterials* **12**, 2031 (2022).
50. Saarimäki, L. A. Toxicogenomics analysis of dynamic dose–response in macrophages highlights molecular alterations relevant for multi-walled carbon nanotube-induced lung fibrosis. *NanoImpact* 100274, https://doi.org/10.1016/j.impact.2020.100274 (2020).
51. Gomes, S. I. L., Scott-Fordsmand, J. J. & Amorim, M. J. B. Profiling transcriptomic response of *Enchytraeus albidus* to Cu and Ni: comparison with Cd and Zn. *Environ. Pollut.* **186**, 75–82 (2014).
52. Sanabria, N. M., Huang, J.-C. & Dubery, I. A. Self/nonself perception in plants in innate immunity and defense. *Self Nonself* **1**, 40–54 (2010).
53. Spoel, S. H. & Dong, X. How do plants achieve immunity? Defence without specialized immune cells. *Nat. Rev. Immunol.* **12**, 89–100 (2012).
54. Curtis, B. et al. Cross-species transcriptomic signatures identify mechanisms related to species sensitivity and common responses to nanomaterials. *Nat. Nanotechnol.* https://doi.org/10.1038/s41565-022-01096-2 (2022).
55. Sharma, V. K., Filip, J., Zboril, R. & Varma, R. S. Natural inorganic nanoparticles—formation, fate, and toxicity in the environment. *Chem. Soc. Rev.* **44**, 8410–8423 (2015).
56. Gislason, S. R. et al. Characterization of Eyjafjallajokull volcanic ash particles and a protocol for rapid risk assessment. *Proc. Natl Acad. Sci. USA* **108**, 7307–7312 (2011).
57. Lespes, G., Faucher, S. & Slaveykova, V. I. Natural nanoparticles, anthropogenic nanoparticles, where is the frontier? *Front. Environ. Sci.* **8**, 71 (2020).
58. Rivetti, C. et al. Vision of a near future: bridging the human health–environment divide. Toward an integrated strategy to understand mechanisms across species for chemical safety assessment. *Toxicol. Vitr.* **62**, 104692 (2020).
59. Jirtle, R. L. & Skinner, M. K. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **8**, 253–262 (2007).
60. Agier, L. et al. Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based HELIX cohort. *Lancet Planet. Health* **3**, e81–e92 (2019).
61. Silver, M. J. et al. Environmentally sensitive hotspots in the methylome of the early human embryo. *eLife* **11**, e72031 (2022).
62. Csoka, A. B. & Szyf, M. Epigenetic side-effects of common pharmaceuticals: a potential new field in medicine and pharmacology. *Med. Hypotheses* **73**, 770–780 (2009).
63. Jeliazkova, N. et al. Towards FAIR nanosafety data. *Nat. Nanotechnol.* **16**, 644–654 (2021).

64. Saarimäki, L. A., Melagraki, G., Afantitis, A., Lynch, I. & Greco, D. Prospects and challenges for FAIR toxicogenomics data. *Nat. Nanotechnol.* **17**, 17–18 (2022).

Methods

Data collection and pre-processing of nanomaterial datasets

The meta-analysis of ENMs toxicogenomics studies in vitro and in vivo can identify common molecular MOA independent from the biological system under evaluation. To this end, we implemented a meta-analysis of 66 transcriptomic datasets derived from the public data collection curated by ref. 8 (https://zenodo.org/record/3949890#.YlPUri0RqH0), supplemented with data previously published in ref. 10 (GSE157266) and ref. 9 (GSE148705) (Supplementary Table 8). From the original collection, we excluded rats' datasets and the ones based on old microarray platforms, as they shared little probes with the more recent versions (Supplementary Table 8).

The datasets GSE148705 and GSE157266 were pre-processed using the eUTOPIA software (version commit December 2021), as previously described[8,65]. Briefly, we filtered probes with a value higher than the 0.8 quantile against the negative control in at least 75% of the samples. Data were normalized between arrays using quantile normalization[66]. No batch correction was needed for the dataset GSE148705, while GSE157266 was corrected for technical variation associated with variables 'dye' and 'slide' using the ComBat method[67]. Finally, we used the limma package (version 3.52.4) to compute the gene expression difference between each exposure in the dataset and the corresponding controls, correcting the *P* value using the Benjamini–Hochberg procedure. The aggregated, normalized and corrected expression matrix was then exported with no additional filtering.

In this study, we selected all pairwise comparisons between time–dose exposures and their respective controls. The final dataset comprised 584 specific exposure conditions (treatment, exposure time and dose, and biological system) and 3,676 genes, ranging across various human and mouse tissues and cell types.

The pool of 3,676 genes represents the intersection of genes present in all the experiments, limiting the mouse–human conversion to 1:1 orthology relationships (that is, where both genes in the pair have only one ortholog in the other species). The ortholog genes were converted using the getLDS function of the biomaRt R package (version 2.52.0)[68]. All data are available in the online Zenodo repository (https://doi.org/10.5281/zenodo.7674574).

Collection and pre-processing of transcriptomics data for drug exposure and chemical compounds. To assess the specificity of the ENM signature, raw microarray data for 158 drugs was downloaded from the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System database[69]. We applied the same pipeline to the in vivo exposures of rats to three dose levels of each drug.

