Improved computational methods and strategies for nanomaterial grouping

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Aileen Bahl

2023

The thesis was created between 2018 und 2023 under supervision of PD Dr. Andrea Haase at the German Federal Institute for Risk Assessment.

1st reviewer: PD Dr. Andrea Haase

2nd reviewer: Prof. Dr. Daniel Klinger

Date of defense: June, 11th 2024

Acknowledgement

First and foremost, I would like to thank PD Dr. Andrea Haase for the continuous support during the preparation of publications, the supervision of my thesis and with my research and work in general. I have learnt a lot during discussions with her and I am very grateful for the helpful comments and suggestions over the years and the efforts she made to keep the projects running.

In addition, I would like to thank Prof. Dr. Daniel Klinger for the supervision of my thesis and the kind support he offered during that time.

I would also like to thank Prof. Dr. Dr. Andreas Luch for the opportunity to complete my doctorate at the German Federal Institute for Risk Assessment in Department of Chemical and Product Safety.

I also want to thank Prof. Dr. Bernhard Renard for supporting my work with very helpful comments and discussions and for the opportunity to be part of his group at the RKI as a guest scientist and to get insights into his inspiring research. I would also like to thank the whole MF1 team. I really enjoyed the time with you all.

In the same way, I had a great time in unit 76 and want to thank all my colleagues for great times we spent, many ups and downs we mastered and great discussions we had. Especially, I would like to thank Verónica, Mario and Anna for always offering support and answering all kind of questions that came to my mind. Also, I would like to thank Anne and Amelie for mutual mental support during some stressful times in the process of completing our theses and for always making me smile again.

In addition, I would like to thank all the great partners we had in the different projects, especially the ones in close collaboration. I have learnt a lot from them and always felt great support.

Finally, I would also like to express a big thank you to my family and friends – my parents who always supported me in any way they could till today, the best kids in the world who always made me forgot any trouble I had, my partner for supporting me during all the stressful times and always believing in me and all my lovely friends who are always there listening to me and motivating or distracting me (whatever was needed).

Declaration of Independence

Herewith, I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me.

Abbreviations

ADME *absorption, distribution, metabolism and excretion*

AI *Artificial intelligence*

AO *Adverse outcome*

BMD *benchmark dose*

BMDL *benchmark dose limit*

DNEL *derived no effect level*

ECHA *European chemicals agency*

ESR *Electron spin resonance*

FAIR *Findable, accessible, interoperable, reusable*

FDR *False discovery rate*

FRAS *Ferric reduction ability of serum*

GSEA *Gene set enrichment analysis*

H0 *Null hypothesis*

IATAs *Integrated approaches to testing and assessment*

KEs *Key events*

kNN k-nearest neighbors

LOAEL *Lowest observed adverse effect level*

LOOCV *Leave-one-out cross-validation*

MIE *Molecular initiating event*

ML *Machine learning*

MoAs *Modes-of-action*

MS *Mass spectrometry*

MWCNTs *multi-walled carbon nanotubes*

NFs *nanoforms*

NMs *Nanomaterials*

NOAEL *No observed adverse effect level*

OECD *Organization for economic co.operation and development*

OP *Oxidative potential*

PBPK *physiologically-based pharmacokinetic*

PCA *Principle component analysis*

PCs *Principle components*

PRIDE *PRoteomics IDEntification*

qAOPs *quantitive adverse outcome pathways*

QSAR *quantitative structure activity relationship*

REACH *Registration, evaluation, authorization and estriction of chemicals*

RF Random forest

RFE Recursive feature elimination

ROS *reactive oxygen species*

SOPs *Standard operating procedures*

SSbD *Safe(r)-and-sustainable-by-design*

STIS *Short-term inhalation studies*

STOT *Specific target organ toxicity*

Outline

Zusammenfassung

Nanomaterialien (NM) können durch Feinabstimmung ihrer physikalisch-chemischen Eigenschaften für verschiedene industrielle Zwecke hergestellt werden. Dies führt zu einer theoretisch unbegrenzten Anzahl von NM-Varianten. Dabei können selbst kleine Variationen in den physikalisch-chemischen Eigenschaften eines NM einen erheblichen Einfluss auf seine Aufnahme, Toxikokinetik und (Öko-)Toxizität haben. Daher muss theoretisch eine separate Risikobewertung für alle Varianten und alle toxikologischen Endpunkte durchgeführt werden. Da dies nicht realistisch umsetzbar ist, sind Gruppierungs- und Read-Across-Ansätze, die eine Übertragung von Informationen zwischen hinreichend ähnlichen NM ermöglichen, vielversprechende Alternativen. Es ist jedoch nicht trivial, zuverlässige Gruppierungsansätze für NM zu entwickeln, da die Beziehung zwischen den einzelnen physikalisch-chemischen Eigenschaften und dem toxikologischen Profil von NM noch nicht hinreichend geklärt ist. Das übergeordnete Ziel dieser Arbeit war es, zu untersuchen, wie maschinelles Lernen (ML) eingesetzt werden kann, um die Gruppierung von NM bei der Suche nach solchen Beziehungen oder den zugrunde liegenden Mustern zu unterstützen. Da die Formulierung einer zuverlässigen Gruppierungshypothese in hohem Maße von einem mechanistischen Verständnis profitieren kann, wurden die zugrundeliegenden Wirkungsweisen oder Modes-of-Action (MoAs) für verschiedene NM auch durch die Untersuchung von Ergebnissen aus Omics-Ansätzen betrachtet.

In der ersten Studie wurde versucht, mit Hilfe von ML die wichtigsten physikalisch-chemischen Eigenschaften zu ermitteln, die die Toxizität von NM beeinflussen. Dazu wurde ein Datensatz von elf NM mit einer umfassenden Beschreibung ihrer physikalisch-chemischen Eigenschaften verwendet. Diese physikalisch-chemischen Eigenschaften wurden dann mit verfügbaren *in vivo* Daten aus Kurzzeit-Inhalationsstudien (STIS) und *in vitro* Toxizitätsdaten verknüpft, die mit dem sogenannten Makrophagen-Assay gemessen wurden. In beiden Fällen wurde die Toxizität als binäre Ergebnisvariable dargestellt, die angibt, ob ein NM in den jeweiligen Toxizitätsstudien "aktiv" oder "passiv" war. Unüberwachte und überwachte ML-Ansätze wurden auf diesem Datensatz trainiert. Im unüberwachten Modell wurde die Hauptkomponentenanalyse (PCA) verwendet, um Informationen darüber abzuleiten, welche physikalisch-chemischen Eigenschaften in den ersten beiden Hauptkomponenten (PC) den stärksten Einfluss haben. Anschließend wurden die Ergebnisse mit Hilfe eines k-Nächste Nachbarn (kNN) Ansatzes mit den definierten Aktivitätsstufen verglichen. Im überwachten Gegenstück wurde eine Random Forest (RF) Analyse mit und ohne rekursive Merkmalseliminierung (RFE) durchgeführt. Die Toxizitätsklassen wurden somit direkt als Kennzeichnungen im Modellbildungsprozess verwendet. Insgesamt wurde das beste Modell mit RF

und RFE erzielt. Es erreichte eine ausgewogene Genauigkeit von 0,82 und wurde auf den drei Parametern Zetapotenzial, Redoxpotenzial und Auflösungsrate aufgebaut. Diese Studie zeigte, wie ML die NM-Gruppierungsansätze unterstützen kann. Gleichzeitig wurde aber auch deutlich, dass die Vorhersagemodellierung, die nur auf physikalisch-chemischen Eigenschaften beruht, erhebliche Einschränkungen aufweist. Während dies heute allgemein anerkannt ist und auch in verschiedenen Empfehlungen erwähnt wird, war dies zum Zeitpunkt der Studie noch nicht der Fall. Bereits mit diesem kleinen Datensatz von NM, der im Vergleich zu vielen anderen Studien eine umfassende Beschreibung der physikalisch-chemischen Eigenschaften aufwies, ist keine perfekte Trennung von "aktiven" und "passiven" NM möglich.

Die zweite Studie zielte darauf ab, neben den physikalisch-chemischen Eigenschaften auch das Oxidationspotenzial (OP) zu untersuchen. Verschiedene Assays zur Messung des OP wurden im Hinblick auf ihre Vorhersagekraft für die NM-Toxizität bewertet. Darüber hinaus war es ein Ziel herauszufinden, ob OP-Assays einander ersetzen können oder ob die Ergebnisse verschiedener OP-Assays aus unterschiedlichen Datensätzen kombiniert werden können. Im Rahmen dieser Studie wurden vier Assays verglichen, nämlich die azellulären Oberflächenreaktivitätsassays Elektronenspinresonanzspektroskopie (ESR) unter Verwendung einer CPH-Spinsonde und einer DMPO-Spinfalle und der Ferric Reduction Ability of Serum (FRAS) Assay sowie der zelluläre Proteincarbonylierungsassay als Marker für oxidative Proteinschäden in NRK-52E-Zellen. Der Vergleich basierte auf einer Fallstudie mit OP-Messungen für 35 NMs. Für die vier OP-Assays wurden die massebasierten Dosen mit den oberflächenbasierten verglichen, Korrelationen und Clustering zwischen den Assays berechnet und ihre Vorhersagekraft für die gleichen Ergebnisvariablen wie in der ersten Studie für einzelne Assays und alle möglichen Kombinationen davon in einem logistischen Regressionsmodell bewertet. Diese Vergleiche ergaben, dass die oberflächenbasierten Dosen eine bessere Vorhersagekraft haben als die massebasierten. Darüber hinaus waren die Korrelationen zwischen den OP-Assays nur moderat. Im Rahmen der logistischen Regressionsanalyse war die Vorhersagekraft bei der Proteincarbonylierung oder bei Kombinationen von Assays, die die Proteincarbonylierung einschließen, am höchsten. Wie erwartet, scheinen also biologische OP-Tests das tatsächliche Toxizitätsergebnis zuverlässiger vorherzusagen. Gleichzeitig ist die Kombination von Datensätzen, bei denen verschiedene OP-Assays verwendet wurden, um robuste ML-Modelle auf der Grundlage großer Datensätze zu erstellen, nicht ohne weiteres möglich, da die Assays nicht hoch korreliert sind. Insgesamt scheint das OP sehr informativ und relevant für NM im Allgemeinen zu sein. Allerdings können auch andere (nicht direkt mit dem OP zusammenhängende) Toxizitätsmechanismen durch die Behandlung mit NM ausgelöst werden, die durch OP-Assays nicht erfasst werden können.

Die dritte Studie konzentrierte sich allgemeiner auf die Aufklärung von MoAs, die der Toxizität von NM zugrunde liegen. Hier sollten proteomische Daten auf ihr Potenzial hin untersucht werden, die MoAs von NM zu entschlüsseln und so deren Gruppierung zu unterstützen. Da jedoch relativ wenige proteomische Daten für NM existieren und die Interpretation aufgrund fehlender Referenzdaten schwierig ist, bestand die Hauptidee darin, proteomische Signaturen, die für NM beobachtet wurden, mit denen von anderen Kompenenten wie Chemikalien, Drogen oder Krankheiten zu vergleichen. Da solche Meta-Analysen vor allem durch die fehlende Standardisierung von proteomischen Daten beeinträchtigt werden, wurde ein Workflow für die harmonisierte Auswertung von öffentlichen proteomischen Daten und deren Integration in eine Meta-Analyse entwickelt. Der PROTEOMAS-Workflow zielt auf die FAIRifizierung (Findable, Accessible, Interoperable, Reusable) von proteomischen Daten ab. In einer ersten Fallstudie wurde PROTEOMAS an 25 proteomischen Datensätzen getestet, um die toxikologischen Wirkungen von NM im Verhältnis zu denen anderer Komponenten in der Lunge zu untersuchen. Proteomische Fingerabdrücke und deren Ähnlichkeiten zwischen den untersuchten Komponenten konnten identifiziert werden. PROTEOMAS war somit nützlich für die Meta-Analyse von Proteomdaten.

Der Übersichtsartikel gibt einen Überblick über die Vielfalt der in der Literatur verfügbaren ML-Modelle und Omics-Ansätze zur Unterstützung der NM-Gruppierung. Entsprechende Modelle wurden gesammelt und einige übergreifende Schlussfolgerungen aus diesen Manuskripten gezogen. Insbesondere die Datenverfügbarkeit und -qualität sind ein großes Problem, das die Entwicklung robuster ML-Modelle für die Vorhersage der Toxizität von NM erschwert. Darüber hinaus sind die Messungen in der Regel nicht gut standardisiert und es werden nur unzureichende Metadaten bereitgestellt, so dass die Datensätze nicht integriert werden können. Insgesamt besteht im Bereich der NM-Sicherheit ein großer Bedarf an FAIRen Daten, was die Entwicklung zuverlässigerer Modelle und die Weiterentwicklung von *in silico* Tools im regulatorischen Kontext ermöglichen würde. Es wurde auch beschrieben, dass die aktuellen Entwicklungen auf dem Gebiet der künstlichen Intelligenz die Schließung von Datenlücken und die Verbesserung der Verfügbarkeit von Metadaten in NM-Datenbanken sowie verknüpfte Datenkonzepte erheblich unterstützen können.

Insgesamt haben sich ML-Modelle und Omics-Methoden als nützlich erwiesen, um NM-Gruppierungsansätze zu unterstützen. Allerdings sind die Datenverfügbarkeit und die Standardisierung der Methoden von größter Bedeutung, um zuverlässige Modelle entwickeln zu können.

Summary

Nanomaterials (NMs) can be manufactured to serve different industrial purposes by fine-tuning their physico-chemical properties. This results in a theoretically unlimited number of NM variants. Thereby, even small variations in the physico-chemical properties of a NM may have substantial influence on their uptake, toxicokinetics as well as (eco-)toxicity. Thus, in theory, risk assessment needs to be performed for all variants and all toxicological endpoint. As this is simply not feasible, grouping and read-across approaches which allow the transfer of information between sufficiently similar NMs are promising alternatives. However, establishing reliable grouping approaches for NMs is not trivial due to the current lack of understanding with respect to the relationship between individual physico-chemical properties and the toxicological profile of NMs. The overall aim of this thesis was to explore how machine learning (ML) approaches can be used to support NM grouping in finding such relationships or underlying patterns. As formulating a reliable grouping hypothesis may largely benefit from mechanistic understanding, the underlying modes-of-action (MoAs) for different NMs were also explored by investigating results from omics approaches.

In the first study, the aim was to use ML for identifying the most important physico-chemical properties influencing the toxicity of NMs. Therefore, a dataset of eleven NMs with comprehensive description of their physico-chemical properties was used. These physico-chemical properties were then linked to available *in vivo* data obtained from short-term inhalation studies (STIS) and *in vitro* toxicity data measured with the so-called macrophage assay. In both cases, toxicity was represented as binary outcome variable indicating whether a NM was 'active' and 'passive' in the respective toxicity studies. Unsupervised and supervised ML approaches were trained on this dataset. In the unsupervised model, principal component analysis (PCA) was used to infer information on which physico-chemical properties have the strongest impact in the first two principal components (PCs). Afterwards, k-nearest neighbors (kNN) was applied to compare results to the defined activity levels. In the supervised counterpart, random forest (RF) analysis with and without recursive feature elimination (RFE) was performed. Toxicity classes were thereby directly used as labels in the model building process. Overall, the best model was obtained using RF with RFE. It reached a balanced accuracy of 0.82 and was built on the three parameters zeta potential, redox potential and dissolution rate. This study showed, how ML could support NM grouping approaches. At the same time, it was also obvious that predictive modeling based solely on physico-chemical properties has severe limitations. While this is widely accepted to date and also mentioned in various recommendations, this was not the case at the time of the study. Already with this small set of NMs

and compared to other studies a comprehensive description of physico-chemical properties, no perfect separation of 'active' and 'passive' NMs was possible.

The second study, aimed at investigating the oxidative potential (OP) in addition to physico-chemical properties. Different assays for measuring the OP were evaluated with respect to their predictivity for NM toxicity. In addition, one goal was to find out whether or not OP assays could replace each other or whether results from different OP assay in different datasets could be combined. Within this study, four assays have been compared, namely the acellular surface reactivity assays electron spin resonance (ESR) spectroscopy using CPH spin probe and DMPO spin trap and the ferric reduction ability of serum (FRAS) assay as well as the cellular protein carbonylation assay as a marker for oxidative protein damage in NRK-52E cells. The comparison was based on a case study holding OP measurements for 35 NMs. For the four OP assays, mass-based doses were compared to surface-based ones, correlations and clustering between assays were computed and their predictivity for the same outcome variables as in the first study was assessed for individual assays and all possible combinations of them in a logistic regression model. As a result of those comparisons, surface-based doses were shown to be more predictive than mass-based ones. In addition, correlations between the OP assays were only moderate. Within the logistic regression model, predictivity was highest for protein carbonylation or combinations of assays which include protein carbonylation. Thus, as expected, biological OP assays seem to predict the actual toxicity outcome more reliably. At the same time, combining datasets which used different OP assays for the purpose of building robust ML models based on large datasets is not easily possible as the assays are not highly correlated. Overall, OP seems to be very informative and relevant for NMs in general. However, also other toxicity mechanisms (not directly related to OP) may be triggered by NM treatment, which cannot be reflected by OP assays.

The third study focused more generally on elucidating MoAs underlying NM toxicity. Here, proteomics data were to be explored for their potential to unravel MoAs of NMs to support NM grouping. However, as proteomics data for NMs are relatively scarce and interpretation is difficult due to missing reference data, the main idea was to integrate proteomics signatures observed for NMs with those from other traits like chemicals, drugs or diseases. As such meta-analyses are mainly hampered by the lack of standardization for proteomics data, a workflow for harmonized evaluation of public proteomics data and their integration in a meta-analysis setting was developed. The workflow PROTEOMAS aims to make proteomics data FAIR (findable, accessible, interoperable, reusable). In an initial case study, PROTEOMAS was tested on 25 proteomics datasets to investigate the toxicological effects of NMs in relation to those of other traits at the lung level. Proteomic

fingerprints and their similarities among the studied traits could be identified. PROTEOMAS was thus useful for meta-analysis of proteomic datasets.

In the review article, an overview on the variety of ML models and omics approaches supporting NM grouping available in literature is provided. Corresponding models were collected and some overarching conclusions were drawn from these manuscripts. Especially, data availability and quality are a major concern preventing the development of robust ML models for NM toxicity prediction. In addition, measurements are usually not well-standardized and insufficient metadata is provided and thus datasets cannot be integrated. Overall, there is a strong need for FAIR data in the NM safety community which would then allow development of more reliable models and advancement of *in silico* tools in a regulatory context. It was also concluded, that recent developments in the field of AI may also greatly support data gap filling and improvement of metadata availability in NM databases as well as linked data concepts.

Overall, ML models as well as omics methods were shown to be useful to support NM grouping approaches. However, data availability and standardization of methods are of utmost importance in order to be able to develop reliable models.

Chapter 1: Introduction

1.1 NMs: Definition, properties and applications

According to the recommendations of the European Commission, a NM is any material that consists of solid particles of which at least 50% in the number-based size distribution have at least one external dimension in the size range of 1 to 100 nm. The particles may thereby be unbound or bound in agglomerates or aggregates[1]. NMs can be produced from a wide range of core materials. Especially, NMs derived from metals and metal oxides, carbon-based materials like carbon nanotubes or graphene as well as organic NMs are frequently used. In addition, the physico-chemical properties of each NM can be varied to optimize their suitability for certain applications. This fine-tuning leads to a multitude of different variants which compared to their corresponding bulk materials, often show enhanced magnetic, electrical, optical, mechanical, catalytic and other properties. For instance, the increased surface area to volume ratio enhances their catalytic activity. In addition, hardness, stiffness as well as thermal stability may be enhanced for NMs compared to their macroscopic equivalents. In case of very small NMs like quantum dots, quantum effects may dominate thus influencing their fluorescent, magnetic, and electrical capacities[2]. These characteristics render them useful for a wide range of applications[3].

NMs are used in different sectors like medicine, cosmetics, textiles, electronics, construction or the food sector. In the textiles industry, graphene-based NMs and carbon nanotubes are used to enhance mechanical and thermal stability[4]. In addition, metal NMs like silver or copper oxide NMs offer antimicrobial[5] and UV-resistant properties[6]. Silica NMs as well as carbon nanotubes can be used to create superhydrophobic coatings which enhance water-repelling properties of surfaces[7]. In cosmetics, titanium dioxide or zinc oxide based NMs are frequently added to sunscreen for effective protection against UVA and UVB radiation[8, 9]. At the same time, nanoclay is employed in food packaging to enhance barrier properties against gases and moisture and provide antimicrobial properties[10]. Fullerenes, nanotubes as well as metal oxide NMs are also applied in the development of solar cells, batteries, and fuel cells as they can improve energy conversion efficiency, storage capacity, and the overall performance of energy systems[11]. Another important sector for the use on NMs is the one of medicine and healthcare. Here, polymeric NMs are well-suited for targeted drug delivery as they enhance the efficacy and specificity of therapeutic agents while at the same time being largely biocompatible[12]. At the same time dendrimers are useful for medical imaging due to properties like high rigidity, low polydispersity and the possibility to easily apply surface

modifications[13]. Similarly, gold NMs can also support medical imaging as they show high x-ray absorption properties and thus can aid computed tomographies[14]. Additionally, NMs are used for tissue engineering of bone, skin, nerve, and dental tissues where they can improve the interaction between artificial implants and biological systems[15]. Finally, also environmental applications can profit from NMs, especially during the removal of pollutants and enhancement of the efficiency of environmental cleaning processes. Here, titanium dioxide NMs or carbon nanotubes can be used for water treatment, air purification, and environmental remediation as their small size and high surface area make them efficient catalysts[16, 17].

Due to this large spectrum of possible applications, a steadily increasing amount of NM has been produced over the last years. This also poses concerns with respect to safety of NMs and requires efficient safety assessment strategies to cope with the steadily increasing number of NMs on the market.

1.2 Safety assessment

While the fine-tuning of NM properties is largely advantageous from an industrial point of view, it also raises concern with respect to the safety of all these new materials and material variants for human health and the environment. A comprehensive safety or risk assessment, which is a systematic process used to identify, evaluate, and estimate the level of risk posed by certain compounds, is necessary for each variant to understand potential implications and risks emerging from the widespread use of NMs.

For NMs, variations in factors like size, shapes or surface characteristics may have substantial effects on their uptake, toxicokinetics or (eco-)toxicity[18]. In general, especially, the large surface area to volume ratio is critical for NM toxicity as this increases the reactivity of the NMs allowing for much larger interactions between NMs and their environment[19]. In addition, the toxicity of NMs can also change in different surrounding media as well as over their lifetime. Due to their small size, NMs may also be capable of crossing biological barriers and interacting with living organisms at the cellular and molecular level[20]. Depending on their biodurability, NMs may also accumulate in biological tissues, thus causing long-term toxic effects. Therefore, substance identity plays a crucial role for risk assessment of NMs. More detailed insights from the regulatory point of view will be provided in the next section.

In general, risk assessment consists of two pillars: hazard assessment and exposure assessment. Hazard thereby relates to negative health effects induced by biological, chemical or physical agents. Hazard assessment includes hazard identification followed by hazard characterization. Hazard identification relates to the question which toxicological endpoints are triggered by a chemical. This directly relates to the MoA that is induced. In addition, hazard characterization addresses the determination of a critical threshold. This threshold defines the maximum dose of a substance that can be applied safely. In order to determine this critical dose level, dose-response studies applying a substance in multiple concentrations need to be carried out.

In addition, exposure is the second pillar of main interest for risk assessment as the hazard of a substance is only relevant if humans or the environment actually get into contact with it. Considering the frequent use of NMs in our daily lifes, a wide range of potential exposure scenarios for humans and environment is conceivable. At the same time, the actual exposure depends on whether a NM may be released or whether it stays firmly bound in a matrix. The primary concern revolves around the intended application of the NM and the probability that humans or the environment get into contact with them. In case, NMs are released from a product, humans can be exposed via three main exposure routes, namely the dermal, oral or inhalative route.

The focus in this work is on hazard identification for NMs in case of human exposure via the inhalative route.

1.2.1 NMs under REACH

In Europe, the overarching regulation for chemicals is REACH (registration, evaluation, authorization, and restriction of chemicals)[21]. In addition, various other regulations for specific sectors are available[22-26]. The aim of REACH is to manage the production, use, and disposal of chemicals to mitigate potential risks and to protect human health as well as the environment. Within REACH, it is stated that each chemical substance produced in or imported into the EU in quantities exceeding one ton per year needs to be registered with the European chemicals agency (ECHA), providing detailed information on their properties, hazards, and safe usage. The bulk form as well as the different NM variants are thereby covered under the same substance registration. Under REACH, the molecular structure and chemical composition are important for determining substance identity. However, for NMs, which are considered as forms of chemical substances, additional nano-specific requirements are needed in order to sufficiently describe them. Therefore, amendments to the REACH annexes[27]

have been introduced to specifically cover information requirements for NMs which exceed those of conventional chemicals. In the revised version of Annex VI of the REACH regulation[27], the concept of nanoforms (NFs) is introduced. A substance can contain several NFs where NFs are distinguished based on differences in the parameters defined under points 2.4.2 to 2.4.5. of Annex VI, namely the size distribution, shape and other morphological characteristics, surface treatment and functionalization as well as the specific surface area of the particles. If two NFs vary in one of these parameters with differences being larger that the batch-to-batch variability during production[28], they are considered to be distinct. According to the amendments to the REACH Annexes[27] which were published in 2018, different NFs of one substance need to be identified and assessed separately in the registration dossier.

One important step when putting new substances and mixtures on the market is their classification with respect to hazards. Different health hazards are relevant in this regard. An overview is given in Table 1. Self-classification with respect to these endpoints needs to be performed by the producers themselves when registering a new chemical. In addition, harmonized classification is performed by the EU in case a substance is identified to be carcinogenic, mutagenic, reprotoxic or a respiratory sensitizer.

The gold standard for hazard assessment for several endpoints are results from *in vivo* testing. The term i*n vivo* study describes experiments carried out with any kind of living animals. In case of toxicity testing for human hazard assessment, this usually refers to studies performed on rodents. In the regulatory context, toxicity testing still largely relies on *in vivo* studies, especially when it comes to more complex higher tier endpoints. Various factors need to be considered when performing an *in vivo* study as they may influence the study result. Especially species, strain and sex of the test animals as well as study duration and number of animals per group are important factors in this regard.
Commonly used test animals in hazard assessment are rats (Rattus norvegicus) or mice (Mus musculus). With respect to duration, one distinguishes acute, sub-acute, sub-chronic and chronic studies. While acute studies focus on short term effects usually induced by one single dose of the test substance, studies with an extended duration and repeated application of the test substance are needed to obtain information on long-term effects. The number of animals per group plays an important role with respect to reliability of the findings in terms of statistical significance. In addition, the influence of biological diversity of animals on the reaction to the test substance can only be tested if the number of animals is sufficiently large. In order to improve the informative value of a toxicological test and to derive suitable limit values, it is also necessary to apply multiple concentrations of the substance.

Table 1: Relevant endpoints for human health under REACH. Which endpoints need is be taken into account in a specific case is depends on the amount in which a chemical is produced (tonnage triggered).

Endpoint	Description
Acute toxicity	Adverse effects after single dose, multiple doses given within 24 hours or inhalation exposure of 4 hours
Skin corrosion/irritation	Irreversible (corrosion) or reversible (irritation) damage to the skin
Serious eye damage/irritation	Serious eye damage: tissue damage in the eye, or serious physical decay of vision, which is not fully reversible within 21 days Eye irritation: changes in the eye, which are fully reversible within 21 days
Respiratory or skin sensitization	Respiratory sensitization: hypersensitivity of the airways following inhalation of the substance Skin sensitization: allergic response following skin contact
Mutagenicity	Alteration of the structure, information content or segregation of DNA
Carcinogenicity	Induction of cancer or increase of its incidence
Reproductive toxicity	Adverse effects on sexual function and fertility in adult males and females or developmental toxicity in the offspring (before or after birth)
Specific target organ toxicity (STOT) – single exposure	Specific, non-lethal target organ toxicity after single exposure
Specific target organ toxicity (STOT) – repeated exposure	Specific, non-lethal target organ toxicity after repeated exposure

Different values can be determined to describe these limit values. The easiest and most straight-forward approach is to determine the NOAEL (no observable adverse effect level) or the LOAEL (lowest observable adverse effect level). The NOAEL is defined as the maximum tested dose for which no adverse effect could be observed while the LOAEL is the lowest test dose showing adverse effects. While being well established in risk assessment, the disadvantage of these measures is that the actual value depends strongly on which doses were actually tested. Depending on the spacing of the doses, these values may therefore be relatively far from the actual limit value and NOAELs or LOAELs from different studies for different chemicals are not directly comparable. An alternative approach is based on benchmark doses (BMD). This approach relies on fitting a quantitative dose-

response curve to the measured data. Based on this fitted curve, one can determine the BMD which is the dose leading to a certain level of response or the benchmark dose limit (BMDL) which in addition considers a confidence interval around the BMD. One important advantage of this method is that it also considers the shape of the dose-response curve across the whole range of tested doses. However, this approach can only be used for effects showing a certain minimum level of response.
In order to obtain reliable safety margins for human exposure limits from animal studies, one needs to consider some uncertainty factors in addition[29]. First, uncertainties may result from inter-species variation from the test animal to humans. Taking this uncertainty factor into account allows to extrapolate from the average animal response to the average human response. In addition, intra-species variation between humans also needs to be considered in order to guarantee that also sensitive humans are safely covered by the allowed limit concentrations. Other uncertainty factors may also play a role, e.g., uncertainties due to extrapolation from short-term to chronic exposure. Taking into account these uncertainty factors, one can obtain the derived no effect level (DNEL).

1.2.2 Inhalation as the main route of exposure

Among the three routes of exposure, inhalation is considered the most critical one for NMs as penetration into the deep lungs is comparatively easy and can potentially lead to respiratory toxicity, systemic absorption and accumulation in secondary organs[30]. Therefore, studying the pulmonary effects of NMs plays a major role in the field of nanotoxicology and is also the main focus within this thesis.

An overview on the lung structure[31] is given in Figure 1. The human respiratory tract is separated into the upper airways containing the nasopharynx as well as the lower airways with the tracheobronchial area and the alveolar region. The nasopharynx is comprised of the nasal cavity and mouth as well as the pharynx and larynx. The lower respiratory tract starts with the trachea, followed by the two main bronchi which further separate into smaller bronchioles in the left and right lung. Bronchioles finally terminate into alveoli in which the gas exchange takes place.

NMs are of particular concern with respect to inhalation due to their potential to reach deep into the lower airways where they mainly deposit in the alveoli. Particles with aerodynamic diameters in the micrometer range remain in the upper airways and are cleared via the mucociliary clearance[31]. In this process, cilia on the membrane covered by a viscous mucus layer move the particles upwards allowing for transfer into the gastrointestinal tract and subsequent excretion. Instead, particles

below 100 nm can reach the distal lung and the alveoli in which the gas exchange takes place[30-32]. The most important cell types being in contact with the NM in this part of the lung are alveolar macrophages and epithelial cells[31].

The alveolar macrophages are responsible for clearance from the alveolar part of the lung via phagocytosis[33-35]. Macrophages engulf particles that reached the alveoli and are then cleared via the mucociliary escalator[36] or the lymph nodes[37]. According to their role, the uptake of NMs into macrophages is very high compared to other cell types. However, the uptake varies depending on the physico-chemical properties of the NMs under consideration. Different studies investigated the behavior of macrophages in terms of NM uptake and responses to NM treatment[38-41]. NMs which are not recognized by macrophages fast enough can reach the alveolar epithelial cells.

Alveolar epithelial cells can be divided into two types: Type I cells form a layer which functions as a barrier between the gas phase within the lung and the blood stream and are responsible for the gas exchange[42, 43]. Type II cells secrete pulmonary surfactant which coats the alveoli and prevents them from collapsing[44]. Epithelial cells can also take up NMs *via* transcytosis which allows NMs to reach the interstitial space[45]. From there, translocation into the blood circulation or the lymph nodes may take place which may potentially cause adverse effects in secondary organs in the end[46-48]. Translocation from the alveolar space to the capillary system and thus systemic availability is however size-dependent and comparatively low[47, 49, 50]. Whether or not NMs accumulate in secondary organs or are cleared from the system depends on their biodurability. In addition, lung epithelial cells are known to respond to toxic external stimuli by releasing chemokines and cytokines triggering the activation of the immune system and inflammatory processes[51]. In long term, this may cause damage of lung tissue or the development of inflammatory lung diseases like COPD or asthma[52, 53].

Therefore, alveolar macrophages and epithelial cells are of high relevance for investigating NM toxicity *in vitro* and the derivation of corresponding computational models. Due to the fact that the majority of particles is taken up by macrophages, however, their response may be expected to be much larger compared to epithelial cells.

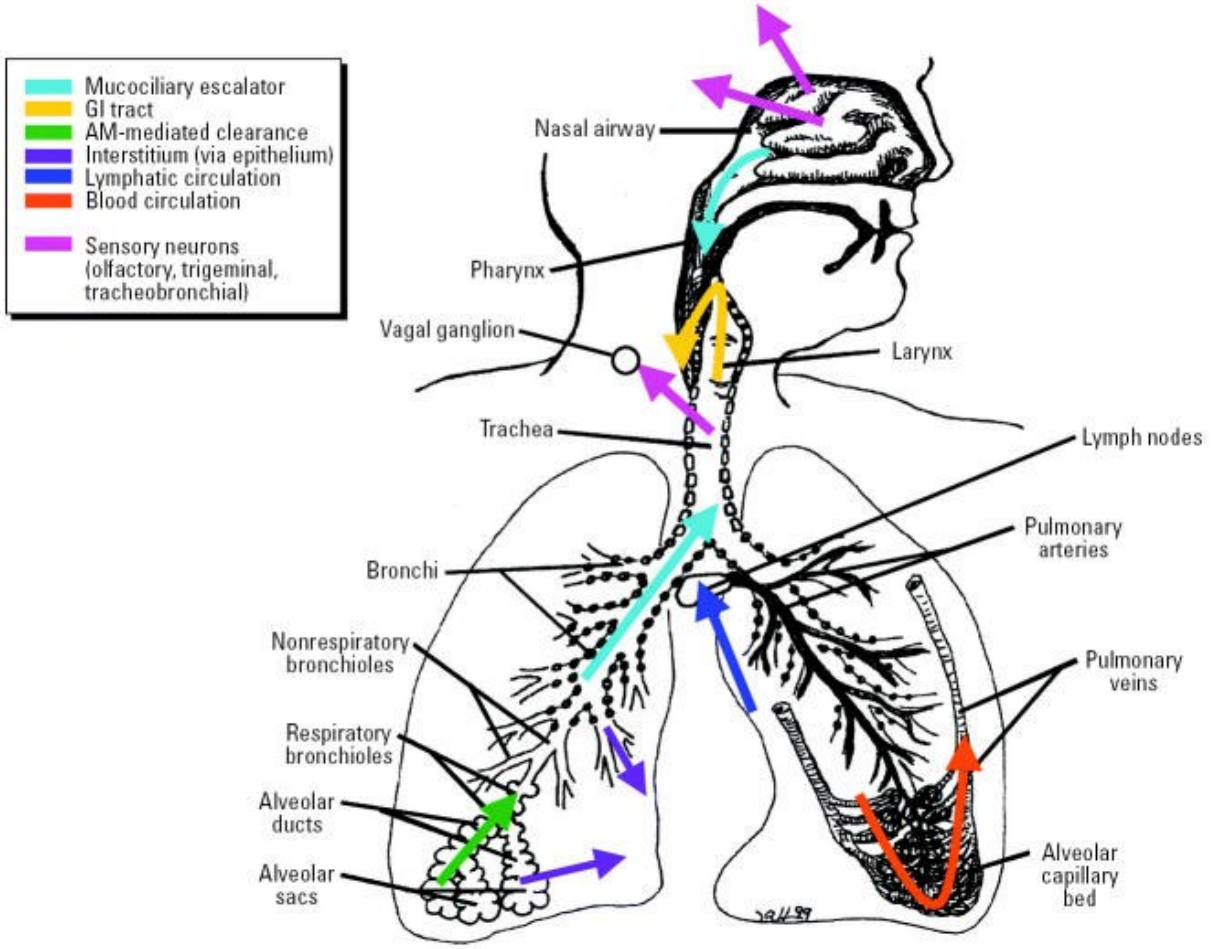


Figure 1: Structure of the respiratory tract. The figure is reprinted from Oberdörster et al.[31]

1.2.3 MoAs

The way in which a substance acts and the changes it induces is called the MoA. MoAs describe specific biological processes which lead to the observed physiological effects induced by a substance. Different MoAs may be invoked by NMs upon human exposure. Which MoA is relevant in a specific case, largely depends on the physico-chemical properties of the NM under consideration.

Firstly, NMs with high biopersistence are of concern as they are not efficiently cleared from the body over a long time. In that case, generic particle effects may be induced after chronic exposure to high doses of insoluble particles with low cytotoxicity[54]. Within the lungs, an overload condition of particles may occur where NMs cannot be taken up fast enough by macrophages[55]. In addition, NMs may also be translocated to secondary organs and accumulate there[56, 57]. This constant presence of particles may lead to chronic inflammation in the lung which might in turn cause pulmonary fibrosis and ultimately lung cancer or respective outcomes in the secondary organs[52, 53, 58]. On the other

hand, if a NM quickly dissolves, it releases ions which may be toxic. While this is not a NM-specific toxicity mechanism, many NMs on the market are metals or metal oxides for which such a toxic ion release frequently occurs. Released ions can directly interact with and damage cellular components like DNA, proteins or phospholipids, thereby inhibiting their intended functionality and disrupting cellular processes. As a secondary mechanism, ions can also catalyze the generation of reactive oxygen species (ROS) and induce oxidative stress[59].

In general, oxidative stress is another very important mechanism for NM toxicity[60]. Among others, the higher relative surface area can lead to increased surface reactivity[61]. NMs with high surface reactivity cause an excess in the production of ROS. While low levels of ROS can be counterbalanced by the cellular antioxidant system, higher concentrations of ROS cannot be sufficiently handled. In that case, ROS can damage cellular components like proteins or DNA or induce inflammation. It is worth mentioning here that oxidative stress and inflammation are adaptive responses of the organism which can be resolved over time. However, if they are not resolved and persist over longer periods, they may lead to manifested adverse effects like genotoxicity[62]. The OP of NMs will be further discussed in the section on physico-chemical properties in Chapter 1.7.

Another factor influencing toxicity is the morphology, as has been well-documented for fibre-like NMs. This is the best understood MoA for NMs. In general, the fibre toxicity paradigm[63] states that fibres exceeding certain thresholds with respect to their dimensions may cause asbestos-like responses like lung fibrosis and on the long term induce the severe cancer type mesothelioma. The biological changes related to inhalation of asbestos-like fibres are described in the Adverse Outcome Pathways AOP303 (https://aopwiki.org/aops/303), AOP409 (https://aopwiki.org/aops/409) and AOP171 (https://aopwiki.org/aops/171). AOP303 relates to the formation of lung cancer in general, while AOP409 and AOP171 focus on mesothelioma more specifically. AOP303 and AOP409 start with the induction of the so-called 'frustrated phagocytosis'[64-66]. After inhalation of asbestos-like fibres, macrophages are recruited as a defense system against the foreign matter. However, unlike particles or other small entities, asbestos-like fibres are too long to be engulfed by the macrophage and instead get stuck and lyse the macrophage. Thereby, different molecules are released, e.g., ROS or mediators like interleukins or tumor necrosis factor α[67]. This causes local inflammation as well as an unspecific immune response which again recruits other macrophages and immune cells. These cells fail to degrade the asbestos-like fibres causing a cyclic increase. Over time, this causes persistent inflammation, the formation of granuloma and pulmonary fibrosis and cancer in the lung or mesotheliom[68, 69]. AOP171 starts from persistent cytotoxicity inducing chronic inflammation and oxidative stress finally leading to the formation of mesothelioma. The World Health Organization

assumes fibres with length > 5 µm, diameter < 3 µm and aspect ratio (length/ diameter) > 3:1 to fall under the classical fibre toxicity paradigm which states that respirable, long and biopersistent fibres are carcinogenic[63]. For nanofibers, this classical fibre paradigm needs adaptation as it was observed that rigidity plays an important role in this case[70]. Very thin fibres are usually able to tangle and coil-up and thus behave like particles. Instead, thicker fibres stay unbent and thus if they also fulfill the length and width criteria, are assumed to show a fibre MoA. However, measuring rigidity and determining a suitable threshold is not trivial and at present still a matter of debate. For multi-walled carbon nanotubes (MWCNTs), a diameter of 30 nm has been defined as the cut-off between entangled and rigid forms[71].

Finally, NMs may also induce specific toxic effects which are distinct for the material under consideration. This specific toxicity arises from the specific physico-chemical properties of the NM or its interactions with the biological target. Often, this specific toxicity is related to effects in particular organs, e.g., liver or the central nervous system. For specific toxicity, a dose-dependency can usually be observed. As an example, some silver NMs or quantum dots have shown specific neurotoxic effects which can occur due to the fact that these NMs can cross the blood-brain barrier[56, 72, 73]. Other NMs based on zinc oxide, titanium dioxide or silver accumulate in liver cells thus inducing hepatoxicity[74-76].

1.3 NAMs

The number of new NFs on the market is rapidly increasing, all demanding for proper risk assessment. Theoretically, as the differences in physico-chemical properties may influence the toxicological profile, each NF would require a separate assessment with respect to its exposure, toxicokinetics, fate and (eco)toxicity. However, this is simply not feasible with respect to time, money and resources as well as the societal demands and legal requirements to reduce animal testing, given that, to date, this assessment still largely relies on *in vivo* studies to obtain reliable results. Another factor to be considered are potential inter-species differences in toxicodynamics and –kinetics which can result in uncertainties especially for negative results. This makes the use of alternative testing strategies, so-called NAMs, especially those focusing on mechanistic understanding of the processes underlying toxicity, unavoidable. Accordingly, a general paradigm shift from conventional animal testing towards NAMs can currently be observed. Under REACH, the use of NAMs is favored with respect to generating new data and should be used to avoid unnecessary animal testing. NAMs are

comprised of a variety of different *in chemico, in silico* and *in vitro* methods. This includes also high-throughput screening allowing to test multiple substances at a time as well as high-content methods like omics approaches.

1.3.1 *In silico*

In silico modeling gained more and more importance in recent years[77]. The idea of *in silico* approaches is to develop computational models or tools inferring or predicting unknown information based on existing knowledge obtained in laboratory-based studies. This is critical as with the growing amount of chemicals on the market, interpretation of complex datasets, reuse of data to fill-in data gaps as well as meta-analyses to obtain a more general understanding of toxicity patterns and their relation to structural features are of utmost importance. The advantages of *in silico* models are that they are cheap, easy to adapt and well standardizable. However, one critical point to be considered is that *in silico* tools typically depend on *in vivo* and *in vitro* data and thus their performance is highly dependent on the availability and quality of such data.

Under REACH, the use of *in silico* tools is strongly recommended. One frequently used *in silico* approach is quantitative structure activity relationship (QSAR) modeling. In QSAR models, the quantitative relationship between relevant physico-chemical properties of chemicals and their biological activity is determined by means of mathematical functions. Usually, this relationship only holds true for a certain class of materials called the applicability domain of the model[78, 79]. Most QSAR models, however, are only developed for simple toxicity endpoints like cytotoxicity as finding mathematical functions modeling complex endpoints sufficiently well is usually not feasible. For NMs, additional challenges arise due to their complexity and specialties which requires development of nano-specific descriptors and curated experimental datasets[80, 81]. Another frequently used tool is physiologically-based pharmacokinetic (PBPK) modeling which is a mathematical modeling technique for predicting the absorption, distribution, metabolism and excretion (ADME) of substances within the body[82]. PBPK modeling, thus, supports the prediction of systemic deposition or target organ exposure. In addition to these data-driven strategies, physics-based models are also useful tools supporting hazard assessment. Among the most frequently used physics-based modeling techniques is molecular docking which allows to predict the interactions between NMs and biomolecules based on existing knowledge on their three-dimensional structures. Its potential for toxicity prediction of various NMs has been shown in some case studies[83]. In addition, molecular dynamics simulations can be used to model time-dependent movements of atoms and molecules and to inform on

thermodynamic and kinetic properties at the atomic level. Thus, it can be used to simulate conformational changes of macromolecules and NMs. However, the tools are computationally very demanding and therefore, to date, only able to exemplify their potential use[84].

One of the main advantages of using *in silico* tools for risk assessment is that in contrast to *in vivo* and *in vitro* testing they may be used to predict potential hazards which means that an assessment can be performed prior to development and testing. Different ML techniques may be used to solve this task of predictive modeling. Thus, moving towards *in silico* tools means that it would be possible to predict and potentially also prevent hazards in safe(r)-and-sustainable-by-design (SSbD) approaches instead of just assessing them retrospectively. This may be useful for prioritization of NMs to be tested further as well as directly in regulatory decision making once the models are mature enough. To date, *in silico* tools are not accepted as stand-alone tools for regulatory purposes. However, they can be used as exploratory and supportive tools[85, 86]. As such they may be included in NAM frameworks like integrated approaches to testing and assessment (IATAs).

1.3.2 *In chemico*

In chemico approaches focus on the evaluation of potential hazards using chemical methods and reactions. Therefore, they assess the biological activity and potential toxicity of substances in an acellular environment. The focus of the corresponding methods is usually on some key mechanisms relevant for toxicity like the generation on ROS, the binding of proteins or the peroxidation of lipids. For the generation of ROS, various *in chemico* assays exist, e.g., the FRAS assay or the cell-free versions of the ESR and the Dichlorofluorescein assay. *In chemico* methods show various advantages compared to *in vivo* or *in vitro* approaches like simplicity, low cost and fast results. On the other hand, obtaining high correlations between *in chemico* methods and actual biological effects is still challenging, especially in case of NMs which have large potential to interact with the biological environment potentially leading to tremendous changes in their behavior.

1.3.3 *In vitro*

In vitro assays comprise studies performed outside living organisms using cell cultures. As with *in vivo* studies, different settings are possible in this context. Different cell lines as well as primary cells can

be treated with the test chemical to study its toxicity with respect to certain toxicological endpoints. Here, different cell models from different origin (human or animals) as well as different organs can be used. Cell lines may be used as single cultures or co-cultures using multiple cell lines at ones as well as in undifferentiated or differentiated state.

In a classical setting, cell cultures are exposed to chemicals under submerged conditions. Here, chemicals are dispersed in a medium and then applied on top of the cell culture. Usually, the dispersion medium is supplemented with 10% fetal calf serum. For NMs, this poses a challenge with respect to particle identity as it will lead to the immediate binding of proteins present in the serum and thereby to the formation of a protein corona. An alternative way for exposing cells to NMs is the use of an air liquid interface. Here, particles are applied in form of aerosols and thus the challenges of interactions between serum and particles as well as the agglomeration of particles can be circumvented. However, this also does not reflect reality as under physiological conditions NMs would interact with body fluids like pulmonary surfactant.

Various *in vitro* assays exist and may be used to study the influence of NMs on cells. The challenge, however, is that *in vitro* assays usually only describe relatively simple effects like cytotoxicity or the generation of ROS. Specifically, due to the short lifespan of cells, *in vitro* tests only represent acute effects of NMs while outcomes of chronic exposure cannot be assessed. Also factors like biodistribution, bioaccumulation and excretion cannot be represented sufficiently well.

One specifically powerful NAM in the field of *in vitro* methods are omics techniques. The term omics refers to studies measuring changes in different biological molecules or components, e.g., genes in genomics, mRNAs levels describing gene expression in transcriptomics, proteins levels in proteomics or metabolite levels in metabolomics. Data from omics measurements give detailed insight into molecular changes and can be used to identify key molecular targets, pathways, and biomarkers associated with toxicity. Omics techniques exhibit various advantaged compared to conventional toxicity studies: 1) Omics yield high-content analyses with many molecules being measured in one single run and can be easily scaled to be used in high-throughput approaches with costs steadily decreasing. Thus, they allow for rapid and cost-effective analysis of biological changes; 2) They are not restricted to one single endpoint but instead allow for detection of all changes happening within the cells or tissue at ones and thus, under the very same conditions. Therefore, concerted effects from different signaling pathways may be observed which otherwise would stay undetected; 3) They enable early detection of changes before the actual outcome manifests meaning at earlier time points as well as at lower doses. This also helps avoiding the use of extrapolations from very high-

dose apical endpoints which are typically used in *in vivo* studies[87]. This is important as it has been shown that in many studies tested concentrations are unrealistically high leading to altered toxicity mechanisms which are not relevant in real exposure scenarios[88, 89]; 4) They also make biomarker identification possible which can help in early and easy detection of induced effects; and 5) Omics data can be mapped to AOPs and can be used to gain detailed insights into toxicity mechanisms induced by a treatment[90]. While it already became clear that omics approaches are valuable tools in the light of toxicological research, their use in regulatory toxicity is still not accepted mainly due to missing standardization of measurement and analysis techniques for omics data as well as challenges in data interpretation.

In general, the best-studied omics layer is transcriptomics. For transcriptomics, comprehensive databases[91] are available and the evaluation of such datasets is comparatively well standardized. In addition, a reporting template for transcriptomics has already been introduced by the organization for economic co-operation and development (OECD)[92, 93]. Other omics layers like proteomics or metabolomics have been less explored in the past. However, due to their closer relation to the actual phenotype, their use is steadily increasing. For proteomics, the number of available raw data in databases like the PRoteomics IDEntification (PRIDE) database[94] is rapidly growing. While for metabolomics, the number of available datasets is still comparatively low, an OECD reporting template similar to the one for transcriptomics has been developed[93, 95]. As each omics technique only explains part of the induced changes in the cells and has its own advantages and challenges, the most complete and representative picture is yielded if different omics layers are combined in a systems biological approach. At the same time, this also strengthens the reliability of the results as it allows to separate noise from findings with actual relevance. This global view is often referred to as toxicogenomics. While systems toxicological approaches are highly relevant, they are challenging with respect to time, cost and complexity and thus not well suited for general screening approaches. However, they are very useful for gaining mechanistic insights supporting regulatory strategies like grouping and read-across as well as SSbD approaches.

1.4 NAM Frameworks

For rather simple endpoints like skin or eye corrosion/ irritation, NAMs can already reliably replace animal testing[96]. However, for higher tier endpoints which are more complex like carcinogenicity or reproductive toxicity, suitable NAMs are only emerging. In these cases, NAMs may still be useful for

reducing animal testing by integrating them in NAM frameworks. Among the most frequently used NAM frameworks are so-called IATAs. IATAs combine various assays and toxicological tests which, in combination, may be suitable to describe complex endpoints. Thereby, animal studies may be combined with different NAMs, e.g., *in vitro* or *in silico* approaches. One critical factor for the successful implementation of IATAs is obtaining knowledge on underlying MoAs which are closely connected to AOPs. In addition, grouping and read-across may be used for waiving testing for certain materials. The different concepts are described in more detail below.

1.4.1 IATAs

Especially for higher tier toxicological endpoints, it is necessary to combine multiple assays to sufficiently describe the single steps leading to the respective outcome. One concept commonly used to set up larger test batteries for complex endpoints are IATAs[97]. IATAs integrate all relevant information on certain aspects of the induced biological effects and weigh them in a weight-of-evidence approach. It can also be used for targeted generation of new hazard data avoiding unnecessary testing. IATAs usually include not only *in vitro* tests but also information from *in vivo* studies or *in silico* models as well as physico-chemical properties or human data and arrange them in a structured manner usually resulting in a specific decision tree. The most comprehensive battery of IATAs for NMs has been developed in the EU project GRACIOUS[98]. In order to determine, which events and tests need to be included in a specific IATA, mechanistic understanding of the underlying biological changes is urgently needed. Especially for NMs whose behavior is very complex, it has proven that integration of mechanistic knowledge is necessary to obtain reliable models[99]. At the same time, this is also likely to improve regulatory acceptance of NAMs. Thus, IATAs are closely connected to MoAs and AOPs.

1.4.2 AOPs

In order to develop reliable alternative methods and IATAs, a solid mechanistic understanding of the underlying biological changes induced by the treatment with a certain chemical or NM is of great advantage. This is especially needed for higher-tier endpoints with complex biological processes being induced. Knowledge on the MoA of substances can be used to support the development and refinement of the closely related AOPs.

AOPs have first been described by Ankley et al.[100] and then adopted by the OECD. AOPs are conceptual frameworks that relate different biological events at various levels of biological organization to an adverse effect induced by a substance where events are arranged in a causally linked, sequential order. AOPs consist of a molecular initiating event (MIE) followed by multiple key events (KEs) ultimately leading to the adverse outcome (AO)[101]. The MIE describes the initial interaction between the substance and a biological target. Subsequently, several KEs may be induced on sub-cellular, cellular, tissue, organ or even whole-body level. The AO represents the final outcome on organ or whole-body level, e.g., lung fibrosis or cancer. The AOP concept is shown in Figure 2. Various AOPs have been described for a number of different AOs and can be found in the AOP-Wiki (https://aopwiki.org/). The first AOP that has been released by the OECD describes skin sensitization[102]. However, nano-specific AOPs are only beginning to emerge[103-106]. One may hypothesize that MIEs could potentially vary between NMs and conventional chemicals while the sequence of KEs may be expected to be the same. However, this hypothesis has to be tested in future approaches. More recently, the concept of AOPs has been extended to quantitative AOPs (qAOPs) which allow not only for qualitative but also for quantitative hazard assessment[107, 108]. This can be achieved using approaches like weight-of-evidence, probabilistic or mechanistic models. Overall, it is expected that AOPs may facilitate extrapolations between species as well as from *in vitro* to *in vivo* testing. In order not to stay at the conceptual level, the actual described changes must be measured in real-life scenarios. One of the most promising tools to do so is by means of omics approaches describing the single KEs[109, 110].

Moreover, information on the MoA should be included into NM grouping approaches in order to make them reliable. Predictive toxicogenomics offer great opportunities with respect to establishing such NM grouping approaches or substantiating existing NM grouping hypothesis based on mechanistic knowledge. This knowledge may be obtained from newly conducted experiments as well as from literature and database searches.

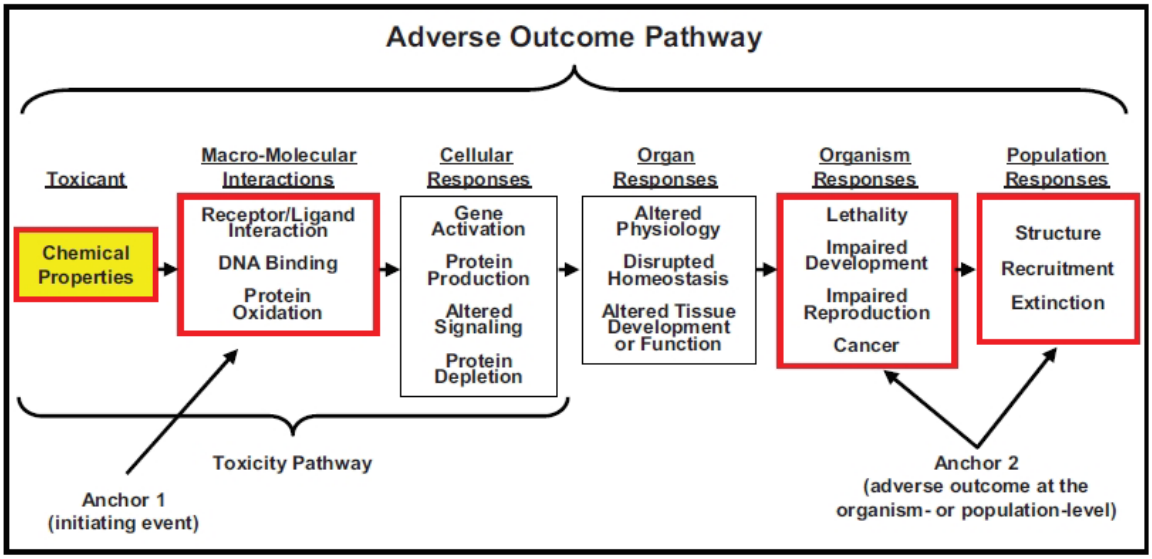


Figure 2: Schematic representation of the AOP approach. The figure is reprinted from Ankley et al.[100]

1.4.3 Grouping

Risk assessment for NMs requires detailed characterization of each single NF with respect to physico-chemical properties, hazard and exposure. As this assessment is simply infeasible in a reasonable amount of time and resources for the large number of existing and emerging NFs, alternative methods are urgently needed. In a regulatory context, grouping and read-across are frequently used to justify the waiving of specific tests or to fill-in data gaps through read-across[102]. In addition, these approaches may also be used in the context of prioritization for further testing, for ranking of NMs with respect to a certain toxicological outcome or in weight-of-evidence approaches as well as for supporting SSbD. Guidance documents for grouping and read-across in general[111, 112] but also specifically for NMs[113] are available.

Grouping is defined as 'the general approach for considering more than one chemical at the same time'[111]. Two main approaches are described for grouping: the category and the analogue approach[111, 114, 115]. In the category approach, chemicals for which the physico-chemical and (eco)toxicological properties are likely to be similar or follow a regular pattern due to structural similarities are grouped together. Trends or patterns within one category should thereby be identified across several materials in a consistent manner. Thus, establishment of a robust category should include a sufficient number of chemicals. Instead, if only a small number of chemicals is investigated and no clear trends can be observed, the analogue approach may be more valid. In that case, structural similarity of chemicals is still the key factor for grouping but additional expert

judgement is needed to identify whether the grouping hypothesis is still valid in the specific case under consideration. Importantly, grouping is endpoint-specific and thus established groups are not globally valid. For each hazard endpoint, a specific grouping hypothesis has to be formulated. This hypothesis needs to include an explanation on which key properties are important for establishing a group and how these are linked to the toxicological endpoint. The grouping hypothesis then needs to be justified in case studies. The most hazardous substance within one group should thereby always be tested in order to set the upper boundary.

Once, a category is established or a suitable analogue is identified, data gaps within this group can be filled. In the regulatory field, this is usually done using a read-across approach. However, also other tools like QSAR or trend analysis are possible options to fulfill this task. Read-across uses information from a data-rich source substances to predict the same property or endpoint for one or several data-poor target substances within the same group for which this data is missing. Thereby, read-across avoids testing each and every single variant thus saving resources. Robust and valid read-across can only be guaranteed in case of a clear underlying rationale for the established grouping. Comparisons to benchmark materials can also be used to reduce uncertainties.

Therefore, the main question underlying every grouping activity is which structural features actually relate to physico-chemical properties or (eco)toxicological events and thus should be used as the basis for establishing a group. While for conventional chemicals, usually common functional groups, precursors or breakdown products are good candidates for potential similarities, for NMs the situation is more complex as there is a much larger variety in factors that may play a role in this regard and their exact structure is not known. The main difficulty is posed by the huge number of physico-chemical properties which could all potentially affect NM toxicity. In addition, changes in different media due to binding of biomolecules or agglomeration as well as aging effects over the whole life cycle need to be taken into account. The understanding on how changes in all these properties relate to toxicity, toxicokinetics as well as uptake behavior is only beginning to emerge.

For NMs, REACH defines two goals for grouping: 1) Grouping can be performed to identify sets of similar NFs. In that case, similarity of NFs has to be shown in order to justify that hazard and exposure assessment can be performed jointly[116]; and 2) Grouping can also be performed for the purpose of read-across similar to conventional chemicals. A key factor of major importance for NM grouping is the unambiguous identification, characterization and naming of NFs. One difficulty here is that many physico-chemical properties can vary and therefore need to be assessed to find out whether two batches actually comprise the same NF or not. As for conventional chemicals,

properties like chemical composition, degree of purity and quantitative information on impurities or additives form the basis for NM characterization and identification[21]. In addition, the description of a NF also requires information on the number-based particle size distribution, surface functionalization or treatment, shape in terms of aspect ratio and particle morphology as well as the specific surface area[27, 28]. On top of that, physical properties like dissolution rate, state of agglomeration or aggregation and changes in surface chemistry as well as other higher-level parameters such as surface reactivity may also be relevant. While some of these properties are intrinsic to the NM itself, others vary depending on the medium surrounding the NM and are thus extrinsic. Therefore, in order to obtain a reliable grouping, characterization has to be performed in the relevant biological medium used for toxicity testing as well.

While a number of physico-chemical properties may influence NM toxicity, no simple linear correlation between these factors and the toxicity outcome could be observed as each of these properties can influence uptake, toxicokinetics and/or (eco)toxicity in a complex way and also interdependencies between the properties can be observed[115]. In addition, uncertainties in measurements of physico-chemical properties and toxicity are still high for NMs. While some test guidelines have already been adapted to NMs, this is still an on-going process of standardization and validation. Especially for extrinsic properties, measurements are difficult to perform as they need to be carried out in complex media. Due to these uncertainties, reliability and comparability of existing datasets is still one of the largest bottlenecks with respect to developing robust grouping approaches. Thus, grouping for NMs remains a major challenge where simple structure-activity relationships cannot be established easily.

ECHA has also generated a guidance document on grouping and read-across between NFs, or between NFs and non-NFs of the same substance[28]. The ECHA guidance clarifies the need to consider similarities of not just physico-chemical properties like aspect ratio, particle size, shape or solubility, but also toxicokinetic behavior and fate, as well as (eco)toxicological behavior between different NFs. The guidance indicates that it is possible to use physico-chemical parameters and/or *in vitro* screening methods to develop a robust scientific explanation of why different forms of the substance are sufficiently similar to be grouped when considering their hazard[117]. Although this guidance addresses only read-across for different forms of the same substance, it does not preclude read-across between NFs of different substances. In addition, identifying NMs sharing a common MoA aids the justification of a grouping hypothesis.

Due to the complexity of the task of NM grouping, various grouping frameworks have been described in past which build on each other and have constantly been improved over the years[118-121].

Comprehensive overviews on different grouping and risk assessment framework were given in Oomen et al.[119] and Giusti et al.[115]. The most recent and most advanced grouping framework for NMs was published as a result of the EU project GRACIOUS[98]. This framework supports grouping and read-across by identifying suitable hypothesis describing key similarities between NFs while including relevant physico-chemical characteristics, route of exposure and hazard endpoints. Within the framework, 40 pre-defined grouping hypotheses were generated based on well-defined toxicokinetic pathways or MoAs[122]. Along with the grouping hypotheses, the GRACIOUS framework also gives advise on how to test each of them. In order to do so, each grouping hypothesis is coupled to a specific IATA guiding the gathering of evidence to test the hypothesis and to determine whether or not different NFs can be grouped together. Within IATAs, evidence from different sources like literature, *in silico*, *in vitro* or *in vivo* assays [123] are assembled and arranged in decision trees that support the decision whether the grouping hypothesis should be accepted or rejected. At the same time, IATAs also guide the decision on which new data may need to be generated in order to fill data gaps. In addition, the framework also comprises a template which can be used to generate own hypothesis. This hypothesis-driven approach is essential for aligning with the REACH requirements and sets the scientific justification for the grouping decision. The collection of some basic information on the NM identity and the intended use allows choosing a suitable grouping hypothesis and the associated IATA. The grouping hypothesis of interest can then be tested on a provisional group of NFs. The IATA enables acquisition of suitable evidence for acceptance or rejection of the grouping hypothesis. The GRACIOUS framework considers intrinsic as well as extrinsic properties of NM and is strongly focused on providing practical guidance for grouping and read-across instead of staying on a conceptual level. It also considers different exposure scenarios, applications and stages of life cycle to obtain a holistic picture. Wherever possible, also IATAs use well-established methodologies or are aligned with OECD testing guidelines to enhance standardization.

While case studies[121, 124-130] have shown that the application of the developed grouping frameworks is useful, it also became obvious that usage of these theoretical frameworks is not that easy and fails in many cases. Especially the fact that no simple correlations can be observed between physico-chemical properties of the NMs and the outcome of *in vivo* or *in vitro* experiments, renders grouping a difficult task. This missing link may be due to the fact that multiple physico-chemical properties influence the toxicological outcome in a non-linear manner. Thus, the bottleneck for making grouping frameworks applicable for read-across is to determine the set of descriptors that is most predictive for the toxicological outcome under consideration as well as the most suitable measurement techniques for reflecting these descriptors. Descriptors in this case may be simple physico-chemical properties of the NMs but also other higher-level descriptors like surface reactivity

or biological descriptors like those derived from omics techniques. ML and bioinformatics tools may aid refinement of the grouping frameworks by identifying the most relevant descriptors for NM grouping.

1.5 Challenges for NAMs for regulatory application

The development of suitable NAMs, is an ongoing process which still poses various challenges that need to be tackled in order to increase their suitability for regulatory purposes. First and most importantly, NAMs need to predict biological systems as close as possible in order to be useful for safety assessment. Due to the large complexity of biological systems, this is often a major hurdle. In addition, standardization and validation of methods are key factors required to ensure reliability and reproducibility of the results. This is directly related to regulatory acceptance of such methods. Also, data interpretation and integration need to be harmonized and technological limitations need to be acknowledged.

For NMs, the situation is even more challenging as summarized before[131] and successful implementation of NAMs requires adaptation of existing test guidelines for conventional chemicals which is currently still an on-going process led by the OECD. The most relevant challenges with respect to interpretation of findings and comparison between different NMs as well as for extrapolations from *in vitro* to *in vivo* to be mentioned in this context are NM dispersion stability, dosimetry and interactions of NMs with their biological environment. In the past, various organizations like the OECD and research projects like NANoREG, NanoHarmony, GRACIOUS or SmartNanoTox have tackled these challenges by developing and standardizing specific methods for characterization, handling and hazard assessment for NMs. The different challenges are described in more detail below.

First of all, dispersion and dispersion stability over time is a critical factor in the context of NMs. Dispersion refers to the ability of particles to uniformly distribute within a medium leading to a homogeneous suspension. The process of NM dispersion usually involves breaking down aggregates into smaller entities or ideally single particles and spreading them throughout the medium. Frequently, NM dispersion is achieved by using sonication techniques based on the application of ultrasonic waves. Dispersion stability describes the ability of NMs to maintain this dispersed state over time. Dispersion is a critical factor for successful implementation of toxicological studies.

Especially in *in vitro* studies, it is crucial to apply stably dispersed NMs to the cells as otherwise proper exposure of the cells to the NMs cannot be guaranteed and results may not be representative as well as reproducible. How well a NM can be dispersed and how stable this dispersion is, depends on the physico-chemical properties of the NM under consideration. As an example, coatings on the surface of NMs can prevent agglomeration by reducing interactions between particles. Also, zeta potential is critical for the dispersion of NMs with high zeta potential generally indicating greater dispersion stability due to higher electrostatic repulsion between particles. Stabilizers may be used in addition to reduce agglomeration. The quality and stability of the dispersion should always be verified during experiments using characterization techniques like Dynamic Light Scattering. As a first step, test guidelines for inhalation toxicity[132, 133] and NM dispersion[134] have been updated by the OECD. However, standardizing measurements for NMs is still an on-going process.

In addition, doses can be measured and compared using different metrics. Commonly, doses are given on mass basis using units like mg/ ml or mg/ kg body weight. However, due to differences in size, same doses with respect to mass may result in a largely differing number of particles contained. At the same time, this means that overall larger surface areas are available for reactions to take place in case of smaller particles. Therefore, it has been hypothesized that comparisons based on same surface doses are more appropriate especially for assays measuring surface-related properties like reactivity or dissolution[135-137]. In that case, doses will be given in mg/ m^2. In addition, the dose per cell is also important implying that the seeding density should be considered when comparing findings[136, 138, 139].

Another critical factor with respect to dosimetry is the effective dose. *In vitro* most variation between different types of NMs exists for cell cultures with adhesive cells under submerged conditions. In that case, NMs are dispersed in medium and then applied to cell culture dishes. Over time, NMs start sedimenting to the bottom of the cell culture dish where they may be in contact and potentially taken up by the cells. However, how fast NMs sediment again depends on their physico-chemical properties, mainly their density and their agglomeration state. Within an experiment, these differences usually cannot directly be influenced or taken into account and thus all results will be measured at the same time point leading to varying amounts of particles having reached the bottom. If one then compares NMs with different sedimentation speed, the results may not be representative as in one case more particles would have reached the cells. The deposited dose may be measured directly using analytical ultracentrifugation. However, the equipment for these techniques may not be available in many laboratories. Therefore, computational models have been developed to compute correction factors for this situation. Hinterliter et al.[140] have developed the *In vitro* Sedimentation, Diffusion and Dosimetry model which is mainly based on sedimentation and diffusion

of the particles to determine how many particles reach the bottom of the dish within a certain time. This model has been refined to also take into account agglomeration and dissolution[141]. The main limitation with this model is that it does not consider desorption from the cell membrane which may be the case for particles showing only weak adhesion or are slowly taken up by the cells. The Distorted Grid model introduced by DeLoid et al.[142] overcomes exactly this limitation by introducing an additional parameter for the absorption strength and allowing for particle diffusion back to the upper layers. Another factor to be considered is that the fact that particles have sedimented, does not automatically imply that they are also taken up by the cells. In many cases particles will only stay attached to the cell surface instead of being internalized. However, usually only particles which are taken up into the cells induce biological responses potentially leading to AOs.

Similarly, internal doses differ from the applied dose in the *in vivo* situation as well. Looking at exposure via the inhalative route, whether or not NMs reach the lower respiratory tract and deposit within the airways again depends on their physico-chemical properties. Higher deposition efficiency of particles also increases the chances of penetration through the alveolar epithelium into the bloodstream and thus systemic availability[143-145]. Oberdoerster et al.[145] have shown that the deposition in the respiratory tract strongly depends on size differences. This size-dependent deposition is also considered within the Multiple Path Particle Dosimetry model which can be used to predict particle dosages in human and rat airways[146-148]. In addition, PBPK modeling is important for estimating the exposure of secondary organs as it predicts the ADME properties of substances within the body.

In addition to these general challenges for the estimation of effective doses of NMs, other factors are also critical for comparison of *in vitro* and *in vivo* results and may hamper the successful development of reliable *in vitro* methods. Among them, the binding of biomolecules from the surrounding medium can influence cellular uptake and thereby potentially also toxicity. Thus, not only the proper characterization of the pristine NM but also the characterization of relevant biological fluids plays an important role. In addition, NMs are known to interfere with spectrophotometric read-outs or enzymatically-catalyzed reactions which renders such methods inappropriate for NMs[149, 150]. Another challenge is the limited availability of *in vivo* reference data for NMs, which hampers proper validation of newly developed NAMs. In addition, the limited understanding on how variations in physico-chemical properties relate to changes in toxicity poses another challenge in the development of reliable NAMs. This challenge may be overcome by detecting underlying MoAs leading to NM toxicity which is one of the most promising approaches for developing reliable NAMs. While great advances have already been achieved over the last years, the

understanding of the biological effects of most NMs is still limited. Further enhancing this knowledge will enable the development of robust NAMs allowing also for prioritization of NMs which require further testing, generating reliable grouping hypotheses and the support of SSbD approaches. Bioinformatics and ML bare great potential for supporting the development of reliable NAMs and NAM frameworks and for unraveling MoA for NMs.

1.6 Bioinformatics and ML models

Bioinformatics and ML tools show great potential to support current risk assessment and for establishing new safety assessment paradigms in various ways. The main advantages of these tools are their capabilities for analyzing large and complex datasets, predicting potential hazards and providing valuable insights into MoAs thereby supporting risk assessment. Bioinformatics is an interdisciplinary field which uses various mathematical and computational methods for interpreting and analyzing biological data. Among those methods, there are various techniques from statistics and ML. In addition, omics data analysis forms one central pillar of bioinformatics. The most important concepts and techniques within these three fields are described below.

With respect to NM grouping, bioinformatics tools can be applied in different ways. First, ML methods may be used to predict toxicity outcomes by linking these outcomes to physico-chemical properties. This can be achieved by the ability of ML models to quickly analyze vast amounts of data in a systematic way which might unravel hidden patterns in the data that cannot easily be detect by humans. In addition, omics analyses may be coupled with these predictive models to gain knowledge on the underlying toxicity mechanisms thereby improving the predictive ability of the models. Finally, bioinformatics tools may also be employed to integrate and manage information on NMs from different sources or databases. Altogether, these tools and integrated data may then aid regulators, researchers or industry in decision making with respect to safety aspects of NMs.

The most critical challenge for modeling approaches in the context of NMs is the scarcity of existing data. At the same time, the most relevant question to be answered is the one of suitable descriptors of NM toxicity. These descriptors might be physico-chemical properties which can directly be influenced during design and production but also other surrogates which can be easily measured and potentially used for regulatory purposes at some point. Therefore, computational models should as much as possible try to fulfill the following criteria: 1) They should be able to cope with limited data availability. 2) They should be able to handle non-linear relationships. 3) They should not show high

risk of overfitting. 4) They should allow for feature selection. 5) They should be understandable and interpretable. 6) They should be easy to use for experts in the field. 7) They should not be highly dependent on changes of hyperparameters[151].

1.6.1 Data collection and pre-processing

The first step in any computational approach is to assemble a dataset containing the necessary information. Therefore, usually data are collected from literature or one of the various available databases. In addition, new data directly obtained from experimental work may also be included in the dataset. For grouping purposes, these datasets should contain information on the physico-chemical properties as well as on the toxicological endpoint under consideration. In addition to the pure numerical or qualitative values of the measured property or assay, the dataset should also contain information on metadata describing the experimental conditions under which the values were obtained. This might include information on whether results stem from *in vivo* or *in vitro* testing, which species or cell lines were used, which doses and time points were studied and so on. Even though this metadata might not be used directly in the model, it will still inform on whether results are comparable or not and will aid interpretation of the modeling outcomes.

After collecting all the data, the dataset needs to be pre-processed in order to be usable for computational modeling. Pre-processing methods include techniques for feature reduction and feature selection, normalization, transformation, imputation, class balancing and others[152]. Which of these methods are needed in a specific case depends on the type of input data as well as on the algorithms applied to the data. Thus, not all steps are necessarily performed in all analyses.

Feature reduction can be used to remove irrelevant or redundant information. This usually refers to variables which are non-informative due to low variance, a high number of missing values or high correlation with other variables in the dataset. In addition, feature selection may be useful for reducing the number of considered variables in order to avoid overfitting of the model[153]. Overfitting describes the situation in which the model learnt the underlying pattern of the data in too much detail such that it is not able to generalize well to new data anymore. This typically happens if the number of considered samples is relatively small while the number of variables with potential influence on the outcome is rather large. This is the typical situation in NM grouping approaches. Therefore, the number of parameters considered for modeling needs to be carefully chosen and adjusted using feature selection.

In addition, transformation and normalization are performed in order to scale variables or measurements such that they are comparable to each other. Log_2 transformation is commonly applied to remove skewness of the data, to compress large dynamic ranges into linear patterns and to equalize variances. Especially for fold changes in omics data analysis, interpretation of the data may also be easier due to linearity which means that a two-fold upregulation would result in a value of +1 and a two-fold down-regulation in a value of -1 instead of 2 and 0.5, respectively. In addition, different normalization techniques are available. Z-score normalization is one of the most frequently used methods and standardizes the data to have a mean value of zero and a variance of one. While z-score normalization makes variables comparable and is robust to outliers, the shapes of the distributions are not necessarily conserved which may make interpretation more complicated. Another option is min-max-normalization which scales the data to a fixed range, typically between 0 and 1. In contrast to z-score normalization, this method preserves the shape of the distribution and restricts the values to concrete boundaries. However, it does not work well if the dataset contains outliers. In the case of omics data, median normalization and quantile normalization are very prominent methods. They are particularly useful for comparing samples from different experiments or platforms which may have different intensity distributions due to systematic biases. In median normalization, the distributions of all samples are shifted to have the same median value. In quantile normalization, data points for each sample or column are ranked from the smallest to the largest value, then replace by the row mean and re-order to the original order. Thereby, it is guaranteed that all samples show exactly the same distributions instead of just same median values. The advantage of quantile normalization is that it is more robust to outliers, however, if the extreme values are of biological relevance, this information may not be preserved. Therefore, normalization is a critical factor in data pre-processing and no universal recommendations fitting all datasets may be given.

