



Review

# Transcriptomics in Toxicogenomics, Part I: Experimental Design, Technologies, Publicly Available Data, and Regulatory Aspects

Pia Anneli Sofia Kinaret <sup>1,2,3,†</sup> , Angela Serra <sup>1,2,†</sup> , Antonio Federico <sup>1,2</sup> , Pekka Kohonen <sup>4,5</sup> , Penny Nymark <sup>4,5</sup> , Irene Liampa <sup>6</sup> , My Kieu Ha <sup>7,8,9</sup>, Jang-Sik Choi <sup>7,8,9</sup>, Karolina Jagiello <sup>10,11</sup> , Natasha Sanabria <sup>12</sup> , Georgia Melagraki <sup>13</sup> , Luca Cattelani <sup>1,2</sup> , Michele Fratello <sup>1,2</sup> , Haralambos Sarimveis <sup>6</sup>, Antreas Afantitis <sup>13</sup> , Tae-Hyun Yoon <sup>7,8,9</sup> , Mary Gulumian <sup>12,14</sup>, Roland Grafström <sup>4,5</sup>, Tomasz Puzyn <sup>10,11</sup> and Dario Greco <sup>1,2,3,\*</sup>

<sup>1</sup> Faculty of Medicine and Health Technology, Tampere University, 33200 Tampere, Finland; pia.kinaret@helsinki.fi (P.A.S.K.); angela.serra@tuni.fi (A.S.); antonio.federico@tuni.fi (A.F.); luca.cattelani@tuni.fi (L.C.); michele.fratello@tuni.fi (M.F.)

<sup>2</sup> BioMediTech Institute, Tampere University, 33200 Tampere, Finland

<sup>3</sup> Institute of Biotechnology, University of Helsinki, 00790 Helsinki, Finland

<sup>4</sup> Institute of Environmental Medicine, Karolinska Institutet, 17177 Stockholm, Sweden; pkpekka@gmail.com (P.K.); penny.nymark@ki.se (P.N.); grafstromrc@gmail.com (R.G.)

<sup>5</sup> Division of Toxicology, Misvik Biology, 20520 Turku, Finland

<sup>6</sup> School of Chemical Engineering, National Technical University of Athens, 15780 Athens, Greece; irini.liampa@gmail.com (I.L.); hsrarimv@central.ntua.gr (H.S.)

<sup>7</sup> Center for Next Generation Cytometry, Hanyang University, Seoul 04763, Korea; hakieumy12@gmail.com (M.K.H.); gksakdma0529@gmail.com (J.-S.C.); taeyoon@hanyang.ac.kr (T.-H.Y.)

<sup>8</sup> Department of Chemistry, College of Natural Sciences, Hanyang University, Seoul 04763, Korea

<sup>9</sup> Institute of Next Generation Material Design, Hanyang University, Seoul 04763, Korea

<sup>10</sup> QSAR Lab Ltd., Aleja Grunwaldzka 190/102, 80-266 Gdansk, Poland; k.jagiello@qsarlab.com (K.J.); t.puzyn@qsarlab.com (T.P.)

<sup>11</sup> Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

<sup>12</sup> National Institute for Occupational Health, Johannesburg 2000, South Africa; natashaS@nioh.ac.za (N.S.); maryG@nioh.ac.za (M.G.)

<sup>13</sup> Nanoinformatics Department, NovaMechanics Ltd., Nicosia 1065, Cyprus; melagraki@novamechanics.com (G.M.); afantitis@novamechanics.com (A.A.)

<sup>14</sup> Haematology and Molecular Medicine Department, School of Pathology, University of the Witwatersrand, Johannesburg 2000, South Africa

\* Correspondence: dario.greco@tuni.fi

† These authors contributed equally to this work.

Received: 10 March 2020; Accepted: 3 April 2020; Published: 15 April 2020



**Abstract:** The starting point of successful hazard assessment is the generation of unbiased and trustworthy data. Conventional toxicity testing deals with extensive observations of phenotypic endpoints in vivo and complementing in vitro models. The increasing development of novel materials and chemical compounds dictates the need for a better understanding of the molecular changes occurring in exposed biological systems. Transcriptomics enables the exploration of organisms' responses to environmental, chemical, and physical agents by observing the molecular alterations in more detail. Toxicogenomics integrates classical toxicology with omics assays, thus allowing the characterization of the mechanism of action (MOA) of chemical compounds, novel small molecules, and engineered nanomaterials (ENMs). Lack of standardization in data generation and analysis currently hampers the full exploitation of toxicogenomics-based evidence in risk assessment. To fill this gap, TGx methods need to take into account appropriate experimental design and possible pitfalls in the transcriptomic analyses as well as data generation and sharing that adhere to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. In this review, we summarize the

recent advancements in the design and analysis of DNA microarray, RNA sequencing (RNA-Seq), and single-cell RNA-Seq (scRNA-Seq) data. We provide guidelines on exposure time, dose and complex endpoint selection, sample quality considerations and sample randomization. Furthermore, we summarize publicly available data resources and highlight applications of TGx data to understand and predict chemical toxicity potential. Additionally, we discuss the efforts to implement TGx into regulatory decision making to promote alternative methods for risk assessment and to support the 3R (reduction, refinement, and replacement) concept. This review is the first part of a three-article series on Transcriptomics in Toxicogenomics. These initial considerations on Experimental Design, Technologies, Publicly Available Data, Regulatory Aspects, are the starting point for further rigorous and reliable data preprocessing and modeling, described in the second and third part of the review series.

**Keywords:** transcriptomics; toxicogenomics (TGx); high throughput; microarrays; sequencing; experimental design; engineered nanomaterials (ENM); toxicology; alternative risk assessment

---

## 1. Introduction

Toxicology aims to understand how various agents can cause harm to humans and the environment. Traditionally, this means extensive testing strategies including exposure to animal models and consequently observing the outcome in the means of physical symptoms, or other measurable parameters. In Europe, the legislative restrictions on animal testing [1] as well as the advent of ENMs led to the realization that conventional toxicity testing is no longer feasible. Toxicity assays have been developed originally for bulk chemicals and pharmaceuticals. Instead, ENMs and other advanced materials are a too broad and novel category to be tested in a similar manner as the more traditional substances. ENMs are defined as structures with at least one dimension smaller than 100 nanometers [2]. Thanks to their unique intrinsic properties, nanomaterials are increasingly utilized in several industrial fields. ENMs are known to interfere with the traditional assays, making conventional testing approaches less feasible [3–5]. Thus, new strategies are needed to keep up the pace of the new substances and materials entering the market as well as to support the reduction of animal experiments. To overcome this bottleneck in the testing of the adverse effects, TGx approaches have been suggested as a complementing method for traditional safety assessment [6,7]. This has led to the establishment of several prominent programs in the EU and worldwide aiming to generate omics-based big data. These could be used in systems biology-based approaches to build predictive models, simultaneously shifting the focus from *in vivo* animal testing towards *in vitro* methodologies and *in silico* modeling [8]. Studying gene expression changes in response to a foreign substance or a compound allows the building of a detailed picture of the organisms molecular levels of alteration and consequently understanding the MOA of the possible toxicant [7]. Many relatively mature methods for examining transcriptional changes already exist. Microarrays and RNA-Seq allow relatively thorough analyses of the transcriptome, and their costs have been steadily decreasing to a level where these assays are now accessible to most researchers. Moreover, large-scale public databases are constantly expanding when more and more data are being produced. The emergence of omics technologies has resulted also in the expansion of methods and tools to handle the complex and large data sets. What the community of toxicologists, biologists, computational scientists, and other professionals is still lacking are more unified and consistent strategies for experimental set-up and design, exposure conditions, and data analysis, all needed steps to efficient interpretation of the outcome and possible hazard identification. Only using harmonized methodologies and pipelines can aid achieving more comprehensive cooperation with regulatory agencies. In this review, we focus on how to create, handle and analyze global, large-scale transcriptomics data from a toxicity point of view, especially emphasizing the new era of nanomaterial research and hazard assessment.

Careful planning and study design are key elements for successful transcriptomics experiments. To obtain valid and unbiased data, as well as to prevent the inclusion of measurement uncertainties and data that misrepresents the changes in expression, several key steps need to be carefully planned and handled. To evaluate and ensure significant and reliable outcomes, robust statistical analyses with a proper amount of sample replicates and appropriate exposure time points and doses are essential together with proper reference (control) groups [9]. Most of all, recognizing the possible pitfalls in experimental design and statistical uncertainties is crucial. Furthermore, transparency and proper data recording are essential for subsequent data analysis and interlaboratory repeatability.

The existing literature about TGx methodologies is currently scattered and focused on particular facets of the workflow. In this review, we discuss all the relevant aspects of using transcriptomics in toxicogenomics, which could be utilized to build more robust strategies to further promote transcriptomics approaches also in regulatory assessment. To this end, in this review series we cover all the steps needed for performing successful toxicogenomics experimentation, from study design to robust and comprehensive computational approaches.

## 2. Replicates and Reference Samples

To afford reliable statistical power to the downstream analysis, the number of sample replicates is an important factor to consider before any further steps are taken. Some tools and methods have been suggested to estimate the statistical power of a transcriptomic experiment with respect to the number of samples included [10–12]. For example, in case of a microarray experiment aiming at identifying differentially expressed genes, the minimum sample size can be estimated by means of permutation method proposed by Tibishirami et al. [13]. This allows identification of the minimum number of samples required to achieve a specific sensitivity (the proportion of truly differentially expressed genes) at a specified false discovery rate level [14]. In the same way, for RNASeq analysis the compromise between the number of biological replicates and sequencing depth should be identified [15]. The price per sample varies depending on the technology [7]. Together with the possible lack of biological replicates especially of *in vivo* samples, the price is usually considered as a limiting factor in transcriptomics. In general, the larger the variance between the biological replicates, the more replicates are needed. For example, exposure to a genetically similar model organism, such as cell lines or inbred mouse strains, provides a relatively homogeneous genetic background with marginal bias derived from the individuals, allowing measurement of fewer replicates as compared to models with complex and varying genotype or growth conditions [16]. Patient samples, instead, possess a highly variable genetic background, and thus require correspondingly more replicates to confirm the measured changes. As with all experiments, in order to understand the change in gene expression due to the treatment or exposure, a baseline needs to be determined by comparisons against a negative reference group. It should be also kept in mind that, for example, gender or age might cause varying responses, especially when patient samples or animals are used to study transcriptional changes. New materials, such as ENMs, for example, bring additional challenges to the testing conditions. The number of replicates depends on the experimental setup and there are no universal standardized guidelines. Transcriptomics datasets derived from ENM-exposed samples deposited in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) generally comprise a number of replicates ranging between 3 and 8. For example, Bajat et al. studied the gold nanoparticles (AuNP) toxicity on *in vitro* human models. In this study, Caco-2 cells were exposed to spherical gold nanoparticles of two different sizes at two different concentrations and two time points. They used three replicates for each experimental condition, and their results indicate that, at high concentrations, smaller AuNPs induce metal exposure and oxidative stress signaling pathways [17]. Moreover, Poulsen et al. [18] studied the inflammatory responses of two different multiwalled carbon nanotubes MWCNTs by having an experimental setup with three doses and three time points and five replicates. The same experimental set-up was used by Bourdon et al. [19] to investigate the effect of the exposure to carbon black (CB) nanoparticles in mice. Another example is the study performed by Dymacek et al. [20] on the effect

of MWCNT in lung inflammation and pathogenesis. The analyzed dataset included four different doses and four time points with eight replicates for each experimental setup. As classical *in vitro* assays might be susceptible to interference originating from the unique physicochemical properties of ENMs (see section “Other technical considerations”), the most common control includes the addition of ENMs to the assay components alone, followed by measurement of the intrinsic fluorescence or absorbance of the ENMs [21]. Thereafter, additional controls are added, such as the inclusion of the ENM with the analyte, which is again measured, and so on. Additionally, comparison against positive reference might be beneficial for validating the experimental procedures [22,23]. Positive reference group could be for example a known toxic substance or bacterial component, depending on the expected outcome or cellular activation status. For example, in case of ENM, studying the toxic potential of hollow carbon nanotube structures, a positive or negative control could be an exposure to well-known substances [22–24]. Experimental planning is an integrated part of TGx analysis. In this context, given the interdisciplinary nature of TGx, it is crucial to consult experts from complementary fields to coherently link all the necessary steps of the experiment.