Raw data were imported into R using the justRMA function from the R library Affy (version 1.60.0)[70]. Probe annotation to Ensembl genes was performed by the custom annotation files rat2302rnensgcdf (v. 22.0.0), downloaded from the brain array website, leading to 12,153 genes. The expression values were quantile normalized by means of the normalizeQuantile function from the R limma library (version 3.52.4)[66]. Differential expression analysis was performed for each drug, for each combination of dose level and time point (1,839 comparisons). The analyses were performed comparing the treated samples to the matched control samples of the same time point. As a result, $\log_2$ fold changes, *P* values and adjusted *P* values (by means of the false discovery rate (FDR) correction) were retrieved for all genes for each comparison.

To make a comparable evaluation, we only selected genes that had been included in the meta-analysis rank. As for chemical compounds, we downloaded, from the CTD, 142 gene signatures of ionic and covalent compounds[71]. We discarded gene sets smaller than 15 and bigger than 1,000, as it would have altered the statistics of the test, as well as hydrocarbons, alcohols and ethers, as they are classified as organic. For each of the remaining gene sets, we performed a gene set enrichment analysis (GSEA) of the ENM-associated rank and considered a significance threshold of *P* value adjusted to 0.05.

Collection of ecotoxicological transcriptomics data. To test the translatability of our model to non-mammal species of eco-toxicological interest, we verified whether genes altered in response to ENM in other non-mammal species are regulated by C_2H_2–ZNF transcription factors. To this aim, we downloaded from the Gene Expression Omnibus (GEO) seven datasets covering 17 exposures to well-known eco-toxicological model organisms (*D. rerio*, *C. elegans*, *E. albidus* and *A. thaliana*) and report the lists of differentially expressed genes (GSE80461, GSE32521, GSE70509, GSE73427, GSE77148, GSE41333, GSE47662)[11,51] (Supplementary Table 11). The genes have been used to perform promoter analysis as previously described for the discovery collection (compare with 'Promoter analysis').

Characterization of the 584 experimental conditions

From the initial collection, we identified 584 experimental conditions that are unique because of the biological system, ENM or experimental condition used. We manually annotated each experimental condition by curating the information in the original publications (Supplementary Table 3). First, we grouped samples according to the biological system and exposure setting. We also included the exposure duration by grouping samples into short, intermediate and long exposures. Different thresholds were defined for in vivo and in vitro. In detail, for in vitro experiments, exposures have been considered short at 24 h, intermediate between 24 and 72 h and long after 72 h, respectively, as we expected the (sub-)acute toxicity to be observed within the first days of the exposure. Indeed, ref. 72 recently reported that in vitro systems tested between 6 and 72 h reproduce a scenario accounting for acute toxicity assessment of chemicals.

For in vivo exposures, thresholds were 3 days (short), between 3 days and 1 month (intermediate) and more than 1 month (long-term). In the health chemical evaluation, the 28 day exposure Organisation for Economic Cooperation and Development protocols are considered as the preliminary tests to assess long-term toxicity, while chronic toxicity studies should have a length of 12 months[73–76] (Supplementary Fig. 10).

As for doses, only nominal doses for each experiment were available. It is noteworthy that nominal doses are not comparable between the exposures, further complicated by the heterogeneity of the measuring units reported (Supplementary Fig. 2b). We investigated that the majority of the studies included here tested sub-toxic dose exposures via a semi-automatic pipeline to scan the original manuscripts. We retrieved all the PubMed Central identifiers of the original articles. When the experimental details were not available, we considered the cited protocol in its references.

We used the BioPython Entrez (version 1.81)[77] api to retrieve the documents through the PubMed Central IDs. Finally, we parsed the resulting XML documents and searched the abstract and/or the whole article for keywords (Supplementary Table 8) referring to sub-toxic doses. Each positive result was returned with its context and manually validated. For a reduced number of datasets, the corresponding articles could not be retrieved automatically, so they were manually checked. We were able to find indications of sub-toxic doses in 47 out of 58 publications.

The nanomaterials used in the experiment were classified according to the chemical characteristics and the presence or absence of functionalized groups.

A panel of information was extracted from the original publications (when possible), covering crystal phase, purity, absence of endotoxins, coating, stabilizer and supplier information, as well as protocol information.

Finally, when characterized, data were reported regarding the nominal diameter, length and specific surface area; transmission electron microscopy diameter, width and length; Brunauer-Emmett-Teller surface area; number of walls; dynamic light scattering mean diameter and polydispersity index in water and medium; and zeta potential in water and medium.

Computation of molecular descriptors

A set of 159 ENM descriptors covering both molecular and electronic structure properties was computed. Liquid drop model molecular attributes[78] are calculated assuming that ENM can be represented as a spherical drop, where elementary molecules are tightly packed, while the density of clusters is equal to the particle mass density[78,79]. We computed the Wigner–Seitz radius (r_w), the number of ENMs in the analysed agglomerate (n), the number of surface elements (S), the surface–volume ratio (SV) and the aggregation parameter (AP)[78]. The Wigner–Seitz radius characterizes the minimum radius of interactions between individual molecules and is represented by equation (1):

$$r_w = \left(\frac{3M}{4\pi\rho N_A}\right)^{\frac{1}{3}} \quad (1)$$

where M is the molecular weight, ρ is mass density, and N_A is the Avogadro constant.