Depending on the aim of the model, discretization especially of the outcome variable might also be necessary. In that case, numerical values will be binned into discrete ones representing a certain range of values. As an example, numerical values of a specific assay may be transferred into binary format setting a certain threshold. All values below that threshold would then belong to the first class and values above the threshold represent a second class. These classes may, for instance, represent NMs which are ‘active’ or ‘passive’ with respect to a certain endpoint. In this setting, also class balancing may play a role. This is the case, if one class is substantially overrepresented meaning that its sample size is much larger compared to the second class. For NMs, this is frequently observed as usually there are more non-toxic than toxic compounds. In that case, the trained model is often biased towards the majority class which means that new samples will be identified as belonging to the majority class with higher probability independent of their nature. Therefore, in case of highly

imbalanced datasets, methods handling this problem should be applied. Here, resampling methods can be useful for oversampling the minority class by adding more instances or undersampling the majority class by removing instances. Frequently, the SMOTE algorithm[154, 155] is applied for this task.
In addition, many datasets contain missing values which cannot be handled by frequently used methods like PCA. In that case, there are two main possibilities to handle this problem. One option is to simply remove variables containing any missing values. However, in most cases this will lead to a huge loss of information. Alternatively, one can use imputation. During imputation, missing values are replaced by those predicted or estimated from the available information in the dataset. Different imputation techniques with different objectives as well as strengths and limitations exist. One option is to replace missing values by the mean or median of the available values for this variable. However, this produces the same value for all missing values of that variable and thus leads to non-normally distributed data. Here, RF imputation[156] may aid by adding random noise. Another technique frequently used in kNN imputation[157] where missing values are estimated based on the values of the k closest data points. However, in all these methods it is expected that missing values are in the same range as the existing ones. However, especially for results from experimental techniques, another option is that values are missing because they fall under the limit of detection or the limit of quantification. In such cases, missing values are expected to be very small and therefore other imputation strategies should be used, e.g., imputation by drawing values from a down-shifted, shrunk normal distribution. Independent of which imputation method is used, potentially introduced uncertainties should always be kept in mind during further analysis.

Once all the relevant data is collected and pre-processed different statistical approaches or modeling techniques may be applied to it.

1.6.2 Statistics - Important methods and measures

Various statistical methods and measures are of importance in the context of this thesis. First and most basic is statistical testing. Statistical testing is used to determine whether observed differences or relationships in the data are likely to be real or whether they just result from variations in the normal error range of the measurements. Therefore, a null hypothesis (H0) is defined which can either be accepted or rejected during statistical testing. The test statistic is a numerical value summarizing the distances between groups compared in the test. Calculation of the p-value is the main criterion for determining whether the observed differences are statistically significant. It represents the probability of obtaining results as extreme as the ones observed in the test statistic

when assuming that H0 is true. Thus, it tells how likely it is to obtain the observed result if in reality there is no difference in the populations. Small p-values suggest that observed result are rather unlikely under H0 and thus there actually is a difference between groups. P-values smaller than a certain significance level, commonly smaller 0.05, are considered as statistically significant and H0 is rejected. In other words, the probability that H0 is actually correct is lower than 5% in that case. Which statistical test needs to be chosen, depends on various factors like the type of data meaning whether variables consist of continuous or categorical values, the number of groups being compared or the study design which relates to independent or paired samples. In addition, the assumptions which can be made on the data play an important role for choosing the most suitable statistical test.

The most frequently used statistical test is the t-test. The t-test is a parametric test comparing the means of two sample populations. As it is a parametric test, it assumes that the data in the dataset is normally distributed. In addition, the variances of the two samples should be approximately equal. Samples used in the t-test can be either independent if measurements are obtained from different populations or paired if data from the same or matched individuals are used in both groups. If the sample size is small, often one does not know whether the underlying distribution is really normally distributed as too few data points are sampled. In those cases, the use of non-parametric tests may be more appropriate. One such test is the Wilcoxon rank-sum test. The Wilcoxon test can be used for continuous but also for ordinal data. It ranks the combined data from both groups and computes the sum of the ranks in each group. These sums are then compared to calculate the test statistic. The Shapiro-Wilks test or Quantile-Quantile plots can be used to test whether the data is actually normally distributed. Other statistical tests also exist but will not be discussed further here.

For high-dimensional datasets, multiple hypotheses are tested simultaneously. This is the case for omics data where the differential status of each transcript, protein or metabolite is assessed. Assuming a significance level of 0.05, if 100 tests are performed one may expect to obtain five false-positive results just by chance. As for omics data usually multiple thousand molecules are assessed, this would result in a substantial number of false-positive results. Therefore, multiple testing correction has been introduced. One of the most reliable and frequently used methods is computing the false discovery rate (FDR) using the Benjamini-Hochberg method. This method sorts the p-values in ascending order and divides each observed p-value by its percentile rank. Then again, a significance cut-off, usually 0.05 or 0.01, is set.

Another important statistical measure is correlation. Correlation measures the strength and direction of the relationship between two variables. Pearson correlation thereby considers linear relationships

between continuous variables. Values range from -1 in case of a perfectly inverse relationship over 0 which indicates no linear relationship at all to 1 for perfect linear relationship in the same direction. While Pearson correlation is very informative, it has some limitations, namely it assumes that the data is normally distributed and it is sensitive to outliers. In the same manner as for statistical testing, a non-parametric alternative is available. Spearman correlation measures the strength and direction of monotonic relationships. This refers to consistent increase or decrease of one variable along with the other one. Again, values of the Spearman correlation coefficient range from -1 to 1 where positive correlation coefficients indicate positive monotonic relationships, and negative correlation coefficients relate to negative monotonic relationships. Pearson correlation is also useful when working with ordinal data.

1.6.3 ML techniques

In addition to statistical methods, also various ML tools exist. ML is a subset of artificial intelligence (AI) focusing on models and algorithms enabling computers to learn patterns from given data and make prediction based on those for new datasets in an exploratory manner. Usually, a set of descriptors that may be relevant for the outcome is used as features or input variables. In case of NM grouping, these features may be physico-chemical properties, omics data or some surrogate measurements. The outcome variable might relate to information on a certain toxicity endpoint. In general, ML models can be divided into supervised and unsupervised methods. Supervised methods thereby learn from labeled data while unsupervised methods simply detect patterns in the data without any prior knowledge on class membership. Depending on the nature of the outcome variable, ML methods are separated further into classification models with categorical outcomes and regression models in case of a continuous outcome variable. Model development in ML is generally divided into two phases: a training and a testing phase. In the first step, the data is therefore split into a training and a test set. The model is than fitted on the training data in order to learn any underlying patterns by adjusting its parameters. In the subsequent testing phase, the performance of the model for unseen data is accessed to judge its generalizability. The most important challenge in this regard is finding the right balance such that neither overfitting nor underfitting occurs[158]. In case of underfitting, the model does not represent the training data very well as it is too simple. In case of overfitting, the training data is learned too well and in too much detail and thus cannot generalize to new data. Once a model is trained and tested and shown to be robust it can be deployed for real-world data and applications. Usually, this is an iterative process in which the model is improved over

time. Feature engineering can be performed to select or transform features thereby improving model performance.

Various ML algorithms are available (see Figure 3). The most common and relevant ones in case of NM grouping approaches will be briefly described here. Unsupervised ML techniques comprise tools for dimensionality reduction and clustering. One of the most frequently used technique is PCA. PCA transforms a set of potentially correlated variables into linearly uncorrelated ones. These new variables are called PCs. By analyzing the coefficients of the PCs, one can determine the importance of the different descriptors. In the case of NMs, the materials may be represented in a lower-dimensional space based on the first few PCs. Each PCs would be a linear combination of all assessed physico-chemical properties. The properties related to the highest coefficients are then assumed to be of highest relevance. In addition, cluster analysis may be used to organize objects based on their similarity with different cluster algorithms being available. Most frequently, hierarchical clustering is used where the two nearest clusters are combined into a new common cluster in an iterative fashion building up a tree-like structure. Typically, Euclidean distances are used to compute similarity of clusters. However, additional distance measures exist and may even be more useful. Apart from that, exclusive clustering is also frequently used. Here, data points are forced to end up in separate clusters. A common example is k-means clustering where all data points are assigned to one of k cluster points. Fuzzy clustering instead allows for multiple class memberships each with a certain degree of membership probability. Overall, clustering is a useful tool for exploratory data analysis and may provide insights into similarity of NMs.

In addition, various supervised techniques exist as well. Among them, RFs[159] are of special interest for NM grouping as they are a non-parametric technique which can handle small datasets relatively well with comparatively low risk of overfitting and at the same time holding inherent methods for feature selection. RFs combine multiple binary decision trees built on bootstrap samples of the original data. Each decision tree makes splits based on the explanatory variables in a way that they separate groups in the outcome variable as well as possible. With respect to feature selection, stepwise removal of the most unimportant features may be performed based on inherent criteria like the Gini impurity or the mean decrease of accuracy[160].

Linear and logistic regression are important techniques modeling relationships between one dependent outcome variables and one or more independent predictor variables. In linear regression, the aim is to find the best-fitting linear equation describing this relationship between dependent and independent variables. The outcome variable thus needs to be continuous. Slope and intercept the

best fitting line are estimated by minimizing the sum of squared differences between the observed and predicted values. Instead, logistic regression is used if the outcome variable is categorical where binary classification with outcomes 0 and 1 is most frequently performed. Logistic regression models the probability that the dependent variable belongs to a particular category based on one or more independent variables. A sigmoid function is thereby used to transform linear combination of the predictor variables into a value between 0 and 1.

Additionally, performing feature selection may be very useful for NM toxicity prediction. Feature selection has two main advantages: in can be used to avoid overfitting of the training data and it can improve the comprehensibility of the model for humans. In general, methods like LASSO[161] or ElasticNets[162] are most frequently used to perform feature selection for ML models. However, simpler methods like RFE[163] may also be used. In that case the least important parameters are removed from the model in a recursive way until the model performance keeps constant or drop again.

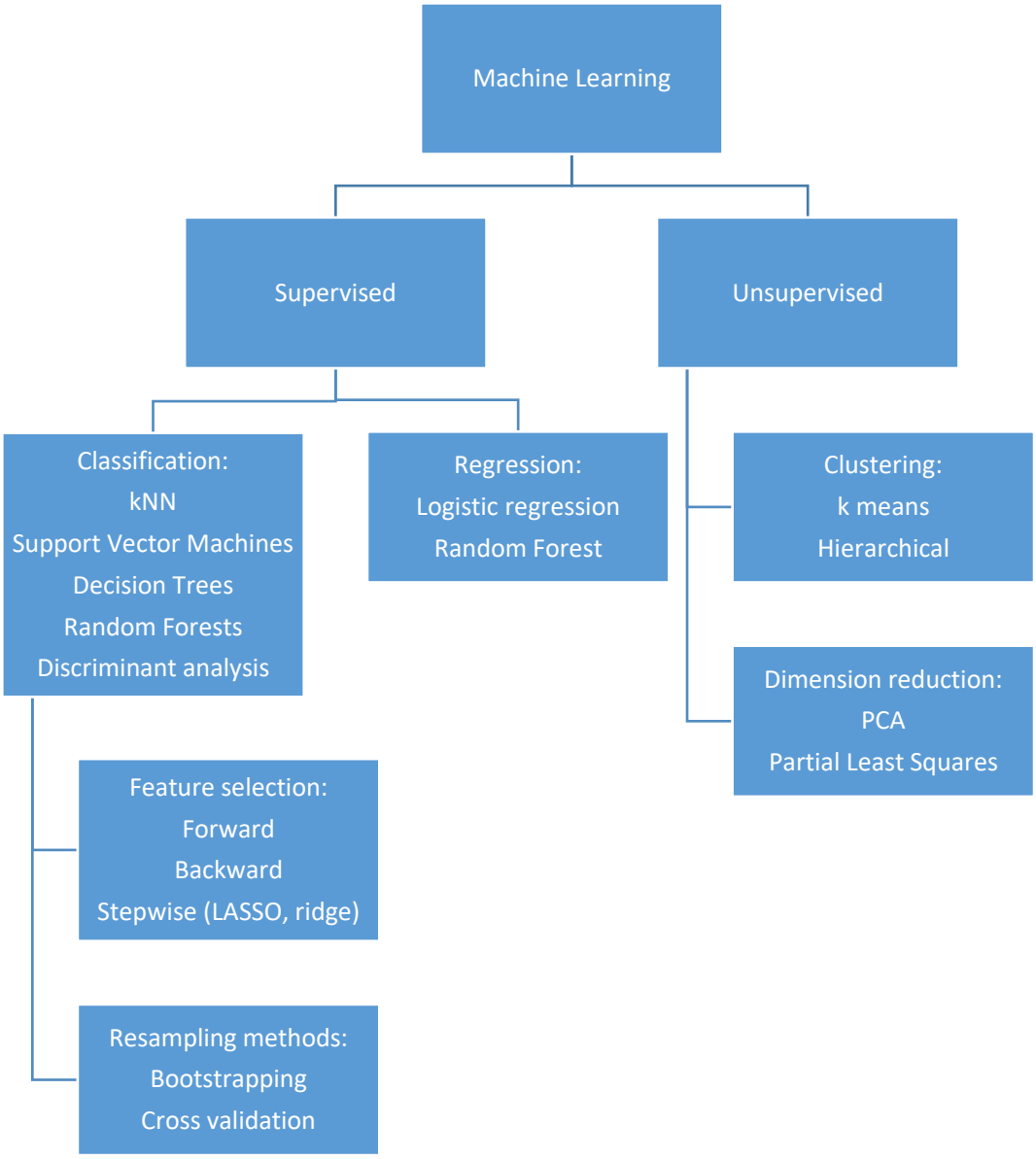


Figure 3: Overview on ML algorithms.

Once ML models supporting NM grouping and toxicity prediction have been developed, it needs to be assessed how well they perform. Several measures and techniques are suitable for this task. For regression models, performance is usually measured in terms of the Goodness-of-fit. Therefore, the squared correlation coefficient or the standard error of estimation may be computed. The performance is measured simply within the original training data. Instead, for supervised models the model performance is assessed by applying the model to separate test sets. After model training the model is used to predict the labels of the data in the test set. Predicted labels are then compared to the true class. From this comparison, counts of true- and false-positives and -negatives can be obtained. These counts can be used to calculate the sensitivity and specificity of the model, meaning its ability to predict correct labels for instances from both classes. The balanced accuracy can be

calculated by taking the mean of sensitivity and specificity. This value is usually used as the final measure of model performance. In case the dataset is small, like it is usually the case for NMs, instead of splitting the data into training and test set, one may use leave-one-out cross-validation (LOOCV)[152]. In that case, one holdout sample is defined, the model is trained on the remaining samples and then the class is predicted for the holdout sample. This is performed in an iterative manner such that in the end the class of each sample is predicted based on all other ones. The performance can then be calculated in the same fashion as before. In a similar fashion, also k-fold cross-validation is possible where a certain number of subsets are used for validation.

Various tools and ML models have been developed to support NM grouping. A comprehensive overview is given in the fourth publication included in this thesis.

1.6.4 Omics techniques and relevant tools

In the first step, omics data is usually pre-processed to allow detection of altered transcripts, proteins, metabolites and so on. This comprises data cleaning, transformation, normalization or imputation. Subsequently, the significance of the differential expression or abundance of each of the molecules can determined using statistical testing or linear modeling. As many molecules are tested at once, multiple testing correction needs to be performed. In addition, fold changes are frequently computed to get insights into the quantitative changes. As omics techniques are prone to false-positive detection due to their high-dimensionality, conclusions are usually not drawn at the level of single molecules. Instead, pathway or gene set enrichment analysis (GSEA)[164] may be performed. Thereby molecules are mapped to predefined gene sets or pathways like KEGG pathways[165, 166], HALLMARK pathways[167], Reactome pathways[168, 169] or GO terms[170]. Only if a sufficient number of proteins are changed, the pathway or gene set will be identified as significantly perturbed which increases the reliability of the finding. Results of omics analysis are frequently visualized in volcano plots which highlight significantly altered molecules with sufficiently large fold changes. Heatmaps instead can show intensities across different samples along with their hierarchical clustering results. Boxplot can be used to visualize data distributions and assess for example the success of data normalization. Various tools for NM toxicity assessment based on omics data have been introduced in the past. A collection of such tools is presented in the review paper which is part of this thesis.

1.7 Available data for application of ML for NM grouping

In order to create ML models which are able to support NM grouping, different types of data are needed: 1) Hazard data from *in vitro* or *in vivo* studies can be used as the output variable of the ML models. In supervised approaches they may represent labels; 2) Physico-chemical properties are used as input variables in classical grouping approaches; and 3) Omics data can support the justification of a grouping hypothesis by informing on shared underlying MoAs or induced AOPs. An overview on available data for these data types in the field of NMs is given below.

1.7.1 Hazard data

Experimental hazard data can be derived from *in vivo* or *in vitro* approaches. *In vivo* 90-sub-chronic or 28-day sub-acute studies with repeated doses applied to rodents thereby comprises the gold standard for inhalation toxicity. As an alternative supporting the 3R principles, the OECD Working Party on Manufactured NMs proposed STIS. The corresponding protocol was developed in the projects NANOSAFE2 and nanoCare and is an adaptation of the OECD Test Guideline 412 for 28-day sub-acute studies[132]. In STIS, rats are exposed for five days followed by a recovery period of two to 13 weeks during which animals are observed. STIS are able to provide information on early NM-induced pathogenesis as well as reversibility or persistence of effects. STIS have been performed for a variety of NMs[171-174] and data from broncho-alveolar lung fluid, lung histopathology and clinical parameters have been collected. Those studies have shown that STIS provide reliable results with respect to toxicity induction in the respiratory tract. From these studies NOAECs can be derived. Most NMs like barium sulfate, many silicon dioxide NMs, graphite nanoplatelets or carbon black show no or only moderate effects in case of unmodified amorphous silicon dioxide in STIS. Only few materials show strong effects in STIS, for instance rigid MWCNTs[175]. While STIS greatly improves testing requirements with respect to time and cost, it still remains an *in vivo* approach.

As an attempt to circumvent the need for *in vivo* studies, various *in vitro* alternatives have been developed. However, most of them correlate poorly with *in vivo* results. The most comprehensive testing strategy so far is the so-called macrophage assay[176] which reflects STIS results well across a large number of tested NMs. The macrophage assay studies the biological responses of alveolar macrophages which are the most important cells in the immune defense against exogenous substances or organisms taken up *via* the inhalative route. Four assays performed with the rat

alveolar macrophage cell line NR8383 are combined in this approach: 1) LDH release informing on particle-induced cytotoxicity, 2) glucuronidase release indicating activation of the immune defense, 3) H_2O_2 release as an indicator for cell-mediated oxidative properties of NMs and 4) TNF-α levels as a surrogate for pro-inflammatory reactions.

Within the original study, the macrophage assay was tested with respect to its applicability for use in a tiered approach for regulatory hazard assessment of NMs. As an example, it might be used in the context of the DF4nanoGrouping framework to distinguish 'active' and 'passive' NMs which form two of the four groups in this framework thereby also prioritizing NMs for further higher tier testing. This requires that the macrophage assay provides results which are in line with the corresponding *in vivo* findings. In order to prove predictivity of the macrophage assay for the *in vivo* outcome, a prediction model was established. Therefore, an initial set of 20 STIS results were categorized as 'active' or 'passive' based on their NOAEC. While STIS and macrophage assays were performed with mass-based doses, surface area seems to be the more relevant metric with respect to NM effects on macrophages[177, 178]. Therefore, mass-based concentrations were converted into surface-based ones and surface-based thresholds for distinguishing 'active' and 'passive' NMs were introduced. For STIS, a threshold of 10 mg/m^3 was set to distinguish NMs 'active' NMs with NOAECs smaller than this threshold from 'passive' NMs with larger NOAECs. This categorization was used as a reference and compared to findings of the macrophage assay. For the macrophage assay, results from all four assays were evaluated separately in terms of statistical significance. Doses selected for testing in these assays are chosen such that they can reflect macrophage loading in STIS in order to facilitate interpretation of the results. For each assay, the LOAEC was determined. A threshold of 6000 mm^2/ml was set to separate 'active' and 'passive' NMs in the macrophage assay. This threshold has been derived from *in vivo* findings on lung overload conditions and reflects the highest *in vitro* non-overload dose of 4000 µm^2/NR8383 cell[176]. In order to exclude borderline reactions, only NMs for which at least two out of four assays showed significant LOAECs below the threshold, were assigned an 'active' label. Testing the predictivity of the macrophage assay for STIS results in the first case study led to an accuracy of 0.95. Meanwhile, case studies for testing the predictivity of the macrophage assay with respect to STIS results have been extended to various additional materials[179] and thereby the usefulness of this assay has been further substantiated. The macrophage assay thus seems to be a good *in vitro* alternative informing on inhalation toxicity.

1.7.2 Physico-chemical properties

Various physico-chemical properties have been shown to influence NM toxicity and thus might be important in the context of NM grouping. Those properties may either be intrinsic and thus only depend on the nature of the NM itself or extrinsic meaning that they can change along with the surrounding conditions of the NM like the medium they are contained in. The most important properties along with the way they might influence toxicity are briefly described here.

The first important property is the size of NMs. The primary particle size is the one typically provided by the manufacturer and can be obtained from transmission electron microscopy or scanning electron microscopy in powder state after synthesis. Once suspended, the hydrodynamic diameter of the NM can be determined. As NMs in suspension are covered by a hydrate shell and ions, this value is typically higher that the PPS. The hydrodynamic diameter may be measured using different technique. Among them, the most common one is dynamic light scattering. In relevant medium, the hydrodynamic diameter is expected to increase further due to the binding of proteins onto the surface of the NMs called the protein corona. Due to their small size, NMs may be taken up and distributed throughout the body thereby potentially also passing biological barriers more easily[31]. However, size is not a single parameter but actually the size distribution is also of interest. Differences in size result from the fact that particles agglomerate or bind other molecules in different amounts. This information may be important as agglomerates may behave substantially different also with regards to induction of toxic effects compared to unbound particles.

One property directly related to size of NMs is the surface area. The smaller a NM is, the larger becomes the ratio between surface area and mass. As many reactions of NMs with their environments take place at the surface of the NMs, the relative surface area is expected to be an important factor with respect to toxicity. The surface area is frequently measured using Brunauer-Emmett-Teller analysis. Another property to be assessed is surface functionalization or treatment. This may lead to electrostatic stabilization by charge or steric stabilization by polymers. Thus, functionalization is expected to decrease toxicity. In addition, zeta potential describes the potential of the interface between layers and is indicative for NM suspension stability. Zeta potential is dependent on the pH, the ion strength and concentration as well as the presence of biomolecules. Thus, the zeta potential is an extrinsic property of NMs and may vary depending on the surrounding medium.

Surface reactivity or the OP is considered an important surrogate variable strongly related to NM toxicity. It describes the ability of NMs to induce oxidative stress by generation of ROS. Oxidative

stress is a commonly occurring toxicity mechanism for NMs[180]. It may cause inflammation or on the long-term AOs like fibrosis or genotoxicity. Various acellular and cellular assays exist for measuring the OP. Among the acellular assays, ESR is frequently used. However, other acellular assays like the FRAS assay exist and have the advantage that they do not rely on expensive equipment. Cellular OP assays are more complicated to standardized and thus less frequently used but have the advantage that they assess the OP in a biological context which might be closer to real exposure situations.

1.7.3 Omics data

Most efforts in the field of omics analyses for NMs have been performed on the level of transcriptomics. A large number of transcriptomics datasets including those on NMs are stores in databases like NCBI GEO[181]. However, other omics layers like proteomics or metabolomics may yield information closer to the phenotype and are therefore of interest for risk assessment. Unfortunately, proteomics or metabolomics datasets for NMs in public databases like PRIDE are relatively scarce. Comprehensive datasets which analyzed omics data across multiple layers are even less explored and only few studies are available in literature and public databases[182-188]. The advantage of omics data, however, is that NM-specific datasets may potentially be integrated with those on other traits. Here, huge collections of datasets are available and can be integrated given that proper harmonization could be achieved.

Chapter 2: Aim of the work

Due to fine-tuning of NMs for industrial purposes, plenty of NM variants are available on the market. As risk assessment for all variants with respect to all toxicological endpoint is not feasible, NAM frameworks like grouping and read-across are urgently needed. However, currently, reliable NM grouping is still a major challenge due to the missing knowledge with respect to the link between physico-chemical properties and toxicological outcomes. The overarching aim of this thesis was to explore the various possibilities by which ML models and bioinformatics approaches can support NM grouping. This was addressed in three different studies. The most important open question for establishing reliable NM grouping approaches is which properties are best suited to define similarity with respect to a specific toxicity endpoint. Therefore, different properties as well as approaches were compared across the publications contained in this work.

First, a variety of physico-chemical properties was assessed with respect to their predictivity for inhalation toxicity in order to find out which of those properties might be most relevant in the context of NM grouping. For this purpose, unsupervised and supervised approaches were compared for a set of eleven NMs with comprehensive description of their physico-chemical properties with respect to their suitability to reduce the set of physico-chemical properties to only the most relevant ones and feature selection was incorporated to test its suitability for model improvement.

In the next step, functional assays should be considered in addition to physico-chemical parameters. For this purpose, the OP of NMs was considered in more detail as: 1) It has been shown to be highly relevant in the context of NM toxicity; and 2) It can serve as a surrogate for multiple physico-chemical properties, thus reducing the characterization efforts. As various assays for measuring the OP exist and are frequently used, the main aim was to compare the outcomes of different OP assays. This has multiple implications for how datasets from different studies holding information on the OP could potentially be combined. In case of high correlations among assays, datasets may directly be integrated in order to increase the number of NMs considered during modeling. Instead, in case of low to medium correlations, combining various OP assays may yield higher predictivity for the toxicity outcome as different aspects of reactivity may be considered. Also, the influence of mass- *vs.* surface-based dose metrics was explored with respect to the predictivity for toxicity.

As NM grouping may also greatly benefit from knowledge on the underlying MoAs, studying the related proteomics signatures to infer correlations between NMs was another key aim of this work. Due to the sparsity of NM-related proteomics datasets, the idea was to integrate NM-specific

proteomics data with those related to other traits like chemicals, drugs or diseases, thereby benefiting from the more profound knowledge on them. However, such meta-analyses are only possible if the applied analysis workflows are harmonized. Thus, the aim was to provide such a workflow which allows for automated evaluation of proteomics datasets in a harmonized manner finally enabling the direct comparison of study results across various publicly available proteomics datasets. As such meta-analyses are mainly hampered by the lack of standardization for proteomics data, a workflow for harmonized evaluation of public proteomics data and their integration in a meta-analysis setting needs to be developed. The workflow PROTEOMAS aims to achieve FAIRification of proteomics data.

Finally, existing ML and omics approaches from the literature were also summarized in order to highlight their great potential for supporting NM grouping.

Chapter 3: Results

3.1 Recursive feature elimination in random forest classification supports nanomaterial grouping

Aileen Bahl[a,b,*], Bryan Hellack[c], Mihaela Balas[d], Anca Dinischiotu[d], Martin Wiemann[e], Joep Brinkmann[f], Andreas Luch[a], Bernhard Y. Renard[b], Andrea Haase[a]

a *German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany*

b *Robert Koch Institute (RKI), Bioinformatics Unit (MF 1), Berlin, Germany*

c *Institute for Energy and Environmental Technology e.V. (IUTA), Duisburg, Germany*

d *University of Bucharest, Bucharest, Romania*

e *IBE R&D Institute for Lung Health gGmbH, Muenster, Germany*

f *Evonik Resource Efficiency GmbH, Hanau, Germany*

This chapter was published as a research article in *NanoImpact* in 2019 (https://doi.org/10.1016/j.impact.2019.100179).

The supplementary information for this publication can be found in Annex I.

NanoImpact 15 (2019) 100179

Contents lists available at ScienceDirect

NanoImpact

journal homepage: www.elsevier.com/locate/nanoimpact

Research paper

Recursive feature elimination in random forest classification supports nanomaterial grouping

Aileen Bahl[a,b,*], Bryan Hellack[c], Mihaela Balas[d], Anca Dinischiotu[d], Martin Wiemann[e], Joep Brinkmann[f], Andreas Luch[a], Bernhard Y. Renard[b], Andrea Haase[a]

[a] German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany
[b] Robert Koch Institute (RKI), Bioinformatics Unit (MF 1), Berlin, Germany
[c] Institute for Energy and Environmental Technology e.V. (IUTA), Duisburg, Germany
[d] University of Bucharest, Bucharest, Romania
[e] IBE R&D Institute for Lung Health gGmbH, Muenster, Germany
[f] Evonik Resource Efficiency GmbH, Hanau, Germany

ARTICLE INFO

Editor: Bernd Nowack

Keywords:
Random forest
Recursive feature elimination
Feature selection
Principal component analysis
Machine learning
Nanomaterial grouping
Toxicity prediction
Physico-chemical properties

ABSTRACT

Nanomaterials (NMs) can be produced in numerous different variants of the same chemical substance. An in-depth safety assessment for each variant by generating test data will simply not be feasible. Thus, NM grouping approaches that would significantly reduce the time and amount of testing for novel NMs are urgently needed. However, identifying structurally similar NM variants remains challenging as many physico-chemical properties could be relevant.

Here, we aimed at emphasizing on the value of machine learning models in the process of NM grouping by considering a case study on eleven selected, well-characterized NMs. To that end, we linked physico-chemical properties of these NMs to characterized hallmarks for inhalation toxicity. We applied unsupervised and supervised machine learning techniques to determine which combination of properties is most predictive. First, we assessed NM similarity in an unsupervised manner using principal component analysis (PCA) followed by subsequent superposition of activity labels combined with a k-nearest neighbors approach. Then, we used random forests (RFs) as a supervised machine learning technique which directly uses the knowledge on the activity class in the process of defining NM similarity. Thus, similarity was defined only on those properties showing the highest correlation with the activity and therefore had the highest discriminative power. In order to improve the performance, we then used recursive feature elimination (RFE) to delete uninformative features biasing the results. The best performance was achieved by the reduced RF model based on RFE where a balanced accuracy of 0.82 was obtained. Out of eleven different properties we determined zeta potential, redox potential and dissolution rate to have the strongest predicting impact on biological NM activity in the present dataset. Though the dataset is too small with respect to the number of NMs studied and the applicability domain is expected to be very limited due to the fact that only few material classes were covered, our study demonstrates how machine learning and feature selection methods can be implemented for identifying the most relevant physico-chemical NM properties with respect to toxicity. We suggest that once the most relevant properties have been detected in a model built on a sufficient number of different NMs and across multiple NM classes, they should obtain special emphasis in future grouping approaches.

1. Introduction

Nanomaterials (NMs) can be manufactured with various functionalities serving different industrial purposes (Forster et al., 2011). In theory, an infinite number of different variants can be obtained for each material type by altering physico-chemical properties such as size, shape or by applying chemical surface coatings. However, altering physico-chemical properties does not only influence the functionality of

Abbreviations: NM, Nanomaterial; PCA, Principle component analysis; PC, Principle component; RF, Random forest; RFE, Reverse feature elimination; kNN, k-nearest neighbors; STIS, Short-term inhalation study; LOAEC, Lowest observable adverse effect concentration; MDA, Mean decrease in accuracy

* Corresponding author at: German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany.
E-mail address: aileen.bahl@bfr.bund.de (A. Bahl).

https://doi.org/10.1016/j.impact.2019.100179
Received 1 February 2019; Received in revised form 25 May 2019; Accepted 28 June 2019
Available online 09 July 2019

the NM but at the same time may also have an impact on its biological interactions by affecting for example cellular uptake, toxikokinetics or (eco-)toxicity (Marzaioli et al., 2014; Froehlich, 2012; Braakhuis et al., 2014). Even slight changes in some properties may drastically alter a NM's toxicological profile while other properties may have a lower impact on the toxicity. Unfortunately, a proper understanding of how changes in certain physico-chemical properties are associated with changes in toxicity, toxikokinetics or uptake is only beginning to emerge. Thus, currently each NM variant requires a detailed case-by-case evaluation that includes a thorough characterization of the physico-chemical properties as well as an in-depth assessment of the toxicological profile. Given the huge number of variants and the high demands with respect to time, laboratory animals and cost needed for these analyses such an approach is not feasible to be followed for all variants (ECHA, 2016). Instead, alternative methods aiming at reducing the amount of testing needed to address the question of potential hazards of NMs, such as grouping and read-across are urgently needed (OECD, 2016;).

For chemicals, grouping concepts have already been well established (OECD, 2016; ECHA, 2008). Two strategies are proposed and the decision which one to use mainly depends on the number of available similar source chemicals (OECD, 2016; ECHA, 2017). If a sufficient number of similar chemicals is available the category approach can be used According to the guidance documents on grouping released by OECD (2016) and ECHA (2008, 2017), a chemical category is a group of chemicals whose physicochemical and (eco-)toxicological properties and/or environmental fate are likely to be similar or follow a regular pattern, usually as a result of structural similarity. For new chemicals to be added to such an established group, the toxicity can then be predicted using tools, such as read-across, trend analysis, quantitative structure activity relationships (QSARs) (EU US Roadmap Nanoinformatics 2030, 2018). If only a smaller number of source chemicals is available, the analogue approach may become appropriate. In that case, trends may not become apparent, such that this approach is more dependent on expert judgement. In any case, key features for assuming similarity of chemicals are *e.g.* common functional groups, common breakdown products or a trend between potency and properties of interest across the group (OECD, 2016).

For NMs, several grouping approaches have been published already (Oomen et al., 2014; Oomen et al., 2015; Sellers et al., 2015; Arts et al., 2015; Dekkers et al., 2016). However, in the absence of case studies most of these approaches stay conceptual at this stage and grouping of NMs remains still challenging (Lamon et al., 2018). One of the main challenges is that one needs a much higher number of physico-chemical properties to describe a NM compared to a conventional chemical. NMs are characterized not only by many material-specific, so-called intrinsic properties, but also by properties that vary in dependence of the surrounding medium (extrinsic properties). All of these properties can potentially influence NM (eco-)toxicity, uptake or fate. However, the specific influence of each of these properties on the observed toxicity as well as a proper understanding on how they may be linked to each other and to the toxicity is only currently emerging. In addition, properties of NMs may also change during their lifetime, for example due to aging, agglomeration or aggregation, corona formation, or dissolution (ECHA, 2017; EU US Roadmap Nanoinformatics 2030, 2018; Oomen et al., 2014). Thus, also the toxicity profile of a NM could change over time. Another important factor is that the current uncertainties with respect to measuring physico-chemical properties and toxicity are high. Many of the test methods are still in the process of being adapted and validated for NMs (Gao and Lowry, 2018). In particular, extrinsic properties, which may change depending on the environmental conditions the NM is exposed to, are difficult to obtain because measurements have to be carried out in complex biological fluids. However, only if both, intrinsic and extrinsic properties of a NM are carefully characterized, information on the transformations of the NM under different conditions can be modeled reliably and used for outcome prediction. Thus, currently the largest bottleneck for establishing grouping approaches for NMs is the lack of systematic and reliable data sets suited for establishing solid linkages between physico-chemical properties and observed toxicity.

Several NM grouping schemes have been proposed already. The most comprehensive ones are the MARINA approach (Sellers et al., 2015), the RIVM approach (Oomen et al., 2015), the DF4nanoGrouping approach (Arts et al., 2015) and the NanoREG approach (Dekkers et al., 2016). However, only one of them, the DF4nanoGrouping framework, has been verified in a number of case studies (Arts et al., 2016). The DF4nanoGrouping approach covers intrinsic and extrinsic properties of the NMs as well as biopersistence, uptake, biodistribution, cellular and apical toxicity. This framework uses a tiered approach to distinguish four different groups of NMs. The first group comprises water-soluble NMs which can be assumed to be non-biopersistent. Group 2 consists of biopersistent high aspect ratio (HAR) NMs. As the DF4nanoGrouping approach focusses on inhalation toxicity, HAR NMs have to be considered separately from other NMs as they are expected to have a much higher hazard potential compared to NMs with lower aspect ratio in the lung. All other NMs are subsequently categorized as either passive or active NMs. The distinction between the groups can be based on the outcome of *in vitro* toxicity tests (alveolar macrophage assay (Wiemann et al., 2016)), as well as on surface reactivity (Ferric Reducing Ability of Serum assay (FRAS) (Gandon et al., 2017) or a cytochrome C assay (Delaval et al., 2017)). While the separation of the first two groups is made based on intrinsic and extrinsic properties of the NMs, the distinction between groups three and four are mainly based on toxicity testing data. The DF4nanoGrouping framework was used as a starting point in this work. Our aim was to identify physico-chemical properties that may guide the distinction between active and passive NMs which is one necessary step for applying the DF4nanoGrouping framework. Several challenges had thereby to be overcome.

Not all physico-chemical properties will necessarily be equally important for discriminating between active and passive materials. Moreover, the relevance of a particular property may be endpoint-specific. Thus, the main challenge is to weigh the physico-chemical properties based on their relevance for a certain toxicity endpoint and to identify combinations of the most relevant properties of a NM, which are predictive for an observed toxicological effect. It can be expected that a prediction of toxicity should be possible with a reduced set of properties (Gao and Lowry, 2018). The knowledge on which physico-chemical properties are predictive for a specific endpoint will not only facilitate grouping approaches and risk assessment for NMs but, at the same time, may also be supportive for Safe-by-Design.

Machine learning techniques are generally well-suited for solving the tasks of parameter selection and parameter ranking in a data-driven, exploratory way. Often, unsupervised approaches, such as principle component analysis (PCA) are suggested to be suited for NM grouping (Lynch et al., 2014; Sayes et al., 2013; Aschberger et al., 2019). PCA reduces the dimensionality of the input feature space to only a few linear combinations of the original input variables that show highest variability across the dataset, the so-called principle components (PCs). However, PCA has some drawbacks in the context of NM grouping and prioritization of certain physico-chemical properties of the NMs with higher importance for the toxicity outcome. As PCA is an unsupervised method, the PCs reflecting the directions of highest variation are not necessarily related to changes in the outcome variable. Some physico-chemical properties may highly vary between a set of NMs without having large influence on their toxicity outcome. In addition, the reduction of the representational space using linear combinations of the input properties makes the interpretation of the resulting PCs difficult. Another limitation of PCA is the assumption of a linear relationship between the PCs and the input space, as well as assuming statistically normal distributed variables, which might not necessarily be true for NM properties.

In order to overcome these drawbacks, non-parametric supervised

Table 1
Physico-chemical properties and measurement techniques used in this study.

Property	Measurement technique
Relative density or specific density of the material (mass per volume)	Literature-based
Primary particle size (SEM) in nm	Scanning electron microscopy
Surface area (BET) in [m^2/g]	Brunnauer-Emmet-Teller
Zeta potential at pH 7.4 in mV	Electrophoretic light scattering
Hydrodynamic diameter (z.average) in nm	Dynamic light scattering
Dissolution rate in [%w]	Solubility/chemical analysis of the supernatant by ICP-OES
Isoelectric point (pH value of no surface charge)	Electrophoretic light scattering
Band gap	Literature-based
Redox potential in mV	Pt-cathode normalized to standard hydrogen electrode
ESR CPH (mass-based)	Electron spin resonance spectroscopy using the spin probe CPH, NMs are applied at same mass concentration, sample to blank ratio
ESR CPH (surface-based)	Electron spin resonance spectroscopy using the spin probe CPH, NMs are applied at same surface area concentration, sample to blank ratio
ESR DMPO (mass-based)	Electron spin resonance spectroscopy using the spin trap DMPO, NMs are applied at same mass concentration, sample to blank ratio
ESR DMPO (surface-based)	Electron spin resonance spectroscopy using the spin trap DMPO, NMs are applied at same surface area concentration, sample to blank ratio

machine learning techniques which do not make strict assumptions on the properties of the input data and at the same time do use labeled data for training can be used instead. One such method is random forest (RF) classification (Breiman, 2001). RFs are a collection of binary decision trees, which are built on bootstrap samples of the original sample space. Every decision tree combines the explanatory variables in such a way that they are best linked to an outcome variable.

Though other supervised methods are available as well, RFs show several advantages for NM toxicity prediction: As the trees built during the RF approach show a rather low correlation among each other due to random choices of sample and variable sets, predictions are rather robust even for relatively small sample sizes and overfitting does not occur as frequently as it does with other methods like single decision trees (Amaratunga et al., 2008). In addition, RFs are non-parametric and thus well suited for different kinds of data properties and relationships between input variables and the outcome. Another advantage of RFs is that they use internal variable importance measures. The variable importance may be directly used to select the subset of NM properties that is most predictive for the toxicity outcome. We assume that establishing NM grouping concepts on only those most predictive NM properties may be more robust then including all NM properties which partly may be unrelated to toxicity.

A few studies have already applied RFs in the context of NM toxicity prediction (Lamon et al., 2018; Sizochenko et al., 2014; Cassano et al., 2016; Ha et al., 2018). However, these studies include all features in the model building step or perform feature selection only based on correlations between the input properties (Lamon et al., 2018). Due to random choices of subsets of variables being made at each split, a large number of noise variables which are unrelated to the outcome variable may have an impact on the performance on RFs. Therefore, feature selection based on feature importance prior to building the final RF model can be highly useful to improve the prediction accuracy (Genuer et al., 2010) and should be assessed. In the current study we use an approach based on recursive feature elimination (RFE) to remove unimportant features in a stepwise manner. Goldberg et al. (2015) already showed the advantages of such an approach for the prediction of the NM transport behavior. In a similar fashion, Findlay et al. (2018) used RFE to improve models predicting protein corona formation on silver NMs based on their physico-chemical properties. Other studies (Darst et al., 2018; Gregorutti et al., 2017) have shown that RFE in general is useful in case of correlated predictors. For NMs, many of the physico-chemical properties are not independent of each other and thus RFE is assumed to be useful to improve RF models for NM toxicity prediction.

There are two main goals to be achieved during feature selection: 1) One may want to determine all important variables related to the outcome variable or 2) one may want to obtain a minimal set of variables that gives a good predictive model, which is not overfitted and able to generalize to new datasets. In the case of NM toxicity prediction, the second goal will be most important.

In RFs, feature selection can be performed in a very straightforward way by the stepwise removal of features with the smallest variable importance. This variable importance can, for example, be assessed by the mean decrease of Gini impurity or the mean decrease of accuracy. The Gini impurity measures how often a randomly chosen sample would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. The mean decrease in accuracy is obtained by permuting the values of the feature under consideration and measuring the error increase due to this randomization. In contrast to PCA, the dimension reduction in this approach is achieved by removing complete features instead of combining them to new linear combinations of the original features. Thus, the interpretability of the results is more straightforward.

In the present study, we compared the performance of unsupervised PCA in combination with a k-nearest neighbor (kNN) approach with that of a RF approach for linking physico-chemical properties to toxicity data and to build a predictive model for NM toxicity. Here, PCA was added for comparison reasons only as it is a commonly used method but not all assumptions are necessarily fulfilled in this study. We also compared the performance of full and reduced RF models. Reduction of the number of input variables is assumed to be useful for improving the prediction accuracy of the model as datasets containing only a small number of input variables are prone to overfitting if too many input variables are included (Breiman, 2001). We tested the performance of the aforementioned methods on a dataset of eleven NMs mainly consisting of different silica particles that are systematically varied in size and structure, surface charge and surface hydrophobicity.

2. Materials and methods

2.1. NMs

In the present study, we analyzed a set of eleven different NMs (Table 1). The main case study consists of seven amorphous silica particles altered in a systematic way by changing their surface charge (SiO_2_15_unmod, SiO_2_15_Amino and SiO_2_15_Phospho), size and structure (SiO_2_15_unmod, SiO_2_40, SiO_2_7) as well as hydrophobicity (SiO_2_7, SiO_2_7_TMS2, SiO_2_7_TMS3). The silica NMs were obtained from BASF SE (SiO_2_15_unmod, SiO_2_15_Amino and SiO_2_15_Phospho) and from Evonik Resource Efficiency GmbH (SiO_2_40, SiO_2_7, SiO_2_7_TMS2, SiO_2_7_TMS3).

In addition to the silica case study a few other NMs were included in this study. TiO2 NM-105 from the JRC repository is used as a benchmark material (Nel, 2013) in this study as it has widely been used and carefully been characterized before. Most importantly it has been chosen by the OECD's Working Party on Manufactured Nanomaterials as a benchmark for interlaboratory comparisons and verification of testing methods for NMs.

In addition, CuPhthalocyanine Blue and CuPhthalocyanine Green were added to the set of considered NMs in this study as they form another mini-case study. They are a pair of materials that only differ in one halogenation. Thus, the influence of that halogenation on the toxicity outcome can directly be studied. Both pigments were obtained in technical grade from BASF Colors and Effects.

Mn_2O_3 was bought from Skyspring Nanomaterials and was included in the dataset as well as it has shown strong effects on macrophages previously (unpublished data obtained in the project nanoGRAVUR). Thus, Mn_2O_3 may serve as a positive control in this study.

All NMs were confirmed to be endotoxin-free in a Limulus Amebocyte Lysate Endochrome (LAL) test.

2.2. NM dispersion and characterization of physico-chemical properties

NMs were dispersed at a final concentration of 0.5 mg/ml using a Bandelin Cup Horn (Bandelin, Germany) following the NanoToxClass SOP (-NanoToxClass, 2017). The hydrophilic NMs were dispersed in water or cell culture medium. 10% fetal calf serum (FCS) was added to the cell culture medium after Cup Horn sonication. For the two NMs with hydrophobic surface coatings (SiO_2_7_TMS2, SiO_2_7_TMS3), 100 µg/ml of Pluronic F108 (Sigma-Aldrich, # 542342, Germany) was added before sonication. Final input power applied were 6 W.

All NMs were characterized with respect to their physico-chemical properties using well standardized state of the art approaches (Izak-Nau and Voetz, 2014) that have already been applied and tested in former German and EU projects like nanoGEM, MARINA or nanOximet. The standardized methods and operation procedures of these projects were used for NM characterization (NanOxiMed, 2014 - 2016). An overview of the measured properties along with their measurement techniques is given in Table 1.

Within this study, physico-chemical properties measured in deionized water (dH_2O) were used. However, similar measurements have been performed in two different cell culture media (F-12K and DMEM) and may be explored for their potential to refine the approach. Only those physico-chemical properties not containing any missing values were included in the analyses.

The mean values of the physico-chemical properties that were used in the classification approach are summarized in Table 2.

2.3. NM toxicity testing

Categorization of NMs into active and passive materials was mainly based on literature data. *In vivo* inhalation toxicity was considered most relevant (Christensen et al., 2010). Information on *in vivo* toxicity was obtained from short-term inhalation studies (STIS) in rats performed by Landsiedel et al. (2014). NMs were considered as active if the NOAEC was below 10 mg/m^3 and otherwise classified as passive as explained in Wiemann et al. (2016).

For NMs in the dataset for which no published *in vivo* data was available at the time of the study, we assigned the activity label based on the macrophage assay as suggested in Wiemann et al. This macrophage assay is performed with the rat alveolar macrophage cell line NR8383 and combines four assay measurements, namely LDH, ROS, TNF-α and glucuronidase. High correlations between the outcomes of the *in vitro* macrophage assay and the *in vivo* STIS have been shown already in Wiemann et al. who directly compared the outcomes of the studies for a comprehensive set of NMs. NMs are considered as active if at least two of the assays (*i.e.* LDH, ROS, TNF-α or glucuronidase) show

Table 2
Physico-chemical properties across studied NMs.

NM Property	SiO_2_15_unmod	SiO_2_15_Amino	SiO_2_15_Phos-pho	SiO_2_40	SiO_2_7	SiO_2_7_TMS2	SiO_2_7_TMS3	Cu-Phthalo-cyanine non-haloge-nated	Cu-Phthalo-cyanine halo-genated	TiO_2 NM-105	Mn_2O_3
Relative density	2.65	2.65	2.65	2.65	2.65	2.65	2.65	1.62	2.14	3.89	4.50
Primary particle size [nm]	15.9	16.0	18.2	71.3	17.5	16.4	14.4	26.2	47.2	18.3	50.0
Surface area [m^2/g]	200	200	200	34	249	213	198	49	61	57	58
Zeta potential pH 7.4 [mV]	− 36.7	− 36.4	− 39.5	− 40.9	− 39.8	− 20.7	−7.1	− 15.8	− 20.1	− 25.1	− 40.7
Hydro-dynamic diameter [nm]	48	47	49	373	243	175	468	649	472	394	175
Dissolution rate [%w]	0.5	3.7	0.91	0.11	0.5	2.49	0.76	0.01	0.01	0.01	0.01
Isoelectric point	1.51	4.36	1.92	2.07	3.32	3.56	4.30	3.01	3.65	4.91	2.46
Band gap	8.9	8.9	8.9	2.15	2.15	2.15	2.15	4.14	4.43	3.1	1.66
Redox potential [mV]	254	216	219	260	258	283	290	215	291	352	536
ESR CPH (mass-based)	0.82	0.92	1.21	0.68	0.93	1.64	1.52	2.08	0.72	0.69	16.9
ESR CPH (surface- based)	0.004	0.005	0.006	0.02	0.004	0.008	0.008	0.042	0.012	0.012	0.291
ESR DMPO (mass-based)	0.57	0.97	0.84	0.98	0.85	0.99	1.53	1.31	0.88	1.01	9.12
ESR DMPO (surface-based)	0.003	0.005	0.004	0.029	0.003	0.005	0.008	0.027	0.014	0.018	0.157

Table 3
In vivo and *in vitro* categorization of the NMs. Activity categories were assigned based on previous finding from STIS (Landsiedel et al., 2014). NMs that were not tested in this study were categorized based on the results of the macrophage assay (Wiemann et al., 2016).

NM	*In vivo* categorization (STIS)	*In vitro* categorization (macrophage assay)
SiO_2_15_unmod	Active	Active
SiO_2_15_Amino	Passive	Passive
SiO_2_15_Phospho	Passive	Passive
SiO_2_40	/	Active[a]
SiO_2_7	/	Active[a]
SiO_2_7_TMS2	/	Passive[a]
SiO_2_7_TMS3	/	Passive[a]
CuPhthalocyanine Blue	Passive	Active
CuPhthalocyanine Green	/	Active[a]
TiO_2 NM-105	Active	Active
Mn_2O_3	/	Active[a]

[a] Obtained within NanoToxClass.

a LOAEC (Lowest Observable Adverse Effect Concentration) below 6000 mm^2/ml and as passive otherwise in accordance with Wiemann et al.

For NMs not studied in Wiemann et al. (see Table 3), the macrophage assay was performed within the study following the method descriptions in Wiemann et al. The assays (*i.e.* LDH, ROS, TNF-α or glucuronidase) were basically performed as described in Wiemann et al. with only two exceptions: 1) The TNF-α assay was replaced by an ELISA test (BMS622, Invitrogen) and 2) the NR8383 cells were seeded at a density of 5×10^5 cells/ml in 96-well plates. The cells were then exposed to 22.5, 45, 90 and 180 µg/ml NMs concentrations in serum free Ham's F-12K medium with 1% penicilin/streptomycin for 16 h and respectively 1.5 h in case of the ROS assay. Blanks (cell free medium ± NMs) corresponding to each sample were used to eliminate any interference of NMs.

2.4. Machine learning approaches

We used an approach based on a PCA combined with a kNN classifier to address the problem of NM toxicity prediction in an unsupervised manner. PCA is commonly used to project high-dimensional data into a lower-dimensional space which still holds as much information as possible. Therefore, one has to determine the PCs of the corresponding dataset. The linear combination representing the direction of highest variability of the data is called the first PC. All remaining PCs are orthogonal vectors of highest variability in that direction. Here, the first two PCs were used to define similarity between NMs and as input for the kNN approach. The kNN reads-across the toxicity value from the k NMs that were determined to be most similar to each other. In this study, the parameter k was set to one and thus the toxicity label was obtained in a read-across manner from the NM that is the nearest neighbor of the target NM. The similarity was defined based on the first two PCs and is visualized in Fig. 2.

RF classification was used for supervised learning. RFs build up a number of decision trees based on bootstrap samples of the original data. Within each decision tree, the input variables, here the physicochemical properties, are combined in such a way that they separate both classes from each other as well as possible. In this step, another layer of randomness is added by considering only a subset of the input variables as potential split criteria for each split. Which descriptor is finally chosen to set the split criterion depends on their separation performance. Common choices to select the split criterion are the Gini impurity or the prediction accuracy (also called permutation error) (Breiman, 2003). Both criteria are described in more detail below in the paragraph on RFE.

In order to assess the generalizability of the constructed RF, the dataset should be divided into a training set, which is used to build the RF and a test set, which is used to assess how well the RF performs on a set of data that the RF has not seen before. Here, we used cross-validation in a leave-one-out manner. Thus, for each NM, the class label was predicted by the RF generated on all other NMs. The final prediction of toxicity for the test NM is based on a majority voting of all trees in the RF. As here RF classification is used, the outcome variable holds class labels for each sample.

For reduction of the number of input variables of the RF, we used backward recursive feature elimination (RFE) (Guyon et al., 2002) based on the mean decrease of accuracy (MDA) importance. The MDA is computed by randomly permuting the values of each input variable, one at a time, and assessing how much the prediction accuracy drops by doing so. Larger decreases in the prediction accuracy correspond to higher importance of the input variable under consideration. The feature with the minimum value for MDA corresponding to the least important variable was removed from the input set and a new RF was built based on this reduced set of variables. The minimal set of input variables giving an optimal balanced accuracy was determined. Equivalently, RFE was performed based on the Gini importance as the variable exclusion criterion. Gini importance measures how well the samples can be assigned to the two output classes by making a split on the variable under consideration at a specific node. The higher that value is, the better is the separation of the instances into the two classes and the higher is the importance of the inspected feature. Means and standard deviations for the MDA as well as Gini importance values for each feature were calculated in order to infer knowledge on the importance of each physico-chemical property on the toxicity outcome. Variability estimates of the importance result from the leave-one-out cross-validation in which variable importance was assessed within each RF model and then averaged across all models.

The performance of the classification models was assessed based on the numbers of correct and incorrect class predictions. A material is correctly classified if the class predicted by the model is the same as the label that the NM was originally assigned based on the results of the STIS or the macrophage assay. Sensitivity (true predictions as 'active' (true positives)/all predictions as 'active' (positives)), specificity (true predictions as 'passive' (true negatives)/all predictions as 'passive' (negatives)) and balanced accuracy (sensitivity + specificity/2) were assessed by comparing the assigned class label to the predicted one.

The implementation of RFs from the R package 'randomForest' was used with the number of trees generated being set to 5000 and the number of features assessed at each split being set to the default value. An R package implementing the methods presented here is available at https://github.com/AileenBahl/ML_Tox.

3. Results

3.1. Assignment of toxicity labels

In vivo STIS results are present for five of the NMs. In four cases they match the results from the macrophage assay. Only CuPhthalocyanine Blue is false-positive *in vitro* but passive *in vivo*. For six of the NMs, no results from the macrophage assay have been published before. Thus, we performed the macrophage assays for those NMs. All results are summarized in Table 3. For all cases in which an *in vivo* categorization was available, we assigned this as the class label. For the other cases, we used the *in vitro* categorization. The only exception is CuPhthalocyanine Green which due to its similarity to CuPhthalocyanine Blue was assumed to be passive *in vivo*. CuPhthalocyanine Green was obtained from CuPhthalocyanine Blue by halogenation. Both materials differ only by this halogenation. As we do not have information from *in vivo* studies for CuPhthalocyanine Green, the passive behavior *in vivo* is only an assumption.

3.2. Physico-chemical properties across NMs

Fig. 1 shows the distribution of values of all studied physico-chemical properties across NMs in a heatmap. The values for the physico-chemical properties of the NMs were translated into colors (ranging from dark blue for the lowest values to dark red for the highest values). The values of all properties were scaled to guarantee comparability between properties with differing ranges of values. A comparison of the left part of the heatmap consisting of the physico-chemical properties belonging to the set of NMs with class label 'active' (column names colored in black) with the right part consisting of passive NMs (column names colored in purple) shows that there is no single variable that can perfectly distinguish between the active and the passive group. However, for some of the properties tendencies are visible. One such example is the zeta potential which is higher in almost all passive NMs than compared to the active ones. In addition, the dendrogram on the left side of the figure shows how similar the different physico-chemical properties are to each other across all tested NMs. In Supp. Fig. 1, the clustering of NMs across all physico-chemical properties is shown. Active and passive NMs cannot be separated from each other and do not cluster together based on all assessed physico-chemical properties with equal weights.

3.3. Unsupervised learning approach - PCA and kNN

The kNN read-across like approach was based on the first two principal components (PCs) obtained from a PCA. These two PCs explain 69.8% of the total variance. Fig. 2 shows the contributions of each input parameter to each of the two PCs, as well as the location of each NM within the space spanned by these two PCs. The first PC is strongly related to the reactivity of the NMs (ESR and redox potential) and to a lesser extent also to the relative density. The second PC is highly influenced by the hydrodynamic diameter followed by zeta potential, surface area, band gap and dissolution rate.

Training a kNN with k = 1, so reading across the toxicity class from the NM that is most similar to the one that should be predicted with respect to the first two PCs, we obtained seven correct predictions, while four NMs were misclassified (namely SiO_2_15_unmod, SiO_2_15_Phospho, SiO_2_7, SiO_2_7_TMS2). This corresponds to a sensitivity of 0.6, a specificity of 0.67 and a balanced accuracy of 0.64.

3.4. Supervised learning approach - random forest

3.4.1. Full model

As a starting point, we created a full RF model by incorporating all assessed physico-chemical properties as input variables. This leads to a correct prediction of the toxicity of six NMs and a misclassification of five NM (see Table 4). The sensitivity of that classifier is 0.4, the specificity is 0.67 and the balanced accuracy is 0.54. The stability of the correct predictions as assessed by the ratio between the correct and the incorrect votes is roughly the same as compared to the stability of the incorrect predictions.

The importance of each input variable in that RF is assessed by the MDA or mean decrease in Gini importance, respectively (see Fig. 3 and Table 5). In both cases, zeta potential, dissolution rate, surface-based ESR DMPO, and redox potential are among the top 5 highest-ranking variables. For Gini importance the set of top 5 variables is completed by mass-based ESR CPH, for accuracy it is the relative density.

3.4.2. Reduced models

As the performance of RFs is drastically reduced if a lot of noise variables not highly related to the outcome variable are included in the prediction, we reduced the number of input variables to see whether the performance of the predictor can be improved. We assessed

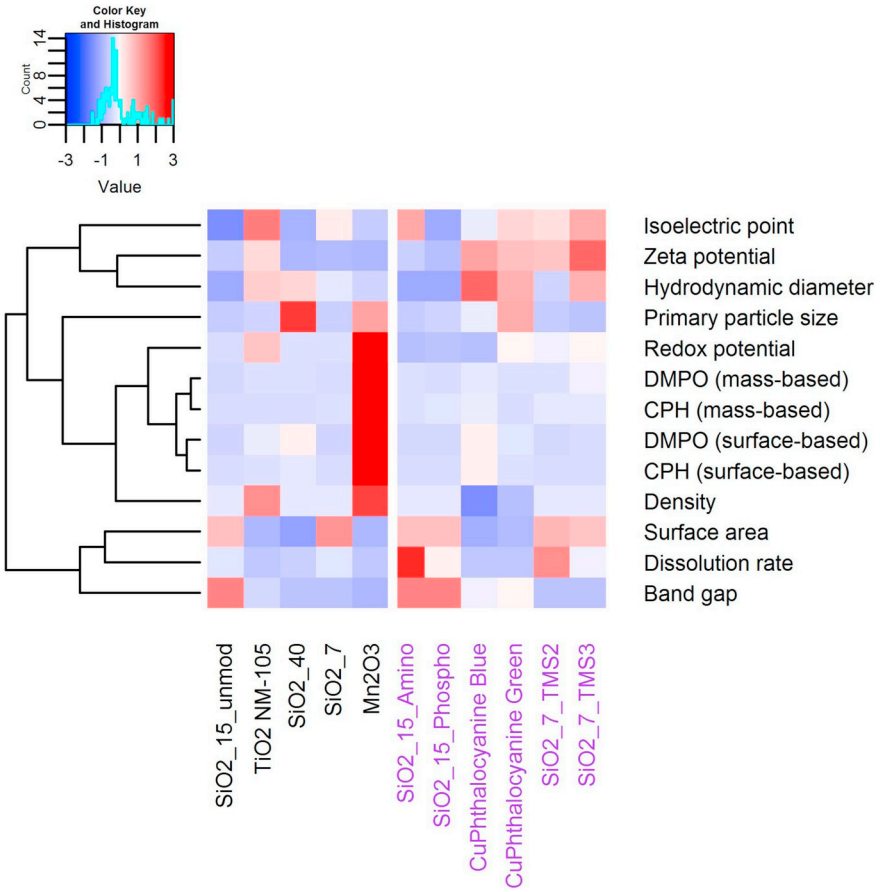


Fig. 1. Heatmap of physico-chemical properties across NMs. The table of physico-chemical properties was translated into colors ranging from dark blue for the smalles values to dark red for the highest values. All properties were scaled across NMs in order to make them comparable and to avoid overrepresentation of those properties having larger values in general in the clustering step. The black labels on the left side of the x-axis correspond to active NMs, the purple ones on the right side correspond to passive NMs. Comparing both sides shows that none of the physico-chemical properties alone is able to seperate active from passive NMs. The dendrogram shows the similarity of the physico-chemical properties across all studied NMs. The length of the branches indicates how closely correlated the properties are. Shorter branches represent higher similarity and thus higher correlation.

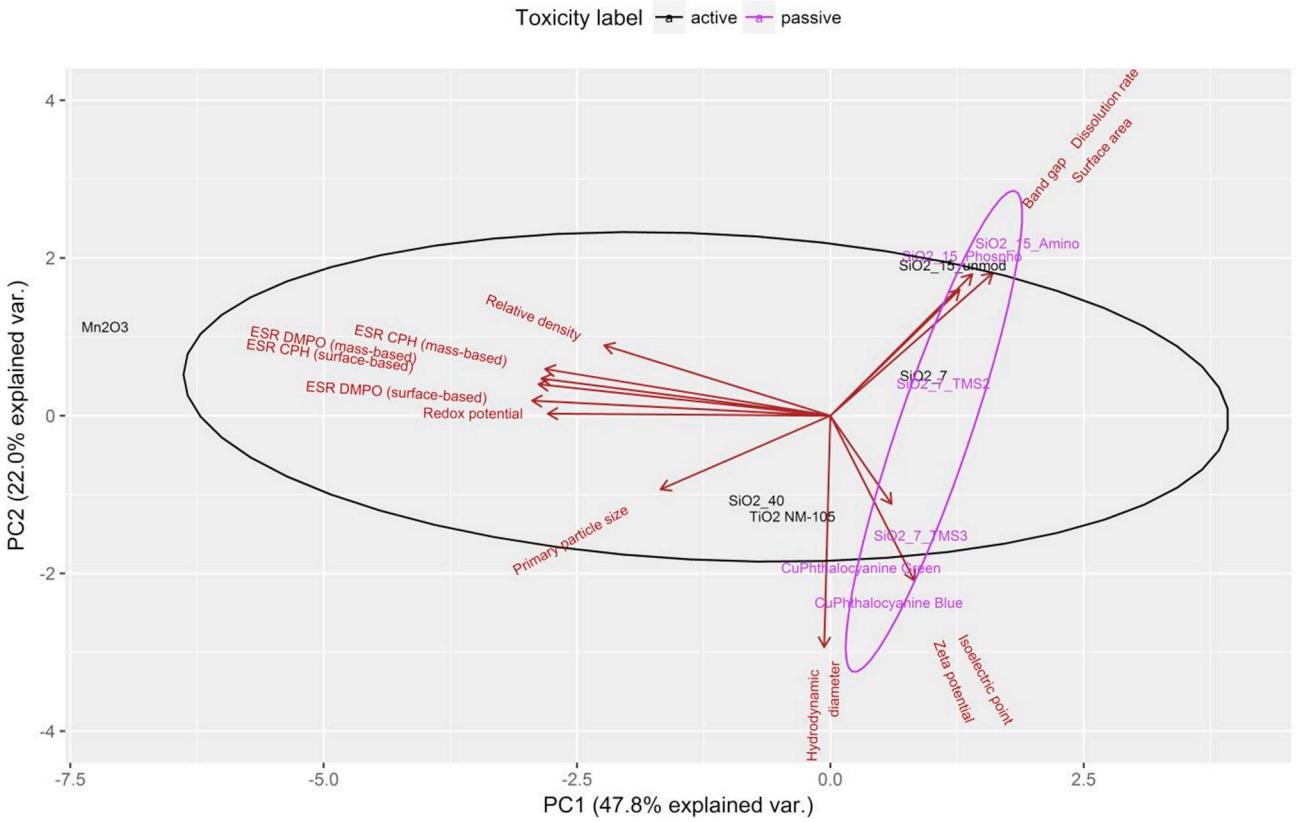


Fig. 2. PCA biplot of the first two principle components (PCs). The figure displays the variable loadings of the physico-chemical properties and PC scores of the NMs across the first two principle components. Values on the x-axis correspond to the scores of each NM as well as to the scaled variable loadings of the physico-chemical properties in PC1. The y-axis represents the same properties for PC2. The arrows represent the weights of each physico-chemical property in the linear combination of each of the two principle components. Higher absolute values of these weights indicate higher importance of the property for the associated PC. The lengths of the arrows relate to the importance of the corresponding properties within the first two PCs with longer arrows representing more important properties. The direction of the arrow indicates whether the particular property is more important in PC1 (horizontal arrows) or in PC2 (vertical arrows). The location of each of the NMs within this reduced space is indicated by black labels for active NMs and purple labels for passive NMs.

Table 4
Classification result for the full RF model based. All assessed physico-chemical properties were used as input for the generation of a RF classifier. Internal model validation was performed using leave-one-out cross-validation. Empirical frequencies are the same for the RF model based on the mean decrease in accuracy and the mean decrease in Gini importance.

NM	True class	Predicted Class	Empirical frequency of votes for label 'active'	Empirical frequency of votes for label 'passive'
SiO_2_15_unmod	Active	Passive	0.34	0.66
SiO_2_15_Amino	Passive	Passive	0.44	0.56
SiO_2_15_Phospho	Passive	Active	0.67	0.33
SiO_2_40	Active	Active	0.61	0.39
SiO_2_7	Active	Passive	0.42	0.58
SiO_2_7_TMS2	Passive	Passive	0.30	0.70
SiO_2_7_TMS3	Passive	Passive	0.25	0.75
CuPhthalocyanine Blue	Passive	Passive	0.48	0.52
CuPhthalocyanine Green	Passive	Active	0.59	0.41
TiO_2 NM-105	Active	Passive	0.26	0.74
Mn_2O_3	Active	Active	0.57	0.43

backward RFE based on MDA as well as on Gini importance prior to the actual model building step.

First, we reduced the number of input variables in the model based on the MDA with re-evaluation. In each step of the RFE, we built a RF, ranked the input variables according to their MDA, removed the variable with lowest importance and created a new RF based on this reduced set of input variables. We then determined the minimal set of input variables leading to the highest balanced accuracy of the model.

The best model was obtained using zeta potential, dissolution rate and redox potential. In that case, only two NMs were misclassified (SiO_2_15_Phospho and TiO_2 NM-105) leading to a sensitivity of 0.8, a specificity of 0.83 and a balanced accuracy of 0.82. The empirical frequencies of votes for both classes in the prediction of SiO_2_15_Phospho were almost equal (55% of all votes for 'active' and 45% of all votes for 'passive') and the difference in the number of correct and incorrect votes was much smaller than for most other NMs (exceptions: SiO_2_15_unmod and CuPhthalocyanine Green which also had almost equally many votes for either of the classes, see Table 6). However, in the case of TiO_2 NM-105, the prediction of the incorrect class label is rather stable (85% of all votes suggested 'passive' as the correct class label). The variable importance for the zeta potential is 20.50 ± 7.11, for dissolution rate it is 16.49 ± 5.59 and for redox potential it is

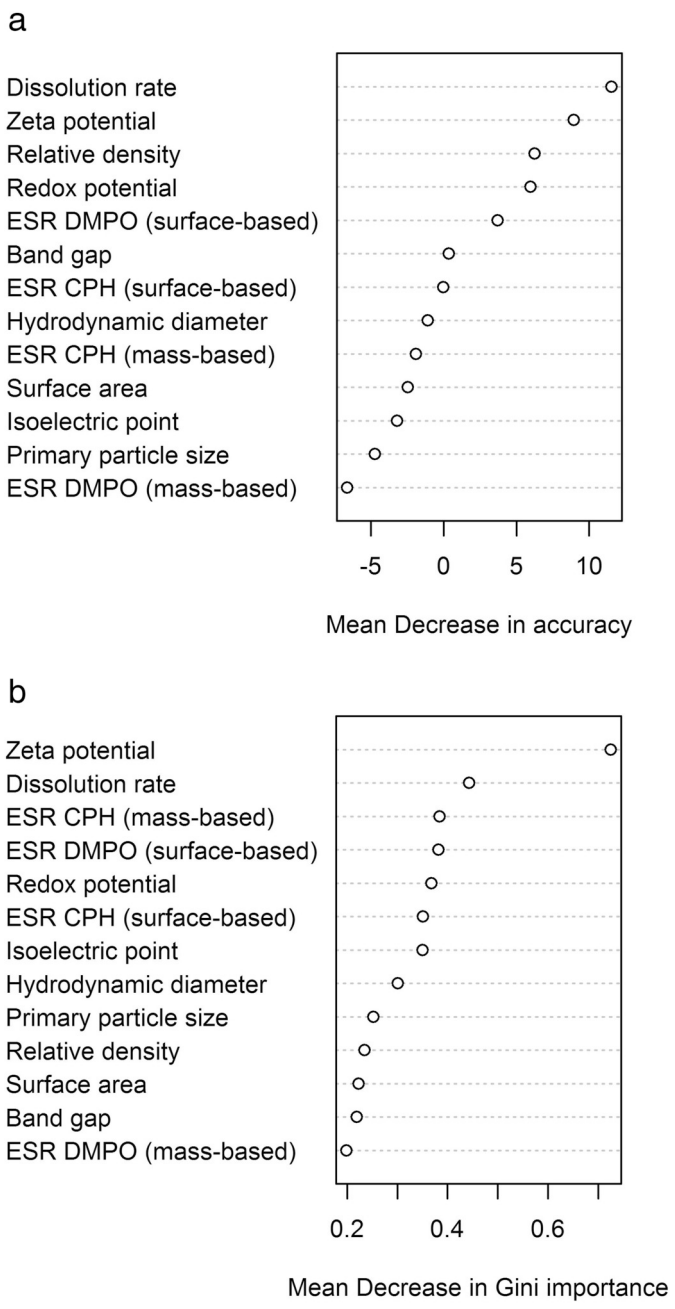


Fig. 3. Variable importance of each parameter within the full RF model. a) Mean decrease in accuracy and b) mean decrease in Gini importance are depicted for each physico-chemical property. Properties at the top of the plot are of highest importance in the particular model.

15.28 ± 3.88. The spatial distribution of NMs across the space spanned by the three variables is depicted in Fig. 4.

If we use the mean decrease in Gini importance instead of the MDA to rank the features, the balanced accuracy of the model drops to 0.73 with SiO_2_15_unmod being misclassified in addition to the previous two materials. The empirical frequencies of votes do not improve and the surface-based ESR measurement with the DMPO spin trap is needed in addition to the three parameters from the model based on the MDA. Thus, using the mean decrease of accuracy leads to better results in that case.

The spatial distribution of NMs across the three variables is depicted in Fig. 4.

Table 5
Variable importance values in the full RF model. Means and standard deviations for the mean decrease of accuracy as well as Gini importance values for each feature are given. Standard deviations result from the leave-one-out cross-validation in which variable importance was assessed within each RF model and then averaged across all models.

Physico-chemical parameter	Mean decrease in accuracy	Mean decrease in Gini importance
Dissolution rate	11.83 ± 3.18	0.45 ± 0.12
Zeta potential	9.66 ± 5.28	0.72 ± 0.12
Relative density	6.44 ± 2.25	0.23 ± 0.07
Redox potential	6.00 ± 2.96	0.38 ± 0.09
ESR DMPO (surface-based)	0.94 ± 3.41	0.39 ± 0.06
Band gap	−0.24 ± 4.85	0.23 ± 0.09
ESR CPH (surface-based)	−0.54 ± 2.89	0.28 ± 0.05
Hydrodynamic diameter	−1.51 ± 4.29	0.30 ± 0.07
ESR CPH (mass-based)	−2.19 ± 5.03	0.40 ± 0.13
Surface area	−2.41 ± 3.48	0.23 ± 0.06
Isoelectric point	−3.65 ± 4.65	0.36 ± 0.09
Primary particle size	−4.42 ± 3.32	0.26 ± 0.04
ESR DMPO (mass-based)	−6.32 ± 1.73	0.21 ± 0.03

4. Discussion

In this study, we evaluated the performance of an unsupervised machine learning approach based on a PCA in combination with a kNN classifier, as well as a supervised strategy based on RFs with and without feature selection for the prediction of the inhalation toxicity of eleven NMs. While the prediction performance of the full RF model was even lower than that of the unsupervised approach, backward RFE prior to building the final RF model strongly improved the accuracy of the model leading to improved results compared to those obtained with PCA. At the same time, our approach allowed to identify the physico-chemical properties having highest predictivity for the outcome of inhalation toxicity based on our dataset. For the most powerful approach of RF with RFE, a systematic removal of the most uninformative property in each step led to a correct prediction for nine out of eleven NMs. Zeta potential, redox potential and dissolution rate were thereby determined to be the best discriminating features.

Overall, zeta potential, redox potential as well as dissolution rate were among the most powerful predictors in all generated models using supervised as well as unsupervised approaches. These properties are also in-line with existing hypotheses.

The zeta potential is a measure for the surface charge of a NM. This can be regarded as a proxy for the stability of NM dispersions *in vitro* and predicts the likelihood of NM interactions as well as interactions of NMs with other charged molecules like proteins (Liu et al., 2015; Cho et al., 2012). Therefore, zeta potential plays an important role in NM agglomeration and the formation of a protein corona. With increasing absolute values of the zeta potential, the repulsion forces between particles increase thereby lowering the potential for aggregation leading to a more stable suspension. This may, for example, improve cellular uptake and induce stronger biological effects.

The redox potential of a NM is associated with its ability to form reactive oxygen species (ROS) (Hellack et al., 2017). ROS are known to react with DNA, proteins, lipids or other cellular compounds and to damage them or hamper their functionality by inducing conformational changes. Thus, NMs with a higher redox potential can likely be assumed to be more active.

The dissolution rate of NMs can potentially affect their toxicity in different ways. On one hand, fast dissolving NMs produce a high amount of ions. The toxicological outcome, of course, will depend on whether these ions are toxic (Cho et al., 2012; Cho et al., 2011). On the other hand, dissolution also affects the bioavailability and biopersistence of NMs and can thus influence the toxicity of particles indirectly as well (Utembe et al., 2015).

Table 6
Classification result for the reduced RF model based after backward recursive feature elimination. The input parameters, comprised of the physico-chemical properties in this case, were reduced in a stepwise manner removing the most unimportant feature in each step. RFs were sequentially built on these reduced sets of input parameters. The RF with the best balanced accuracy and the minimal number of input features was selected as the final model. Internal model validation was performed using leave-one-out cross-validation. The best model was obtained with the three physico-chemical properties zeta potential, dissolution rate and redox potential as input parameters and variable importance being addressed by the mean decrease in accuracy. Results for this model are shown in this table.