### 3. Time and Dose Selection

One major step in the toxicological assessment involves the outlining of the relationship between the amount of exposure and the biological response. Together with the selection of the appropriate model organism and a sufficient number of replicates, the doses and time points need to be carefully considered [25].

Dose selection for the TGx experiments largely depends on the exposure route (oral, airways, dermal, etc.), exposure frequency and duration, exposure system, as well as the model organism. Different computational models for ENM dose estimation in the respiratory tract of animals and for *in vitro* models have been developed [26,27]. As with other substances and pharmaceuticals, also in the case of ENM the possible toxic effects depend on several factors, and thus no universal guidelines exist for dose metrics. Poulsen et al. exposed mice with MWCNT by intratracheal instillation with 6, 18, and 54  $\mu\text{g}$  per mouse [28], whereas Wallin et al. exposed mice with titanium dioxide nanoparticles with dose of 18, 54, and 162  $\mu\text{g}$  per mouse [29]. Kinaret et al. instead, mimicked chronic exposure model by exposing mice by oropharyngeal aspiration for four consecutive days to 10 and 40  $\mu\text{g}$  of MWCNT and by inhalation exposure for 4 hours a day for four consecutive days with 6.2–8.2  $\text{mg}/\text{m}^3$  of MWCNT [30]. Moreover, Scala et al. exposed THP-1, A549 and BEAS-2B cells to eight different ENMs with 0.1, 0.5, 1, 5, 10, 50, 100, and 500  $\mu\text{g}/\text{mL}$  for 48 hours and concluded that the concentration of 10  $\mu\text{g}/\text{mL}$  induced low levels of cytotoxicity and high cell viability, enabling recording of the fine molecular changes caused by the ENM instead of the more dominant apoptotic and cellular stress signals [31]. In conclusion, the dose selection requires careful considerations and planning on the ENM-specific aspects and should be designed distinctively based on the research question.

In toxicology, the dose–response assessment is crucial in determining the dose thresholds in which a low-toxic substance becomes harmful, or a highly toxic substance does not cause harm. In more traditional dose–response assessment, to determine the highest dose, which does not cause statistically significant adverse effects, the No observed adverse effect level dose (NOAEL) can be used to determine the point of departure (POD) for the chemical. More recently, the benchmark dose (BMD) approach has been introduced, in which a dose–response model is fitted to the data and the response rate is recorded. These approaches are dependent on the dose selection and sample size. Traditionally, these methods require time- and resource-consuming methodologies of biological parameters such as tumor incidence, cell proliferation, or body weight changes. TGx, instead, provides faster strategies for determining BMD levels, with less screening and animal experiments [32]. Moreover, the goal of BMD is to build estimates about safe daily exposure levels, and for this, TGx approach can provide faster predictions allowing complex modeling of gene expression changes in thousands of features simultaneously. Further details on BMD modeling can be found in the third part of this review: “Transcriptomics in toxicogenomics, Part III: Data Modeling”.

Several efforts have been made to demonstrate that BMD values from transcriptional data are comparable with apical data [33–37]. TGx analyses are still relatively expensive; thus, a reasonable strategy to optimize between doses and specific end points of interest needs to be considered. Useful TGx profiles can be obtained by testing one dose in one time-point, although, in order to cover the boundaries of toxic effects (i.e., POD), more doses are required [38]. Nonetheless, one dose and one time point exposures will provide information about specific and direct MOA [39]. Alternatively, using multiple doses will help to disentangle complex mechanisms of dose-dependent molecular alteration but will not allow any interpretation of time course, or long-term effects. One dose with long term follow-up instead, will help to highlight the pathways of recovery. Multiple exposure time points, as well as doses, are beneficial in defining dose-dependent effects and transcriptional PODs [40]. To disentangle the kinetic patterns of molecular alterations exerted by an exposure, analyzing multiple time points becomes of paramount importance. As toxicity testing involves evaluation of acute phase response as well as possible chronic effects, exposure time points need to be cautiously considered in order to recognize the course of molecular activities. In general, microarrays and sequencing experiments provide information about a specific response at a given time. For this reason, one exposure time point results in only revealing a small fraction of the whole biological outcome. Additionally, enough dose and time points need to be measured in order to predict long-term toxic effects through advanced computational approaches and machine learning algorithms [41,42], as also discussed in detail in the third part of this review. Nonetheless, most of the gene expression changes are measurable in hours or days after exposure, rather than months or years. Studying the changes in expression profiles after relatively short exposure to toxicants is useful for investigating the relationships between acute and chronic mechanisms of toxicity. Furthermore, when relationships between exposure duration and the produced gene expression profiles are compared, secondary effects can be identified [43].

Other important issues that need to be considered are the comparable doses applied during the *in vitro* and *in vivo* studies. Before the comparison of the results, one needs to ensure that the tested *in vitro* doses are suitably equivalent for the *in vivo* situation [44]. Especially in the case of ENMs, the proper dose metrics are not easily defined, with surface area-to-volume ratio having a significant effect on ENMs properties and reactivity [45].

### 3.1. Exposures and Sample Quality

Following the ancient dictum “dose makes the poison”, when deciding whether an agent presents a toxicity concern, exposure assessment is central. Conversely, toxicity testing needs to be done at biologically realistic concentrations; otherwise, a hazard that is detected may not be meaningful. For thorough TGx analysis that aims at answering the questions about dose-dependency and safe limits, it is therefore important to consider multiple exposure doses. Recommendations for dose and endpoint selection, target organism, route and method of administration, and time of exposure are indicated in guidelines provided by regulatory entities like the European Commission’s science and knowledge service and the The Organisation for Economic Co-operation and Development (OECD) [46]. Before proceeding to the TGx experimentation, other measurements should be considered to fine-tune proper dose and time selection. For example, controlling viability, confluency, and cytotoxicity in *in vitro* models is recommended in order to avoid biased molecular effects caused by induced apoptotic signaling cascades and excess cellular stress caused by the growth conditions [47]. This might not be simple, as discussed in the subsequent paragraph “Other technical considerations”.

Chemical exposures for toxicological studies need to be performed in a highly controlled manner and with careful administration of the toxicant. Controlled exposure conditions with minimal effects derived from different environmental factors are crucial. Changes for example in temperature, pH, and agglomeration status are all having an impact on the outcome. Again, as an example, much discussion and arguments about proper testing methods for ENMs have risen due to properties such as

hydrophobicity, charge, and reactivity, causing for example agglomeration, protein corona formation, and other unexpected behaviors in biological systems [48].

Furthermore, after successful exposure, before performing extensive microarray or RNAseq experiments, the quality of the obtained RNA needs to be verified. One important step in sample processing, together with the proper storage conditions, is the purification of the mRNA samples. It needs to be ensured that no contamination for example from genomic DNA or RNase enzymes will interfere with the analysis. For this, spectrophotometric analysis gives an estimation of RNA purity and concentration. In order to confirm the integrity of the RNA molecules, more advanced measurements, such as microcapillary electrophoretic RNA separation, are recommended [49].

In toxicology, the route (exposure pathway) and duration of exposure, together with appropriate doses and chemical composition of the possible toxicant, affect the chosen exposure method and model organism. The principles of toxicology exposure can be found in guidelines from governmental institutions and the European Chemicals Agency (ECHA) and are outside the scope of this review.

### 3.2. Minimizing the Variation

Undesirable variation caused by technical and other non-biological sources are called batch effects. Well-known batch effects in transcriptomics data (further discussed in the second part of this review) include technical, environmental, and individual factors, such as changes in time, personnel, instruments, sample storage conditions and period, reagent slots, and slides/chips/wells. Technical artifacts and batch effects cannot be entirely avoided in high-throughput experiments, but careful design and sample randomization can reduce these effects significantly [50]. Therefore, batches and other sources of variation should be uniformly distributed across the biological groups [51]. In practice, after quality check, the samples need to be randomized in a controlled manner. This means, that, in order to reduce additional biases, samples, slides, and dyes are distributed (organized) to cover all possible sample combinations [52]. For example, if all control samples are labeled with the same dye, and treated samples with another dye, the additional effect caused by the labeling cannot be removed from the data. For this reason, carefully controlled randomization is a necessity. Additionally, if enough replicates are obtainable, some of the technical biases can be reduced by pooling several samples together before the experiment. Nonetheless, pooling should not be performed on biological samples, such as cancer tissue samples of patients, in which the individual differences are of high importance [25,53].

An often neglected part of transcriptional data management is the accurate recording of the metadata. This includes recording of all the possible sources of variation and batch effects such as changes in operators, exposure and procedure dates, sample storage and processing conditions, sample labeling, dyes, array and slot/position information, changes in reagent slots/batches, quality measurements, and other possible deviations. In order to ensure replicability of the experiment and the data analysis, accurate information about the samples need to be recorded and reported. For this, uniform phenotypic data files are essential, and need to be created as shown in the example in Table 1. Besides, major public databases such as the Gene Expression Omnibus (GEO) and ArrayExpress have adopted the Minimum Information About a Microarray Experiment (MIAME) and Minimum Information about a high throughput SEQuencing Experiment (MINSEQE) standards [54–56]. Thus, they offer full support for the submission of the genomic experiment data that these standards define as the minimum information that enables the interpretation and reproducibility of the results of relevant experiments. Brazma et al., when first proposed the MIAME documentation, argued that the initial gap in the information related to the quality and reliability of publicly available microarray data would affect not only the understanding of the outcome of their analysis but also any future effort to use the stored data for automated data analysis or mining. Furthermore, FAIR (Findable, Accessible, Interoperable and Reusable) Data Principle guidelines for data management, are highly recommended, allowing to promote the maximum reuse of research data (see section “Publicly

available datasets for toxicogenomics” paragraph) [57]. Supplementary standardized information relevant to the exposure details can further facilitate the reusability of TGx data [58].

**Table 1.** Example of the phenodata table reporting.

Sample	Sample ID	Date	Material	Dose	Time	Day	Array.N	Hybr Date	Slot	Slid Bar Code	Operator
12	121	12/03/2019	TiO <sub>2</sub>	10	24	Cy5	2	18.3.	1_2	757480412533_S05	PKI

### 3.3. Model Systems

An urgent issue in substituting animal toxicity testing with *in vitro* experimentations has been raised by Poulsen et al. [44]. The same concern extends to transcriptomics studies [44,59,60]. Efforts have been made to assess the relevance of the *in vitro*-based gene expression changes (DNA microarray as well as RNA-Seq data) to *in vivo* health effects for chemicals [44,59–61]. The replacement of *in vivo* transcriptomics studies with *in vitro* tests would allow investigation of specific biomarkers and pathways in a less expensive, more ethical and time-effective manner. *In vitro* approaches further allow several time points and doses to be tested with a more reasonable time scale when compared to extensive, time-consuming *in vivo* experimentation.