The number of ENMs in the agglomerate (n) is represented using equation (2):

$$n = \left(\frac{r_0}{r_w}\right)^3 \quad (2)$$

where r_0 is the radius of each ENM.

The number of surface elements (S) is represented by equation (3):

$$S = 4n^{-\frac{1}{3}} \quad (3)$$

where S describes the ratio of surface molecules to molecules in the volume (or surface ENMs in agglomerates).

The surface–volume ratio (SV) is represented using equation (4):

$$SV = \frac{S}{1-S} \quad (4)$$

where SV is the feature that describes the ratio of surface molecules to molecules in volume (or surface ENMs in agglomerates).

Size-dependent interfacial thickness (h) was calculated with equation (5)

$$h = 0.01 \times (T - 273) \times r^{0.35} \quad (5)$$

where r is the nominal size of the ENM and T is temperature[80].

The ENM electronic structure descriptors were computed by density functional theory and semi-empirical quantum chemical methods, while the Hamaker constants were evaluated from atomistic force fields and a continuum method[81,82]. ENMs interact via long-range van der Waals interaction, which is a major contribution to calculating the adsorption energies of biomolecules in water. Therefore, Hamaker constants are evaluated to describe bio–nano interactions in water through an atomistic force field approach and via Lifschitz theory[82]. In the Lifschitz theory[82] two materials are interacting through a medium; the Hamaker constant for the ENM and a biomolecule in water is calculated from optical parameters that are experimentally determined (Supplementary Table 13), while in the force field approach long-range dispersion interaction is calculated using the Lorentz–Berthelot rules for sigma (atom size) and epsilon (atom–atom interaction amplitude)[83,84]. For metal ENMs, we used CHARMM force field parameters[85]. For metal oxides and carbon ENMs as well as amino acids, lipids and sugars, the force fields are reported in ref. 86. Considering all atom–atom interactions between two molecular entities, the Hamaker constant is derived by an approximation of the combined sigma and epsilon dispersion parameters[87]. In this work, we also considered the interaction between two ENM pieces in water. The geometric structures of the bulk ENMs were optimized with density functional theory and the Perdew-Burke-Ernzerhof functional[88] using the SIESTA code[89]. The band gaps were also calculated by PBE[88], while the heat of formation, electronegativity, absolute hardness, dispersion energy per atom, dipole moment and static polarizability descriptors[81] were obtained on the self-consistent field level through the semi-empirical code MOPAC (http://OpenMOPAC.net) using the PM6-D3[90] parametrization. Finally, ionization potentials, electron affinities, and the global electrophilicity index were computed through self-consistent charge calculations (ΔSCC calculation) for the electronic states of the neutral and ion ENMs via the GFN1-xTB parameterization of the GFN-xTB code[91–95].

We further included a set of atomistic descriptors that are based on the chemical composition, potential energy, lattice energy, topology, size and force vectors[96,97]. Constitutional descriptors are the counts of atoms of different identity and/or location. Potential energy descriptors are derived from the force-field calculations, corresponding to the arithmetic means of the potential energies for specific atom types and/or locations in the ENM. Lattice energies are based on the same potential energies but presented as per metal oxide nominal units (MxOy) and describe the energy needed to rip away said unit from the ENM surface. The coordination number of atoms is defined as the count of the neighbouring atoms that lie inside the radius,

$$R = 1.2 \times (R_M \text{ and } R_O) \quad (6)$$

where R_M and R_O are the ionic radii of metal and oxygen ions, respectively. A low coordination number indicates that some atoms have missing neighbours and thus makes the ENM more unstable. The size was derived from the actual calculated ENM diameter. The force vector lengths have been derived from the structure optimization. For example, to derive the average length (V) of the surface normal component of the force vector for a shell region atom, its coordinates (x,y,z), force vector components (f_x,f_y,f_z) and distance from the centre of mass (d) are used:

$$V = \frac{(xf_x + yf_y + zf_z)}{d} \quad (7)$$

Sample values for TiO_2 (10 nm) are reported (Supplementary Table 14). For amorphous ENMs a multi-step procedure was used requiring the simulation of bulk metal or metal oxide materials above their melting temperature, the extraction of ENMs with the desired size and shape, and their subsequent cooling at the temperature of interest with a prescribed rate. Such a procedure has been applied to build spherical amorphous ENMs with the aid of the automated Enalos Demokritos KNIME nodes. All the data are hosted at the NanoPharos database (db.nanopharos.eu) and were converted into a ready-for-modelling format.

Meta-analysis implementation

We implemented a consensus of three algorithms for meta-analysis to prioritize the shared 3,676 genes.

As previously proposed[98,99], our pipeline is based on the effect-size, P value-based and rank-product methods. Usually, meta-analysis frameworks are based on effect-size methods, assessing within- and between-study variations across multiple studies. These methods outperform others when there is large between-study variation and small sample sizes. To implement it, the 'effect_sizes' function from the esc R package (version 0.5.1) was used with the P-value argument and the 'chi_esc' function[100]. The Fisher's sum of logs method combines individual P values. Fisher's sum of logs method was implemented by using the 'sumlog' function of the R package metap (version 1.8), giving as input the P values of each gene[101]. Finally, the rank product is a non-parametric statistical method to combine differential gene expression analysis results from individual studies based on the within-study gene ranks. To this end, the genes in each experiment were ranked based on the relevance of their associated P values, and

the 'RP.advance' function of the RankProd R package (version 3.24.0) was used to merge them via one-class analysis of the rank-product method[102,103]. This function allows combining data coming from different studies, such as in the case of datasets generated by different laboratories. For each method, a rank was generated. Finally, all the ranks were combined through the Borda function of the TopKlists R package (version 1.0.8)[104]. The final mean rank is reported in Supplementary Table 1.

