

Summary

The work described in this dissertation develops a predictive framework that links properties of engineered metal- and carbon-based nanomaterials to quantitatively predictable changes in transcriptomic profiles, with relevance to lung toxicity. This work tackles a key challenge in the field of nanosafety: traditional animal studies are costly and tend to focus on late-stage endpoints, while transcriptomic readouts are able to detect early molecular outcomes in the body, but the sheer complexity and dimensionality of gene-expression data make it difficult to use such data in routine risk assessments. To address this limitation, the dissertation integrates characterizations of nanomaterials, data from toxicogenomic studies, and machine learning approaches along with Organization for Economic Cooperation and Development (OECD)-approved validation strategies into a single, coherent computational workflow. In this way, the dissertation contributes to ongoing efforts to establish machine learning approaches that are applicable across diverse classes of engineered nanomaterials, while combining data-driven predictive modeling with mechanistic interpretability to support the development of next-generation *in silico* New Approach Methodologies (NAMs) for safety assessment.

This work examined two representative classes of toxicologically relevant nanomaterials – titanium dioxide nanoparticles (representatives of metal-based/metal-oxide nanomaterials) and multi-walled carbon nanotubes (representatives of high-aspect-ratio, carbon-based nanomaterials) – that differ markedly in chemical composition, shape, longevity, and anticipated mechanisms of action. These nanomaterials, therefore, provide a useful test case for assessing whether a broadly applicable machine learning framework can be established and optimized for predicting transcriptomic alterations across diverse classes of engineered nanomaterials. The central hypothesis of the dissertation is that high-dimensional transcriptomic data can be reduced to predictive, biologically interpretable endpoints suitable for mechanistic hazard assessment and potency ranking, and that these endpoints, together with selected transcriptomic biomarkers, can be modeled from nanomaterial physicochemical properties and exposure-related descriptors using validated machine learning approaches. Collectively, the work described in this dissertation extends the reach of nanosafety modeling beyond single endpoint assays and singular gene predictions, demonstrating the feasibility of developing and validating machine learning (ML) workflows that predict both particular transcriptomic biomarkers and broad indices of global transcriptional

activity from nanomaterial descriptors. The developed framework simplifies the predictions of early molecular responses in processes related to inflammation, fibrosis, oxidative stress, and tissue repair. Importantly, the predictive models were constructed according to the regulatory-established guidelines of internal and external validation, and the estimation of the applicability domain of the trained models. The rigorous validation underlines the potential applicability of the developed framework in the context of regulatory toxicology, Integrated Approaches to Testing and Assessment (IATA), and hazard evaluation based on non-animal methods.

This dissertation links the fields of materials chemistry and systems toxicology by establishing a scalable methodology that integrates physicochemical characterization of exposed nanomaterials with omics datasets, dimensionality-reduction techniques, and validated machine learning approaches, in order to provide an applicable workflow for converting complex transcriptomic information into a mechanistically interpretable endpoint. The resulting approach has the potential to reduce the need for large-scale animal experimentation, aid early warning of potentially hazardous properties of nanomaterials, contribute to the ranking of their potencies, and help to inform the development of safer nanomaterials.