Unfortunately, the task is not simple nor straightforward and, at present, *in vitro* models cannot fully replicate *in vivo* outcomes. For example, the comparison of the *in vivo* inhalation of rats with the *in vitro* exposures using three cell lines (rat lung epithelial cells, rat primary alveolar macrophages and co-culture of lung epithelial and primary alveolar macrophages) performed for five different types of nanoparticles indicates only a little correlation in their toxicological behavior [44]. Similar observations were reported by Warheit et al. and Sayes et al., where the authors assessed the predictability of the *in vitro* lung response induced by nano-sized particles (nanoscale zinc oxide and fullerenes) for *in vivo* pulmonary hazard potential [62,63]. The authors postulated that these differences may be caused by the fact that *in vivo* responses evaluate the toxicity in whole organisms while the *in vitro* primarily focuses on the response of a single isolated cell type [44,62–64]. However, on the opposite side, there are studies with a good agreement between the *in vitro* and *in vivo* models in case of the genotoxic potential of ENMs [65]. This might indicate that proper selection of the cell type as well as doses and toxicity endpoints would be the crucial factors influencing this comparison [44], although genotoxicity is a special case involving the singular and universal molecular target, DNA. According to recommendations by Poulsen et al., another important issue should be the choice of data analysis [44]. They observed that, besides the fact that the regulations of most individual genes measured in the *in vivo* and *in vitro* approaches are different, at the pathway level the common cellular functions (i.e., the oxidative stress and fibrosis) were activated. Similarly, Kinaret et al. demonstrated that when proper data analysis is utilized and the different ENM properties are taken into account, similarities in transcriptomics responses between mouse lung *in vivo* and human macrophage cell lines after ENM exposures can be observed [66]. Despite a little overlap between perturbed genes *in vivo* and *in vitro* was observed, similar molecular and cellular functions were identified. Thus, as a strategy, they proposed a network-based response modules that project the knowledge from individual genes to transcriptome-based biological functions [67]. In effect, they postulated that the proper strategy of data analysis gives the possibility to use the transcriptomic signature measured *in vitro* to characterize the outcome for the *in vivo* scenario [66]. Although some studies highlight the correlation between exposure systems, more effort is needed to unify the different aspects between the systems, such as doses and time points, as well as data handling and transcriptomics analyses. Advanced co-culture systems and 3D models are already extensively explored, and will provide more realistic comparisons also for toxicogenomics. In this manner it should be possible to gain a more complete appreciation of differences between *in vitro* and *in vivo* TGx outcomes.

### 3.4. Other Technical Considerations

Especially in the case of ENMs, many additional technical considerations should be taken into account. A way to control the quality of the data and to ensure correct measuring procedures is to validate the assay in order to assess whether or not the target of interest is actually being detected, i.e., instead of false-positive results. This is an important step related to standardization in data generation and analysis, since most optical signal-based assays are subject to a form of assay interference, which is inadvertently caused by the presence of ENMs within the cell or cytosol [21,68–70]. For example, the most common molecule used in molecular cell biology assays would be an enzyme, which is a protein. Ong and colleagues reported that ENMs can bind to proteins and dyes and alter their structure and/or function [71–77]. They went further to explain that this may adversely influence reactions and cause significant changes in enzyme activity, fluorescence, and/or the absorbance characteristics of indicator molecules [71–76,78,79]. Thus, the false positive and/or negative results in conventional screening assays are due to interference caused by intracellular ENMs that affect the color, chemical activity, light scattering, etc. [80–82].

Some studies have tried to compensate for this phenomenon by including more controls, or by verifying one method with another similar method. Examples include the testing of the cytotoxicity of ENMs such as SiO<sub>2</sub> in A549 human alveolar epithelial cells, which was assessed by using two different toxicity assays [83–86]. As will be discussed below, microarray gene expression results are commonly validated by quantitative real-time PCR (qPCR). For example, the global transcriptomic analyses of exposures to surface-modified gold nanoparticles (Au-NPs) were investigated by Grzincic and colleagues. The authors tested gene expression in human dermal fibroblast (HDF) and prostate cancer (PIC3) cell lines via RNA microarray, which was confirmed via qPCR [87]. However, assay interference has been reported for various types of PCR [88–90]. Specifically, Li and colleagues reported a generalized hindrance of the PCR process [88]. Humes and colleagues concluded that assay interference occurred when the surface-oxidized multiwalled carbon nanotubes (oxMWCNTs) inhibited the reverse transcription step of RT-qPCR used to assess gene expression [88–90]. However, pristine MWCNTs did not cause the same level of interference, which emphasizes the role of surface chemistry in these types of analyses. Au-NPs were also reported to affect the reverse transcription efficiency in RT-qPCR [91]. Gao and colleagues later found that the efficiency of the high-fidelity DNA polymerase (Phusion) was significantly inhibited by some of the major types of metal oxide NPs (e.g., Fe<sub>2</sub>O<sub>3</sub>, ZnO, CeO<sub>2</sub>, FeO<sub>4</sub>, Al<sub>2</sub>O<sub>3</sub>, CuO, TiO<sub>2</sub>), but that this did not introduce mutations, i.e., the overall error rate was not significantly different and single nucleotide polymorphisms were not introduced [90]. The time-dependent quenching effect of ZnO NPs on fluorescence emissions was also investigated by Zhang and Lai [92]. These authors went further to determine the interference caused by other metal oxide nanoparticles in combination with the porphyrin quenching of ZnO NPs, i.e., an example of interactions between NPs in mixtures. These studies mentioned above support the need to assess the interactions between, especially metal oxide, intracellular NPs and genetic material, so as to explain the inconsistencies found in the literature regarding ENM-related TGx and/or toxicity.

In general, Omics techniques have been supposed to exhibit lower interference due to the removal of the ENMs during the isolation of the analyte [81]. Unfortunately, even if the ENMs are removed by repeated wash-steps within the isolation procedure, the damage might have already been done when intracellular or cell membrane-bound ENMs interfere with the biological molecules that are used to assess toxicity [93]. ENMs could also remain within the cells, or bound to the cell membranes [69,94]. Again, the recommendation is to test for the unintended interactions and to also validate results with multiple techniques that use different detection methods. In this manner, more stringent conditions for the use of these types of assays can be employed to prevent misinformation from being produced in nanotoxicity testing, which has substantial implications for our understanding of, as well as confidence in, the reported bioactivity of ENMs [21].



## 4. Transcriptomic Technologies in Toxicogenomics

### 4.1. Polymerase Chain Reaction, PCR

PCR is considered as a state-of-the-art method to study gene transcription of known sequences. The relative change in expression is measured by first synthesizing cDNA from the mRNA transcripts with reverse transcriptase enzyme and then using PCR to amplify the genes of interest. Quantitative PCR is a valuable assay for gene expression analysis, but has limits in number of analytes, enabling assay of only a limited number of genes at a time, making it time consuming and impractical from a toxicogenomics point of view. However, it is still important, and highly recommended to confirm and validate the obtained results from high content experiments. PCR-based multiplex methods have been developed, enabling the amplification of several sequences simultaneously. Nonetheless, from a toxicogenomics point of view, a genome-wide analysis is needed, for in-depth coverage of the changes in the whole genome.

### 4.2. DNA Microarray

DNA microarrays were first proposed by Schena et al. in 1995 [95]. In microarray technology, thousands of probes are fixed to a surface and samples (targets) are labeled with fluorescent dyes for detection after hybridization. Laser light is used to excite the fluorescent dyes and collect relative signal intensities. The hybridization intensity is represented by the amount of fluorescent emission, which gives an estimate of the relative amount of the different transcripts that are represented. mRNA from biological samples is reverse transcribed and labeled with one or more fluorescent dyes. After hybridization, fluorescence is measured separately and captured in two images. These are merged to produce a composite image, which goes through preprocessing before expression values are analyzed.

Microarrays have been widely applied in TGx studies and related predictive efforts [30,31,66,96–99]. DNA microarrays can capture relevant transcriptomic responses with concordant biological signals, independently from the platforms and laboratories [100,101]. A further advantage of microarray technology is that it is relatively mature, with numerous well-established commercial and open source data analysis tools [102,103]. Nonetheless, one of the limitations of this technology is the inability to detect unknown transcripts, as the probes are designed on the basis of known nucleotide sequences. Moreover, cross-hybridization events may occur when working with highly repetitive genomes, such as human's, resulting in an inaccurate observed expression value of the gene compared to the real expression [104]. From the TGx point of view, distinct sample handling and data preprocessing steps might limit the reproducibility of the results in different laboratory environments.

### 4.3. RNA-Sequencing

Although the advent of high-throughput hybridization-based technologies, such as DNA microarrays, significantly boosted the generation of large scale gene expression profiles, the recent advances in sequencing technologies further improved such capability. In fact, the so-called Next-Generation Sequencing (NGS) technologies are currently employed in multiple fields of the life sciences, including the toxicogenomics. In this context, the RNA-sequencing (RNA-Seq) allowed overcoming the technical limitations imposed by the hybridization-based techniques. For instance, the RNA-Seq allows the detection of gene expression with increased dynamic range, solving the problem of probes saturation for highly expressed transcripts [105]. Furthermore, it does not need a priori knowledge of the genomic sequence of the studied organism and it does not suffer from the above mentioned cross-hybridization events, especially in the analysis of complex genomes. As a consequence, RNA-Seq allows to perform a de novo transcript discovery in order to 1-identify unannotated transcripts, and 2-characterize new transcripts generated by alternative splicing [106]. Conversely, an appropriate analysis plan should be employed in order to avoid or mitigate certain biases that could occur during the data management and analysis. For instance, previous works

showed that standard normalization procedures can affect the sensitivity of the differential expression analysis, reflecting the behavior of a relatively small number of either high-count or ubiquitous genes [107,108]. RNA sequencing typically produces bigger and more complex data, which need longer time and more sophisticated analytical approaches, as compared for instance with DNA microarray experiments. Although TGx experiments are increasingly employing transcriptome profiling, the analysis pipelines are still far from being standardized. To date, a benchmark of the optimal analytic procedures in the transcriptome profiling in TGx experiments has not been formulated. More recently, the reduction of the costs for the analysis of a single transcriptome made possible the accomplishment of big scale studies, carried out by international programs, such as CMAP, TOX21, and LINCS1000 aimed at understanding the responses of hundreds of cell models upon treatment with several classes of chemicals/compounds.

#### 4.4. Single Cell RNA-Sequencing

Since the first single-cell RNA sequencing (scRNA-Seq) protocol was published in 2009 [109], there has been increasing interest in applying this technique in various areas of research, especially in cellular heterogeneity, such as in tumors [110], the immune system [111], the brain [112], and embryo development [113], which is often masked in the typical RNA-Seq. In TGx studies, this technique has been extensively used to analyze drug tolerance in cancer cells [114]. The scRNA-Seq, as its name indicates, analyses the transcriptome of an individual cell and is mainly used to resolve the transcriptional heterogeneity and to identify rare cell populations that would otherwise be hidden in bulk-cell analyses [115–117]. Currently, there are approximately 20 different scRNA-Seq protocols that have been reported to address various research purposes.