GSEA and feature selection step

To select the most biologically relevant portion of the rank, we performed a GSEA of the meta-analysis rank on five databases (Wikipathways[105], Gene Ontology[106], Reactome[107], Kyoto Encyclopedia of Genes and Genomes[108] and MsigDB[109]). In each case, the 'fgsea' function from the fgsea R package (version 1.22.0) was used[110]. For each test, we identified the position of the rank having the highest peak value of cumulative enrichment statistics. We created a reduced representation of the meta-analysis gene rank by setting as a threshold the top 10th percentile of such values (1,873 genes).

Computation of the frequency score and hierarchical clustering

To find genes associated with in vitro and in vivo long-term exposures for each gene of the reduced meta-analysis rank, we calculated a frequency score as the percentage of samples in which the gene was statistically significant.

The genes were clustered according to the Euclidean distance of their frequency scores. The hierarchical clustering algorithm, with Ward linkage method, implemented into the 'hclust' function of the R package 'stats' (version 4.2.0) was used.

For each type of exposure system, we selected the cluster with the most frequently deregulated genes. To functionally annotate them, we performed a pathway enrichment analysis through the EnrichR online tool (accessed in 2021), using the MsigDB and Reactome databases[107,109,111–113].

Computation of the molecular descriptors–gene expression correlation

To identify associations between ENM chemical properties and molecular alterations induced in cells and organisms by their exposure, the Pearson correlation coefficient was computed for each pair of gene and molecular descriptor, after a pre-processing step. In particular, a Winsorize function of the DescTools R package (version 0.99.43)[114] was used to replace extreme values of $\log_2$ fold changes with less extreme ones. Moreover, a cube root transformation was applied to the molecular descriptor values.

As the molecular descriptor data layer contains missing data, the Pearson correlation was computed for the subset of samples where values were available.

For each descriptor the top 10% of the most correlated genes were selected. First, the gene sets were enriched against the Kyoto Encyclopedia of Genes and Genomes pathways by means of the FunMappOne tool (version commit December 2021)[115]. Only pathways with FDR-corrected $P < 0.05$ were considered significantly enriched. The molecular descriptors were further clustered in nine groups based on the Jaccard Index similarity of the shared enriched pathways. Lastly, the fgsea R package (version 1.22.0)[110] was used to perform a GSEA analysis and identify the molecular descriptors whose set of associated genes is enriched on the top of the ranked list of genes identified with the meta-analysis approach. Only molecular descriptors with an adjusted $P < 0.01$ were selected.

Promoter analysis. For each gene in the subset of interest, the sequence of the promoter region [−500 bp, +100 bp] around the TSS was downloaded using the biomart package and the getSequence function in 'coding_gene_flank' mode[68]. In this modality the function returns the flanking region of the gene including the untranslated regions.

Motif discovery was conducted with the MEME software suite (version 5.5.1)[116]. The motif site distribution was set as any number of repetitions; the search was restricted to motifs ranging between 6 and 15 bases and the *P*-value threshold to 0.05.

For each result, the Factorbook database was interrogated to explore the TFBS (https://www.factorbook.org/). Factorbook is a transcription factor-centric web-based repository associated with ENCODE ChIP-seq data, as well as multiple databases of TFBSs. We selected the TFBS that best matches the query according to the tool.

For each organism (*Homo sapiens* and *Mus musculus* in the discovery set, and *D. rerio*, *C. elegans*, *E. albidus* and *A. thaliana* in the eco-toxicological comparison) we annotated whether the TFBS would be recognized by a C_2H_2–ZNF member. To evaluate the statistical significance of C_2H_2–ZNF overrepresentation, we performed a Fisher test with the fisher.test function in the stats R package. We used as a background the set of non-redundant transcription factor binding profiles provided in the JASPAR database[117]. The contingency matrix was built by using the set of TFBS of the C_2H_2–ZNF family members and all the others, respectively.