NM	True class	Predicted class	Empirical frequency of votes for label 'active'	Empirical frequency of votes for label 'passive'
SiO_2_15_unmod	Active	Active	0.56	0.44
SiO_2_15_Amino	Passive	Passive	0.24	0.76
SiO_2_15_Phospho	Passive	Active	0.55	0.45
SiO_2_40	Active	Active	0.82	0.18
SiO_2_7	Active	Active	0.74	0.26
SiO_2_7_TMS2	Passive	Passive	0.13	0.87
SiO_2_7_TMS3	Passive	Passive	0.20	0.80
CuPhthalocyanine Blue	Passive	Passive	0.31	0.69
CuPhthalocyanine Green	Passive	Passive	0.43	0.57
TiO_2 NM-105	Active	Passive	0.15	0.85
Mn_2O_3	Active	Active	0.81	0.19

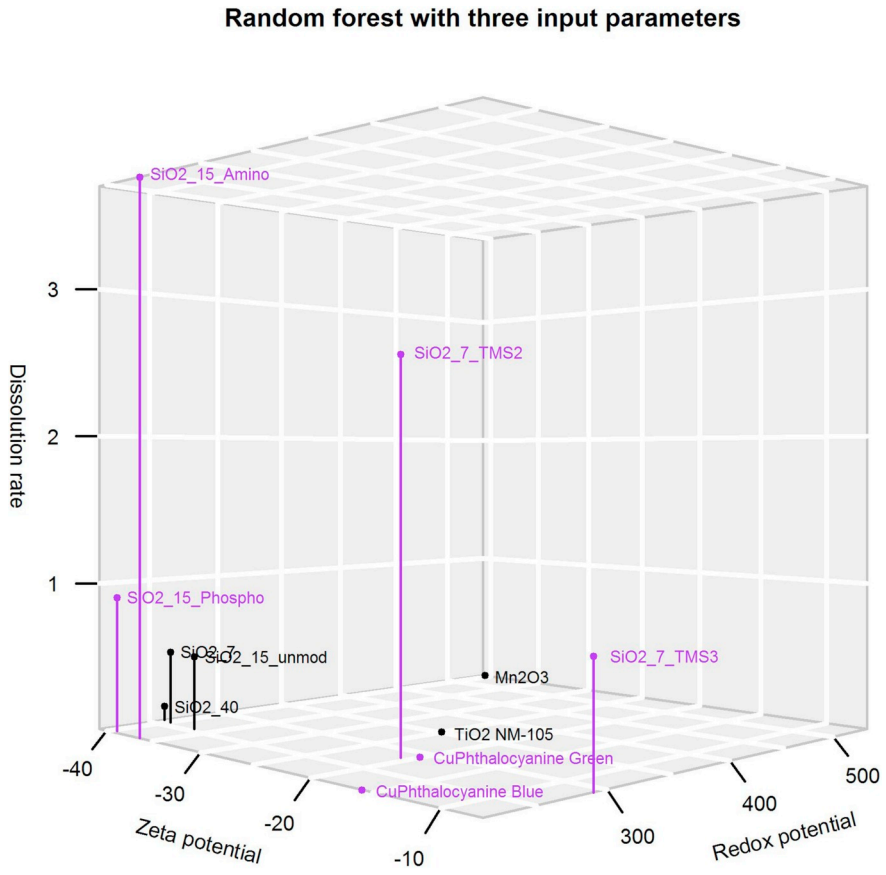


Fig. 4. Scatterplot of the input variables of the RF model reduced by recursive feature elimination (RFE). The best RF model after RFE contains only three of the physico-chemical properties as input properties: zeta potential, redox potential and dissolution rate. Their mean values for each NM are shown here.

There are a number of studies that found, at least partially, the same properties to be important for NM toxicity. Burello (2017) performed a regression analysis on 43 oxide NMs to relate physico-chemical properties at the level of neutrophiles in BALF. He identified reactivity, surface charge, wettability and dissolution rate as the most predictive properties. This is in accordance with our findings. Cassano et al. (2016) predicted cytotoxicity assessed with different cell lines for 19 silica particles and found aspect ratio and zeta potential to yield the most important associations. This is not contradictory to our findings, because we did not systematically vary the aspect ratio of the particles in our study. In the publication of Singh and Gupta (2014), zeta potential turned out to be the most important property for linking the results of an apoptosis assay with 44 NM having different metal cores to the NM properties. The study of Cho et al. (2012) revealed zeta potential and dissolution as the most important properties influencing lung inflammation. In addition, Warheit et al. (2007a, 2007b) and Sayes et al. (2006) have also shown a correlation between high surface reactivity of NMs and inhalation toxicity. Drew et al. (2017) applied RFs to predict the potency group for pulmonary toxicity for six NMs and found zeta potential to be among the most predictive physico-chemical properties. Our results also confirm the rather generic statement of Arts et al. (2015) that intrinsic material properties like size or surface area alone are not sufficient to group NMs for predicting their toxicity.

Comparing the empirical frequencies of the votes, the full model

shows very strong preferences for the incorrect group for misclassifications of SiO_2_15_unmod, SiO_2_15_Phospho and TiO_2 NM-105. For the misclassification of SiO_2_15_unmod and SiO_2_15_Phospho, one potential reason might be related to the fact that across all physico-chemical properties these two NMs are very similar (see Supp. Fig. 1). As for the prediction of the class of SiO_2_15_unmod, this NM is left out in the leave-one-out cross-validation, it probably follows the same paths down the trees in the RF that SiO_2_15_Phospho took in the training step in many cases. As SiO_2_15_Phospho belongs to the opposite category this will result in a misclassification of SiO_2_15_unmod. The same is true in the other direction as well. In addition, zeta potential is among the most important properties in the full model. For TiO_2 NM-105, all NMs with a similar zeta potential belong to the passive class (see Supp. Fig. 1). This might be the major reason for the strong tendency to assign a passive label to it (and thus to misclassify it). Especially, the high empirical frequency of the wrong category of TiO_2 NM-105 is also retained in the reduced model. This might be due to the fact that the zeta potential is the most important predictor in that model as well.

Another reason why the model performs very poorly for TiO_2 NM-105 might be that the classifier trained here is highly biased towards silica-based NM. Thus, the applicability domain might also be limited to silica-based NMs or materials behaving very similar. Adding more titania to the training set or in general increasing the range of different core materials covered, might improve the classification performance for TiO_2 NM-105 (and other underrepresented material classes). This bias may also have substantial influence on the selection of the most relevant physico-chemical properties. Using a different set of NMs may therefore change the set of parameters leading to the best predictive model. However, this is not a limitation of the method but rather a limitation due to the fact that only few datasets exist and that those datasets that do exist are not standardized in the way they assess physico-chemical properties and/or toxicity and thus cannot easily be integrated. Independent of which machine learning tool is used, stable and reliable results for the selection of the most important properties for NM grouping may only be obtained if models are based on a larger number of NMs and material classes.

In addition to the limited dataset, unknown *in vivo* behavior for some of the NMs is a potential source of error as well. It is possible that some of the NMs which have not been assessed *in vivo* so far were assigned to the wrong category and thus the assumed ground truth may actually not reflect the reality completely. In that case, the model performance and detected most important physico-chemical properties might change drastically especially as we tested only a very limited number of NMs here. However, Wiemann et al. showed that there is quite good agreement between the macrophage assay and STIS results in general and thus we assume that most NMs are assigned to the correct category here.

Another important point when building a predictive model is the representation of the outcome variable. In this approach we used a binary categorization into active and passive materials. Another possibility would be the representation of the toxicity as a continuous variable. One commonly used method to obtain a continuous outcome variable for toxicity data is benchmark dose (BMD) modeling (EPA, 2012). While BMDs might improve the model, we did not use them in this study as several challenges exist. Most importantly, the BMD approach was not suitable to compare results obtained for the four assays performed in the macrophage assay. For these assays, we observed very different dose responses such that obtained BMDs might not necessarily be comparable. The dose-response curves thereby deviated in their shape as well as in the amount of change observed. In addition, *in vitro in vivo* correlations of BMDs would also have to be assessed to be able to compare the results for all NMs.

As mentioned before, another difficulty for reliably linking physico-chemical properties with toxicity is the fact that many techniques for measuring physico-chemical properties are not sufficiently adapted and tested for NMs and thus their results may not be reproducible or comparable between studies. Also, the best metric for the comparison of the toxicity effects of NMs is still discussed (Oberdoerster and Kuhlbusch, 2018). Doses corresponding to the same surface area are frequently assumed to be of higher relevance when comparing NM effects. However, so far no final conclusions have been drawn in that regard. Also, depending on the choice of metric, the most important physico-chemical properties predicted by the model may vary. Depending on whether the outcome variable is represented as a binary or as a continuous variable and whether a discriminating or a clustering approach is applied, the link between physico-chemical properties and toxicity might change as well (Aschberger et al., 2019; Drew et al., 2017).

In future models, the fact that misclassification of active NMs as passive is much more costly than *vice versa* should also be considered. This is due to the fact, that overlooking and not testing a hazardous NM may have drastic consequences while this is not true for misclassifying a passive NM as active and simply testing that NM without necessity. Thus, adapting the misclassification cost in such a way that the penalty for misclassifying an active material as passive is much higher than that for misclassifying a passive NM as active should be included in the model building process. In RF approaches, this can be achieved in different ways. Usually, weighting the misclassification costs differently for different classes is based on sampling or thresholding techniques (Drew et al., 2017) and can be easily included into the approach presented here.

With respect to the assessment of feature importance, the Gini importance is known to favor predictor variables with more categories over those with fewer categories (Strobl et al., 2007). Here, we assessed only continuous variables, such that this is not an issue in the present study. However, should additional categorical parameters be included in future models, this fact has to be considered. Variable importance values retrieved from MDA are more reliable on the one hand, but seem to overestimate the variable importance in case of highly correlated variables on the other (Strobl et al., 2008). In the case of NM toxicity prediction, down voting of highly correlated variables is not problematic, because one just aims to find a minimal set of predictive features and does not necessarily need all good predictors. Also, RFE has been shown to decrease issues arising due to highly correlated input variables (Darst et al., 2018; Gregorutti et al., 2017). For more complex RF modeling and including more diverse input parameters it might be necessary to explore more sophisticated methods of measuring variable importance and performing feature selection as presented by *e.g.* Strobl et al. (2007, 2008).

As mentioned earlier, results of the PCA are only reliable if certain assumptions are fulfilled, *e.g.* a linear relationship between the principle components and the input space, as well as statistically normal distributed variables. Here, these assumptions have not been assessed in detail. However, for some of the physico-chemical properties it can easily be seen that for the limited set of NMs assessed here the assumption of normally distributed values does not hold true. This is the case for variables like the relative density of the NMs which is the same for most materials studied here. Also, we cannot exclude the possibility of non-linear relationships between principle components or higher-order correlation which may not be resolved by PCA. Thus, results obtained by the PCA analyses should be handled with care. Instead, the RF approach does not make such strong assumptions and might thus lead to more reliable results in that case.

Apart from PCA, one could also apply other methods which are simpler than RFs but do not rely on strong assumptions like linearity or normality. One such method that would be able to relate the values of the physico-chemical properties to the toxicity of the NM is logistic regression. However, logistic regression has some limitations compared to RFs: While in RFs the importance of each physico-chemical property is automatically assessed in the context of all other available properties, in logistic regression each possible interaction has to be integrated as a separate term into the regression formula. While this is still possible for

low-dimensional data, more advanced models will have to include more potential descriptors from which the most predictive set of features has to be chosen afterwards. Thus, for high-dimensional data, using logistic regression is impractical. In that way, RFs are much more flexible compared to logistic regression. A benchmark study on a large set of different datasets has also shown better performance of RFs compared to logistic regression, especially in the case of a large number of input variables relative to the number of samples (Couronné et al., 2018). Another advantage of RFs is that categorical input variables, which are very likely to occur in NM toxicity prediction, can be integrated much easier than in logistic regression.

Independent of uncertainties in the results, this study was able to show how machine learning and feature selection strategies can be used for linking physico-chemical properties of NMs to their toxicity. These extracted physico-chemical properties may then be used to detect NMs which are similar in terms of their toxicity effect, *i.e.* for establishing grouping with respect to NM hazards. Here, we used a categorization into active and passive materials in accordance with previous studies (Wiemann et al., 2016; Landsiedel et al., 2014). However, the prediction algorithm may be extended to the case of multiple class labels or even continuous outcomes once a larger number of consistent data is collected. A better understanding of which of the many possible physio-chemical properties actually drive toxicity will certainly enable the selection of sufficiently similar NMs to achieve robust grouping. The properties and models developed in this study should be regarded as a first basis for how to further develop NM grouping procedures and how to better understand and interpret similarity between NM variants. However, further refinement of the models and external validation with more and different materials will be necessary for obtaining reliable predictions and resolving the misclassifications presented here. Multiple improvement strategies will be tested in our future work.

As not all misclassifications have been resolved yet, the predictors we included in our dataset do not seem to be sufficient to explain the complete underlying differences in mechanisms of toxicity. Thus, additional variables may have to be included in the model building step to improve the prediction accuracy. Therefore, in future we are planning to extend our set of input parameters by adding more computed theoretical descriptors (EU US Roadmap Nanoinformatics 2030, 2018; s.r.l., K.C., n.d.; SCC, n.d.), as well as descriptors based on protein coronas and multi-omics data. Furthermore, we are specifically searching for similar data sets that might be helpful for data integration and external validation of our model. In addition, we will also include NM descriptors measured in relevant media or bio-fluids.

5. Conclusion

This work aimed to demonstrate how machine learning approaches can be used to determine sets of physico-chemical properties which are predictive for certain NM toxicity endpoints. Here, we applied different machine learning tools to a set of eleven NMs to identify the combination of physico-chemical properties that is most predictive for inhalation toxicity within this set. This was done for two different purposes 1) to identify physico-chemical properties that strongly correlate with toxicity and 2) to propose a reduced set of physico-chemical properties that will not only facilitate NM grouping but at the same time may support further research as well.

To achieve this, we assessed the suitability of an unsupervised approach based on PCA combined with kNN as well as a supervised approached based on RFs with and without prior feature selection for predicting NM toxicity. The best performance in terms of balanced accuracy of the prediction model was obtained with the reduced RF model after backward RFE. Variable selection based on the MDA led to equal or better results than Gini importance in all cases. The three most important features zeta potential, redox potential, and dissolution rate were among the highest ranking variables in unsupervised as well as supervised analyses with both the full as well as the reduced model after RFE.

However, in this study only a very limited set of NMs with a material focus on silica NMs was tested. In order to obtain reliable and generalizable results, the number of studied NMs has to be extended and a range of different material classes has to be tested. The pre-requisite for this is the standardization of measurements and the improvement of predictive assays which allow for meta-analyses of the results of multiple studies. The incorporation of benchmark materials like TiO_2 NM-105 is very useful in this regard. As so far, test methods for the analysis of NMs are not standardized, uncertainties with respect to suitability and reliability of the methods exist. Thus, uncertainties with respect to data quality are present in NM datasets in general. Studying benchmark materials allows for comparison with results from other studies and for estimating the reliability and reproducibility of applied methods. In addition, TiO_2 NM-105 has been used in many other case studies and can thus be used to compare and integrate datasets from different studies. This forms the basis for developing more reliable and robust predictive models incorporating a wide range of core materials and different nanoforms for each material class.

This study should be understood as a proof-of-concept study on how to use machine learning tools to build predictive models and to detect physico-chemical properties that are of high importance for NM toxicity and can be used for NM grouping. With extended datasets, the evaluation strategy presented here may add a significant contribution to understanding how physico-chemical properties of NMs may be linked to toxicity in future. The study provides valuable insights into which methods may be applied and further developed to decrease the complexity of input parameters in order to facilitate NM grouping.

Crucial next steps will be the enlargement of datasets and useful descriptors and external validation of the predicted major descriptors once reliable models on extended datasets have been built.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.impact.2019.100179.

Conflict of interest

The authors declare no conflicts of interests.

Acknowledgements

We thank BASF SE and Evonik Resource Efficiency GmbH for providing NMs. Special thanks go to Wendel Wohlleben for supporting us with helpful discussions and giving useful advice for this manuscript.

Funding

This work was supported by of the SIINN ERA-NET, funded under the ERA-NET scheme of the Seventh Framework Program of the European Commission and is part of the NanoToxClass project (BMBF Grant Agreement No. 03XP0008). Part of the study was also funded within the project nanoGRAVUR (BMBF Grant Agreement No.03XP0002).

References

Amaratunga, D., Cabrera, J., Lee, Y.-S., 2008. Enriched random forests. Bioinformatics 24 (18), 2010–2014.

Arts, J.H., et al., 2015. A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). Regul. Toxicol. Pharmacol. 71 (2 Suppl), S1–27.

Arts, J.H.E., et al., 2016. Case studies putting the decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping) into practice. Regul. Toxicol. Pharmacol. 76, 234–261.

Aschberger, K., et al., 2019. Grouping of multi-walled carbon nanotubes to read-across genotoxicity: a case study to evaluate the applicability of regulatory guidance. Computational Toxicology 9, 22–35.

Braakhuis, H.M., et al., 2014. Physicochemical characteristics of nanomaterials that affect pulmonary inflammation. Part Fibre Toxicol 11, 18.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breiman, L., *Manual for Setting Up, Using, and Understanding Random Forest V4.0.* 2003.

Burello, E., 2017. A mechanistic model for predicting lung Inflammogenicity of oxide nanoparticles. Toxicol. Sci. 159 (2), 339–353.
Cassano, A., et al., 2016. Comparing the CORAL and Random Forest approaches for modelling the in vitro cytotoxicity of silica nanomaterials. Altern. Lab. Anim 44 (6), 533–556.
Cho, W.S., et al., 2011. Progressive severe lung injury by zinc oxide nanoparticles; the role of Zn2+ dissolution inside lysosomes. Part Fibre Toxicol 8, 27.
Cho, W.S., et al., 2012. Zeta potential and solubility to toxic ions as mechanisms of lung inflammation caused by metal/metal oxide nanoparticles. Toxicol. Sci. 126 (2), 469–477.
Christensen, F.M., et al., 2010. Nano-silver - feasibility and challenges for human health risk assessment based on open literature. Nanotoxicology 4 (3), 284–295.
Couronné, R., Probst, P., Boulesteix, A.-L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 19 (1), 270.
Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet. 19 (1), 65.
Dekkers, S., et al., 2016. Towards a nanospecific approach for risk assessment. Regul. Toxicol. Pharmacol. 80, 46–59.
Delaval, M., et al., 2017. Assessment of the oxidative potential of nanoparticles by the cytochrome c assay: assay improvement and development of a high-throughput method to predict the toxicity of nanoparticles. Arch. Toxicol. 91 (1), 163–177.
Drew, N.M., et al., 2017. A quantitative framework to group nanoscale and microscale particles by hazard potency to derive occupational exposure limits: proof of concept evaluation. Regul. Toxicol. Pharmacol. 89, 253–267.
ECHA, *Guidance on information requirements and chemical safety assessment,*. 2008, European chemicals agency (ECHA) Helsinki, Finland.
ECHA, 2016. JRC, and RIVM, *Usage of (eco)toxicological data for bridging data gaps between and grouping of nanoforms of the same substance, Elements to consider*.
ECHA, *Appendix R.6–1 for nanomaterials applicable to the Guidance on QSARs and Grouping of Chemicals*. 2017, European chemicals agency (ECHA): Helsinki, Finland.
EPA, *Benchmark Dose Technical Guidance*. 2012, U.S. Environmental Protection Agency: Washington, USA.
Findlay, M.R., et al., 2018. Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. Environmental Science: Nano 5 (1), 64–71.
Forster, S., S. Olveira, and S. Seeger, *Nanotechnology in the market: Promises and realities*. Vol. 8. 2011.
Froehlich, E., 2012. The role of surface charge in cellular uptake and cytotoxicity of medical nanoparticles. Int. J. Nanomedicine 7, 5577–5591.
Gandon, A., et al., 2017. Surface reactivity measurements as required for grouping and read-across: an advanced FRAS protocol. J. Phys. Conf. Ser. 838, 012033.
Gao, X., Lowry, G.V., 2018. Progress towards standardized and validated characterizations for measuring physicochemical properties of manufactured nanomaterials relevant to nano health and safety risks. NanoImpact 9, 14–30.
Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recogn. Lett. 31 (14), 2225–2236.
Goldberg, E., et al., 2015. Prediction of nanoparticle transport behavior from physicochemical properties: machine learning provides insights to guide the next generation of transport models. Environmental Science: Nano 2 (4), 352–360.
Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. Stat. Comput. 27 (3), 659–678.
Guyon, I., et al., 2002. Gene selection for Cancer classification using support vector machines. Mach. Learn. 46 (1), 389–422.
Ha, M.K., et al., 2018. Toxicity classification of oxide nanomaterials: effects of data gap filling and PChem score-based screening approaches. Sci. Rep. 8 (1), 3141.
EU US Roadmap Nanoinformatics 2030. https://doi.org/10.5281/zenodo.1486012.
Hellack, B., et al., 2017. Analytical methods to assess the oxidative potential of nanoparticles: a review. Environmental Science: Nano 4 (10), 1920–1934.
Izak-Nau, E. and M. Voetz, *As-produced: intrinsic physico-chemical properties and appropriate characterization tools*, in *Safety of Nanomaterials along Their Lifecycle*, W. Wohlleben, Kuhlbusch, T., Schnekenburger, J., Lehr, C.M., editor. 2014, CRC Press: Boca Raton.
Lamon, L., et al., 2018. Grouping of nanomaterials to read-across hazard endpoints: from data collection to assessment of the grouping hypothesis by application of chemoinformatic techniques. Part Fibre Toxicol 15 (1), 37.
Landsiedel, R., et al., 2014. Application of short-term inhalation studies to assess the inhalation toxicity of nanomaterials. Part Fibre Toxicol 11, 16.
Liu, R., et al., 2015. Prediction of nanoparticles-cell association based on corona proteins and physicochemical properties. Nanoscale 7 (21), 9664–9675.
Lynch, I., Weiss, C., Valsami-Jones, E., 2014. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. Nano Today 9 (3), 266–270.
Marzaioli, V., et al., 2014. Surface modifications of silica nanoparticles are crucial for their inert versus proinflammatory and immunomodulatory properties. Int. J. Nanomedicine 9, 2815–2832.
NanoToxClass. *SOP Dispersion*. Available from: https://www.nanopartikel.info/files/projekte/NanoToxClass/NanoToxClass-SOP_Dispersion_by_cup_horn_sonication_V2.0.pdf, 2017.
NanOxiMed. *SOPs*. Available from: https://www.nanopartikel.info/projekte/era-net-siinn/nanoximet/veroeffentlichungen-nanoximet, 2014 - 2016.
Nel, A.E., 2013. Implementation of alternative test strategies for the safety assessment of engineered nanomaterials. J. Intern. Med. 274 (6), 561–577.
Oberdoerster, G., Kuhlbusch, T.A.J., 2018. In vivo effects: methodologies and biokinetics of inhaled nanomaterials. NanoImpact 10, 38–60.
OECD, *Guidance on grouping of chemicals*. 2016, Organisation for economic cooperation and development (OECD): Paris, France.
Oomen, A.G., et al., 2014. Concern-driven integrated approaches to nanomaterial testing and assessment–report of the NanoSafety Cluster Working Group 10. Nanotoxicology 8 (3), 334–348.
Oomen, A.G., et al., 2015. Grouping and read-across approaches for risk assessment of nanomaterials. Int. J. Environ. Res. Public Health 12 (10), 13415–13434.
s.r.l., K.C. *DRAGON 7.0*. Available from: https://chm.kode-solutions.net/products_dragon.php.
Sayes, C.M., et al., 2006. Correlating nanoscale titania structure with toxicity: a cytotoxicity and inflammatory response study with human dermal fibroblasts and human lung epithelial cells. Toxicol. Sci. 92 (1), 174–185.
Sayes, C.M., Smith, P.A., Ivanov, I.V., 2013. A framework for grouping nanoparticles based on their measurable characteristics. Int. J. Nanomedicine 8 (Suppl. 1), 45–56.
(SCC), S.C.C. *MOPAC2016*. Available from: http://openmopac.net/MOPAC2016.html.
Sellers, K., et al., 2015. Grouping Nanomaterials - a Strategy towards Grouping and Read-across.
Singh, K.P., Gupta, S., 2014. Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. RSC Adv. 4 (26), 13215–13230.
Sizochenko, N., et al., 2014. From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. Nanoscale 6 (22), 13986–13993.
Strobl, C., et al., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8 (1), 25.
Strobl, C., et al., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9 (1), 307.
Utembe, W., et al., 2015. Dissolution and biodurability: important parameters needed for risk assessment of nanomaterials. Particle and Fibre Toxicology 12 (1), 11.
Warheit, D.B., et al., 2007a. Pulmonary toxicity study in rats with three forms of ultrafine-TiO2 particles: differential responses related to surface properties. Toxicology 230 (1), 90–104.
Warheit, D.B., et al., 2007b. Pulmonary bioassay studies with nanoscale and fine-quartz particles in rats: toxicity is not dependent upon particle size but on surface characteristics. Toxicol. Sci. 95 (1), 270–280.
Wiemann, M., et al., 2016. An in vitro alveolar macrophage assay for predicting the short-term inhalation toxicity of nanomaterials. J Nanobiotechnology 14, 16.

3.2 Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods

Aileen Bahl[a], Bryan Hellack[b,c], Martin Wiemann[d], Anna Giusti[a], Kai Werle[e], Andrea Haase[a,*,1], Wendel Wohlleben[e,**,1]

a *German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany*

b *Institute for Energy and Environmental Technology e.V. (IUTA), Duisburg, Germany*

c *German Environment Agency (UBA), Dessau, Germany*

d *IBE R&D Institute for Lung Health gGmbH, Münster, Germany*

e *BASF SE, Department of Material Physics, Ludwigshafen, Germany*

This chapter was published as a research article in *NanoImpact* in 2020 (https://doi.org/10.1016/j.impact.2020.100234).

The supplementary information for this publication can be found in Annex II.

NanoImpact 19 (2020) 100234

Contents lists available at ScienceDirect

NanoImpact

journal homepage: www.elsevier.com/locate/nanoimpact

Research paper

Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods

Aileen Bahl[a], Bryan Hellack[b,c], Martin Wiemann[d], Anna Giusti[a], Kai Werle[e], Andrea Haase[a,*,1], Wendel Wohlleben[e,**,1]

[a] German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany
[b] Institute for Energy and Environmental Technology e.V. (IUTA), Duisburg, Germany
[c] German Environment Agency (UBA), Dessau, Germany
[d] IBE R&D Institute for Lung Health gGmbH, Münster, Germany
[e] BASF SE, Department of Material Physics, Ludwigshafen, Germany

ARTICLE INFO

Editor: Bernd Nowack

Keywords:
Nanomaterials
Surface reactivity
Oxidative potential
Functional assays
Test strategy
Logistic regression

ABSTRACT

Nanomaterials (NMs) can be manufactured in plenty of variants differing in their physicochemical properties. Functional assays can be highly useful to cope with the enormous variability by supporting prioritization and categorization. Oxidative potential (OP) seems to be in particular important in this context and different assays are available. However, their reliability and predictivity are not well-characterized.

This study compares four different test methods for measuring NM OP. Reactive oxygen species (ROS) generation was measured on a set of 35 different materials, all extensively characterized with respect to physicochemical properties and most of them with respect to toxicity. Different acellular assays were applied, namely electron spin resonance (ESR) spectroscopy using CPH spin probe and DMPO spin trap, and the ferric reduction ability of serum (FRAS) assay. In addition, protein carbonylation as a marker for oxidative protein damage was analyzed in NRK-52E cells. All assays were assessed individually for their predictivity compared to established toxicological endpoints. We also aimed to identify the optimal assay combination using multivariate logistic regression and other statistical measures.

BET surface area-based doses were more suitable to relate surface reactivity to toxicity. In addition, normalization to the deposited dose was advantageous for cellular assays as it improved the predictivity for *in vitro* as well as *in vivo* toxicity. The carbonylation assay, potentially in combination with ESR (DMPO spin trap) or FRAS assay, led to the best predictive performance.

In summary, we propose a testing strategy for NM OP and demonstrated the applicability in an extended case study on 35 materials. This work is an important contribution towards reliable grouping and testing strategies for NMs.

1. Introduction

Nanomaterials (NMs) can be produced from different chemical substances and their physicochemical properties can be precisely fine-tuned to meet specific functional needs. This results in a theoretically unlimited number of NM variants, differing in physicochemical properties such as size, shape or surface chemistry. Hence, they may also show different toxicological profiles rendering hazard and risk assessment time and resource intense. Grouping and read-across are powerful tools to reduce the amount of necessary experimental testing and thus gained huge interest in the last decade (ECHA, 2008; OECD, 2014). NMs with similar physicochemical and toxicological properties can be assessed as a group. Within an established group available data from data-rich source materials can be used to predict properties and/or toxicities of data-poor target materials. Within a group, comparisons to known benchmark materials can be used as a powerful approach to reduce the uncertainty related to health risks of innovative, new materials. In addition, such approaches are also useful for safety assessment of well-established materials such as fillers and pigments that are produced in megaton quantities in many different (nano-)forms (l'Environment Md, 2015; Wohlleben et al., 2017).

* Correspondence to: A. Haase, German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Germany.
** Correspondence to: W. Wohlleben, BASF SE, Department of Material Physics and Department of Experimental Toxicology, Germany.
E-mail addresses: andrea.haase@bfr.bund.de (A. Haase), wendel.wohlleben@basf.com (W. Wohlleben).
[1] authors contributed equally

https://doi.org/10.1016/j.impact.2020.100234
Received 17 February 2020; Received in revised form 18 June 2020; Accepted 18 June 2020
Available online 06 July 2020

Several frameworks for NM grouping have already been proposed such as the approach developed by the US-Canada Regulatory Cooperation Council (RCC RCC, 2003), the DF4NanoGrouping framework (Arts et al., 2015), the MARINA grouping and read across approach (Oomen et al., 2015) and others, as summarized in several review articles (Giusti et al., 2019; Oomen et al., 2018). Almost every NM grouping framework acknowledges surface reactivity as a central parameter (Arts et al., 2015; Oomen et al., 2018; Arts et al., 2014; Arts et al., 2016; Braakhuis et al., 2014; Collier et al., 2015; Kuempel et al., 2012; Nel et al., 2013). The underlying idea is that intrinsic factors like shape, size, coating, composition, crystallinity, impurities, etc. may modulate the surface reactivity and thus account for differences in the toxicological potency between different (nano-)forms of a substance. As the relationship between intrinsic physicochemical properties and surface reactivity is not trivial and hardly can be predicted, the measurement of surface reactivity is a good starting point to be implemented in grouping and categorization frameworks. A high surface reactivity is assumed to cause the generation of high quantities of reactive oxygen species (ROS) that may then lead to non-selective oxidation of biomolecules, to oxidative cell damage and to oxidative stress (Sies, 2015; Stone et al., 2007; Xia et al., 2007). NMs can produce ROS by different mechanisms. Fenton-like reactions leading to the generation of hydroxyl radicals are among the most common ones. Other relevant mechanisms are catalytic processes at the NM surface or radical production via dissolved (metal) ions (Driessen et al., 2015; He et al., 2014). ROS can be divided into two classes of molecules. The first type comprises very reactive and thus short-lived species, such as singlet oxygen or free radicals (e.g. superoxide or hydroxyl radicals). They can only reach nearby cellular targets and interact with them. In addition, there are more stable and long-lived species (such as hydrogen peroxide), which can thus also affect distant targets (Hellack et al., 2017a).

The imbalance between ROS generation and ROS detoxification leads to elevated ROS levels within cells and is called oxidative stress. This can induce various adverse outcomes such as cytotoxicity, genotoxicity or inflammation (Unfried et al., 2007; Marano et al., 2011; Lanone et al., 2009; Halliwell and Whiteman, 2004; Fu et al., 2014; Borm and Müller-Schulte, 2006). Most of these responses are triggered by alterations in cellular pathways and/or by oxidative damage of biomolecules like proteins, lipids or DNA (Nel et al., 2013; Xia et al., 2007; Driessen et al., 2015; Hellack et al., 2017a). Proteins are the most common targets for cellular ROS. It has been estimated that up to 70% of all ROS finally end up as direct or indirect protein adducts (Davies, 2005; Dalle-Donne et al., 2006). Many different NMs are able to induce oxidative stress (He et al., 2014).

The term ROS comprises different short-lived radicals and molecular species. It is neither efficient nor straightforward to detect all of them, even if this would allow for important insights into the underlying mode of action (MoA). The measurements are in particular difficult in dynamic biological environments (He et al., 2014). Accordingly, functional assays are useful to determine the production of specific ROS and of specific consequences of oxidative stress. In order to obtain reliable results, the outcomes of different assays have to be compared and potentially combined to form one parameter, referred to herein as the oxidative potential (OP). The OP as obtained from individual or combined assays should then be correlated to the outcomes of different *in vivo* and *in vitro* toxicity assays to assess the predictive power of surface reactivity for the specific toxicological endpoint.

Electron spin resonance (ESR) spectroscopy allows qualitative identification and quantitative measurements of free radical species in acellular and cellular environments. It measures the transitions between electron spin states of paramagnetic molecules and it can thus be used to study species with at least one unpaired electron. The ESR spectrum is specific for each radical and its intensity is proportional to the amount of radicals. Two main challenges affect ROS detection: their high reactivity (and thus short life time), and their very low concentrations (He et al., 2014). Therefore, one usually uses spin probes/labels which react with non-radical ROS to make them visible by ESR or spin traps to stabilize the free short-living radicals. In the current study, the spin probe CPH (1-hydroxy-3-carboxy-pyrrolidine) (Hellack et al., 2017b) and the spin trap DMPO (5,5-dimethylpyrroline N-oxide) (Shi et al., 2003) were used. CPH is able to react mainly with singlet oxygen, superoxide radicals and peroxynitrite to give the stable nitroxide radical 3-carboxy-2,2,5,5-tetramethylpyrrolidine 1-oxyl, which then can be detected by ESR spectroscopy (He et al., 2014). DMPO is a nitrone spin trap that can trap mainly hydroxyl and superoxide radicals. The advantage of ESR is that one gets very detailed insights into the detected ROS species. However, special equipment is needed, which is not available in most laboratories.

Ferric reduction ability of serum (FRAS) assay measures ROS formation in human blood serum (HBS). Thus, it uses a biologically relevant medium. The FRAS assay is a more indirect read-out of ROS generation, measuring the total antioxidant depletion in HBS. The surface reactivity is derived from the ability of the NM-preincubated HBS to reduce Fe^{3+} to Fe^{2+} detected by a color change of the solution. The FRAS assay is able to differentiate between NMs at the lower end of the reactivity scale with higher sensitivity compared to ESR (Gandon et al., 2017). It detects a broad range of ROS species at once.

Differently to acellular ESR and FRAS, the measurement of protein carbonylation is a cell-based assay, which measures an important and widely assessed biomarker for protein oxidation (Stadtman, 2006). Protein carbonylation is an irreversible post-translational modification, often related to a loss or a decrease of protein function (Driessen et al., 2015). Carbonylated proteins accumulate in cells under oxidative stress conditions and are associated with several human diseases (Dalle-Donne et al., 2003) such as neurodegenerative diseases (for example Alzheimer's or Parkinson's diseases (Smith et al., 1991; Uttara et al., 2009)), diabetes (Telci et al., 2000), chronic lung diseases or with natural processes, such as aging (Rudzińska et al., 2020). The measurement of the amount of protein carbonylation induced in cells as a consequence of NM exposure is considered as an indirect measure of ROS production (Hellack et al., 2017a). The assay exploits the specific derivatization of the carbonyl moieties with dinitrophenylhydrazine. The resulting dinitrophenylhydrazone is then revealed using an immunoassay (Giusti and Haase, 2020). While cellular assays are assumed to be of higher biological relevance, they are more difficult to standardize and they are affected by biological variability. In addition, they do not represent an intrinsic property of the NM but instead are highly dependent on the cell line and the cell culture conditions.

Several studies already suggested combining multiple assays for measuring NM OP and/or oxidative stress. Riebeling et al. (2016) described a tiered approach in which first an acellular assay like ESR or FRAS should be performed (tier 1) followed by an *in vitro* assay, e.g. protein carbonylation (tier 2) and, in the last step, an approach giving more mechanistic insights, e.g. mass spectrometry of carbonylated proteins and analysis of regulation of the pathways they belong to (tier 3). Similarly, Hellack et al. (2017a) suggested using acellular assays like ESR or FRAS assays, which might then be accompanied by cellular assays for complementary read-outs and omics approaches for other oxidative stress independent responses as well as additional evidence for false positive and false negative results. While the last step of mechanistic insights is not further evaluated in this study, we assessed the suitability of different acellular and cellular assays for relating OP to toxicity outcomes using a larger number of materials. In total 35 different materials were included in this study, all of which are extensively characterized with respect to their physicochemical properties. Moreover, most of the materials also have been well-characterized for their toxicological profiles *in vivo* and/or *in vitro* (Wiemann et al., 2016).

The surface reactivity of NMs is a strong candidate to contribute to adverse reactions in lung tissue. A high surface reactivity is assumed to cause the generation of high quantities of reactive oxygen species (ROS), which may then lead to non-selective oxidation of biomolecules, thus to oxidative cell damage and eventually to oxidative stress and

inflammation. Therefore, the results from the above mentioned four tests (that is ESR with CPH spin probe or DMPO spin trap, FRAS assay, carbonylation assay) were correlated with the results of the short-term inhalation studies (STIS), a method which describes multiple kinds of acute adverse effects of inhaled particles on the lung and which differentiates between active and passive NM. Another correlation was made with *in vitro* data from the alveolar macrophage test. This strategy was chosen for three reasons.

Firstly, NM inhalation is considered the exposure route of the highest concern. Accordingly, we built our study on data describing the lung toxicity of NMs following inhalation. Secondly, numerous STIS results up to now provide the largest pool of reliable inhalation toxicity data all of which were obtained under the conditions of good laboratory practice (GLP). In these inhalation studies all types of adverse reactions evident from BALF data, lung histopathology, and also gross biopsies are routinely considered. According to the evaluation routines of STIS a NM is considered as to be "active" if a particle concentration of < 10 mg per m^3 (administered for 5 days, nose only inhalation exposure, 21 days follow up) leads to any significant adverse effects. These may include all signs of inflammation, granuloma formation, beginning fibrosis, or other pathological findings. A high oxidative potential of a NM may directly or indirectly contribute to this panel of adverse reactions and, therefore, was correlated with the STIS data. Thirdly, the alveolar macrophage assay based on NR8383 cells, as an *in vitro* test, has a considerable potential to predict the results obtained with STIS. Accordingly, the assay is increasingly being used as a simple tool to differentiate between active and passive NMs. Since not only inflammation but also a damage of the macrophage population and local H_2O_2 production (e.g. via DNA adduct formation) may lead to a damage to the lung epithelium, the macrophage assay tests for (1.) particle-induced cytotoxicity (via LDH release), (2.) activation (glucuronidase release), (3.) cell-mediated oxidative properties of NMs (H_2O_2 release), and (4.) pro-inflammatory properties (release of TNF-α). The cellular particle dose used in this assay, under ideal conditions (i.e. complete gravitational settling), overlaps with the macrophage loading in STIS, which facilitates the interpretation of results. Nevertheless, evaluation and interpretation of the four single results and especially the subsequent active/passive classification of NM (in analogy to STIS), demanded the introduction of a surface-based threshold value, and also of a 2-out-of-4 criterion. As outlined and explained in detail by Wiemann et al. (2016), this is an empiric and also pragmatic *in vitro* approach. However, it reduced the number of false positives, excluded false negatives, and resulted in a 95% accuracy in predicting the active/passive categorization derived from the well-established animal inhalation experiments. Due to its predictive properties we used the active/passive allocation of the macrophage assay in parallel to STIS results and explored possible correlations with the oxidative potential of NMs measured by ESR, FRAS and carbonylation.

In order to consider the influence of NM properties on the specific assays, we selected several materials of different chemical compositions and different variants (e.g. size, shape, crystallinity, surface chemistry) of the same chemical composition. We ranked and compared the NM variants of the same chemical composition according to their OP and also compared results over a wider range of materials. Our study further justifies the proposal of positive and negative benchmark materials for surface reactivity, which are of fundamental importance for the development of reliable and accepted test methods.

2. Materials and methods

2.1. (Nano-) materials

We studied several materials including nano- and non-nanomaterials of in total twelve different material classes. These were different variants of silica, aluminosilicates, iron oxides, titania, ceria, zinc oxides, diketopyrrolopyrrol pigments, carbon-, copper- and wolfram-based materials. In addition, manganese oxide and barium sulfate were included as positive (Arts et al., 2015; Arts et al., 2016) and negative (Arts et al., 2015; Arts et al., 2016; Buesen et al., 2014; Landsiedel et al., 2014) benchmarks, respectively. An overview of all materials, including the supplier information and a summary of the most important physicochemical properties is given in Table 1. For a more complete overview see Wohlleben et al. (2019) and Maser et al. (2015). Please notice that Cu-Phthalocyanine non-halogenated is called Pigment Blue in Wiemann et al. (2016). The NMs were obtained from different suppliers, which are listed in Supplementary Table 1a.

2.2. Material dispersion

Unless otherwise noted, all materials were dispersed following a SOP jointly developed by the projects nanoGRAVUR and NanoToxClass based on cup horn sonication (https://www.nanopartikel.info/files/projekte/NanoToxClass/NanoToxClass-SOP_Dispersion_by_cup_horn_sonication_V2.0.pdf). In brief, a final power of 6 W was applied. Stock solutions of hydrophilic NMs were generated with a concentration of 0.5 mg/ml in water or in cell culture medium (DMEM w/o phenol red and L-glutamine, high glucose, PAN Biotech GmbH supplemented with 2 mM L-glutamine, 0.1 mg/ml penicilline/streptomycine, 25 mM hepes buffer) without serum. For assessing protein carbonylation, the stock solutions were diluted to the final concentrations using DMEM medium supplemented with 10% non-heat inactivated fetal calf serum (FBS Good from PAN Biotech). In the case of the two CuPhthalo NMs, 100 µg/ml Pluronic was added in order to obtain a better dispersion.

2.3. Material characterization

2.4. Analytical ultracentrifugation

Analytical ultracentrifugation was used to determine the amount of material deposited to the bottom of a cell dish. These values were used to normalize the values of the protein carbonylation assay.

By synchronizing an optical detection system to the centrifugal frequency, analytical ultracentrifugation can track colloids during their settling under increased gravitational acceleration. The sample preparation at a concentration of 0.5 mg/ml in DMEM with 10% FCS was identical to the preparation for the protein carbonylation assay. Here we used Beckman XLI machines and evaluated the distribution of sedimentation coefficients by the SedFit v15 software. The distribution of sedimentation coefficients can be simply integrated over the dose that is deposited by sedimentation during incubation (Sauer et al., 2015). This approach to dosimetry neglects transport by diffusion, which may increase the deposited dose above the predicted level especially for the colloidal SiO_2 materials (Sauer et al., 2015). However, the approach has the advantage that for evaluation one does not need to know the effective density but inherently incorporates the modulation of sedimentation behavior due to protein corona and agglomeration and even distributions thereof which is not possible in modeling approaches (DeLoid et al., 2014).

2.5. Assays for determination of OP

2.5.1. Electron spin resonance (ESR) spectroscopy

All measurements were performed in deionized water (dH_2O) using the spin probe 1-hydroxy-3-carboxy-pyrrolidine (CPH) (Hellack et al., 2017b) or the spin trap5,5-dimethylpyrroline N-oxide (DMPO) (Shi et al., 2003). CPH is able to react with singlet oxygen, superoxide radicals and peroxynitrite to the stable nitroxide radical 3-carboxy-2,2,5,5-tetramethylpyrrolidine 1-oxyl (He et al., 2014). DMPO is a nitrone spin trap that can trap hydroxyl and superoxide radicals. Surface reactivity values were obtained as the ratio of radical formation with

Table 1
Overview of tested materials including key physicochemical properties (Wohlleben et al., 2019). "ns" indicates that the result was not significant against the limit of detection. *data from JRC (2013).

Class	Material	Primary particle size (TEM/SEM mean diameter) (nm)	Specific surface area (BET) [m^2/g]	Surface charge (zeta potential) at pH 7 (mV)	Hydrophobicity (water contact angle)	Solubility in water (OECD screening LoD or value, metal ion) (ppm)	Dissolution rate in relevant human medium (lysosomal dissolution rate k) [ng/cm^2/h]
Silica	SiO_2_15_unmod	15	200	−39	0	27	0.2
	SiO_2_15_Amino	15	200	0	0	13	0.27
	SiO_2_15_Phospho	15	200	−43	0	12	0.45
	SiO_2 NM-200*	50	189	−45			
	SiO_2 NM-203	26.2	213	−24	44	56	0.4
	SiO_2Levasil 50	55	50	−39			
	SiO_2Levasil 100	30	100				
	SiO_2Levasil 300	15	200	−32.7			
	SiO_2Aerosil 200	9	300				
Aluminosilicates	Kaolin	279	24	−53	10	0.8	1.3
	Bentonite	1362	52	−31	10	0.1	0.65
	DQ12 (Quartz)						
Iron oxides	Fe_2O_3nanoform A	12	107	−27	10	ns	0.04
	Fe_2O_3nanoform B	37	30	18	10	ns	0.04
	Fe_2O_3larger	48	12	−55	10	ns	0.1
Titania	TiO_2 NM-102	34.8	80	−33	10	ns	0
	TiO_2 NM-105	21	51	−17	60	ns	0.013
	TiO_2 non-nano	204	15	36	10	ns	0
Ceria	CeO_2 NM-211	15	66	−24	10	0.1	0.14
	CeO_2 NM-212	40	27	15	60	0.1	0.06
Copper-based	CuO	24	34	−34	10	97	283
	Cu-Phthalocyanine	17	53	−11	138	ns	0.76
	non-halogenated	39	69	−38	163	ns	0.42
	Cu-Phthalocyanine halogenated						
Zinc oxides	ZnO NM-110	42	12	30	10	3.6	204
	ZnO NM-111	80	14	−25	152	3.3	177
Diketopyrrolopyrrol pigments	DPP_premixed	400	17	−30.4	103		
	DPP nano	43	94	−16	135		
	DPP non-nano	233	16	−41	136		
Carbon-based	Carbon black		56.5	−16.5	148		
	Graphene oxide			−16.2			
	Graphene 1-layer		559	−45.4	93		
	Graphene multilayer		17.6	−40.7	79		
Tungsten-based	WS_2	100		−49.2			
Manganese-based	Mn_2O_3 (pos. control)	36	20	−5		0.2	2
Barium-based	$BaSO_4$ NM-220 (neg. control)	32	41	−37	10	6	10

and without the NM. CPH was used in PBS in the presence of the chelator desferroxamine (DFO, 0.1 mM) to prevent transition metal-induced reactions (Driessen et al., 2015) and incubated for 10 min at 37 °C before measurements. DMPO was mixed with PBS and H_2O_2 to induce Fenton-like reactions and incubated for 15 min at 37 °C in a shaking water bath in the dark before measurements. ESR measurements were performed on a ESR 300spectrometer MiniScope (Magnettech, Berlin, Germany) at room temperature (RT) with the following settings: magnetic field 3365 G, sweep width 100 G, scan time 30 s, number of scans 3, modulation amplitude 1.975 G, and receiver gain 1000 (Hellack et al., 2017b).

ESR measurements were performed with same mass concentration for all NMs (CPH: 0.5 mg/ml and DMPO: 0.25 mg/ml) as well as with same BET surface area for all NMs (1 m^2/ml, exception: NM-110 and NM-600: 0.5 m^2/g (results extrapolated to 1 m^2/g)).

2.5.2. FRAS

In the current study, the multi-dose protocol of the FRAS assay presented in Gandon et al. (2017) is used as it was shown to be much more sensitive especially for small to medium responses and details can be found there. In brief, NMs are added to human blood serum (HBS) and removed by centrifugation after 90 min of incubation. A solution containing Fe^{3+} is then added to the HBS. After incubation, the amount of reduction from Fe^{3+} (transparent) to Fe^{2+} (blue) is detected optically in terms of color change (absorbance at 593 nm). The darker the blue color detected in this process, the more antioxidant species are used which is (at least partly) caused by ROS scavenging. This usage of antioxidant species is measured as the biological oxidative damage (BOD) and is expressed in Trolox equivalent units (TEU). Trolox, a water-soluble analog of vitamin E, is used for calibration of the absorption signal.

The FRAS assay was performed with uniform mass doses with five different concentrations over all NMs (0.75, 2, 5.5,15 and 40 mg/ml) as well as with same BET surface area doses for each NM (1 m^2/ml). The data analysis was then performed as described in Section 2.8.

2.5.3. Protein carbonylation

In the current study, protein carbonylation is quantified similarly to what was described in Driessen et al. (2015), following the procedure in Giusti and Haase (2020). NMs were tested at final concentrations of 10, 25 and 50 μg/ml in NRK-52E cell cultures (obtained from DSMZ, German collection of microorganisms and cell cultures). Cells were seeded in 6-well plates at a density of 0.15 × 10^6 cells/well. After a settling time of 24 h, cells were incubated with NMs for another 6 h and then lysed with modified RIPA buffer. Protein concentrations were measured using a Bradford assay according to manufacturer instructions. Protein carbonyls were assessed using the Oxiblot® kit (Millipore), which uses the derivatization reagent 2,4-dinitrophenylhydrazine (DNPH). DNPH reacts with the carbonyl groups and

is transformed to 2,4-dinitrophenylhydrazones which can be detected with specific antibodies (Riebeling et al., 2016). Protein carbonyls were assessed here using a Dot Blot technique, which permits to assess the signal of protein carbonylation and to normalize it to the total deposited protein determined by Colloidal Gold Total Protein Stain (Bio-Rad). The results are expressed as fold increase in comparison to untreated cells. Semi-quantification relative to a negative control (signal from untreated cells) was performed using Image Lab™ software (Bio-Rad) with a global background subtraction. For each NM, carbonylation of three independent biological replicates was measured.

Protein carbonylation measurements were only performed with uniform masses (10, 25 and 50 μg/ml) tested for each NM. In the protein carbonylation assay, three different concentrations of NMs were applied to the cells (Supplementary Fig. 1). For comparison to the other assays, we only considered data obtained for the medium dose as these already reveal differences between the different NMs while still keeping cytotoxicity at an acceptable level. Uniform surface values were not measured but instead calculated by normalizing to 1 m^2/ml BET surface using the formula

$$\text{value for } 1\frac{m^2}{ml} = \frac{\text{obtained value}}{\text{applied concentration}\left[\frac{g}{ml}\right] * \text{BET}\left[\frac{m^2}{g}\right]}$$

with obtained values being those from the medium applied dose.

2.6. Toxicity data and molecular descriptors

Most of the materials used here have already been well characterized with respect to their toxicity with results from *in vitro* macrophage assay and *in vivo* short-term inhalation studies (STIS) being available (Wiemann et al., 2016; Landsiedel et al., 2014; Wiemann et al., 2018). *In vitro* toxicity data were obtained using published and well-established standard operating procedures (SOPs). *In vivo* data were obtained under the conditions of good laboratory practice (GLP). The results of these studies were compared against those of the four reactivity assays performed here.

In vitro categories were obtained from Wiemann et al. (2016). Here, four different assays are performed with the supernatants from particle-exposed NR8383 rat alveolar macrophages, namely LDH, ROS, TNF-α and glucuronidase. Different doses of NMs between 22.5 and 180 μg/ml are tested. NMs are considered as being active if for at least two assays the LOAEC (Lowest Observable Adverse Effect Concentration) is reached at a particle surface concentration < 6000 mm^2/ml, and otherwise as passive. As explained in Wiemann et al. (2016) the threshold of 6000 mm^2/ml was derived from 4000 μm^2/NR8383 cell, which would approximately correspond to 0.04–0.08 m^2 per rat lung or 0.04–0.08 m^2/g lung tissue, and which is in line with previous estimations of lung overload for TiO_2 and $BaSO_4$ NMs in rats. A total of 28 NMs under consideration in the current study have been studied in the macrophage assay; 14 of them were assessed in all surface reactivity assays and are used for comparisons in the regression models.

In addition, 15 of the NMs under study were tested in *in vivo* STIS (Arts et al., 2016; Landsiedel et al., 2014; Keller et al., 2014; Ma-Hock et al., 2009). Rats were exposed to aerosols of different NM concentrations ranging from 0.5 to 50 mg/m^3 for 6 h/day on five consecutive days. Rats were sacrificed between 3 and 21 days after the end of exposure. Blood samples, bronchoalveolar lavage fluid (BALF) and histopathological sections from different parts of the respiratory tract were analyzed. Based on the combination of all these analyses, a common NOAEC (No Observable Adverse Effect Concentration) was assigned to each NM. This NOAEC represents the NM concentration at which no effects were observed in any of the analyses. In accordance with the categorization used in Wiemann et al. (2016), NMs were considered as active if the NOAEC is < 10 mg/m^3, otherwise they are considered as passive.

In a last step, reactivity data was integrated with the calculated descriptor LUMO (Lowest Unoccupied Molecular Orbital). The energy level of the LUMO describes the electron affinity of the NM and is related to its surface redox activity. Thus, a high correlation between the energy of the LUMO and the surface reactivity, especially the ESR results, which directly measure ROS formation of the NMs, may be expected. Data on the LUMO energy is taken from Gajewicz et al. (2018).

2.7. Deposited dose

A key factor for comparability of results for different NMs in cellular assays with adhesive cells is the effective dose. NMs sediment to the bottom of the cell culture dishes over time. However, the speed of this sedimentation differs between NMs depending on their physicochemical properties, mainly density and state of agglomeration. The amount of NMs actually deposited at the time of performing the assay is critical for evaluating the toxicity of the NM. In this study, the deposited dose was directly measured by analytical ultracentrifugation. Measurements were performed after 6 h of incubation at a concentration of 0.5 mg/ml.

2.8. Analysis and statistics

First, we addressed the question whether to compare NM responses based on same mass-based doses or based on same BET surface area doses. Therefore, results of assays for which multiple concentrations of NMs were assessed, were converted into a single value. For uniform mass FRAS data, four-parameter log-logistic dose-response models were fitted and the area under the curve (AUC) was computed. This value was used in subsequent analyses. For protein carbonylation, three different concentrations were used. We decided to use the medium concentration of 25 μg/ml in downstream analyses as for the highest concentration some materials show cytotoxic effects already while for the lowest concentration only very slight differences in the carbonylation is observable across the NMs. For all other assays and metrics, only single doses were tested.

We then split the NMs into active and passive materials based on their categorization with respect to the macrophage assay as well as STIS results. The distributions of values within these two groups were compared to each other for each assay using both dose metrics. For the protein carbonylation assay, the values were normalized to the deposited dose in addition.

Assays were then compared with regards to their similarity across the tested NMs. Heatmaps were drawn for scaled data and hierarchical clustering was performed using a complete linkage setting on Euclidean distances. Concordance in rankings of NMs between different assays was assessed using Spearman correlation. We chose Spearman correlation as the considered data was skewed and we do not necessarily expect a linear relationship between assays. In an additional step, we also compared NMs within well-studied material classes to get more detailed insights into rankings and assay comparisons.

In addition, univariate and multivariate logistic regression models were fit using the R packages glm and brglm. The dependent variable was comprised of *in vitro* or *in vivo* categories as described above. Each NM was assigned a label ‘active’ or ‘passive’. The active class was chosen as the positive class. Dependent variables were used in a surface-based manner with normalization to the deposited dose in case of the carbonylation assay. First, all univariate regression models were fit and their performance was evaluated using the Akaike Information Criterion (AIC) where smaller values indicate a better model fit. In order to account for small sample biases and the problem of complete separation, Firth's bias reduction was performed. The regression models were then used to predict the classes of each NM and the predictive performance of the models was assessed in a leave-one-out cross-validation (LOOCV).

In a last step, we also compared the results obtained by ESR to computed molecular descriptors (LUMO) from the literature to assess whether this might be a useful theoretical descriptor for NM grouping describing the reactivity of NMs.

All computations and visualizations were performed with R and Bioconductor. Codes are available on github under AileenBahl/logRegReactivity.

3. Results

3.1. Reactivity data

Within this study, we investigated the OP of various different materials using four different assays. We assessed how useful these assays are for the purpose of NM ranking and categorization. To that end, we included 35 well-characterized materials of twelve different chemical classes, comprising of nano- and non-nano materials (see Table 1). The OP was assessed using electron spin resonance (ESR) spectroscopy using CPH spin probe and DMPO spin trap, ferric reduction ability of serum (FRAS) assay and a cell-based protein carbonylation assay. NMs were applied with same mass doses (all assays) as well as same surface doses (only acellular assays). For the cellular carbonylation assay, obtained values were normalized to the deposited dose as measured by analytical ultracentrifugation in addition. It is worth noting, that Pluronic was also tested in the carbonylation assay to exclude any influence on the assay (data not shown). The values across all assays and metrics are shown in Table 2a (mass-based) and Table 2b (surface-based) and in more detail in Supplementary Table 1a and 1b. Data availability is summarized in Supplementary Table 2.

3.2. Correlation of surface reactivity assay results with available toxicity data

First, we addressed the question of the most suitable metric to obtain useful descriptors of the OP for NM categorization. To that end we compared data from same mass doses with those from same surface doses. The NMs were categorized into active and passive materials (Table 3) using previous categorizations based on the macrophage assay (Wiemann et al., 2016) and (if available) STIS (Landsiedel et al., 2014).

Wiemann et al. (2016) assessed *in vitro* toxicity of most NMs used in this study in the rat alveolar macrophage lung cell line NR8383. The macrophage assay combines four different endpoints: LDH, ROS by H_2O_2 production, TNF-α and glucuronidase. If at least two of these assays give positive results, the NM is considered to be active. The outcome of the macrophage assay across all NMs is shown in Table 3. Next, for each OP assay, the distribution of the values within the active and the passive group of materials as assessed by the macrophage assay is depicted in the boxplots in Fig. 1. Regardless of the assay and metric, the NMs that are active in the macrophage assay, show a tendency for elevated mean values compared to those in the passive group. In ESR with CPH and DMPO, NMs with low and high reactivity appear in both, the active and the passive group. However, especially for the CPH probe, the mean values of both groups are very far apart and thus a clear tendency for higher values in the active group is visible. In the

Table 2a
Mass-based values obtained from the four OP assays.

Class	Material	ESR (CPH) [Ratio sample/blank at 0.5 mg/ml]	ESR (DMPO) [Ratio sample/blank at 0.25 mg/ml]	FRAS [AUC]	Carbonyls [Ratio sample/control at 25 μg/ml]	Carbonyls [Ratio sample/control at 25 μg/ml] (normalized)
Silica	SiO_2_15_unmod	0.82	0.97	890,786	1.73	86.25
	SiO_2_15_Amino	0.92	0.97	1,300,879	1.38	3.06
	SiO_2_15_Phospho	1.21	0.83	1,288,551	0.58	3.19
	SiO_2 NM-200				0.97	
	SiO_2 NM-203			718,407	1.05	2.92
	SiO_2Levasil 50					
	SiO_2Levasil 100					
	SiO_2Levasil 300					
	SiO_2Aerosil 200	0.92	0.84			
Alumino-silicates	Kaolin	1.60	0.99	503,136	1.65	3.06
	Bentonite			2,915,633	1.36	7.17
	DQ12 (Quartz)			93,404		
Iron oxides	Fe_2O_3nanoform A	0.51	0.75	3,220,889	4.29	4.52
	Fe_2O_3nanoform B	0.82	1.13	150,508	1.8	2.54
	Fe_2O_3larger	13.86	4.33	357,083	2.51	2.79
Titania	TiO_2 NM-102	0.63	0.88	829,911		
	TiO_2 NM-105	0.69	1.01	677,867	1.48	3.1
	TiO_2 non-nano	0.94	0.91	257,420		
Ceria	CeO_2 NM-211	1.80	1.28	837,353		
	CeO_2 NM-212	1.42	2.07	570,057		
Copper-based	CuO	178.12	21.21	10,293,942	10.74	12.78
	Cu-Phthalocyanine non-halogenated	1.22	1.71	408,695	0.66	1.04
	Cu-Phthalocyanine halogenated	1.03	0.67	734,588	0.81	1.11
Zinc oxides	ZnO NM-110	1.51	2.20	2,898,432	2.53	4.76
	ZnO NM-111			1,069,431	1.89	
Diketo-pyrrolo-pyrrolPigments	DPP_premixed	0.88	0.77	−337,683	0.77	
	DPP nano	0.84	0.77	92,638	0.41	
	DPP non-nano	0.82	1.06	52,079	1.43	
Carbon-based	Carbon black	1.21	0.98	2,100,127	1.33	
	Graphene oxide				0.57	
	Graphene 1-layer	10.22	0.75	516,294		
	Graphene Multilayer	2.42	1.15	2,018,319		
Tungsten-based	WS_2	0.71	1.22	674,129		
Manganese-based	Mn_2O_3 (pos. control)	16.79	2.27	9,525,211	2.48	
Barium-based	$BaSO_4$NM-220 (neg. control)			6232	1.14	1.43

Table 2b
Surface-based values obtained from the four OP assays.

Class	Material	ESR (CPH) [Ratio sample/blank at 0.5 mg/ml]	ESR (DMPO) [Ratio sample/blank at 0.25 mg/ml]	FRAS [AUC]	Carbonyls [Ratio sample/control at 25 μg/ml]	Carbonyls [Ratio sample/control at 25 μg/ml] (normalized)
Silica	SiO_2_15_unmod	0.87	1.03	14.05	345	17,250
	SiO_2_15_Amino	0.92	1.07	18.95	275	611
	SiO_2_15_Phospho	0.91	1.21	8.01	115	639
	SiO_2 NM-200				204	
	SiO_2 NM-203			23.20	197	548
	SiO_2 Levasil 50			4.76		
	SiO_2 Levasil 100			13.23		
	SiO_2 Levasil 300			13.94		
	SiO_2 Aerosil 200			23.20		
Alumino-silicates	Kaolin	44.45	7.28	16.39	2754	5100
	Bentonite	6.86	1.28	86.99	1048	5516
	DQ12 (Quartz)					
Iron oxides	Fe_2O_3 nanoform A	15.21	0.87	44.20	1604	1688
	Fe_2O_3 nanoform B	20.19	1.04	15.33	2400	3380
	Fe_2O_3 larger	383.33	9.24	34.46	8376	9307
Titania	TiO_2 NM-102	1.29	1.79	6.30		
	TiO_2 NM-105	1.09	1.66	18.64	1942	2428
	TiO_2 non-nano	1.21	1.58	13.77		
Ceria	CeO_2 NM-211	268.60	1.37	14.16		
	CeO_2 NM-212	189.41	2.32	12.90		
Copper-Based	CuO	567.23	5.42	268.75	12,631	15,037
	Cu-Phthalocyanine non-halogenated	11.34	7.43	11.51	496	787
	Cu-Phthalocyanine Halogenated	185.32	1.44	18.15	468	641
Zinc oxides	ZnO NM-110	63.38	7.36	150.99	8417	15,881
	ZnO NM-111			20.31	5030	
Diketopyrrolo-pyrrolPigments	DPP_premixed	4.97	0.64	−23.11	1812	
	DPP nano	1.75	0.58	3.39	174	
	DPP non-nano	1.15	0.59	1.83	3563	
Carbon-based	Carbon black				942	
	Graphene oxide				104	
	Graphene 1-layer					
	Graphene multilayer					
Tungsten-based	WS_2					
Manganese-based	Mn_2O_3 (pos. control)	461.42	2.04	209.76	6080	
Barium-based	$BaSO_4$ NM-220 (neg. control)	2.30	1.31	6.65	56	65

FRAS assay, NMs in both boxes overlap at the lower end of the plot and thus, NMs with low surface reactivities appear in both, the active and the passive group. However, materials with high reactivity in the FRAS assay only appear in the active group. For the protein carbonylation assay, the overlap between the distributions is higher in the high activity range. NMs with very low reactivity values in the carbonylation assay only appear in the passive group. To take into account that only the fraction of NMs depositing at the bottom of the wells contributes to the cells exposure, we normalized the data to deposited doses calculated by using the distribution of sedimentation coefficients measured by analytical ultracentrifugation. A consequence of the normalization to the deposited dose is a more pronounced separation between the active and the passive group. Statistical differences between values for active and passive materials can be observed for ESR with DMPO in the mass-based analysis as well as for protein carbonylation after normalization for the surface-based analysis (Wilcoxon $p < 0.05$). However, in the case of the mass-based ESR with DMPO, most sample-to-blank ratios are close to 1 and thus these results may not be reliable. Overall, the use of a surface-based dose metric has as a consequence an increased separation between active and passive NMs.

In the same way, available *in vivo* data from STIS were compared to results of the current study (Fig. 2). NMs which are active in STIS, show overall higher mean values across the assays with the only exception being ESR with CPH. For the CPH and carbonylation assay, the picture is very similar to the *in vitro* situation seen in the macrophage assay above. Again, a consequence of the normalization to the deposited dose is a better separation between the groups. For the FRAS assay and the carbonylation assay, there is a clear separation between both groups with especially low values being only present in the passive group. For the ESR the picture is not as clear as in the macrophage assay. Here, no clear separation between both groups is possible. However, this may not only be an indication for the higher complexity of *in vivo* studies but results also from the fact that here the positive class is only comprised of three materials. Thus, generalization from this set is very difficult and results may not necessarily be robust for larger case studies.

Respective boxplots for only those cases with complete information across all assays are shown in Supplementary Fig. 2 for the macrophage assay and Supplementary Fig. 3 for STIS.

Overall, surface-based values led to a better separation of active and passive materials. In addition, normalizing the protein carbonylation data to the deposited dose also improved the separation of the NMs in the two categories. In the subsequent section, only those values will be considered in the main manuscript. All remaining analyses are shown in the supplementary files.

3.3. Correlations between reactivity assays

Next, we compared data across all test materials to assess the overall

Table 3
Toxicity categories obtained from the *in vitro* macrophage assay as well as from *in vivo* STIS.

Class	Material	Categorization *in vitro* (Wiemann et al., 2016; Wohlleben et al., 2019)	Categorization *in vivo* (Landsiedel et al., 2014; Wiemann et al., 2018)
Silica	SiO_2_15_unmod	active	active
	SiO_2_15_Amino	passive	passive
	SiO_2_15_Phospho	passive	passive
	SiO_2 NM-200	active	
	SiO_2 NM-203	active	active
	SiO_2 Levasil 50	active	
	SiO_2 Levasil 100	active	
	SiO_2 Levasil 300	passive	
	SiO_2 Aerosil 200		
Aluminosilicates	Kaolin	active	
	Bentonite	active	
	DQ12 (Quartz)		
Iron oxides	Fe_2O_3 nanoform A	passive	passive
	Fe_2O_3 nanoform B	active	
	Fe_2O_3 larger	passive	passive
Titania	TiO_2 NM-102	passive	
	TiO_2 NM-105	active	active
	TiO_2 non-nano	passive	
Ceria	CeO_2 NM-211	active	active
	CeO_2 NM-212	active	active
Copper-based	CuO	active	active
	Cu-Phthalocyanine non-halogenated	active	passive
	Cu-Phthalocyanine Halogenated	passive	
Zinc oxides	ZnO NM-110	active	
	ZnO NM-111	active	active
Diketopyrrolo-pyrrol Pigments	DPP_premixed	passive	
	DPP nano	passive	passive
	DPP non-nano	active	passive
Carbon-based	Carbon black		
	Graphene oxide		
	Graphene 1-layer		
	Graphene multilayer		
Tungsten-based	WS_2		
Manganese-based	Mn_2O_3 (pos. control)	active	
Barium-based	$BaSO_4$ NM-220 (neg. control)	passive	passive

correlation between the different assays. Spearman correlations between all pairs of assays over all tested materials show how well the ranks of the NMs in the corresponding assays correlate (Table 4). The highest correlation can be seen between the two ESR assays ($\rho = 0.552$). In addition, there are significant correlations at a significance level of 0.05 between ESR with CPH and FRAS ($\rho = 0.530$), ESR with DMPO and FRAS ($\rho = 0.529$) as well as FRAS and protein carbonylation ($\rho = 0.544$).

The heatmap in Fig. 3 gives more detailed insights into the comparability of results for the different materials. It also provides a ranking of all materials considering data of all assays while Supplementary Fig. 4 depicts a ranking of all materials using the data of the FRAS assay only.

It becomes obvious that highly reactive materials (e.g. CuO, Mn_2O_3, ZnO NM-110) tend to be reactive in most assays while non-reactive materials (e.g. TiO_2 NM-105, SiO_2_15_Amino, SiO_2_15_Phospho, $BaSO_4$ NM-220) similarly tend to be negative in all assays. In between there are materials with medium reactivity, for which the results of the different assays may differ more. CuO overall clearly is the material giving the highest response in ESR with CPH, FRAS and the carbonylation assay. Fe_2O_3 larger and ZnO NM-110 also respond very strongly in three of the four assays. However, the least responsive assay is not the same in both cases (FRAS for Fe_2O_3 larger and ESR with CPH for ZnO NM-110). Kaolin and CuPhthalo non-halogenated show high responses only in the ESR with DMPO, while values of the other assays are only in the medium to low range. Apart from the NMs at the lower end of the heatmap which show very low values in all assays, Kaolin and CuPhthalo non-halogenated is the pair of NMs with the most similar activation pattern.

3.4. Model performance of single and combined reactivity assays

In order to compare how well each of the reactivity assays explains the toxicity categorization and whether a combination of multiple assays improves the predictions, we fitted univariate and multivariate logistic regression models. The outcome for each NM was defined by a categorization into active and passive materials with respect to results from the macrophage assay or STIS as explained above. Due to biases because of small datasets and complete separation in some of the models, we decided to use Firth's bias reduction method (FIRTH, 1993). In addition, the fitted regression models were used to predict the class of each material after model fitting in a LOOCV approach. The results of the univariate and multivariate models are summarized in Table 5 and Supplementary Table 3.

In a first approach, we build the regression models on all materials with complete information available across all four assays. While the performance of the best model for the macrophage assay was good (prediction accuracy of 0.81), we observed that one of the misclassified materials was CuO. As CuO shows a very high reactivity in all assays and moreover is active *in vitro* and *in vivo*, a good prediction model

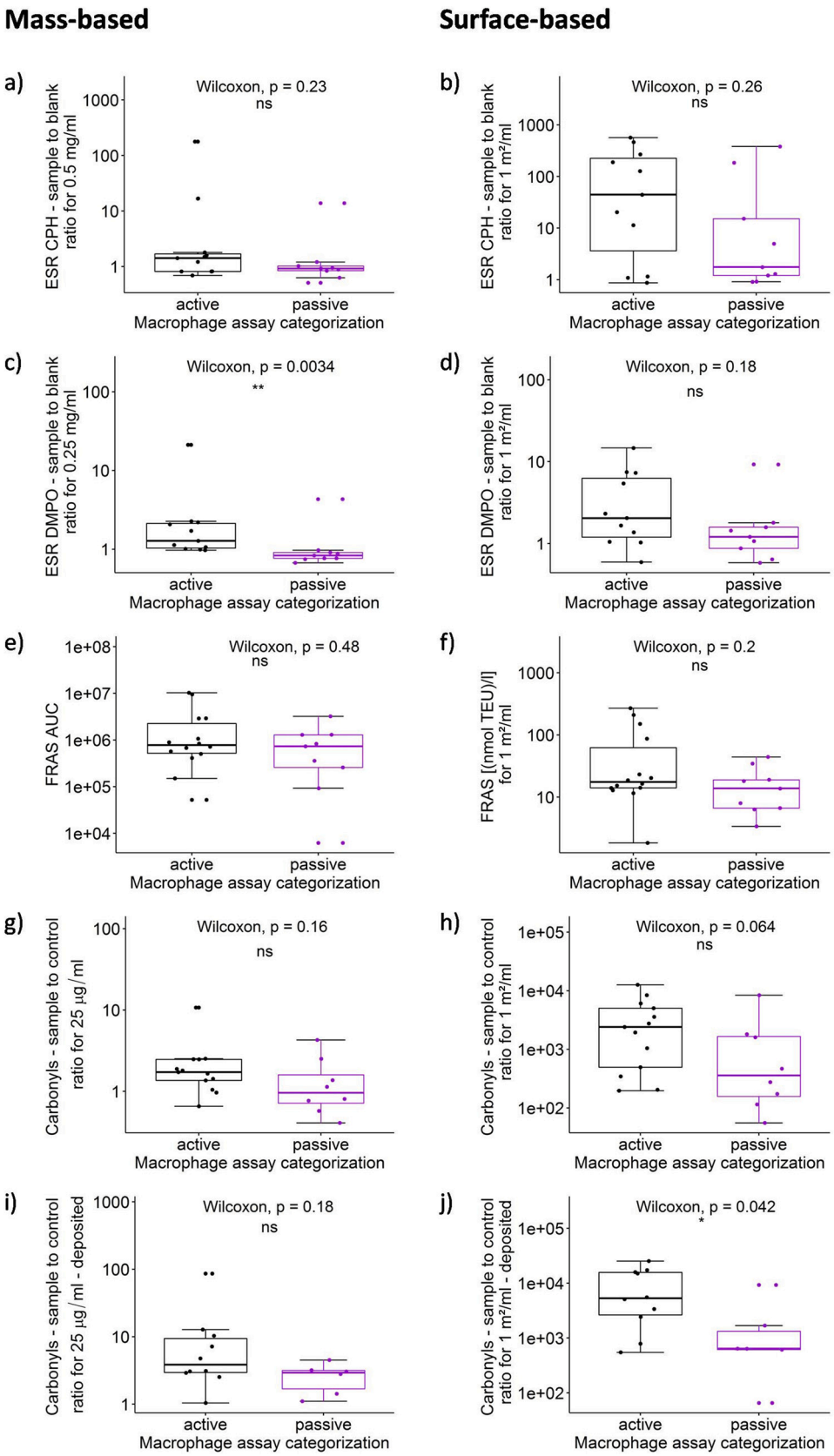


Fig. 1. Boxplots comparing mass- and surface-based values from each of the four assays comparing active versus passive NMs as assessed in the macrophage assay. Depicted are values from ESR with CPH probe with same mass (a) and same BET surface area doses (b), ESR with DMPO trap with same mass (c) and same BET surface area doses (d), FRAS assay with same mass (e) and same BET surface area doses (f) and carbonylation assay with same mass (g) and same BET surface area doses (h). In addition, carbonyl values were normalized to the deposited dose for same mass (i) and same BET surface area doses (j).

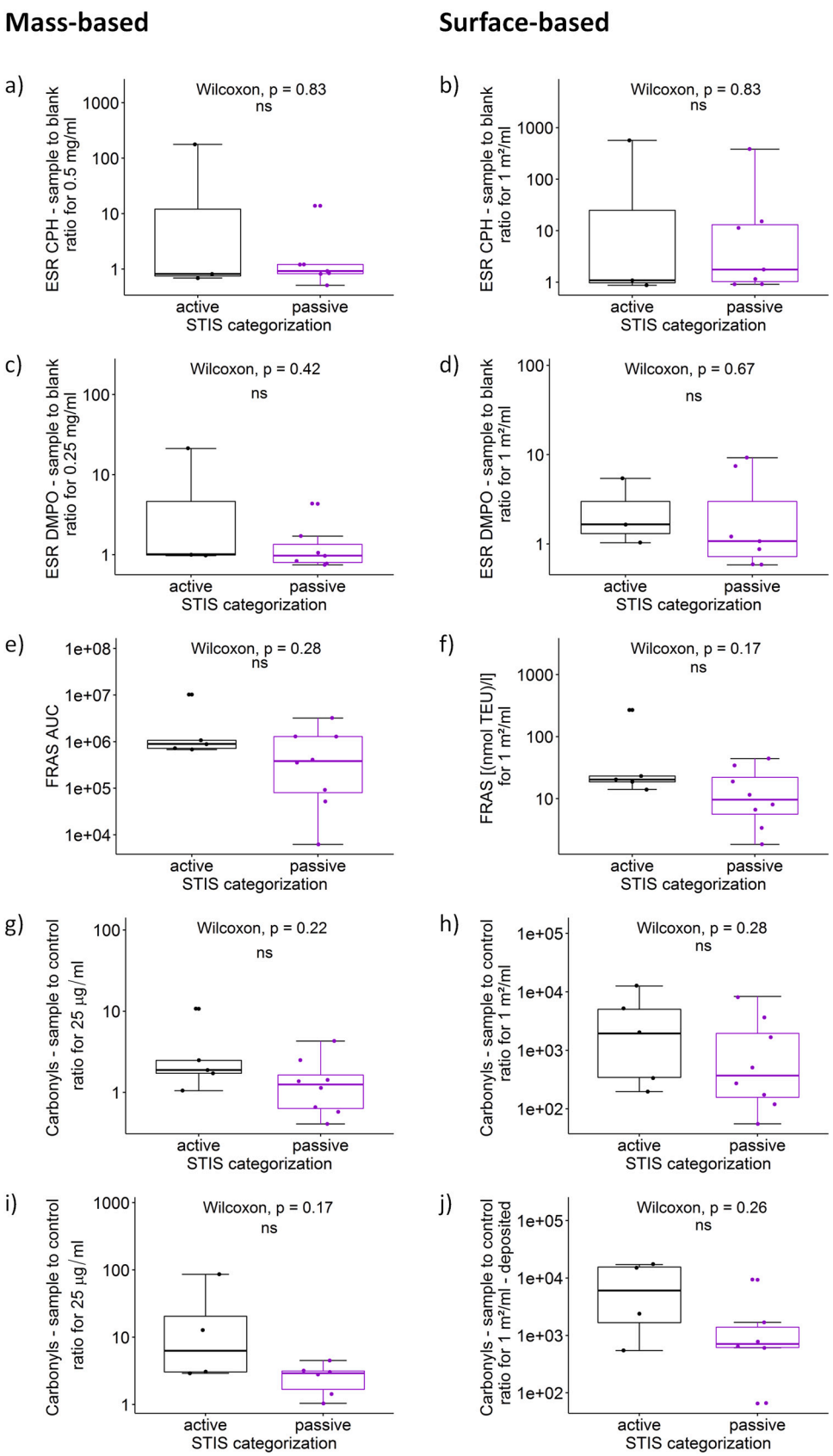


Fig. 2. Boxplots comparing mass- and surface-based values from each of the four assays comparing active versus passive NMs as assessed in STIS. Depicted are values from ESR with CPH probe with same mass (a) and same BET surface area doses (b), ESR with DMPO trap with same mass (c) and same BET surface area doses (d), FRAS assay with same mass (e) and same BET surface area doses (f) and carbonylation assay with same mass (g) and same BET surface area doses (h). In addition, carbonyl values were normalized to the deposited dose for same mass doses (i) and same BET surface area doses (j).