Full-length methods, such as SMART-seq, cover the whole transcriptome and thereby increase the number of mappable reads as well as the detected genes per cell. This is useful for analyzing genes that are expressed at low levels [118]. However, one major drawback of full-length methods is that they do not support multiplexing and thus require intensive cost and labor as each sample should have their own sequencing library. In contrast, 3'-end counting methods, such as Drop-seq, MARS-seq, and CEL-seq, only cover the 3'-end of the transcripts [119]. The transcripts, in this case, can be indexed with barcodes and unique molecular identifiers (UMIs) to allow multiplexing of samples and quantification of gene expression. However, as the reads are limited to the 3'-end of the transcripts, the overall number of detected genes per cell is lower than in full-length methods. In terms of single-cell isolation, all these technologies separate single cells into nanoliter reaction chambers. In the droplet-based methods, each cell is partitioned in a hydrogel droplet [120]. In microfluidics-based methods, cells are individually captured on a microfluidic chip, whereas in plate-based methods, individual cells are placed into wells on a 96-well plate by micropipetting or fluorescence-activated cell sorting [121]. Of all the scRNA-Seq sample preparation platforms, 10X Genomics' Chromium is one of the most widely used commercial solutions. It is a droplet-based 3'-end counting system. On the 10X Genomics' Chromium, each cell is captured in a water-in-oil emulsion droplet that contains a gel bead labeled with sequencing adapters, 16-nucleotide barcodes for cell identification, 10-nucleotide UMIs for transcript identification, and poly[T]primers for mRNA capture.

Although there are various scRNA-Seq methods available for different research questions, most of them adhere to a similar workflow. The first and most important step is the isolation of viable, single cells from the tissue of interest. Next, the isolated individual cells are lysed and the mRNA molecules are captured using the poly[T]-primed oligomers that bind to the poly[A] tail of the mRNA. Analysis of non-poly[A]-tailed mRNA will require more specialised protocols [122,123]. The complementary DNA (cDNA) is then synthesized by reverse transcription. Depending on each method, the reverse transcription primers can have additional nucleotide sequences (i.e., barcodes, UMIs, etc.). Subsequently, the cDNA is amplified to acquire an adequate amount for sequencing by either PCR, or in vitro transcription. Then, libraries are prepared by pooling and fragmenting the amplified cDNA.

Libraries are finally sequenced via NGS. The preprocessing, analysis, and interpretation of the data is a challenging research topic on its own and will be further discussed in later parts.

#### 4.5. High-Throughput Transcriptomics

High-throughput transcriptomics technologies dramatically increase the throughput of gene expression data generation. In particular, they allow a low-cost profiling of dose responses. These methods are very diverse and include microarray-based, RNA sequencing-based as well as other technologies. For example, the L1000 (Library of Integrated Network-Based Cellular Signatures 1000) platform that is used in the LINCS CMAP database uses flow cytometry and labeled beads [124]. It measures only approximately 1000 genes, but these carefully selected “landmark genes” can be used to predict the expression of nearly 11,000 genes. This illustrates the principle of redundancy and co-regulation within gene expression. The S1500+ platform designed by the US National Toxicology Program has 2753 probe sets or genes, including all the L1000 genes [125]. The larger number of genes is used to improve the accuracy of gene estimation and to measure directly many toxicologically relevant and pathway-associated genes. The S1500+ platform uses a TempO-Seq targeted RNA-Seq method [126]. Compared to the 10-20 million counts in a regular RNA-Seq, it only needs 0.5-3 million reads per profile. An example of an array-based high-throughput platform is the Affymetrix/ThermoFisher Scientific Clariom GO Screen assay that includes approximately 20,000 gene ontology-related genes in a 384-multiplex format. In general, the same or very similar methods can be used to analyze high-throughput and regular expression experiments. If the data is continuous or array-based, microarray methods and RNA-Seq methods with count data are recommended. Due to the larger volume of data, automated R/Bioconductor or other workflows need to be run [127]. Here, tools such as eUTOPIA (A solUTION for Omics data Preprocessing and Analysis) can be helpful in reducing the coding workload [128].

### 5. Publicly Available Datasets for Toxicogenomics

The scientific advancement in the toxicogenomics field is strongly dependent on the availability of high-quality TGx datasets. Thus, in the last years, many efforts have been made to produce good quality databases such as Chemical Effects in Biological Systems (CEBS) [129,130], Connectivity Map (CMAP) [131], LINCS 1000 [124], DrugMatrix [132], Open TG-GATEs[98], ArrayExpress [133], and Gene Expression Omnibus (GEO) [54,134].

CEBS is a comprehensive toxicogenomics resource compiled by the National Center for Toxicogenomics (NCT) within the National Institute of Environmental Health Science (NIEHS). It includes the study design, the timeline, the clinical chemistry, and histopathology findings, as well as microarray and proteomics data from in vivo and in vitro exposure [130]. CEBS freely available online at <http://tools.niehs.nih.gov/cebs3/ui/>.

The CMAP project started in 2006 with gene expression profiles of 164 small-molecule compounds and was later updated to “build 2”, containing gene expression profiles from three cultured human cells treated with 1309 bioactive small molecules, aimed at finding connections among small molecules sharing a mechanism of action, chemical and physiological processes, diseases, and drugs [131].

The L1000 Connectivity Map is a Library of Integrated Network-based Cellular Signatures (LINCS) and was created as a large-scale expansion of the original CMAP, in which almost 20,000 small-molecules have been profiled on up to 77 cell lines at 96 hours after the knockdown [124]. The experiments were performed by using the L1000 platform, which aims to capture the greatest amount of variation while measuring only a subset of 978 genes, named “landmark genes” [124]. This subset of genes was chosen since they are able to capture the greatest proportion of the variance in expression [124,135,136].

The DrugMatrix dataset contains gene expression response to compound treatments in rat tissues [132]. Approximately 600 different compounds have been profiled in up to seven different tissues of rats, representing over 3200 different drug–dose–time–tissue combinations. The dataset

also provides histopathological, hematologic and clinical chemistry data associated with compound treatments, allowing specific forms of toxicity to be investigated. The DrugMatrix dataset is available through the CEBS database using the CEBS Guided Search tools [130].

Open TG-GATEs is a TGx database that stores gene expression profiles and traditional toxicological data derived from in vivo (rat) and in vitro (primary rat hepatocytes, primary human hepatocytes) exposure to 170 compounds at multiple doses and time points [98]. The gene expression data for a test compound were derived from the administration of individual compounds at up to four dose levels and eight time-points (corresponding to four single-dose studies and four repeated-dose studies).

ArrayExpress repository stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community [133].

Gene Expression Omnibus (GEO) is a public repository for high-throughput microarray and next-generation sequencing data sets, carried out by the research community. The resource supports archiving of raw data, processed data and metadata which are indexed, cross-linked, and searchable. All data is freely available for download in a variety of formats [54,134].

As introduced above, usable, high-quality databases for TGx data are already available. In order to deal with the growing number of large datasets and to be able to fully utilize the potential of data sharing in toxicology, more comprehensive and integrated infrastructure should be still created to handle the data according to the FAIR principles. In practice, this requires proper metadata handling, unique identifiers, trusted repositories for easy access and data handling, clear usage licenses, and opportunities for interoperability [57].

## 6. Regulatory Aspects

The concept of systems biology has been known already for decades, but the emergence of high-throughput methodologies and omics approaches rapidly expanded in the beginning of the 21st century. Since then, transcriptional profiling has become a well recognized field of bioscience. Due to reducing costs and well-established infrastructure, today omics approaches are available for a wide research community. The ample amount of available omics-based exposure data enables researchers to combine and interpret complex and large datasets in a meaningful way. Additionally, in Europe, directives on animal testing limitations and marketing restrictions for example in case of cosmetics, encourage the development of alternative testing methods.

Toxicogenomics is recognized as a promising approach to characterize biological activity caused by exposure to different stressors. For example, in the ECHA's Guidance "Information Requirements and Chemical Safety Assessment", TGx-related technologies and toxicity testing are discussed in the context of in vitro data and Adverse Outcome Pathways (AOPs). The guidance states that the value of the in vitro models in toxicity testing could be increased by incorporating molecular based markers through the application of proteomic and toxicogenomic approaches [137]. Also the European Food Safety Authority (EFSA) has recognized the potential of TGx [138]. Moreover, the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulation emphasizes the utilization of TGx approaches in hazard assessment:

*"The Commission, Member States, industry and other stakeholders should continue to contribute to the promotion of alternative test methods on an international and national level including computer supported methodologies, in vitro methodologies, as appropriate, those based on toxicogenomics, and other relevant methodologies. The Community's strategy to promote alternative test methods is a priority and the Commission should ensure that within its future Research Framework Programmes and initiatives such as the Community Action Plan on the Protection and Welfare of Animals 2006 to 2010 this remains a priority topic. Participation of stakeholders and initiatives involving all interested parties should be sought [139]."*

Although ongoing effort is being made in Europe to bring TGx approaches as part of toxicity and hazard assessment, no formal guidance yet exists. Nonetheless, there are several ongoing projects to produce frameworks for omics data-based risk assessment: In the European Union, the European Centre for the Validation of Alternative Methods (ECVAM), an integral part of the Joint Research Centre (JRC) and Extended Advisory Group for Molecular Screening and Toxicogenomics (EAGMST) in OECD are working towards TGx approaches for example by using the concept of AOP [140,141]. OECD supports the development and improvement of AOPs with TGx approaches, by filling the data gaps and supporting the available mechanistic toxicity data. In 2012, the OECD established a new programme on the development of Adverse Outcome Pathways. Since the beginning of the programme in 2012, OECD has released 16 publications on AOP, from which seven were published in the year 2019 [142]. In addition, the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) has together with ECVAM developed a Transcriptomics Reporting Framework for omics data processing and analysis [140]. The framework does not dictate the methodology to be used, but rather lays the basis for what should be reported when performing bioinformatics with the aim of supporting risk assessment. Thus, the current three-article series provides a basis for refining and extending the framework with guidelines based on the suggested state-of-the-art methods and approaches. Importantly, we suggest here a shift in thinking regarding the current focus on differentially expressed genes in the framework, as will be discussed in detail in Part II of this article series. Analysis of differentially expressed genes is often coupled to arbitrary *p*-value limits and fold-change thresholds, which leads to one of the largest sources of bias within TGx, i.e., the use of high irrelevant doses.

In the U.S., The National Toxicology Program (NTP), an interagency program composed by the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH), the National Center for Toxicological Research (NCTR) of the Food and Drug Administration (FDA), and The National Institute for Occupational Safety and Health (NIOSH) of the Centers for Disease Control and Prevention (CDC), has established a toxicogenomics testing program consisting of 5-day rodent chemical exposure studies (<https://ntp.niehs.nih.gov/whatwestudy/testpgm/toxicogenomics/index.html>—Retrieved 3 March 2020). All the TGx data generated by this collaboration are stored in the CEBS database and publicly available to the community. The agencies across continents are also collaborating in scientific advisory boards towards legitimate validation and international harmonization of testing strategies based on alternative test methods including TGx [143].

Data for adverse effects and toxicity are classically retrieved from epidemiological studies of humans or more commonly from experimental systems including *in vitro* cell culture models and *in vivo* animal models. The data generated by traditional toxicity tests are the cornerstone for human health risk assessment, determined from the identified changes and phenotypic effects. Nonetheless, new chemicals, and especially the new era of ENMs, require extensive testing and screening, for which conventional toxicity testing is not optimal. This problem has been widely recognized, and efforts for integrated approaches together with high-throughput methods and 3R principles aim to identify early molecular markers and pathways of toxicity. Instead of traditional endpoints, TGx provides a possibility to test the entire genome of a cell or tissue, and thus, reveal earlier effects leading to toxic outcomes [40].