Dual luciferase assay. BEAS-2B cells (ATCC, CRL-9606) were grown in BEGM (Lonza, CC-3170). Cells were cultivated in 75 cm^2 culture flasks at 37 °C with a humidified atmosphere of 5% CO_2. For all experiments, 500 µl of cells was seeded at a density of 3.75×10^3 cells per ml in 48-well plates. Cells were then left to rest overnight before transfections. One hour before transfection, the media was replaced with 225 µl of fresh BEGM per well. Two vectors were used for transfection: human cytomegalovirus (CMV) (positive vector control) and the ZNF362, created with VectorBuilder (Supplementary Fig. 7b). Per well, 0.25 µg of DNA vector in Opti MEM reduced serum medium (Gibco, 31985062) was added 1:1 with Lipofectamine 3000 reagent 6% V/V (ThermoFisher Scientific, 15338030) and P3000 enhancer reagent 4% V/V in Opti MEM reduced serum medium (Gibco, 31985062). About 25 µl of this transfection solution was added per well and mixed by gentle agitation of the plate. Twenty-four hours post-transfection, cells were exposed to one of the following: NM401 (JRC MWCNTs-NM401-JRCNM04001) or carbon black (CB, Orion Engineered Carbons, Printex 90) nanomaterials at either low (20 µg ml^{-1}) or high (100 µg ml^{-1}) concentration; or nefazodone hydrochloride (Sigma-Aldrich, N5536) at low (25 µM) or high (50 µM) concentration. NM401 and CB nanomaterials were prepared according to the nanogenotox protocol (https://safenano.re.kr/download.do?SEQ=175). Briefly, in a glass vial, 0.5% of final stock volume of ethanol was added to the initial weighed nanomaterial powder; 0.05%W/V BSA-BEGM was added for a final stock concentration of 0.2 mg ml^{-1}. The vial was then sonicated 2 times for 15 min in a water bath; this stock solution was then diluted in 0.05%W/V BSA-BEGM to create final solutions, and final solutions were sonicated for 15 min before addition to the wells. Vehicle control (VC) for nanomaterials was 0.05%W/V BSA-BEGM. Nefazodone hydrochloride was prepared in 0.05%W/V BSA-BEGM and DMSO (final DMSO concentration in well of 0.5%). VC for nefazodone hydrochloride was 0.5% DMSO in 0.05%W/V BSA-BEGM. Exposures were performed for 24 and 48 h.

The Dual-Glo luciferase Assay System (Promega, E2920) was used as per the manufacturer's guidelines to measure firefly and renilla luciferase activity, on a Spark multiplate reader (Tecan).

There were three samples measured for each vector and vehicle control, for each exposure. The mean signal of three background wells (cells only) was used to subtract background from luciferase measurements. The firefly luciferase activity was normalized to renilla luciferase and power transformed. When present, outliers were removed with the boxplot function in R. The *t*-test was used to investigate the differences between the experimental and control samples.

Dose-dependent gene analysis. To verify if the C_2H_2–ZNF model can explain the dose-dependent portion of the ENM response, a dose–response analysis of 62 studies derived from 33 datasets (all initially included in the analysis but GSE146708) was performed following the strategy implemented in the BMDx tool (version commit February 2022)[118]. Briefly, multiple models were fitted, and the optimal model was selected as the one with the lowest Akaike information criterion. The effective doses (BMD, BMDL and BMDU) were estimated under the assumption of constant variance. The benchmark response was identified by means of the standard deviation approach with a benchmark response factor (BMRF) of 1.349, corresponding to a minimum of 10% difference with respect to the controls. Only genes with lack-of-fit $P > 0.01$ and with estimated benchmark dose (BMD), benchmark dose lower confidence limit (BMDL) and benchmark dose upper confidence limit (BMDU) values were deemed relevant. Genes with BMD or BMDU values higher than the highest exposure dose were removed. Furthermore, genes whose ratio between the predicted doses is higher than the suggested values (BMD/BMDL > 20, BMDU/BMD > 20 and BMDU/BMDL > 40) were removed from the analysis.

AOP enrichment analysis. To enrich key events and AOPs, we exploited the recently curated annotation from ref. 48. For each gene set, we performed an enrichment of the C_2H_2–ZNF targets as derived from ref. 119. For individual events, the significance threshold was set at P value adjusted to 0.05.

As for the complete AOPs, we first enriched as described the genes annotated to each pathway and discarded P values higher than 0.05. In a next step, we evaluated if at least one-third of the events contained in it passed the same threshold of significance (0.05).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The pre-processed version of the transcriptomic datasets included in the discovery datasets, that is, ENM exposures of human and mouse samples, have been previously deposited at https://zenodo.org/record/3949890#.YlPUri0RqH0. The original datasets can be accessed at Array Express (https://www.ebi.ac.uk/biostudies/arrayexpress) with the entry code EMTAB6396 and at GEO (https://www.ncbi.nlm.nih.gov/) under accession numbers GSE103101, GSE112780, GSE113088, GSE117056, GSE122197, GSE127773, GSE146708, GSE148705, GSE157266, GSE16727, GSE17676, GSE19487, GSE20692, GSE29042, GSE35193, GSE39330, GSE41041, GSE42066, GSE42067, GSE42068, GSE43515, GSE45322, GSE45598, GSE4567, GSE46998, GSE46999, GSE50176, GSE51186, GSE51417, GSE51421, GSE51636, GSE53700, GSE55286, GSE55349, GSE56324, GSE56325, GSE60797, GSE60798, GSE60799, GSE60800, GSE61366, GSE62253, GSE62769, GSE63552, GSE63806, GSE68036, GSE75429, GSE79766, GSE81564, GSE81565, GSE81566, GSE81567, GSE81568, GSE81569, GSE82062, GSE84982, GSE85711, GSE88786, GSE92563, GSE92900, GSE92987, GSE96720, GSE98236 and GSE99929.

Transcriptomic datasets used for the eco-toxicological analysis are freely available at GEO under accession numbers GSE80461, GSE32521, GSE70509, GSE73427, GSE77148, GSE41333 and GSE47662. Transcriptomic datasets of small molecule exposure (Open-TG GATEs) have been downloaded from https://dbarchive.biosciencedbc.jp/en/open-tggates/download.html in November 2020. Functional data were downloaded from https://www.gsea-msigdb.org/gsea/msigdb/ version 7.2.