Table 4
Spearman correlation to assess comparability between the different assays.

	ESR CPH	ESR DMPO	FRAS	Carbonylation
ESR CPH	1	0.552* (n = 23)	0.530* (n = 22)	0.475 (n = 15)
ESR DMPO		1	0.529* (n = 22)	0.357 (n = 15)
FRAS			1	0.544* (n = 16)
Carbonylation				1

* $p < 0.05$.

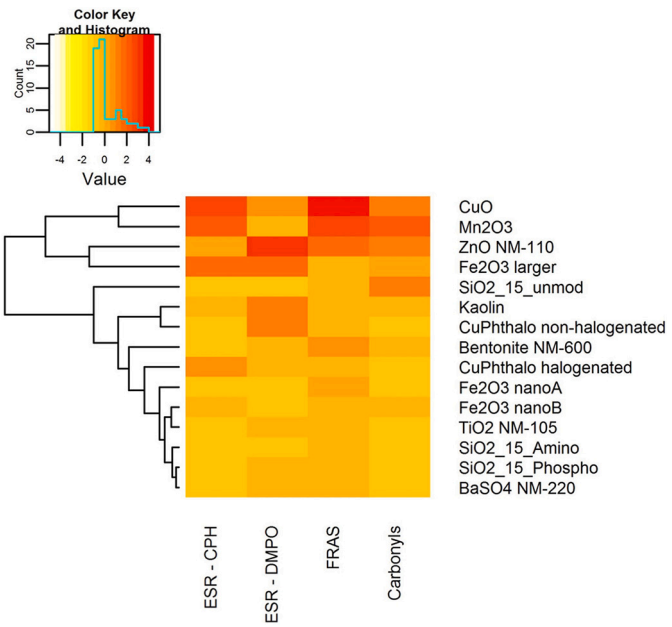


Fig. 3. Heatmap of scaled outcomes from all four assays. Hierarchical clustering is performed based on Euclidean distance.

should definitely classify this material correctly. We thus investigated this misclassification in more detail and observed it mainly results from the inclusion of Fe_2O_3 larger in the regression model. Fe_2O_3 larger shows relatively high reactivity in the ESR measurements while being passive *in vitro* and *in vivo*. The behavior of Fe_2O_3 larger deserves a deeper investigation. An interesting observation is that Fe_2O_3 larger shows a very unusual morphology of a spiked, irregular surface in TEM scans, which clearly distinguishes it from all other materials, which are closer to spheroidal geometry. In our study we decided to exclude this material from the final analysis as in general a false negative classification is considered a severe drawback of a model and we expected a good prediction model to correctly classify materials like CuO.

For the macrophage assay, the best models reached a prediction accuracy of 0.89 (see Table 5). This accuracy was achieved when ESR with the DMPO probe or FRAS assay results were combined with carbonylation assay data or in the case when all assays were included. Due to the lower model complexity, the first two models obtain a lower AIC (14.68 ± 0.82 and 13.22 ± 3.16, respectively, compared to 18.77 ± 0.76 for the complete set of assays) and are thus preferable. In the model with all descriptors as well as in the model including ESR with DMPO and carbonylation, the misclassified materials are Fe_2O_3 nanoformB and TiO_2 NM-105 which are both classified as passive while being active *in vitro* and in the case of TiO_2 NM-105 also *in vivo* (see Fig. 4 and Supplementary Fig. 5a). In the model including FRAS assay and carbonylation (see Supplementary Fig. 5b), Fe_2O_3 nanoformB and CuPhthalo non-halogenated are misclassified as passive. The best model based on a single assay is the one based on the carbonylation data with an AIC of 12.35 ± 2.28 and a balanced accuracy of 0.84. In this model, CuPhthalo non-halogenated (passive instead of active) and Fe_2O_3 nanoformA (active instead of passive) are misclassified (see Supplementary Fig. 5c).

For the STIS categorization, the same models as in the macrophage assay except the one including all surface reactivity assays show the best performance. Here, the model with only the carbonylation assay, even performs exactly as good (prediction accuracy of 0.83) as the other two which include in addition ESR with DMPO or FRAS results, respectively. All three models misclassify only TiO_2 NM-105 (passive instead of active) (see Supplementary Fig. 6).

Thus, the best testing strategy in our case study is to include ESR with DMPO or FRAS together with the carbonylation assay (based on outcome of the *in vitro* macrophage assay) or using the carbonylation assay alone (based on the outcome of the *in vivo* STIS results).

In the next part, we want to investigate the different material classes in more detail with the goal of ranking the individual materials within the classes.

3.5. Case studies

3.5.1. Silica case study

We considered nine variants of silica NMs and three aluminosilicates. Results from all assays were present for three silica particles and one aluminosilicate (see Supplementary Tables 1a, 1b, 2).

In the first step, we used the assays with the same mass concentrations (referred to as uniform mass data). For the FRAS assay, five different concentrations of SiO_2_15_unmod, SiO_2_15_Amino, SiO_2_15_Phospho, SiO_2NM-203, Bentonite, Kaolin and Quartz NMs were tested (see Fig. 5). We performed dose-response modeling for better comparability between NMs. The results are shown in Fig. 5. The results of the supernatants are shown in Supplementary Fig. 7. As expected, the dose-response curves of all silica particles lie between the one of the negative control ($BaSO_4$ NM-220) and the positive control (Mn_2O_3). For all NMs except quartz, a clear dose-dependent response is observed. The micron-sized quartz gives a low intensity, concentration independent response. Compared to all other silica particles and the aluminosilicates, quartz has the lowest OP. Kaolin shows roughly the same OP as SiO_2 NM-203 at the highest tested dose. However, while the dose-response curve already reached a saturation point for NM-203, it is expected to further increase for Kaolin if higher concentrations were tested. This would result in Kaolin being ranked higher than SiO_2 NM-203 at higher concentrations. This is actually a general phenomenon that can be observed among all particles: While the dose-response curves for some materials are very steep at low doses and then level off very fast (e.g. SiO_2 NM-203 or Mn_2O_3), others continue growing until the highest dose that was applied here but at lower rates (e.g. Kaolin or Bentonite). Within the group of silica particles, SiO_2_15_Amino and SiO_2_15_Phospho behave very similar over the complete dose range and cause more oxidative damage in FRAS assay than SiO_2_15_unmod, which in turn shows a higher response compared to SiO_2 NM-203. Bentonite gives the highest response of all silica-based materials. Using the outcome obtained for the highest tested dose, the NMs can be ranked from lowest to highest response in FRAS assay: $BaSO_4$ NM-220 < quartz < SiO_2 NM-203 ≈ Kaolin < SiO_2_15_untreated < SiO_2_15_Amino ≈ SiO_2_15_Phospho < Bentonite < Mn_2O_3.

In order to compare the results obtained from different assays we ranked all materials from lowest to highest responses for each assay (Supplementary Fig. 8). As expected, the positive control Mn_2O_3 is ranked highest across all assays but the carbonylation assay where it only obtains the second highest rank after SiO_2_15_unmod. The negative control $BaSO_4$ has the lowest rank in the FRAS assays and the carbonylation assay. In the mass-based ESR assay, no measurements for $BaSO_4$ were performed. Bentonite NM-600, which shows the highest response of all silica-based materials in the FRAS assay, shows only medium response in the carbonylation assay. While SiO_2_15_Phospho,

Table 5
Logistic regression models are fit for each combination of predictors. Sensitivity, specificity as well as the balanced accuracy of each assay alone as well as all possible combinations of them are assessed in a Leave-One-Out Cross-Validation using Firth' regression. The best models, typically achieved by combining several assays, are highlighted in green. The best models relying on one assay only are highlighted in yellow. The ticks in columns 2 to 5 indicate which predictors were used in each logistic regression model. The AIC relates to the goodness-of-fit of the regression model relative to its complexity. Smaller values of the AIC indicate higher suitability of the model. The sensitivity gives the proportion of active materials correctly predicted as active while the specificity relates to correct predictions of passive materials. The balanced accuracy is the trade-off between sensitivity and specificity.

Outcome variable	Predictors				AIC (mean ± sd) from LOOCV	Balanced prediction accuracy	Sensitivity	Specificity
	CPH	DMPO	FRAS	Carbonyl				
Macro-phage assay	✓				20.02 ± 0.65	0.5	1	0
		✓			16.68 ± 0.88	0.43	0.67	0.2
			✓		18.49 ± 0.66	0.22	0.44	0
				✓	12.35 ± 2.28	0.84	0.89	0.8
	✓	✓			18.98 ± 0.81	0.38	0.56	0.2
	✓		✓		20.12 ± 0.76	0.28	0.56	0
	✓			✓	13.69 ± 1.79	0.79	0.78	0.8
		✓	✓		18.20 ± 0.94	0.73	0.67	0.8
		✓		✓	14.68 ± 0.82	0.89	0.78	1
			✓	✓	13.22 ± 3.16	0.89	0.78	1
	✓	✓	✓		19.69 ± 1.05	0.48	0.56	0.4
	✓	✓		✓	16.32 ± 0.66	0.83	0.67	1
	✓		✓	✓	18.30 ± 2.01	0.52	0.44	0.6
		✓	✓	✓	17.63 ± 0.79	0.83	0.67	1
	✓	✓	✓	✓	18.77 ± 0.76	0.89	0.78	1
STIS	✓				11.62 ± 0.99	0.5	0	1
		✓			12.86 ± 0.77	0.4	0	0.8
			✓		11.76 ± 1.02	0.5	0	1
				✓	8.78 ± 1.28	0.83	0.67	1
	✓	✓			13.46 ± 0.79	0.4	0	0.8
	✓		✓		13.38 ± 1.01	0.4	0	0.8
	✓			✓	10.78 ± 1.59	0.67	0.33	1
		✓	✓		13.61 ± 0.83	0.4	0	0.8
		✓		✓	11.09 ± 1.19	0.83	0.67	1
			✓	✓	10.80 ± 1.64	0.83	0.67	1
	✓	✓	✓		15.13 ± 0.72	0.2	0	0.4
	✓	✓		✓	12.69 ± 1.57	0.57	0.33	0.8
	✓		✓	✓	12.78 ± 1.41	0.57	0.33	0.8
		✓	✓	✓	12.92 ± 1.45	0.73	0.67	0.8
	✓	✓	✓	✓	14.64 ± 1.32	0.47	0.33	0.6

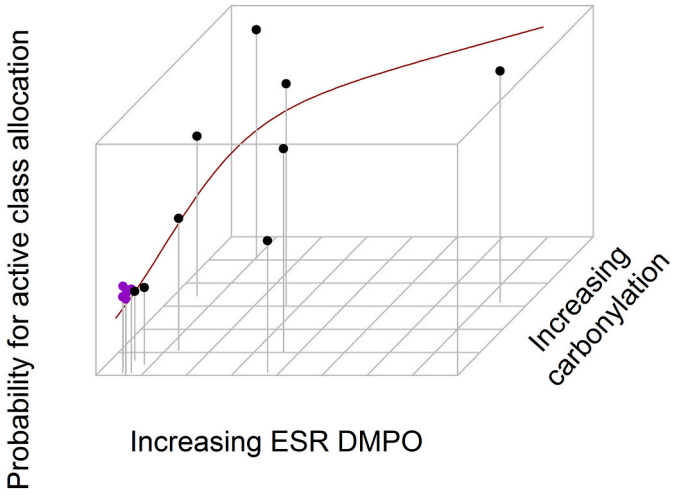


Fig. 4. Scatterplot showing the result of the logistic regression model including ESR with DMPO and carbonylation data based on the data holding macrophage assay labels. On the y-axis, the probability of each NM to belong to the active class as given by the regression model is depicted. NMs at the lower end of this axis are predicted to be passive, the ones at the top are assigned to the active class. Purple dots represent NMs that were assigned to the passive class by the macrophage assay, black dots represent active NMs with respect to the macrophage assay. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

SiO_2_15_Amino and SiO_2_15_unmod show very similar response in FRAS and ESR, the protein carbonylation shows a much higher response for SiO_2_15_unmod. This difference in response in the protein carbonylation is in-line with the fact that SiO_2_15_unmod is classified as active *in vitro* an *in vivo*, while the other two are passive. In the mass-based assays, Kaolin shows relatively low reactivity even though it is active *in vitro*.

In accordance with the mass-based data, we also performed the ranking of BET surface area-based measurements of the silica-based materials. The results are shown in Fig. 6. The positive control Mn_2O_3 still has the highest rank in all assays but ESR with DMPO, where it only gives the second highest response value. The negative control $BaSO_4$ shows the lowest response in the carbonylation assay as well as the second lowest rank in the FRAS assay. In the ESR assay, $BaSO_4$ ranks in the middle of all materials. However, the sample to blank ratio is still relatively close to 1 and thus $BaSO_4$ is also not very reactive in ESR. In the surface-based carbonylation assay (normalized to the deposited dose), we observe almost perfect separation of active and passive materials with the only exception being SiO_2 NM-203 which shows low carbonylation while being active *in vitro* and *in vivo*. While in the FRAS assay SiO_2 NM-203 is ranked higher, this assay is not separating SiO_2_15_unmod well from its two coated variants. Similarly, in both ESR assays, the aluminosilicates are ranked very high (corresponding to their categorization as active), while the three silica show very similar response even though SiO_2_15_unmod is active and the other two are not.

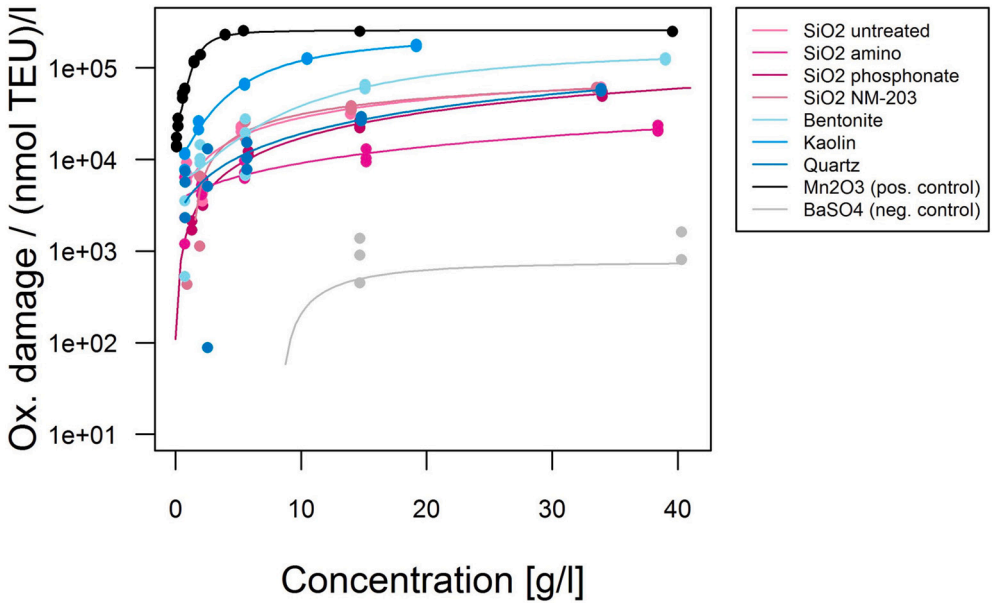


Fig. 5. Dose-response curves for different silicas obtained by the FRAS assay.

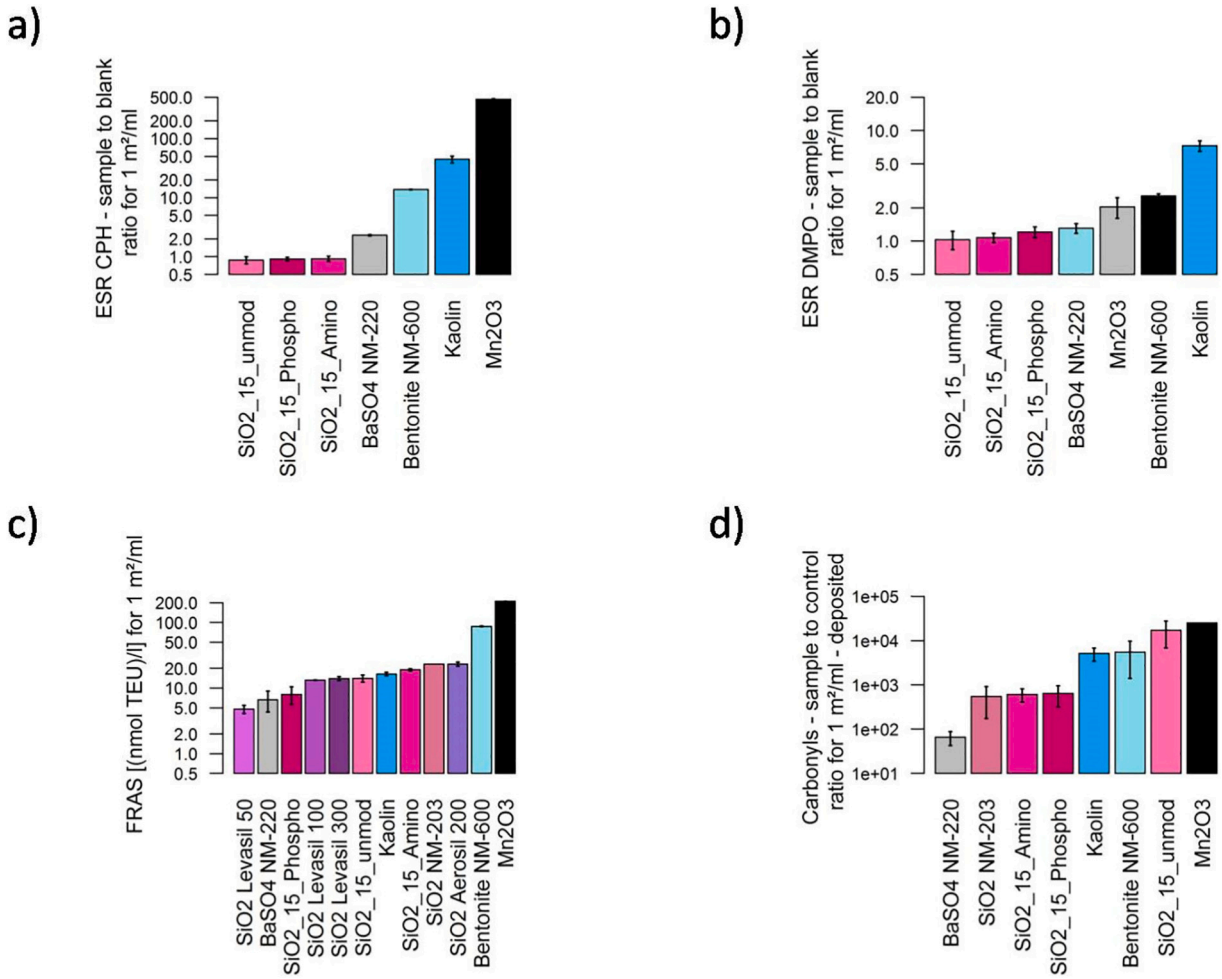


Fig. 6. Surface reactivity of silica particles measured (a, b, c) or calculated (d) according to different assays and uniform BET surface area. NMs within each plot are ranked with particle showing lowest reactivity on the left end to particles with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO trap, c) FRAS and d) protein carbonylation assay results are depicted.

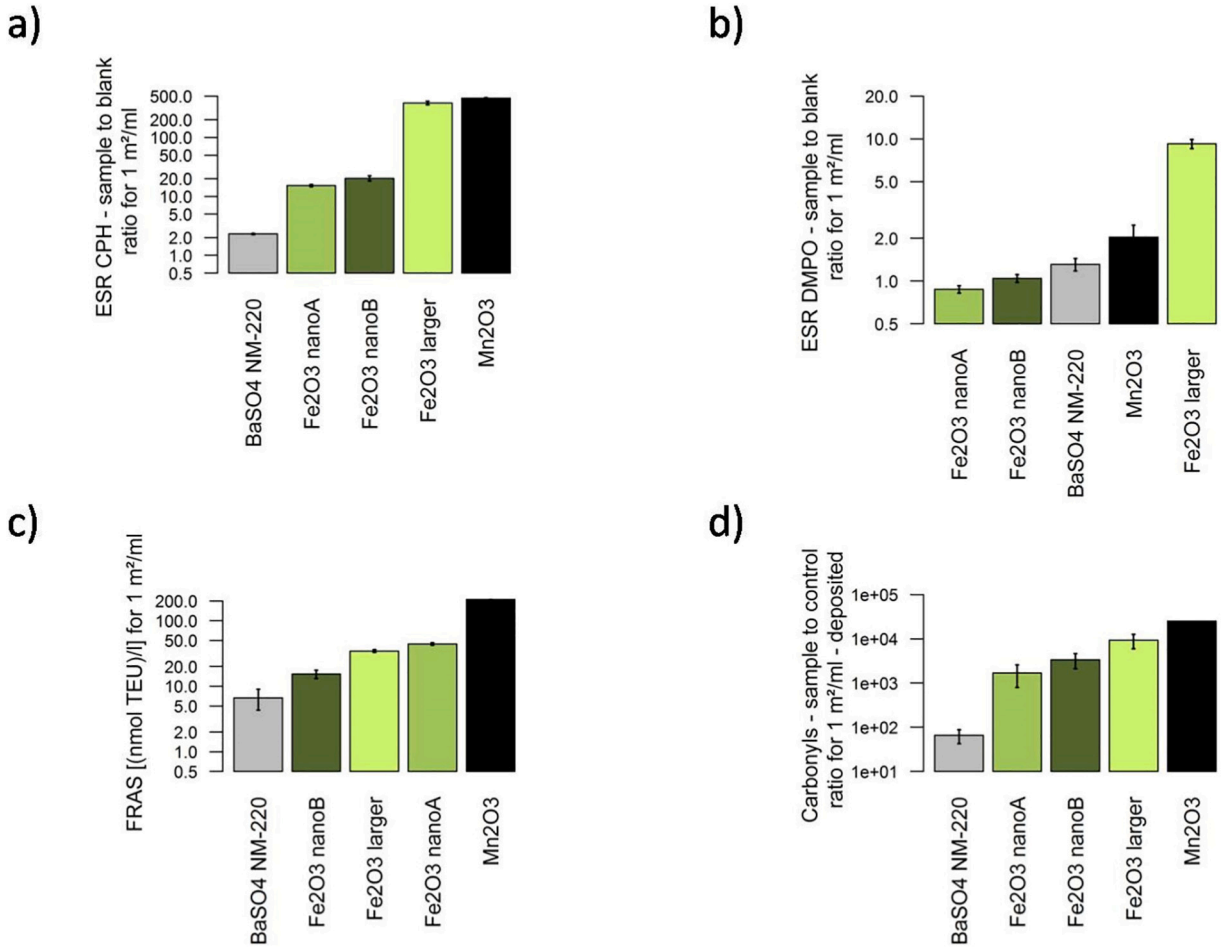


Fig. 7. Surface reactivity of iron oxide particles measured (a, b, c) and calculated (d) according to different assays and uniform BET surface area. NMs in the plot are sorted by ranks, from the particle with lowest reactivity on the left end to the particle with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO probe, c) FRAS and d) protein carbonylation assay results are depicted.

Overall, these results suggest that within the silica-based materials the surface-based measurements also perform slightly better than mass-based assays as well as that even here a combination of the different assays might be advantageous.

3.5.2. Iron oxide case study

In the case of the surface-based measurements for iron oxide, ESR with CPH and DMPO as well as the carbonylation assay both rank the particles in the same way, namely Fe_2O_3 nanoformA < Fe_2O_3 nanoformB < Fe_2O_3 larger (Fig. 7). The difference in response between the two smaller nanoforms is much smaller than the one between the two small nanoforms and the larger one.

In the FRAS assay, Fe_2O_3 nanoformB shows less response compared to Fe_2O_3 larger and Fe_2O_3 nanoformA (see also Supplementary Fig. 9c). The carbonylation assay ranks Fe_2O_3 nanoformB higher than nanoformA and Fe_2O_3 larger obtains the highest rank again. In all assays, Fe_2O_3 larger is ranked higher that the two smaller nanoforms, in case of ESR with DMPO even higher than Mn_2O_3. Comparing the two smaller nanoforms among each other, the carbonylation assay as well as both ESR assays rank Fe_2O_3 nanoformB which is active higher that the passive Fe_2O_3 nanoformA (even though for ESR with DMPO both values are close to 1). While for ESR this ordering is preserved in the mass-based measurements (see Supplementary Fig. 9), for the carbonylation assay it is not.

3.5.3. Zinc oxide

For zinc oxide all assays show elevated signals compared to $BaSO_4$ (see Supplementary Figs. 10 and 11). This matches the fact that ZnO NM-110 and ZnO NM-111 are both classified active. Especially for ESR, this effect is even more pronounced in the surface-based measurements. Only in the FRAS assay, both zinc oxides are tested and there ZnO NM-110 shows slightly higher response that ZnO NM-111.

3.6. The potential of the theoretical descriptor 'LUMO' in the OP detection

Another concept frequently used in NM grouping and QSAR approaches is the calculation of theoretical descriptors. It has been shown by Burello and Worth (2011) and by Zhang et al. (2012) that it is possible to predict the electron transfer from/to NMs to/from biological environment by comparing the bottom of the conduction band (E_c) with the cellular redox potential (from −4.12 to −4.84 eV). An electron transfer from biological systems to NMs can only take place when the cellular redox potential is higher than E_c. If this is not the case, it is always possible that an electron transfer from an aqueous donor to the NM and from this to the biological system takes place until reaching a steady state. Gajewicz et al. (2018) applied such an approach within a grouping framework for NMs. Considering that E_c for semiconducting NMs can be approximated with the energy of the lowest unoccupied molecular orbital (LUMO). Gajewicz et al. considered a threshold for the LUMO of −2.40 eV, NMs with LUMO at lower energy are considered active and all the others are considered passive.

The same approach could be used to interpret the release of radicals in aqueous environment, as for example in ESR experiments using different spin traps. Therefore, we compared the results obtained by

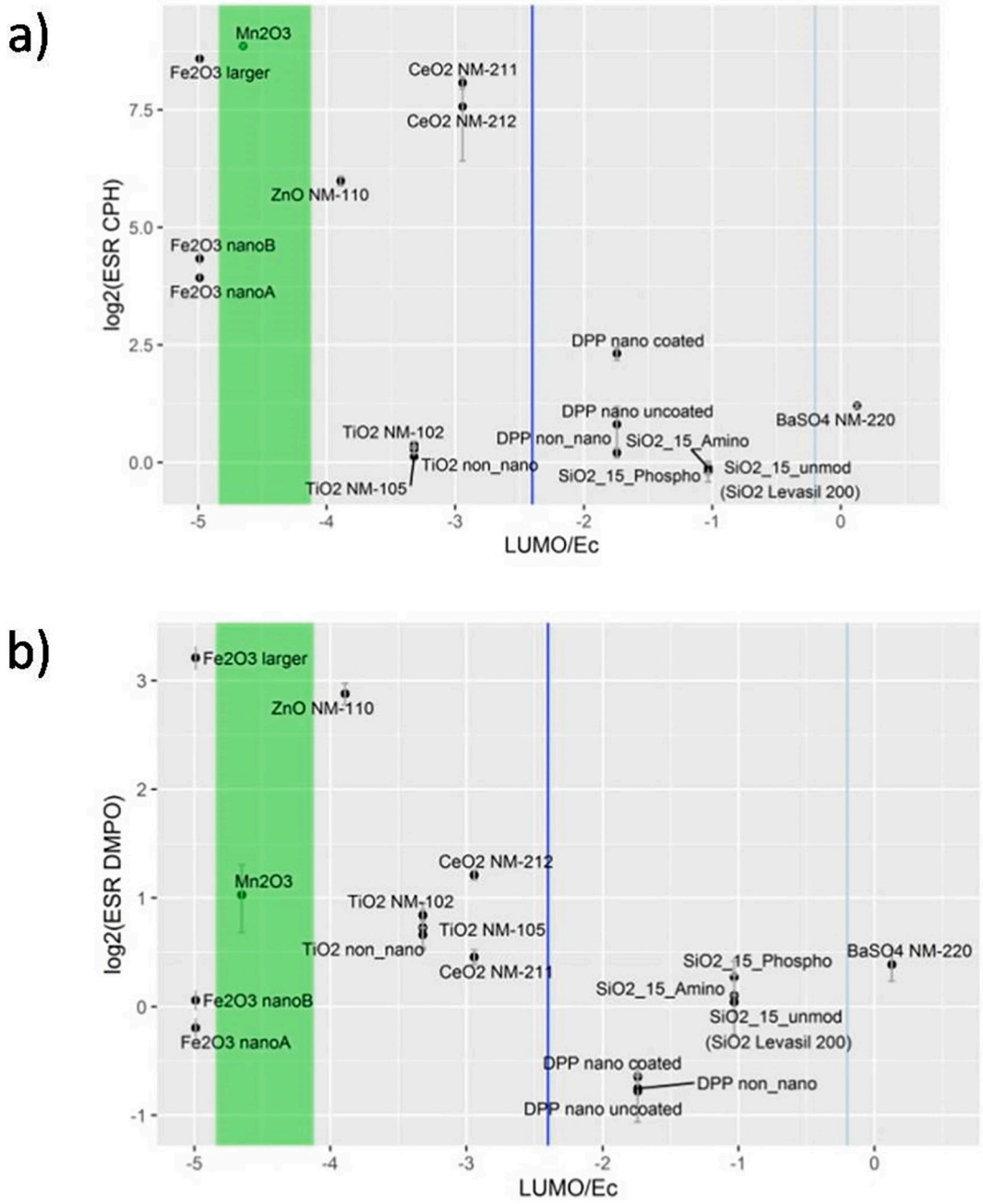


Fig. 8. Scatterplot comparing the outcomes of a) ESR with CPH probe and b) ESR with DMPO probe with the LUMO of the NMs. The area highlighted in green is the cellular redox potential (Zhang et al., 2012). The light blue vertical lines correspond to the redox potential of hydroxyl radicals on the right (Wang et al., 2017) and the redox potential of superoxide radicals on the left. The dark blue vertical line represents the LUMO values determined as a cut-off between active and passive NMs by Gajewicz et al. (2018). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ESR with CPH spin probe and DMPO spin trap with the LUMO or E_c obtained from literature (Gajewicz et al., 2018; Burello and Worth, 2011; Zhang et al., 2012; FIRTH, 1993) (see Fig. 8). A slight trend towards higher ESR values for materials with LUMO or E_c close to or within the region of the cellular redox potential can be observed, especially in the case of the CPH probe. However, there are some outliers visible like CeO_2 NM-211 and NM-212 which show high values in ESR with CPH while having a LUMO clearly larger than −4.12 eV or Fe_2O_3 nanoformA and nanoformB which have a LUMO close to the cellular redox potential but rather low response in ESR with DMPO.

4. Discussion

Previously, different assays have been used to measure the OP of NMs. All assays rely on different principles and thus, different techniques have different advantages and disadvantages. Their use depends on what exactly shall be measured, the generation of free radicals, which can be performed in acellular as well as in cell-based environments or different cell-based consequences. In addition, there are also technical reasons why different laboratories may favor specific techniques. For instance, ESR seems to be a very reliable assay but needs

special equipment, which is very expensive and thus available in only few laboratories.

In order to obtain reliable grouping concepts and a suitable testing strategy and to be able to apply statistical and computational methods that can aid this process, large amounts of data representing a great variety of existing NMs are necessary. For combining all the available data, we need to assess how comparable results from different assays are and how they can be integrated. Therefore, we need to address the questions of how well assays are suitable to rank materials within chemical families with respect to know toxicity data from previous studies as well as how well this works over all material classes.

Here, we compared the results of four different assays measuring the OP of NMs, namely ESR with the spin probe CPH as well as with the spin trap DMPO, the FRAS assay and a cell-based protein carbonylation assay. We assessed differences in the assay results if NMs were applied with the same mass dose or with the same surface dose and compared how well both metrics can explain results of the *in vitro* macrophage assay and *in vivo* STIS. Logistic regression was used to obtain the combination of assays that is most suitable to predict toxicity outcomes and to compare the results to those from analyses in which only one of the assays was used. Investigation of different NM families allowed for more detailed insights.

The comparison of assay results based on same mass doses and same surface doses, respectively, showed that measurements based on same surface doses are more suitable to distinguish between active and passive NMs as categorized by the macrophage assay. For STIS categorization, the advantages of the surface-based measurements were still visible overall although much less pronounced. It should be noted that the number of investigated materials *in vivo* was much lower. The better outcome for surface-based dose metrics is most probably related to the fact that ROS generation can only take place at the surface of the NMs and thus, surface area is the more important metric. Similar findings were also described in previous studies (Mottier et al., 2016; Schmid and Stoeger, 2016; Simkó et al., 2014). The normalization of the cell-based carbonylation assay results to the deposited dose was advantageous for both, the prediction of the *in vitro* macrophage assay and the *in vivo* STIS results. This is related to the fact that only NMs that reach the cells may cause protein carbonylation in them. Other studies also highlighted the importance of the deposited dose or even better internal dose for the prediction of toxicological responses (Simkó et al., 2014; Cohen et al., 2014; Schmid and Cassee, 2017).

Comparing the different assays for determining the OP of NMs showed that ESR with CPH and DMPO and FRAS assay had pairwise significant correlations ($p < 0.05$) among each other and also FRAS assay was significantly correlated with the carbonylation assay. However, correlation coefficients only lie between 0.5 and 0.6. This might be explained by the fact that all assays detect different types of ROS, which indicates that each assay may have a different specificity and a combination of various assays might improve the overall predictive performance. With respect to the specificity of the assays, CPH spin probes in ESR are most sensitive to singlet oxygen, superoxide radicals and peroxynitrite. In contrast, DMPO spin traps trap mainly hydroxyl and superoxide radicals. The FRAS assay measures antioxidant depletion in general which might be triggered by different ROS but also by reactive nitrogen species. In protein carbonylation, the result depends on the amount of different kinds of produced ROS but also on the compensating mechanisms reducing ROS in the cell.

In addition, the assays are performed in different environments with increasing systems complexity. ESR directly measures the ROS production in dH_2O. While dH_2O is a very simple environment which does not induce much interference with the measurement technique, it may not represent the physiological conditions in which the NM acts very well. Important properties like the formation of a protein corona are not accounted for by ESR. FRAS assay instead uses human serum, which is a more physiological environment. While this might be a more relevant scenario then measurements in dH_2O, FRAS is still an acellular assay and does not account for additional factors important for toxicity of NMs, e.g. different amounts of NM uptake or cell-based ROS defense activities. Cellular assays like protein carbonylation do consider those factors and thus represent a more relevant read-out from a toxicological point of view. However, biological variability may scatter the outcomes of cellular assays as the outcome will depend on many details such as selected cell model, passage number, cell seeding density, *etc.* This, together with the unknown internal dose makes the interpretation of the results difficult. Based on the described differences between the assays the observed differences in the ranking of materials are not unexpected or astonishing.

The idea that a combination of assays might be advantageous for predicting the toxicity category of the NM is also supported by the fact that in the heatmap, the combination of all assays leads to a clustering with most of the active materials being in the upper part of the heatmap for which at least one of the assays gives a strong response while most passive materials are at the bottom with very low responses across all assays.

While this combination of all assays in the heatmap already gives first hints that using multiple assays at the same time might be beneficial, we wanted to compare how well models based on single assays perform compared to multivariate models and which combination of assays gives best results. Therefore, we fitted logistic regression models with different assay combinations as well as single assays as the input. In the case of macrophage assay categories, the best model consisted of ESR with DMPO spin trap or FRAS together with the carbonylation assay. This is in-line with other studies which also suggested combinations of acellular and cellular assays to determine the OP of NMs previously (Riebeling et al., 2016). The best model based on one single assay is the one including only carbonylation data with a drop in accuracy from 89% to 84%. For STIS categories, the best model performance (83% accuracy) could be achieved even with the carbonylation assay alone. Given the fact that this assay is much less complicated and needs no special equipment, it seems to be a reasonable choice to use protein carbonyls to describe the OP of NMs.

Just a small set of NMs are subject to misclassification in the best models. TiO_2 NM-105 is misclassified in the optimal STIS models as well as in the macrophage assay model including all assays or ESR with DMPO and carbonylation assay. Some crystalline forms of titanium dioxide are photoactive, including the anatase-dominated NM tested here. As the OP was detected in dark which might not be the situation *in vitro* and especially *in vivo*, discrepancies in the categorization might result from this fact. In addition, CuPhthalo non-halogenated is misclassified in the macrophage assay models with FRAS and carbonylation as well as in the model with only the carbonylation data. However, CuPhthalo non-halogenated is known to be false-positive *in vitro* already (Arts et al., 2016; Wiemann et al., 2016) (see Pigment Blue in Table 3 in (Wiemann et al., 2016)). Thus, the misclassification detected here fits to the overall picture for this particular material. In the macrophage assay models also Fe_2O_3 nanoformA and Fe_2O_3 nanoformB are frequently misclassified, as they are passive in the inhalation test (Hofmann et al., 2016).

Comparing results of different assays within families of NMs, we saw that even within different families same-BET-surface-area-based measurements performed slightly better than mass-based assays. Again, the carbonylation assay was most useful to separate active and passive materials. However, in most cases other assays are needed for complete separation.

This study provided useful insights into a potential testing strategy for OP of NMs. Even though our case study containing 35 different materials was quite extended, overall only a limited set of materials was tested, which might impair generalizability of the results. Larger collections of materials are needed to rerun the approach provided here.

In addition, the same surface data for the carbonylation assay were only calculated in this approach. These results could be blurred by deviations in the agglomeration state, deposition, etc. Same surface

data should thus be obtained by directly measuring NMs at same surface doses, which is however more challenging in cell-based assays as the BET surfaces vary a lot between different materials. Thus, it may become very challenging to select a suitable uniform surface-area dose for cell-based assays.

5. Conclusion

Here, we compared four different assays measuring the oxidative potential (OP) of 35 nanomaterials from twelve different chemical classes. These assays, namely ESR (with CPH spin probe and DMPO spin trap), FRAS assay and a cell-based protein carbonylation assay have been used in previous studies and also in some NM grouping approaches. Overall, for very reactive and non-reactive NMs we saw good agreement among the assays. However, for NMs with intermediate reactivity, the consistency was lower. This is not surprising, considering that the assays measure different types of ROS, they are conducted in cellular or acellular environment, and reflect ROS production either directly or indirectly. A combination of assays, or the justified selection of one assay, may better predict *in vivo* outcomes. In our dataset we observed a good agreement with previous *in vitro* and *in vivo* findings as well. This was especially the case when NMs were applied at same surface doses and if, for the carbonylation assay, results were normalized to the deposited dose. Logistic regression revealed that a combination of ESR with DMPO or FRAS together with the carbonylation assay was most predictive for *in vitro* outcomes, whereas for *in vivo* outcomes the model based on carbonylation data only gave the highest predictivity, especially when surface dose is used as a dose metric. In any case, suitable benchmark materials with high (CuO, Mn_2O_3, ZnO NM-110) and low ($BaSO_4$ NM-220) OP were identified. In a next step, further validation of the approach should be performed by extending and completing the data set. This may also result in more specific recommendations as to whether single or combined assays are more appropriate to predict the toxicologically relevant OP of nanomaterials.

CRediT authorship contribution statement

Aileen Bahl:Formal analysis, Investigation, Writing - original draft.**Bryan Hellack**:Investigation, Formal analysis, Conceptualization.**Martin Wiemann**:Formal analysis.**Anna Giusti**:Investigation, Formal analysis.**Kai Werle**:Investigation, Formal analysis. **Andrea Haase**:Conceptualization, Writing - original draft.**Wendel Wohlleben**:Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge funding from the German BMBF (grant agreement number 03XP0002). The authors also thank all the institutions for support. In addition, Doreen Wittke is acknowledged for excellent technical support.

Abbreviations

AIC	Akaike information criterion
AUC	area under the curve
BOD	biological oxidative damage
CPH	1-hydroxy-3-carboxy-pyrrolidine
DMPO	5,5-dimethylpyrroline N-oxide
ESR	electron spin resonance
FRAS	ferric reduction ability of serum
HBS	human blood serum
LDH	lactate dehydrogenase
LOAEC	lowest observable adverse effect concentration
LOOCV	leave-one-outcross-validation
LUMO	lowest unoccupied molecular orbital
NM	nanomaterial
NOAEC	no observable adverse effect concentration
OP	oxidative potential
ROS	reactive oxygen species

STIS – short-term inhalation study
TEU – Trolox equivalent unit
TNF-α – tumor necrosis factor α

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.impact.2020.100234.

References

Arts, J.H.E., Hadi, M., Keene, A.M., Kreiling, R., Lyon, D., Maier, M., Michel, K., Petry, T., Sauer, U.G., Warheit, D., et al., 2014. A critical appraisal of existing concepts for the grouping of nanomaterials. Regul. Toxicol. Pharmacol. 70 (2), 492–506.
Arts, J.H.E., Hadi, M., Irfan, M.A., Keene, A.M., Kreiling, R., Lyon, D., Maier, M., Michel, K., Petry, T., Sauer, U.G., et al., 2015. A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). Regul. Toxicol. Pharmacol. 71 (2), S1–S27.
Arts, J.H.E., Irfan, M.A., Keene, A.M., Kreiling, R., Lyon, D., Maier, M., Michel, K., Neubauer, N., Petry, T., Sauer, U.G., et al., 2016. Case studies putting the decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping) into practice. Regul. Toxicol. Pharmacol. 76, 234–261.
Borm, P.J., Müller-Schulte, D., 2006. Nanoparticles in drug delivery and environmental exposure: same size, same risks? Nanomedicine 1 (2), 235–249.
Braakhuis, H.M., Park, M.V.D.Z., Gosens, I., De Jong, W.H., Cassee, F.R., 2014. Physicochemical characteristics of nanomaterials that affect pulmonary inflammation. Particle and fibre toxicology 11, 18.
Buesen, R., Landsiedel, R., Sauer, U.G., Wohlleben, W., Groeters, S., Strauss, V., Kamp, H., van Ravenzwaay, B., 2014. Effects of SiO2, ZrO2, and BaSO4 nanomaterials with or without surface functionalization upon 28-day oral exposure to rats. Arch. Toxicol. 88 (10), 1881–1906.
Burello, E., Worth, A.P., 2011. A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. Nanotoxicology 5 (2), 228–235.
Cohen, J.M., Teeguarden, J.G., Demokritou, P., 2014. An integrated approach for the in vitro dosimetry of engineered nanomaterials. Particle and Fibre Toxicology 11 (1), 20.
Collier, Z.A., Kennedy, A.J., Poda, A.R., Cuddy, M.F., Moser, R.D., MacCuspie, R.I., Harmon, A., Plourde, K., Haines, C.D., Steevens, J.A., 2015. Tiered guidance for risk-informed environmental health and safety testing of nanotechnologies. J. Nanopart. Res. 17 (3), 155.
Dalle-Donne, I., Giustarini, D., Colombo, R., Rossi, R., Milzani, A., 2003. Protein carbonylation in human diseases. Trends Mol. Med. 9.
Dalle-Donne, I., Aldini, G., Carini, M., Colombo, R., Rossi, R., Milzani, A., 2006. Protein carbonylation, cellular dysfunction, and disease progression. J. Cell. Mol. Med. 10.
Davies, M.J., 2005. The oxidative environment and protein damage. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1703.
DeLoid, G., Cohen, J.M., Darrah, T., Derk, R., Rojanasakul, L., Pyrgiotakis, G., Wohlleben, W., Demokritou, P., 2014. Estimating the effective density of engineered nanomaterials for in vitro dosimetry. Nat. Commun. 5, 3514.
Driessen, M.D., Mues, S., Vennemann, A., Hellack, B., Bannuscher, A., Vimalakanthan, V., Riebeling, C., Ossig, R., Wiemann, M., Schnekenburger, J., et al., 2015. Proteomic analysis of protein carbonylation: a useful tool to unravel nanoparticle toxicity mechanisms. Particle and Fibre Toxicology 12 (1), 36.
ECHA, 2008. Guidance on Information Requirements and Chemical Safety Assessment Chapter R.6: QSARs and Grouping of Chemicals.
FIRTH, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80 (1), 27–38.
Fu, P.P., Xia, Q., Hwang, H.-M., Ray, P.C., Yu, H., 2014. Mechanisms of nanotoxicity: generation of reactive oxygen species. J. Food Drug Anal. 22 (1), 64–75.
Gajewicz, A., Puzyn, T., Odziomek, K., Urbaszek, P., Haase, A., Riebeling, C., Luch, A., Irfan, M.A., Landsiedel, R., van der Zande, M., et al., 2018. Decision tree models to classify nanomaterials according to the DF4nanoGrouping scheme. Nanotoxicology 12 (1), 1–17.
Gandon, A., Werle, K., Neubauer, N., Wohlleben, W., 2017. Surface reactivity measurements as required for grouping and read-across: an advanced FRAS protocol. J. Phys. Conf. Ser. 838, 012033.
Giusti, A., Haase, A., 2020. SOP: Determining Protein Carbonylation. *zenodo*.
Giusti, A., Atluri, R., Tsekovska, R., Gajewicz, A., Apostolova, M.D., Battistelli, C.L., Bleeker, E.A.J., Bossa, C., Bouillard, J., Dusinska, M., et al., 2019. Nanomaterial

grouping: existing approaches and future recommendations. NanoImpact 16, 100182.
Halliwell, B., Whiteman, M., 2004. Measuring reactive species and oxidative damage in vivo and in cell culture: how should you do it and what do the results mean? Br. J. Pharmacol. 142.
He, W., Liu, Y., Wamer, W.G., Yin, J.J., 2014. Electron spin resonance spectroscopy for the study of nanomaterial-mediated generation of reactive oxygen species. J. Food Drug Anal. 22.
Hellack, B., Nickel, C., Albrecht, C., Kuhlbusch, T.A.J., Boland, S., Baeza-Squiban, A., Wohlleben, W., Schins, R.P.F., 2017a. Analytical methods to assess the oxidative potential of nanoparticles: a review. Environmental Science: Nano 4 (10), 1920–1934.
Hellack, B., Nickel, C., Schins, R.P.F., 2017b. Oxidative potential of silver nanoparticles measured by electron paramagnetic resonance spectroscopy. J. Nanopart. Res. 19 (12), 404.
Hofmann, T., Ma-Hock, L., Strauss, V., Treumann, S., Rey Moreno, M., Neubauer, N., Wohlleben, W., Gröters, S., Wiench, K., Veith, U., et al., 2016. Comparative short-term inhalation toxicity of five organic diketopyrrolopyrrole pigments and two inorganic iron-oxide-based pigments. Inhal. Toxicol. 28 (10), 463–479.
JRC, 2013. Synthetic Amorphous Silicon Dioxide (NM-200, NM-201, NM-202, NM-203, NM-204): Characterisation and Physico-Chemical Properties.
Keller, J., Wohlleben, W., Ma-Hock, L., Strauss, V., Gröters, S., Küttler, K., Wiench, K., Herden, C., Oberdörster, G., van Ravenzwaay, B., et al., 2014. Time course of lung retention and toxicity of inhaled particles: short-term exposure to nano-ceria. Arch. Toxicol. 88 (11), 2033–2059.
Kuempel, E.D., Castranova, V., Geraci, C.L., Schulte, P.A., 2012. Development of risk-based nanomaterial groups for occupational exposure control. J. Nanopart. Res. 14 (9).
Landsiedel, R., Ma-Hock, L., Hofmann, T., Wiemann, M., Strauss, V., Treumann, S., 2014. Application of short-term inhalation studies to assess the inhalation toxicity of nanomaterials. Part Fibre Toxicol 11.
Lanone, S., Rogerieux, F., Geys, J., Dupont, A., Maillot-Marechal, E., Boczkowski, J., Lacroix, G., Hoet, P., 2009. Comparative toxicity of 24 manufactured nanoparticles in human alveolar epithelial and macrophage cell lines. Particle and fibre toxicology 6, 14.
l'Environment Md, 2015. Eléments issus des déclarations des substances à l'état nanoparticulaire: Exercise 2015.
Ma-Hock, L., Burkhardt, S., Strauss, V., Gamer, A.O., Wiench, K., van Ravenzwaay, B., Landsiedel, R., 2009. Development of a short-term inhalation test in the rat using nano-titanium dioxide as a model substance. Inhal. Toxicol. 21 (2), 102–118.
Marano, F., Hussain, S., Rodrigues-Lima, F., Baeza-Squiban, A., Boland, S., 2011. Nanoparticles: molecular targets and cell signalling. Arch. Toxicol. 85 (7), 733–741.
Maser, E., Schulz, M., Sauer, U.G., Wiemann, M., Ma-Hock, L., Wohlleben, W., Hartwig, A., Landsiedel, R., 2015. In vitro and in vivo genotoxicity investigations of differently sized amorphous SiO2 nanomaterials. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 794, 57–74.
Mottier, A., Mouchet, F., Laplanche, C., Cadarsi, S., Lagier, L., Arnault, J.C., Girard, H.A., León, V., Vázquez, E., Sarrieu, C., et al., 2016. Surface area of carbon nanoparticles: a dose metric for a more realistic ecotoxicological assessment. Nano Lett. 16 (6), 3514–3518.
Nel, A.E., Nasser, E., Godwin, H., Avery, D., Bahadori, T., Bergeson, L., Beryt, E., Bonner, J.C., Boverhof, D., Carter, J., et al., 2013. A multi-stakeholder perspective on the use of alternative test strategies for nanomaterial safety assessment. ACS Nano 7 (8), 6422–6433.
OECD, 2014. Guidance on grouping of chemicals. In: Series on Testing & Assessment no 194, second edition. .
Oomen, A.G., Bleeker, E.A.J., Bos, P.M.J., van Broekhuizen, F., Gottardo, S., Groenewold, M., Hristozov, D., Hund-Rinke, K., Irfan, M.A., Marcomini, A., et al., 2015. Grouping and read-across approaches for risk assessment of nanomaterials. Int. J. Environ. Res. Public Health 12 (10), 13415–13434.
Oomen, A.G., Steinhäuser, K.G., Bleeker, E.A.J., van Broekhuizen, F., Sips, A., Dekkers, S., Wijnhoven, S.W.P., Sayre, P.G., 2018. Risk assessment frameworks for nanomaterials: scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency. NanoImpact 9, 1–13.
RCC RCC, 2003. Work Element 2, Development of a Classification Scheme for Nanomaterials Regulated under the New Substances Programs of Canada and the United States.
Riebeling, C., Wiemann, M., Schnekenburger, J., Kuhlbusch, T.A.J., Wohlleben, W., Luch, A., Haase, A., 2016. A redox proteomics approach to investigate the mode of action of nanomaterials. Toxicol. Appl. Pharmacol. 299, 24–29.
Rudzińska M, Parodi A, Balakireva AV, Chepikova OE, Venanzi FM, Zamyatnin AA, Jr.: Cellular aging characteristics and their association with age-related disorders. Antioxidants (Basel) 2020, 9(2):94.
Sauer, U.G., Aumann, A., Ma-Hock, L., Landsiedel, R., Wohlleben, W., 2015. Influence of dispersive agent on nanomaterial agglomeration and implications for biological effects in vivo or in vitro. Toxicology in vitro : an international journal published in association with BIBRA 29 (1), 182–186.
Schmid, O., Cassee, F.R., 2017. On the pivotal role of dose for particle toxicology and risk assessment: exposure is a poor surrogate for delivered dose. Particle and Fibre Toxicology 14 (1), 52.
Schmid, O., Stoeger, T., 2016. Surface area is the biologically most effective dose metric for acute nanoparticle toxicity in the lung. J. Aerosol Sci. 99, 133–143.
Shi, T., Schins, R.P.F., Knaapen, A.M., Kuhlbusch, T., Pitz, M., Heinrich, J., 2003. Hydroxyl radical generation by electron paramagnetic resonance as a new method to monitor ambient particulate matter composition. J. Environ. Monit. 5.
Sies, H., 2015. Oxidative stress: a concept in redox biology and medicine. Redox Biol. 4, 180–183.
Simkó, M., Nosske, D., Kreyling, W.G., 2014. Metrics, dose, and dose concept: the need for a proper dose concept in the risk assessment of nanoparticles. Int. J. Environ. Res. Public Health 11 (4), 4026–4048.
C. D. Smith JMC, P. E. Starke-Reed, C. N. Oliver, E. R. Stadtman, R. A. Floyd and W. R. Markesbery: Excess brain protein oxidation and enzyme dysfunction in normal aging and in Alzheimer disease. Proc Nati Acad Sci USA 1991, 88:10540–10543.
Stadtman, E.R., 2006. Protein oxidation and aging. Free Radic. Res. 40.
Stone, V., Johnston, H., Clift, M.J.D., 2007. Air pollution, ultrafine and nanoparticle toxicology: cellular and molecular interactions. IEEE Transactions on NanoBioscience 6 (4), 331–340.
Telci, A., Çakatay, U., Kayali, R., Erdoğan, C., Orhan, Y., Sivas, A., Akçay, T., 2000. Oxidative protein damage in plasma of type 2 diabetic patients. Horm. Metab. Res. 32 (01), 40–43.
Unfried, K., Albrecht, C., Klotz, L.-O., Von Mikecz, A., Grether-Beck, S., Schins, R.P.F., 2007. Cellular responses to nanoparticles: target structures and mechanisms. Nanotoxicology 1 (1), 52–71.
Uttara, B., Singh, A.V., Zamboni, P., Mahajan, R.T., 2009. Oxidative stress and neurodegenerative diseases: a review of upstream and downstream antioxidant therapeutic options. Curr. Neuropharmacol. 7 (1), 65–74.
Wang, G., Zhang, J., He, X., Zhang, Z., Zhao, Y., 2017. Ceria nanoparticles as enzyme mimetics. Chin. J. Chem. 35 (6), 791–800.
Wiemann, M., Vennemann, A., Sauer, U.G., Wiench, K., Ma-Hock, L., Landsiedel, R., 2016. An in vitro alveolar macrophage assay for predicting the short-term inhalation toxicity of nanomaterials. Journal of Nanobiotechnology 14, 16.
Wiemann, M., Sauer, U.G., Vennemann, A., Bäcker, S., Keller, J.-G., Ma-Hock, L., Wohlleben, W., Landsiedel, R., 2018. In vitro and in vivo short-term pulmonary toxicity of differently sized colloidal amorphous SiO_2. Nanomaterials (Basel) 8.
Wohlleben, W., Punckt, C., Aghassi-Hagmann, J., Siebers, F., Menzel, F., Esken, D., Drexel, C.-P., Zoz, H., Benz, H.U., Weier, A., et al., 2017. Nanoenabled products: categories, manufacture, and applications. In: Metrology and Standardization of Nanotechnology, pp. 409–464.
Wohlleben, W., Hellack, B., Nickel, C., Herrchen, M., Hund-Rinke, K., Kettler, K., Riebeling, C., Haase, A., Funk, B., Kühnel, D., et al., 2019. The nanoGRAVUR framework to group (nano)materials for their occupational, consumer, environmental risks based on a harmonized set of material properties, applied to 34 case studies. Nanoscale 11 (38), 17637–17654.
Xia, T., Kovochich, M., Nel, A.E., 2007. Impairment of mitochondrial function by particulate matter (PM) and their toxic components: implications for PM-induced cardiovascular and lung disease. Front. Biosci. 12, 1238–1246.
Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., Liu, R., Pokhrel, S., Lin, S., Wang, X., Liao, Y.-P., et al., 2012. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. ACS Nano 6 (5), 4349–4368.

3.3 PROTEOMAS: a workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping

Aileen Bahl[1], Celine Ibrahim[1], Kristina Plate[1], Andrea Haase[1], Jörn Dengjel[2], Penny Nymark[3] and Verónica I. Dumit[1*]

1 Department of Chemicals and Product Safety, German Federal Institute for Risk Assessment (BfR), Berlin, Germany

2 University of Fribourg, Fribourg, Switzerland

3 Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

This chapter was published as a research article in the *Journal of Cheminformatics* in 2023 (https://doi.org/10.1186/s13321-023-00710-2).

The supplementary information for this publication can be found in Annex III.

Journal of Cheminformatics

RESEARCH

Open Access

PROTEOMAS: a workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping

Aileen Bahl[1], Celine Ibrahim[1], Kristina Plate[1], Andrea Haase[1], Jörn Dengjel[2], Penny Nymark[3] and Verónica I. Dumit[1*]

Abstract

Toxicological evaluation of substances in regulation still often relies on animal experiments. Understanding the substances' mode-of-action is crucial to develop alternative test strategies. Omics methods are promising tools to achieve this goal. Until now, most attention was focused on transcriptomics, while proteomics is not yet routinely applied in toxicology despite the large number of datasets available in public repositories. Exploiting the full potential of these datasets is hampered by differences in measurement procedures and follow-up data processing. Here we present the tool PROTEOMAS, which allows meta-analysis of proteomic data from public origin. The workflow was designed for analyzing proteomic studies in a harmonized way and to ensure transparency in the analysis of proteomic data for regulatory purposes. It agrees with the Omics Reporting Framework guidelines of the OECD with the intention to integrate proteomics to other omic methods in regulatory toxicology. The overarching aim is to contribute to the development of AOPs and to understand the mode of action of substances. To demonstrate the robustness and reliability of our workflow we compared our results to those of the original studies. As a case study, we performed a meta-analysis of 25 proteomic datasets to investigate the toxicological effects of nanomaterials at the lung level. PROTEOMAS is an important contribution to the development of alternative test strategies enabling robust meta-analysis of proteomic data. This workflow commits to the FAIR principles (Findable, Accessible, Interoperable and Reusable) of computational protocols.

Keywords Proteomics, Harmonized proteomics data analysis, Meta-analysis, Mode-of-action (MoA), Adverse outcome pathways (AOP), Nanomaterials, FAIR data

Introduction

Animal testing is still key in risk assessment of chemical substances but in vivo experiments imply exorbitant costs. The high number of different toxicological endpoints that need to be evaluated is also a bottleneck when assessing substance toxicity. The increasing number of substances to be introduced in the market calls for the development of reliable alternative methods. The most commonly used experimental alternative models are in vitro tests based on cell cultures that are typically used to assess acute effects. However, to adequately cover more complex endpoints and in particular chronic effects, integrated test strategies that combine a series of different assays are needed. Developing such test strategies requires mechanistic understanding of the underlying biological changes caused by the substances.

*Correspondence:
Verónica I. Dumit
Veronica.Dumit@bfr.bund.de
[1] Department of Chemicals and Product Safety, German Federal Institute for Risk Assessment (BfR), Berlin, Germany
[2] University of Fribourg, Fribourg, Switzerland
[3] Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

In toxicology, a key concept to depict mechanistic knowledge of the effect of a substance at different biological levels is the concept of Adverse Outcome Pathway (AOP), which is a robust framework to contribute to regulatory decision making [1, 2]. AOPs address the alterations induced by a substance at the molecular, cellular, organ and organism level [3] and aim to describe the substance mode-of-action (MoA) [4] as a series of key events. Different in vitro and in silico technologies can then be applied to evaluate the key events preceeding the adverse outcome.

Omics-based technologies became important in toxicology because they allow to investigate toxicity mechanisms in a holistic manner. In this way, they account for the generation of vast datasets at different biological levels [5]. Although these approaches can provide detailed insights into MoA at molecular and cellular levels [6, 7], omics technologies are not yet part of the routine methods in regulatory hazard assessment procedures because standardization of the computational models for interpretation of the datasets is still needed [8, 9]. Workflows for harmonized analysis of omic data contribute directly to facilitate the use of omics in regulatory-decision making.

Among all omic techniques, transcriptomics has an immediate potential in this field, because data generation and analysis can be well harmonized and results allow for straightforward comparison between experiments. However, the major drawback of transcriptomics is the relatively indirect relationship between the measured effects and the respective phenotype. Proteomics, despite being able to describe closer the phenotype, is not generally performed in a harmonized manner. Next to inherent technical challenges, several factors contribute to the lack of uniformity of proteomics measurements: there are no unified experimental design nor sample preparation protocols, and the different degrees of sophistication of the measuring devices result in high level of noise. Additionally, datasets available on public repositories frequently suffer from insufficient metadata, hindering the assignment of the correct experimental condition to each file within the dataset. Moreover, different methods for analyzing and modeling the data often lead to different results, hampering the comparison of the data originated from separate studies. Although analytic methods are equally valid, their pipelines are usually adapted to fit the datasets generating an impact on the outcome. These challenges call for attention if publicly available data is meant to be reused [10].

In this work, we introduce PROTEOMAS, a workflow designed to analyze proteomic studies in a harmonized and transparent manner with the aim to increase their potential for (re)use in toxicological regulatory processes. The workflow follows the Omics Reporting Framework by the Organisation for Economic Cooperation and Development (OECD) [11, 12], precisely to the Data Acquisition and Processing Reporting Module (DAPRM) and to the Data Analysis Reporting Module (DARM) for discovery of differently abundant molecules. It intends to integrate proteomics to transcriptomics and metabolomics, which are so far the only omic techniques further accepted in regulatory matters. During the analysis, a log file collecting all relevant information according to the Omics Reporting Framework is created which guarantees transparency of all steps and results. The overarching aim of PROTEOMAS is to contribute to the understanding of the MoA of substances and to the development of AOPs. Notably, our workflow complies with the FAIR principles (Findable, Accessible, Interoperable and Reusable) of bioinformatics tools, and contributes to data FAIRness of proteomics studies [13].

Results

The PRIDE Archive, one of the main public repositories for proteomic data, currently hosts over 20,000 projects. This large amount of data has great potential in toxicology. However, it is difficult to use these datasets to compare the outcomes of different projects due to their heterogenic nature. Apart from the technical differences, the large variety of analytic workflows and interpretation tools hinders comparability of the results. The tool that we introduce in this work, PROTEOMAS (PROTEOmics Meta-AnalysiS), can perform automated and harmonized meta-analyses of data-dependent acquisition (DDA) proteomic datasets using the popular and commonly used label-free quantification (LFQ) algorithm of the freeware MaxQuant [14]. PROTEOMAS functions independently of technical specifications and of metadata availability. PROTEOMAS can process results obtained from different devices, with the only condition that generated files can be analyzed with MaxQuant. Currently supported file formats are *.wiff (ABSciex), *.mzxml (MzXml), *.raw (Thermo), *.uimf (UIMF), and *.d (Agilent and Bruker). This approach then enables the comparison of results from different studies.

Figure 1 provides an overview on the different processing steps included in the workflow. The **numbers** in the flowchart indicate the different steps which are discussed in detail below. Some steps are decision-based and depend on different criteria regarding the characteristics of the respective datasets and their associated metadata. The workflow can be applied to publicly available datasets from repositories or to newly generated ones.

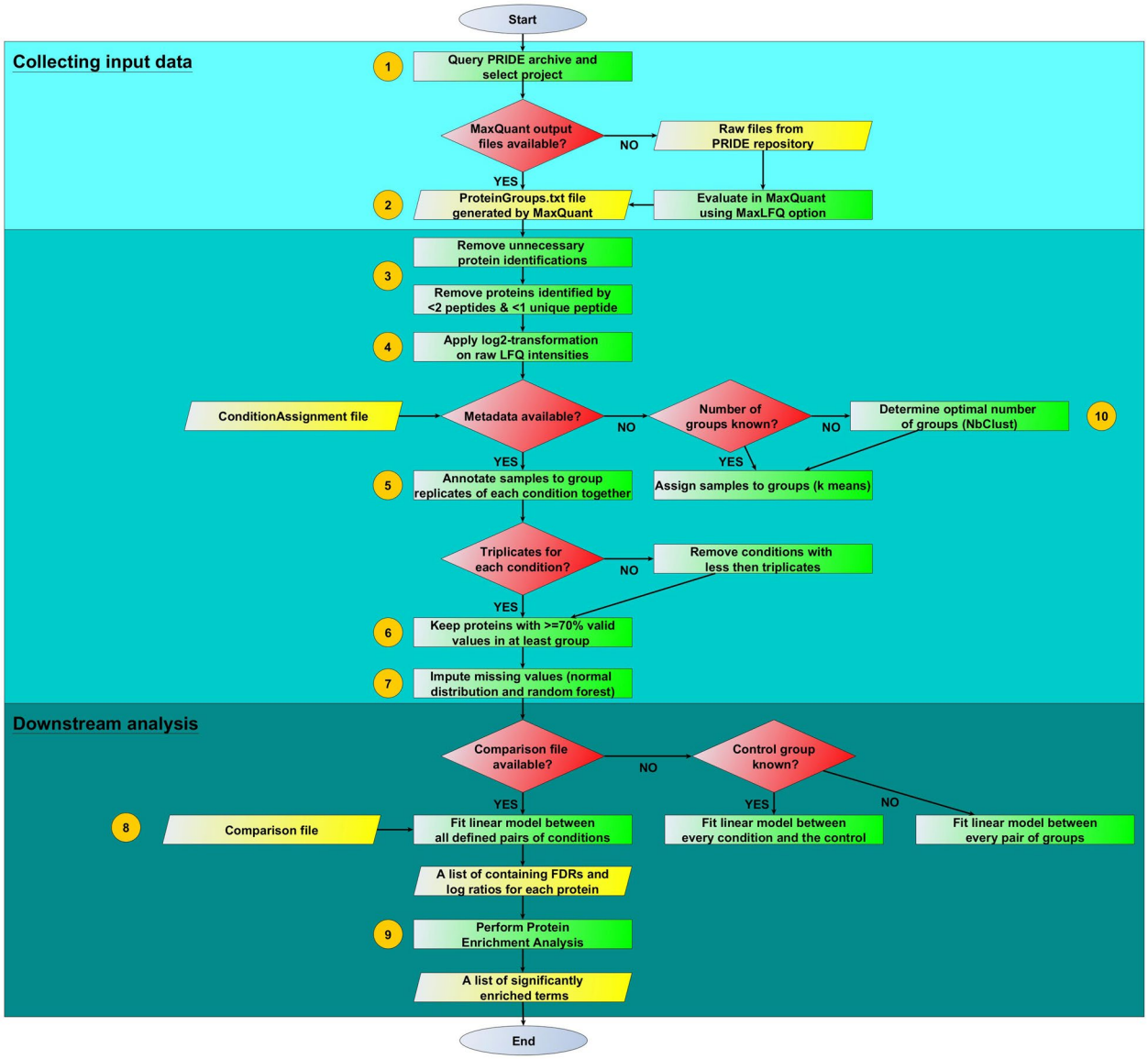


Fig. 1 Flowchart showing the processing steps within the workflow. Each implemented step is represented by a separate rectangle (green). Decisions based on certain criteria are represented by diamond shapes (red), while input and output files are shown in parallelograms (yellow). Numbers correspond to the processing steps performed and are further explained in the text

Preparation of the proteomic datasets

PROTEOMAS can be used for any dataset generated by a label-free proteomics approach (Fig. 1, step 1). The *sine qua non* condition for the dataset is the possibility to run MaxQuant on it. MaxQuant [14] is a widely used proteomics software for identification and quantification of proteins analyzed by mass spectrometry (MS). Raw files downloaded from the PRIDE repository have to be analyzed with MaxQuant including the option 'LFQ intensities' before running PROTEOMAS. This setting allows for a generic normalization and quantification technique called MaxLFQ [15]. MaxLFQ performs delayed normalization in combination with maximum peptide ratio extraction. Thereby, it solves two common problems occurring during quantification of label-free proteomics data: (1) Delayed normalization removes biases occurring from slight differences in handling and MS performance between sample fractions. The only assumption here is that most proteins do only change minimally between experimental conditions. (2) The maximum peptide ratio extraction algorithm defines the selection of peptide signals which contribute to the overall protein signal

across samples. It calculates all pairwise protein ratios among samples based on all shared peptides belonging to the protein of interest. By default, at least two peptide ratios are needed to obtain a valid protein ratio. This default value was not changed in the analysis. As a last step, LFQ intensity profiles are calculated for each protein such that all pairwise peptide comparisons are satisfied and the best estimate is obtained. The underlying assumption of MaxLFQ is that the majority of proteins is not changing between analyzed conditions. However, in the original publication of MaxLFQ, the authors tested this assumption in a benchmark dataset in which more than 30% of all identified proteins were changed. While there was a shift in total log ratios between changed and non-changed proteins, changed proteins could still be detected and quantified as such [15].

MaxQuant generates, among others, a 'proteinGroups.txt' file as output, which is the main input required for running PROTEOMAS (Fig. 1, step 2). In addition to the 'proteinGroups.txt' file, the user can create three additional optional files, where information about each measured sample can be described: (1) The 'ConditionAssignment.csv' should be used if each treatment condition is known for each sample. Within this file each original sample name is assigned to the associated condition. (2) A more precise specification of the comparisons to be evaluated (e.g. treatment 1 *vs*. control 1 and treatment 2 *vs*. control 2) can be transferred to the workflow through the next optional file called 'Comparisons.csv'. (3) If, on the other hand, information about the treatment conditions is unknown, the 'ClusterNumber.csv' file may be used if the total number of different treatments conditions is known.

Data pre-processing

Within PROTEOMAS data pre-processing starts by loading the 'proteinGroups.txt' file. The input file is preprocessed to remove non-relevant data (Fig. 1, step 3), precisely, proteins marked as 'contaminants', 'identified only by site' or 'reverse'. To minimize protein misidentification, only proteins identified by at least two peptides with at least one of them being unique were kept for downstream analysis. In addition, data was log2-transformed (Fig. 1, step 4). In a technically sound dataset, one would expect that log2-transformed values of the LFQ intensities show a normal distribution when plotted as histograms, which in the next steps supports the use of downstream analysis methods which often assume normally distributed data. The workflow produces the corresponding figures and collects them directly in the output folder. This information can be used for analyzing data quality. No external data normalization is performed as this step is inherently done with MaxQuant using the MaxLFQ option.

DDA proteomic datasets typically contain a large number of missing values, which are listed as zeros in the output files. A missing value in the dataset does not necessarily mean that the respective protein was not present in the sample; it means that there were too few data points for proper quantification [16]. After a log2-transformation, values equal to zero will be converted to non-assigned numbers (NaN). In case sufficient metadata is available, all samples will be assigned to their corresponding condition as indicated in the 'ConditionAssignment.csv' file before continuing the analysis (Fig. 1, step 5). The case of missing metadata is described later on. After assigning samples to their respective conditions, it is checked whether at least triplicates are present in each group as this is a minimal pre-requisite for successful outlier detection and statistical testing. If this is not the case for some of the conditions, those are deleted from the dataset.

Dealing with missing values

A threshold of minimal valid values for each protein entry of 70% in at least one condition group was set as default (Fig. 1, step 6). After filtering proteins, some missing values will very likely remain in the datasets. These values can be replaced by valid values by a process called imputation (Fig. 1, step 7). Imputation allows to retain the full sample size of detected proteins [17], which can contribute to improving the proteome coverage and the determination of enriched descriptors.

There are different imputation methods, which can be divided into two classes: MCAR and MNAR methods [17, 18]. In MCAR methods, values are assumed to be missing completely at random. In the case of proteomics, this would mean that only by chance peptides were not detected by the mass spectrometer. As an example, this could happen if a more abundant peptide elutes at the same time and overshadows the presence of another peptide, which goes undetected. In that case, missing values would optimally be replaced by values, which are in the same range as those of the other replicates within this condition for the protein under consideration. Available methods comprise, e.g. the k-nearest neighbors (kNN) method [19] or the random forest (RF) method [20]. In contrast, MNAR methods assume that values are missing not at random and thus the protein is truly absent. Common examples of MNAR imputation methods are replacement by LOD (limit of detection) values or sampling from a downshifted and shrunk normal distribution which means that missing values are replaced by small values. In proteomic datasets, one would usually expect

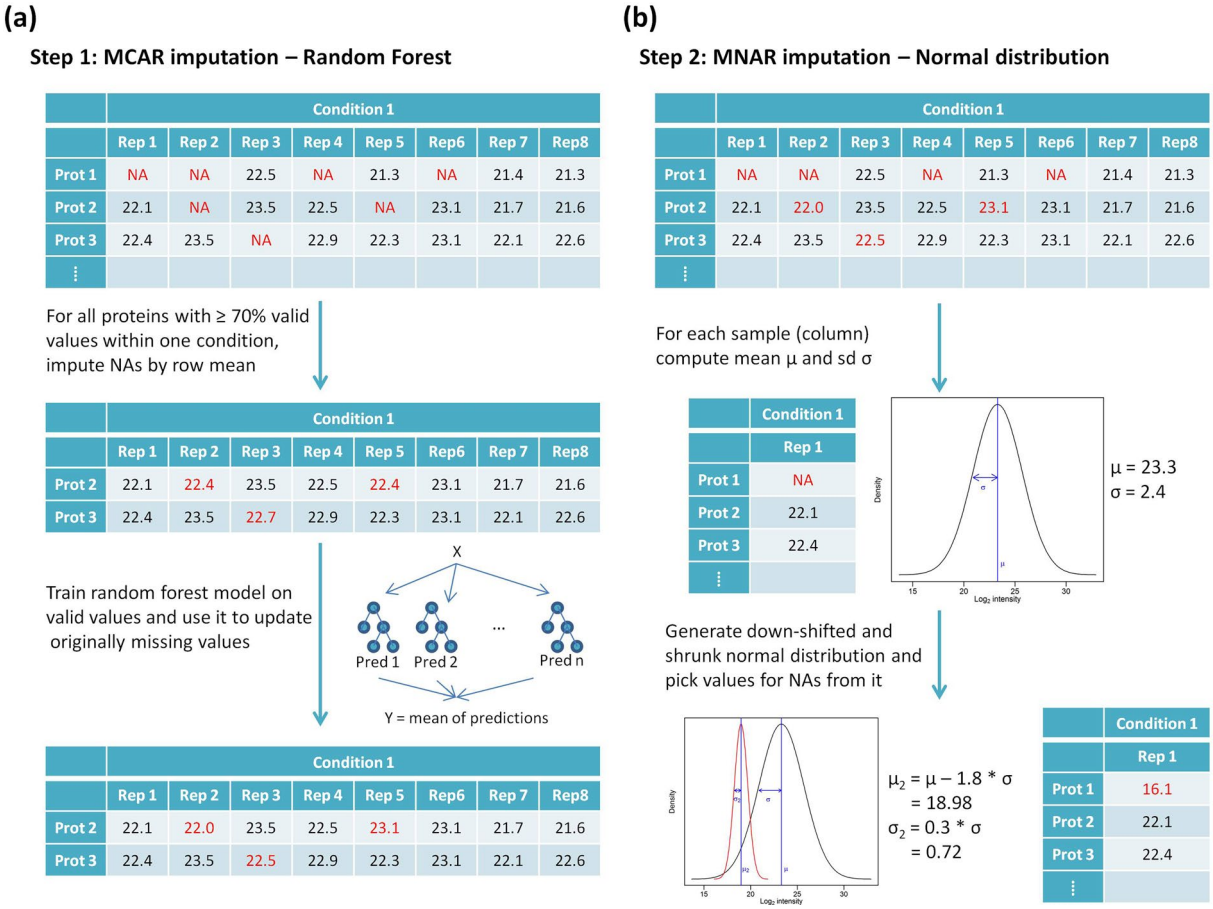


Fig. 2 MCAR *vs.* MNAR imputation. Within the PROTEOMAS workflow these are implemented in terms of **a** Step 1: random forest imputation and **b** Step 2: imputation from a down-shifted and shrunk normal distribution

to see both, MCAR and MNAR values; however, it is impossible to determine the exact type for each missing value. Therefore, within the workflow, we used a combination of MCAR and MNAR methods and the decision on which one to use is based on the amount of missing values across samples within a condition.

First, we extract all proteins which have at most 30% missing values within the specified condition. We assume that these are actually MCAR values as they were detected in most replicates of the same condition. For those, missing values are replaced with random forest imputation as in that case, we would actually expect the protein to be present in all samples of that condition (in concordance with the filtering based on valid values) (Fig. 2). Random forest imputation starts with replacing all missing values by the mean value of that protein within a given condition and then generating random forest models each time leaving out one of the originally missing values. Each random forest model then predicts a new value which replaces the mean value. This step is done iteratively in order to obtain better results.