Toxicogenomic approaches are supported in EU-programmes such as Framework Programme 7 (FP7) and Horizon 2020 (H2020), but to the best of our knowledge, no specific legislation exists for TGx approaches. Currently, toxicogenomics approaches can be used to complement more conventional methods. Required testing methods depend on the chemical and its composition, and in general, requirements for pharmaceuticals are more demanding than for industrial chemicals and require extended testing related to efficacy, safety, and quality control. In the European Commission's Technical Guidance Document on Risk Assessment, the use of methods such as quantitative structure–activity relationship (QSAR) in human health risk assessment is recognized as a potentially contributing factor

only for certain endpoints [144]. Moreover, the guidance states that modeled data should only be used for risk assessment purposes where it is supported by other strands of evidence [145].

Understandably more investigations are needed to further embed TGx-based evidence in regulatory decision making. International unification and harmonization of methods and data generation and sharing is extremely important in developing toxicogenomics approaches for hazard assessment [146]. Furthermore, in addition to regulatory implementation of TGx, targeting the industry with the methods as support for safety assessment at early stages of technological innovation and development (aiming for so called safe by design strategies) may support uptake into the generally more slow-adapting (as compared to industrial processes) regulatory processes (<https://doi.org/10.1002/sml.201904749>).

## 7. Conclusions

In this first part of the review, we consider the crucial aspects needed to perform unbiased and reliable toxicogenomics analysis. We discuss the necessary steps for designing the experiments, deciding the sample groups, timepoints and doses. Furthermore, we discuss the public databases and transcriptomics related technologies. TGx has a great potential to overcome the limitations of traditional hazard assessment and to reduce the amount of animal experimentations. Additionally, when repeatable and reliable transcriptomics results are achieved, TGx approach allows in vitro methods to be utilized in regulatory toxicology. This includes better utilization of known methods such as AOPs and dose-response assessment, as well as better characterization of the biological processes related to QSAR models.

Reliable TGx data are the starting point to elucidate compounds MOA and predicting toxicity. Even though public TGx datasets can be found, toxicogenomics advancement would benefit from the availability of more comprehensive gene expression data and toxic endpoint annotations. Open data access, open protocols and publicly available meta-data annotations are also important to make data findable, accessible, interoperable and reusable (FAIR Data Principles) [57].

In conclusion, TGx methods are already favored by a wide community of researchers and pharmaceutical industries to characterize the molecular mechanisms underlying toxicity by the means of AOPs and to further predict compound toxicity. Furthermore, the field is evolving and currently working to reduce animal tests by replacing them with in vitro experiments. Developing TGx towards regulatory decision making depends on multiple technical and conceptual solutions that are currently coming together [6,147]. It is difficult to say when these methods will replace animal tests, but they can already be used as convincing supporting evidence. Altogether, TGx will have a strong impact on in vitro-based workflows in risk-assessment, on drug development strategies and on clinical decision making.

**Author Contributions:** Conceptualization, D.G., P.A.S.K., and A.S. Methodology, P.A.S.K. and A.S. Investigation, P.A.S.K. and A.S. Writing—original draft preparation, P.A.S.K., A.S., A.F., P.K. P.N., I.L., M.K.H., J.-S.C., K.J., and N.S., Writing—review and editing, P.A.S.K., A.S., A.F., P.K. P.N., I.L., M.K.H., J.-S.C., K.J., N.S., G.M., L.C., M.F., H.S., A.A., T.-H.Y., M.G., R.G., T.P., and D.G. Visualization, P.A.S.K., Supervision, D.G. Funding acquisition, D.G., A.A., R.G., H.S., T.-H.Y., T.P., and M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Academy of Finland [grant number 322761], the EU H2020 NanoSolveIT project [grant number 814572] and the H2020 NanoinformaTIX [grant number 814426].

**Acknowledgments:** The authors would like to thank David Winkler for critical comments to the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AOP	adverse outcome pathways
BMD	benchmark dose
CDC	Centers for Disease Control and Prevention
CEBS	Chemical Effects in Biological Systems
cDNA	complementary DNA
CMAP	Connectivity Map
EAGMST	Extended Advisory Group for Molecular Screening and Toxicogenomics
ECHA	European chemicals agency
ECVAM	European Centre for the Validation of Alternative Methods
EFSA	European Food Safety Authority
eUTOPIA	A solUTion for Omics data PreprocessIng and Analysis
FAIR	Findable, Accessible, Interoperable and Reusable Data Principle
FDA	Food and Drug Administration
GEO	Gene Expression Omnibus
GO	gene ontology
GTE <sub>x</sub>	Genotype-Tissue Expression
JRC	Joint Research Centre
L1000	Library of Integrated Network-Based Cellular Signatures 1000
MIAME	Minimum Information About a Microarray Experiment
MINSEQE	Minimal Information about a high throughput SEQuencing Experiment
MOA	mechanism of action
NCT	National Center for Toxicogenomics
NCTR	National Center for Toxicological Research
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	The National Institute for Occupational Safety and Health
NTP	The National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
TG-GATE	Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System
PCR	polymerase chain reaction
POD	point of departure
QSAR	quantitative structure activity relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RNA-Seq	RNA sequencing
scRNA-Seq	single-cell RNA sequencing
TG <sub>x</sub>	toxicogenomics
3R	Replacement, Reduction and Refinement in animal testing

## References

1. EU. Regulation (EC) No 1223/2009 of the European Parliament and of the Council. *Off. J. Eur. Union L* **2009**, *342*, 59.
2. Nel, A.; Xia, T.; Mädler, L.; Li, N. Toxic potential of materials at the nanolevel. *Science* **2006**, *311*, 622–627. [[CrossRef](#)] [[PubMed](#)]
3. Poland, C.A.; Duffin, R.; Kinloch, I.; Maynard, A.; Wallace, W.A.; Seaton, A.; Stone, V.; Brown, S.; MacNee, W.; Donaldson, K. Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study. *Nat. Nanotechnol.* **2008**, *3*, 423. [[CrossRef](#)] [[PubMed](#)]
4. Sweeney, S.; Berhanu, D.; Misra, S.K.; Thorley, A.J.; Valsami-Jones, E.; Tetley, T.D. Multi-walled carbon nanotube length as a critical determinant of bioreactivity with primary human pulmonary alveolar cells. *Carbon* **2014**, *78*, 26–37. [[CrossRef](#)] [[PubMed](#)]

5. Kettler, K.; Veltman, K.; van de Meent, D.; van Wezel, A.; Hendriks, A.J. Cellular uptake of nanoparticles as determined by particle properties, experimental conditions, and cell type. *Environ. Toxicol. Chem.* **2014**, *33*, 481–492. [[CrossRef](#)]
6. Liu, Z.; Huang, R.; Roberts, R.; Tong, W. Toxicogenomics: A 2020 Vision. *Trends Pharmacol. Sci.* **2019**, *40*, 92–103. [[CrossRef](#)]
7. Alexander-Dann, B.; Pruteanu, L.L.; Oerton, E.; Sharma, N.; Berindan-Neagoe, I.; Módos, D.; Bender, A. Developments in toxicogenomics: Understanding and predicting compound-induced toxicity from gene expression data. *Mol. Omics* **2018**, *14*, 218–236. [[CrossRef](#)]
8. Grafström, R.C.; Nymark, P.; Hongisto, V.; Spjuth, O.; Ceder, R.; Willighagen, E.; Hardy, B.; Kaski, S.; Kohonen, P. Toward the replacement of animal experiments through the bioinformatics-driven analysis of ‘omics’ data from human cell cultures. *Altern. Lab. Anim.* **2015**, *43*, 325–332. [[CrossRef](#)]
9. Noordzij, M.; Dekker, F.W.; Zoccali, C.; Jager, K.J. Sample size calculations. *Nephron Clin. Pract.* **2011**, *118*, c319–c323. [[CrossRef](#)]
10. Li, C.I.; Samuels, D.C.; Zhao, Y.Y.; Shyr, Y.; Guo, Y. Power and sample size calculations for high-throughput sequencing-based experiments. *Brief. Bioinform.* **2018**, *19*, 1247–1255. [[CrossRef](#)]
11. Son, D.S.; Lee, D.; Lee, K.; Jung, S.H.; Ahn, T.; Lee, E.; Sohn, I.; Chung, J.; Park, W.; Huh, N.; et al. Practical approach to determine sample size for building logistic prediction models using high-throughput data. *J. Biomed. Inform.* **2015**, *53*, 355–362. [[CrossRef](#)] [[PubMed](#)]
12. Alin, A. Minitab. *Wiley Interdisc. Rev. Comput. Stat.* **2010**, *2*, 723–727. [[CrossRef](#)]
13. Tibshirani, R. A simple method for assessing sample sizes in microarray experiments. *BMC Bioinform.* **2006**, *7*, 106. [[CrossRef](#)] [[PubMed](#)]
14. Lin, W.J.; Hsueh, H.M.; Chen, J.J. Power and sample size estimation in microarray studies. *BMC Bioinform.* **2010**, *11*, 48. [[CrossRef](#)]
15. Liu, Y.; Zhou, J.; White, K.P. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics* **2014**, *30*, 301–304. [[CrossRef](#)]
16. Frattini, A.; Fabbri, M.; Valli, R.; De Paoli, E.; Montalbano, G.; Gribaldo, L.; Pasquali, F.; Maserati, E. High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci. Rep.* **2015**, *5*, 15377. [[CrossRef](#)]
17. Bajak, E.; Fabbri, M.; Ponti, J.; Gioria, S.; Ojea-Jiménez, I.; Collotta, A.; Mariani, V.; Gilliland, D.; Rossi, F.; Gribaldo, L. Changes in Caco-2 cells transcriptome profiles upon exposure to gold nanoparticles. *Toxicol. Lett.* **2015**, *233*, 187–199. [[CrossRef](#)]
18. Poulsen, S.S.; Saber, A.T.; Williams, A.; Andersen, O.; Købler, C.; Atluri, R.; Pozzebon, M.E.; Mucelli, S.P.; Simion, M.; Rickerby, D.; et al. MWCNTs of different physicochemical properties cause similar inflammatory responses, but differences in transcriptional and histological markers of fibrosis in mouse lungs. *Toxicol. Appl. Pharmacol.* **2015**, *284*, 16–32. [[CrossRef](#)]
19. Bourdon, J.A.; Halappanavar, S.; Saber, A.T.; Jacobsen, N.R.; Williams, A.; Wallin, H.; Vogel, U.; Yauk, C.L. Hepatic and pulmonary toxicogenomic profiles in mice intratracheally instilled with carbon black nanoparticles reveal pulmonary inflammation, acute phase response, and alterations in lipid homeostasis. *Toxicol. Sci.* **2012**, *127*, 474–484. [[CrossRef](#)]
20. Dymacek, J.; Snyder-Talkington, B.N.; Porter, D.W.; Mercer, R.R.; Wolfarth, M.G.; Castranova, V.; Qian, Y.; Guo, N.L. mRNA and miRNA regulatory networks reflective of multi-walled carbon nanotube-induced lung inflammatory and fibrotic pathologies in mice. *Toxicol. Sci.* **2015**, *144*, 51–64. [[CrossRef](#)]
21. Ong, K.J.; MacCormack, T.J.; Clark, R.J.; Ede, J.D.; Ortega, V.A.; Felix, L.C.; Dang, M.K.; Ma, G.; Fenniri, H.; Veinot, J.G.; et al. Widespread nanoparticle-assay interference: Implications for nanotoxicity testing. *PLoS ONE* **2014**, *9*, e90650. [[CrossRef](#)] [[PubMed](#)]
22. Rydman, E.M.; Ilves, M.; Vanhala, E.; Vippola, M.; Lehto, M.; Kinaret, P.A.; Pylkänen, L.; Happonen, M.; Hirvonen, M.R.; Greco, D.; et al. A single aspiration of rod-like carbon nanotubes induces asbestos-like pulmonary inflammation mediated in part by the IL-1 receptor. *Toxicol. Sci.* **2015**, *147*, 140–155. [[CrossRef](#)] [[PubMed](#)]
23. Palomäki, J.; Sund, J.; Vippola, M.; Kinaret, P.; Greco, D.; Savolainen, K.; Puustinen, A.; Alenius, H. A secretomics analysis reveals major differences in the macrophage responses towards different types of carbon nanotubes. *Nanotoxicology* **2015**, *9*, 719–728. [[CrossRef](#)] [[PubMed](#)]