All the other relevant data and data supporting the findings of this study have been deposited in the online Zenodo repository (https://doi.org/10.5281/zenodo.7674574).

Code availability

All the relevant and custom code supporting the findings of this study has been deposited in the online Zenodo repository (https://doi.org/10.5281/zenodo.7674574) and on Github at https://github.com/fhaive/metanalysis_toxicogenomic_data.

References

65. Marwah, V. S. et al. eUTOPIA: solution for omics data preprocessing and analysis. *Source Code Biol. Med.* **14**, 1 (2019).
66. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
67. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
68. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
69. Igarashi, Y. et al. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* **43**, D921–D927 (2015).
70. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
71. Davis, A. P. et al. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkac833 (2022).
72. Macko, P., Palosaari, T. & Whelan, M. Extrapolating from acute to chronic toxicity in vitro. *Toxicol. Vitr.* **76**, 105206 (2021).
73. *Test No. 407: Repeated Dose 28-Day Oral Toxicity Study in Rodents* (OECD, 2008); https://doi.org/10.1787/9789264070684-en
74. *Test No. 412: Subacute Inhalation Toxicity: 28-Day Study* (OECD, 2018); https://doi.org/10.1787/9789264070783-en
75. *Test No. 452: Chronic Toxicity Studies* (OECD, 2018); https://doi.org/10.1787/9789264071209-en
76. *Test No. 410: Repeated Dose Dermal Toxicity: 21/28-Day Study* (OECD, 1981); https://doi.org/10.1787/9789264070745-en
77. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
78. Sizochenko, N. et al. From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. *Nanoscale* **6**, 13986–13993 (2014).
79. Sizochenko, N. et al. How the toxicity of nanomaterials towards different species could be simultaneously evaluated: a novel multi-nano-read-across approach. *Nanoscale* **10**, 582–591 (2018).
80. Sizochenko, N., Syzochenko, M., Gajewicz, A., Leszczynski, J. & Puzyn, T. Predicting physical properties of nanofluids by computational modeling. *J. Phys. Chem. C.* **121**, 1910–1917 (2017).
81. Rouse, I. et al. First principles characterisation of bio–nano interface. *Phys. Chem. Chem. Phys.* **23**, 13473–13482 (2021).
82. *Intermolecular and Surface Forces* (Elsevier, 2011); https://doi.org/10.1016/C2011-0-05119-0
83. Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* **248**, 127–136 (1881).
84. Berthelot, D. Sur le mélange des gaz. in *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* 126 1703–1855 (1966).
85. Martin, L., Bilek, M. M., Weiss, A. S. & Kuyucak, S. Force fields for simulating the interaction of surfaces with biological molecules. *Interface Focus* **6**, 20150045 (2016).
86. Brandt, E. G. & Lyubartsev, A. P. Molecular dynamics simulations of adsorption of amino acid side chain analogues and a titanium binding peptide on the tio_2 (100) surface. *J. Phys. Chem. C.* **119**, 18126–18139 (2015).

87. Munaò, G., Correa, A., Pizzirusso, A. & Milano, G. On the calculation of the potential of mean force between atomistic nanoparticles. *Eur. Phys. J. E Soft Matter* **41**, 38 (2018).
88. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
89. Soler, J. M. et al. The SIESTA method for ab initio order-N materials simulation. *Phys. Condens. Matter* **14**, 2745 (2002).
90. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
91. Grimme, S., Bannwarth, C. & Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (*Z*=1–86). *J. Chem. Theory Comput.* **13**, 1989–2009 (2017).
92. Bannwarth, C. Extended tight-binding quantum chemistry methods. WIREs *Comput. Mol. Sci*. https://doi.org/10.1002/wcms.1493 (2020).
93. Pracht, P., Caldeweyher, E., Ehlert, S. & Grimme, S. A robust non-self-consistent tight-binding quantum chemistry method for large molecules. Preprint at *ChemRxiv* https://doi.org/10.26434/chemrxiv.8326202.v1 (2019).
94. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
95. Spicher, S. & Grimme, S. Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem. Int. Ed.* **59**, 15665–15673 (2020).
96. Burk, J. et al. Fe-doped ZnO nanoparticle toxicity: assessment by a new generation of nanodescriptors. *Nanoscale* **10**, 21985–21993 (2018).
97. Tämm, K. et al. Parametrization of nanoparticles: development of full-particle nanodescriptors. *Nanoscale* **8**, 16243–16250 (2016).
98. Kan, M., Shumyatcher, M., Diwadkar, A., Soliman, G. & Himes, B. E. Integration of transcriptomic data identifies global and cell-specific asthma-related gene expression signatures. *AMIA Annu. Symp. Proc.* **2018**, 1338–1347 (2018).
99. Hong, F. & Breitling, R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* **24**, 374–382 (2008).
100. Lüdecke, D. esc: effect size computation for meta analysis. *CRAN* (2019).
101. Dewey, M. metap: meta-analysis of significance values. *CRAN* (2022).
102. Del Carratore, F. et al. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics* **33**, 2774–2775 (2017).
103. Hong, F. et al. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22**, 2825–2827 (2006).
104. Lin, S. Space oriented rank-based data integration. *Stat. Appl. Genet. Mol. Biol.* **9**, Article20 (2010).
105. Martens, M. et al. WikiPathways: connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).
106. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
107. Jassal, B. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
108. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
109. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
110. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://doi.org/10.1101/060012 (2016).
111. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
112. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
113. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **14**, 128 (2013).
114. Signorell, A. DescTools: tools for descriptive statistics. *R Package Version 0.99.43* (2021).
115. Scala, G., Serra, A., Marwah, V. S., Saarimäki, L. A. & Greco, D. FunMappOne: a tool to hierarchically organize and visually navigate functional gene annotations in multiple experiments. *BMC Bioinform.* **20**, 79 (2019).
116. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
117. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
118. Serra, A., Saarimäki, L. A., Fratello, M., Marwah, V. S. & Greco, D. BMDx: a graphical Shiny application to perform Benchmark Dose analysis for transcriptomics data. *Bioinformatics* **36**, 2932–2933 (2020).
119. Han, H. et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **46**, D380–D386 (2018).