After random forest imputation, all other missing values are assumed to be MNAR values and are thus replaced by small values obtained from an imputation based on drawing values from the downshifted and shrunk normal distribution (Fig. 3). In this approach, the width and the center of each sample are calculated separately to simulate random values, which are used to fill the missing values of each sample, such that the width of the distribution will shrink to a factor of 0.3 (default) and the distribution will be downshifted by 1.8 (default) standard deviations (sampling from the left side of the distribution) for each sample. Histograms can be used to check the imputation and dataset quality. An example of histograms before and after imputation is shown in Fig. 3. In addition, Fig. 4 shows a boxplot confirming that samples are comparable and no further normalization is needed after MaxLFQ. The complete collection of quality control plots can be found in the GitHub repository.

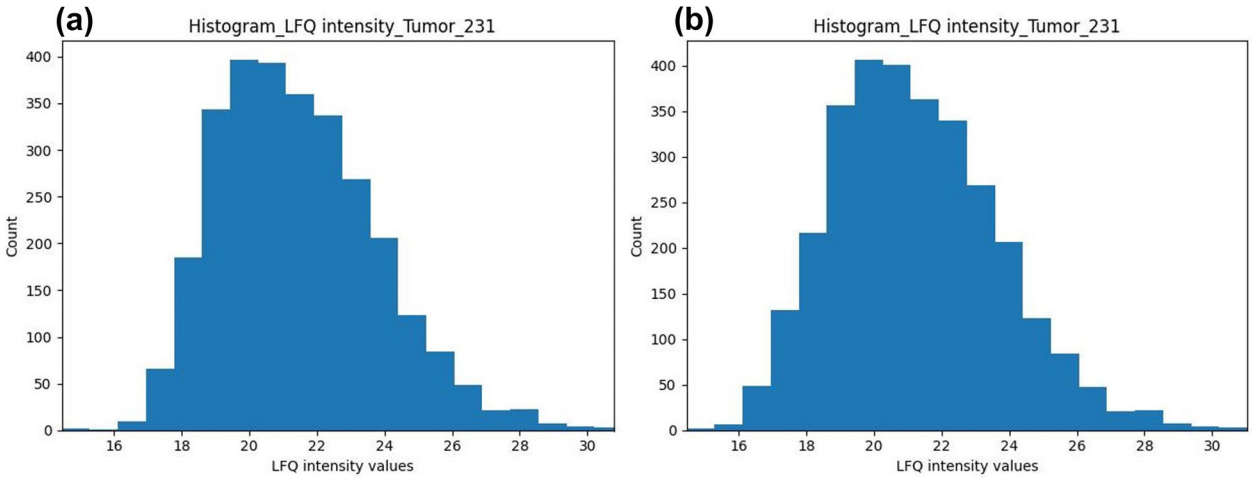


Fig. 3 Histograms comparing the data distribution **a** before and **b** after imputation

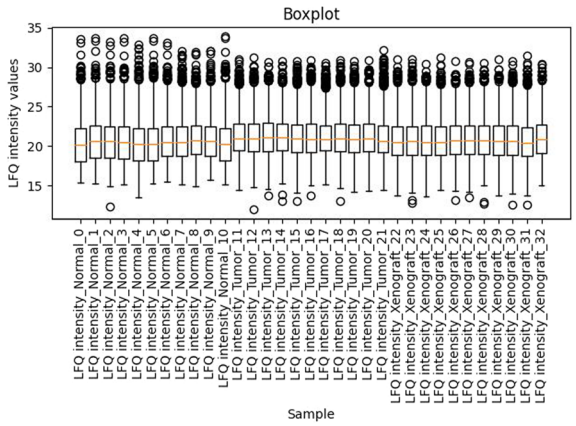


Fig. 4 Boxplot showing the comparability of samples after MaxLFQ normalization

Differential analysis of quantitative changes in protein levels

Remaining proteins including both valid and imputed values are subjected to differential analysis. This allows to detect changes in protein levels between different samples or conditions, while determining along the degree of statistical significance. Here, we used linear modeling (Fig. 1, step 8) to identify proteins which show a significant difference in abundance between two conditions. A Benjamini–Hochberg FDR threshold of 0.05 is used to correct for multiple testing. A fold change threshold of 1.5 up- or downregulation was set for determining significantly changed protein levels between two conditions. For visual inspection of the results, the workflow creates PCA plots and heatmap (Fig. 5) showing the clustering of all groups within one project as well as volcano plots for each comparison (Fig. 6). Plots for all projects are provided in the GitHub repository.

The conditions to be tested against each other can be defined in the 'Condition.csv' file. If no such file is specified, PROTEOMAS will look for any condition named 'control' and compares all conditions against this one. In case no 'Condition.csv' file, as well as no 'control' condition is available, all pairs of conditions are compared against each other.

Dealing with (missing) metadata

As it is often the case in repositories, the lack of metadata adjoining the datasets hampers proper comparison among treatments or conditions, as the relationship between raw files and corresponding measured samples is not clear. It is still possible to identify clusters of samples according to similarities of protein patterns, but typically criteria to separate treatment or condition groups remains subjective. An additional difficulty arises if the number of conditions evaluated in the dataset is unknown. PROTEOMAS, on the other hand, is able to perform assignment of conditions to each sample in an automated and objective fashion, without subjective bias.

In case metadata is not sufficient to directly assign experimental conditions to each sample, an additional automated condition assignment step is included in PROTEOMAS (Fig. 1, step 10). Here, each sample will be assigned to its condition group using a k-means clustering approach. In k-means clustering, k random cluster centers are defined and each sample is assigned to its nearest cluster center based on Euclidean distance. Then cluster centers are recalculated based on the assigned samples and samples are reassigned to the new center means. This is continued in an iterative fashion until the

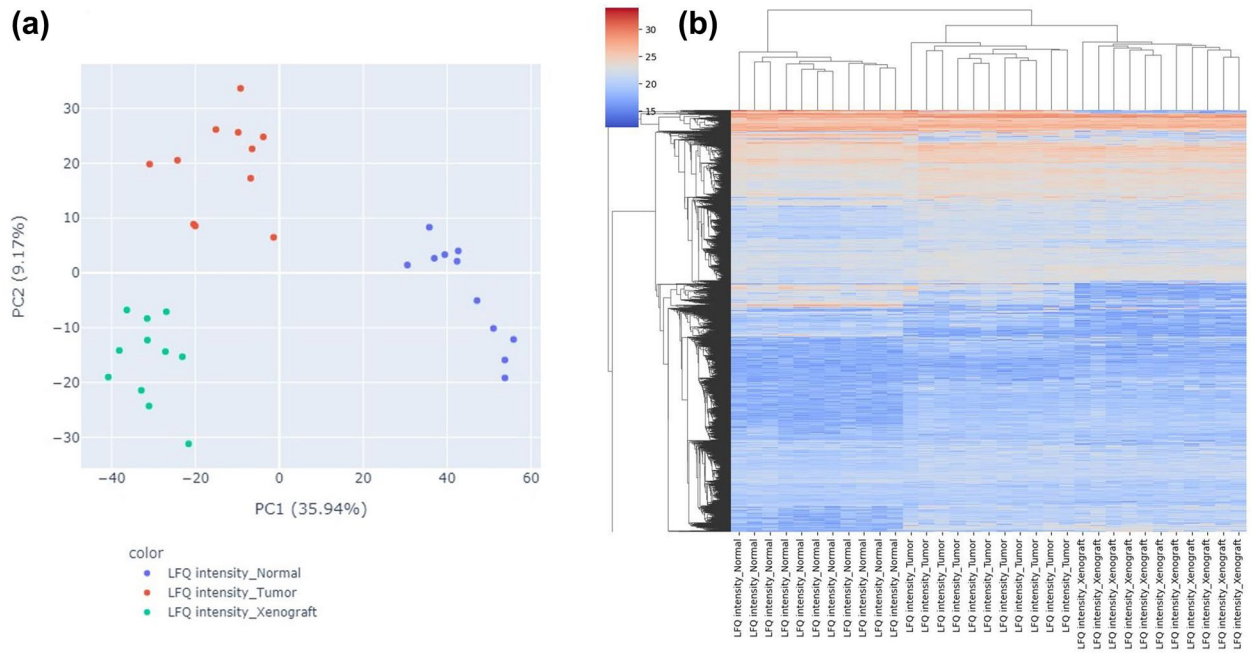


Fig. 5 **a** PCA plot and **b** heatmap showing clustering of samples within the different conditions of project 'PXD000853'

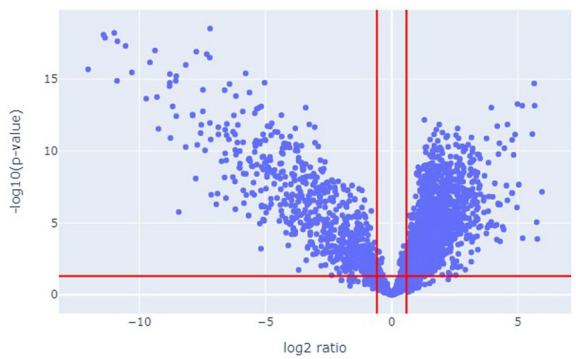


Fig. 6 Volcano plot showing differentially abundant proteins between healthy humans and mice with xenograft tumors for project 'PXD000853'

algorithm converges and group assignments no longer change.

The crucial point in k-means clustering is the value of k, which is the number of clusters to define. In case group assignments cannot directly be obtained from the metadata but still the number of groups is known, k-means algorithm can be performed directly. Otherwise, if the number of groups is also unknown, k first has to be determined. Although the determination of the optimal number of groups could be done by visual inspection of hierarchical clustering or PCA plots, this option is not feasible when processing a large number of projects, and it implies a subjective bias. Therefore, the determination of the optimal number of clusters k is done automatically in this workflow. Multiple methods for detecting the optimal number of k exist and a number of them are implemented in the R-package 'NbClust' [21].

To find the most suitable method for determining the optimal k and at the same time also assess the quality of the condition assignment using k-means, we blinded all studies considered for the case study below, which do have sufficient metadata and compared the outcomes in terms of significant proteins (Fig. 7) and KEGG pathways of the blinded and the nonblinded approach (Fig. 8). For the final implementation of PROTEOMAS, cindex was chosen for determining the optimal number of clusters k as it shows the highest recovery of KEGG pathways. Figure 8 shows the amount of KEGG pathways found to be significantly altered in the nonblinded and blinded setting, as well as their overlap for each analyzed project. For blinded analyses, the determined number of k is used as the number of cluster centers to be used for k-means clustering. Each sample of the dataset is then assigned to one of the clusters. An example of the condition assignment by k-means is shown in Fig. 9. Index 'fixedK' corresponds to the case when the number of conditions is set manually by the user using the 'ClusterNumber.csv'. Other plots

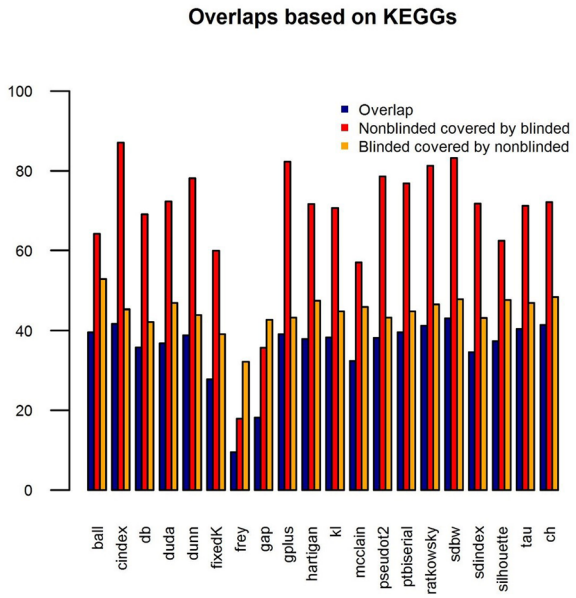


Fig. 7 Barplot showing the percentage of overlapping identified KEGG pathways between blinded and nonblinded studies for different indices used to determine the optimal number of clusters k. Each bar represents the mean percentage across all analyzed projects. For each index, the overlap between nonblinded and blinded is shown in blue, the percentage of KEGG pathways in the nonblinded dataset that could also be determined with the blinded setting is shown in red and the percentage of KEGG pathways in the blinded setting that was also present in the original nonblinded setting is depicted in yellow

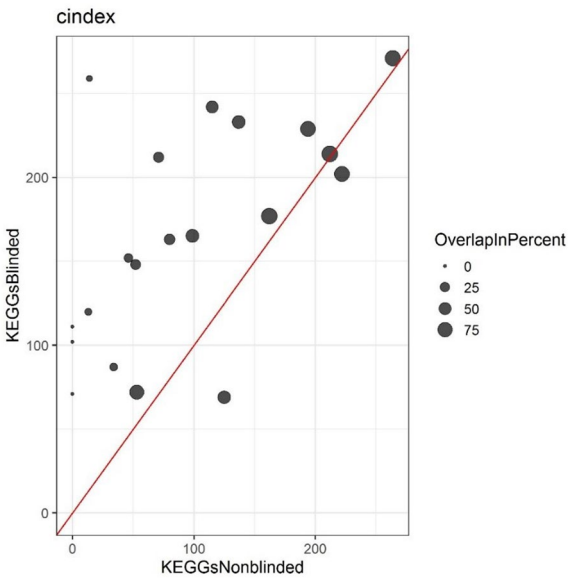


Fig. 8 Amount of KEGG pathways found to be significantly altered in the nonblinded, as well as the blinded setting for each analyzed project. Percent overlaps are shown by the size of the dots

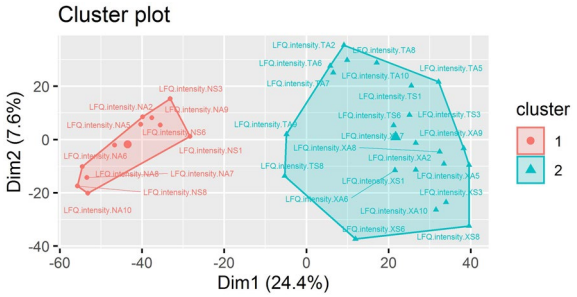


Fig. 9 Example of automatic cluster detection (project 'PXD000853'). The optimal number of clusters k is obtained from method cindex. This choice of k is then used for group assignment using a k-means clustering approach. The cluster plot indicates the group assignment of each sample after k-means clustering with the predicted number of conditions (in this case 2)

corresponding to this step can be found on the GitHub repository.

Protein enrichment analysis

Gene set enrichment analysis (GSEA) was used to identify enriched KEGG, Reactome and HALLMARK pathways as well as GO terms (Fig. 1, step 9). For each condition, the list of proteins is sorted in ascending order using the following formula: -log10 (FDR) * abs (log ratio). Thereby, proteins with small FDR and large log ratio are shifted to the beginning of the list, while those with large FDR and small log ratio occur at the end of the list. GSEA then identifies gene set which show over-representation at the top of the protein list. If enough proteins of a gene set are ranked at the top of the list, the gene set will show significant enrichment. These significantly enriched gene sets are collected in the final output file of PROTEOMAS and constitute the proteomic fingerprint of the studied condition.

Comparison with the original findings

In order to compare results from the analysis with our workflow to the original ones, we extracted lists of significantly altered proteins from the publications for randomly selected studies [22–24], and performed enrichment analysis of KEGG pathways. Original results were then compared to the lists we obtained using PROTEOMAS. Results of these comparisons are shown in Fig. 10. For all projects under consideration, we see very similar trends: PROTEOMAS consistently finds a similar set of significantly altered proteins and KEGG pathways like the original publications. In all cases, the major part of proteins and pathways are shared between the original publications and PROTEOMAS. Overlaps between original findings and those from our workflow range from

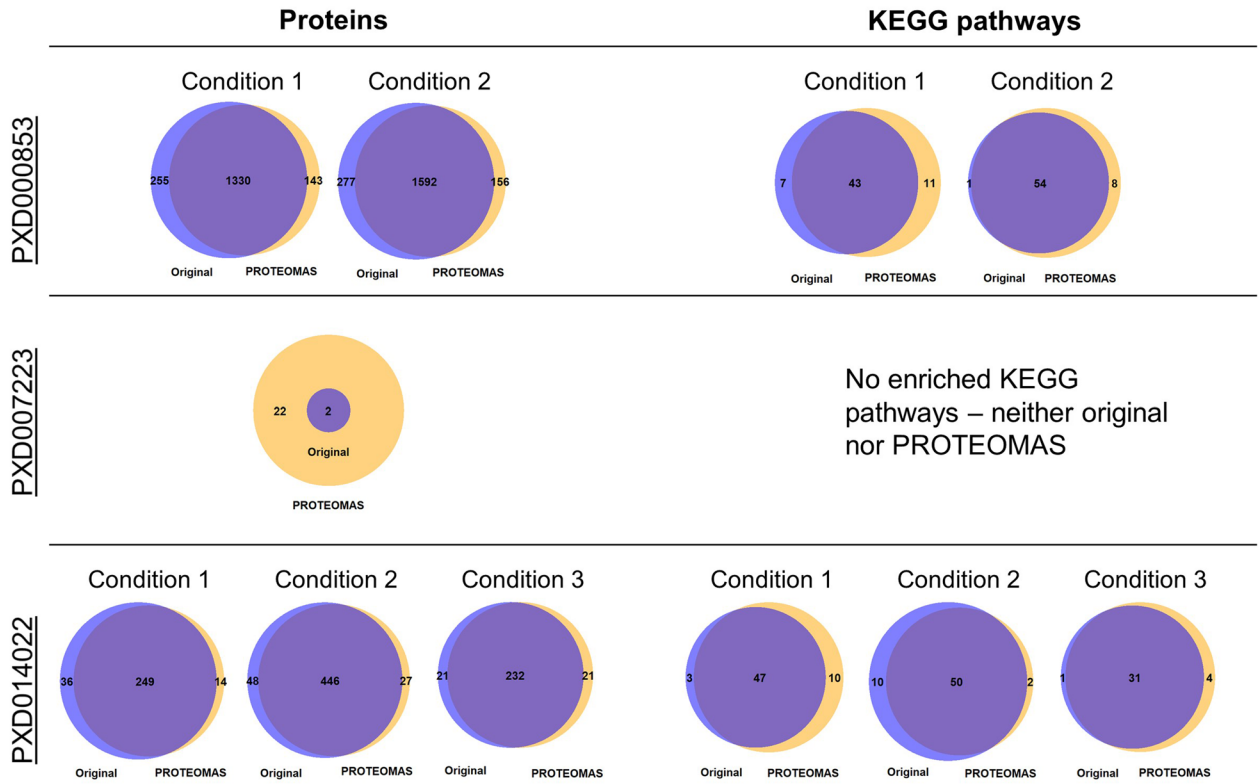


Fig. 10 Venn diagrams showing the overlap of the original findings for projects 'PXD007223', 'PXD000853' and 'PXD014022', and those obtained from PROTEOMAS

19 to 86% at the protein level and 38% to 85% for KEGG pathways. Therefore, especially on the level of KEGG pathways, which is assumed to be less prone to false-positive findings, PROTEOMAS gives very similar results. In addition, PROTEOMAS detects only very few additional pathways and thus creates only minimal noise.

Case study: comparison of the proteomic fingerprint of different projects related to lung conditions

As a case study, we tested the workflow to evaluate the toxicological effects of nanomaterials (NMs) at the lung level. NMs consist of particles of which at least 50% are 1 to 100 nm in size in terms of at least one external dimension [25]. Several comprehensive projects have studied the effect of NMs by omics measurements from in vivo and in vitro experiments. Although these results have contributed to the understanding the NMs hazards, the collected information is still not yet sufficient to conclusively unravel different MoA in detail, since the number of NM omic datasets is still comparatively low. However, one may assume that NMs will to a large extent share common MoA with conventional chemicals or other conditions. It is likely that NMs will have unique initiating events, but the following downstream physiological changes are likely to be shared by other effectors. As for NMs inhalation is considered the most critical route of exposure, we kept the focus of our work on lung proteomic datasets. Within this case study, we demonstrate how by means of PROTEOMAS, we are able to extract mechanistic information from different proteomic studies publicly available.

We analyzed 25 lung-related proteomics studies obtained from the PRIDE Archive within this case study. These include studies on lung cancer, pulmonary fibrosis, invasive pulmonary aspergillosis (IPA), chronic obstructive pulmonary disease (COPD), SARS-CoV-2 (Covid-19), and various NM treatments. Table 1 provides information on some project characteristics, as well as the total number of identified proteins, as obtained from the 'proteinGroups.txt' file, and it indicates whether metadata to the corresponding dataset is available. Table 2 shows the number of significantly altered proteins as well as enriched descriptors, which constitute the proteomic fingerprint for each comparison.

Comparing proteomic signatures across multiple datasets

Evaluated projects usually contain more than one condition, since different treatment, time-points,

Table 1 Overview of proteomic datasets used in this case study and their characteristics.

Project id	Species	In vitro/ in vivo model	Trait	Total number of proteins in raw data	Sufficient metadata available?	Number of pairs of conditions compared
PXD007223	Human	A549	Lung cancer	2008	Yes	1
PXD000861	Human	BEAS-2B	Lung cancer	3670	Yes	4
PXD018895	Human	A549	Lung cancer	3744	Yes	1
PXD000853	Human	A549	Lung cancer	5197	Yes	2
PXD005698	Human	A549, H358	Lung cancer	942	Yes	2
PXD005733	Human	Lung cancer and adjacent tissue	Lung cancer	1936	Yes	1
PXD007137*	Human	NCI-H650	Lung cancer	1321	No	2
PXD004818*	Human	Lung tissue	Lung cancer	2811	No	6
PXD007180	Human	A549	Smoking	2590	Yes	4
PXD020470	Human	HPA-HULEC co-culture	SARS CoV-2	6753	Yes	2
PXD021685	Human	THP-1	SARS CoV-2	1787	Yes	2
PXD007148h	Human	A549	COPD	466	Yes	2
PXD007148m	Mouse	Lung tissue	COPD	875	Yes	2
PXD016664h	Human	Lung tissue and BALF	IPA	5118	Yes	1
PXD016664m	Mouse	Lung tissue and BALF	IPA	3054	Yes	2
PXD014022	Human	A549	IPA and P. aeruginosa infection	4184		3
PXD005834*	Mouse	A549	IPA	2790	No	4
PXD018569	Human	NCI-H2030	> 30 drugs	8773	Yes	27
PXD023041*	Mouse	Lung tissue	Influenza	3440	No	0
PXD013244	Mouse	Blood serum	Gu-Ben-Fang-Xiao decoction (GBFXD)	3429	Yes	2
PXD016148	Mouse	BALF	NMs (Fe, Co, CB)	1525	Yes	22
PXD019267	Human	THP-1	31 NMs	3665	Yes	33
PXD018900	Rat	BALF	NM-401 (MWCNT)	1223	Yes	8
PXD005970	Human	HBEC-3KT	NM-400 (MWCNT)	5483	Yes	2
PXD025423	Human	HBEC-3KT	NM-62002a (TiO2)	5483	Yes	2

Projects marked with * had insufficient metadata

concentrations, etc., belong to the same dataset. For the present analysis, we have merged the conditions into traits, resulting in the following categories: lung cancer, aspergillosis, COPD, different drug treatments, viral infection, different NMs like carbon, ion releasing and TiO_2 NMs, among others. Each trait was normalized by the amount of conditions included for an equilibrated comparability.

Enriched HALLMARK pathways within the different traits were compared in a meta-analysis and results are depicted as a heatmap in Fig. 11. Hierarchical clustering was performed among traits as well as HALLMARK pathways. For this case study we have added a pathway that we created especially for this analysis: the "Lung Inflammation Key Event". This pathway includes proteins and genes known to be regulated in lungs undergoing inflammation, as collected from 35 papers addressing explicitly this topic. The list of 266 proteins and genes, as well as the citation to the original articles, are included in the Additional file 1: Table S1. The aim of the "Lung Inflammation Key Event" pathway was to gain a comprehensive description of an important key event often present in different AOPs, and particularly in the AOP for lung fibrosis. LPS serves as a positive control for activation of inflammatory response, and it proved to strongly regulate this pathway. Figure 11 shows that different types of NMs exhibit particular behaviors, and the caused alterations resemble distinct traits. This is true also for the "Lung Inflammation Key Event" pathway.

Table 2 Overview on the number of significantly enriched gene set for different background datasets for each analyzed trait in the case study

Trait	# Significantly altered GO terms	# Significantly altered KEGG pathways	# Significantly altered REACTOME pathways	# Significantly altered HALLMARK sets
Lung cancer	1462	46	181	35
IPA	1065	42	189	22
CNT NM-400	396	11	44	12
COPD	312	6	110	8
Drug	1739	47	329	25
CB	19	0	6	0
Co NM	6	1	7	2
Fe NM	80	1	18	0
FeCo NM	127	1	30	0
CNT NM-403	201	2	13	3
Ag NM	185	6	21	5
Au NM	417	10	93	9
Other CNT	300	6	79	8
CuO NM	300	3	65	9
ND	145	1	10	1
LPS	140	4	19	7
QD	276	8	66	8
TiO2	696	33	171	17
Virus	446	21	73	16

When comparing proteomic signatures, most reliable results are obtained when considering as many projects as possible. PROTEOMAS' ability to process them in a harmonized and automated manner permits to deal with such a high number of datasets. The ever-increasing number of omic studies being publicly available will allow to develop an increasing understanding of the biological alterations caused by studied traits.

Discussion

New methodological developments to contribute to the advance of AOPs are crucial in establishing reliable alternative methods for toxicology in line with the 3R-principles of reducing, refining or replacing animal testing. Omic techniques are very promising methods in this regard, precisely due to the potential to provide plenty of information on the MoA of evaluated substances. Currently, most omic-based approaches to unravel toxicity mechanisms rely on transcriptomics [26–29]. Transcriptomics has already proven its potential, e.g. by its contribution to the lung fibrosis AOP [30, 31] and by its involvement in the Genomic Allergen Rapid Detection (GARD) approach for skin sensitization [32, 33].

Proteomics, on the other hand, has the potential to be more descriptive of the adverse outcome, since this method can closer describe the phenotype than transcriptomics. The drawback of proteomics is the heterogeneity among proteomic datasets due to the high variability of methods and instruments used to generate the data. As opposed to microarray- or RNAseq-based transcriptomics, proteomic output does not necessarily contain information on the same set of molecules, i.e. includes a larger number of missing values, which makes comparison of different experiments more challenging.

Here, we present an automated workflow to process proteomic data which allows analysis in a high-throughput manner without subjective bias. Although the workflow can be used to process data from a single dataset as well, its main benefit lies in the possibility of processing a large number of them, for example those found in public repositories. Once a collection of datasets is retrieved from the repository, PROTEOMAS allows in a simple and harmonized way, to process the datasets in a sequential manner.

For each dataset, the workflow identifies a group of proteins that appear altered among evaluated conditions within the dataset, and assigns a series of descriptors, like protein IDs, GO terms, KEGG, HALLMARK and Reactome pathways, among others, altogether highlighting the proteomic signature of each particular dataset, which can be linked to relevant biological changes and by this to phenotypic differences. Such a systematic

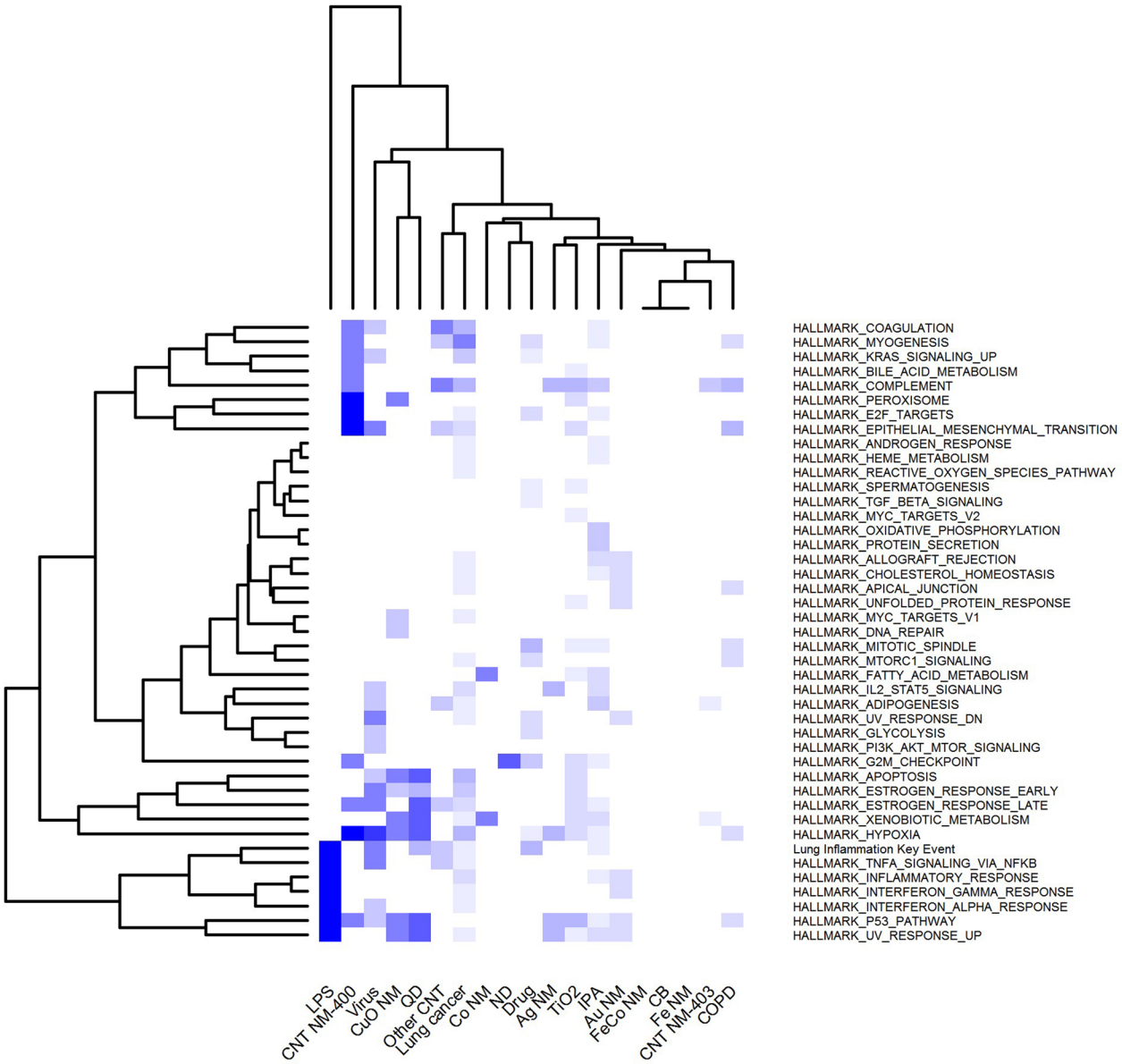


Fig. 11 Heatmap comparing HALLMARK pathways across different traits. Each row corresponds to a certain HALLMARK pathway. Different grades of blue indicate the degree to which the corresponding pathway is altered for that trait. Clustering was performed using Euclidean distance and average linkage. The results of the clustering between projects are represented in the column dendrogram; the clustering between HALLMARK pathways across projects is shown in the row dendrogram

and harmonized data analysis allows the comparison of results from many different proteomic projects, i.e. by mapping their proteomic signatures. Additionally, it contributes to the reuse of proteomic data, which can be then more easily integrated to the outcome of other omic techniques, like the wealth of transcriptomic and metabolomic results already publicly available.

In the present work, we show that original results from the projects taken from the repository vary only minimally from those originated by PROTEOMAS. In the framework of a case study, we proved the utility of PROTEOMAS for comparing NM-related proteomic data with other lung-related studies. Since this workflow is versatile in processing a large amount of proteomic datasets, we could easily compare the proteomic signature from different NM treatments, to those of other various effectors. A special emphasis was put to inflammation as key event attempting to contribute to the

development of AOPs. The same procedure however, could easily be followed to investigate the effect of other substances on other organs. Results can advance directly the development of AOPs and the understanding of the MoA.

In parallel, our workflow aims at facilitating the application of artificial intelligence strategies to describe the effect of evaluated treatments, thus contributing to make proteomic data analysis more FAIR [34, 35]. Simultaneously, our workflow was developed to comply with the DAPRM and DARM of the Omics Reporting Frameworks (TRF and MRF) by the OECD, in order to increase the transparency of proteomic data analysis for regulatory purposes.

Conclusion

Hazard assessment of chemicals relies mostly on very expensive and time consuming in vivo experiments. The high number of substances which are placed in the market requires the development of alternative methods. However, their adequacy depends on the deep understanding of the substance's mechanistic effects. Omic studies are extremely useful to provide the required mechanistic knowledge, since they provide a comprehensive description of caused alterations at different molecular levels. However, they are not yet considered as routine methods in regulatory assessments due to the lack of standardization of the computational analysis of the datasets. Workflows for harmonization of the analysis of omic data contribute directly to facilitate the use of omics in regulatory decision making. Most of the efforts in this regard have been made in the field of transcriptomics. Proteomic experiments on the other hand, besides being more descriptive of the phenotype, are not performed in a manner that allows straightforward comparison of results, because the experimental setup and measuring methods do not belong to established platforms, as for transcriptomics. To address this challenge, in this work we introduced a workflow called PROTEOMAS for harmonized proteomic data analysis, precisely intended to facilitate the use of omics in regulatory decision making. Thus the main utility of our workflow is that it can perform meta-analysis of proteomic data from public origin, allowing the comparison of results from different experimental sources, while increasing the transparency of the analysis. Additionally, it is in agreement with Omics Reporting Framework guidelines of the OECD to integrate proteomics to other omic methods used in regulatory toxicology.

In this work to show the robustness and reliability of PROTEOMAS, we run our workflow on 25 different datasets from public origins and obtained comparable results with the source publications. Additionally, we developed a case study, where we performed a meta-analysis to study the toxicological effect of nanomaterials at the lung level, with a particular focus set on inflammation. Altogether, PROTEOMAS is a contribution to the development of alternative test strategies by facilitating the integration of proteomic experiments, while committing to the FAIR principles (Findable, Accessible, Interoperable and Reusable) of computational protocols.

Methods

Workflow characteristics

PROTEOMAS is a workflow for efficient processing of MS-based proteomic datasets in a high throughput manner. The workflow is fully automated and implemented in Python (version 3.5) and R (version 4.1.0) in a platform-independent manner (usable under Windows, Linux and MAC). In addition, it can be applied on any dataset, either publicly available or de novo generated by an LFQ approach, which includes multiple replicates for each condition or treatment ($n \geq 3$). The corresponding code can be found under https://github.com/AileenBahl/PROTEOMAS.

In brief, PROTEOMAS starts from MaxQuant output files and performs a series of statistical steps, which are explained in more detail in the Results section. The workflow starts with typical data processing steps like filtering, transformation, normalization, imputation and outlier removal. Subsequently, proteins which are significantly altered among conditions are identified. Protein set enrichment analysis is used to identified enriched KEGG, Reactome and HALLMARK pathways as well as GO terms. A flowchart (Fig. 1) summarizing the workflow steps was created using the yEd tool https://www.yworks.com/products/yed.

Obtaining input data

PROTEOMAS can be used to analyze the user's own as well as public proteomic datasets. Public datasets may be retrieved from the PRIDE [36] (PRoteomics IDEntifications) Archive, which is a public data repository of MS-based proteomic data (https://www.ebi.ac.uk/pride/archive). The PRIDE Archive includes currently over 20.000 (state November 2022) projects and this number is rapidly increasing. From PRIDE, the user may download raw files for each project of interest and subject them to a MaxQuant analysis. Instead, for many projects MaxQuant output files are available on PRIDE along the corresponding raw data which can be used directly as input for the workflow.

Raw data analysis

In case only raw data is available for a project of interest, MaxQuant has to be run before PROTEOMAS.

Table 3 Selection of requirements of the OECD Transcriptomics Reporting Framework which are relevant for PROTEOMAS

Task	Required information
Normalization	- Normalization method - Background data subtraction - Method of background calculation - Weighting procedure - Log transformation - Data trimmed? - Control samples removed before normalization? - Formulas - Link/repository/accession number for deposited normalized data + format + description of raw data tables
Data filtering	- Low signal intensities - High variability between technical replicates - Which methods? - Which cut-offs?
Outlier removal	- Method for identification and thresholds - Exclusion at which processing step - List of samples excluded and per sample Justification - Removal before or after normalization and Justification
Discovery of differentially abundant molecules (DAMs)	- Name and version of software - Operating system - Name and version of additional libraries - Availability of software, hyperlinks or source codes - Table of all contrasts / conditions compared for DAM identification - Table of number of samples in each group for DAM identification - Identification of samples with expected covariances (due to shared conditions during processing) - Identification of technical replicates - Name and description of statistical approach - Data transformation performed - For effects models: Specification of effects models used and effects that were modelled - For pairwise comparison approaches: specification of test and values (any transformation or adjustment) being used - Specification of decision criteria (nominal alpha value, p-value threshold, multiple testing correction method, adjusted threshold value, log fold-change cut-off level) including exact order of operations - Output and supporting files according to the file manifest, list all files including a description, describe rows and columns of tables, analysis scripts, software configurations or tables of metadata

MaxQuant [14] is an established proteomics software, which is primarily used for protein identification and quantification, using algorithms specifically developed for the analysis of high-resolution quantitative MS data. It performs data integration and statistical validation for protein inference by using false discovery rates (FDR). MaxQuant output files are tables of the detected peptides, proteins and protein groups. MaxQuant (version 1.6.14) was used in this work to process raw MS-based proteomics files by searching either against human, rat or mouse Uniprot databases (State: March 2021), respectively. The false discovery rate was set to 1% (default value). For advanced protein identification, the 'Match between runs' parameter was enabled. Protein normalization and quantification was done in MaxQuant by applying the LFQ parameter, in which the minimum number of unique peptides was set to 1. The workflow's input is the 'proteinGroups.txt' output file generated by MaxQuant analysis, which contains the identified protein groups, all-, razor- and unique peptides, as well as LFQ intensities. Normalized LFQ intensities generated by MaxQuant are exported from the 'proteinGroups.txt' file and used for further analysis.

Statistical analysis

All data cleaning, transformation and filtering steps were performed using basic Python (version 3.8), as well as some standard additional packages like pandas and numpy. In addition, for the statistical analysis some more advanced R packages (R version 4.1.0) were embedded into the Python code using the rpy2 package. For the imputation of values missing completely at random (MCAR) we used the R package missForest [20] with the number of trees set to 30 and the maximum number of iterations set to 3. In addition, the PCAGrid [37, 38] method from the rrcov package is used to automatically detect outlier samples. All arguments were set to default. In case of insufficient metadata, the NbClust [21] package is used for prediction of the optimal number of groups (k) and group assignment using k-means algorithm. Linear modeling was performed using the lm() function from R and false discovery rates (FDRs) are computed using Python's statsmodels.stats.multitest package. This results in a list of significantly altered proteins for each analyzed dataset with cut-offs set to FDR < 0.05 and log ratio of abundances $\geq \log2(1.5)$.

Protein set enrichment analysis

Protein set enrichment analysis is a method used for the biological interpretation of the obtained sets of proteins with significantly altered abundances. Different databases or ontologies can be used to this end. In this work, protein enrichment analysis was performed

using the R-package 'fgsea' [39]. Background sets were obtained from the Human Molecular Signatures Database (MSigDB) [40] (version 2022, human). Uniprot IDs of the analyzed proteomic datasets were mapped to gene names using the Uniprot.ws package from R. Mouse and rat gene names were mapped to human ones using the msigdbr() package. After these id transformations, results from the proteomic experiments are ready to be compared to the background databases.

Different databases are used to obtain information on enriched gene sets. Kyoto Encyclopedia of Genes and Genomes or shortly **KEGG** (www.kegg.jp/kegg/pathway.html) is a bioinformatics database resource for understanding biological and cellular functions as well as biological pathways from a genomic perspective [41]. The database is online available and can be used to analyze and classify genes into their respective functional pathways, which are a collection of reference maps that correspond to a known functional or biological network. The 'KEGG PATHWAY' category represents pathway maps in various types of molecular networks, such as reaction and interaction networks for metabolism, cellular processes networks, disrupted reaction and interaction networks of human diseases, as well as chemical structure transformation networks for drug development. Similarly, the **REACTOME** database (https://reactome.org/) [42] contains manually curated pathways describing various molecular processes. For REACTOME gene set in the MSigdb, the original REACTOME pathways have been filtered to remove redundancy between the different sets. The Gene Ontology (**GO**) knowledgebase (www.geneontology.org) describes biological information based on three main layers: biological process (BP), cellular component (CC) and molecular function (MF). **HALLMARK** gene sets represent a collection of well-defined biological states or processes which show coherent expression [40].

In addition to the established databases and gene sets described above, we specifically created the so-called "Lung Inflammation Key Event" gene set. This set incorporates 266 genes that are known to be regulated in lungs undergoing inflammation. These genes were extracted from 35 papers addressing explicitly this topic. The list of included genes, as well as the citation to the original articles, is given in Additional file 1: Table S1. With this gene set we aim at comprehensively describing the important key event of inflammation which is present in many different AOPs, and particularly in the AOP for lung fibrosis.

The Python script created in this work obtains enrichment scores for all of the aforementioned databases. All significantly enriched terms having a FDR less than 0.05 are collected in a single file, which includes the category (e.g. Process, Function, KEGG), the term (e.g. GO identifier), the description, as well as the p-value and the FDR values for each enriched term. The enriched terms generated by PROTEOMAS were used for data interpretation. Within this work, we mainly concentrated on the interpretation of HALLMARK pathways.

Datasets and application of the workflow

We randomly selected publicly available proteomic datasets from the PRIDE Archive repository, which originate from studies on lung alterations. We focused on pulmonary alterations because we intend to investigate the inhalative toxicological effects of NM in future studies. First, we prioritize cancerogeneous effects induced by NM. Therefore, we compiled a collection of 25 proteomic datasets (Table 1) to generate a preliminary map of lung alterations, eight of which are related to lung cancer and lung cancer treatments. The other projects cover different pulmonary traits as a background set of alteration as well as five studies on NM treatments. All studies were analyzed in an automated manner by the PROTEOMAS workflow. Venn diagrams comparing original findings against those obtained from PROTEOMAS were generated using R's VennDiagram package.

Report generation according to requirements of the OECD transcriptomics reporting framework

During data evaluation PROTEOMAS automatically generates a report summarizing relevant information on the data analysis. The recorded information is in line with the requirements laid down in the transcriptomics reporting framework of the OECD. The requirements are summarized in Table 3.

Abbreviations

AOP	Adverse outcome pathways
DAM	Discovery of differentially abundant molecules
DAPRM	Data acquisition and processing reporting module
DARM	Data analysis reporting module
DDA	Data-dependent acquisition
FAIR	Findable, accessible, interoperable and reusable
FDR	False discovery rate
GARD	Genomic allergen rapid detection
GSEA	Gene set enrichment analysis
kNN	K-nearest neighbors
LFQ	Label-free quantification
LOD	Limit of detection
MCAR	Missing completely at random
MoA	Mode-of-action
MRF	Metabolomic reporting frameworks
MS	Mass spectrometry
OECD	Organisation for Economic Cooperation and Development
TRF	Transcriptomic reporting frameworks

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-023-00710-2.

Additional file 1. Supplemental Table: Proteins and genes known to be regulated in lungs under inflammation conditions. It was built from 35 publications addressing explicitly this subject. The list includes 266 proteins and genes, as well as the citation to the original articles.

Acknowledgements
AB and VID are very grateful to Dr. Mario Pink for fruitful discussions in the field of nanomaterial risk assessment.

Author contributions
AB developed the research idea, contributed to methods and interpretation of the data, implemented the PROTEOMAS workflow, analyzed the data and wrote the manuscript. CI contributed to methods and interpretation of the data, created an initial version of the workflow, assembled the datasets for the case study and commented the manuscript. KP contributed to methods and interpretation of the data and commented the manuscript. AH provided scientific supervision and revised the manuscript. JD provided scientific insights from the proteomic field. PN provided scientific insights from the transcriptomic field. VID developed the research idea, led the scientific supervision and wrote the manuscript. All authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This work was supported by the EU H2020 project NanoInformaTIX (Grant agreement No 814426) and by the BfR *Sonderforschungsprojekt* 1322–777. JD is thankful for support by the Swiss National Science Foundation, NRP79 "Advancing 3R—Animals, Research and Society" (Grant # 407940), by the canton and the University of Fribourg, and by the SKINTEGRITY.CH collaborative research initiative. PN acknowledges funding from the Swedish Fund for Research Without Animal Experiments (Grant number F2021-0005).

Availability of data and materials
ALL raw data can be found on the PRIDE archive under the following accessions: PXD007223 [23], PXD000861 [43], PXD018895 [44], PXD000853 [22], PXD005698 [45], PXD005733 [46], PXD007137 [47], PXD004818 [48], PXD007180 [49, 50], PXD020470 [51], PXD021685 [52], PXD007148 [50], PXD016664, PXD014022 [24], PXD005834 [53], PXD018569 [54], PXD023041 [55], PXD013244 [56], PXD016148 [57], PXD019267 [58], PXD018900 [59], PXD005970 [60], PXD025423 [61]. To support FAIR principles, the processed datasets and results from downstream analyses can be accessed via https://github.com/AileenBahl/PROTEOMAS/AnalysisResults. All PROTEOMAS scripts can be found on GitHub under the following link: https://github.com/AileenBahl/PROTEOMAS.

Declarations

Competing interests
The authors declare no competing of interests.

Received: 23 December 2022 Accepted: 13 March 2023
Published online: 19 March 2023

References

1. OECD (2018) Users' Handbook supplement to the Guidance Document for developing and assessing Adverse Outcome Pathways. https://doi.org/10.1787/5jlv1m9d1g32-en
2. Halappanavar S et al (2020) Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. Part Fibre Toxicol 17:16. https://doi.org/10.1186/s12989-020-00344-4
3. Villeneuve DL et al (2014) Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142:312–320. https://doi.org/10.1093/toxsci/kfu199
4. Krewski D et al (2014) A framework for the next generation of risk science. Environ Health Perspect 122:796–805. https://doi.org/10.1289/ehp.1307260
5. Stagljar I (2016) The power of OMICs. Biochem Biophys Res Commun 479:607–609. https://doi.org/10.1016/j.bbrc.2016.09.095
6. Ayers D, Day PJ (2015) Systems medicine: the application of systems biology approaches for modern medical research and drug development. Mol Biol Int. https://doi.org/10.1155/2015/698169
7. Simões T, Novais S, Natal-da-Luz T, Devreese B, de Boer T, Roelofs D, Sousa JP, van Straalen NM, Lemos MFL (2018) An integrative omics approach to unravel toxicity mechanisms of environmental chemicals: effects of a formulated herbicide. Sci Rep. https://doi.org/10.1038/s41598-018-29662-6
8. Sauer UG et al (2017) The challenge of the application of 'omics technologies in chemicals risk assessment: background and outlook. Regul Toxicol Pharmacol 91:14–26. https://doi.org/10.1016/j.yrtph.2017.09.020
9. Harrill JA et al (2021) Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. Regulatory Toxicol Pharmacol 125:105020. https://doi.org/10.1016/j.yrtph.2021.105020
10. Saarimäki LA, Melagraki G, Afantitis A, Lynch I, Greco D (2022) Prospects and challenges for FAIR toxicogenomics data. Nat Nanotechnol 17:17–18. https://doi.org/10.1038/s41565-021-01049-1
11. OECD (2021)Transcriptomic Reporting Framework (TRF). www.oecd.org/chemicalsafety/testing/transcriptomic-reporting-framework.pdf
12. OECD (2021) Metabolomic Reporting Framework (MRF). www.oecd.org/chemicalsafety/testing/metabolomics-reporting-framework.pdf
13. Jeliazkova N et al (2021) Towards FAIR nanosafety data. Nat Nanotechnol 16:644–654. https://doi.org/10.1038/s41565-021-00911-6
14. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. https://doi.org/10.1038/nbt.1511
15. Cox J et al (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526. https://doi.org/10.1074/mcp.M113.031591
16. Karpievitch YV, Dabney AR, Smith RD (2012) Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinformatics 13:S5. https://doi.org/10.1186/1471-2105-13-S16-S5
17. Jin L et al (2021) A comparative study of evaluating missing value imputation methods in label-free proteomics. Sci Rep 11:1760. https://doi.org/10.1038/s41598-021-81279-4
18. Lazar C, Gatto L, Ferro M, Bruley C, Burger T (2016) Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. J Proteome Res 15:1116–1125. https://doi.org/10.1021/acs.jproteome.5b00981
19. Troyanskaya O et al (2001) Missing value estimation methods for DNA microarrays. Bioinformatics 17:520–525. https://doi.org/10.1093/bioinformatics/17.6.520
20. Stekhoven DJ, Bühlmann P (2011) MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28:112–118. https://doi.org/10.1093/bioinformatics/btr597
21. Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: an r package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36. https://doi.org/10.18637/jss.v061.i06
22. Li L et al (2014) Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. Nat Commun 5:5469. https://doi.org/10.1038/ncomms6469
23. Tomin T et al (2018) Deletion of adipose triglyceride lipase links triacylglycerol accumulation to a more-aggressive phenotype in A549 Lung carcinoma cells. J Proteome Res 17:1415–1425. https://doi.org/10.1021/acs.jproteome.7b00782
24. Margalit A, Kavanagh K, Carolan JC (2020) Characterization of the proteomic response of A549 cells following sequential exposure to *Aspergillus fumigatus* and *Pseudomonas aeruginosa*. J Proteome Res 19:279–291. https://doi.org/10.1021/acs.jproteome.9b00520

25. European Union (2011) EU 2011/696 Commission Recommendation on the definition of nanomaterial. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011H0696&from=EN.
26. Nikota J et al (2016) Meta-analysis of transcriptomic responses as a means to identify pulmonary disease outcomes for engineered nanomaterials. Part Fibre Toxicol 13:25. https://doi.org/10.1186/s12989-016-0137-5
27. Kohonen P et al (2017) A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. Nat Commun 8:15932. https://doi.org/10.1038/ncomms15932
28. Serra A et al (2019) INSIdE NANO: a systems biology framework to contextualize the mechanism-of-action of engineered nanomaterials. Sci Rep 9:179. https://doi.org/10.1038/s41598-018-37411-y
29. Saarimäki LA et al (2021) Manually curated transcriptomics data collection for toxicogenomic assessment of engineered nanomaterials. Sci Data 8:49. https://doi.org/10.1038/s41597-021-00808-y
30. Labib S et al (2016) Nano-risk science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. Part Fibre Toxicol 13:15–15. https://doi.org/10.1186/s12989-016-0125-9
31. Jagiello K et al (2021) Transcriptomics-based and AOP-informed structure-activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. Small 17:2003465. https://doi.org/10.1002/smll.202003465
32. Johansson H, Albrekt A-S, Borrebaeck CAK, Lindstedt M (2013) The GARD assay for assessment of chemical skin sensitizers. Toxicol In Vitro 27:1163–1169. https://doi.org/10.1016/j.tiv.2012.05.019
33. Zeller KS et al (2017) The GARD platform for potency assessment of skin sensitizing chemicals. Altex 34:539–559. https://doi.org/10.14573/altex.1701101
34. Wilkinson MD et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18
35. Jeliazkova N et al (2021) Towards FAIR nanosafety data. Nat Nanotechnol 16:644–654. https://doi.org/10.1038/s41565-021-00911-6
36. Perez-Riverol Y et al (2018) The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47:D442–D450. https://doi.org/10.1093/nar/gky1106
37. Croux C, Filzmoser P, Oliveira MR (2007) Algorithms for projection-pursuit robust principal component analysis. Chemom Intell Lab Syst 87:218–225. https://doi.org/10.1016/j.chemolab.2007.01.004
38. Todorov V, Filzmoser P (2009) An object-oriented framework for robust multivariate analysis. J Stat Softw 32:1–47. https://doi.org/10.18637/jss.v032.i03
39. Korotkevich G et al (2021) Fast gene set enrichment analysis. bioRxiv. https://doi.org/10.1101/060012
40. Liberzon A et al (2015) The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 1:417–425. https://doi.org/10.1016/j.cels.2015.12.004
41. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27
42. Gillespie M et al (2022) The reactome pathway knowledgebase 2022. Nucleic Acids Res 50:D687-d692. https://doi.org/10.1093/nar/gkab1028
43. Wang Y-T et al (2017) Phosphoproteomics reveals HMGA1, a CK2 substrate, as a drug-resistant target in non-small cell lung cancer. Sci Rep 7:44021. https://doi.org/10.1038/srep44021
44. Poschmann G, Brenig K, Lenz T, Stühler K (2021) Comparative secretomics gives access to high confident secretome data: evaluation of different methods for the determination of bona fide secreted proteins. Proteomics 21:e2000178. https://doi.org/10.1002/pmic.202000178
45. Wiredja DD et al (2017) Phosphoproteomics profiling of nonsmall cell lung cancer cells treated with a novel phosphatase activator. Proteomics. https://doi.org/10.1002/pmic.201700214
46. Stewart PA et al (2017) Relative protein quantification and accessible biology in lung tumor proteomes from four LC-MS/MS discovery platforms. Proteomics. https://doi.org/10.1002/pmic.201600300
47. Kuenzi BM et al (2017) Polypharmacology-based ceritinib repurposing using integrated functional proteomics. Nat Chem Biol 13:1222–1231. https://doi.org/10.1038/nchembio.2489
48. Großerueschkamp F et al (2017) Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics. Sci Rep 7:44829
49. Mossina A et al (2017) Cigarette smoke alters the secretome of lung epithelial cells. Proteomics. https://doi.org/10.1002/pmic.201600243
50. Kammerl IE et al (2019) Dissecting the molecular effects of cigarette smoke on proteasome function. J Proteomics 193:1–9. https://doi.org/10.1016/j.jprot.2018.12.015
51. Wang P et al (2020) A cross-talk between epithelium and endothelium mediates human alveolar-capillary injury during SARS-CoV-2 infection. Cell Death Dis 11:1042–1042. https://doi.org/10.1038/s41419-020-03252-9
52. Dalskov L et al (2020) SARS-CoV-2 evades immune detection in alveolar macrophages. EMBO Rep 21:e51252–e51252. https://doi.org/10.15252/embr.202051252
53. Seddigh P et al (2017) Quantitative analysis of proteome modulations in alveolar epithelial type II cells in response to pulmonary *Aspergillus fumigatus* infection. Mol Celi Proteom 16:2184–2198. https://doi.org/10.1074/mcp.ra117.000072
54. Ruprecht B et al (2020) A mass spectrometry-based proteome map of drug action in lung cancer cell lines. Nat Chem Biol 16:1111–1119. https://doi.org/10.1038/s41589-020-0572-3
55. Deliyannis G et al (2021) TLR2-mediated activation of innate responses in the upper airways confers antiviral protection of the lungs. JCI Insight. https://doi.org/10.1172/jci.insight.140267
56. Xing Q-Q et al (2019) Serum proteomics analysis based on label-free revealed the protective effect of Chinese herbal formula Gu-Ben-Fang-Xiao. Biomed Pharmacother 119:109390. https://doi.org/10.1016/j.biopha.2019.109390
57. Billing AM et al (2020) Fast and robust proteome screening platform identifies neutrophil extracellular trap formation in the lung in response to cobalt ferrite nanoparticles. ACS Nano 14:4096–4110. https://doi.org/10.1021/acsnano.9b08818
58. Gallud A et al (2020) Multiparametric profiling of engineered nanomaterials: unmasking the surface coating effect. Adv Sci 7:2002221
59. Seidel C et al (2021) Inhaled multi-walled carbon nanotubes differently modulate global gene and protein expression in rat lungs. Nanotoxicology 15:238–256. https://doi.org/10.1080/17435390.2020.1851418
60. Phuyal S et al (2018) Characterization of the proteome and lipidome profiles of human lung cells after low dose and chronic exposure to multiwalled carbon nanotubes. Nanotoxicology 12:138–152. https://doi.org/10.1080/17435390.2018.1425500
61. Alswady-Hoff M et al (2021) Long-term exposure to nanosized TiO2 triggers stress responses and cell death pathways in pulmonary epithelial cells. Int J Mol Sci. https://doi.org/10.3390/ijms22105349

Publisher's Note

3.4 Bioinformatics and machine learning to support nanomaterial grouping

Aileen Bahl[1,2,3,*], Sabina Halappanavar[4], Wendel Wohlleben[5], Penny Nymark[6], Pekka Kohonen[6], Håkan Wallin[7,8], Ulla Vogel[9], Andrea Haase[1,3,*]

1 German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany

2 German Federal Institute for Risk Assessment (BfR), Department of Biological Safety, Berlin, Germany

3 Freie Universität Berlin, Institute of Pharmacy, Berlin, Germany

4 Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada

5 BASF SE, Dept. Analytical and Material Science and Dept. Experimental Toxicology and Ecology, Ludwigshafen, Germany

6 Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

7 National Institute of Occupational Health, Dept. of chemical and biological risk factors, Oslo, Norway

8 Department of Public Health, Copenhagen University, Copenhagen, Denmark

9 National Research Centre for the Working Environment, Copenhagen, Denmark

This chapter was submitted as a review article to *Nanotoxicology*.

Bioinformatics and machine learning to support nanomaterial grouping

Aileen Bahl[1,2,3], Sabina Halappanavar[4], Wendel Wohlleben[5], Penny Nymark[6], Pekka Kohonen[6], Håkan Wallin[7,8], Ulla Vogel[9], Andrea Haase[1,3]

1 German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany

2 German Federal Institute for Risk Assessment (BfR), Department of Biological Safety, Berlin, Germany

3 Freie Universität Berlin, Institute of Pharmacy, Berlin, Germany

4 Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada

5 BASF SE, Dept. Analytical and Material Science and Dept. Experimental Toxicology and Ecology, Ludwigshafen, Germany

6 Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

7 National Institute of Occupational Health, Dept. of chemical and biological risk factors, Oslo, Norway

8 Department of Public Health, Copenhagen University, Copenhagen, Denmark

9 National Research Centre for the Working Environment, Copenhagen, Denmark

Corresponding authors:

Aileen Bahl

E-Mail: aileen.bahl@bfr.bund.de

Andrea Haase

E-Mail: andrea.haase@bfr.bund.de

1 Abstract

Nanomaterials (NMs) offer plenty of novel functionalities. Moreover, their physicochemical properties can be fine-tuned to meet the needs of specific applications, leading to virtually unlimited numbers of NM variants. Hence, efficient hazard and risk assessment strategies building on new approach methodologies (NAMs) become indispensable. Indeed, the design, the development and implementation of NAMs has been a major topic in a substantial number of research projects. One of the promising strategies that can help dealing with the high number of NMs variants is grouping and read-across. Based on demonstrated structural and physico-chemical similarity, NMs can be grouped and assessed together. Within an established NM group, read-across may be performed to fill in data gaps for data-poor variants using existing data for NMs within the group. Establishing a group requires a sound justification, usually based on a grouping hypothesis that links specific physico-chemical properties to well-defined hazard endpoints. However, for NMs these interrelationships are only beginning to be understood. The aim of this review is to demonstrate the power of bioinformatics with a specific focus on machine learning (ML) approaches to unravel the NM Modes-of-Action (MoA) and identify the properties that are relevant to specific hazards, in support of grouping strategies. This review emphasizes on the following messages: 1) ML supports identification of the most relevant properties contributing to specific hazards; 2) ML supports analysis of large omics datasets and identification of MoA patterns in support of hypothesis formulation in grouping approaches; 3) omics approaches are useful for shifting away from consideration of single endpoints towards a more mechanistic understanding across multiple endpoints gained from one experiment; and 4) approaches from other fields of Artificial Intelligence (AI) like Natural Language Processing or image analysis may support automated extraction and interlinkage of information related to NM toxicity. Here, existing ML models for predicting NM toxicity and for analyzing omics data in support of NM grouping are reviewed.

Keywords:

Nanomaterial Grouping, Machine Learning, Omics, Artificial Intelligence, New Approach Methodologies

2 Introduction

Engineering physicochemical properties of nanomaterials (NMs) such as size, morphology or surface chemistries has become common practice in order to meet the needs of specific applications. This has resulted in a large variety and a steadily increasing number of different NMs or nanoforms (NFs), as defined in specific regulatory frameworks[1]. However, adjusting NM physicochemical properties does not only impact their desired functionalities but the fine tuning also influences their original or expected behavior in biological milieu including, their uptake by cells, biodistribution, dissolution rate and/or toxicity to humans or the environment, and more. To overcome the need to fully characterize each and every NM variant for all possible toxicological outcomes, European chemicals legislation REACH[1] allows for grouping and read-across, to either justify waiving specific tests or to fill in data gaps.

For chemicals, grouping is well established as a 'general approach for considering more than one chemical at the same time'[2,3]. The idea behind grouping approaches is that chemicals which are similar enough with respect to certain criteria (e.g., structural, physico-chemical properties, etc.) can be considered as a group. Chemicals within one group are then expected to show similar (eco-)toxicological and/or environmental fate behaviour. Within this group, data gaps on toxicological behaviour for a certain member of the group can therefore be filled by read-across using information from the other members in the group. In general, grouping may support risk assessment as well as Safe(r)-and-Sustainable-by-Design (SSbD) approaches. Groups are established initially on the basis of structural similarity, which can be based on various principles such as common functional groups, precursors, breakdown products or a constant incremental change of the properties of interest across the group[2,4]. However, it then has to be demonstrated that these structural similarities result in a similar fate and/or (eco-)toxicity. Thus, knowledge of a common toxic mechanism or Mode-of-Action (MoA) can strongly facilitate grouping, since grouping always requires a proper scientific justification which is mainly supported by establishing a link between specific properties and the toxicological endpoint of interest. In addition, grouping is endpoint-specific meaning that group membership may vary depending on which toxicity endpoint is considered.

In the last decade several grouping frameworks have been developed for NMs, e.g. the MARINA grouping and read-across approach[5], the DF4nano Grouping Framework[6], which are comprehensively summarized in Oomen et al.[7] and Giusti et al.[8] The most recent and comprehensive framework is the GRACIOUS[9] grouping framework. Additional insights into NM grouping in the context of the EU chemical legislation are detailed in Mech et al.[10] The most recent GRACIOUS framework is based on a hypothesis-driven approach. It proposes several grouping hypotheses, which link specific physicochemical properties with specific fates and/or toxicities, it tested the hypotheses, in case studies[11-16] and lastly, it has developed guiding principles to support users to formulate their own grouping hypotheses[17].

Nevertheless, grouping of NMs remains a challenge. In particular, unravelling relationships between specific physicochemical properties and toxicities is not trivial due to the large panel of interdependent physicochemical properties that are needed to describe a single NM[18]. The number- based particle size distribution, surface functionalization or treatment, shape or morphology as well as surface area are certainly the most central ones[19,20]. Dissolution rate, state of agglomeration/aggregation and surface reactivity have also been shown to be of high relevance[21,22]. However, plenty of other NM properties

exist that are not tested and, several of the NM properties are polydisperse already after production and additionally have the potential to change during the life cycle, depending on the environment/ biological medium in which they are suspended or incorporated. This renders the physicochemical characterization of NMs both in their dry state and as applied, a complex task. Overall, identifying which physicochemical parameters are driving toxicity remains the key challenge in NM grouping[23,24].

Omics approaches are a very promising tool which have already frequently been applied for chemicals including NMs. Different omics layers can be investigated, e.g., levels of gene transcription (transcriptomics), protein abundances (proteomics) or levels of small molecule metabolites (metabolomics). Among them, transcriptomics is by far the most studied omics level for describing the molecular changes induced by NMs. This is mainly due to the fact that transcriptomics technologies are highly advanced, the evaluation is well standardized and interpretation of the results is relatively straightforward due to the well-studied and well-annotated state of this particular level of biological organization[25-27] and well-established tools. On the other hand, while being closer to the actual phenotype, indicating the potential for more direct causal association with the adverse outcome (AO), i.e., endpoint of interest, proteomics or metabolomics are much complex and do not at present have standardized protocols for the analysis and interpretation of data. The OECD reporting frameworks have been established for both transcriptomics[28] and metabolomics[29]; however, standardized evaluation and interpretation for metabolomics and proteomics is still lagging behind. While for now, it is possible to individually assess and report the single omics end point (transcriptomics or metabolomics or proteomics), their combined evaluation, which may be important to obtain a holistic view of the biological response to an exposure remains difficult. Thus, developing models that are based on omics data is also consistent with the major shift away from pure consideration of single endpoints towards a more mechanistic understanding which can be observed within the field of toxicology. In general, omics approaches yield the advantage that multiple endpoints are considered at the same time as a whole panel of cellular changes is measured in one single experiment, allowing for investigation of dependencies between events and endpoints.

Within NM experiments, usually omics data for a few materials are obtained and analyzed subsequently[30,31]. While this provides useful insights into various biological changes in that specific experiment, it is difficult to derive general patterns and to extract which changes are most relevant in terms of grouping NMs with respect to a certain endpoint. Especially, the high-dimensional setting of omics experiments with the number of parameters being much larger than the number of samples is a major challenge. Here, meta-analyses may be very useful to counterbalance this situation. Due to data sharing rules in the community and corresponding journals, huge databases of publicly available datasets exist for different omics levels like the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO)[32] for transcriptomics or PROteomic IDEntification(PRIDE) database[33] for proteomics. While these datasets are naturally covering a wide range of studied traits like chemicals, drugs or diseases, omics datasets should theoretically allow integration across those traits. Such meta-analyses settings may be beneficial for the detection of patterns as they are based on a greater knowledge base and noise occurring in single studies may be cancelled out.

Machine learning (ML), a subfield of artificial intelligence (AI), may be highly valuable for addressing this task. The greatest advantage of ML is that such models are able to automatically learn patterns in large datasets and derive associated predictions. Generally, data to be modeled is described in a feature vector or matrix. This comprises values for measured input features like physicochemical properties or others. Additionally, a respective outcome variable, describing for instance a certain

toxicity outcome can be linked to the input data. ML algorithms range from classical linear or logistic regression models to more complex ones such as Random Forests (RFs) or Support Vector Machines (SVMs) and Deep Learning models like Neural Networks (NNs). The choice of ML algorithm used in a study should be based on the amount and complexity of the available data as well as the specific goal in order to obtain robust trustworthy results. In an optimal dataset for the development of ML models, the number of samples should be much larger than the number of features describing these samples. However, in biological settings this is rarely ever the case. Therefore, the main goal for implementing useful ML models is to find the right balance between model complexity and generalizability – thus, avoiding overfitting. Model complexity means that a model is complex enough to be able to describe a set of so-called training data, at the same time, be able to predict outcomes for previously unseen test data. Also, there is usually a trade-off between flexibility/power and interpretability/inference capacity of a ML method[34,35]. This means that simplest methods are also the most easily interpretable ones, such as linear regression (and variants such as elastic net regression) or decision trees. More complex methods such as RF and SVMs are significantly less interpretable, and the least interpretable models are neural nets and the multilayered deep neural networks (deep learning models). Explainable AI (XAI) is a subset of AI and ML techniques focused on making AI systems more understandable, transparent, and interpretable for humans. This has implications also for regulatory toxicology, as models should be maximally interpretable for reasons of transparency and governance.

Again, ML are optimally suited to aid the analysis of omics data for several reasons. Omics datasets are high-dimensional and rich in describing many cellular alterations within each single measurement, in which patterns and relationships can be detected and that may not be visible using traditional statistical methods. However, the high dimensionality of the data can also be challenging. Here, moving from datasets limited to single experiment or materials towards integration of various datasets in a meta-analysis setting may be beneficial. In such cases, ML is well-suited to handle the complexity of the datasets, unraveling the hidden interactions or dependencies between molecular components. ML can also be used to automatically select the most relevant features such as transcripts, proteins or metabolites related to specific molecular events or key event (KE), thereby reducing the dimensionality and improving the interpretability, resulting in identification of potential biomarkers. ML can also directly be used for omics-driven predictive modeling. Additionally, meta-analyses are great tools for handling noise in the data and distinguishing true from random signals[36]. ML can support integration of multiple omics layers. Finally, a great advantage of ML methods is that they can easily be adapted by re-training once new data become available. Thus, ML models are well suited for reducing the complexity of analysis of omics datasets, revealing the important biological traits perturbed after exposure. Moving forward, once NMs with similar MoAs have been identified, one may investigate corresponding similarities and differences in physico-chemical properties and elucidate the most relevant properties for formulating a grouping hypothesis, and also to provide design principles for NMs with reduced hazard. Advances in high-throughput transcriptomics facilitate the creation of large and uniform data sets that are ideally suited for ML and grouping applications as the available technologies reduce the cost of transcriptome profiling by up to 10-20-fold[37]. Coupling together omics and high-throughput screening technologies in a tiered approach increases the granularity and informativeness of the data further[38]. XAI may also facilitates MoA discovery as it helps to interpret the decision made by ML models. In the context of transcriptomics-derived MoAs, tools need to be developed, in part, to conform to XAI principles while maximizing predictive ability.

Some recent reviews have summarized ML models in the NM field[9,39-41]. However, they have not or only scarcely considered the potential of omics data in the context of NM grouping. The aim of this review is to provide an overview of existing ML models that enable complex data integration and analysis in support of NM grouping and to shed light on how omics data may be used in conjunction with ML models in the development of reliable grouping frameworks. The review focuses only on hazard endpoints which are considered relevant under REACH (see Table 1) for human toxicity. ML models for other endpoints like cytotoxicity or ecotoxicity as well as those predicting NM uptake or protein corona formation are excluded from this review.

3 Relevant hazard endpoints

Grouping allows waiving of tests or filling the data gaps related to a target substance by using data from a previously tested source substance. In the European Union, the most important overarching regulation for chemicals is REACH. The Annexes VII to X of the REACH regulation[1] specify which information the manufacturers need to provide when registering a new substance. These requirements are dependent on the tonnage of production per year. Thus, higher tonnages lead to more extensive toxicological testing requirements. In order to tackle specific information requirements for NMs, the REACH annexes VII to X have been updated to specifically consider nano-specific information needs[21]. In addition, work is underway to modify or adapt test guidelines in consideration of specific properties and property specific effects of NMs[42].

Apart from REACH, other specific legislations exist for chemicals with specific applications, e.g., cosmetics[43], food contact materials[44] or pesticides[45]. All these regulations consider (to a large extent) similar toxicological endpoints. The relevant endpoints for several legislation for NMs have also been collected in the European Union Observatory for NMs (EUON) report on novel approach methodologies (NAMs, refer to a set of innovative techniques or testing strategies used in toxicology and risk assessment to evaluate the safety of chemicals.) for NMs[46] as well as in a recent publication from Bleeker et al.[47] which intends to support harmonization of the testing requirements for NMs across EU legislations. An overview on relevant hazard classes for human health as described in these documents is given in Table 1.

While for some of the above-mentioned endpoints such as skin corrosion/irritation or serious eye damage/irritation, NAMs are well-established and NAMs data is accepted for regulatory decision making, for other more complex endpoints, NAM development is still on-going. Although a push for replacing animal tests with validated NAMs is in full swing internationally[48] to date, toxicological testing for complex endpoints still largely relies on animal studies.

Table 1: Relevant endpoints for human health under REACH. The endpoints to consider depends on the production amount of the substance in tonnage (tonnage triggered).

Endpoint	Description
Acute toxicity	Adverse effects after single dose, multiple doses given within 24 hours or inhalation exposure of 4 hours
Skin corrosion/irritation	Irreversible (corrosion) or reversible (irritation) damage to the skin
Serious eye damage/irritation	Serious eye damage: tissue damage in the eye, or serious physical decay of vision, which is not fully reversible within 21 days Eye irritation: changes in the eye, which are fully reversible within 21 days
Respiratory or skin sensitization	Respiratory sensitization: hypersensitivity of the airways following inhalation of the substance Skin sensitization: allergic response following skin contact
Mutagenicity / Genotoxicity	Alteration of the structure, information content or segregation of DNA
Carcinogenicity	Induction of cancer or increase of its incidence
Reproductive toxicity	Adverse effects on sexual function and fertility in adult males and females Developmental toxicity in the offspring (before or after birth)
Specific target organ toxicity (STOT) – single exposure	Specific, non-lethal target organ toxicity after single exposure
Specific target organ toxicity (STOT) – repeated exposure	Specific, non-lethal target organ toxicity after repeated exposure
(Developmental) neurotoxicity	Disruption of the nervous system of the individual or *in utero* or early postnatal development
Effects on gut microbiome	Effects on gut microbiome variability and dysbiosis
Endocrine disruption	Adverse effects connected to endocrine system like developmental malformations, disorders of immune and nervous systems functions or increased cancer risk
Hypersensitivity / food intolerance	Adverse reaction to food including or excluding the immune system
(Developmental) immunotoxicity	Adverse effects on the structure and function of the immune system in the individual or its offspring
Phototoxicity	Toxic response after exposure to environmental light

4 NAMs and NAM frameworks

For many chemicals, and especially for NMs, the data for complex endpoints is very scarce and an increasing number of new materials is entering the market every year. Filling the data gaps and testing new materials using animal studies raises not only ethical concerns but is also limited with respect to time and cost efficiency. In addition, human relevance of animal models is frequently questioned and

the underlying toxicity mechanisms are often less obvious in animal models. Therefore, NAMs are becoming increasingly important for safety assessment in the light of the 3R principles for reducing, refining and replacing animal studies. Several NAMs exist and are based on *in vitro*, *in chemico* and *in silico* methods. This also includes high-throughput screening, allowing for testing of multiple chemicals at a time and high-content methods like omics approaches that enable comprehensive understanding of the underlying mechanisms[49].

One major advantage of NAMs is that they allow unraveling toxicity mechanisms which may greatly improve hazard and risk assessment in the future. However, single NAMs are not sufficient to describe an AO. Instead, a battery of NAMs may be required to sufficiently assess an AO. Therefore, NAM frameworks combining multiple individual NAMs are needed.

To date, several NAM frameworks have been developed. Within REACH, one of the most important alternatives relying on NAMs frameworks is the concept of grouping and read-across. For chemicals in general, grouping is already well established and frequently used in the regulatory decision making, with structural similarities or common functional groups being some of the key parameters defining similarity. In contrast, for NMs, the situation is much more complex and establishing reliable grouping approaches is not trivial. This stems from the facts that 1) the number of physico-chemical properties needed to sufficiently describe a NM is much large and to date, no simple relationship between any single property and the toxicological outcomes has been consistently observed. This may also be due to the polydispersity of NM (or any other particle) in their properties, such that not all particles in one test preparation exhibit the exactly identical properties; 2) the NM physico-chemical properties change during the lifecycle and/or in different environments; 3) other tasks such as exposure characterisation and dose estimation are an issue; and 4) NMs interfere with some assay components, requiring optimisation of existing methods or development of novel methods. All of this has resulted in inconsistent results and reporting, leading to an inability to generalise the observed results across NMs of similar properties.
Instead of purely relying on read-outs from *in vivo* or *in vitro* testing, information on common MoAs or toxicity mechanisms may greatly advance NM grouping approaches and can help to justify an existing grouping hypothesis[2]. The knowledge of underlying MoAs can then in turn greatly advance two important concepts: Adverse Outcome Pathways (AOPs) and Integrated Approaches to Testing and Assessment (IATAs).

AOPs (Ankley et al. 2010) are conceptual frameworks which aim to causally link certain biological events in a sequential manner, starting from a molecular initiating event (MIE), inducing multiple KEs and finally leading to an AO. The MIE thereby represents the interaction of a NM with a biomolecule or an event after its first interaction. That is followed by KEs at multiple levels of biological organization that are essential to the disease progression and can be measured, e.g., at the cellular, tissue or organ level. The AO may then be one of the endpoints mentioned in Table 1. Several AOPs have been proposed for chemical induced toxicity and may also be applicable or adaptable to NMs[50-53]. One such example is AOP173 from the AOPwiki (https://aopwiki.org/aops/173) which describes the development of pulmonary fibrosis after substance interaction with pulmonary resident cell membrane components, which is relevant to NMs.

Often AOPs are used as a basis to establish IATAs, which are frameworks for evaluating complex hazard endpoints by integrating multiple sources of information for studying various aspects of the toxicity endpoint under consideration[53,54]. A proper understanding of the underlying MoA and related AOPs is

important for developing reliable IATAs, as they enable identification of the right assays reflecting KEs to be tested. IATAs allow integration of the different assay outcomes and provide information on a potential hazard in a weight-of-evidence manner. Within an IATA, different kinds of NAMs can be combined.

In the following sections, an overview of existing approaches based on ML and omics supporting NM grouping and potentially also other NAM frameworks is given.