24. Nymark, P.; Wijshoff, P.; Cavill, R.; van Herwijnen, M.; Coonen, M.L.; Claessen, S.; Catalán, J.; Norppa, H.; Kleinjans, J.C.; Briede, J.J. Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells. *Nanotoxicology* **2015**, *9*, 624–635. [[CrossRef](#)]
25. Council, N.R. *Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment*; National Academies Press: Washington, DC, USA, 2007.
26. Lamon, L.; Asturiol, D.; Vilchez, A.; Cabellos, J.; Damásio, J.; Janer, G.; Richarz, A.; Worth, A. Physiologically based mathematical models of nanomaterials for regulatory toxicology: A review. *Comput. Toxicol.* **2019**, *9*, 133–142. [[CrossRef](#)]
27. Warheit, D.B. Hazard and risk assessment strategies for nanoparticle exposures: How far have we come in the past 10 years? *F1000Research* **2018**, *7*. [[CrossRef](#)]
28. Poulsen, S.S.; Jackson, P.; Kling, K.; Knudsen, K.B.; Skaug, V.; Kyjovska, Z.O.; Thomsen, B.L.; Clausen, P.A.; Atluri, R.; Berthing, T.; et al. Multi-walled carbon nanotube physicochemical properties predict pulmonary inflammation and genotoxicity. *Nanotoxicology* **2016**, *10*, 1263–1275. [[CrossRef](#)]
29. Wallin, H.; Kyjovska, Z.O.; Poulsen, S.S.; Jacobsen, N.R.; Saber, A.T.; Bengtson, S.; Jackson, P.; Vogel, U. Surface modification does not influence the genotoxic and inflammatory effects of TiO<sub>2</sub> nanoparticles after pulmonary exposure by instillation in mice. *Mutagenesis* **2017**, *32*, 47–57. [[CrossRef](#)]
30. Kinaret, P.; Ilves, M.; Fortino, V.; Rydman, E.; Karisola, P.; Lahde, A.; Koivisto, J.; Jokiniemi, J.; Wolff, H.; Savolainen, K.; et al. Inhalation and oropharyngeal aspiration exposure to rod-like carbon nanotubes induce similar airway inflammation and biological responses in mouse lungs. *ACS Nano* **2017**, *11*, 291–303. [[CrossRef](#)]
31. Scala, G.; Kinaret, P.; Marwah, V.; Sund, J.; Fortino, V.; Greco, D. Multi-omics analysis of ten carbon nanomaterials effects highlights cell type specific patterns of molecular regulation and adaptation. *NanoImpact* **2018**, *11*, 99–108. [[CrossRef](#)]
32. Phillips, J.R.; Svoboda, D.L.; Tandon, A.; Patel, S.; Sedykh, A.; Mav, D.; Kuo, B.; Yauk, C.L.; Yang, L.; Thomas, R.S.; et al. BMDExpress 2: Enhanced transcriptomic dose-response analysis workflow. *Bioinformatics* **2019**, *35*, 1780–1782. [[CrossRef](#)] [[PubMed](#)]
33. Farmahin, R.; Williams, A.; Kuo, B.; Chepelev, N.L.; Thomas, R.S.; Barton-Maclaren, T.S.; Curran, I.H.; Nong, A.; Wade, M.G.; Yauk, C.L. Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Arch. Toxicol.* **2017**, *91*, 2045–2065. [[CrossRef](#)] [[PubMed](#)]
34. Jackson, A.F.; Williams, A.; Recio, L.; Waters, M.D.; Lambert, I.B.; Yauk, C.L. Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicol. Appl. Pharmacol.* **2014**, *274*, 63–77. [[CrossRef](#)]
35. Moffat, I.; Chepelev, N.L.; Labib, S.; Bourdon-Lacombe, J.; Kuo, B.; Buick, J.K.; Lemieux, F.; Williams, A.; Halappanavar, S.; Malik, A.I.; et al. Comparison of toxicogenomics and traditional approaches to inform mode of action and points of departure in human health risk assessment of benzo [a] pyrene in drinking water. *Crit. Rev. Toxicol.* **2015**, *45*, 1–43. [[CrossRef](#)] [[PubMed](#)]
36. Webster, A.F.; Chepelev, N.; Gagné, R.; Kuo, B.; Recio, L.; Williams, A.; Yauk, C.L. Impact of genomics platform and statistical filtering on transcriptional benchmark doses (BMD) and multiple approaches for selection of chemical point of departure (PoD). *PLoS ONE* **2015**, *10*, e0136764. [[CrossRef](#)]
37. Labib, S.; Williams, A.; Yauk, C.L.; Nikota, J.K.; Wallin, H.; Vogel, U.; Halappanavar, S. Nano-risk Science: Application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Part. Fibre Toxicol.* **2015**, *13*, 15. [[CrossRef](#)]
38. Zhang, Q.; Li, J.; Middleton, A.; Bhattacharya, S.; Conolly, R.B. Bridging the data gap from in vitro toxicity testing to chemical safety assessment through computational modeling. *Front. Public Health* **2018**, *6*, 261. [[CrossRef](#)]
39. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; et al. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935. [[CrossRef](#)]
40. Yauk, C.L.; Cheung, C.; Barton-Maclaren, T.S.; Boucher, S.; Bourdon-Lacombe, J.; Chauhan, V.; Gagné, M.; Gillespie, Z.; Halappanavar, S.; Honeyman, M.; et al. Toxicogenomic applications in risk assessment at Health Canada. *Curr. Opin. Toxicol.* **2019**, *18*, 34–45. [[CrossRef](#)]

41. Ciallella, H.L.; Zhu, H. Advancing computational toxicology in the big data era by artificial intelligence: Data-driven and mechanism-driven modeling for chemical toxicity. *Chem. Res. Toxicol.* **2019**, *32*, 536–547. [[CrossRef](#)]
42. Vo, A.H.; Van Vleet, T.R.; Gupta, R.R.; Liguori, M.J.; Rao, M.S. An Overview of Machine Learning and Big Data for Drug Toxicity Evaluation. *Chem. Res. Toxicol.* **2019**, *33*, 20–37. [[CrossRef](#)] [[PubMed](#)]
43. Nuwaysir, E.F.; Bittner, M.; Trent, J.; Barrett, J.C.; Afshari, C.A. Microarrays and toxicology: The advent of toxicogenomics. *Mol. Carcinog.* **1999**, *24*, 153–159. [[CrossRef](#)]
44. Poulsen, S.S.; Jacobsen, N.R.; Labib, S.; Wu, D.; Husain, M.; Williams, A.; Bøgelund, J.P.; Andersen, O.; Købler, C.; Mølhav, K.; et al. Transcriptomic analysis reveals novel mechanistic insight into murine biological responses to multi-walled carbon nanotubes in lungs and cultured lung epithelial cells. *PLoS ONE* **2013**, *8*, e80452. [[CrossRef](#)] [[PubMed](#)]
45. Schmid, O.; Cassee, F.R. On the Pivotal Role of Dose for Particle Toxicology and Risk Assessment: Exposure Is a Poor Surrogate for Delivered Dose. *Part Fibre Toxicol.* **2017**, *14*, 52
46. Organisation for Economic Co-operation and Development. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q) SAR] Models*; OECD Publishing: Paris, France, 2014.
47. Aslantürk, Ö.S. *In Vitro Cytotoxicity and Cell Viability Assays: Principles, Advantages, and Disadvantages*; InTechOpen: London, UK, 2018; Volume 2, Chapter 1.
48. Drasler, B.; Sayre, P.; Steinhäuser, K.G.; Petri-Fink, A.; Rothen-Rutishauser, B. In vitro approaches to assess the hazard of nanomaterials. *NanoImpact* **2017**, *8*, 99–116. [[CrossRef](#)]
49. Schroeder, A.; Mueller, O.; Stocker, S.; Salowsky, R.; Leiber, M.; Gassmann, M.; Lightfoot, S.; Menzel, W.; Granzow, M.; Ragg, T. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **2006**, *7*, 3. [[CrossRef](#)]
50. Scherer, A. *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*; John Wiley & Sons: London, UK, 2009; Volume 868.
51. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739. [[CrossRef](#)]
52. Verdugo, R.A.; Deschepper, C.F.; Muñoz, G.; Pomp, D.; Churchill, G.A. Importance of randomization in microarray experimental designs with Illumina platforms. *Nucleic Acids Res.* **2009**, *37*, 5610–5618. [[CrossRef](#)]
53. Naidoo, S.; Denby, K.J.; Berger, D.K. Microarray experiments: Considerations for experimental design. *S. Afr. J. Sci.* **2005**, *101*, 347–354.
54. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2012**, *41*, D991–D995. [[CrossRef](#)]
55. Parkinson, H.; Kapushesky, M.; Shojatalab, M.; Abeygunawardena, N.; Coulson, R.; Farne, A.; Holloway, E.; Kolesnykov, N.; Lilja, P.; Lukk, M.; et al. ArrayExpress—A public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* **2007**, *35*, D747–D750. [[CrossRef](#)] [[PubMed](#)]
56. Brazma, A.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, C.A.; Causton, H.C.; et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **2001**, *29*, 365. [[CrossRef](#)] [[PubMed](#)]
57. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*. [[CrossRef](#)]
58. Williams-Devane, C.R.; Wolf, M.A.; Richard, A.M. Toward a public toxicogenomics capability for supporting predictive toxicology: Survey of current resources and chemical indexing of experiments in GEO and ArrayExpress. *Toxicol. Sci.* **2009**, *109*, 358–371. [[CrossRef](#)] [[PubMed](#)]
59. Sayes, C.M.; Reed, K.L.; Warheit, D.B. Assessing toxicity of fine and nanoparticles: Comparing in vitro measurements to in vivo pulmonary toxicity profiles. *Toxicological sciences* **2007**, *97*, 163–180. [[CrossRef](#)] [[PubMed](#)]
60. Seagrave, J.; McDonald, J.D.; Mauderly, J.L. In vitro versus in vivo exposure to combustion emissions. *Exp. Toxicol. Pathol.* **2005**, *57*, 233–238. [[CrossRef](#)]