Acknowledgements

This study was supported by the Academy of Finland project UNICAST NANO (322761) (D.G.), European Research Council (ERC) programme, Consolidator project ARCHIMEDES (101043848) (D.G), EU Horizon 2020 project NanoSolveIT (814572) (D.G) and NanoInformaTIX (814426) (D.G). A.S. and A.F. were supported by the Tampere Institute for Advanced Study. K.J. was supported by the National Science Center, Poland (TransNano project, grant number UMO-2020/37/B/ST5/01894). A.M. was supported by European Union's Horizon 2020 research and innovation programme (DIAGONAL Project, grant number 953152 and NanoInformaTIX project, grant number 814426). K.K. and V.L. were supported by Science Foundation Ireland (Bio-Interface project, Grant No. 16/IA/4506). D.G. thanks the principal investigators of the research cluster of Cell and Molecular Physiology (CMP) of the faculty of Medicine and Health Technology (MET) of Tampere University for the insightful discussion.

Author contributions

Conceptualization, D.G.; methodology, G.d.G., A.S., M.F. and A.F.; data generation, M.E.A., J.M., K.K., I.R., S.A.C., K.J., A.M., A.G.P., E.V., G.M., A.A., K.T., T.P. and V.L.; data curation, G.d.G., L.A.S., M.F. and P.A.S.K.; investigation, G.d.G., D.G., K.K., I.R. and V.L.; visualization, G.d.G. and A.S.; funding acquisition, D.G.; project administration D.G.; supervision, D.G., A.S. and A.F.; writing (original draft), G.d.G., D.G., A.S., L.A.S., M.F., J.M., K.K., S.A.C., I.R., K.J., A.M., A.G.P., A.A., K.T. and V.L.; writing (review and editing), G.d.G., N.S., M.G., I.L., A.F. and D.G.

Funding

Open access funding provided by Tampere University including Tampere University Hospital, Tampere University of Applied Sciences (TUNI).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41565-023-01393-4.

Correspondence and requests for materials should be addressed to D. Greco.

Peer review information *Nature Nanotechnology* thanks Scott Glaberman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Dario Greco

Last updated by author(s): Mar 17, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to download omics data. The original publications manuscripts were downloaded with Biopython API (www.biopython.org).
Data analysis	Preprocessing of transcriptomic data has been performed using eUTOPIA (https://github.com/Greco-Lab/eUTOPIA) version commit December 2021. Dose dependent analysis was performed using the BMDx tool (https://github.com/Greco-Lab/BMDx) version commit February 2022. Functional enrichment was performed with the FunMappOne tool (https://github.com/Greco-Lab/FunMappOne) version commit December 2021. Computation of molecular descriptors has been performed with the semi-empirical code MOPAC version PM6 (http://OpenMOPAC.net) , the SIESTA code (as described by Soler et al.) and the GFN-xTB method (reported in Bannwarth et al.). Promoter analysis was performed with the MEME suite version 5.5.1 (https://meme-suite.org/meme/) and annotation to transcription factor binding sites was performed with FactorBook (https://www.factorbook.org, version 1). All the custom analysis and scripts have been generated in the R environment version 4.0.5. The following R packages have been used: esc, version 0.5.1, metap version 1.8, RankProd version 3.24.0, TopKList version 1.0.8, Fgsea version 1.22.0, stats version 4.2.0, DescTools version 0.99.43, limma version 3.52.4, Biomart version 2.52.0, Affy version 1.60.0 . The code has been deposited in the online Zenodo repository (https://doi.org/10.5281/zenodo.7674574) and on Github under https://github.com/fhaive/metanalysis_toxicogenomic_data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The preprocessed version of the transcriptomic datasets included in the discovery datasets, i.e., ENM exposures of human and mouse samples, have been previously deposited at https://zenodo.org/record/3949890#.YlPUri0RqH0. The original datasets can be accessed at Array Express (https://www.ebi.ac.uk/biostudies/arrayexpress) with the entry code EMTAB6396, and at GEO (https://www.ncbi.nlm.nih.gov/) under accession number GSE103101, GSE112780, GSE113088, GSE117056, GSE122197,GSE127773,GSE148705, GSE157266, GSE16727, GSE17676, GSE19487, GSE20692, GSE29042, GSE35193, GSE39330, GSE41041, GSE42066, GSE42067, GSE42068, GSE43515, GSE45322, GSE45598, GSE4567, GSE46998, GSE46999, GSE50176, GSE51186, GSE51417, GSE51421, GSE51636, GSE53700, GSE55286, GSE55349, GSE56324, GSE56325, GSE60797, GSE60798, GSE60799, GSE60800, GSE61366, GSE62253, GSE62769, GSE63552, GSE63806, GSE68036, GSE75429, GSE79766, GSE81564, GSE81565, GSE81566, GSE81567, GSE81568, GSE81569, GSE82062, GSE84982, GSE85711, GSE88786, GSE92563, GSE92900, GSE92987, GSE96720, GSE98236, GSE99929.
Transcriptomic datasets used for the eco-toxicological analysis are freely available at GEO under accession numbers GSE80461, GSE32521, GSE70509, GSE73427, GSE77148, GSE41333, GSE47662. Drug exposed datasets have been downloaded from https://dbarchive.biosciencedbc.jp/en/open-tggates/download.html in November 2020. Functional data were downloaded from https://www.gsea-msigdb.org/gsea/msigdb/ version 7.2.
All the other relevant data and supporting the findings of this study have been deposited in the online Zenodo repository (https://doi.org/10.5281/zenodo.7674574).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	does not apply
Population characteristics	does not apply
Recruitment	does not apply
Ethics oversight	does not apply