5 ML and existing ML models for NMs

ML models are well suited for predicting outcomes with respect to NM toxicity and for selecting the most relevant descriptors influencing toxicity. In general, supervised and unsupervised ML models, as well as some mixed types exist. In supervised models, the goal is to map labels assigned to each sample to variances observed in the input data. In the case of NM grouping, these labels are usually representing a toxicity endpoint, e.g., the outcome of an *in vivo* study or an *in vitro* assay. Depending on the nature of this response variable, ML models can be divided into regression models in which the outcome variable is continuous and classification models with discrete outcome variables. Instead, unsupervised models use unlabeled data and thus, only rely on variances in the input data and seek to find patterns therein. An overview of frequently used ML algorithms[34,35] is given in Figure 1.

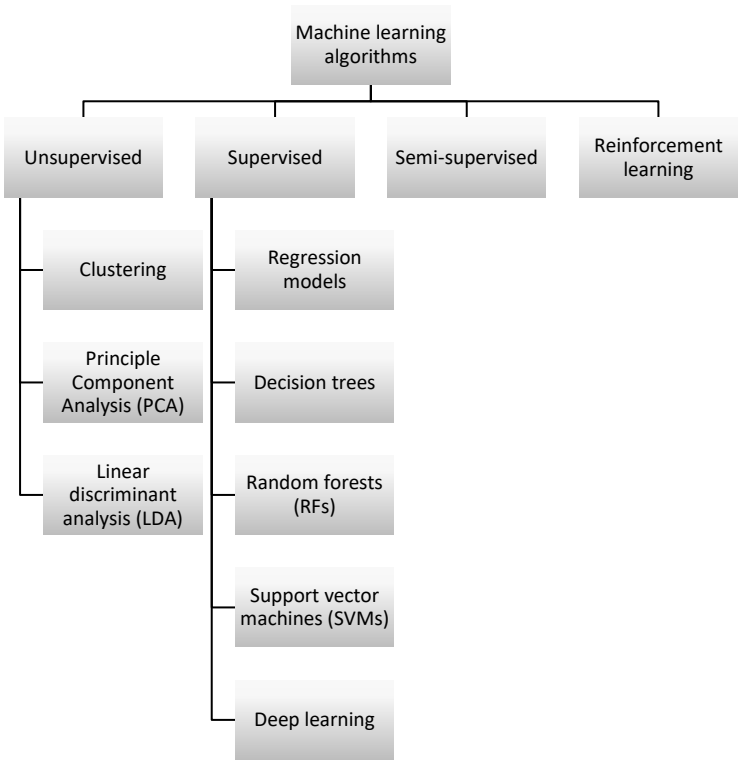


Figure 1: Overview on common ML algorithms.

To identify relevant studies, a search in Scopus using the following search query: TITLE-ABS-KEY ((*nanoparticle* OR *nanomaterial OR nanoparticles* OR *nanomaterials*) AND (*"in silico"* OR *computational* OR *"machine learning"* OR *"case study"*) AND *toxicity*), was conducted. This search matched 988 publications of which 657 publications were tagged as primary publications (Stand: 11/2023). From these publications, those using a ML model to predict one of the relevant endpoints in Table 1 were identified. In addition, relevant studies from previous reviews on ML models in the NM field[9,39-41] were also added. An overview of the relevant approaches is provided in Table 2.

Except for one study, all identified approaches concentrate either on mutagenicity and genotoxicity or on STOT as the modelled endpoint. In addition, almost all models include supervised approaches. Most frequently, tree-based approaches, namely decision trees or RFs, are used to predict toxicity. Often, these are preceded by an unsupervised analysis using hierarchical clustering or Principle Component Analysis (PCA). Overall, the predictive performance of the models was found quite good (0.7 to 1.0). However, it is also easily visible from Table 2 that the number of available datasets is very small in most cases and usually not sufficient to build robust ML models. In addition to predicting toxicity outcomes, many studies also perform feature selection to reduce the model to only the most relevant descriptors. From Table 2, it becomes obvious that the selected descriptors vary largely across studies. Thus, even though most models show relatively high predictive performance, it may be expected that their applicability domain is rather limited. With respect to selected materials, there is a strong focus on metal oxide NMs and multi-walled carbon nanotubes (MWCNTs). Other materials are not sufficiently covered so far and thus, cannot easily be assessed with the available models. As properties by which different material classes and materials of different shapes can be described may differ, integrating various types of NMs in a common model is not straightforward.

This raises the question, whether establishing NM grouping approaches on the level of intrinsic physico-chemical properties is sufficient. Predictive models and grouping approaches based on intrinsic physicochemical properties -describing the chemical and physical structures have two major advantages: 1) they can mostly be controlled directly during the production process and 2) many of the intrinsic properties can be measured more easily compared to extrinsic properties, which often require more complex methodologies, which are not well-standardized. However, intrinsic physico-chemical parameters alone are insufficient to group NMs in a reliable manner. Extrinsic descriptors reflecting the biological activity of NMs are superior compared to approaches based on simple physico-chemical parameters[55] and they need to be derived to separate distinct NM hazard groups sufficiently well. However, grouping of NMs is still not perfect. In addition, intrinsic as well as extrinsic physico-chemical properties suffer from the fact that the applicability domain usually restricts models to very specific subsets of NMs. Due to all these reasons, NM grouping may be viewed as a complex endevour and further efforts are needed to develop effective grouping strategies.

In a more general approach, one may try to group NMs based on a common MoA. This is closely related to the concept of AOPs and the analysis of omics experiments which are described in the next section. Omics measurements may especially be helpful to separate different MoAs and thus solve the difficulties for NM grouping that result from the fact that some AOs in vivo or in vitro may feature different MoAs and in other cases, different MOAs may lead to the same AO.

Table 2: Predictive ML models for NM toxicity and grouping in mammals or mammalian cell models

Modelled endpoint	**Reference**	**Study type**	**Model type**	**Feature selection performed**	**Computational validation performed**	**Study design / dataset**	**Selected descriptors for best model**	**Predictive performance of best model**
Germ cell mutagenicity	Lamon et al., 2018[56]	*In vitro*	Hierarchical clustering, PCA, RF	Gini index	/	6 TiO_2, testing RAAF workflow, read-across for Comet assay	Content of organic matter, total non-TiO_2, biodurability in different media	/
	Aschberger et al., 2019[57]	*In vitro / in vivo* (different rodents)	Hierarchical clustering, PCA	PCA loadings	/	19 MWCNTs, testing RAAF workflow, different genotoxicity assays (Comet, micronucleus) *in vitro* and *in vivo*	Length, SSA, CEA:H, and CEA:N	/
	Murugadoss et al., 2021[58]	*In vitro*	PCA, regression and RF	/	Trainings and test set	TiO_2 case study	Agglomerate size	R^2 = 0.658
	Kotzabasaki et al., 2021[59]	*In vitro / in vivo* (different rodents)	PCA, SVM, RF, Logistic regression, Naïve Bayes	Recursive Feature Elimination	Trainings and test dataset (Kennard-Stone algorithm)	15 MWCNTs from literature	Length, zeta potential, Purity, Polydispersity index	Accuracy = 0.8
	Sizochenko, 2019[60]	*In vitro / in vivo*	Support vector machines	/	/	21 MO NMs, genotoxicity in	Electronegativity of the metal and the charge of its ion	Balanced accuracy = 0.75 for Comet assay

	(Wistar rats)	(SVM), Naïve Bayes, k-nearest neighbors (KNN), and Decision Tree (DT), Self-organizing maps (SOM)			Comet assay and Ames test		and 0.83 for Ames test
Ambure, 2020[61]	*In vitro / in vivo* (different rodents)	LDA, RF	Genetic algorithm (GA) and best subset selection (BSS)	Internal (training set) and external (test set) validation, Y-randomization test and 10-fold cross-validation for RF	7 metal oxides nanoparticles (SiO_2, ZnO, TiO_2, CuO, Fe_2O_3, Fe_3O_4, Al_2O_3,) classification-based multi-tasking (mtk)-QSAR model predicting genotoxicity	Concentration exposed, core size of nanoparticle, exposure time, experimental protocols, and cell lines + 5 calculated descriptors	Accuracy = 0.94
El Yamani et al., 2022[62]	*In vitro*	PLS			seventeen NMs derived from titanium dioxide (TiO_2), zinc oxide (ZnO), silver	H-L energy, Ionization potential, HOMO energy, pristine (TEM) size	

						(Ag) and silica (SiO_2), comet assay with Fpg		
	Halder et al., 2020[63]	*In vitro*	Perturbation theory machine learning (PTML) based QSTR approach		Training and test set + external validation data	78 metal oxide NMs, *in vitro* Comet assay	Number of heavy atoms and the size of the NMs, graph density	Accuracy = 97.81%
Reproductive toxicity	Ban et al., 2018[64]	*In vivo* (different rodents)	RF regression, Similarity network		training data sets (from 20% to 100%) were randomly selected from original data sets, with 10 repetitions	Based on public data for male rodents; carbon nanotubes, and Ag and TiO_2 NPs		R^2 = 0.624, RMSE = 0.198
Specific target organ toxicity (STOT) – single exposure	González-Durruthy et al., 2019[65]	*Ex vivo* (Wistar rats)	LDA and ANN based on linear and non-linear classification algorithms		Training and validation set	Prediction of mitotoxicity in rat kidney, 9 types of CNTs, based on fractal SEM nanodescrip-tors		Specificity = 99.5% and sensitivity = 99.2%
Specific target organ toxicity	Gajewicz et al., 2018[66]	*In vivo*	Decision tree model	Gini importance	Training and test sets, multiple splits	19 NM, NOAEC from STIS, protein	Size, surface area, presence of a coating	Balanced accuracy = 0.8 for protein

(STOT) – repeated exposure						carbonylation and IOP (FRAS assay)		carbonylation and 1.0 for IOP
	Bahl et al., 2019[67]	*In vitro / in vivo* (Wistar rats)	PCA with kNN, RF	RFE	LOOCV	11 NM (mainly SiO_2), predicted outcomes: Macrophage assay and STIS, classification into active and passive	Zeta potential, redox potential and dissolution rate	Balanced accuracy = 0.82
	Bahl et al., 2020[68]	*In vitro / in vivo* (Wistar rats)	Prediction model based on multivariate logistic regression with Firth's bias reduction method	Greedy search	LOOCV	14 NM from various material classes, categorization based on surface reactivity, predicted outcomes: Macrophage assay and STIS, classification into active and passive	Carbonylation assay potentially in combination with either FRAS or ESR with DMPO	Balanced accuracy = 0.89 for the macrophage assay and 0.83 for STIS

Marvin et al., 2017[69]	*In vitro* / *in vivo* (different rodents)	Bayesian network (network structure based on expert judgement)	/	Validated with independent data	Metal- and metal-oxide nanomaterials, based on international expert consultation and the scientific literature (e.g., *in vitro* / *in vivo* data)	Elemental composition, surface coating, surface area, aggregation and particle size	Prediction accuracy: 0.72 for hazard potential and 0.71 for biological effect
Sheehan et al., 2018[70]	*In vitro* / *in vivo* (different rodents)	Bayesian networks (combined with weight of evidence with MCDA (multi-criteria decision analysis) methodology)	/	Validated with independent data	See Marvin et al., 2017[75]	See Marvin et al., 2017[75]	Prediction accuracy = 0.67
Drew et al., 2017[71]	*In vivo* (different rodents)	Dose-response-modeling, hierarchical clustering, Random Forest	Mean squared error	External validation set	Data collected from 25 *in vivo* studies, various materials, pulmonary inflammation	Density, surface area, and diameter	Balanced accuracy = 0.8

		regression (for initial potency class assignment) and RF classification (for prediction)			in rodents, materials classified into four potency groups		
Furxhi et al., 2020[72]	*In vitro*	RF	Information gain analysis	k-fold cross validation and external validation set	Neurotoxicity prediction	Exposure dose and duration, toxicological assay, cell type, zeta potential	Accuracy: 0.98
Gernand and Casman, 2014[73]	*In vivo* (different rodents)	Regression trees and RF	Mean variance reduction	10-fold cross-validation	Meta-analysis across 17 CNTs, polymorpho-nuclear neutrophils (PMNs), macrophages (MAC), lactate dehydrogenase (LDH), and total protein (TP) modelled separately; reflect	PMN + macrophages: median diameter, mass mode aerodynamic diameter (MMAD), and cobalt content LDH: short median length, cobalt content TP: geometric variables, median length and median diameter, and MMAD	R^2 between 0.83 for neutrophils and 0.95 for total protein

					immune response and cell membrane damage and death		
Gernand and Casman, 2016[74]	*In vivo* (different rodents)	RF, Monte Carlo resampling technique	Mean variance reduction	Out-of-bag error used	Meta-analysis of pulmonary nanoparticle toxicity, concentration of LDH or the number of PMNs in bronchoalveolar lavage (BAL) fluid; CNT, TiO_2, SiO_2, ZnO, MnO	Varies with modelled material classes, total mass more important than all physico-chemical parameters	$R^2 = 0.97$

6 Omics approaches revealing NM MoA

As described before, omics data have the potential to support NM grouping approaches by informing about underlying MoAs and induced AOPs. Thus, this review also focuses on predictive models for NM toxicity which include omics data as descriptors as well as other omics approaches which may potentially support NM grouping. Therefore, all omics-based tools and approaches were searched and below, important primary NM omics studies are described.

Several studies were performed to describe the changes induced by NMs on the level of transcriptomics. Various *in vitro*[75-77] as well as *in vivo*[78-83] approaches have been described. Although scarcer, literature on the effects of NM treatment on other omics layers also exists. For proteomics, mass spectrometry gives the most comprehensive results and is therefore frequently used nowadays[84-87]. Also, metabolomics changes are addressed in multiple studies[88-90]. In addition, some approaches considered multiple omics layers at the same time[91-96]. While this is not a comprehensive list of available studies, it is already clear, that these several omics datasets shed light on molecular changes induced by various NMs *in vitro* and *in vivo* from different perspectives, using different omics layers, techniques and methods, cell models, species and so on. The main question that remains is, how to use this existing information to support NM grouping. An overview of predictive models as well as other useful approaches will be given in the next section. In addition, the different obstacles rendering the development of omics-based models non-trivial is discussed.

As mentioned previously, one of the major advantages with respect to omics data is that almost all journals require study authors to make the raw datasets belonging to a publication available. Therefore, a large amount of data is available in public databases, e.g., NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) for transcriptomic data or PRIDE archive (https://www.ebi.ac.uk/pride/archive/) for proteomic data. As ML models require large datasets with respect to the number of samples, reuse of this data to train or test models is of great value for developing robust approaches and has frequently been applied. In addition, meta-analyses of several omics-based studies may also broaden the understanding of molecular mechanisms and MoAs of NMs. Predictive ML models and other useful tools developed in the field of omics-based NM grouping are summarized in Tables 3 and 4, respectively.

From the literature review performed here, we identified a few models that aim to predict NM toxicity or grouping based on omics measurements (see Table 3). The models use different ML algorithms to predict either *in vitro* or *in vivo* outcomes or to directly suggest grouping on NMs. All models yield high predictive performances of at least 0.7 for the validation set. To reduce the number of descriptors, two studies use feature selection [97,98] for reducing the parameters included in the final model thereby improving explainability.

In addition, multiple other tools have been developed or used to support omics-based analysis of NM toxicity. Kohonen et al.[99] developed a 'predictive toxicogenomics space' (PTGS) tool which yields a predictive signature for drug-induced liver injury (DILI). This tool has also recently been applied successfully to NMs[100]. Serra et al.[101] created the INSIdE NANO tools which contextualizes transcriptomic changes of NMs with those induced by drugs, chemicals and diseases. A similar approach is followed by Bahl et al.[102] who developed PROTEOMAS, a harmonized proteomic workflow

which is applied to a case study in a very similar fashion. In addition, tools calculating benchmark doses (BMDs) based on omics data are also useful to support NM toxicity evaluation and grouping. Proposals for such tools have been made by Halappanavar et al.[103], Gromelski et al.[52] and Serra et al.[104] Others have attempted to link physico-chemical properties of NMs to observations of changes in omics data. Kinaret et al.[105] used coexpression networks and Bannuscher et al.[106] and Karkossa et al.[92] used Weighted Gene Correlation Network Analysis (WGCNA)[107]. Jagiello et al. developed a QSAR models for predicting transcriptomic pathway level response[108]. A complete list of identified approaches with more details is shown in Table 4. In addition, this table also lists several meta-analyses which may provide useful datasets for further model development or validation.

Table 4: Predictive ML models for NM toxicity and grouping based on omics data

Omics layer	**Reference**	**Model type**	**Feature selection performed**	**Validation performed**	**Study design**	**Predictive performance of best model or proposed grouping**
Transcriptomics	Furxhi et al., 2018[109]	Bayesian network		Internal 10-fold cross-validation + external test set + reliability validation set	Eight different types of NM, certain pathways included into BN	Depending on the endpoint, ~ 0.9 - 1.0 for the test set and 0.7-0.9 for reliability validation set
Proteomics	Yanamala et al., 2019[97]	HCA, LCA, L1-l2-norm, SVM, t-test	RFE	External validation set	Carboneous NMs, pulmonary toxicity (BALF) in mice	Prediction accuracy = 0.9 - 1.0
	Billing et al., 2020[86]	Fuzzy c-means algorithm, PCA			Fe_3O_4 nanoparticles doped with increasing amounts of cobalt, in BALF, modeling inflammation, cytotoxicity and genotoxicity	PCA based on three protein signatures (NP response, NETosis, and NP response) as proposed grouping
Multiomics	Fortino et al., 2022[98]	Bayesian information criterion, Similar Network Fusion, Logistic regression, RF	PCA, LASSO, varSelRF, GARBO	External validation set, 70% training and 30% test set	31 industrially relevant ENMs, metals / metal oxides and carboneous materials, mRNA, miRNA, proteins and protein corona, THP-1 and BEAS-2B, mouse lung tissue, EC10, classification based on cytotoxicity and neutrophil infiltration	Accuracy ~1.0 for cytotoxicity and ~0.95 for in vivo toxicity in the best case

Table 5: Additional omics approaches that could support NM grouping

Omics layer	Reference	Approach / Model type	Study design	Descriptors suitable for use in NM grouping
Transcriptomics	Kohonen et al., 2017[99]	Predictive toxicogenomics space (PTGS) Unsupervised probabilistic component modelling	- Input: transcriptomic data of 1,300 compounds (CMap) and dose-dependent cytotoxicity data (NCI-60) - Prediction of chemically-induced pathological states in liver - PTGS composed of 1,331 genes distributed over 14 overlapping cytotoxicity-related gene space components	Predictive signature may be useful for NM-induced liver toxicity and approach may be used to find predictive signatures for other relevant endpoints
	Serra et al., 2019[101]	INSIdE NANO Interaction network	- Input: over 3,000 biological entities including 28 NMs (mainly metal-oxides) - Systems biology framework for contextualization of MoA of NMs - Infer knowledge on NMs from drugs, chemicals and diseases	Biosignatures and similarities with other entities within the network
	Serra et al., 2020[104]	BMDx Benchmark Dose (BMD) Modelling	- Input: gene expression matrix and phenotype table - Computes BMDs, related values and IC50/EC50 estimations	BMD based on transcriptional changes
	Serra et al., 2020[110]	TinderMix Time- and dose-response	- Input: Gene log fold changes - Simultaneously models the effects of time and dose on the transcriptome	Identification of genes showing a dynamic (time-dependent) and dose-dependent response; responsive genes labelled

	modeling for transcriptomics	- Fits different integrated time and dose models to each gene, selects the optimal one, and computes time and dose effect map - Genes with time- and dose-dependent response	according to integrated time and dose point of departure
Halappanavar et al., 2019[103]	BMDt / BMDneu BMD modelling	- Input: Genome-wide lung transcriptomic responses for 10 different MWCNTs, 9 variants of nano TiO_2 and one Carbon Black type - BMD values for transcriptional and apical endpoints	BMD for transcriptional changes used to rank NMs by their potency (median BMD of all or most relevant pathways)
Marwah et al., 2019[111]	eUTOPIA Transcriptomic analysis workflow	- Processing, visualization and interpretation of results - R shiny app	Differential gene expression results
Gromelski et al., 2022[52]	Nano-QSAR model	- Input: Length and diameter of MWCNTs - Web-based application that enables to predict the transcriptomic pathway-level response - Predict doses that initiate inflammation	BMD(L) value by AOP-anchored Nano-QSAR model and expectations with respect to up- and downregulation of genes
Halappanavar et al., 2021[112]	Review on AOPs	- Systematic review of nanotoxicology literature for identifying KEs of relevance to NMs - Development of Nano-AOP database - Case studies using the database to describe key events for AOPs	Description of effects related to certain key events, possibility of calculating e.g., BMDs
Kinaret et al., 2017[105]	Coexpression networks	- Correlations between gene expression profiles and physico-chemical properties - Significantly correlated genes selected - Gene-gene coexpression networks inferred - Genes in the network ranked based on previous correlations - Three different network response modules extrapolated (5,10 and 20 top-ranked genes)	List of most important genes and GO terms from coexpression networks and correlations with physico-chemical properties

		- GO terms progressively enriched across the response modules chosen	
Labib et al., 2016[113]	Transcriptional BMD	- Input: gene expression profiles for three MWCNTs - Significantly perturbed pathways categorized along key events in lung fibrosis AOP - Benchmark doses (BMDs) calculated for each perturbed pathway - Overall transcriptional BMDs for each MWCNT derived	Overall transcriptional BMDs for each NM
Jagiello et al., 2021[108]	Nano-QSAR model	- Input: Transcriptomics data - Selection of relevant pathways for modeling - Generation of AOP-informed Nano-QSAR model - Applied to MWCNT case study	AOP-informed Nano-QSAR modeling approach
Saarimäki et al., 2023[114]	AOP fingerprint	- Multi-step strategy to annotate AOPs is developed - Highlight relevant AOs for chemical exposures with strong *in vitro* and *in vivo* convergence - Supporting chemical grouping and other data-driven approaches - Panel of AOP-derived in vitro biomarkers for pulmonary fibrosis identified and experimentally validated	AOP fingerprint and AOP-driver biomarkers
Williams et al., 2015[127]	Bi-clustering	- Input: transcriptional data for TiO_2 NMs, CB and CNTs - Meta-analysis on public microarray datasets for pulmonary diseases in mouse models following substance exposure - Similar gene expression profiles identified - Bi-clusters used for GSEA - Determination of disease significance of these data-driven gene sets	Toxicity fingerprints for lung diseases (List of gene sets)

	Halappanavar et al., 2015[78]	Transcriptional profiling	- Transcriptional profiling for different TiO_2 NM variants - Determination of differentially expressed genes	List of differentially expressed genes (DEGs)
	Balfourier et al., 2023[115]	Meta-analysis	- Input: 56 GEO microarray datasets on 8 metals and 2 non-metals - Forward selection of relevant properties	List of differentially expressed genes (DEGs) and pathway enrichment; clustering of NMs in meta-analysis
	Saarimäki et al., 2021[116]	Meta-analysis	- Manually curated database on transcriptomic profiles on 101 NM-related datasets from GEO and Array express	Curated and FAIRified collection of NM transcriptomics data including their physicochemical characteristics
	Ghojavand, Bagheri and Tanha, 2019[117]	Meta-analysis	- Publicly available microarray dataset treated by either Ag NPs or Ag ions - Protein-protein-interaction network analysis	List of DEGs and up- and down-regulated hub genes
	Nikota et al., 2016[118]	Meta-analysis, clustering	- Input: Seven toxicogenomics studies on mouse pulmonary responses for carbon nanotubes (CNTs), carbon black, and TiO_2 NPs - mRNA profiles compared to publicly available datasets of 15 other mouse models of lung injury/diseases induced by various agents - Implications of ENM-perturbed biological processes to disease pathogenesis in lungs	List of DEGs; clustering of traits in meta-analysis
Proteomics	Bahl et al., 2023[102]	PROTEOMAS Analysis workflow	- Proteomic analysis workflow for harmonized meta-analyses - Exemplified using 25 datasets from PRIDE archive on lung-specific alterations	List of altered proteins and pathway enrichment; clustering of traits in meta-analysis
	Basak et al., 2016[119]	Computation of similarity index	- Input: MWCNTs and TiO_2 nanobelts, Caco-2/HT29-MTX cells in co-culture	Similarity index telling how similar abundances of a certain

				protein are between two conditions
	Varsou et al., 2018[120]	toxFlow Model-based connectivity mapping	- Web-tool integrating physicochemical, omics and biology information for read-across prediction - Based on gene set variation analysis (GSVA) - Applied in a case study based on 129 protein corona fingerprints (PCF) for 84 gold NMs to predict cell association with human A549 cells (R_{LOO}^2 = 0.97)	Integrated network of physicochemical, omics and biology information data for read-across prediction
Metabolomics	Enea et al., 2019[121]	PCA and OPLS-DA	- Input: Gold nanospheres versus gold nanostars in rat liver	Discriminating metabolites
Multiomics	Shin et al., 2021[122]	PCA and kNN	- Input: 11 NMs with different core materials - Transcriptomics, proteomics and metabolomics - Afterwards 19 drugs applied to see how they affect nanotoxicity effects	Predictive signature
	Canzler et al., 2020[123]	multiGSEA	- Tool to perform sequential GSEA integrating multiple omics layers	Pathway enrichment across omics levels
	Bannuscher et al., 2019[106]	Weighted Gene Correlation Network Analysis (WGCNA)	- Input: Proteomics and metabolomics for 7 NMs in NR8383 cells	Correlations between omics and physico-chemical properties and identification of key drivers
	Karkossa et al., 2019[92]	Weighted Gene Correlation Network Analysis (WGCNA)	- Input: Proteomics and metabolomics for 11 NMs in RLE-6TN cells	Correlations between omics and physico-chemical properties and identification of key drivers
	Dumit et al., 2023[124]	Meta-analysis, Random Forest	- Meta-analysis across publicly available proteomic and transcriptomic data on MWCNTs	Identification of most discriminant HALLMARK pathways, prediction of

			expected behavior with respect to inducing a fibre-specific MoA
Martens et al., 2018[125]	WikiPathways	- Make omics data interoperable with the AOP-Wiki	Mapping of omics data to AOPs
Cai et al., 2018[126]	Multi-hierarchical nano-SAR profiling	- Input: Fe_2O_3 NMs - Heatmap correlating metabolomic and proteomic results to seven basic physico-chemical properties - Metabolic pathways from MetaboAnalyst for metabolomics and KEGG pathways for proteomics	Correlations between omics, physico-chemical properties and toxicity outcomes

7 The value of AI for supporting NAMs

AI is a subfield of computer science in which algorithms and models are developed to mimic cognitive functions of human intelligence, such as learning and problem-solving. AI can aid risk assessment of NMs in various ways, especially by automating and improving commonly used processes. The major field of classical ML mainly dealing with pattern recognition and predictive modelling has already been introduced in detail above. However, this type of modeling and pattern recognition is only one part of existing AI methods. Other applications of AI may also be relevant for risk assessment of NMs and are briefly introduced below.

Automated (meta)data extraction and linked data

One important task in risk assessment is to gather all available data on the NM under study. This is a difficult and time-consuming task as data might be spread across various databases, tables or even be only present in scientific publications as unstructured texts. AI may support this task in terms of data and text mining using automated information retrieval and natural language processing (NLP). Different approaches for mining chemical and biological data have been developed in the past. Swain and Cole developed the ChemDataExtractor for automated extraction of chemical information from scientific literature[128]. Also tmChem[129] can perform chemical named entity recognition from texts. In addition, CD-REST[130] is able to extract chemical-induced disease relations from literature. This automated data extraction may support integration of individual findings across multiple publications which was not obvious previously. Also, relationships between diseases and genes[131] or proteins and drugs[132] could be identified by text mining approaches. Similar approaches may also be used in the field of nanotoxicology. Especially, with recent developments in the field of Large Language Models (LLMs) this field is expected to be of major relevance in future research. Integration of such LLMs with knowledge graphs may further support data retrieval and storage in a structured way. This may also aid the curation of databases with respect to toxicological results as well as metadata. In addition, linking different databases holding certain information on NMs may also be facilitated by LLMs due to automated recognition of similar terms and alternative naming. As an example, AI may be useful for linking NM-specific information stored in databases like eNanoMapper[133] to omics databases like GEO or PRIDE in an efficient way even if the naming schema and underlying ontologies differ. In the broader context of risk assessment, toxicity information may be automatically integrated with other information such as their intended use or information based on different exposure scenarios. This may also be useful in terms of prioritization of NMs to be investigated and ML models may be developed specifically for this task. Additionally, if trained well, AI models can evaluate multiple risk factors and dependencies between them simultaneously. This may be of high value when evaluating complex mixtures of NMs or chemicals.

Data curation

Data curation is of utmost importance for developing reliable approaches and models for NM toxicity[134]. However, if performed manually this is a highly time-consuming task with numerous

challenges as shown in various projects before. AI may provide useful tools to facilitate this task[135]. Here, anomaly detection is one of the most well-known information enrichment techniques which enables identification of outliers or patterns differing from the rest of the data in an automated way. Suitable models may significantly speed-up the identification of inconsistencies in the data. In addition, NLP may also improve the detection of data gaps in registration dossiers for NMs on the market in the context of regulatory process optimization by automated screening strategies. At the same time, recommendations based on historical data may automatically be generated. In addition to advantages in time consumption, AI methods are also less prone to errors compared to humans processing large amounts of data, given that the input data is of high quality. While not every published study can be considered as high quality, AI methods can directly support the identification of quality issues and assess the data quality. Anomaly detection is one very prominent example of data quality checks and is frequently supported by computational tools like Isolation Forests[136] or autoencoders[137].

Image analysis

Another field which provides great opportunities for the application of AI is the field of image analysis. Automated image analysis may be enabled by deep learning approaches like convolutional neural networks. An example of such an approach has been provided by Aversa et al.[138] who automatically identified NMs in Transmission Electron Microscopy (TEM)/ Scanning Electron Microscope (SEM) images and derived their size and number. In addition, Karatzas et al.[139] used deep learning models to predict the effects of NMs on Daphnia magna. Similar approaches may also be derived from videos instead of images. However, the main challenge for image recognition is that large datasets need to be labeled before training the model.

Support during omics data analysis

Omics data are especially useful in the context of AI as usually many datasets are publicly available and, in addition, integration of NM-specific data with other chemicals or traits is easier. Thus, they allow to obtain more comprehensive insights into the underlying biological consequences of NM treatment. One of the most common applications of AI to omics data is the identification of potential biomarkers[140]. LLM in turn could also support the interpretation of the identified biomarkers by quickly searching for existing literature and extracting relevant information and context. Similarly, one may also elucidate information on perturbed molecular pathways, affected targets or common patterns induced by treatment with different NMs. These developments are supported by the fact that LLMs like chatGPT can access databases such as GenBank, Ensembl and Gene Ontology, KEGG, Reactome, GEO or ArrayExpress thereby directly being able to connect the various information stored in these resources[141]. The integration across different omics layers is another field which may be supported by AI. Commonly, ML and Deep Learning algorithms are applied in the context of multi-omics integration[142,143]. In predictive ML models, omics data may also be used to infer links between physico-chemical properties, molecular changes and toxicity which may support regulatory decision making or Safe(r)-by-Design strategies. Especially, explainability is an important area of research allowing for new insights into outcomes of AI models thereby probably enhancing their acceptance in the field of toxicology[144].

Apart from all these advantages, AI models are highly dependent on the quality and amount of underlying data. Training AI models requires large datasets of high-quality data which are relevant for the studied question and are not subject to biases. One important limiting factor is data availability. Therefore, the implementation of the FAIR (findable, accessible, interoperable and reusable) principles is a key factor in the development of reliable AI models[145].

8 Remaining challenges and future requirements

Data availability

The first and most critical challenge in developing robust NM grouping frameworks is the limited availability of data. Any grouping approach can only be generalizable to different kinds of NMs if it was developed based on NMs belonging to different classes and property combinations. This requires many datasets to be available and to be comparable to each other in a way that they can be integrated. Especially, ML models can only make robust predictions if they have been trained on large sets of data containing as much variability as possible[146]. Although high-throughput omics technologies are facilitating the generation of ever larger training data sets even by academic researchers, no single project will be able to generate sufficiently extensive datasets, and data reuse is the only option. In order to allow proper data reuse, the FAIR data principles[145,147] should be considered when publishing any results. Briefly, the principles refer to e.g., clear and harmonized use of persistent identifiers, and that data as well as corresponding metadata can easily be found by both humans and machines at the same time. Data should also be accessible, either openly or through means of authorization or authentication depending on the need for restriction or not. Furthermore, data should be interoperable in order to allow for integration across highly diverse data sets. Overall, the principles guide in terms of making data reusable in the sense that descriptions of the data and metadata are sufficiently detailed to allow for reproducibility, automated assessments of for example data quality and large-scale integration allowing for novel interpretation opportunities.

Unique identification of NMs

Another central question in the case of NMs, which is directly related to the FAIR data principles, is how to uniquely identify a specific NM. Each NM has a multitude of different physico-chemical properties which may also change during their lifecycle. Small changes in few of these properties may have large impacts on the toxicity outcome. Even for benchmark materials, variations may occur if they are from different batches. Therefore, it is important to know exactly which NF was studied in a certain experiment in order to be able to interpret and reuse data. Due to the complexity of NMs, this is, however, not straightforward. Some attempts have been made to structure the nomenclature in such a way that each NM can be uniquely identified, e.g. the NInChI[148] or labelled with a unique identifier even to the level of new batches, e.g. the European Materials Registry[149]. These first approaches will have to be fine-tuned in the future and, importantly, also must be used within the community such that this information will be provided along with published data.

Reliability of data

Reliability of studies on NMs is an important factor. This is mainly because the handling of NMs is not easy. Also, the best way to disperse NMs is still a matter of debate and thus handled differently among researchers. In many cases, there are no standardized ways on how to perform certain measurements of physico-chemical properties or toxicity assays. Many of the OECD test guidelines are not yet adapted for NMs[42,47]. In addition, interferences of toxicological assays with the tested NMs may occur and suitable replacements are not always available at the time being. All these uncertainties, missing adaptations to NMs and lacking standardizations lead to high variability of the resulting measurements and renders integration of results from different studies a difficult task.

Standardization and regulatory acceptance of omics data

Some of the uncertainties induced by grouping based on physico-chemical properties alone can be overcome by adding omics data as these add an additional layer of mechanistic understanding. However, omics data often lack sufficient comparability. This is mainly because different measurement devices and analysis pipelines are in use. Therefore, also in the case of omics data harmonization and standardization of data reporting and analysis workflows are of utmost importance. The transcriptomic[28] and metabolomic[29] reporting frameworks developed by the OECD are highly valuable tools in this context. These frameworks define certain information that needs to be provided for transcriptomic and metabolomic experiments thereby supporting the implementation of the FAIR data principles.

Obtaining robust models

Finally, the question on how to combine and compare NM-related data in order to obtain robust models is another important aspect for grouping of NMs. Here, different aspects need to be taken into consideration: 1) Different measures of similarity exist and can be applied. A number of tools for quantifying similarity between NMs were recently compared by Jeliazkova et al.[150] However, the different measures have different advantages and disadvantages and may eventually lead to different results. The use of high-dimensional omics datasets may even pose additional challenges in this regard; 2) Instead of focusing solely on NMs, integration with other more extensively studied traits like chemical, drugs or diseases may be supporting the acceptance of a certain grouping hypotheses as they may provide additional insights. While this is complicated with regards to physico-chemical properties as those may be very different between NMs and other traits, on the level of omics data this approach seems to be reasonable. ML models like those based on transfer learning may be very valuable in this context; 3) AI in general is expected to form an important pillar in the generation of robust models as it will not be feasible to search, evaluate and integrate the huge and ever-increasing amount of data available manually; and 4) Once ML models have been developed on well standardized data and methods, extensive validation is indispensable.

9 Conclusion

ML models for NM grouping

ML is a valuable tool supporting NM grouping. Especially, the ability to extract the most important parameters describing toxicity in an automated, objective fashion is of great use. The extracted properties can then be used to determine the similarity of NMs. ML models can aid NM grouping approaches in several ways: 1) ML models are able to derive information and patterns from complex and high-dimensional datasets which cannot be easily detected by human inspection. Thus, they may capture more complex interactions between physical, chemical, and biological properties and can potentially enable more accurate classification and grouping of NMs; 2) ML models can predict the behavior of NMs under different conditions which may reduce time and cost when assessing safety, toxicity, and environmental impact of NMs and may guide SSbD approaches; 3) ML models are very time and cost efficient. Once a model is trained, incoming data for new NMs can usually be processed very quickly; and 4) ML can also aid the discovery of new properties or applications of NMs by identifying common patterns or correlations in large datasets. In addition, combining supervised and unsupervised approaches may also be important. While the prediction task of models is usually based on labeled data and supervised approaches, the advantage of unsupervised methods is that they can also use the large pool of unlabeled data to search for as patterns or reduce dimensionality. As dataset size is of high relevance with respect to developing robust ML models with a large applicability domain, this is a very useful and recommended strategy.

Omics for NM grouping

Usually, NM grouping approaches are based on measured physico-chemical properties or sometimes calculated descriptors. The advantage of directly linking physico-chemical properties with toxicity is that one obtains information on how NMs need to be manipulated to make them safer, e.g., in Safe(r)-and-sustainable-by-design approaches. However, limitations with respect to the generalizability and applicability domain have frequently been discussed and can also be observed from the different models presented in this review. This is because the importance of certain properties largely varies with the material types and shapes under study. Also, similar effects may be obtained from NMs with very different physico-chemical properties. Omics approaches are very well suited to overcome these limitations as they yield additional mechanistic insights which can support NM grouping. Nevertheless, there are a few challenges which need to be overcome in order to successfully integrated omics results into NM grouping approaches. Especially, FAIRification of NM omics data, harmonization and standardization of measurements and analyses workflows, the definition of similarity and validation of findings are major fields in which improvement is still required in order to achieve robust models and regulatory acceptance. Once these hurdles are successfully tackled, omics approaches are a very promising source of information for supporting NM grouping. This is confirmed by the first approaches which obtained very good predictive performance when including omics data in the prediction models for NM toxicity. Thereby, omics approaches hold a number of benefits for NM grouping approaches: 1) Omics data provide detailed information about biological effects and interactions of NMs. This may aid the formulation and testing of a grouping hypothesis and thus support regulatory bodies in making informed decisions regarding the use and control of NM; 2) While grouping approaches based on physico-chemical properties suffer from the

fact that these properties may change depending on their surrounding medium as well as over time, omics provide a more direct read-out of the actual state of the NM that was seen by the cells. In combination with ML patterns and relationships not evident from traditional analysis methods may be detected; and 3) If well performed and analyzed, omics experiments are directly comparable among studies thus well-suited for meta-analyses. This does not only hold true within studies on different NMs but instead meta-analyses including other traits like chemicals, drugs and so on are also possible. The fact that a huge number of such omics datasets is publicly available including raw data as well as metadata, is very promising for the implementation of robust ML models, whose performance largely depends on the amount of available data.

AI for NM grouping

AI in general will be unavoidable in the context of developing reliable grouping approaches to cope with the wealth of information available. The complexity of the task of finding suitable NAMs for NM toxicity assessment leads to a high number of potential applications of different AI methods. Especially NLP is expected to have a great influence on future data processing and retrieval. The following tasks have high potential for support by NLP: 1) Automated review and data extraction from scientific literature. This may also include automated generation of databases; 2) Finding patterns, trends and inconsistencies in research papers or datasets; 3) Standardization of terminologies; 4) Compliance of research documents with regulatory standards; and 5) automated annotation and generation of metadata within databases. However, while AI may provide very helpful tools for obtaining information in an efficient and objective manner, human judgement will be of utmost importance and should not be underestimated. This may especially be important in terms of plausibility considerations, contextual interpretation, decision making, quality checks or results with high uncertainties.

Recommendations for future research investment

In order to implement ML and omics techniques successfully in the process of NM grouping, standardization of test methods including dispersion and quality assessment of produced data are urgently needed in order to allow comparison between results from different studies. In line with that, raw data should be made available to the community once they are published and they should be accompanied by sufficient metadata. This may be supported by further refinement of file and model sharing formats, ontologies, terminologies and data quality assessment tools specific for the needs in the field of NMs. Along these lines, it is also of high relevance to find a solution for unambiguous naming of NMs which allows direct comparison of the data. New developments in AI, especially in the field of LLMs can aid the curation and linkage of existing databases. More reliable and comparable data will automatically support the implementation of more robust ML models with a larger applicability domain. At the same time, it will be necessary to enhance the explainability of ML models in order to derive a grouping hypothesis that can also be tested for validation purposes. Here, ML algorithms need to be explored and adjusted with respect to methods and insights that allow for interpretation of the outcome. While it is important to increase the explainability of the model itself, also unraveling the underlying MoA will support the hypothesis formulation. Here, more omics studies on NMs and meta-analyses will be highly useful for extracting predictive signatures that can be used in hazard assessment. Once predictive signatures and thus relevant transcripts, proteins or metabolites have

been identified, targeted testing of only these entities will be possible allowing for high-throughput screening.

Acknowledgements

This work was supported by the BMBF SIINN ERA-NET project NanoToxClass project (Grant Agreement No. 03XP0008), the BMBF project nanoGRAVUR (Grant Agreement No. 03XP0002), the EU H2020 project NanoReg2 (Grant Agreement No. 646221), the EU H2020 project NanoInformaTIX (Grant agreement No 814426), the BMBF project InnoMat.Life (Grant Agreement No. 03XP0216X), the EU H2020 project GRACIOUS (Grant Agreement No. 760840) and the EU H2020 project HARMLESS (Grant Agreement No. 953183). SH was funded by Health Canada's Genomics Research and Development Initiative. PN was funded by the Swedish Fund for Research Without Animal Experiments (Grant Agreement No. F2021-0005).

Disclosure of interest

The authors report no conflict of interest.

Data availability statement

Data sharing not applicable – no new data generated

References

1 EC. 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). (2006).

2 OECD. Guidance on grouping of chemicals. second edition." In Series on Testing & Assessment. **194** (2014).

3 ECHA. Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals. (2008).

4 ECHA. Read-Across Assessment Framework (RAAF). *Helsinki, Finland* (2017).

5 Oomen, A. G. *et al.* Grouping and Read-Across Approaches for Risk Assessment of Nanomaterials. *International Journal of Environmental Research and Public Health* **12**, 13415-13434 (2015).

6 Arts, J. H. E. *et al.* A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). *Regulatory Toxicology and Pharmacology* **71**, S1-S27, doi:https://doi.org/10.1016/j.yrtph.2015.03.007 (2015).

7 Oomen, A. G. *et al.* Risk assessment frameworks for nanomaterials: Scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency. *NanoImpact* **9**, 1-13, doi:https://doi.org/10.1016/j.impact.2017.09.001 (2018).
8 Giusti, A. *et al.* Nanomaterial grouping: Existing approaches and future recommendations. *NanoImpact* **16**, 100182, doi:https://doi.org/10.1016/j.impact.2019.100182 (2019).
9 Stone, V. *et al.* A framework for grouping and read-across of nanomaterials- supporting innovation and risk assessment. *Nano Today* **35**, 100941, doi:https://doi.org/10.1016/j.nantod.2020.100941 (2020).
10 Mech, A. *et al.* Insights into possibilities for grouping and read-across for nanomaterials in EU chemicals legislation. *Nanotoxicology* **13**, 119-141, doi:10.1080/17435390.2018.1513092 (2019).
11 Ag Seleci, D. *et al.* Determining nanoform similarity via assessment of surface reactivity by abiotic and in vitro assays. *NanoImpact* **26**, 100390, doi:https://doi.org/10.1016/j.impact.2022.100390 (2022).
12 Keller, J. G. *et al.* Variation in dissolution behavior among different nanoforms and its implication for grouping approaches in inhalation toxicity. *NanoImpact* **23**, 100341, doi:https://doi.org/10.1016/j.impact.2021.100341 (2021).
13 Ruggiero, E., Santizo, K. Y., Persson, M., Delpivo, C. & Wohlleben, W. Food contact of paper and plastic products containing SiO2, Cu-Phthalocyanine, Fe2O3, CaCO3: Ranking factors that control the similarity of form and rate of release. *NanoImpact* **25**, 100372, doi:https://doi.org/10.1016/j.impact.2021.100372 (2022).
14 Jeliazkova, N., Ma-Hock, L., Janer, G., Stratmann, H. & Wohlleben, W. Possibilities to group nanomaterials across different substances – A case study on organic pigments. *NanoImpact* **26**, 100391, doi:https://doi.org/10.1016/j.impact.2022.100391 (2022).
15 Cross, R. *et al.* Assessing the similarity of nanoforms based on the biodegradation of organic surface treatment chemicals. *NanoImpact* **26**, 100395, doi:https://doi.org/10.1016/j.impact.2022.100395 (2022).
16 Song, Y., Bleeker, E., Cross, R. K., Vijver, M. G. & Peijnenburg, W. J. G. M. Similarity assessment of metallic nanoparticles within a risk assessment framework: A case study on metallic nanoparticles and lettuce. *NanoImpact* **26**, 100397, doi:https://doi.org/10.1016/j.impact.2022.100397 (2022).
17 Murphy, F. A. *et al.* How to formulate hypotheses and IATAs to support grouping and read-across of nanoforms. *Altex* **40**, 125-140, doi:10.14573/altex.2203241 (2023).
18 Lynch, I., Weiss, C. & Valsami-Jones, E. A strategy for grouping of nanomaterials based on key physico-chemical descriptors as a basis for safer-by-design NMs. *Nano Today* **9**, 266-270, doi:https://doi.org/10.1016/j.nantod.2014.05.001 (2014).
19 Cosnier, F. *et al.* Retained particle surface area dose drives inflammation in rat lungs following acute, subacute, and subchronic inhalation of nanomaterials. *Particle and Fibre Toxicology* **18**, 29, doi:10.1186/s12989-021-00419-w (2021).
20 Poulsen, S. S. *et al.* Multi-walled carbon nanotube physicochemical properties predict pulmonary inflammation and genotoxicity. *Nanotoxicology* **10**, 1263-1275, doi:10.1080/17435390.2016.1202351 (2016).
21 EC. EU 2018/1881 amending Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). (2018).
22 Gutierrez, C. T. *et al.* Acute phase response following pulmonary exposure to soluble and insoluble metal oxide nanomaterials in mice. *Part Fibre Toxicol* **20**, 4, doi:10.1186/s12989-023-00514-0 (2023).
23 Ribeiro, A. R. *et al.* Challenges on the toxicological predictions of engineered nanoparticles. *NanoImpact* **8**, 59-72, doi:https://doi.org/10.1016/j.impact.2017.07.006 (2017).
24 Drasler, B., Sayre, P., Steinhäuser, K. G., Petri-Fink, A. & Rothen-Rutishauser, B. In vitro approaches to assess the hazard of nanomaterials. *NanoImpact* **8**, 99-116, doi:https://doi.org/10.1016/j.impact.2017.08.002 (2017).

25 Kinaret, P. A. S. *et al.* Transcriptomics in Toxicogenomics, Part I: Experimental Design, Technologies, Publicly Available Data, and Regulatory Aspects. *Nanomaterials* **10**, 750 (2020).
26 Federico, A. *et al.* Transcriptomics in Toxicogenomics, Part II: Preprocessing and Differential Expression Analysis for High Quality Data. *Nanomaterials* **10**, 903 (2020).
27 Serra, A. *et al.* Transcriptomics in Toxicogenomics, Part III: Data Modelling for Risk Assessment. *Nanomaterials* **10**, 708 (2020).
28 OECD. Transcriptomic Reporting Framework (TRF). (2021).
29 OECD. Metabolomic Reporting Framework. (2021).
30 Ma, N. L. *et al.* Use, exposure and omics characterisation of potential hazard in nanomaterials. *Materials Today Advances* **17**, 100341, doi:10.1016/j.mtadv.2023.100341 (2023).
31 Nguyen, N. H. A. & Falagan-Lotsch, P. Mechanistic Insights into the Biological Effects of Engineered Nanomaterials: A Focus on Gold Nanoparticles. *Int J Mol Sci* **24**, doi:10.3390/ijms24044109 (2023).
32 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210, doi:10.1093/nar/30.1.207 (2002).
33 Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research* **50**, D543-D552, doi:10.1093/nar/gkab1038 (2021).
34 Gareth James, D. W., Trevor Hastie, Robert Tibshirani. *An introduction to statistical learning : with applications in R*. (Springer Texts in Statistics, 2021).
35 Gareth James, D. W., Trevor Hastie, Robert Tibshirani, Jonathan Taylor. *An introduction to statistical learning : with applications in Python*. (Springer Texts in Statistics, 2023).
36 Cervantes-Gracia, K., Chahwan, R. & Husi, H. Integrative OMICS Data-Driven Procedure Using a Derivatized Meta-Analysis Approach. *Frontiers in Genetics* **13**, doi:10.3389/fgene.2022.828786 (2022).
37 Ye, C. *et al.* DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nature Communications* **9**, doi:10.1038/s41467-018-06500-x (2018).
38 Grafström, R. C. *et al.* Toward the Replacement of Animal Experiments through the Bioinformatics-driven Analysis of 'Omics' Data from Human Cell Cultures. *Altern Lab Anim* **43**, 325-332, doi:10.1177/026119291504300506 (2015).
39 Lamon, L., Aschberger, K., Asturiol, D., Richarz, A. & Worth, A. Grouping of nanomaterials to read-across hazard endpoints: a review. *Nanotoxicology* **13**, 100-118, doi:10.1080/17435390.2018.1506060 (2019).
40 Basei, G. *et al.* Making use of available and emerging data to predict the hazards of engineered nanomaterials by means of in silico tools: A critical review. *NanoImpact* **13**, 76-99, doi:https://doi.org/10.1016/j.impact.2019.01.003 (2019).
41 Furxhi, I., Murphy, F., Mullins, M., Arvanitis, A. & Poland, C. A. Practices and Trends of Machine Learning Application in Nanotoxicology. *Nanomaterials (Basel)* **10**, doi:10.3390/nano10010116 (2020).
42 Rasmussen, K., Rauscher, H., Kearns, P., González, M. & Riego Sintes, J. Developing OECD test guidelines for regulatory testing of nanomaterials to ensure mutual acceptance of test data. *Regulatory Toxicology and Pharmacology* **104**, 74-83, doi:https://doi.org/10.1016/j.yrtph.2019.02.008 (2019).
43 EC. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. (2009).
44 EC. Regulation (EC) No 1935/2004 of the European Parliament and of the Council of 27 October 2004 on materials and articles intended to come into contact with food and repealing Directives 80/590/EEC and 89/109/EEC. (2004).
45 EC. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 con-cerning the placing of plant protection products on the market. . (2009).

46 Jagiello K., S. A., Stępnik M., Gromelski M., Płonka K. Nano-specific alternative methods in human hazard/safety assessment under different EU regulations, considering the animal testing bans already in place for cosmetics and their ingredients. *EUON report* (2022).
47 Bleeker, E. A. J. *et al.* Towards harmonisation of testing of nanomaterials for EU regulatory requirements on chemical safety - A proposal for further actions. (2023).
48 Schmeisser, S. *et al.* New approach methodologies in human regulatory toxicology – Not if, but how and when! *Environment International* **178**, 108082, doi:https://doi.org/10.1016/j.envint.2023.108082 (2023).
49 ECHA. New Approach Methodologies in Regulatory Science, Proceedings of a scientific workshop Helsinki, 19–20 April 2016 doi:10.2823/543644 (2016).
50 Halappanavar, S. *et al.* Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. *Particle and Fibre Toxicology* **17**, 16, doi:10.1186/s12989-020-00344-4 (2020).
51 Gerloff, K. *et al.* The Adverse Outcome Pathway approach in nanotoxicology. *Computational Toxicology* **1**, 3-11, doi:https://doi.org/10.1016/j.comtox.2016.07.001 (2017).
52 Gromelski, M. *et al.* AOP173 key event associated pathway predictor – online application for the prediction of benchmark dose lower bound (BMDLs) of a transcriptomic pathway involved in MWCNTs-induced lung fibrosis. *Nanotoxicology* **16**, 183-194, doi:10.1080/17435390.2022.2064250 (2022).
53 Nymark, P., Karlsson, H. L., Halappanavar, S. & Vogel, U. Adverse Outcome Pathway Development for Assessment of Lung Carcinogenicity by Nanoparticles. *Front Toxicol* **3**, 653386, doi:10.3389/ftox.2021.653386 (2021).
54 Halappanavar, S. *et al.* Non-Animal Strategies for Toxicity Assessment of Nanoscale Materials: Role of Adverse Outcome Pathways in the Selection of Endpoints. *Small* **17**, e2007628, doi:10.1002/smll.202007628 (2021).
55 Wohlleben, W., Mehling, A. & Landsiedel, R. Lessons Learned from the Grouping of Chemicals to Assess Risks to Human Health. Angew Chem Int Ed Engl 62, e202210651, doi:10.1002/anie.202210651 (2023).
56 Lamon, L. *et al.* Grouping of nanomaterials to read-across hazard endpoints: from data collection to assessment of the grouping hypothesis by application of chemoinformatic techniques. *Particle and Fibre Toxicology* **15**, 37, doi:10.1186/s12989-018-0273-1 (2018).
57 Aschberger, K. *et al.* Grouping of multi-walled carbon nanotubes to read-across genotoxicity: A case study to evaluate the applicability of regulatory guidance. *Computational Toxicology* **9**, 22-35, doi:https://doi.org/10.1016/j.comtox.2018.10.001 (2019).
58 Murugadoss, S. *et al.* Identifying nanodescriptors to predict the toxicity of nanomaterials: a case study on titanium dioxide. *Environmental Science: Nano* **8**, 580-590, doi:10.1039/D0EN01031F (2021).
59 Kotzabasaki, M., Sotiropoulos, I., Charitidis, C. & Sarimveis, H. Machine learning methods for multi-walled carbon nanotubes (MWCNT) genotoxicity prediction. *Nanoscale Advances* **3**, 3167-3176, doi:10.1039/D0NA00600A (2021).
60 Sizochenko, N., Syzochenko, M., Fjodorova, N., Rasulev, B. & Leszczynski, J. Evaluating genotoxicity of metal oxide nanoparticles: Application of advanced supervised and unsupervised machine learning techniques. *Ecotoxicology and Environmental Safety* **185**, 109733, doi:https://doi.org/10.1016/j.ecoenv.2019.109733 (2019).
61 Ambure, P. *et al.* Development of Generalized QSAR Models for Predicting Cytotoxicity and Genotoxicity of Metal Oxides Nanoparticles. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)* **5**, 83-100, doi:10.4018/IJQSPR.20201001.oa2 (2020).
62 El Yamani, N. *et al.* Hazard identification of nanomaterials: In silico unraveling of descriptors for cytotoxicity and genotoxicity. *Nano Today* **46**, 101581, doi:https://doi.org/10.1016/j.nantod.2022.101581 (2022).
63 Halder, A. K., Melo, A. & Cordeiro, M. A unified in silico model based on perturbation theory for assessing the genotoxicity of metal oxide nanoparticles. *Chemosphere* **244**, 125489, doi:10.1016/j.chemosphere.2019.125489 (2020).

64 Ban, Z., Zhou, Q., Sun, A., Mu, L. & Hu, X. Screening Priority Factors Determining and Predicting the Reproductive Toxicity of Various Nanoparticles. *Environmental Science & Technology* **52**, 9666-9676, doi:10.1021/acs.est.8b02757 (2018).
65 González-Durruthy, M. *et al.* MitoTarget Modeling Using ANN-Classification Models Based on Fractal SEM Nano-Descriptors: Carbon Nanotubes as Mitochondrial F0F1-ATPase Inhibitors. *Journal of Chemical Information and Modeling* **59**, 86-97, doi:10.1021/acs.jcim.8b00631 (2019).
66 Gajewicz, A. *et al.* Decision tree models to classify nanomaterials according to the DF4nanoGrouping scheme. *Nanotoxicology* **12**, 1-17, doi:10.1080/17435390.2017.1415388 (2018).
67 Bahl, A. *et al.* Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact* **15**, 100179, doi:https://doi.org/10.1016/j.impact.2019.100179 (2019).
68 Bahl, A. *et al.* Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods. *NanoImpact* **19**, 100234, doi:https://doi.org/10.1016/j.impact.2020.100234 (2020).
69 Marvin, H. J. *et al.* Application of Bayesian networks for hazard ranking of nanomaterials to support human health risk assessment. *Nanotoxicology* **11**, 123-133, doi:10.1080/17435390.2016.1278481 (2017).
70 Sheehan, B. *et al.* Hazard Screening Methods for Nanomaterials: A Comparative Study. *Int J Mol Sci* **19**, doi:10.3390/ijms19030649 (2018).
71 Drew, N. M., Kuempel, E. D., Pei, Y. & Yang, F. A quantitative framework to group nanoscale and microscale particles by hazard potency to derive occupational exposure limits: Proof of concept evaluation. *Regul Toxicol Pharmacol* **89**, 253-267, doi:10.1016/j.yrtph.2017.08.003 (2017).
72 Furxhi, I. & Murphy, F. Predicting In Vitro Neurotoxicity Induced by Nanoparticles Using Machine Learning. *Int J Mol Sci* **21**, doi:10.3390/ijms21155280 (2020).
73 Gernand, J. M. & Casman, E. A. A meta-analysis of carbon nanotube pulmonary toxicity studies--how physical dimensions and impurities affect the toxicity of carbon nanotubes. *Risk Anal* **34**, 583-597, doi:10.1111/risa.12109 (2014).
74 Gernand, J. M. & Casman, E. A. Nanoparticle Characteristic Interaction Effects on Pulmonary Toxicity: A Random Forest Modeling Framework to Compare Risks of Nanomaterial Variants. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg* **2**, doi:10.1115/1.4031216 (2016).
75 Bajak, E. *et al.* Changes in Caco-2 cells transcriptome profiles upon exposure to gold nanoparticles. *Toxicology Letters* **233**, 187-199, doi:https://doi.org/10.1016/j.toxlet.2014.12.008 (2015).
76 Boyadzhiev, A. *et al.* Impact of copper oxide particle dissolution on lung epithelial cell toxicity: response characterization using global transcriptional analysis. *Nanotoxicology* **15**, 380-399, doi:10.1080/17435390.2021.1872114 (2021).
77 Snyder-Talkington, B. N. *et al.* Multi-walled carbon nanotube-induced gene expression in vitro: concordance with in vivo studies. *Toxicology* **328**, 66-74, doi:10.1016/j.tox.2014.12.012 (2015).
78 Halappanavar, S. *et al.* Transcriptional profiling identifies physicochemical properties of nanomaterials that are determinants of the in vivo pulmonary response. *Environ Mol Mutagen* **56**, 245-264, doi:10.1002/em.21936 (2015).
79 Chézeau, L. *et al.* Short- and long-term gene expression profiles induced by inhaled TiO(2) nanostructured aerosol in rat lung. *Toxicol Appl Pharmacol* **356**, 54-64, doi:10.1016/j.taap.2018.07.013 (2018).
80 Poulsen, S. S. *et al.* A transcriptomic overview of lung and liver changes one day after pulmonary exposure to graphene and graphene oxide. *Toxicology and Applied Pharmacology* **410**, 115343, doi:https://doi.org/10.1016/j.taap.2020.115343 (2021).

81 Decan, N. *et al.* Characterization of in vitro genotoxic, cytotoxic and transcriptomic responses following exposures to amorphous silica of different sizes. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **796**, 8-22, doi:https://doi.org/10.1016/j.mrgentox.2015.11.011 (2016).
82 Gosens, I. *et al.* Pulmonary toxicity and gene expression changes after short-term inhalation exposure to surface-modified copper oxide nanoparticles. *NanoImpact* **22**, 100313, doi:https://doi.org/10.1016/j.impact.2021.100313 (2021).
83 Solorio-Rodriguez, S. A. *et al.* Single-Walled vs. Multi-Walled Carbon Nanotubes: Influence of Physico-Chemical Properties on Toxicogenomics Responses in Mouse Lungs. *Nanomaterials* **13**, 1059 (2023).
84 Gallud, A. *et al.* Multiparametric Profiling of Engineered Nanomaterials: Unmasking the Surface Coating Effect. *Adv Sci (Weinh)* **7**, 2002221, doi:10.1002/advs.202002221 (2020).
85 Torres, A. *et al.* Repeated Exposure of Macrophages to Synthetic Amorphous Silica Induces Adaptive Proteome Changes and a Moderate Cell Activation. *Nanomaterials (Basel)* **12**, doi:10.3390/nano12091424 (2022).
86 Billing, A. M. *et al.* Fast and Robust Proteome Screening Platform Identifies Neutrophil Extracellular Trap Formation in the Lung in Response to Cobalt Ferrite Nanoparticles. *ACS Nano* **14**, 4096-4110, doi:10.1021/acsnano.9b08818 (2020).
87 Miranda, R. R., Oliveira, A. C. S., Skytte, L., Rasmussen, K. L. & Kjeldsen, F. Proteome-wide analysis reveals molecular pathways affected by AgNP in a ROS-dependent manner. *Nanotoxicology* **16**, 73-87, doi:10.1080/17435390.2022.2036844 (2022).
88 Bannuscher, A. *et al.* Metabolomics profiling to investigate nanomaterial toxicity in vitro and in vivo. *Nanotoxicology* **14**, 807-826, doi:10.1080/17435390.2020.1764123 (2020).
89 Cui, L., Wang, X., Sun, B., Xia, T. & Hu, S. Predictive Metabolomic Signatures for Safety Assessment of Metal Oxide Nanoparticles. *ACS Nano* **13**, 13065-13082, doi:10.1021/acsnano.9b05793 (2019).
90 Chen, Z., Zhou, D., Han, S., Zhou, S. & Jia, G. Hepatotoxicity and the role of the gut-liver axis in rats after oral administration of titanium dioxide nanoparticles. *Particle and Fibre Toxicology* **16**, 48, doi:10.1186/s12989-019-0332-2 (2019).
91 Bannuscher, A. *et al.* A multi-omics approach reveals mechanisms of nanomaterial toxicity and structure-activity relationships in alveolar macrophages. *Nanotoxicology* **14**, 181-195, doi:10.1080/17435390.2019.1684592 (2020).
92 Karkossa, I. *et al.* An in-depth multi-omics analysis in RLE-6TN rat alveolar epithelial cells allows for nanomaterial categorization. *Particle and Fibre Toxicology* **16**, 38, doi:10.1186/s12989-019-0321-5 (2019).
93 Nymark, P. *et al.* Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells. *Nanotoxicology* **9**, 624-635, doi:10.3109/17435390.2015.1017022 (2015).
94 Seidel, C. *et al.* Inhaled multi-walled carbon nanotubes differently modulate global gene and protein expression in rat lungs. *Nanotoxicology* **15**, 238-256, doi:10.1080/17435390.2020.1851418 (2021).
95 Aragoneses-Cazorla, G., Buendia-Nacarino, M. P., Mena, M. L. & Luque-Garcia, J. L. A Multi-Omics Approach to Evaluate the Toxicity Mechanisms Associated with Silver Nanoparticles Exposure. *Nanomaterials (Basel)* **12**, doi:10.3390/nano12101762 (2022).
96 Dekkers, S. *et al.* Multi-omics approaches confirm metal ions mediate the main toxicological pathways of metal-bearing nanoparticles in lung epithelial A549 cells. *Environmental science. Nano* **5**, 1506-1517 (2018).
97 Yanamala, N. *et al.* Grouping of carbonaceous nanomaterials based on association of patterns of inflammatory markers in BAL fluid with adverse outcomes in lungs. *Nanotoxicology* **13**, 1102-1116, doi:10.1080/17435390.2019.1640911 (2019).
98 Fortino, V. *et al.* Biomarkers of nanomaterials hazard from multi-layer data. *Nature Communications* **13**, 3798, doi:10.1038/s41467-022-31609-5 (2022).

99 Kohonen, P. *et al.* A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nature Communications* **8**, 15932, doi:10.1038/ncomms15932 (2017).
100 Abstracts of the 55th Congress of the European Societies of Toxicology (EUROTOX 2019) TOXICOLOGY SCIENCE PROVIDING SOLUTIONS. *Toxicology Letters* **314**, S1-S309, doi:https://doi.org/10.1016/j.toxlet.2019.09.002 (2019).
101 Serra, A. *et al.* INSIdE NANO: a systems biology framework to contextualize the mechanism-of-action of engineered nanomaterials. *Scientific Reports* **9**, 179, doi:10.1038/s41598-018-37411-y (2019).
102 Bahl, A. *et al.* PROTEOMAS: A workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping. *Journal of Cheminformatics* (2023).
103 Halappanavar, S., Ede, J. D., Shatkin, J. A. & Krug, H. F. A systematic process for identifying key events for advancing the development of nanomaterial relevant adverse outcome pathways. *NanoImpact* **15**, 100178, doi:https://doi.org/10.1016/j.impact.2019.100178 (2019).
104 Serra, A., Saarimäki, L. A., Fratello, M., Marwah, V. S. & Greco, D. BMDx: a graphical Shiny application to perform Benchmark Dose analysis for transcriptomics data. *Bioinformatics* **36**, 2932-2933, doi:10.1093/bioinformatics/btaa030 (2020).
105 Kinaret, P. *et al.* Network Analysis Reveals Similar Transcriptomic Responses to Intrinsic Properties of Carbon Nanomaterials in Vitro and in Vivo. *ACS Nano* **11**, 3786-3796, doi:10.1021/acsnano.6b08650 (2017).
106 Bannuscher, A. *et al.* A multi-omics approach reveals mechanisms of nanomaterial toxicity and structure–activity relationships in alveolar macrophages. *Nanotoxicology* **14**, 1-15 (2019).
107 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
108 Jagiello K, H. S., Rybińska-Fryca A, Willliams A, Vogel U, Puzyn T. Transcriptomics-Based and AOP-Informed Structure-Activity Relationshipsto Predict Pulmonary Pathology Induced by Multiwalled Carbon Nanotubes. *Small*, doi:doi: 10.1002/smll.202003465 (2021).
109 Furxhi, I., Murphy, F., Sheehan, B., Mullins, M. & Mantecca, P. in *2018 IEEE 18th International Conference on Nanotechnology (IEEE-NANO).* 1-4.
110 Serra, A. *et al.* TinderMIX: Time-dose integrated modelling of toxicogenomics data. *GigaScience* **9**, doi:10.1093/gigascience/giaa055 (2020).
111 Marwah, V. S. *et al.* eUTOPIA: solUTion for Omics data PreprocessIng and Analysis. *Source Code for Biology and Medicine* **14**, 1, doi:10.1186/s13029-019-0071-7 (2019).
112 Halappanavar, S. *et al.* A methodology for developing key events to advance nanomaterial-relevant adverse outcome pathways to inform risk assessment. *Nanotoxicology* **15**, 289-310, doi:10.1080/17435390.2020.1851419 (2021).
113 Labib, S. *et al.* Nano-risk Science: application of toxicogenomics in an adverse outcome pathway framework for risk assessment of multi-walled carbon nanotubes. *Particle and Fibre Toxicology* **13**, 15, doi:10.1186/s12989-016-0125-9 (2016).
114 Saarimäki, L. A. *et al.* Toxicogenomics Data for Chemical Safety Assessment and Development of New Approach Methodologies: An Adverse Outcome Pathway-Based Approach. *Advanced Science* **10**, 2203984, doi:https://doi.org/10.1002/advs.202203984 (2023).
115 Balfourier, A., Marty, A.-P. & Gazeau, F. Importance of Metal Biotransformation in Cell Response to Metallic Nanoparticles: A Transcriptomic Meta-analysis Study. *ACS Nanoscience Au* **3**, 46-57, doi:10.1021/acsnanoscienceau.2c00035 (2023).
116 Saarimäki, L. A. *et al.* Manually curated transcriptomics data collection for toxicogenomic assessment of engineered nanomaterials. *Scientific Data* **8**, 49, doi:10.1038/s41597-021-00808-y (2021).
117 Ghojavand, S., Bagheri, F. & Mesrian Tanha, H. Integrative meta-analysis of publically available microarray datasets of several epithelial cell lines identifies biological processes

affected by silver nanoparticles exposure. *Comp Biochem Physiol C Toxicol Pharmacol* **216**, 67-74, doi:10.1016/j.cbpc.2018.11.003 (2019).
118 Nikota, J. *et al.* Meta-analysis of transcriptomic responses as a means to identify pulmonary disease outcomes for engineered nanomaterials. *Part Fibre Toxicol* **13**, 25, doi:10.1186/s12989-016-0137-5 (2016).
119 Basak, S. C., Vracko, M. & Witzmann, F. A. Mathematical Nanotoxicoproteomics: Quantitative Characterization of Effects of Multi-walled Carbon Nanotubes (MWCNT) and TiO2 Nanobelts (TiO2-NB) on Protein Expression Patterns in Human Intestinal Cells. *Curr Comput Aided Drug Des* **12**, 259-264, doi:10.2174/1573409912666160824145722 (2016).
120 Varsou, D.-D. *et al.* toxFlow: A Web-Based Application for Read-Across Toxicity Prediction Using Omics and Physicochemical Data. *Journal of Chemical Information and Modeling* **58**, 543-549, doi:10.1021/acs.jcim.7b00160 (2018).
121 Enea, M. *et al.* A Metabolomic Approach for the In Vivo Study of Gold Nanospheres and Nanostars after a Single-Dose Intravenous Administration to Wistar Rats. *Nanomaterials (Basel)* **9**, doi:10.3390/nano9111606 (2019).
122 Shin, T. H. *et al.* Silica-coated magnetic-nanoparticle-induced cytotoxicity is reduced in microglia by glutathione and citrate identified using integrated omics. *Particle and Fibre Toxicology* **18**, 42, doi:10.1186/s12989-021-00433-y (2021).
123 Canzler, S. & Hackermüller, J. multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinformatics* **21**, 561, doi:10.1186/s12859-020-03910-x (2020).
124 Dumit, V. I. *et al.* Meta-analysis of integrated proteomic and transcriptomic data discerns structure-activity-relationship of carbon materials with different morphologies. *Advanced Science*, doi:10.1002/advs.202306268 (2023).
125 Martens, M. *et al.* Introducing WikiPathways as a Data-Source to Support Adverse Outcome Pathways for Regulatory Risk Assessment of Chemicals and Nanomaterials. *Front Genet* **9**, 661, doi:10.3389/fgene.2018.00661 (2018).
126 Cai, X. *et al.* Multi-hierarchical profiling the structure-activity relationships of engineered nanomaterials at nano-bio interfaces. *Nat Commun* **9**, 4416, doi:10.1038/s41467-018-06869-9 (2018).
127 Williams A, H. S. Application of biclustering of gene expression data and gene set enrichment analysis methods to identify potentially disease causing nanomaterials. Beilstein J Nanotechnol. , doi:10.3762/bjnano.6.252 (2015).
128 Swain, M. C. & Cole, J. M. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *Journal of Chemical Information and Modeling* **56**, 1894-1904, doi:10.1021/acs.jcim.6b00207 (2016).
129 Leaman, R., Wei, C.-H. & Lu, Z. tmChem: a high performance approach for chemical named entity recognition and normalization. *Journal of Cheminformatics* **7**, S3, doi:10.1186/1758-2946-7-S1-S3 (2015).
130 Xu, J. *et al.* CD-REST: a system for extracting chemical-induced disease relation in literature. *Database* **2016**, baw036, doi:10.1093/database/baw036 (2016).
131 Lever, J. A.-O., Zhao, E. A.-O., Grewal, J. A.-O., Jones, M. A.-O. X. & Jones, S. A.-O. CancerMine: a literature-mined resource for drivers, oncogenes and tumor suppressors in cancer.
132 Zheng, S., Dharssi, S., Wu, M., Li, J. & Lu, Z. Text Mining for Drug Discovery.
133 Jeliazkova, N. *et al.* The eNanoMapper database for nanomaterial safety information. *Beilstein Journal of Nanotechnology* **6**, 1609-1634, doi:10.3762/bjnano.6.165 (2015).
134 Alves, V. M. *et al.* Curated Data In - Trustworthy In Silico Models Out: The Impact of Data Quality on the Reliability of Artificial Intelligence Models as Alternatives to Animal Testing. *Altern Lab Anim* **49**, 73-82, doi:10.1177/02611929211029635 (2021).
135 Martín, L., Sánchez, L., Lanza, J. & Sotres, P. Development and evaluation of Artificial Intelligence techniques for IoT data quality assessment and curation. *Internet of Things* **22**, 100779, doi:https://doi.org/10.1016/j.iot.2023.100779 (2023).

136 Liu, F. T., Ting, K. M. & Zhou, Z. H. in *2008 Eighth IEEE International Conference on Data Mining.* 413-422.
137 Shvetsova, N., Bakker, B., Fedulova, I., Schulz, H. & Dylov, D. V. Anomaly Detection in Medical Imaging With Deep Perceptual Autoencoders. *IEEE Access* **9**, 118571-118583, doi:10.1109/ACCESS.2021.3107163 (2021).
138 Aversa, R., Modarres, M. H., Cozzini, S., Ciancio, R. & Chiusole, A. The first annotated set of scanning electron microscopy images for nanoscience. *Scientific Data* **5**, 180172, doi:10.1038/sdata.2018.172 (2018).
139 Karatzas, P. *et al.* Development of Deep Learning Models for Predicting the Effects of Exposure to Engineered Nanomaterials on Daphnia magna.
140 Michelhaugh, S. A. & Januzzi, J. L., Jr. Using Artificial Intelligence to Better Predict and Develop Biomarkers. *Heart Fail Clin* **18**, 275-285, doi:10.1016/j.hfc.2021.11.004 (2022).
141 Ng, S., Masarone, S., Watson, D. & Barnes, M. R. The benefits and pitfalls of machine learning for biomarker discovery. *Cell and Tissue Research*, doi:10.1007/s00441-023-03816-z (2023).
142 Picard, M., Scott-Boyer, M.-P., Bodein, A., Périn, O. & Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal* **19**, 3735-3746, doi:https://doi.org/10.1016/j.csbj.2021.06.030 (2021).
143 Biswas, N. & Chakrabarti, S. Artificial Intelligence (AI)-Based Systems Biology Approaches in Multi-Omics Data Analysis of Cancer. *Front Oncol* **10**, 588221, doi:10.3389/fonc.2020.588221 (2020).
144 Santorsola, M. & Lescai, F. The promise of explainable deep learning for omics data analysis: Adding new discovery tools to AI. *New Biotechnology* **77**, 1-11, doi:https://doi.org/10.1016/j.nbt.2023.06.002 (2023).
145 Jeliazkova, N. *et al.* Towards FAIR nanosafety data. *Nature Nanotechnology* **16**, 644-654, doi:10.1038/s41565-021-00911-6 (2021).
146 Duan, Y., Edwards, J. S. & Dwivedi, Y. K. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management* **48**, 63-71, doi:https://doi.org/10.1016/j.ijinfomgt.2019.01.021 (2019).
147 Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018, doi:10.1038/sdata.2016.18 (2016).
148 Lynch, I. *et al.* Can an InChI for Nano Address the Need for a Simplified Representation of Complex Nanomaterials across Experimental and Nanoinformatics Studies? *Nanomaterials* **10**, 2493 (2020).
149 van Rijn, J. *et al.* European Registry of Materials: global, unique identifiers for (undisclosed) nanomaterials. *Journal of Cheminformatics* **14**, 57, doi:10.1186/s13321-022-00614-7 (2022).
150 Jeliazkova, N. *et al.* How can we justify grouping of nanoforms for hazard assessment? Concepts and tools to quantify similarity. *NanoImpact* **25**, 100366, doi:https://doi.org/10.1016/j.impact.2021.100366 (2022).

Chapter 4: Discussion

4.1 ML and omics are highly useful tools to support NM grouping

Within this work, the value of using various bioinformatics and ML models on different kind of data with the aim to support NM grouping approaches and toxicity prediction was emphasized. Different models have been applied and the outcomes of each study will be discussed briefly.

In the first study, unsupervised and supervised ML were used to predict NM toxicity based on physico-chemical properties. Therefore, a classifier based on PCA and kNN was compared to one deploying RFs with and without feature selection. The models were used to predict inhalation toxicity from a set of eleven NMs described by various physico-chemical parameters. The supervised RF model with backward RFE showed higher prediction accuracy. Importantly, the approaches do not only allow for prediction of NM toxicity but at the same time also for identification of physico-chemical properties which are strongly related to the toxicity outcome. This is a major requirement for implementing robust NM grouping approaches. The ML models were applied to a case study consisting mainly of systematically varied silica NMs and some additional materials for benchmarking and checking transferability to other material classes. For this specific case study investigated in the first publication, nine out of eleven NMs could be predicted correctly with respect to inhalation toxicity. The inhalation response was thereby expressed as a binary variable with labels ‘active’ and ‘passive’ based on previous results from STIS and the macrophage assay. The final model was based on three explanatory variables, namely zeta potential, redox potential and dissolution rate which showed highest impact on the toxicity outcome.