61. Sayes, C.M.; Reed, K.L.; Subramoney, S.; Abrams, L.; Warheit, D.B. Can in vitro assays substitute for in vivo studies in assessing the pulmonary hazards of fine and nanoscale materials? *J. Nanopart. Res.* **2009**, *11*, 421–431. [[CrossRef](#)]
62. Warheit, D.; Sayes, C.; Reed, K. Nanoscale and fine zinc oxide particles: Can in vitro assays accurately forecast lung hazards following inhalation exposures? *Environ. Sci. Technol.* **2009**, *43*, 7939–7945. [[CrossRef](#)]
63. Sayes, C.M.; Marchione, A.A.; Reed, K.L.; Warheit, D.B. Comparative pulmonary toxicity assessments of C60 water suspensions in rats: Few differences in fullerene toxicity in vivo in contrast to in vitro profiles. *Nano Lett.* **2007**, *7*, 2399–2406. [[CrossRef](#)]
64. Johnston, H.J.; Hutchison, G.R.; Christensen, F.M.; Peters, S.; Hankin, S.; Aschberger, K.; Stone, V. A critical review of the biological mechanisms underlying the in vivo and in vitro toxicity of carbon nanotubes: The contribution of physico-chemical characteristics. *Nanotoxicology* **2010**, *4*, 207–246. [[CrossRef](#)]
65. Dhawan, A.; Sharma, V. Toxicity assessment of nanomaterials: Methods and challenges. *Anal. Bioanal. Chem.* **2010**, *398*, 589–605. [[CrossRef](#)] [[PubMed](#)]
66. Kinaret, P.; Marwah, V.; Fortino, V.; Ilves, M.; Wolff, H.; Ruokolainen, L.; Auvinen, P.; Savolainen, K.; Alenius, H.; Greco, D. Network analysis reveals similar transcriptomic responses to intrinsic properties of carbon nanomaterials in vitro and in vivo. *ACS Nano* **2017**, *11*, 3786–3796. [[CrossRef](#)] [[PubMed](#)]
67. Marwah, V.S.; Kinaret, P.A.S.; Serra, A.; Scala, G.; Lauerma, A.; Fortino, V.; Greco, D. Inform: Inference of network response modules. *Bioinformatics* **2018**, *34*, 2136–2138. [[CrossRef](#)]
68. Kroll, A.; Pillukat, M.H.; Hahn, D.; Schneckeburger, J. Current in vitro methods in nanoparticle risk assessment: Limitations and challenges. *Eur. J. Pharm. Biopharm.* **2009**, *72*, 370–377. [[CrossRef](#)] [[PubMed](#)]
69. Monteiro-Riviere, N.; Inman, A.; Zhang, L. Limitations and relative utility of screening assays to assess engineered nanoparticle toxicity in a human cell line. *Toxicol. Appl. Pharmacol.* **2009**, *234*, 222–235. [[CrossRef](#)]
70. Pedersen, E.; Fant, K. *Guidance Document on Good In Vitro Method Practices (GIVIMP): Series on Testing and Assessment No. 286*; OECD Publishing: Paris, France, 2018.
71. Fei, L.; Perrett, S. Effect of nanoparticles on protein folding and fibrillogenesis. *Int. J. Mol. Sci.* **2009**, *10*, 646–655. [[CrossRef](#)]
72. Asuri, P.; Bale, S.S.; Karajanagi, S.S.; Kane, R.S. The protein–nanomaterial interface. *Curr. Opin. Biotechnol.* **2006**, *17*, 562–568. [[CrossRef](#)]
73. Kane, R.S.; Stroock, A.D. Nanobiotechnology: Protein–nanomaterial interactions. *Biotechnol. Prog.* **2007**, *23*, 316–319. [[CrossRef](#)]
74. Lord, M.S.; Foss, M.; Besenbacher, F. Influence of nanoscale surface topography on protein adsorption and cellular response. *Nano Today* **2010**, *5*, 66–78. [[CrossRef](#)]
75. Lynch, I.; Dawson, K.A. Protein–nanoparticle interactions. *Nano Today* **2008**, *3*, 40–47. [[CrossRef](#)]
76. Stueker, O.; Ortega, V.A.; Goss, G.G.; Stepanova, M. Understanding interactions of functionalized nanoparticles with proteins: A case study on lactate dehydrogenase. *Small* **2014**, *10*, 2006–2021. [[CrossRef](#)] [[PubMed](#)]
77. Casey, A.; Herzog, E.; Lyng, F.; Byrne, H.; Chambers, G.; Davoren, M. Single walled carbon nanotubes induce indirect cytotoxicity by medium depletion in A549 lung cells. *Toxicol. Lett.* **2008**, *179*, 78–84. [[CrossRef](#)] [[PubMed](#)]
78. Hedderman, T.G.; Keogh, S.M.; Chambers, G.; Byrne, H.J. Solubilization of SWNTs with organic dye molecules. *J. Phys. Chem. B* **2004**, *108*, 18860–18865. [[CrossRef](#)]
79. Ramakrishna, G.; Ghosh, H.N. Emission from the charge transfer state of xanthene dye-sensitized TiO<sub>2</sub> nanoparticles: A new approach to determining back electron transfer rate and verifying the marcus inverted regime. *J. Phys. Chem. B* **2001**, *105*, 7000–7008. [[CrossRef](#)]
80. Fröhlich, E. Cellular targets and mechanisms in the cytotoxic action of non-biodegradable engineered nanoparticles. *Curr. Drug Metab.* **2013**, *14*, 976–988. [[CrossRef](#)]
81. Fröhlich, E. Role of omics techniques in the toxicity testing of nanoparticles. *J. Nanobiotechnol.* **2017**, *15*, 84. [[CrossRef](#)]
82. Fröhlich, E.; Meindl, C.; Pieber, T.R. Important issues in the cytotoxicity screening of nano-sized materials. *EURO-NanoTox-Lett.* **2010**, *2*, 1–5.
83. Pisani, C.; Gaillard, J.C.; Nouvel, V.; Odorico, M.; Armengaud, J.; Prat, O. High-throughput, quantitative assessment of the effects of low-dose silica nanoparticles on lung cells: Grasping complex toxicity with a great depth of field. *BMC Genom.* **2015**, *16*, 315. [[CrossRef](#)]

84. Foth, H. Abstracts of the 46th Congress of the European Societies of Toxicology (EUROTOX2009). Dresden, Germany. 13–16 September 2009. *Toxicol. Lett.* **2009**, *189*, S1–S2. [[CrossRef](#)] [[PubMed](#)]
85. Hongisto, V.; Nymark, P.; Kohonen, J.; Hattara, J.; Grafstrom, R. *High-Throughput Hazard-Based Scoring, Ranking and Grouping of Engineered Nanomaterials*; Toxicology Letters; Elsevier: Amsterdam, The Netherlands, 2019; Volume 314, pp. S202–S203.
86. Nymark, P. *Grouping of Representative Nanomaterials Is Efficiently Executed By Combining High-Throughput-Generated Biological Data With Physicochemical Data*; Toxicology Letters, Elsevier: Amsterdam, The Netherlands, 2019, Volume 314, pp. 1–310.
87. Grzincic, E.; Yang, J.; Drnevich, J.; Falagan-Lotsch, P.; Murphy, C. Global transcriptomic analysis of model human cell lines exposed to surface-modified gold nanoparticles: The effect of surface chemistry. *Nanoscale* **2015**, *7*, 1349–1362. [[CrossRef](#)]
88. Li, K.; Zhao, X.; Hammer, K.B.; Du, S.; Chen, Y. Nanoparticles inhibit DNA replication by binding to DNA: Modeling and experimental validation. *ACS Nano* **2013**, *7*, 9664–9674. [[CrossRef](#)] [[PubMed](#)]
89. Humes, S.T.; Hentschel, S.; Lavelle, C.M.; Smith, L.C.; Lednický, J.A.; Saleh, N.B.; Sabo-Attwood, T. Overcoming qRT-PCR interference by select carbon nanotubes in assessments of gene expression. *BioTechniques* **2017**, *63*, 81–84. [[CrossRef](#)] [[PubMed](#)]
90. Gao, C.H.; Mortimer, M.; Zhang, M.; Holden, P.A.; Cai, P.; Wu, S.; Xin, Y.; Wu, Y.; Huang, Q. Impact of metal oxide nanoparticles on in vitro DNA amplification. *PeerJ* **2019**, *7*, e7228. [[CrossRef](#)] [[PubMed](#)]
91. Sanabria, N.M.; Gulumian, M. The presence of residual gold nanoparticles in samples interferes with the RT-qPCR assay used for gene expression profiling. *J. Nanobiotechnol.* **2017**, *15*, 72. [[CrossRef](#)]
92. Zhang, W.; Lai, E.P. Fluorescence detection of Zinc oxide nanoparticles in water contamination analysis based on surface reactivity with porphyrin. *AIMS Environ. Sci.* **2018**, *5*, 67. [[CrossRef](#)]
93. Sanabria, N.M.; Vetten, M.; Andraos, C.; Boodhia, K.; Gulumian, M. Gold nanoparticle interference study during the isolation, quantification, purity and integrity analysis of RNA. *PLoS ONE* **2014**, *9*, e114123. [[CrossRef](#)]
94. Carmichael, H.; Nelson, L.; Russell, R.; Chandra, V.; Lyon, A.; Cochran, K. Effect of prostaglandin 15 (R) 15 methyl-E2 methyl ester on aspirin and taurocholic acid-induced gastric mucosal haemorrhage in rats. *Gut* **1976**, *17*, 33–36. [[CrossRef](#)]
95. Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**, *270*, 467–470. [[CrossRef](#)]
96. Yasokawa, D.; Iwahashi, H. Toxicogenomics using yeast DNA microarrays. *J. Biosci. Bioeng.* **2010**, *110*, 511–522. [[CrossRef](#)]
97. Lettieri, T. Recent applications of DNA microarray technology to toxicology and ecotoxicology. *Environ. Health Perspect.* **2006**, *114*, 4–9. [[CrossRef](#)]
98. Igarashi, Y.; Nakatsu, N.; Yamashita, T.; Ono, A.; Ohno, Y.; Urushidani, T.; Yamada, H. Open TG-GATEs: A large-scale toxicogenomics database. *Nucleic Acids Res.* **2015**, *43*, D921–D927. [[CrossRef](#)] [[PubMed](#)]
99. Shi, L.; Campbell, G.; Jones, W.D.; Campagne, F.; Wen, Z.; Walker, S.J.; Su, Z.; Chu, T.M.; Goodsaid, F.M.; Pusztai, L.; et al. The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* **2010**, *28*, 827.
100. Shi, L.; Reid, L.H.; Jones, W.D.; Shippy, R.; Warrington, J.A.; Baker, S.C.; Collins, P.J.; De Longueville, F.; Kawasaki, E.S.; Lee, K.Y.; et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* **2006**, *24*, 1151. [[PubMed](#)]
101. Guo, L.; Lobenhofer, E.K.; Wang, C.; Shippy, R.; Harris, S.C.; Zhang, L.; Mei, N.; Chen, T.; Herman, D.; Goodsaid, F.M.; et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat. Biotechnol.* **2006**, *24*, 1162–1169. [[CrossRef](#)] [[PubMed](#)]
102. Nitsche, B.M.; Ram, A.F.; Meyer, V. The use of open source bioinformatics tools to dissect transcriptomic data. In *Plant Fungal Pathogens*; Springer: Berlin, Germany, 2012; pp. 311–331.
103. Slonim, D.K.; Yanai, I. Getting started in gene expression microarray analysis. *PLoS Comput. Biol.* **2009**, *5*. [[CrossRef](#)] [[PubMed](#)]
104. Draghici, S.; Khatrı, P.; Eklund, A.C.; Szallasi, Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* **2006**, *22*, 101–109. [[CrossRef](#)] [[PubMed](#)]