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This study included 68 datasets of ENMs exposure to human and mouse previously manually curated and preprocessed according to standardized protocols (https://zenodo.org/record/3949890#.YlPUri0RqH0 , Saarimäki et al. DOI: 10.1038/s41597-021-00808-y). From the datasets we studied 584 experimental instances (ENM-dose-time point specific) and covered the expression of 3,676 genes. As for the drug exposures, expression of the same genes has been explored in 150 individual exposures which were freely available at the Open TGgates datates (https://dbarchive.biosciencedbc.jp/en/open-tggates/download.html). This study used all the ENMs exposure datasets that were available at the moment of the analysis that would be considered FAIR. Therefore this study relies on the largest curated collection of ENMs transcriptomic data currently available at the moment of the study. Indeed, all the datasets included have been selected as satisfying number of replicates (n=3), presence of untreated control samples (negative controls), and rigorous experimental design. Similarly, as for eco-toxicological datasets retrival, upon literature reviews, all datasets of ENMs exposures in non mammal species which would contain differentially expressed and annotated genes were selected for the study. Therefore no statistical methods were used to calculate the samples size because every datasets satisfying the quality requirements was taken into account.
Data exclusions	Datasets have been excluded in case the microarray platform used in the experiment was not commercially available (custom designed), or marginally represented.
Replication	Original omic data was performed in at least triplicates successfully. The raw files of the original files underwent strict quality control and only the samples passing the quality standard were included (Saarimäki et al. DOI: 10.1038/s41597-021-00808-y). Statistical evaluation of our results was internally performed through the application of the meta-analitical approaches, which represents a quantitative synthesis of the

individual experimental instances. Furthermore the results pointing to the zinc fingers models were successfully replicated in multiple datasets and species (GSE80461, GSE32521, GSE70509, GSE73427, GSE77148, GSE41333, GSE47662 and https://dbarchive.biosciencedbc.jp/en/open-tggates/download.html), supporting the reproducibility of this study.

Randomization

Randomization protocols for omics data were described in the original publications reporting each dataset (EMTAB6396,GSE103101, GSE112780, GSE113088, GSE117056, GSE122197,GSE127773,GSE148705, GSE157266, GSE16727, GSE17676, GSE19487, GSE20692, GSE29042, GSE35193, GSE39330, GSE41041, GSE42066, GSE42067, GSE42068, GSE43515, GSE45322, GSE45598, GSE4567, GSE46998, GSE46999, GSE50176, GSE51186, GSE51417, GSE51421, GSE51636, GSE53700, GSE55286, GSE55349, GSE56324, GSE56325, GSE60797, GSE60798, GSE60799, GSE60800, GSE61366, GSE62253, GSE62769, GSE63552, GSE63806, GSE68036, GSE75429, GSE79766, GSE81564, GSE81565, GSE81566, GSE81567, GSE81568, GSE81569, GSE82062, GSE84982, GSE85711, GSE88786, GSE92563, GSE92900, GSE92987, GSE96720, GSE98236, GSE99929). Samples in this study have not be allocated in experimental groups. As the aim of the study was to highlight commonalities between ENMs, this study does not perform any statistical analysis between groups but considers all exposures as ENMs based. A thorough characterization of the experimental samples has been provided and is based on the experimental information reported in the original studies and the physichochemical properties of the ENMs as reported in the original publications.

Blinding

Blinding techniques were not needed for this study. Blinding techniques are needed to minimize allocation bias. As explained, samples in this study have not been allocated in experimental groups. All samples have been considered as belonging to the ENMs groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

The human acute monocytic leukemia cell line THP-1 was purchased from the American Type Culture Collection (ATCC).

Authentication

The cells were used for up to 30 passages and were tested regularly using MycoAlert® mycoplasma detection kit (Lonza).

THP-1 cells were not authenticated by us but ATCC has performed cell line authentication by using STR analysis/profiling.

Mycoplasma contamination

All cell lines were tested negative for mycoplasma

Commonly misidentified lines
(See ICLAC register)

None