While the predictive performance of the model is quite good, it already becomes obvious that generalization to other types of NMs may be challenging. Finding additional sets of NMs for which similar properties have been measured was not possible at the time of the publication due to a lack of standardization of assays back then. This situation has largely improved in the last years with many standardization efforts being performed by standardization organizations like the OECD, ISO or CEN as well as large research projects like NanoDefine, NanoTest, NanoValid, NANoREG, PATROLS or NanoHarmony which are also laid down in several guidance documents. Given the small sample size, the model is not expected to be robust across a larger applicability domain. However, the RF approach itself seems to be very well-suited even with a small number of tested NMs and the technique can easily be adapted to new case studies. Therefore, the models established in this study

can be considered as an initial foundation and a proof-of-concept for advancing NM grouping approaches and gaining better comprehension of similarities between NM variants while making use[189, 190] of the power of computational modeling. Further adaptations will be needed to obtain reliable models that generalize well across a wide range of NMs. Namely, larger and more heterogeneous trainings and test datasets as well as additional properties which are more focused on describing the underlying MoA of the NMs need to be considered. Also, standardization efforts with more and more TGs being adapted for NMs have been made since the publication of this study and are currently on-going within the OECD[191]. This standardization of methods being used will allow for combining datasets across multiple studies for meta-analyses in future, thereby, leading to a more comprehensive data foundation for computational modeling approaches. An inclusion of benchmark materials also aids in assessing whether or not results from different studies are comparable. Finally, once reliable models are generated, external validation is also necessary.

In addition to simple physico-chemical descriptors, the read-outs of higher-level functional assays describing the OP of NMs were used in the second study. This was a promising approach as many studies had observed that OP has a great impact on NM toxicity which was also supported by our own findings from the first study in which redox potential was among the most relevant properties for predicting NM toxicity. Using such surrogate assays may be useful especially for NMs with their complex and variable properties which are difficult to be sufficiently reflected in all details. Thus, they may drastically reduce characterization efforts and at the same time potentially simplify the combination of datasets for meta-analyses. The main difficulty for OP is that various assays based on different principles exist and are frequently used. Therefore, the focus of the second study was on how comparable different OP assays are and how well they predict, alone or in combination, NM toxicity. In this study, the OP of a set of 35 NMs with a variety of different core materials was determined using four different assays, namely ESR with the spin probe CPH as well as with the spin trap DMPO, the FRAS assay and a cell-based protein carbonylation assay. We compared the results of the different assays with respect to their predictivity for a binary outcome variable reflecting inhalation toxicity in terms of mass-based as well as surface-based dose metrics. As a result, we found that surface-based doses show better separation of ‘active’ and ‘passive’ materials which is assumed to be related to the fact that ROS generation only takes place at the surface of the NMs. Also, normalization to the actual deposited dose in the cell-based carbonylation assay was advantageous in terms of predictivity. With respect to the comparability of assay results, we saw moderate correlations between 0.5 and 0.8 for all assay pairs. Also, within NM families, the order of NMs by OP was not perfectly preserved across NMs. Especially those NMs with intermediate OP showed inconsistencies between assays. Comparing all assay combinations with respect to their

ability to separate 'active' and 'passive' NMs in a logistic regression model, carbonylation combined with either ESR or FRAS resulted in the best separation of the two classes.

By combining multiple OP assays, well-performing models for toxicity prediction could already be generated. However, most studies measure OP only with one certain technique. This is challenging as the moderate correlation between the different OP assays prevents combination of datasets for which the OP was detected using different measurement techniques. This hampers meta-analysis settings in which sufficiently large datasets for ML modeling may be generated. In addition, while it is a very common one, ROS generation due to high OP reflects only one specific way in which NMs may induce toxicity. However, other possible MoAs also need to be considered in order to establish robust NM grouping approaches. High-content approaches yielding broader information like omics may be very useful tools in this regard.

Omics techniques are well-suited to inform on various MoAs which could be induced by NMs at the same time. They are useful tools as they provide mechanistic insights into molecular alterations induced by compounds. Currently most omics studies in toxicology focus on transcriptomics. At the same time, other omics layers like proteomics and metabolomics are expected to reflect the actual phenotypes much closer. While for transcriptomics measurements and analysis are already quite well standardized, the analysis of other omics layers is more complex and less harmonized. However, harmonization is a pre-requisite for meta-analyses which would allow for larger datasets necessary for robust modeling. In addition, the lack of standardization currently prevents regulatory acceptance of omics analyses. An advantage of omics techniques is that there are common descriptors that are frequently computed and can be directly compared among studies, e.g., KEGG pathways[165, 166] or GO terms[170, 192]. The same descriptors can also be used across different omics layers allowing for even more comprehensive results. Strikingly, comparisons are not only possible between NM-related studies but can also include results for other traits like chemical treatments, drugs or diseases on which much more knowledge on underlying MoAs is available. For these traits, a huge number of datasets is available in public repositories like GEO or PRIDE.

In order to make use of this available wealth of data, the third study focused on harmonizing proteomics data analysis which is a pre-requisite enabling meta-analyses across different studies. The main challenge for proteomics is the large heterogeneity of datasets caused by the fact that various methods and instruments are used for measuring abundances of proteins in different labs and no standard analysis workflows exist so far. One main difference to transcriptomics is the fact that only a subset of all proteins is detected in each run during mass spectrometry. Different machines thereby

show different sensitivity which leads to substantially varying amounts of identified proteins. Therefore, the amount and identity of identified proteins is highly variable and lots of missing values are present in each dataset. This becomes a major challenging when comparing multiple datasets in a meta-analysis. In order to tackle this challenge, we developed a workflow called PROTEOMAS which allows to automatically process a large number of proteomic datasets without introducing any subjective biases. For each dataset, PROTEOMAS processes the data allowing for identification of altered proteins and subsequently deriving more overarching information on GO terms[170, 192] as well as KEGG[165, 166], Reactome[168, 169] or HALLMARK[167] pathways. These descriptors then comprise the proteomic signature of the dataset. These signatures may be compared among various traits allowing to integrate proteomic results for NMs generated in different projects as well as contextualization using other proteomics datasets from traits for which more profound knowledge exists. While the methods used in each step are kept rather simple, not necessarily modeling every detail of each dataset, the expectation is that the most important patterns would be unraveled once a large number of datasets for the same trait have been investigated while noise cancels out. The usefulness of the workflow was shown in a case study which exemplified a meta-analysis studying the toxicological effects of NMs at the lung level with particular focus on inflammation based on 25 publicly available datasets of different traits.

In the final review publication, different computational approaches which can support NM grouping approaches were assembled. Here, we only considered models for toxicological endpoints relevant under REACH. Apart from general computational models, the focus was put on models which integrate omics results and omics analysis tools in the field of NMs which were not directly developed in the context of NM grouping but might still largely support such kind of approaches. The review showed that in general ML models and omics are very useful tools supporting NM grouping. At the same time, limitations with respect to generalizability and a restricted applicability domain have been acknowledged in most publications. In addition, approaches integrating omics result emphasize on the added value due to these approaches. While some of the models already show good performance with respect to toxicity prediction, most also acknowledge that a number of challenges remain with regard to developing robust prediction models. Major issues that need to be tackled comprise: 1) the limited number of available datasets, 2) the necessary improvement of FAIRification for NM-related data including omics data, 3) the harmonization and standardization of measurements and analysis workflows, 4) the comparison of different measures for NM similarity and 5) the external validation of models to improve regulatory acceptance. Given the wealth and complexity of information available, AI may greatly support the generation of linked data as well as model development. These points will be discussed further in the following subsections.

4.2 Remaining challenges for developing robust ML approaches for NM grouping

4.2.1 Data availability

The most important requirement for developing robust computational models supporting NM grouping are large, high-quality datasets describing physico-chemical properties as well as NM toxicity. These datasets should contain measurements from standardized assays including standardized data evaluation procedures. ML methods can only generalize well to new NMs if they have been trained on a sufficiently large number of samples with varying properties such that they can learn which patterns are consistently present in the data. If the number of samples is small, it is very probable that some features co-occur with a certain toxicity outcome by chance. Once larger datasets are used, this randomly occurring similarities can be detected much easier. In addition, if the model shall be able to predict NM toxicity across various types of NMs, this heterogeneity should also be present in the training data. This property is referred to as the applicability domain of the model. If the training dataset is very homogeneous with respect to NM types or otherwise restricted to only subsets of the occurring ranges of certain physico-chemical properties, the applicability domain is expected to be small and the predictivity for NM types or properties not present in the training data will probably be poor.

In the world of chemicals, large-scale systematic datasets exist and led to the development of more powerful computational models. In 2007, the US EPA launched the ToxCast project[193] in which high-throughput screening techniques and computational toxicology approaches were applied to a large number of chemicals. The dataset contains data for more than 1,000 chemicals tested for more than 700 high-throughput assay endpoints covering various cell-based as well as biochemical *in vitro* assays. This dataset is publicly available and can be used for model development and testing. Together with other federal US agencies this initial dataset was then enlarged in the Tox21 program to more than 10,000 chemicals tested in approximately 50 assays. The data is stored in a MySQL database and can be accessed programmatically. Based on this data, it is possible to identify patterns in compound-induced biological responses, rank and prioritize chemicals or develop predictive toxicity models. While reaching this dimension of available FAIR data might not be realistic in near future for NMs, it still shows how powerful bioinformatics and ML tools can be if a sufficient number of systematic data is available.

In the field of NM toxicity, these large, reliable datasets with diversity in studied NM types are rarely available[77]. This is especially the case for datasets which study endpoints with regulatory relevance and comprise information on a sufficient number of physico-chemical properties to characterize the different NMs well enough. Typically, only few NMs are tested within one study due to time and financial constraints. Therefore, development of accurate and robust ML models in the field of NM toxicity is currently still a major challenge[194]. In order to successfully implement computational modeling approaches for NM grouping and toxicity prediction, sufficiently large high-quality datasets with respect to physico-chemical properties and effects to be modelled are urgently needed. Optimally, these datasets should also be present in a unified format and freely accessible in public databases[151].

In order to assemble large datasets suitable for ML modeling, data from different sources may need to be integrated. Here, data on different aspects of the same NM may need to be collected with the main challenge being the identification of different sources and corresponding matching of NM identifiers. In addition, one may also want to collect data on different NMs in which case comparability of the measurements is the most critical factor. Both cases will be discussed below.

4.2.2 Data quality

Another critical aspect with regards to obtaining reliable ML models is data quality. Only if measured data for physico-chemical properties, NM toxicity or any other parameter included in the model are accurately measured with low technical errors, the ML model can detect underlying patterns in a reliable manner. Reliability of measurements, expected technical or biological variability and potential other uncertainty factors are important in this regard. However, for NMs various challenges with respect to reliability of measurements for material characterization, a lack of validated toxicity assay as well as a missing understanding of MoAs are observed[195]. Criteria for estimating the quality of NM experimental data have been introduced for guiding scientists and modellers in judging reliability of data[196]. Integrating such data quality checks into the modeling process may at the same time improve the robustness of developed models.

4.2.3 Integration of data on one particular NM

As many factors have to be considered in order to fully characterize a NM including physico-chemical properties, interactions with the environment, different exposure scenarios and toxicity assays as well as omics measurements, information on one particular NM may be scattered. Thereby, data originating from different studies or projects may be stored in different databases or even be only existent in scientific literature. Unambiguous identification and characterization of NMs and missing unique representations or naming of NMs are the main challenge for the purpose of data integration. Unambiguous identification is mainly hampered by the fact that many physico-chemical properties can vary and therefore need to be assessed to find out whether two NMs are actually the same or not. As for conventional chemicals, properties like chemical composition, degree of purity and quantitative information on impurities or additives form the basis for NM characterization and identification[21]. In addition, the description of a NM also requires information on the number-based particle size distribution, surface functionalization or treatment, shape in terms of aspect ratio and particle morphology as well as the specific surface area. On top of that, physical properties like dissolution rate, state of agglomeration or aggregation and changes in surface chemistry as well as other higher-level parameters such as surface reactivity may also be relevant[27]. While some of these properties are intrinsic to the NM itself, others vary depending on the medium surrounding the NM and are thus extrinsic. Therefore, in order to obtain a reliable grouping, characterization has to be performed in the relevant biological medium used for toxicity testing.

In addition, unique naming is important for reuse and integration of data as it is required for mapping data from different sources. For NMs, no clear naming standards are agreed on in the community as their complex structure does not directly allow for trivial terminology. However, attempts have been made to establish a common notation, recently. Lynch et al.[197] introduced NInChIs which are line notations describing NMs and discriminating different NFs. NInChIs are machine-readable and thus allow easy identification and integration of data from various sources. Inclusion of this standard representation as a generally accepted and used identifier among the nanosafety but also material producing and material modeling communities is a challenging task. However, achieving this goal would have great impact on successfully integrating data and performing NM grouping.

4.2.4 Integration of data across NMs

Due to the complexity of the characterization of NMs and toxicity testing, it is usually the case that only a handful of NMs are considered within one study. In order to create sufficiently large datasets for ML modeling, integration of data across various NMs from multiple studies is therefore necessary. For robust modeling, comparability of measurements is critical with respect to combining data from different laboratories or projects. If different assays or standard operating procedures (SOPs) have been used, it is very difficult for a model to learn patterns on such data which hold additionally introduced differences. Here, careful characterization of dispersions and expected effective doses play an important role. A sufficient number of replicates and tested doses in case of toxicity assays are also a pre-requisite for developing robust models with low uncertainties. If datasets from different sources are sufficiently comparable and well-described in standard formats, they may be combined in modeling approaches which can largely improve the quality of the model. However, in the field of NMs, standardization of measurements is still an on-going process. While for many physico-chemical properties TGs have been adapted already or are currently being adapted, standardization of toxicity assays is still more challenging. Frequently, discrepancies and inconsistencies have been observed in the data. Main factors hampering comparability are difficulties in handling of NMs with respect to factors like dispersion and dosimetry as well as interactions with biological molecules in the surrounding of the NM. Also, systematically varied NMs and benchmark materials have not been included in many studies which renders implementation of NM grouping approaches and comparisons between studies challenging tasks.

One important step towards standardized measurements across projects are interlaboratory comparisons which assess whether or not similar results are obtained by different laboratories if they perform the same assay using the same SOP. That implies that processes need to be well-described in any detail and that local differences should not introduce large variations impacting results. Various projects have performed interlaboratory comparisons for NMs, e.g., GRACIOUS or NanoHarmony. Another important factor is community-wide acceptance of standardized protocols as only if researchers actually perform measurements according to SOPs, comparability of datasets can be achieved. As ML models are critically dependent on reliable and comparable underlying data, this is a major requirement for exploiting the full potential of these methods. The major challenge here is to agree on and establish common standards, terminologies and harmonized infrastructures across various stakeholders including researchers, regulators and industry.

4.2.5 Challenges with respect to single omics datasets

Omics may have great impact on risk assessment in the light of the current strategic shift towards NAMs. The major advantage is that they give detailed insights into MoAs and can inform AOPs. Due to highly advanced data sharing policies for omics data, a large amount of publicly available datasets exists. While only a small part of these omics datasets is related to NMs, integration of NM-specific omics datasets with those on other traits like chemicals, drugs or diseases may still support interpretation and unraveling of underlying MoAs. However, the use of omics data in the context of risk assessment in general as well as integration of public omics datasets are not trivial for various reasons.

First, omics analyses are prone to technical and biological noise due to the complexity of measurements and the exposed system. Also, the huge variability in applied analysis steps with regard to normalization, imputation or multivariate statistical or ML methods may lead to substantial differences in biomarker detection. Therefore, reproducibility of toxicogenomics signatures is a major concern which needs to be carefully investigated[198]. One important factor in this regard is statistical power which is directly related to the number of replicates in a study. Benchmark datasets for comparing different platforms, cell types, species and so on may support reproducibility.

In addition, dimensionality is a major concern. In omics analysis, the number of features is much larger than the number of replicate measurements. This large number of potential descriptors like single transcripts, proteins or metabolites is challenging in terms of false-positive or false-negative results. Therefore, computing FDRs and observation of consistent trends across pathways or other defined sets is of utmost importance for avoiding artifacts. For predictive modeling, methods like MaNGA[199] may be used to find the best minimal set of features with high predictivity and stability and a wide applicability domain using a multi-objective optimization strategy.

The advantage of omics is that molecular changes are measurable already few hours or days after treatment which renders it very efficient and appealing for predicting long-term effects compared to *in vivo* studies which may even take years. While single dose and time point analyses can inform on MoAs, deriving a Point of Departure or BMD requires testing of multiple doses. Variations of the investigated timepoint are also important to study kinetic patterns of molecular alterations and for gaining insights into potential chronic effects or recovery mechanisms. However, while in conventional toxicity studies at least multiple doses are frequently tested, for omics usually these

comprehensive assessments are not performed. Still, some approaches for estimating BMDs or PODs from omics data have been proposed[200, 201].

Another challenge is that it is currently not known in detail, which changes are indicative with respect to adverse effects and what are relevant signatures or patterns in the different omics layers. Especially for complex cases where one substance induces changes related to multiple hazard endpoints or if mixtures of substances are considered not much is known about underlying signatures. In addition, signatures may also include defense or repair mechanisms which need to be separated from the adverse ones. While this is partly covered during enrichment analysis, this knowledge is far from complete to date.

4.2.6 Challenges for integration of various omics datasets or layers

While analyzing single omics datasets is already not trivial, more challenges occur when datasets shall be combined. Omics datasets may be generated by different types of machines and analyzed in various ways with numerous commercial and open-source software tools and packages being available. Here, the variability strongly depends on the omics layer under study. In transcriptomics, data is usually derived from RNA-seq or formerly microarrays. While many variations in protocols and analyses can be observed, recommendations for best practices exist and standardization effort have been made[202, 203]. An additional advantage is the fact that in transcriptomics the complete or a specified fixed set of transcripts, respectively, is detected which facilitates comparisons between studies. Interlaboratory comparisons have shown good reproducibility of gene expression profiles induced upon exposure to toxicants. However, while the complete transcriptome is detected in RNA-seq experiments, not all transcripts will actually be expressed and yield functional proteins in the cell.

Other omics layers like proteomics and metabolomics are better suited to represent the actual functional changes due to NM treatment on molecular level. Despite the fact that results better reflect the real biological situation, standardization and comparability are much more complicated and less advanced for these fields[204]. Large-scale proteomics and metabolomics are frequently carried out using mass spectrometry (MS)-based approaches. The disadvantage in proteomics approaches is that usually peptides are measured and then mapped back to proteins. Thereby, one loses certain information on factors like isoforms. Proteomics also is not as close to the real phenotypes as for example metabolomics. This is mainly due to the fact that post-translational modifications are highly relevant for protein activity while in the main LC-MS/MS run only the

abundances of proteins are measured. Modifications can also be measured; however, this process is time-consuming. As mentioned before, comparisons between transcriptomics and proteomics are hampered by the fact that in transcriptomics usually a finite set of transcripts is measured while this is not the case for proteomics.

Instead, in metabolomics all small molecules in the cell shall be detected. These small molecules are end products of cellular regulatory processes and thus, this omics layer is closest to the phenotype. In addition, small molecules can easily distribute all over the body and therefore they can be secreted in various biofluids where easy sampling of such biomarkers is possible. However, standardization is complex and almost not tackled so far. Especially, quantification is a very challenging task in this regard.

Systems biology integrates information across various omics layers, thereby also considering interactions between them. Thus, systems biology yields a very broad understanding of NM-induced changes at the molecular level. However, integration of different omics layers is not trivial for several reasons: 1) Systematic studies assessing multiple omics layers under the very same conditions allow for direct comparisons. However, while in the field of NM toxicity some multi-omics studies have been published[182, 183] these studies are still rare; 2) Determination of the most suitable time point to be studied is essential. While effects in transcriptomics are visible comparatively fast after treatment, proteins have to be built first which requires time and thus changes on the proteomic level are seen much later. This has to be taken into account for integration of multiple layers. Thus, testing of multiple time points and doses is very useful but at the same time increases the complexity; 3) Specific databases allowing the deposition of data from multiple omics levels are lacking. Thus, if data from multi-omics experiments are publicly available, the data will be spread across databases and tools for data integration and automated querying are needed; and 4) Usually, interpretation of omics data is performed on the level of affected pathways or GO terms. While for single layers this is quite straightforward, tools integrating information from multiple omics layer during the enrichment analysis are still under development[205].

One additional challenge with respect to integration of omics datasets is the availability of metadata. While huge amounts of data are present in public databases, they are not always well annotated. In many cases, databases do not hold information on which samples belong to which study groups, which concentrations or time points were tested and so on. Sometimes this information can be found in the corresponding publications or their supplementary files. In other cases, the information is missing completely. Here, templates for the upload of omics data are of great use. For

transcriptomics and metabolomics, the OECD already defined rules on how to describe transcriptomics and metabolomics datasets in regulatory toxicology in their reporting frameworks[92, 93, 95]. For proteomics, no such document exists so far. These frameworks can be considered as a good starting point for making omics data ready for regulatory use. However, their benefit relies on the actual usage by researchers who upload their data to databases. In addition, AI may also support the curation of publicly available datasets by automatically extracting metadata from publications.

Overall, the use and integration of omics datasets are of great value for risk assessment but currently hampered by lack of standardization of measurements and harmonized analysis workflows as well as sufficient annotations. FAIRification of available data is thus urgently needed for omics data but also other data related to NM characterization and toxicity.

4.3 The need for FAIR data

Large-scale high-quality data is required for developing reliable computational approaches for risk assessment. However, in reality, studies usually investigate only a small number of NMs. Meta-analyses could strongly improve the data situation for model development. Unfortunately, original data is often not made publicly available or scattered across databases. This hampers proper meta-analysis as data is scarce and often not comparable due to differences in ontologies, metadata standards or missing nano-specific information. Here, making data compliant to the FAIR principles[206] would greatly support these efforts.

Data are findable if they are assigned to globally unique persistent identifiers. They should also be accompanied by rich metadata which can be found in a searchable resource. Accessibility is guaranteed by making sure that data and their corresponding metadata can be retrieved by their assigned identifier using standardized communication protocols like ‘http’ or ‘ftp’. Interoperability describes the fact that data and metadata should use a formal, accessible, shared and broadly applicable language for knowledge representation meaning that it should be human- as well as machine-readable. Finally, reusability relates to describing the data by metadata in such detail that other users can decide whether the data will actually fit their research goal and be of use. It also deals with data licenses, origin and citation issues as well as the question whether or not community standards or best practices for data sharing have been used. Various factors have to be considered in

order to harmonize NM-related data and make them FAIR and were discussed in previous publications[207, 208].

First of all, unique identifiers for NMs need to be developed. This issue has already been discussed above. The main question here is how to represent structurally complex NMs in a standardized machine-readable manner[77, 209]. Currently, only expert judgement in combination with comparisons of physico-chemical properties can be used to estimate whether two datasets actually described the same NM or not. First approaches for generating globally unique identifiers in terms of machine-readable structural representations were based on nano-InChIs[197, 210]. In addition, common terminology and ontologies are needed for integrating data in an automated way. While large efforts have been made to curate metadata in projects like NanoReg2, it became obvious that this is not reasonable without harmonized terminologies. The OECD Harmonized Templates are a great achievement providing standard formats for reporting on various aspects of risk assessment including some physico-chemical properties of NMs. Currently, the most comprehensive collection of information on NMs and NM toxicity exists in the eNanoMapper database[207]. The eNanoMapper database implements the nanosafety community-based eNanoMapper ontology and uses the ISA-TAB-NANO format[211] which is a standard data sharing format for NMs. This allows access to the data using specified ontology terms which is directly related to reporting standards for metadata which need to be implemented in the nanosafety community in order to reduce future data curation efforts. Importantly, links to specific SOPs and TGs should also be provided. In general, accessibility of data and metadata needs to be guaranteed. Currently, it is often necessary to contact experts directly in order to find data and obtain all necessary details regarding characterization and experimental setup. However, this is not efficient for large-scale modeling tasks. Instead, data along with sufficient metadata and detailed descriptions with respect to conditions under which they were derived should be stored in publicly available databases. Here, storing raw data as well as processed data is of additional value for re-interpretation of the data. As various datasets on the same material are expected to be stored in different fit-for-purpose databases, these should be well linked and interoperable to allow for retrieval and integration of related datasets[207]. Awareness of previously published data from different projects may be increased in that way as well.

Omics data and corresponding repositories are generally considered to be FAIR. In the field of omics data, standardized file formats and repositories are commonly used. Also, publications in this area require usage of standardized reporting formats and database upload to public repositories. However, in the nanosafety domain interoperability is hampered due to the lack of domain-specific metadata, ontologies and reporting standards accepted by the community. Thus, nanosafety

information and omics data can often not be linked. However, efforts have been made enable this interlinkage, for instance, a template for connecting nano-specific proteomics datasets obtained from the PRIDE archive to collected NM-specific metadata was generated and allows for upload to eNanoMapper thereby facilitating data reuse.

Overall, FAIR data related to nanosafety allow for sustainable reuse of publicly funded data for the purpose of developing *in silico* models and IATAs. Compliance to the FAIR principles will greatly improve data integration and thus enable the creation of comprehensive datasets covering various NMs, physico-chemical properties and omics data. With these datasets, development of robust models and tools supporting NM grouping will be largely facilitated. To further support the FAIRification of data for NMs and advanced materials, the Advanced Nano FAIR Implementation Network (https://www.go-fair.org/implementation-networks) was established during the EU project Gov4Nano.

4.4 *In vitro* to *in vivo* extrapolation

As mentioned in the introduction, it is critical to keep in mind the differences between applied and effective dose both *in vitro* and *in vivo*. Only if the effective doses are comparable, *in vitro* results will be relevant for drawing conclusions on the *in vivo* situation. Several factors play a role in this regard. Dispersion stability needs to be carefully checked before any *in vitro* treatment in order to make sure that NMs are presented to the cells as a homogeneous suspension in order to guarantee the reproducibility of the results. Dosimetry modeling for NMs is thereby much more complex compared to conventional chemicals as differences in particle density and agglomeration status may have large impact on the deposition of NMs in *in vitro* tests. Direct measurements of particle uptake *in vitro* are also critical as using fluorescence labeling might also change the properties of the NM and thereby its uptake behavior and/ or toxicity[212]. In addition, the choice of the cell model is of high relevance[213]. While some studies show good agreement between *in vitro* and *in vivo* results, it is expected that differences mainly result from the fact that whole organism responses are compared to single, isolated cell types. This is problematic as interactions between cell types cannot be sufficiently modeled and intrinsic circulation of NMs in the body cannot be mimicked. Advanced co-culture systems and 3D models may improve this situation by providing more realistic surrogates. Also, the test methods must be checked for reliability taking into account possible interferences of the NM with the assay read-out. Another question that arises is how reliable the true outcome is in terms of

predictivity for humans. This relates to the fact that not all toxicity mechanisms are conserved across different species and the ADME behavior might also vary between tested animals and humans. However, the percentage of materials for which this is expected to be the case is rather low. For obtaining realistic concentrations reflecting the *in vivo* exposure of a certain tissue, PBPK modeling should optimally be considered jointly with the predictive model in the context of NM grouping.

While acute effects are much easier to detect, chronic effects pose a larger challenge. *In vitro* assays are not suitable for directly predicting chronic effects. However, acute mechanistic responses may still be predictive for chronic effects. In general, oxidative stress and inflammation may be used as indicators for chronic diseases which may be developed upon prolonged exposure to the NM. In addition, some clinical biomarkers are known to be detectable also at the beginning of disease progression as they reflect certain disease mechanisms which contribute to pathogenesis, irrespective of the stage. These changes may also be well reflected in *in vitro* studies. In addition, for NMs, usually only STIS has been performed. However, manifestation of some adverse effects may actually need more time and may not be visible in STIS. Thus, actual chronic effects have not been studied *in vivo* as well and can only be concluded from STIS.

For omics analyses, a few additional factors have to be considered. Poulsen et al.[214] have shown that comparing individual genes between *in vivo* and *in vitro* led to low correlations. Instead, they recommended comparisons on pathway level as similarity was found to be much higher between *in vivo* and *in vitro*. Similarly, Kinaret et al.[215] found transcriptomic similarities on the functional level between mouse lungs *in vivo* and human macrophage cell line results after NM exposure when data were properly analyzed and NM properties are taken into account. In addition, NMs may also trigger non-specific cellular effects like adaptive responses which are not indicative of adverse outcomes.

4.5 Regulatory acceptance and needs

The urgent need for a paradigm shift towards NAMs and mechanistic understanding on adverse effects induced by NMs has been recognized also in regulation as NAMs can enhance risk assessment in terms of efficiency and potentially also accuracy while at the same time reducing animal testing and increasing the understanding of toxicity mechanisms. Regulatory bodies like ECHA emphasize on the integration of NAMs into risk assessment. However, currently a number of hurdles still need to

be overcome in order to successfully integrate NAMs, and especially *in silico* tools and omics data into regulatory risk assessment.

In general, to gain regulatory acceptance, validation of models and tools is necessary. Here, it needs to be demonstrated that methods under consideration are reliable, relevant and predictive. For grouping and read-across, REACH also requires robust scientific justifications and adequate and reliable documentation for the applied methods. First of all, standardization of models and underlying data is needed. With respect to data quality, measurement techniques for physico-chemical properties, toxicological endpoints as well as omics data must use standardized measurement and evaluation protocols and must clearly state the details of the measurement procedure in order to be traceable and reproducible. Quality measures also need to be introduced and data sharing practices should be established in order to ensure consistency and reliability of such data. In addition, developed models need to be robust, accurate and reproducible. Inclusion of benchmark materials may aid in assessing the performance of *in silico* models across datasets. Transparency and interpretability of models is another important factor allowing to understand decisions and predictions made in the model. At the same time, the models should also show biological relevance. Therefore, integration of domain expertise and mechanistic insights are important to ensure biological plausibility. Also, quantification of uncertainties of models and their communication to regulators and stakeholders is crucial for regulatory decision-making. Establishing a robust and sustainable infrastructure will also be needed to allow for integration of systems biological approaches into regulatory testing. To date, for most of these criteria, implementation is still work-in-progress.

Chapter 5: Conclusion and outlook

In this thesis, the value of ML tools and bioinformatics for supporting NM grouping approaches was assessed. It was shown that ML models are very useful for extracting the most important properties describing NM toxicity in an automated and objective fashion. However, in order to build robust ML models, large high-quality datasets in terms of the number of tested NMs are urgently needed.

This directly relates to the largest challenge of the first study described here. In this study, a RF model was used for predicting results from STIS and the macrophage assay based on physico-chemical properties of the NMs. While the accuracy of the developed model was quite good, it already became clear that the number of tested materials is a major limitation with respect to robustness. In case of NMs, a large number of physico-chemical properties is needed to sufficiently cover their high complexity considering that not only intrinsic properties are relevant but also extrinsic ones varying with the surrounding environment or during the life cycle of NMs. All these properties may play an important role for toxicity. Thus, description of each NM would require a large number of measurements which is very time-consuming. Especially for extrinsic properties, measuring them in a reliable and standardized way is not straightforward as assays need to be performed in relevant media which usually causes severe interferences. Difficulties with NM dispersion and computation of effective doses further impair integration of available datasets. Overall, it became clear that establishing NM grouping approaches on physico-chemical properties is not trivial.

The OP of NMs seems to be a well-suited surrogate variable reducing the need for extensive characterization of the NMs, which might allow for the measurement of larger sets of NMs. However, different OP assays exist and need to be studied in terms of their comparability. This was assessed in the second study. Here, different dose metrics were compared, correlations between assays were computed and logistic regression was used to assess predictivity of single assays and assay combinations. In was found that different OP assays show only moderate correlations and that combining multiple assays improved predictivity. However, due to only moderate correlations, integration of datasets from different studies for obtaining sufficiently large datasets for modeling cannot easily be performed. Also, while the OP is considered important for describing how NMs can induce toxicity, other MoAs not related to the OP exist and are not properly represented in OP assays.

Thus, including information on MoAs in terms of omics data is a promising approach. So far, the knowledge on MoAs of NMs is still limited. Especially in case of proteomics, only few NM-specific datasets are available. The advantage with omics data, however, is that NM-specific datasets may directly be integrated with those from other traits like chemicals, drugs or diseases. Thereby, generating larger datasets suitable for ML modeling is facilitated. However, integration of datasets for meta-analyses requires harmonization of analysis workflows and differences in measurement techniques need to be taken into account in order to make results comparable. While for transcriptomics large efforts have been made in terms of standardization, this is not true for proteomics. Therefore, as a first step towards harmonized meta-analysis of proteomics data, PROTEOMAS was introduced in the third study. PROTEOMAS is a harmonized workflow for evaluating proteomics data from public repositories. Its value for comparing NM-specific proteomics signatures to those of other traits like chemicals, drugs or diseases has been shown in a case study on 25 proteomics datasets.

Finally, other computational approaches were reviewed in order to conclude on general challenges and further steps needed. A large number of approaches can be found in literature and have proven to be very valuable for gaining insights into NM toxicity. However, various challenges are still hampering the development of robust models. Data availability and quality were the most critical challenges in this regard. Without sufficient data, developing robust ML models which can generalize well across a large number of NMs is not feasible. As the number of studied NMs in each study is naturally very limited due to the complexity of necessary measurements, integration of results from different studies and projects will be necessary for any robust modeling approaches. This integration is only possible if data is comparable. To achieve comparability, data should comply with the FAIR data principles for data sharing and reuse. The main challenges for NMs with respect to FAIR data are the generation of unified naming schemes, the development and use of suitable SOPs and the provision of sufficient metadata describing experiments.

AI-based tools based on transfer learning, large language models or other techniques are expected to be indispensable for efficient data handling in risk assessment of NMs and in the context of ML model development and omics data analysis supporting NM grouping. Apart from predictive modeling described here, other tasks related to NM risk assessment may be largely profit from available AI tools. As an example, data curation efforts may be accelerated by AI. If performed manually, this process is highly time-consuming. On the other hand, AI models may quickly detect and potentially also fill data gaps in registration dossiers based on historical data thereby optimizing regulatory processes. In addition, automated extraction of relevant information from sources like

research publications can also be performed using, for instance, large language models. This may aid FAIRification of toxicological data as well as metadata of public databases. Also linking data stored in different databases may be facilitated by AI. In the context of risk assessment in general, toxicity information may be automatically linked to other information like physico-chemical characterization, intended use, exposure scenarios or omics data which might be spread across several databases without proper links between them. Additionally, AI models may be capable of evaluating multiple risk factors and dependencies between them simultaneously which may be useful in case of evaluating complex mixtures of NMs or chemicals.

Recent developments in the field of AI including ML as well as continuous improvements in the field of omics data analysis together with on-going efforts in standardization of assays and FAIRification of available NM-related data are expected to be pointing the way to the future for efficient NM risk assessment supported by computational tools.

References

1. EC, Commission Recommendation of 10 June 2022 on the definition of nanomaterial (Text with EEA relevance) 2022/C 229/01. **2022**.

2. Ghosh, C. K., Quantum Effect on Properties of Nanomaterials. In *Introduction to Nano: Basics to Nanoscience and Nanotechnology*, Sengupta, A.; Sarkar, C. K., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2015; pp 73-111.

3. Stark, W. J.; Stoessel, P. R.; Wohlleben, W.; Hafner, A., Industrial applications of nanoparticles. *Chemical Society Reviews* **2015,** *44* (16), 5793-5805.

4. Alhashmi Alamer, F.; Almalki, G. A., Fabrication of Conductive Fabrics Based on SWCNTs, MWCNTs and Graphene and Their Applications: A Review. *Polymers (Basel)* **2022,** *14* (24).

5. Sánchez-López, E.; Gomes, D.; Esteruelas, G.; Bonilla, L.; Lopez-Machado, A. L.; Galindo, R.; Cano, A.; Espina, M.; Ettcheto, M.; Camins, A.; Silva, A. M.; Durazzo, A.; Santini, A.; Garcia, M. L.; Souto, E. B., Metal-Based Nanoparticles as Antimicrobial Agents: An Overview. *Nanomaterials (Basel)* **2020,** *10* (2).

6. Mondal, S., Nanomaterials for UV protective textiles. *Journal of Industrial Textiles* **2022,** *51* (4_suppl), 5592S-5621S.

7. Sheraz, M.; Choi, B.; Kim, J., Enhancing Textile Water Repellency with Octadecyltrichlorosilane (OTS) and Hollow Silica Nanoparticles. *Polymers* **2023,** *15* (20), 4065.

8. Smijs, T. G.; Pavel, S., Titanium dioxide and zinc oxide nanoparticles in sunscreens: focus on their safety and effectiveness. *Nanotechnol Sci Appl* **2011,** *4*, 95-112.

9. Lu, P.-J.; Huang, S.-C.; Chen, Y.-P.; Chiueh, L.-C.; Shih, D. Y.-C., Analysis of titanium dioxide and zinc oxide nanoparticles in cosmetics. *Journal of Food and Drug Analysis* **2015,** *23* (3), 587-594.

10. Perera, K. Y.; Hopkins, M.; Jaiswal, A. K.; Jaiswal, S., Nanoclays-containing bio-based packaging materials: properties, applications, safety, and regulatory issues. *Journal of Nanostructure in Chemistry* **2023**.

11. Elkodous, M. A.; Hamad, H. A.; Maksoud, M. I. A. A.; Ali, G. A. M.; Abboubi, M. E.; Bedir, A. G.; Eldeeb, A. A.; Ayed, A. A.; Gargar, Z.; Zaki, F. S.; Farage, D. A. M.; Matsuda, A.; Abdelnour, M. R.; Sabra, B. A.; Elsayed, A.; Abdelrazek, T. A.; Abdelhameed, S. T.; Gharieb, M. A.; Rabee, M. M.; Aboeldiar, S. A.; Abdo, N. A.; Elwakeel, A. M.; Mahmoud, A. S.; Elsaid, M. M. M.; Omar, W. M.; Hania, R. A.; Mahmoud, N. G.; Elsayed, A. S. S.; Mohamed, T. M.; Sewidan, M. A.; Sayed, M. A. M.; Abbas, A. A.; Elsayed, A. H.; Alazab, A. M.; Basyooni, M. A.; Magdy, M.; Mashaly, E. A.; Salem, O. M.; Saber, S.; Hafez, A. A.; Tan, W. K.; Kawamura, G., Cutting-edge development in waste-recycled nanomaterials for energy storage and conversion applications. *Nanotechnology Reviews* **2022,** *11* (1), 2215-2294.

12. Patra, J. K.; Das, G.; Fraceto, L. F.; Campos, E. V. R.; Rodriguez-Torres, M. d. P.; Acosta-Torres, L. S.; Diaz-Torres, L. A.; Grillo, R.; Swamy, M. K.; Sharma, S.; Habtemariam, S.; Shin, H.-S., Nano based drug delivery systems: recent developments and future prospects. *J Nanobiotechnology* **2018,** *16* (1), 71.

13. Siddique, S.; Chow, J. C. L., Application of Nanomaterials in Biomedical Imaging and Cancer Therapy. *Nanomaterials* **2020,** *10* (9), 1700.

14. Nakagawa, T.; Gonda, K.; Kamei, T.; Cong, L.; Hamada, Y.; Kitamura, N.; Tada, H.; Ishida, T.; Aimiya, T.; Furusawa, N.; Nakano, Y.; Ohuchi, N., X-ray computed tomography imaging of a tumor with high sensitivity using gold nanoparticles conjugated to a cancer-specific antibody via polyethylene glycol chains on their surface. *Science and Technology of Advanced Materials* **2016,** *17* (1), 387-397.

15. Hasan, A.; Morshed, M.; Memic, A.; Hassan, S.; Webster, T. J.; Marei, H. E., Nanoparticles in tissue engineering: applications, challenges and prospects. *Int J Nanomedicine* **2018,** *13*, 5637-5655.

16. Saleem, H.; Zaidi, S. J., Developments in the Application of Nanomaterials for Water Treatment and Their Impact on the Environment. *Nanomaterials* **2020,** *10* (9), 1764.

17. Krishna, R. H.; Chandraprabha, M. N.; Samrat, K.; Krishna Murthy, T. P.; Manjunatha, C.; Kumar, S. G., Carbon nanotubes and graphene-based materials for adsorptive removal of metal ions – A review on surface functionalization and related adsorption mechanism. *Applied Surface Science Advances* **2023,** *16*, 100431.

18. ECHA, How to prepare registration dossiers that cover nanoforms: best practices. **2017**.

19. Nel, A. E.; Mädler, L.; Velegol, D.; Xia, T.; Hoek, E. M. V.; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M., Understanding biophysicochemical interactions at the nano–bio interface. *Nature Materials* **2009,** *8* (7), 543-557.

20. Meng, H.; Leong, W.; Leong, K. W.; Chen, C.; Zhao, Y., Walking the line: The fate of nanomaterials at biological barriers. *Biomaterials* **2018,** *174*, 41-53.

21. EC, 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). **2006**.

22. EC, Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 con-cerning the placing of plant protection products on the market. **2009**.

23. EC, Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. **2009**.

24. EC, Regulation (EC) No 1935/2004 of the European Parliament and of the Council of 27 October 2004 on materials and articles intended to come into contact with food and repealing Directives 80/590/EEC and 89/109/EEC. **2004**.

25. EC, Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys (Text with EEA relevance) **2009**.

26. EC, Regulation (EU) No 1007/2011 – fibre names labelling and the marking of the fibre composition of textile products. **2011**.

27. EC, EU 2018/1881 amending Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). **2018**.

28. ECHA, Appendix for Nanoforms Applicable to the Guidance on Registration and Substance Identification. ECHA Helsinki, Finland: 2019.

29. Dankovic, D. A.; Naumann, B. D.; Maier, A.; Dourson, M. L.; Levy, L. S., The Scientific Basis of Uncertainty Factors Used in Setting Occupational Exposure Limits. *J Occup Environ Hyg* **2015,** *12 Suppl 1* (sup1), S55-68.

30. Yang, W.; Peters, J. I.; Williams, R. O., 3rd, Inhaled nanoparticles--a current review. *Int J Pharm* **2008,** *356* (1-2), 239-47.

31. Oberdörster, G.; Oberdörster, E.; Oberdörster, J., Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. *Environ Health Perspect* **2005,** *113* (7), 823-39.

32. Praphawatvet, T.; Peters, J. I.; Williams, R. O., Inhaled nanoparticles–An updated review. *International Journal of Pharmaceutics* **2020,** *587*, 119671.

33. Gosens, I.; Post, J. A.; de la Fonteyne, L. J.; Jansen, E. H.; Geus, J. W.; Cassee, F. R.; de Jong, W. H., Impact of agglomeration state of nano- and submicron sized gold particles on pulmonary inflammation. *Part Fibre Toxicol* **2010,** *7* (1), 37.

34. Morfeld, P.; Treumann, S.; Ma-Hock, L.; Bruch, J.; Landsiedel, R., Deposition behavior of inhaled nanostructured TiO2 in rats: fractions of particle diameter below 100 nm (nanoscale) and the slicing bias of transmission electron microscopy. *Inhalation Toxicology* **2012,** *24* (14), 939-951.

35. Geiser, M., Morphological aspects of particle uptake by lung phagocytes. *Microscopy Research and Technique* **2002,** *57* (6), 512-522.

36. Clarke, S. W.; Pavia, D., Lung Mucus Production and Mucociliary Clearance: Methods of Assessment. In *Methods in Clinical Pharmacology—Respiratory System*, Howell, J. B. L.; Tattersfield, A. E., Eds. Macmillan Education UK: London, 1981; pp 85-94.

37. Kirby, A. C.; Coles, M. C.; Kaye, P. M., Alveolar macrophages transport pathogens to lung draining lymph nodes. *J Immunol* **2009,** *183* (3), 1983-9.

38. Veith, L.; Vennemann, A.; Breitenstein, D.; Engelhard, C.; Hagenhoff, B.; Wiemann, M. Distribution of Paramagnetic Fe2O3/SiO2–Core/Shell Nanoparticles in the Rat Lung Studied by Time-of-Flight Secondary Ion Mass Spectrometry: No Indication for Rapid Lipid Adsorption *Nanomaterials*, 2018.

39. Harfoush, S.; Hannig, M.; Le, D.-D.; Heck, S.; Leitner, M.; Omlor, A.; Tavernaro, I.; Kraegeloh, A.; Kautenburger, R.; Kickelbick, G.; Beilhack, A.; Bischoff, M.; Nguyen, J.; Sester, M.; Bals, R.; Dinh, Q., High-dose intranasal application of titanium dioxide nanoparticles induces the systemic uptakes and allergic airway inflammation in asthmatic mice. *Respiratory Research* **2020,** *21*.

40. MacParland, S. A.; Tsoi, K. M.; Ouyang, B.; Ma, X.-Z.; Manuel, J.; Fawaz, A.; Ostrowski, M. A.; Alman, B. A.; Zilman, A.; Chan, W. C. W.; McGilvray, I. D., Phenotype Determines Nanoparticle Uptake by Human Macrophages from Liver and Blood. *ACS Nano* **2017,** *11* (3), 2428-2443.

41. Walkey, C. D.; Olsen, J. B.; Guo, H.; Emili, A.; Chan, W. C. W., Nanoparticle Size and Surface Chemistry Determine Serum Protein Adsorption and Macrophage Uptake. *Journal of the American Chemical Society* **2012,** *134* (4), 2139-2147.

42. Fröhlich, E.; Salar-Behzadi, S., Toxicological Assessment of Inhaled Nanoparticles: Role of in Vivo, ex Vivo, in Vitro, and in Silico Studies. *International Journal of Molecular Sciences* **2014,** *15* (3), 4795-4822.

43. Mühlfeld, C.; Rothen-Rutishauser, B.; Blank, F.; Vanhecke, D.; Ochs, M.; Gehr, P., Interactions of nanoparticles with pulmonary structures and cellular responses. *American journal of physiology. Lung cellular and molecular physiology* **2008,** *294*, L817-29.

44. Bernhard, W., Lung surfactant: Function and composition in the context of development and respiratory physiology. *Annals of Anatomy - Anatomischer Anzeiger* **2016,** *208*, 146-150.

45. Detampel, P.; Ganguly, A.; Tehranian, S.; Green, F.; Singha, S.; Santamaria, P.; Jeje, A. A.; Cho, C. S.; Petri, B.; Amrein, M. W., In vivo clearance of nanoparticles by transcytosis across alveolar epithelial cells. *PLOS ONE* **2019,** *14* (9), e0223339.

46. Kreyling, W. G.; Hirn, S.; Schleh, C., Nanoparticles in the lung. *Nature Biotechnology* **2010,** *28* (12), 1275-1276.

47. Kreyling, W. G.; Semmler-Behnke, M.; Takenaka, S.; Möller, W., Differences in the Biokinetics of Inhaled Nano- versus Micrometer-Sized Particles. *Accounts of Chemical Research* **2013,** *46* (3), 714-722.

48. Stone, V.; Miller, M. R.; Clift, M. J. D.; Elder, A.; Mills, N. L.; Møller, P.; Schins, R. P. F.; Vogel, U.; Kreyling, W. G.; Alstrup Jensen, K.; Kuhlbusch, T. A. J.; Schwarze, P. E.; Hoet, P.; Pietroiusti, A.; De Vizcaya-Ruiz, A.; Baeza-Squiban, A.; Teixeira, J. P.; Tran, C. L.; Cassee, F. R., Nanomaterials Versus Ambient Ultrafine Particles: An Opportunity to Exchange Toxicology Knowledge. *Environ Health Perspect* **2017,** *125* (10), 106002.

49. Kreyling, W. G.; Semmler-Behnke, M.; Seitz, J.; Scymczak, W.; Wenk, A.; Mayer, P.; Takenaka, S.; Oberdörster, G., Size dependence of the translocation of inhaled iridium and carbon nanoparticle aggregates from the lung of rats to the blood and secondary target organs. *Inhal Toxicol* **2009,** *21 Suppl 1*, 55-60.

50. Moreno-Horn, M.; Gebel, T., Granular biodurable nanomaterials: No convincing evidence for systemic toxicity. *Crit Rev Toxicol* **2014,** *44* (10), 849-75.

51. Whitsett, J. A.; Alenghat, T., Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nat Immunol* **2015,** *16* (1), 27-35.

52. Barnes, P. J., Immunology of asthma and chronic obstructive pulmonary disease. *Nature Reviews Immunology* **2008,** *8* (3), 183-192.

53. Murdoch, J. R.; Lloyd, C. M., Chronic inflammation and asthma. *Mutat Res* **2010,** *690* (1-2), 24-39.

54. Morrow, P. E., Possible mechanisms to explain dust overloading of the lungs. *Fundamental and Applied Toxicology* **1988,** *10* (3), 369-384.

55. Oberdörster, G.; Ferin, J.; Gelein, R.; Soderholm, S. C.; Finkelstein, J., Role of the alveolar macrophage in lung injury: studies with ultrafine particles. *Environ Health Perspect* **1992,** *97*, 193-9.

56. Tang, J.; Xiong, L.; Wang, S.; Wang, J.; Liu, L.; Li, J.; Yuan, F.; Xi, T., Distribution, translocation and accumulation of silver nanoparticles in rats. *J Nanosci Nanotechnol* **2009,** *9* (8), 4924-32.

57. Shimada, A.; Kawamura, N.; Okajima, M.; Kaewamatawong, T.; Inoue, H.; Morita, T., Translocation pathway of the intratracheally instilled ultrafine particles from the lung into the blood circulation in the mouse. *Toxicol Pathol* **2006,** *34* (7), 949-57.

58. Ryman-Rasmussen, J. P.; Tewksbury, E. W.; Moss, O. R.; Cesta, M. F.; Wong, B. A.; Bonner, J. C., Inhaled multiwalled carbon nanotubes potentiate airway fibrosis in murine allergic asthma. *Am J Respir Cell Mol Biol* **2009,** *40* (3), 349-58.

59. Juan, C. A.; Pérez de la Lastra, J. M.; Plou, F. J.; Pérez-Lebeña, E., The Chemistry of Reactive Oxygen Species (ROS) Revisited: Outlining Their Role in Biological Macromolecules (DNA, Lipids and Proteins) and Induced Pathologies. *Int J Mol Sci* **2021,** *22* (9).

60. Hellack, B.; Nickel, C.; Albrecht, C.; Kuhlbusch, T. A. J.; Boland, S.; Baeza-Squiban, A.; Wohlleben, W.; Schins, R. P. F., Analytical methods to assess the oxidative potential of nanoparticles: a review. *Environmental Science: Nano* **2017,** *4* (10), 1920-1934.

61. Sharifi, S.; Behzadi, S.; Laurent, S.; Forrest, M. L.; Stroeve, P.; Mahmoudi, M., Toxicity of nanomaterials. *Chemical Society Reviews* **2012,** *41* (6), 2323-2343.

62. Mortezaee, K.; Najafi, M.; Samadian, H.; Barabadi, H.; Azarnezhad, A.; Ahmadi, A., Redox interactions and genotoxicity of metal-based nanoparticles: A comprehensive review. *Chemico-Biological Interactions* **2019,** *312*, 108814.

63. World Health Organisation, Determination of airborne fibre number concentrations: A recommended method by phase constrast optical microscopy. *World Health Organisation, Geneva* **1997**.

64. Fortini, R.; Meyer-Plath, A.; Kehren, D.; Gernert, U.; Jácome, L. A.; Sturm, H., Measurement of Flexural Rigidity of Multi-Walled Carbon Nanotubes by Dynamic Scanning Electron Microscopy. *Fibers* **2020,** *8* (5), 31.

65. Donaldson, K.; Murphy, F. A.; Duffin, R.; Poland, C. A., Asbestos, carbon nanotubes and the pleural mesothelium: a review of the hypothesis regarding the role of long fibre retention in the parietal pleura, inflammation and mesothelioma. *Part Fibre Toxicol* **2010,** *7*, 5.

66. Lippmann, M., Toxicological and epidemiological studies on effects of airborne fibers: Coherence and public health implications. *Critical Reviews in Toxicology* **2014,** *44* (8), 643-695.

67. Murphy, F. A.; Schinwald, A.; Poland, C. A.; Donaldson, K., The mechanism of pleural inflammation by long carbon nanotubes: interaction of long fibres with macrophages stimulates them to amplify pro-inflammatory responses in mesothelial cells. *Part Fibre Toxicol* **2012,** *9*, 8.

68. Gaudino, G.; Xue, J.; Yang, H., How asbestos and other fibers cause mesothelioma. *Transl Lung Cancer Res* **2020,** *9* (Suppl 1), S39-s46.

69. Broßell, D.; Meyer-Plath, A.; Kaempf, K.; Plitzko, S.; Wohlleben, W.; Stahlmecke, B.; Wiemann, M.; Haase, A., A human risk banding scheme for high aspect-ratio materials. In *Synthetic Nano- and Microfibers*, **2020**.

70. Kane, A. B.; Hurt, R. H.; Gao, H., The asbestos-carbon nanotube analogy: An update. *Toxicology and Applied Pharmacology* **2018,** *361*, 68-80.

71. BAuA, Proposal for Harmonised Classification and Labelling Multi-Walled Carbon Tubes (synthetic graphite in tubular shape) with a geometric tube diameter range ≥ 30 nm to < 3 µm and a length ≥ 5 µm and aspect ratio ≥ 3:1, including Multi-Walled Carbon Nanotubes, MWC(N)T. *European Chemical Agency (ECHA)* **2021,** *2.0*, p. 1-111.

72. Gao, X.; Tang, M.; Li, Z.; Zha, Y.; Cheng, G.; Yin, S.; Chen, J.; Ruan, D.-y.; Chen, L.; Wang, M., Streptavidin-conjugated CdSe/ZnS quantum dots impaired synaptic plasticity and spatial memory process. *Journal of Nanoparticle Research* **2013,** *15* (4), 1575.

73. Wu, T.; Zhang, T.; Chen, Y.; Tang, M., Research advances on potential neurotoxicity of quantum dots. *J Appl Toxicol* **2016,** *36* (3), 345-51.

74. Tang, H.-Q.; Xu, M.; Rong, Q.; Jin, R.-W.; Liu, Q.-J.; Li, Y.-L., The effect of ZnO nanoparticles on liver function in rats. *International Journal of Nanomedicine* **2016,** *11* (null), 4275-4285.

75. Shukla, R. K.; Kumar, A.; Gurbani, D.; Pandey, A. K.; Singh, S.; Dhawan, A., TiO2 nanoparticles induce oxidative DNA damage and apoptosis in human liver cells. *Nanotoxicology* **2013,** *7* (1), 48-60.

76. Gaiser, B. K.; Hirn, S.; Kermanizadeh, A.; Kanase, N.; Fytianos, K.; Wenk, A.; Haberl, N.; Brunelli, A.; Kreyling, W. G.; Stone, V., Effects of silver nanoparticles on the liver and hepatocytes in vitro. *Toxicological Sciences* **2013,** *131* (2), 537-547.

77. Haase, A.; Klaessig, F.; Nymark, P.; Paul, K.; Greco, D., *EU US Roadmap Nanoinformatics 2030*. 2018.

78. Netzeva, T. I.; Worth, A.; Aldenberg, T.; Benigni, R.; Cronin, M. T.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D.; Schultz, T.; Stanton, D. W.; van de Sandt, J. J.; Tong, W.; Veith, G.; Yang, C., Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern Lab Anim* **2005,** *33* (2), 155-73.

79. Roy, K.; Kar, S.; Ambure, P., On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems* **2015,** *145*, 22-29.

80. Winkler, D. A., Recent advances, and unresolved issues, in the application of computational modelling to the prediction of the biological effects of nanomaterials. *Toxicology and Applied Pharmacology* **2016,** *299*, 96-100.

81. Burello, E., Review of (Q)SAR models for regulatory assessment of nanomaterials risks. *NanoImpact* **2017,** *8*, 48-58.

82. Deepika, D.; Kumar, V., The Role of "Physiologically Based Pharmacokinetic Model (PBPK)" New Approach Methodology (NAM) in Pharmaceuticals and Environmental Chemical Risk Assessment. *Int J Environ Res Public Health* **2023,** *20* (4).

83. Abdelsattar, A. S.; Dawoud, A.; Helal, M. A., Interaction of nanoparticles with biological macromolecules: a review of molecular docking studies. *Nanotoxicology* **2021,** *15* (1), 66-95.

84. Mancardi, G.; Mikolajczyk, A.; Annapoorani, V. K.; Bahl, A.; Blekos, K.; Burk, J.; Çetin, Y. A.; Chairetakis, K.; Dutta, S.; Escorihuela, L.; Jagiello, K.; Singhal, A.; van der Pol, R.; Bañares, M. A.; Buchete, N.-V.; Calatayud, M.; Dumit, V. I.; Gardini, D.; Jeliazkova, N.; Haase, A.; Marcoulaki, E.; Martorell, B.; Puzyn, T.; Agur Sevink, G. J.; Simeone, F. C.; Tämm, K.; Chiavazzo, E., A computational view on nanomaterial intrinsic and extrinsic features for nanosafety and sustainability. *Materials Today* **2023,** *67*, 344-370.

85. ECHA, *Non-animal approaches – Current status of regulatory applicability under the REACH, CLP and biocidal products regulations*. European Chemicals Agency: 2017.

86. Burgdorf, T.; Piersma, A. H.; Landsiedel, R.; Clewell, R.; Kleinstreuer, N.; Oelgeschläger, M.; Desprez, B.; Kienhuis, A.; Bos, P.; de Vries, R.; de Wit, L.; Seidle, T.; Scheel, J.; Schönfelder, G.; van Benthem, J.; Vinggaard, A. M.; Eskes, C.; Ezendam, J., Workshop on the validation and regulatory acceptance of innovative 3R approaches in regulatory toxicology - Evolution versus revolution. *Toxicol In Vitro* **2019,** *59*, 1-11.

87. Hartung, T.; van Vliet, E.; Jaworska, J.; Bonilla, L.; Skinner, N.; Thomas, R., Systems Toxicology. *ALTEX - Alternatives to animal experimentation* **2012,** *29* (2), 119-128.

88. Sewell, F.; Corvaro, M.; Andrus, A.; Burke, J.; Daston, G.; Delaney, B.; Domoradzki, J.; Forlini, C.; Green, M. L.; Hofmann, T.; Jäckel, S.; Lee, M. S.; Temerowski, M.; Whalley, P.; Lewis, R., Recommendations on dose level selection for repeat dose toxicity studies. *Arch Toxicol* **2022,** *96* (7), 1921-1934.

89. Oberdörster, G., Nanotoxicology: In vitro-in vivo dosimetry. *Environmental Health Perspectives* **2012,** *120* (1), A13.

90. Sauer, U.; Deferme, L.; Gribaldo, L.; Hackermüller, J.; Tralau, T.; Van Ravenzwaay, B.; Yauk, C.; Poole, A.; Tong, W.; Gant, T., The challenge of the application of 'omics technologies in chemicals risk assessment: Background and outlook. *Regulatory Toxicology and Pharmacology* **2017,** *91 Suppl 1*.

91. Edgar, R.; Domrachev, M.; Lash, A. E., Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **2002,** *30* (1), 207-10.

92. OECD, Transcriptomic Reporting Framework (TRF), www.oecd.org/chemicalsafety/testing/transcriptomic-reporting-framework.pdf. **2021**.

93. Harrill, J. A.; Viant, M. R.; Yauk, C. L.; Sachana, M.; Gant, T. W.; Auerbach, S. S.; Beger, R. D.; Bouhifd, M.; O'Brien, J.; Burgoon, L.; Caiment, F.; Carpi, D.; Chen, T.; Chorley, B. N.; Colbourne, J.; Corvi, R.; Debrauwer, L.; O'Donovan, C.; Ebbels, T. M. D.; Ekman, D. R.; Faulhammer, F.; Gribaldo, L.; Hilton, G. M.; Jones, S. P.; Kende, A.; Lawson, T. N.; Leite, S. B.; Leonards, P. E. G.; Luijten, M.; Martin, A.; Moussa, L.; Rudaz, S.; Schmitz, O.; Sobanski, T.; Strauss, V.; Vaccari, M.; Vijay, V.; Weber, R. J. M.; Williams, A. J.; Williams, A.; Thomas, R. S.; Whelan, M., Progress towards an OECD reporting framework for transcriptomics and metabolomics in regulatory toxicology. *Regulatory Toxicology and Pharmacology* **2021,** *125*, 105020.

94. Perez-Riverol, Y.; Bai, J.; Bandla, C.; García-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, Deepti J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaíno, Juan A., The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research* **2021,** *50* (D1), D543-D552.

95. OECD, Metabolomics Reporting Framework (MRF), www.oecd.org/chemicalsafety/testing/metabolomics-reporting-framework.pdf. **2021**.

96. SCCS, SCCS/1628/21 - The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation, 11th revision. **2021**.

97. OECD, Guidance document on the reporting of defined approaches to be used within integrated approaches to testing and assessment. **2016**.

98. Stone, V.; Gottardo, S.; Bleeker, E. A. J.; Braakhuis, H.; Dekkers, S.; Fernandes, T.; Haase, A.; Hunt, N.; Hristozov, D.; Jantunen, P.; Jeliazkova, N.; Johnston, H.; Lamon, L.; Murphy, F.; Rasmussen, K.; Rauscher, H.; Jiménez, A. S.; Svendsen, C.; Spurgeon, D.; Vázquez-Campos, S.; Wohlleben, W.;

Oomen, A. G., A framework for grouping and read-across of nanomaterials- supporting innovation and risk assessment. *Nano Today* **2020,** *35*, 100941.

99. Murphy, F. A.; Johnston, H. J.; Dekkers, S.; Bleeker, E. A. J.; Oomen, A. G.; Fernandes, T. F.; Rasmussen, K.; Jantunen, P.; Rauscher, H.; Hunt, N.; di Cristo, L.; Braakhuis, H. M.; Haase, A.; Hristozov, D.; Wohlleben, W.; Sabella, S.; Stone, V., How to formulate hypotheses and IATAs to support grouping and read-across of nanoforms. *Altex* **2023,** *40* (1), 125-140.

100. Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Serrrano, J. A.; Tietge, J. E.; Villeneuve, D. L., Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* **2010,** *29* (3), 730-41.

101. Villeneuve, D.; Crump, D.; Garcia-Reyero, N.; Hecker, M.; Hutchinson, T.; Lalone, C.; Landesmann, B.; Lettieri, T.; Munn, S.; Nepelska, M.; Ottinger, M.; Vergauwen, L.; Whelan, M., Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicological sciences : an official journal of the Society of Toxicology* **2014,** *142*, 312-20.

102. OECD, *The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins*. 2014.

103. Nymark, P.; Karlsson, H. L.; Halappanavar, S.; Vogel, U., Adverse Outcome Pathway Development for Assessment of Lung Carcinogenicity by Nanoparticles. *Front Toxicol* **2021,** *3*, 653386.

104. Halappanavar, S.; van den Brule, S.; Nymark, P.; Gaté, L.; Seidel, C.; Valentino, S.; Zhernovkov, V.; Høgh Danielsen, P.; De Vizcaya, A.; Wolff, H.; Stöger, T.; Boyadziev, A.; Poulsen, S. S.; Sørli, J. B.; Vogel, U., Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. *Part Fibre Toxicol* **2020,** *17* (1), 16.

105. Braakhuis, H. M.; Gosens, I.; Heringa, M. B.; Oomen, A. G.; Vandebriel, R. J.; Groenewold, M.; Cassee, F. R., Mechanism of Action of TiO2: Recommendations to Reduce Uncertainties Related to Carcinogenic Potential. *Annual Review of Pharmacology and Toxicology* **2021,** *61* (1), 203-223.

106. Brand, W.; Peters, R. J. B.; Braakhuis, H. M.; Maślankiewicz, L.; Oomen, A. G., Possible effects of titanium dioxide particles on human liver, intestinal tissue, spleen and kidney after oral exposure. *Nanotoxicology* **2020,** *14* (7), 985-1007.

107. Spinu, N.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Worth, A. P., Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. *Archives of Toxicology* **2020,** *94* (5), 1497-1510.

108. Paini, A.; Campia, I.; Cronin, M. T. D.; Asturiol, D.; Ceriani, L.; Exner, T. E.; Gao, W.; Gomes, C.; Kruisselbrink, J.; Martens, M.; Meek, M. E. B.; Pamies, D.; Pletz, J.; Scholz, S.; Schüttler, A.; Spînu, N.;

Villeneuve, D. L.; Wittwehr, C.; Worth, A.; Luijten, M., Towards a qAOP framework for predictive toxicology - Linking data to decisions. *Computational Toxicology* **2022,** *21*, 100195.

109. Nymark, P.; Rieswijk, L.; Ehrhart, F.; Jeliazkova, N.; Tsiliki, G.; Sarimveis, H.; Evelo, C. T.; Hongisto, V.; Kohonen, P.; Willighagen, E.; Grafström, R. C., A Data Fusion Pipeline for Generating and Enriching Adverse Outcome Pathway Descriptions. *Toxicol Sci* **2018,** *162* (1), 264-275.

110. Martens, M.; Verbruggen, T.; Nymark, P.; Grafström, R.; Burgoon, L. D.; Aladjov, H.; Torres Andón, F.; Evelo, C. T.; Willighagen, E. L., Introducing WikiPathways as a Data-Source to Support Adverse Outcome Pathways for Regulatory Risk Assessment of Chemicals and Nanomaterials. *Frontiers in Genetics* **2018,** *9*.

111. ECHA, Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals. **2008**.

112. ECHA, Read-across assessment framework (RAAF). *ECHA-17-R-01-EN.* **2017**.

113. ECHA, Guidance on information requirements and chemical safety assessment Appendix R.6-1: Recommendations for nanomaterials applicable to the Guidance on QSARs and Grouping of Chemicals. **2016**.

114. Mech, A.; Rasmussen, K.; Jantunen, P.; Aicher, L.; Alessandrelli, M.; Bernauer, U.; Bleeker, E. A. J.; Bouillard, J.; Di Prospero Fanghella, P.; Draisci, R.; Dusinska, M.; Encheva, G.; Flament, G.; Haase, A.; Handzhiyski, Y.; Herzberg, F.; Huwyler, J.; Jacobsen, N. R.; Jeliazkov, V.; Jeliazkova, N.; Nymark, P.; Grafström, R.; Oomen, A. G.; Polci, M. L.; Riebeling, C.; Sandström, J.; Shivachev, B.; Stateva, S.; Tanasescu, S.; Tsekovska, R.; Wallin, H.; Wilks, M. F.; Zellmer, S.; Apostolova, M. D., Insights into possibilities for grouping and read-across for nanomaterials in EU chemicals legislation. *Nanotoxicology* **2019,** *13* (1), 119-141.

115. Giusti, A.; Atluri, R.; Tsekovska, R.; Gajewicz, A.; Apostolova, M. D.; Battistelli, C. L.; Bleeker, E. A. J.; Bossa, C.; Bouillard, J.; Dusinska, M.; Gómez-Fernández, P.; Grafström, R.; Gromelski, M.; Handzhiyski, Y.; Jacobsen, N. R.; Jantunen, P.; Jensen, K. A.; Mech, A.; Navas, J. M.; Nymark, P.; Oomen, A. G.; Puzyn, T.; Rasmussen, K.; Riebeling, C.; Rodriguez-Llopis, I.; Sabella, S.; Sintes, J. R.; Suarez-Merino, B.; Tanasescu, S.; Wallin, H.; Haase, A., Nanomaterial grouping: Existing approaches and future recommendations. *NanoImpact* **2019,** *16*, 100182.

116. Jeliazkova, N.; Bleeker, E.; Cross, R.; Haase, A.; Janer, G.; Peijnenburg, W.; Pink, M.; Rauscher, H.; Svendsen, C.; Tsiliki, G.; Zabeo, A.; Hristozov, D.; Stone, V.; Wohlleben, W., How can we justify grouping of nanoforms for hazard assessment? Concepts and tools to quantify similarity. *NanoImpact* **2022,** *25*, 100366.

117. ECHA, Guidance on information requirements and chemical safety assessment, Appendix R.6-1 for nanoforms applicable to the Guidance on QSARs and Grouping of Chemicals. **2019**.

118. Sellers, K.; Deleebeeck, N.; Messiaen, M.; Jackson, M.; Bleeker, E. A. J.; Sijm, D.; Broekhuizen, F., *Grouping Nanomaterials - A strategy towards grouping and read-across*. 2015.

119. Oomen, A. G.; Steinhäuser, K. G.; Bleeker, E. A. J.; van Broekhuizen, F.; Sips, A.; Dekkers, S.; Wijnhoven, S. W. P.; Sayre, P. G., Risk assessment frameworks for nanomaterials: Scope, link to regulations, applicability, and outline for future directions in view of needed increase in efficiency. *NanoImpact* **2018,** *9*, 1-13.

120. Arts, J. H. E.; Hadi, M.; Irfan, M.-A.; Keene, A. M.; Kreiling, R.; Lyon, D.; Maier, M.; Michel, K.; Petry, T.; Sauer, U. G.; Warheit, D.; Wiench, K.; Wohlleben, W.; Landsiedel, R., A decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping). *Regulatory Toxicology and Pharmacology* **2015,** *71* (2, Supplement), S1-S27.

121. Wohlleben, W.; Hellack, B.; Nickel, C.; Herrchen, M.; Hund-Rinke, K.; Kettler, K.; Riebeling, C.; Haase, A.; Funk, B.; Kühnel, D.; Göhler, D.; Stintz, M.; Schumacher, C.; Wiemann, M.; Keller, J.; Landsiedel, R.; Broßell, D.; Pitzko, S.; Kuhlbusch, T. A. J., The nanoGRAVUR framework to group (nano)materials for their occupational, consumer, environmental risks based on a harmonized set of material properties, applied to 34 case studies. *Nanoscale* **2019,** *11* (38), 17637-17654.

122. Murphy, F.; Dekkers, S.; Braakhuis, H.; Ma-Hock, L.; Johnston, H.; Janer, G.; di Cristo, L.; Sabella, S.; Jacobsen, N. R.; Oomen, A. G.; Haase, A.; Fernandes, T.; Stone, V., An integrated approach to testing and assessment of high aspect ratio nanomaterials and its application for grouping based on a common mesothelioma hazard. *NanoImpact* **2021,** *22*, 100314.

123. OECD, *Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment*. 2017.

124. Arts, J. H. E.; Irfan, M.-A.; Keene, A. M.; Kreiling, R.; Lyon, D.; Maier, M.; Michel, K.; Neubauer, N.; Petry, T.; Sauer, U. G.; Warheit, D.; Wiench, K.; Wohlleben, W.; Landsiedel, R., Case studies putting the decision-making framework for the grouping and testing of nanomaterials (DF4nanoGrouping) into practice. *Regulatory Toxicology and Pharmacology* **2016,** *76*, 234-261.

125. Ag Seleci, D.; Tsiliki, G.; Werle, K.; Elam, D. A.; Okpowe, O.; Seidel, K.; Bi, X.; Westerhoff, P.; Innes, E.; Boyles, M.; Miller, M.; Giusti, A.; Murphy, F.; Haase, A.; Stone, V.; Wohlleben, W., Determining nanoform similarity via assessment of surface reactivity by abiotic and in vitro assays. *NanoImpact* **2022,** *26*, 100390.

126. Jeliazkova, N.; Ma-Hock, L.; Janer, G.; Stratmann, H.; Wohlleben, W., Possibilities to group nanomaterials across different substances – A case study on organic pigments. *NanoImpact* **2022,** *26*, 100391.

127. Keller, J. G.; Persson, M.; Müller, P.; Ma-Hock, L.; Werle, K.; Arts, J.; Landsiedel, R.; Wohlleben, W., Variation in dissolution behavior among different nanoforms and its implication for grouping approaches in inhalation toxicity. *NanoImpact* **2021,** *23*, 100341.

128. Ruggiero, E.; Santizo, K. Y.; Persson, M.; Delpivo, C.; Wohlleben, W., Food contact of paper and plastic products containing SiO2, Cu-Phthalocyanine, Fe2O3, CaCO3: Ranking factors that control the similarity of form and rate of release. *NanoImpact* **2022,** *25*, 100372.

129. Cross, R.; Matzke, M.; Spurgeon, D.; Diez, M.; Andres, V. G.; Galvez, E. C.; Esponda, M. F.; Belinga-Desaunay-Nault, M.-F.; Lynch, I.; Jeliazkova, N.; Svendsen, C., Assessing the similarity of nanoforms based on the biodegradation of organic surface treatment chemicals. *NanoImpact* **2022,** *26*, 100395.

130. Song, Y.; Bleeker, E.; Cross, R. K.; Vijver, M. G.; Peijnenburg, W. J. G. M., Similarity assessment of metallic nanoparticles within a risk assessment framework: A case study on metallic nanoparticles and lettuce. *NanoImpact* **2022,** *26*, 100397.

131. Petersen, E. J.; Ceger, P.; Allen, D. G.; Coyle, J.; Derk, R.; Garcia-Reyero, N.; Gordon, J.; Kleinstreuer, N. C.; Matheson, J.; McShan, D.; Nelson, B. C.; Patri, A. K.; Rice, P.; Rojanasakul, L.; Sasidharan, A.; Scarano, L.; Chang, X., U.S. Federal Agency interests and key considerations for new approach methodologies for nanomaterials. *Altex* **2022,** *39* (2), 183–206.

132. OECD, *Test No. 412: Subacute Inhalation Toxicity: 28-Day Study*. 2018.

133. OECD, *Test No. 413: Subchronic Inhalation Toxicity: 90-day Study*. 2018.

134. OECD, *Test No. 318: Dispersion Stability of Nanomaterials in Simulated Environmental Media*. 2017.

135. Schmid, O.; Stoeger, T., Surface area is the biologically most effective dose metric for acute nanoparticle toxicity in the lung. *Journal of Aerosol Science* **2016,** *99*, 133-143.

136. Simkó, M.; Nosske, D.; Kreyling, W. G., Metrics, Dose, and Dose Concept: The Need for a Proper Dose Concept in the Risk Assessment of Nanoparticles. *International Journal of Environmental Research and Public Health* **2014,** *11* (4), 4026-4048.

137. Mottier, A.; Mouchet, F.; Laplanche, C.; Cadarsi, S.; Lagier, L.; Arnault, J.-C.; Girard, H. A.; León, V.; Vázquez, E.; Sarrieu, C.; Pinelli, É.; Gauthier, L.; Flahaut, E., Surface Area of Carbon Nanoparticles: A Dose Metric for a More Realistic Ecotoxicological Assessment. *Nano Letters* **2016,** *16* (6), 3514-3518.

138. Cohen, J. M.; Teeguarden, J. G.; Demokritou, P., An integrated approach for the in vitro dosimetry of engineered nanomaterials. *Particle and Fibre Toxicology* **2014,** *11* (1), 20.

139. 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* (1), 52.

140. Hinderliter, P. M.; Minard, K. R.; Orr, G.; Chrisler, W. B.; Thrall, B. D.; Pounds, J. G.; Teeguarden, J. G., ISDD: A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies. *Part Fibre Toxicol* **2010,** *7* (1), 36.

141. Thomas, D. G.; Smith, J. N.; Thrall, B. D.; Baer, D. R.; Jolley, H.; Munusamy, P.; Kodali, V.; Demokritou, P.; Cohen, J.; Teeguarden, J. G., ISD3: a particokinetic model for predicting the combined effects of particle sedimentation, diffusion and dissolution on cellular dosimetry for in vitro systems. *Part Fibre Toxicol* **2018,** *15* (1), 6.

142. DeLoid, G. M.; Cohen, J. M.; Pyrgiotakis, G.; Pirela, S. V.; Pal, A.; Liu, J.; Srebric, J.; Demokritou, P., Advanced computational modeling for in vitro nanomaterial dosimetry. *Particle and Fibre Toxicology* **2015,** *12* (1), 32.

143. Nemmar, A.; Vanbilloen, H.; Hoylaerts, M. F.; Hoet, P. H.; Verbruggen, A.; Nemery, B., Passage of intratracheally instilled ultrafine particles from the lung into the systemic circulation in hamster. *Am J Respir Crit Care Med* **2001,** *164* (9), 1665-8.

144. Nemmar, A.; Hoet, P. H.; Vanquickenborne, B.; Dinsdale, D.; Thomeer, M.; Hoylaerts, M. F.; Vanbilloen, H.; Mortelmans, L.; Nemery, B., Passage of inhaled particles into the blood circulation in humans. *Circulation* **2002,** *105* (4), 411-4.

145. Oberdörster, G.; Sharp, Z.; Atudorei, V.; Elder, A.; Gelein, R.; Kreyling, W.; Cox, C., Translocation of inhaled ultrafine particles to the brain. *Inhal Toxicol* **2004,** *16* (6-7), 437-45.

146. Anjilvel, S.; Asgharian, B., A Multiple-Path Model of Particle Deposition in the Rat Lung. *Fundamental and Applied Toxicology* **1995,** *28* (1), 41-50.