105. Rao, M.S.; Van Vleet, T.R.; Ciurlionis, R.; Buck, W.R.; Mittelstadt, S.W.; Blomme, E.A.; Liguori, M.J. Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. *Front. Genet.* **2019**, *9*, 636. [[CrossRef](#)]
106. Scarpato, M.; Federico, A.; Ciccodicola, A.; Costa, V. Novel transcription factor variants through RNA-sequencing: The importance of being “alternative”. *Int. J. Mol. Sci.* **2015**, *16*, 1755–1771. [[CrossRef](#)]
107. Bullard, J.H.; Purdom, E.; Hansen, K.D.; Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform.* **2010**, *11*, 94. [[CrossRef](#)]
108. Ramsköld, D.; Wang, E.T.; Burge, C.B.; Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **2009**, *5*. [[CrossRef](#)]
109. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **2009**, *6*, 377. [[CrossRef](#)] [[PubMed](#)]
110. Müller, S.; Diaz, A. Single-Cell mRNA sequencing in cancer research: Integrating the genomic fingerprint. *Front. Genet.* **2017**, *8*, 73. [[CrossRef](#)] [[PubMed](#)]
111. Yang, J.; Tanaka, Y.; Seay, M.; Li, Z.; Jin, J.; Garmire, L.X.; Zhu, X.; Taylor, A.; Li, W.; Euskirchen, G.; et al. Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res.* **2017**, *45*, 1281–1296. [[CrossRef](#)] [[PubMed](#)]
112. Ofengeim, D.; Giagtzoglou, N.; Huh, D.; Zou, C.; Yuan, J. Single-cell RNA sequencing: Unraveling the brain one cell at a time. *Trends Mol. Med.* **2017**, *23*, 563–576. [[CrossRef](#)]
113. Mohammed, H.; Hernando-Herraez, I.; Savino, A.; Scialdone, A.; Macaulay, I.; Mulas, C.; Chandra, T.; Voet, T.; Dean, W.; Nichols, J.; et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **2017**, *20*, 1215–1228. [[CrossRef](#)]
114. Lee, M.C.W.; Lopez-Diaz, F.J.; Khan, S.Y.; Tariq, M.A.; Dayn, Y.; Vaske, C.J.; Radenbaugh, A.J.; Kim, H.J.; Emerson, B.M.; Pourmand, N. Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4726–E4735. [[CrossRef](#)]
115. Shalek, A.K.; Satija, R.; Adiconis, X.; Gertner, R.S.; Gaublumme, J.T.; Raychowdhury, R.; Schwartz, S.; Yosef, N.; Malboeuf, C.; Lu, D.; et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **2013**, *498*, 236–240. [[CrossRef](#)]
116. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [[CrossRef](#)]
117. Zeisel, A.; Muñoz-Manchado, A.B.; Codeluppi, S.; Lönnerberg, P.; La Manno, G.; Juréus, A.; Marques, S.; Munguba, H.; He, L.; Betsholtz, C.; et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **2015**, *347*, 1138–1142. [[CrossRef](#)]
118. Picelli, S.; Faridani, O.R.; Björklund, Å.K.; Winberg, G.; Sagasser, S.; Sandberg, R. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **2014**, *9*, 171. [[CrossRef](#)]
119. Ziegenhain, C.; Vieth, B.; Parekh, S.; Reinius, B.; Guillaumet-Adkins, A.; Smets, M.; Leonhardt, H.; Heyn, H.; Hellmann, I.; Enard, W. Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell* **2017**, *65*, 631–643. [[CrossRef](#)] [[PubMed](#)]
120. Huang, H.; Yu, Y.; Hu, Y.; He, X.; Usta, O.B.; Yarmush, M.L. Generation and manipulation of hydrogel microcapsules by droplet-based microfluidics for mammalian cell culture. *Lab Chip* **2017**, *17*, 1913–1932. [[CrossRef](#)] [[PubMed](#)]
121. Yousuff, C.M.; Ho, E.T.W.; Hussain, K.; Hamid, N.H.B. Microfluidic platform for cell isolation and manipulation based on cell properties. *Micromachines* **2017**, *8*, 15. [[CrossRef](#)]
122. Sheng, K.; Cao, W.; Niu, Y.; Deng, Q.; Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* **2017**, *14*, 267. [[CrossRef](#)] [[PubMed](#)]
123. Fan, X.; Zhang, X.; Wu, X.; Guo, H.; Hu, Y.; Tang, F.; Huang, Y. Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biol.* **2015**, *16*, 148. [[CrossRef](#)] [[PubMed](#)]
124. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **2017**, *171*, 1437–1452. [[CrossRef](#)]

125. Ramaiahgari, S.C.; Auerbach, S.S.; Saddler, T.O.; Rice, J.R.; Dunlap, P.E.; Sipes, N.S.; DeVito, M.J.; Shah, R.R.; Bushel, P.R.; Merrick, B.A.; et al. The power of resolution: Contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. *Toxicol. Sci.* **2019**, *169*, 553–566. [[CrossRef](#)]
126. Yeakley, J.M.; Shepard, P.J.; Goyena, D.E.; VanSteenhouse, H.C.; McComb, J.D.; Seligmann, B.E. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *PLoS ONE* **2017**, *12*, e0178302. [[CrossRef](#)]
127. House, J.S.; Grimm, F.A.; Jima, D.D.; Zhou, Y.H.; Rusyn, I.; Wright, F.A. A pipeline for high-throughput concentration response modeling of gene expression for toxicogenomics. *Front. Genet.* **2017**, *8*, 168. [[CrossRef](#)]
128. Marwah, V.S.; Scala, G.; Kinaret, P.A.S.; Serra, A.; Alenius, H.; Fortino, V.; Greco, D. eUTOPIA: SolUTion for Omics data Preprocessing and Analysis. *Source Code Biol. Med.* **2019**, *14*, 1. [[CrossRef](#)]
129. Waters, M.; Stasiewicz, S.; Alex Merrick, B.; Tomer, K.; Bushel, P.; Paules, R.; Stegman, N.; Nehls, G.; Yost, K.J.; Johnson, C.H.; et al. CEBS—Chemical Effects in Biological Systems: A public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res.* **2007**, *36*, D892–D900. [[CrossRef](#)] [[PubMed](#)]
130. Lea, I.A.; Gong, H.; Paleja, A.; Rashid, A.; Fostel, J. CEBS: A comprehensive annotated database of toxicological data. *Nucleic Acids Res.* **2017**, *45*, D964–D971. [[CrossRef](#)] [[PubMed](#)]
131. Lamb, J. The Connectivity Map: A new tool for biomedical research. *Nat. Rev. Cancer* **2007**, *7*, 54–60. [[CrossRef](#)] [[PubMed](#)]
132. Ganter, B.; Snyder, R.D.; Halbert, D.N.; Lee, M.D. Toxicogenomics in Drug Discovery and Development: Mechanistic Analysis of Compound/Class-Dependent Effects Using The DrugMatrix<sup>®</sup> Database; *Pharmacogenomics* **2006**. [[CrossRef](#)]
133. Kolesnikov, N.; Hastings, E.; Keays, M.; Melnichuk, O.; Tang, Y.A.; Williams, E.; Dylag, M.; Kurbatova, N.; Brandizi, M.; Burdett, T.; et al. ArrayExpress update—Simplifying data submissions. *Nucleic Acids Res.* **2015**, *43*, D1113–D1116. [[CrossRef](#)]
134. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [[CrossRef](#)]
135. Duan, Q.; Flynn, C.; Niepel, M.; Hafner, M.; Muhlich, J.L.; Fernandez, N.F.; Rouillard, A.D.; Tan, C.M.; Chen, E.Y.; Golub, T.R.; et al. LINCS Canvas Browser: Interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* **2014**, *42*, W449–W460. [[CrossRef](#)]
136. Cheng, L.; Li, L. Systematic quality control analysis of LINCS data. *CPTv Pharm. Syst. Pharmacol.* **2016**, *5*, 588–598. [[CrossRef](#)]
137. Hansen, S.F.; Sørensen, S.N.; Skjolding, L.M.; Hartmann, N.B.; Baun, A. Revising REACH guidance on information requirements and chemical safety assessment for engineered nanomaterials for aquatic ecotoxicity endpoints: Recommendations from the EnvNano project. *Environ. Sci. Eur.* **2017**, *29*, 14. [[CrossRef](#)]
138. EFSA. EFSA Scientific Network on Harmonisation of Risk Assessment Methodologies. *EFSA Support. Publ.* **2013**, *10*, 489E.
139. CEC. *Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)*; European Parliament: Brussels, Belgium, 2006.
140. Gant, T.W.; Sauer, U.G.; Zhang, S.D.; Chorley, B.N.; Hackermüller, J.; Perdichizzi, S.; Tollefsen, K.E.; Van Ravenzwaay, B.; Yauk, C.; Tong, W.; et al. A generic transcriptomics reporting framework (TRF) for 'omics data processing and analysis. *Regul. Toxicol. Pharmacol.* **2017**, *91*, S36–S45. [[CrossRef](#)] [[PubMed](#)]
141. Sauer, U.G.; Deferme, L.; Gribaldo, L.; Hackermüller, J.; Tralau, T.; van Ravenzwaay, B.; Yauk, C.; Poole, A.; Tong, W.; Gant, T.W. The challenge of the application of 'omics technologies in chemicals risk assessment: Background and outlook. *Regul. Toxicol. Pharmacol.* **2017**, *91*, S14–S26. [[CrossRef](#)] [[PubMed](#)]
142. Farhat, A. *Adverse Outcome Pathway on Aryl Hydrogen Receptor Activation Leading to Early Life Stage Mortality, via Reduced VEGF*; OECD Publishing: Paris, France, 2019.

143. Corvi, R.; Ahr, H.J.; Albertini, S.; Blakey, D.H.; Clerici, L.; Coecke, S.; Douglas, G.R.; Gribaldo, L.; Groten, J.P.; Haase, B.; et al. Meeting report: Validation of toxicogenomics-based test systems: ECVAM-ICCVAM/NICEATM considerations for regulatory use. *Environ. Health Perspect.* **2006**, *114*, 420–429. [[CrossRef](#)] [[PubMed](#)]
144. European Commission. *Technical Guidance Document on Risk Assessment in Support Of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances, Commission Regulation (EC) No 1488/94 on Risk Assessment For Existing Substances, and Directive 98/8/EC of the European Parliament and Of the Council Concerning the Placing of Biocidal Products on the Market*; European Commission Joint Research Centre, EUR: Ispra, Italy, 2003; Volume 20418.
145. European Commission. *Technical Guidance Document on Risk Assessment*. Institute for Health and Consumer Protection, European Chemicals Bureau. Part II. 2003. Available online: [https://echa.europa.eu/documents/10162/16960216/tgdp2\\_2ed\\_en.pdf](https://echa.europa.eu/documents/10162/16960216/tgdp2_2ed_en.pdf) (accessed on 3 April 2020) .
146. OECD. *Environment Directorate, Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology*; OECD: Paris, France.
147. Thomas, R.S.; Bahadori, T.; Buckley, T.J.; Cowden, J.; Deisenroth, C.; Dionisio, K.L.; Frithsen, J.B.; Grulke, C.M.; Gwinn, M.R.; Harrill, J.A.; et al. The next generation blueprint of computational toxicology at the US Environmental Protection Agency. *Toxicol. Sci.* **2019**, *169*, 317–332. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).