147. Price, O. T.; Asgharian, B.; Miller, F. J.; Cassee, F. R.; deWinter-Sorkina, R. In *Multiple Path Particle Dosimetry model (MPPD v1.0): A model for human and rat airway particle dosimetry*, 2002.

148. Miller, F. J.; Asgharian, B.; Schroeter, J. D.; Price, O., Improvements and additions to the Multiple Path Particle Dosimetry model. *Journal of Aerosol Science* **2016,** *99*, 14-26.

149. Kroll, A.; Pillukat, M. H.; Hahn, D.; Schnekenburger, J., Interference of engineered nanoparticles with in vitro toxicity assays. *Arch Toxicol* **2012,** *86* (7), 1123-36.

150. Stone, V.; Johnston, H.; Schins, R. P. F., Development of in vitro systems for nanotoxicology: methodological considerations. *Critical Reviews in Toxicology* **2009,** *39* (7), 613-626.

151. Tran, L.; Bañares, M. A.; Rallo, R., Modelling the Toxicity of Nanoparticles. *Springer Cham* **2017**.

152. James, G.; Witten, D.; Hastie, T.; Tibshirani, R., An Introduction to Statistical Learning with Applications in R. *SpringerLink* **2023**.

153. Hawkins, D. M., The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences* **2004,** *44* (1), 1-12.

154. Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P., SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* **2002,** *16*, 321-357.

155. Elreedy, D.; Atiya, A. F., A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Information Sciences* **2019,** *505*, 32-64.

156. Stekhoven, D. J.; Bühlmann, P., MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2011,** *28* (1), 112-118.

157. Troyanskaya, O.; Cantor, M.; Sherlock, G.; Brown, P.; Hastie, T.; Tibshirani, R.; Botstein, D.; Altman, R. B., Missing value estimation methods for DNA microarrays. *Bioinformatics* **2001,** *17* (6), 520-525.

158. Jabbar, H.; Khan, R. Z., Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices* **2015,** *70* (10.3850), 978-981.

159. Breiman, L., Random Forests. *Machine Learning* **2001,** *45* (1), 5-32.

160. Wang, H.; Yang, F.; Luo, Z., An experimental study of the intrinsic stability of random forest variable importance measures. *BMC Bioinformatics* **2016,** *17* (1), 60.

161. Tibshirani, R., Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **2018,** *58* (1), 267-288.

162. Zou, H.; Hastie, T., Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **2005,** *67* (2), 301-320.

163. Darst, B. F.; Malecki, K. C.; Engelman, C. D., Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics* **2018,** *19*.

164. Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005,** *102* (43), 15545-50.

165. Kanehisa, M.; Goto, S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **2000,** *28* (1), 27-30.

166. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K., KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **2017,** *45* (D1), D353-d361.

167. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J. P.; Tamayo, P., The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **2015,** *1* (6), 417-425.

168. Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; D'Eustachio, P.; Schmidt, E.; de Bono, B.; Jassal, B.; Gopinath, G. R.; Wu, G. R.; Matthews, L.; Lewis, S.; Birney, E.; Stein, L., Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **2005,** *33* (Database issue), D428-32.

169. Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; Deng, C.; Varusai, T.; Ragueneau, E.; Haider, Y.; May, B.; Shamovsky, V.; Weiser, J.; Brunson, T.; Sanati, N.; Beckman, L.; Shao, X.; Fabregat, A.; Sidiropoulos, K.; Murillo, J.; Viteri, G.; Cook, J.; Shorser, S.; Bader, G.; Demir, E.; Sander, C.; Haw, R.; Wu, G.; Stein,

L.; Hermjakob, H.; D'Eustachio, P., The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **2022,** *50* (D1), D687-d692.

170. Consortium, T. G. O.; Aleksander, S. A.; Balhoff, J.; Carbon, S.; Cherry, J. M.; Drabkin, H. J.; Ebert, D.; Feuermann, M.; Gaudet, P.; Harris, N. L.; Hill, D. P.; Lee, R.; Mi, H.; Moxon, S.; Mungall, C. J.; Muruganugan, A.; Mushayahama, T.; Sternberg, P. W.; Thomas, P. D.; Van Auken, K.; Ramsey, J.; Siegele, D. A.; Chisholm, R. L.; Fey, P.; Aspromonte, M. C.; Nugnes, M. V.; Quaglia, F.; Tosatto, S.; Giglio, M.; Nadendla, S.; Antonazzo, G.; Attrill, H.; dos Santos, G.; Marygold, S.; Strelets, V.; Tabone, C. J.; Thurmond, J.; Zhou, P.; Ahmed, S. H.; Asanitthong, P.; Luna Buitrago, D.; Erdol, M. N.; Gage, M. C.; Ali Kadhum, M.; Li, K. Y. C.; Long, M.; Michalak, A.; Pesala, A.; Pritazahra, A.; Saverimuttu, S. C. C.; Su, R.; Thurlow, K. E.; Lovering, R. C.; Logie, C.; Oliferenko, S.; Blake, J.; Christie, K.; Corbani, L.; Dolan, M. E.; Drabkin, H. J.; Hill, D. P.; Ni, L.; Sitnikov, D.; Smith, C.; Cuzick, A.; Seager, J.; Cooper, L.; Elser, J.; Jaiswal, P.; Gupta, P.; Jaiswal, P.; Naithani, S.; Lera-Ramirez, M.; Rutherford, K.; Wood, V.; De Pons, J. L.; Dwinell, M. R.; Hayman, G. T.; Kaldunski, M. L.; Kwitek, A. E.; Laulederkind, S. J. F.; Tutaj, M. A.; Vedi, M.; Wang, S.-J.; D'Eustachio, P.; Aimo, L.; Axelsen, K.; Bridge, A.; Hyka-Nouspikel, N.; Morgat, A.; Aleksander, S. A.; Cherry, J. M.; Engel, S. R.; Karra, K.; Miyasato, S. R.; Nash, R. S.; Skrzypek, M. S.; Weng, S.; Wong, E. D.; Bakker, E.; Berardini, T. Z.; Reiser, L.; Auchincloss, A.; Axelsen, K.; Argoud-Puy, G.; Blatter, M.-C.; Boutet, E.; Breuza, L.; Bridge, A.; Casals-Casas, C.; Coudert, E.; Estreicher, A.; Livia Famiglietti, M.; Feuermann, M.; Gos, A.; Gruaz-Gumowski, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Le Mercier, P.; Lieberherr, D.; Masson, P.; Morgat, A.; Pedruzzi, I.; Pourcel, L.; Poux, S.; Rivoire, C.; Sundaram, S.; Bateman, A.; Bowler-Barnett, E.; Bye-A-Jee, H.; Denny, P.; Ignatchenko, A.; Ishtiaq, R.; Lock, A.; Lussi, Y.; Magrane, M.; Martin, M. J.; Orchard, S.; Raposo, P.; Speretta, E.; Tyagi, N.; Warner, K.; Zaru, R.; Diehl, A. D.; Lee, R.; Chan, J.; Diamantakis, S.; Raciti, D.; Zarowiecki, M.; Fisher, M.; James-Zorn, C.; Ponferrada, V.; Zorn, A.; Ramachandran, S.; Ruzicka, L.; Westerfield, M., The Gene Ontology knowledgebase in 2023. *Genetics* **2023,** *224* (1).

171. Landsiedel, R.; Ma-Hock, L.; Hofmann, T.; Wiemann, M.; Strauss, V.; Treumann, S.; Wohlleben, W.; Gröters, S.; Wiench, K.; van Ravenzwaay, B., Application of short-term inhalation studies to assess the inhalation toxicity of nanomaterials. *Part Fibre Toxicol* **2014,** *11*, 16.

172. Ma-Hock, L.; Burkhardt, S.; Strauss, V.; Gamer, A. O.; Wiench, K.; van Ravenzwaay, B.; Landsiedel, R., Development of a Short-Term Inhalation Test in the Rat Using Nano-Titanium Dioxide as a Model Substance. *Inhalation Toxicology* **2009,** *21* (2), 102-118.

173. Buesen, R.; Landsiedel, R.; Sauer, U. G.; Wohlleben, W.; Groeters, S.; Strauss, V.; Kamp, H.; van Ravenzwaay, B., Effects of SiO2, ZrO2, and BaSO4 nanomaterials with or without surface functionalization upon 28-day oral exposure to rats. *Archives of Toxicology* **2014,** *88* (10), 1881-1906.

174. Arts, J. H.; Muijser, H.; Duistermaat, E.; Junker, K.; Kuper, C. F., Five-day inhalation toxicity study of three types of synthetic amorphous silicas in Wistar rats and post-exposure evaluations for up to 3 months. *Food Chem Toxicol* **2007,** *45* (10), 1856-67.

175. Ma-Hock, L.; Strauss, V.; Treumann, S.; Küttler, K.; Wohlleben, W.; Hofmann, T.; Gröters, S.; Wiench, K.; van Ravenzwaay, B.; Landsiedel, R., Comparative inhalation toxicity of multi-wall carbon nanotubes, graphene, graphite nanoplatelets and low surface carbon black. *Part Fibre Toxicol* **2013,** *10*, 23.

176. Wiemann, M.; Vennemann, A.; Sauer, U. G.; Wiench, K.; Ma-Hock, L.; Landsiedel, R., An in vitro alveolar macrophage assay for predicting the short-term inhalation toxicity of nanomaterials. *J Nanobiotechnology* **2016,** *14*, 16-16.

177. Tran, C. L.; Buchanan, D.; Cullen, R. T.; Searl, A.; Jones, A. D.; Donaldson, K., Inhalation of poorly soluble particles. II. Influence Of particle surface area on inflammation and clearance. *Inhal Toxicol* **2000,** *12* (12), 1113-26.

178. Duffin, R.; Tran, L.; Brown, D.; Stone, V.; Donaldson, K., Proinflammogenic Effects of Low-Toxicity and Metal Nanoparticles In Vivo and In Vitro: Highlighting the Role of Particle Surface Area and Surface Reactivity. *Inhalation Toxicology* **2007,** *19* (10), 849-856.

179. Wiemann, M.; Sauer, U. G.; Vennemann, A.; Bäcker, S.; Keller, J.-G.; Ma-Hock, L.; Wohlleben, W.; Landsiedel, R., In Vitro and In Vivo Short-Term Pulmonary Toxicity of Differently Sized Colloidal Amorphous SiO2. *Nanomaterials* **2018,** *8* (3), 160.

180. Nel, A.; Xia, T.; Mädler, L.; Li, N., Toxic potential of materials at the nanolevel. *Science* **2006,** *311* (5761), 622-7.

181. Saarimäki, L. A.; Federico, A.; Lynch, I.; Papadiamantis, A. G.; Tsoumanis, A.; Melagraki, G.; Afantitis, A.; Serra, A.; Greco, D., Manually curated transcriptomics data collection for toxicogenomic assessment of engineered nanomaterials. *Scientific Data* **2021,** *8* (1), 49.

182. Karkossa, I.; Bannuscher, A.; Hellack, B.; Bahl, A.; Buhs, S.; Nollau, P.; Luch, A.; Schubert, K.; von Bergen, M.; Haase, A., An in-depth multi-omics analysis in RLE-6TN rat alveolar epithelial cells allows for nanomaterial categorization. *Particle and Fibre Toxicology* **2019,** *16* (1), 38.

183. Bannuscher, A.; Karkossa, I.; Buhs, S.; Nollau, P.; Kettler, K.; Balas, M.; Dinischiotu, A.; Hellack, B.; Wiemann, M.; Luch, A.; von Bergen, M.; Haase, A.; Schubert, K., A multi-omics approach reveals mechanisms of nanomaterial toxicity and structure–activity relationships in alveolar macrophages. *Nanotoxicology* **2020,** *14* (2), 181-195.

184. Billing, A. M.; Knudsen, K. B.; Chetwynd, A. J.; Ellis, L.-J. A.; Tang, S. V. Y.; Berthing, T.; Wallin, H.; Lynch, I.; Vogel, U.; Kjeldsen, F., Fast and Robust Proteome Screening Platform Identifies Neutrophil Extracellular Trap Formation in the Lung in Response to Cobalt Ferrite Nanoparticles. *ACS Nano* **2020,** *14* (4), 4096-4110.

185. Gallud, A.; Delaval, M.; Kinaret, P.; Marwah, V. S.; Fortino, V.; Ytterberg, J.; Zubarev, R.; Skoog, T.; Kere, J.; Correia, M.; Loeschner, K.; Al-Ahmady, Z.; Kostarelos, K.; Ruiz, J.; Astruc, D.; Monopoli, M.; Handy, R.; Moya, S.; Savolainen, K.; Alenius, H.; Greco, D.; Fadeel, B. Multiparametric Profiling of Engineered Nanomaterials: Unmasking the Surface Coating Effect *Adv Sci (Weinh)* [Online], 2020, p. 2002221. PubMed. (accessed 2020/11//).

186. Seidel, C.; Zhernovkov, V.; Cassidy, H.; Kholodenko, B.; Matallanas, D.; Cosnier, F.; Gaté, L., Inhaled multi-walled carbon nanotubes differently modulate global gene and protein expression in rat lungs. *Nanotoxicology* **2021,** *15* (2), 238-256.

187. Phuyal, S.; Kasem, M.; Knittelfelder, O.; Sharma, A.; Fonseca, D. d. M.; Vebraite, V.; Shaposhnikov, S.; Slupphaug, G.; Skaug, V.; Zienolddiny, S., Characterization of the proteome and lipidome profiles of human lung cells after low dose and chronic exposure to multiwalled carbon nanotubes. *Nanotoxicology* **2018,** *12* (2), 138-152.

188. Alswady-Hoff, M.; Erdem, J. S.; Phuyal, S.; Knittelfelder, O.; Sharma, A.; Fonseca, D. d. M.; Skare, Ø.; Slupphaug, G.; Zienolddiny, S., Long-Term Exposure to Nanosized TiO2 Triggers Stress Responses and Cell Death Pathways in Pulmonary Epithelial Cells. *International Journal of Molecular Sciences* **2021,** *22* (10), 5349.

189. OECD, *Test No. 125: Nanomaterial Particle Size and Size Distribution of Nanomaterials*. 2023.

190. OECD, *Test No. 126: Determination of the Hydrophobicity Index of Nanomaterials Through an Affinity Measurement*. 2023.

191. OECD, *Test No. 124: Determination of the Volume Specific Surface Area of Manufactured Nanomaterials*. 2022.

192. Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G., Gene Ontology: tool for the unification of biology. *Nature Genetics* **2000,** *25* (1), 25-29.

193. Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J., The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* **2007,** *95* (1), 5-12.

194. Gajewicz, A.; Jagiello, K.; Cronin, M. T. D.; Leszczynski, J.; Puzyn, T., Addressing a bottle neck for regulation of nanomaterials: quantitative read-across (Nano-QRA) algorithm for cases when only limited data is available. *Environmental Science: Nano* **2017,** *4* (2), 346-358.

195. OECD, Important Issues on Risk Assessment of Manufactured Nanomaterials Series on the Safety of Manufactured Nanomaterials No. 103. **2022**.

196. Marchese Robinson, R. L.; Lynch, I.; Peijnenburg, W.; Rumble, J.; Klaessig, F.; Marquardt, C.; Rauscher, H.; Puzyn, T.; Purian, R.; Åberg, C.; Karcher, S.; Vriens, H.; Hoet, P.; Hoover, M. D.; Hendren,

C. O.; Harper, S. L., How should the completeness and quality of curated nanomaterial data be evaluated? *Nanoscale* **2016,** *8* (19), 9919-9943.

197. Lynch, I.; Afantitis, A.; Exner, T.; Himly, M.; Lobaskin, V.; Doganis, P.; Maier, D.; Sanabria, N.; Papadiamantis, A. G.; Rybinska-Fryca, A.; Gromelski, M.; Puzyn, T.; Willighagen, E.; Johnston, B. D.; Gulumian, M.; Matzke, M.; Green Etxabe, A.; Bossa, N.; Serra, A.; Liampa, I.; Harper, S.; Tämm, K.; Jensen, A. C.; Kohonen, P.; Slater, L.; Tsoumanis, A.; Greco, D.; Winkler, D. A.; Sarimveis, H.; Melagraki, G., Can an InChI for Nano Address the Need for a Simplified Representation of Complex Nanomaterials across Experimental and Nanoinformatics Studies? *Nanomaterials* **2020,** *10* (12), 2493.

198. Lustgarten, J. L.; Gopalakrishnan, V.; Visweswaran, S., Measuring stability of feature selection in biomedical datasets. *AMIA Annu Symp Proc* **2009,** *2009*, 406-10.

199. Serra, A.; Önlü, S.; Festa, P.; Fortino, V.; Greco, D., MaNGA: a novel multi-niche multi-objective genetic algorithm for QSAR modelling. *Bioinformatics* **2019,** *36* (1), 145-153.

200. Halappanavar, S.; Rahman, L.; Nikota, J.; Poulsen, S. S.; Ding, Y.; Jackson, P.; Wallin, H.; Schmid, O.; Vogel, U.; Williams, A., Ranking of nanomaterial potency to induce pathway perturbations associated with lung responses. *NanoImpact* **2019,** *14*, 100158.

201. Serra, A.; Saarimäki, L. A.; Fratello, M.; Marwah, V. S.; Greco, D., BMDx: a graphical Shiny application to perform Benchmark Dose analysis for transcriptomics data. *Bioinformatics* **2020,** *36* (9), 2932-2933.

202. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szcześniak, M. W.; Gaffney, D. J.; Elo, L. L.; Zhang, X.; Mortazavi, A., A survey of best practices for RNA-seq data analysis. *Genome Biology* **2016,** *17* (1), 13.

203. 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 for Biology and Medicine* **2019,** *14* (1), 1.

204. Amorim, M. J. B.; Peijnenburg, W.; Greco, D.; Saarimäki, L. A.; Dumit, V. I.; Bahl, A.; Haase, A.; Tran, L.; Hackermüller, J.; Canzler, S.; Scott-Fordsmand, J. J., Systems toxicology to advance human and environmental hazard assessment: A roadmap for advanced materials. *Nano Today* **2023,** *48*, 101735.

205. Canzler, S.; Schor, J.; Busch, W.; Schubert, K.; Rolle-Kampczyk, U. E.; Seitz, H.; Kamp, H.; von Bergen, M.; Buesen, R.; Hackermüller, J., Prospects and challenges of multi-omics data integration in toxicology. *Archives of Toxicology* **2020,** *94* (2), 371-388.

206. 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.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J.

G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B., The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016,** *3* (1), 160018.

207. Jeliazkova, N.; Apostolova, M. D.; Andreoli, C.; Barone, F.; Barrick, A.; Battistelli, C.; Bossa, C.; Botea-Petcu, A.; Châtel, A.; De Angelis, I.; Dusinska, M.; El Yamani, N.; Gheorghe, D.; Giusti, A.; Gómez-Fernández, P.; Grafström, R.; Gromelski, M.; Jacobsen, N. R.; Jeliazkov, V.; Jensen, K. A.; Kochev, N.; Kohonen, P.; Manier, N.; Mariussen, E.; Mech, A.; Navas, J. M.; Paskaleva, V.; Precupas, A.; Puzyn, T.; Rasmussen, K.; Ritchie, P.; Llopis, I. R.; Rundén-Pran, E.; Sandu, R.; Shandilya, N.; Tanasescu, S.; Haase, A.; Nymark, P., Towards FAIR nanosafety data. *Nature Nanotechnology* **2021,** *16* (6), 644-654.

208. Dumit, V. I.; Ammar, A.; Bakker, M. I.; Bañares, M. A.; Bossa, C.; Costa, A.; Cowie, H.; Drobne, D.; Exner, T. E.; Farcal, L.; Friedrichs, S.; Furxhi, I.; Grafström, R.; Haase, A.; Himly, M.; Jeliazkova, N.; Lynch, I.; Maier, D.; Noorlander, C. W.; Shin, H. K.; Soler-Illia, G. J. A. A.; Suarez-Merino, B.; Willighagen, E.; Nymark, P., From principles to reality. FAIR implementation in the nanosafety community. *Nano Today* **2023,** *51*, 101923.

209. Karcher, S.; Willighagen, E. L.; Rumble, J.; Ehrhart, F.; Evelo, C. T.; Fritts, M.; Gaheen, S.; Harper, S. L.; Hoover, M. D.; Jeliazkova, N.; Lewinski, N.; Marchese Robinson, R. L.; Mills, K. C.; Mustad, A. P.; Thomas, D. G.; Tsiliki, G.; Hendren, C. O., Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations. *NanoImpact* **2018,** *9*, 85-101.

210. van Rijn, J.; Afantitis, A.; Culha, M.; Dusinska, M.; Exner, T. E.; Jeliazkova, N.; Longhin, E. M.; Lynch, I.; Melagraki, G.; Nymark, P.; Papadiamantis, A. G.; Winkler, D. A.; Yilmaz, H.; Willighagen, E., European Registry of Materials: global, unique identifiers for (undisclosed) nanomaterials. *Journal of Cheminformatics* **2022,** *14* (1), 57.

211. Gottardo, S.; Quiros Pesudo, L.; Totaro, S.; Riego Sintes, J.; Crutzen, H., *Data logging templates for the environmental, health and safety assessment of nanomaterials*. 2017.

212. Murray, R. A.; Escobar, A.; Bastús, N. G.; Andreozzi, P.; Puntes, V.; Moya, S. E., Fluorescently labelled nanomaterials in nanosafety research: Practical advice to avoid artefacts and trace unbound dye. *NanoImpact* **2018,** *9*, 102-113.

213. Riebeling, C.; Piret, J.-P.; Trouiller, B.; Nelissen, I.; Saout, C.; Toussaint, O.; Haase, A., A guide to nanosafety testing: Considerations on cytotoxicity testing in different cell models. *NanoImpact* **2018,** *10*, 1-10.

214. Søs Poulsen, S.; Jacobsen, N. R.; Labib, S.; Wu, D.; Husain, M.; Williams, A.; Bøgelund, J. P.; Andersen, O.; Købler, C.; Mølhave, K.; Kyjovska, Z. O.; Saber, A. T.; Wallin, H.; Yauk, C. L.; Vogel, U.; Halappanavar, S., 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* (11), e80452.

215. Kinaret, P.; Singh, M.; 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*.

List of publications

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., Haase, A., Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, **15**, 100179, https://doi.org/10.1016/j.impact.2019.100179 (2019).

Contribution: Development of the research idea, data analysis and modelling, writing, review and editing

Kratochvil, I.*, Bannuscher, A.*, Hellack, B., Bahl, A., Buhs, S., Nollau, P., Luch, A., Schubert, K., von Bergen, M., Haase, A., An in-depth multi-omics analysis in RLE-6TN rat alveolar epithelial cells allows for nanomaterial categorization. *Part Fibre Toxicol* 16, 38, https://doi.org/10.1186/s12989-019-0321-5 (2019).

Contribution: Data and statistical analysis, contributions during writing, review and editing

Bahl, A., Hellack, B., Wiemann, M., Giusti, A., Werle, K., Haase, A., Wohlleben, W., Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods. *NanoImpact* **19**, 100234, https://doi.org/10.1016/j.impact.2020.100234 (2020).

Contribution: Formal analysis, investigation and writing, review and editing

Bahl, A., Ibrahim, C., Plate, K., Haase, A., Dengjel, J., Nymark, P., Dumit, V. I., PROTEOMAS: A workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping. *Journal of Cheminformatics* (2023) **15**, 34, https://doi.org/10.1186/s13321-023-00710-2 (2023).

Contribution: Development of the research idea, contribution to methods and interpretation of the data, implementation of the PROTEOMAS workflow, analysis of the data and writing of the manuscript, review and editing

Mancardi, G., Mikolajczyk, A., Annapoorani, V. K., Bahl, A., Blekos, K., Burk, J., Çetin, Y. A., Chairetakis, K., Dutta, S., Escorihuela, L., Jagiello, K., Singhal, A., van der Pol, R., Bañares, M., A., Buchete, N.-V., Calatayud, M., Dumit, V. I., Gardini, D., Jeliazkova, N., Haase, A., Marcoulaki, E., Martorell, B., Puzyn, T., Sevink, G. J. A., Simeone, F. C., Tämm, K., Chiavazzo, E., A computational view on nanomaterial intrinsic and extrinsic features for nanosafety and sustainability. *Materials Today* **67**, 344-370, https://doi.org/10.1016/j.mattod.2023.05.029 (2023).

Contribution: Contribution during writing, review and editing

Amorim, M. J. B., Peijnenburg, W., Greco, D., Saarimäki, L. A., Dumit, V. I., Bahl, A., Haase, A., Tran, L., Hackermüller, J., Canzler, S., Scott-Fordsmand, J. J., Systems toxicology to advance human and environmental hazard assessment: A roadmap for advanced materials. *Nano Today*, **48**, 101735, https://doi.org/10.1016/j.nantod.2022.101735 (2023).

Contribution: Contribution during writing, review and editing

Dumit, V. I. Liu, Y.-C., Bahl, A., Kohonen, P., Grafström, R., Nymark, P., Müller-Graf, C., Haase, A., Pink, M., Meta-Analysis of Integrated Proteomic and Transcriptomic Data Discerns Structure–Activity Relationship of Carbon Materials with Different Morphologies. *Advanced Science*, https://doi.org/10.1002/advs.202306268 (2023).

Contribution: Preparation of proteomics data, RF Analysis, contribution during writing, review and editing

Bahl, A., Halappanavar, S., Wohlleben, W., Nymark, P., Kohonen, P., Wallin, H., Vogel, U., Haase, A., Bioinformatics and machine learning to support nanomaterial grouping. Submitted to *Nanotoxicology*.

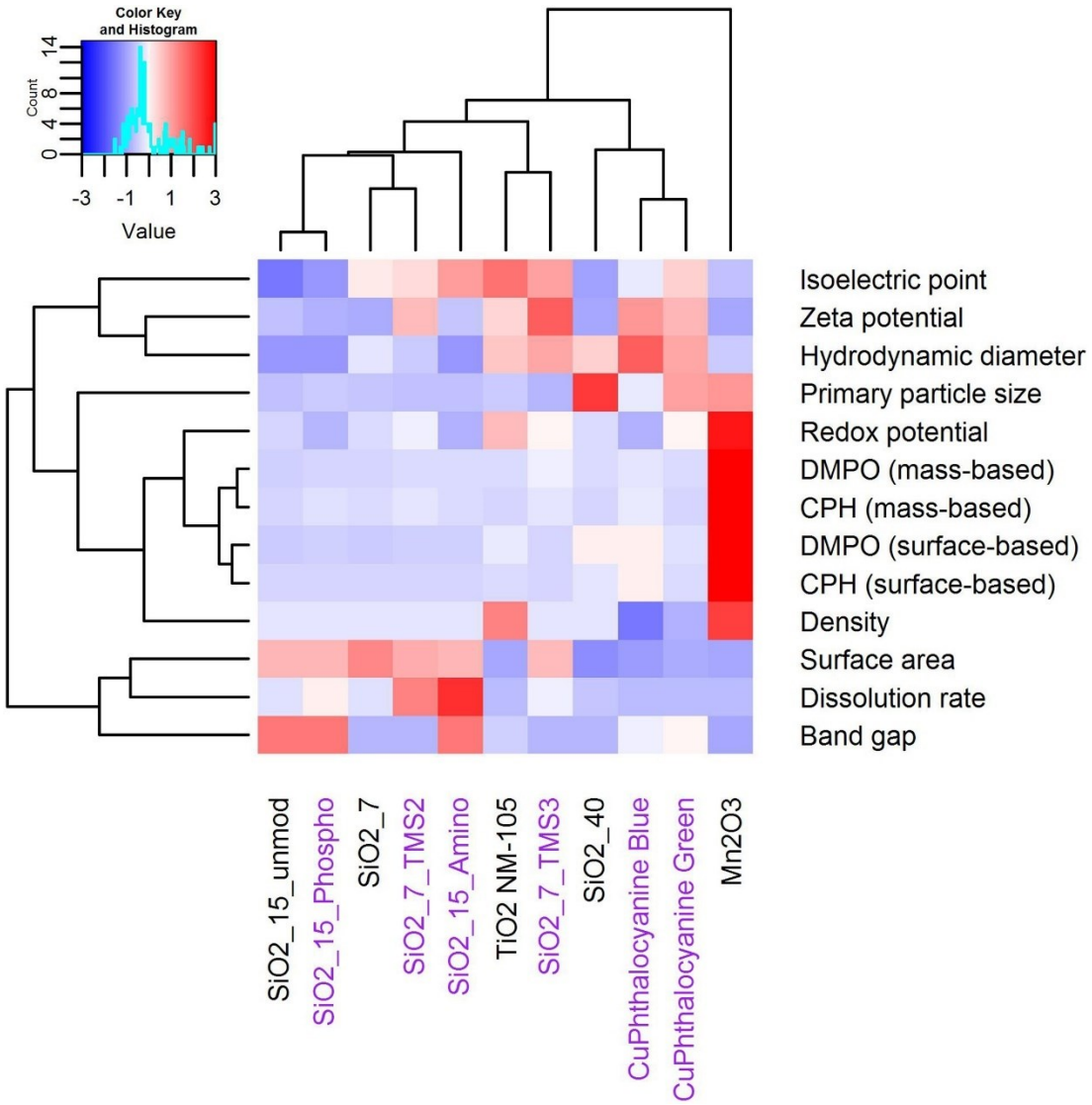
Contribution: Literature research, interpretation, writing, review and editing

Annexes

Annex I

Supplementary information to

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., Haase, A., Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, **15**, 100179, https://doi.org/10.1016/j.impact.2019.100179 (2019).



Supp. Fig. 1. Heatmap of physico-chemical properties across NMs. The table of physico-chemical properties was translated into colors ranging from dark blue for the smallest values to dark red for the highest values. All properties were scaled across NMs in order to make them comparable and avoid overrepresentation of those properties having larger values in general in the clustering step. The dendrogram on the left shows the similarity of the physico-chemical properties across all studied NMs. The dendrogram on the top shows how similar NMs are across all studied physico-chemical properties.

Annex II

Supplementary information to

Bahl, A., Hellack, B., Wiemann, M., Giusti, A., Werle, K., Haase, A., Wohlleben, W., Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods. *NanoImpact* **19**, 100234, https://doi.org/10.1016/j.impact.2020.100234 (2020).

Supplementary Information
Nanomaterial categorization by surface reactivity: A case study comparing 35 materials with four different test methods

Aileen Bahl[1], Bryan Hellack[2,3], Martin Wiemann[4], Anna Giusti[1], Kai Werle[5], Andrea Haase[1]*, Wendel Wohlleben[5]*

[1]*German Federal Institute for Risk Assessment (BfR), Department of Chemical and Product Safety, Berlin, Germany*

[2]*Institute for Energy and Environmental Technology e.V. (IUTA), Duisburg, Germany*

[3]*German Environment Agency (UBA), Dessau, Germany*

[4] *IBE R&D Institute for Lung Health gGmbH, Muenster, Germany*

[5]*BASF SE, Department of Material Physics, Ludwigshafen, Germany*

***** authors contributed equally

Corresponding Authors

Andrea Haase

German Federal Institute for Risk Assessment (BfR)

Department of Chemical and Product Safety

Germany

Email: andrea.haase@bfr.bund.de

Phone: +49-30-18412-27600

Wendel Wohlleben

BASF SE

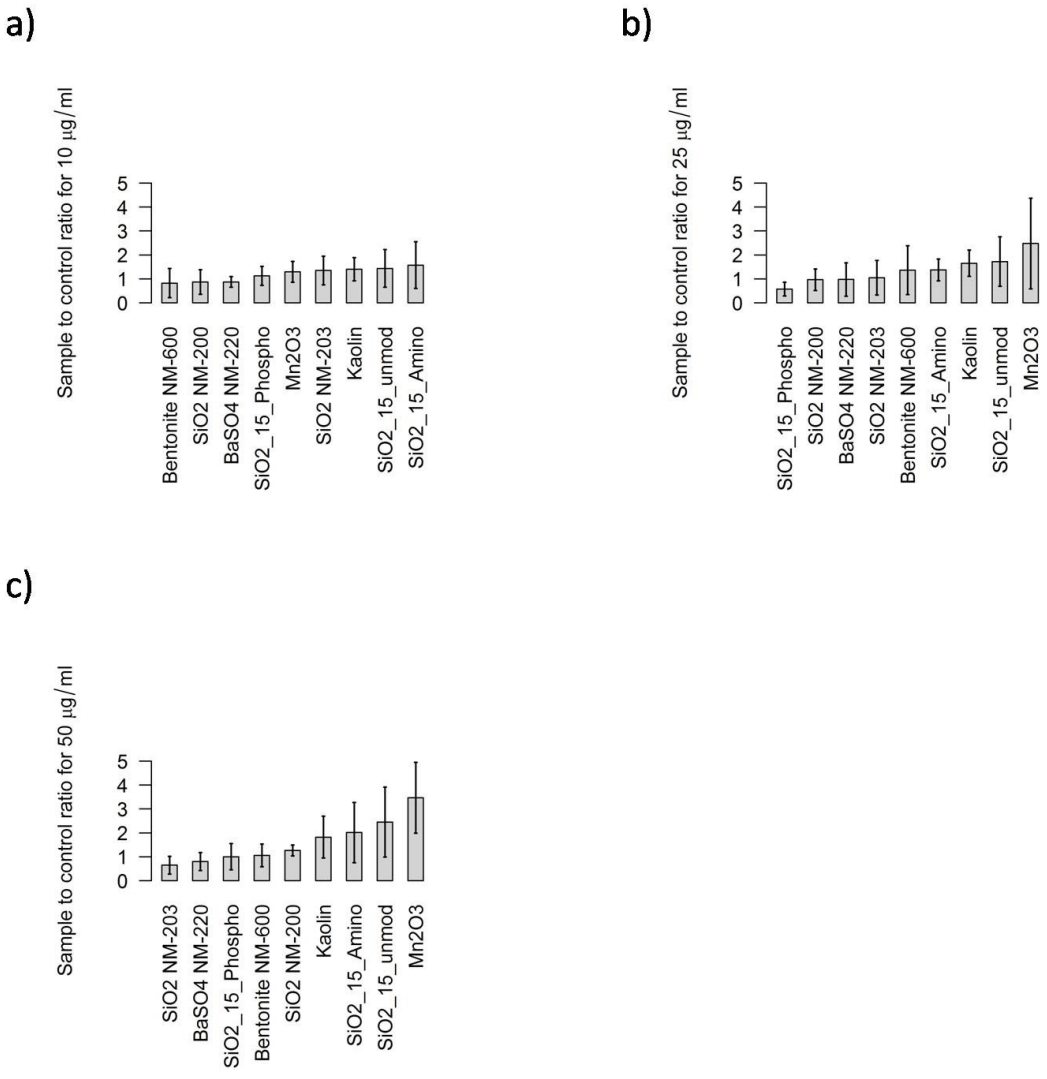
Department of Material Physics and Department of Experimental Toxicology

Germany

Email: wendel.wohlleben@basf.com

Phone: +49-621-6095339

Keywords: nanomaterials, surface reactivity, oxidative potential, functional assays, test strategy, logistic regression



Supplementary Figure 1: Protein carbonylation of silica particles measured at three different concentrations a) 10 µg/ml, b) 25 µg/ml and c) 50 µg/ml.

Supplementary Table 1a: Mass-based values obtained from the four OP assays with standard deviations

Class	Material	Supplier	ESR (CPH)	ESR (DMPO)	FRAS (AUC)	Carbonyls	Carbonyls (normalized)
Silica	SiO_2_15_unmod	BASF SE	0.82±0.14	0.97± 0.01	890786	1.73± 1.04	86.25± 51.86
	SiO_2_15_Amino	BASF SE	0.92± 0.03	0.97± 0.01	1300879	1.38± 0.46	3.06± 1.02
	SiO_2_15_Phospho	BASF SE	1.21± 0.04	0.83± 0.08	1288551	0.58± 0.29	3.19± 1.60
	SiO_2 NM-200	JRC repository				0.97± 0.45	
	SiO_2 NM-203	JRC repository			718407	1.05± 0.72	2.92± 2.00
	SiO_2Levasil 50	Levasil					
	SiO_2Levasil 100	Levasil					
	SiO_2Levasil 300	Levasil					
	SiO_2Aerosil 200	Evonik	0.92± 0.16	0.84± 0.19			
Aluminosilicates	Kaolin	BASF SE	1.60± 0.11	0.99± 0.27	503136	1.65± 0.55	3.06± 1.01
	Bentonite	Zoz GmbH			2915633	1.36± 1.02	7.17± 5.36
	DQ12 (Quartz)	(Archiv Toxikologie BASF)			93404		
Iron oxides	Fe_2O_3nanoform A	BASF Colors and Effects	0.51± 0.02	0.75± 0.04	3220889	4.29± 2.26	4.52± 2.38
	Fe_2O_3nanoform B	BASF Colors and Effects	0.82± 0.12	1.13± 0.05	150508	1.8± 0.67	2.54± 0.95
	Fe_2O_3larger	Huntsman	13.86± 1.16	4.33± 0.11	357083	2.51± 0.90	2.79± 1.00
Titania	TiO_2 NM-102	JRC repository	0.63± 0.08	0.88± 0.15	829911		
	TiO_2 NM-105	JRC repository	0.69± 0.01	1.01± 0.06	677867	1.48± 1.57	3.1± 1.96
	TiO_2 non-nano	Kronos	0.94± 0.09	0.91± 0.18	257420		
Ceria	CeO_2 NM-211	JRC repository	1.80± 0.23	1.28± 0.24	837353		
	CeO_2 NM-212	JRC repository	1.42± 0.18	2.07± 0.32	570057		
Copper-based	CuO	Sigma Aldrich	178.12± 29.37	21.21± 2.63	10293942	10.74± 1.76	12.78± 2.10
	Cu-Phthalocyanine non-halogenated	BASF Colors and Effects	1.22± 0.07	1.71± 0.30	408695	0.66± 0.15	1.04± 0.23
	Cu-Phthalocyanine Halogenated	BASF Colors and Effects	1.03± 0.09	0.67± 0.07	734588	0.81± 0.43	1.11± 0.59
Zinc oxides	ZnO NM-110	JRC repository	1.51± 0.29	2.20± 0.09	2898432	2.53± 1.98	4.76± 3.74
	ZnO NM-111	JRC repository			1069431	1.89± 1.44	
Diketopyrrolopyrrol pigments	DPP_premixed	BASF Colors and Effects	0.88± 0.06	0.77± 0.04	-337683	0.77± 0.69	
	DPP nano	BASF Colors and Effects	0.84± 0.03	0.77± 0.08	92638	0.41± 0.1	
	DPP non-nano	BASF Colors and Effects	0.82± 0.05	1.06± 0.15	52079	1.43± 0.43	
Carbon-based	Carbon black	Ensaco	1.21± 0.16	0.98± 0.11	2100127	1.33± 1.18	
	Graphene oxide	(Archiv IUTA)				0.57± 0.28	

	Graphene 1-layer Graphene Multilayer	ACS nanomaterials ACS nanomaterials	10.22± 0.56 2.42± 0.21	0.75± 0.12 1.15± 0.06	516294 2018319		
Wolfram-based	Wolfram	Apnano, Israel	0.71± 0.01	1.22± 0.02	674129		
Manganese-based	Mn_2O_3 (pos. control)	Skyspring Nanomaterials	16.79± 3.33	2.27± 0.06	9525211	2.48± 1.89	
Barium-based	$BaSO_4$ NM-220 (neg. control)	Solvay			6232	1.14± 0.4	1.43± 0.50

Supplementary Table 1b: Surface-based values obtained from the four OP assays with standard deviations

Class	Material	ESR (CPH)	ESR (DMPO)	FRAS	Carbonyls	Carbonyls (normalized)
Silica	SiO_2_15_unmod	0.87± 0.12	1.03± 0.19	14.05± 1.72	345± 207	17250± 10372
	SiO_2_15_Amino	0.92± 0.09	1.07± 0.10	18.95± 0.79	275± 91	611± 203
	SiO_2_15_Phospho	0.91± 0.06	1.21± 0.13	8.01± 2.43	115± 57	639± 319
	SiO_2 NM-200				204± 95	
	SiO_2 NM-203			23.20	197± 135	548± 375
	SiO_2Levasil 50			4.76± 0.67		
	SiO_2Levasil 100			13.23± 0.10		
	SiO_2Levasil 300			13.94± 0.97		
	SiO_2Aerosil 200			23.20± 1.61		
Aluminosilicates	Kaolin	44.45± 6.02	7.28± 0.79	16.39± 1.07	2754± 912	5100± 1689
	Bentonite	13.72 ± 0.26	2.56 ± 0.23	86.99± 1.15	1048± 783	5516± 4120
	DQ12 (Quartz)					
Iron oxides	Fe_2O_3nanoform A	15.21± 0.61	0.87± 0.05	44.20± 2.03	1604± 846	1688± 890
	Fe_2O_3nanoform B	20.19± 2.01	1.04± 0.07	15.33± 2.24	2400± 899	3380± 1266
	Fe_2O_3larger	383.33± 26.71	9.24± 0.69	34.46± 1.72	8376± 3007	9307± 3341
Titania	TiO_2 NM-102	1.29± 0.07	1.79± 0.14	6.30± 0.60		
	TiO_2 NM-105	1.09± 0.06	1.66± 0.20	18.64± 2.76	1942± 1226	2428± 1532
	TiO_2 non-nano	1.21± 0.06	1.58± 0.15	13.77± 2.21		
Ceria	CeO_2 NM-211	268.60± 22.10	1.37± 0.06	14.16± 2.40		
	CeO_2 NM-212	189.41± 104.20	2.32± 0.07	12.90± 2.17		
Copper-Based	CuO	567.23± 34.34	5.42± 1.06	268.75± 2.07	12631± 2074	15037± 2469
	Cu-Phthalocyanine non-halogenated	11.34± 2.99	7.43± 2.46	11.51± 0.77	496± 110	787± 175
	Cu-Phthalocyanine Halogenated	185.32± 7.84	1.44± 0.18	18.15± 2.09	468± 249	641± 341
Zinc oxides	ZnO NM-110	126.76 ± 8.80	15 ± 1.00	150.99± 2.10	8417± 6607	15881± 12466
	ZnO NM-111			20.31± 2.26	5030± 3827	
DiketopyrrolopyrrolPigments	DPP_premixed	4.97± 0.46	0.64± 0.02	-23.11± 2.66	1812± 1622	
	DPP nano	1.75± 0.52	0.58± 0.06	3.39± 1.80	174± 43	

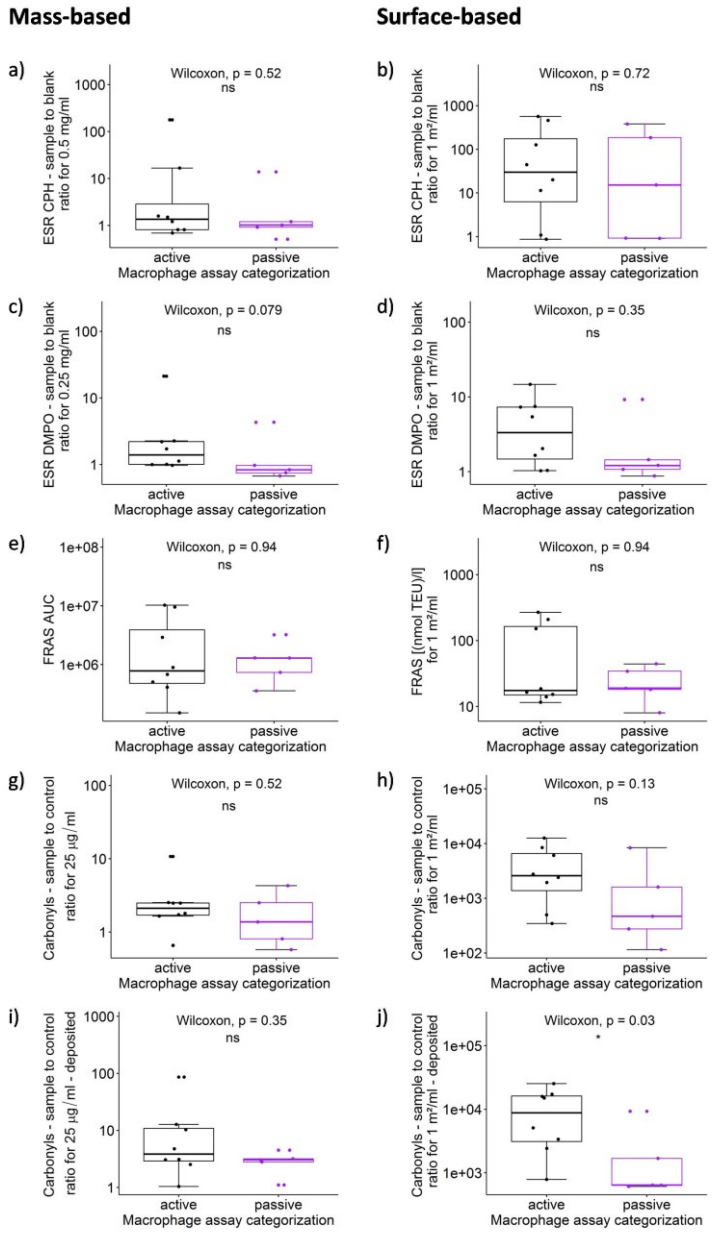
	DPP non-nano	1.15± 0.08	0.59± 0.11	1.83± 2.18	3563± 1061	
Carbon-based	Carbon black Graphene oxide Graphene 1-layer Graphene Multilayer				942± 833 104± 50	
Wolfram-based	Wolfram					
Manganese-based	Mn_2O_3 (pos. control)	461.42± 10.65	2.04± 0.43	209.76±	6080± 4642	
Barium-based	$BaSO_4$ NM-220 (neg. control)	2.30± 0.06	1.31± 0.13	6.65±	56± 20	65± 23

Supplementary Table 2: Overview of tested materials and data availability. Ticks indicate that materials were measure in this assay. (M) means that only values for uniform mass assays exist, (S) means that the material was only measured in assays performed on uniform surface areas.

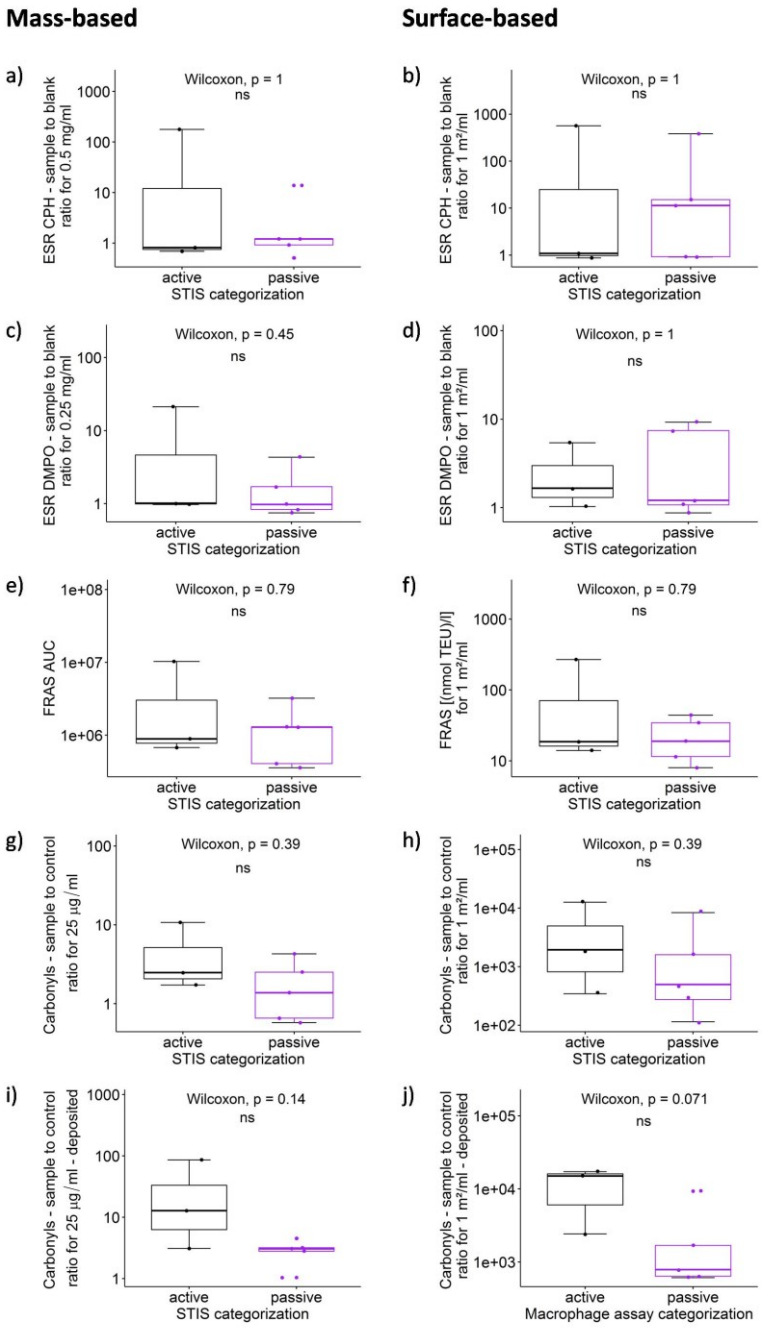
Class	Material	ESR	FRAS	Carbonyl
Silica	SiO_2_15_unmod*	✓	✓	✓
	SiO_2_15_Amino	✓	✓	✓
	SiO_2_15_Phospho	✓	✓	✓
	SiO_2 NM-200			✓
	SiO_2 NM-203		✓ (M)	✓
	SiO_2Levasil 50		✓ (S)	
	SiO_2Levasil 100		✓ (S)	
	SiO_2Levasil 300		✓ (S)	
	SiO_2Aerosil 200	✓ (M)	✓ (S)	
Aluminosilicates	Kaolin	✓	✓	✓
	Bentonite	✓ (S)	✓	✓
	DQ12 (Quartz)		✓ (M)	
Iron oxides	Fe_2O_3nanoformA	✓	✓	✓
	Fe_2O_3nanoformB	✓	✓	✓
	Fe_2O_3larger	✓	✓	✓
Titania	TiO_2 NM-102	✓	✓	
	TiO_2 NM-105	✓	✓	✓
	TiO_2 non-nano	✓	✓	
Ceria	CeO_2 NM-211	✓	✓	
	CeO_2 NM-212	✓	✓	
Copper-based	CuO	✓	✓	✓
	Cu-Phthalocyanine non-halogenated	✓	✓	✓
	Cu-Phthalocyanine halogenated	✓	✓	✓
Zinc oxides	ZnO NM-110	✓	✓	✓
	ZnO NM-111		✓	✓
Diketopyrrolopyrrol Pigments	DPP_premixed**	✓	✓	✓
	DPP nano	✓	✓	✓
	DPP non-nano	✓	✓	✓
Carbon-based	Carbon black	✓ (M)	✓ (M)	✓
	Graphene oxide			✓
	Graphene 1-layer	✓ (M)	✓ (M)	
	Graphene multilayer	✓ (M)	✓ (M)	
Wolfram-based	Wolfram	✓ (M)	✓ (M)	
Manganese-based	Mn_2O_3 (pos. control)	✓	✓	✓
Barium-based	$BaSO_4$NM-220 (neg. control)	✓ (S)	✓	✓

* SiO2_15_unmod ≙ SiO2 Levasil 200

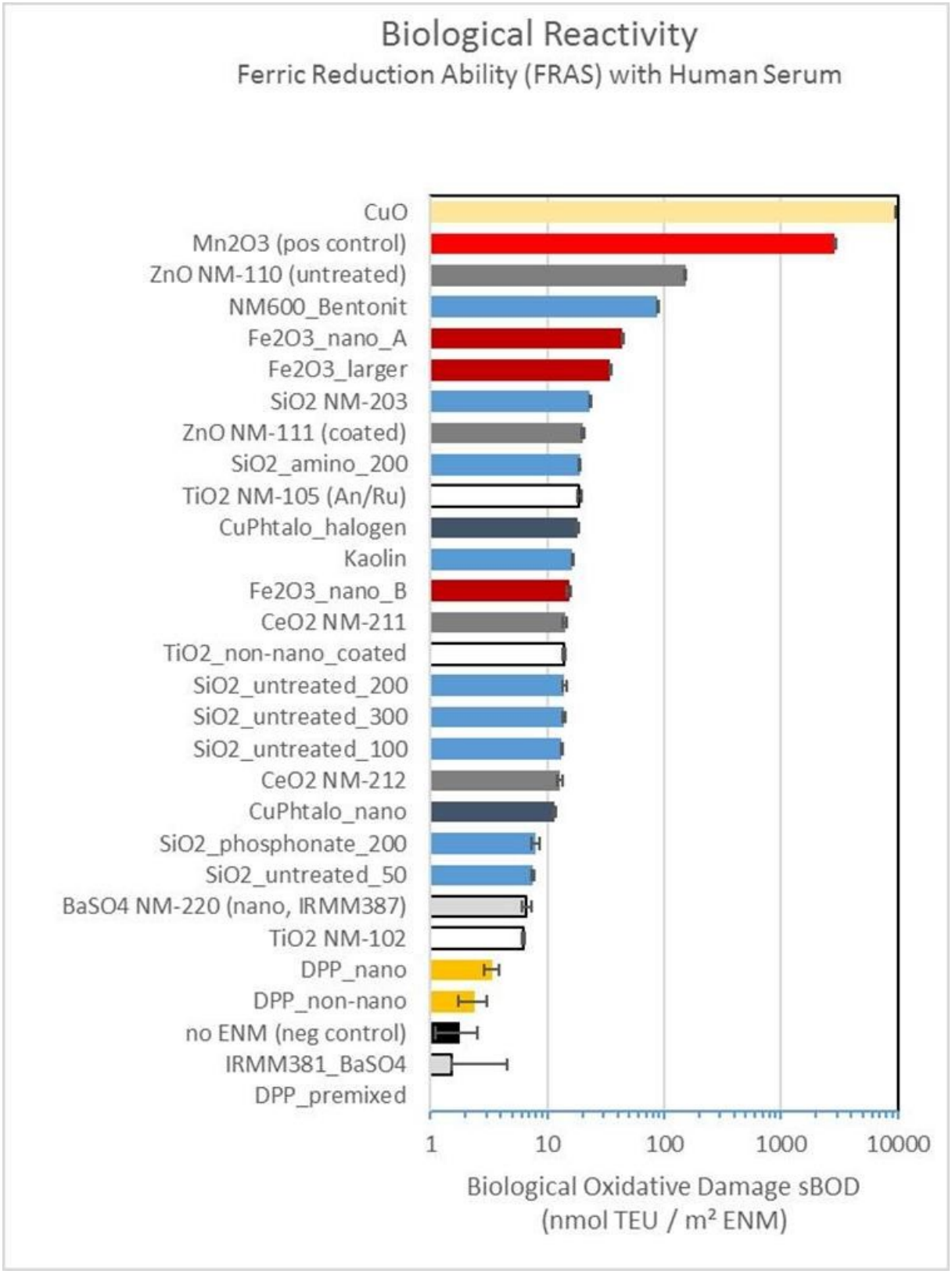
** DPP_premixed is a physical mixture of DPP with acrylic resin to enhance the compatibility of the pigment when it is integrated in certain polymer matrices.



Supplementary Figure 2: Boxplots comparing mass-based and surface-based values of complete cases across all assays from each of the four assays between NMs that were active in the macrophage assay and those that were passive. Depicted are values from ESR with CPH with same mass (a) and same surface area doses applied (b), ESR with DMPO with same mass (c) and same surface area doses applied (d),FRAS assay with same mass (e) and same surface area doses applied (f) and carbonylation assay with same mass (g) and same surface area doses applied (h). In addition, carbonyl values were normalized to the deposited dose of the respective NM for same mass doses (i) and same surface doses (j).



Supplementary Figure 3: Boxplots comparing mass-based and surface-based values of complete cases across all assays from each of the four assays between NMs that were active in STIS and those that were passive. Depicted are values from ESR with CPH with same mass (a) and same surface area doses applied (b), ESR with DMPO with same mass (c) and same surface area doses applied (d),FRAS assay with same mass (e) and same surface area doses applied (f) and carbonylation assay with same mass (g) and same surface area doses applied (h). In addition, carbonyl values were normalized to the deposited dose of the respective NM for same mass doses (i) and same surface doses (j).

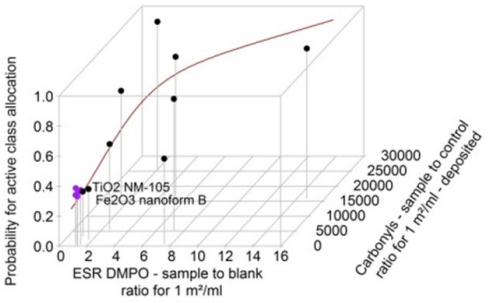


Supplementary Figure 4: Ranking of all materials by the FRAS assay, measured at equal surface dose (1m²/mL). Bentonite NM-600 cannot be grouped with other aluminosilicates. Coated ZnO NM-111 cannot be grouped with untreated ZnO NM-110. For other substance families, the similarity is high among the different (nano)forms of each substance.

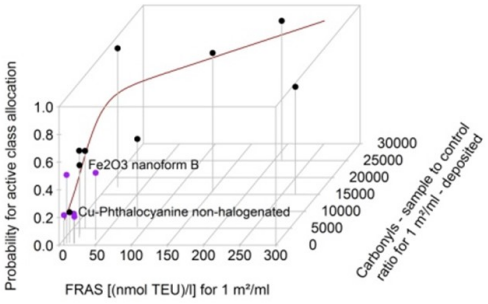
Supplementary Table 3: Outcome of univariate and multivariate logistic regression models with and without Firth' bias reduction.

Outcome variable	Predictors				AIC (glm)	AIC (brglm)
	CPH	DMPO	FRAS	Carbonyl		
Macro-phage assay	✓				23.11	23.12
		✓			22.20	22.25
			✓		20.68	21.30
				✓	18.88	19.23
	✓	✓			23.77	23.88
	✓		✓		20.51	21.29
	✓			✓	13.01	14.07
		✓	✓		22.38	23.06
		✓		✓	20.87	21.37
			✓	✓	20.76	21.67
	✓	✓	✓		19.02	20.41
	✓	✓		✓	8	12.74
	✓		✓	✓	14.94	16.39
		✓	✓	✓	22.72	23.72
	✓	✓	✓	✓	10	19.65
STIS	✓				14.77	14.81
		✓			15.31	15.38
			✓		13.37	13.77
				✓	10.92	11.24
	✓	✓			14.61	15.50
	✓		✓		14.77	15.61
	✓			✓	6	13.33
		✓	✓		14.42	15.26
		✓		✓	6	12.34
			✓	✓	12.74	13.41
	✓	✓	✓		16.42	17.59
	✓	✓		✓	8	14.92
	✓		✓	✓	8	14.59
		✓	✓	✓	8	14.68
	✓	✓	✓	✓	10	16.85

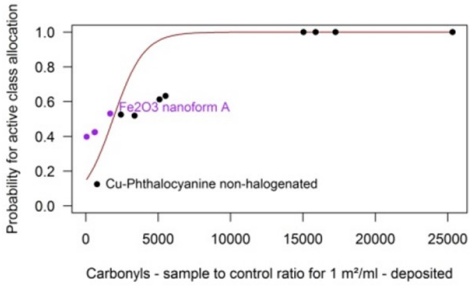
a)



b)

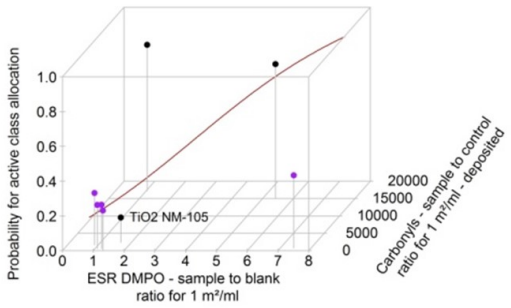


c)

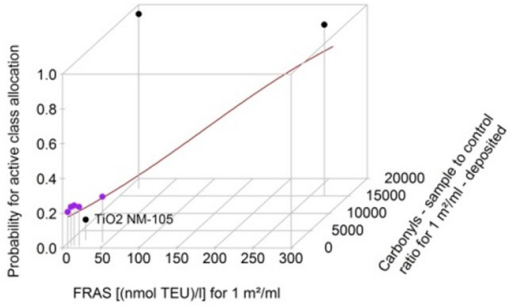


Supplementary Figure 5: Scatterplots showing the result of the logistic regression model including a) ESR DMPO and carbonylation data, b) FRAS and carbonylation data and c) carbonylation data only based on the data holding macrophage assay labels. On the y-axis, the probability of each NM to belong to the active class as given by the regression model is depicted. If this probability exceeds 0.5, the NM is assigned an active class label, otherwise it is predicted to be passive. Purple dots represent NMs that were assigned to the passive class by the macrophage assay, black dots represent active NMs with respect to the macrophage assay.

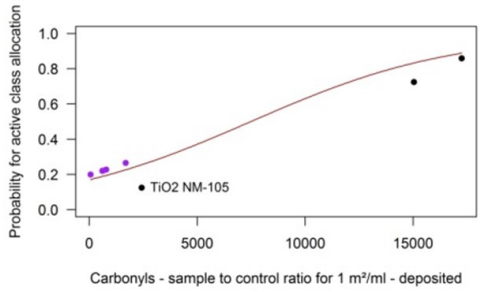
a)



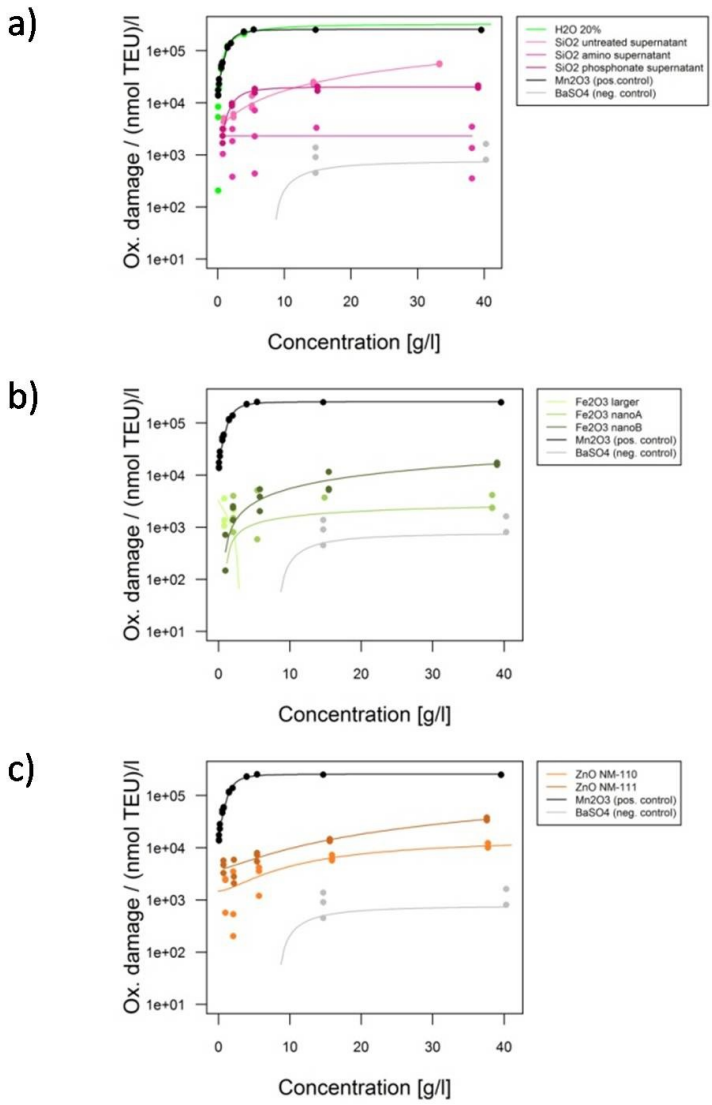
b)



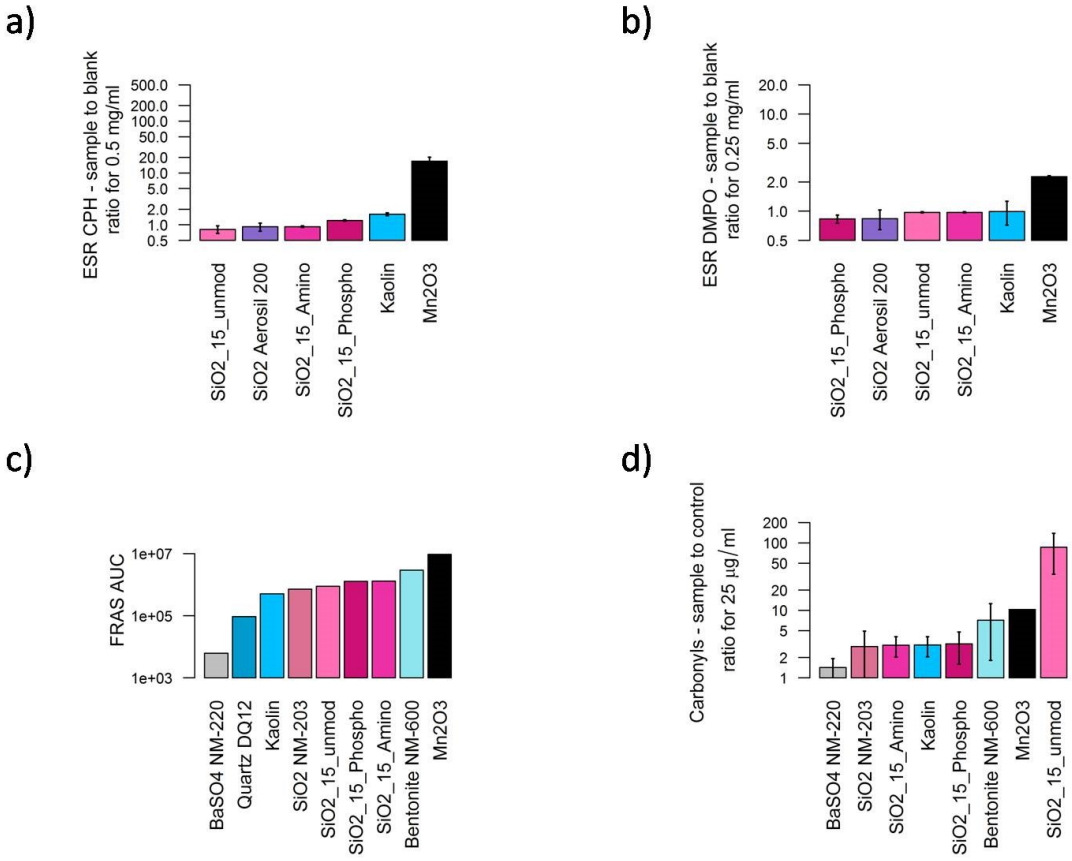
c)



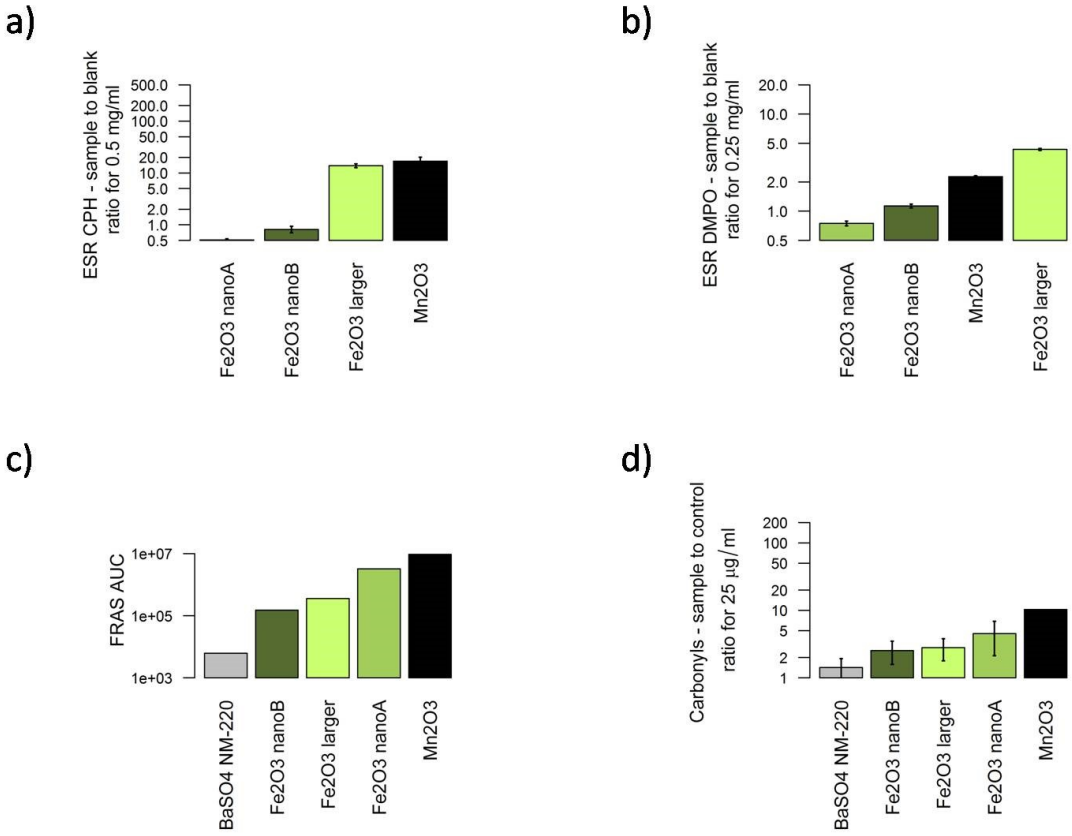
Supplementary Figure 6: Scatterplots showing the result of the logistic regression model including a) ESR DMPO and carbonylation data, b) FRAS and carbonylation data and c) carbonylation data only based on the data holding STIS labels. On the y-axis, the probability of each NM to belong to the active class as given by the regression model is depicted. If this probability exceeds 0.5, the NM is assigned an active class label, otherwise it is predicted to be passive. Purple dots represent NMs that were assigned to the passive class based on STIS results, black dots represent active NMs with respect to STIS.



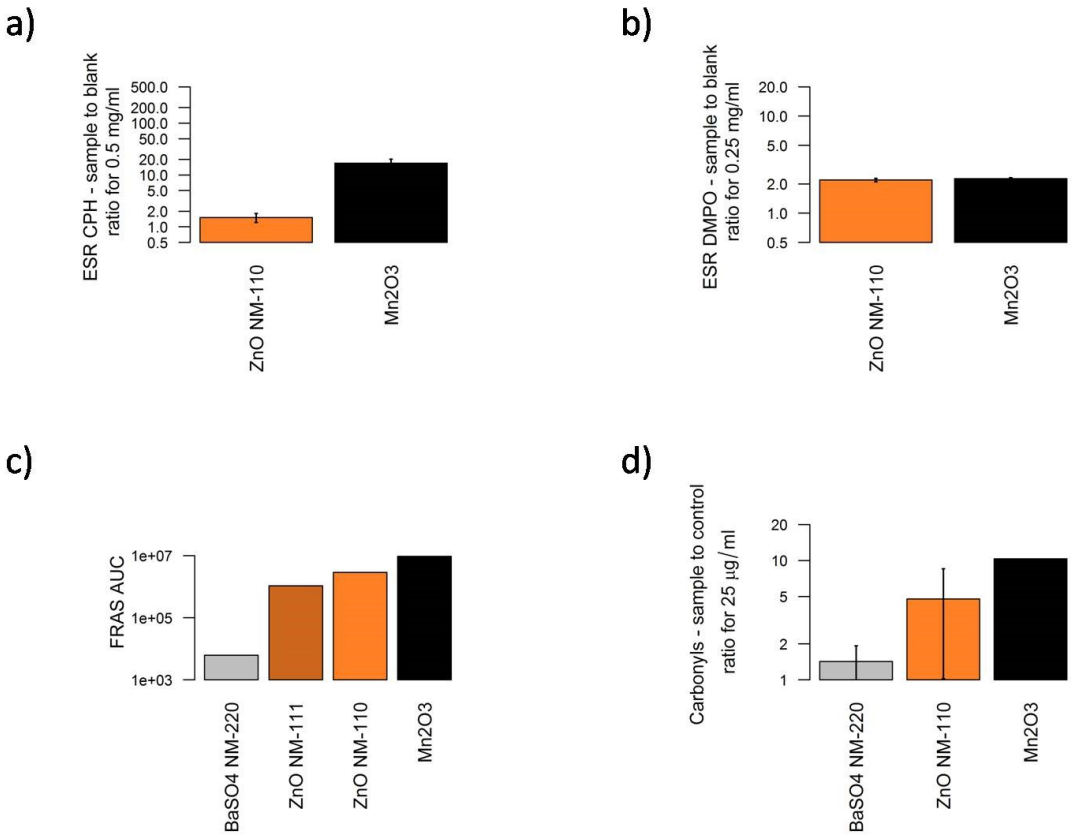
Supplementary Figure 7: Dose-response curves for a) the supernatants of the silica particles and H_2O 20%, b) iron particles and c) zinc particles obtained by the FRAS assay. Each figure also includes a grey dose-response curve for the negative control $BaSO_4$ as well as a black curve for the positive control Mn_2O_3.



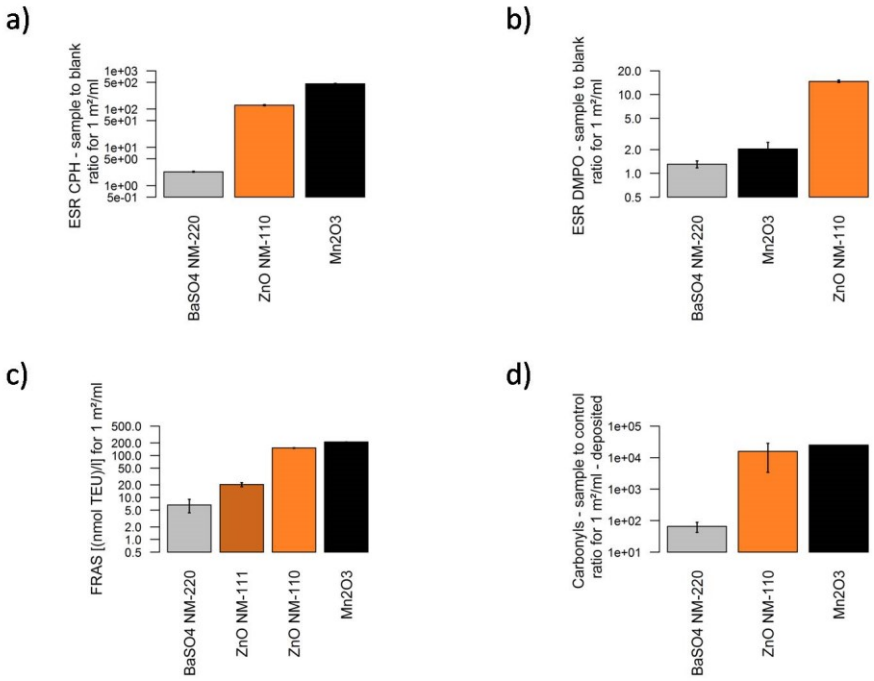
Supplementary Figure 8: Surface reactivity of silica particles measured with different assays and uniform mass. NMs in the plot are sorted by ranks, from the particle with lowest reactivity on the left end to the particle with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO probe, c) FRAS and d) protein carbonylation assay results are depicted.



Supplementary Figure 9: Surface reactivity of iron oxide particles measured with different assays and uniform mass. NMs in the plot are sorted by ranks, from the particle with lowest reactivity on the left end to the particle with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO probe, c) FRAS and d) protein carbonylation assay results are depicted.



Supplementary Figure 10: Surface reactivity of zinc oxide particles measured with different assays and uniform mass. NMs in the plot are sorted by ranks, from the particle with lowest reactivity on the left end to the particle with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO probe, c) FRAS and d) protein carbonylation assay results are depicted.



Supplementary Figure 11: Surface reactivity of zinc oxide particles measured with different assays and uniform surface area. NMs in the plot are sorted by ranks, from the particle with lowest reactivity on the left end to the particle with highest response at the right. a) ESR with CPH probe, b) ESR with DMPO probe, c) FRAS and d) protein carbonylation assay results are depicted.

Annex III

Supplementary information to

Bahl, A., Ibrahim, C., Plate, K., Haase, A., Dengjel, J., Nymark, P., Dumit, V. I., PROTEOMAS: A workflow enabling harmonized proteomic meta-analysis and proteomic signature mapping. *Journal of Cheminformatics* (2023) **15**, 34, https://doi.org/10.1186/s13321-023-00710-2 (2023).

Excel table available online under https://jcheminf.biomedcentral.com/articles/10.1186/s13321-023-00710-2. The table is too large for direct attachment to the thesis.