PERSPECTIVE
Emma L. Schymanski et al.
Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: potential and challenges

ISSN 2050-7887
Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: potential and challenges†

Emma L. Schymanski, a Nancy C. Baker, a Antony J. Williams, c Randolph R. Singh, ad Jean-Pierre Trezzi, ef Paul Wilmes, f Pierre L. Kolber, gh Rejko Kruger, gh Nicole Paczia, i Carole L. Linster i and Rudi Balling i j

Connecting chemical exposures over a lifetime to complex chronic diseases with multifactorial causes such as neurodegenerative diseases is an immense challenge requiring a long-term, interdisciplinary approach. Rapid developments in analytical and data technologies, such as non-target high resolution mass spectrometry (NT-HR-MS), have opened up new possibilities to accomplish this, inconceivable 20 years ago. While NT-HR-MS is being applied to increasingly complex research questions, there are still many unidentified chemicals and uncertainties in linking exposures to human health outcomes and environmental impacts. In this perspective, we explore the possibilities and challenges involved in using cheminformatics and NT-HR-MS to answer complex questions that cross many scientific disciplines, taking the identification of potential (small molecule) neurotoxicants in environmental or biological matrices as a case study. We explore capturing literature knowledge and patient exposure information in a form amenable to high-throughput data mining, and the related cheminformatic challenges. We then briefly cover which sample matrices are available, which method(s) could potentially be used to detect these chemicals in various matrices and what remains beyond the reach of NT-HR-MS. We touch on the potential for biological validation systems to contribute to mechanistic understanding of observations and explore which sampling and data archiving strategies may be required to form an accurate, sustained picture of small molecule signatures on extensive cohorts of patients with chronic neurodegenerative disorders. Finally, we reflect on how NT-HR-MS can support unravelling the contribution of the environment to complex diseases.

Environmental significance

Non-target high resolution mass spectrometry has attracted immense interest regarding the potential for increased characterisation of chemicals and exposures in environmental studies. However, high quality studies remain difficult to perform and often yield fewer successful identifications than desired. We present our multi-disciplinary perspective on tackling potential causes for this discrepancy and the complexity involved in capturing the knowledge needed to investigate the impact of chemicals on human health in the context of long-term diseases. We look at how to mine patient and expert knowledge to understand, measure and validate potential neurotoxicants and explore the potential for NT-HR-MS to provide small molecule data signatures in cohorts and build the knowledge required for greater future understanding of environmental impacts on complex diseases.

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c9em00068b
Introduction

Rapid developments in analytical techniques over the last decades have opened up new opportunities such as non-target high resolution mass spectrometry (NT-HR-MS) to explore the impact of humans on the environment,\(^1\) as well as the effect of environmental contamination on humans\(^2,3\) and other organisms.\(^4,5\) With growing accessibility and maturity in NT-HR-MS analytical methods, these are being applied to increasingly complex research questions and hypotheses.\(^6-8\) In this perspective article, we take a close look at the challenges in identifying “neurotoxins” in the context of chronic neurodegenerative diseases and use this case study to explore the potential and limitations of NT-HR-MS to support answering these questions. Ultimately, solving complex challenges such as these will need a systems approach,\(^9,10\) requiring knowledge that crosses many disciplines. While the main focus of this article, determining which (relevant) chemicals are present in environmental or biological samples (let alone their mode of action), is already a daunting task, it is a necessary foundation to enable the elucidation of functional effects and generation of hypotheses for a greater understanding of the causes of complex diseases and environmental impacts.

NT-HR-MS, typically coupled with separation techniques such as liquid or gas chromatography, is becoming an increasingly popular approach for the broad screening of complex environmental samples.\(^1\) NT-HR-MS can be performed on low sample amounts and reach ppb (or lower) detection limits in many cases, with data acquisition now often able to capture both known and unknown chemicals in a single measurement.\(^11\) Ideally in NT-HR-MS, post-acquisition data processing is used to identify (1) the known chemicals, i.e. the “targets” or also “biomarkers” in the context of metabolomics and disease; (2) potential chemicals of interest based on prior knowledge, i.e. lists of chemicals (“suspects”) and (3) the relevant “unknowns” or “non-targets” using some form of prioritisation.\(^1\) Data generated in NT-HR-MS can be archived for later use and exploration (also termed retrospective screening).\(^12\) While NT-HR-MS has been declared “ready to go”,\(^1\) in this article we explore how to gather the knowledge required to take NT-HR-MS to the next level of helping unravel the cause and mechanisms of environmental exposure in the context of late-developing complex diseases, where a lifetime of exposure to thousands of chemicals may potentially impact disease progression. Forming epidemiological connections, however, does not just involve the detection of chemicals of interest, but also gathering information on qualitative aspects and characteristics of potential neurotoxins, as well as time and duration of exposure.

To date, NT-HR-MS of environmental samples has been commonly directed by substance classes. Several examples include using suspect screening approaches to investigate classes such as pharmaceuticals,\(^13-14\) pesticides\(^15,16\) or antibiotics.\(^17-18\) Further studies are now using publicly-available suspect lists in their screening efforts.\(^19,20\) Other early efforts include comprehensive, exploratory characterisation efforts of wastewater that went beyond compound classes.\(^21,22\) As NT-HR-MS matures, effect-directed analysis (EDA) studies are beginning to show some level of success with established bioassay endpoints such as mutagenicity.\(^23\) The growing popularity of “big data” techniques has seen the expansion into more data exploration-based methods such as virtual EDA, clustering, trend analysis and fingerprinting, summarized elsewhere,\(^1\) inspired partially by metabolomics efforts. From a metabolomics perspective, great investments have been made into the detection of biomarkers for certain diseases (discussed further

---

Associate Professor Emma Schymanski is a Luxembourg National Research Fund (FNR) ATTRACT Fellow and head of the Environmental Cheminformatics group at the Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg. Dr Randolph Singh is a postdoctoral fellow in the same group. Dr Nancy Baker is a consultant at Leidos, Research Triangle Park, USA. Dr Antony Williams is a Computational Chemist, National Center of Computational Toxicology, US EPA. Dr Jean-Pierre Trezzi is postdoctoral fellow affiliated with the Integrated Biobank of Luxembourg, Luxembourg Institute of Health as well as the Eco-Systems Biology group at LCSB, which is headed by Associate Professor Paul Wilmes. Dr med. Pierre Kolber, MD is a member of the Neurology department at the Centre Hospitalier de Luxembourg and the Clinical and Experimental Neuroscience group at LCSB, which is headed by Professor Dr med. Rejko Krüger, also an FNR PEARL Fellow and Director of Transversal Translational Medicine at the Luxembourg Institute of Health. Dr Nicole Paczia, former joint-manager of the Metabolomics Platform at LCSB, now heads the Metabolomics Core facility at the Max Planck Institute in Marburg. Dr Carole Linster is head of the Enzymology and Metabolism Group, LCSB. Professor Rudi Balling is Director of the LCSB.
below) and the trend is now towards using these approaches for a more mechanistic understanding of the biology behind the biomarkers, primarily using targeted analysis techniques. Yet the conundrum remains: there are orders of magnitude more “unknowns” than “knowns” in both biological and environmental samples, despite compound databases now containing over 100 million chemicals and impressive improvements to computational mass spectrometry workflows. Additionally, despite greater identification efforts, toxic effects observed in the environment (e.g. in EDA studies) also remain unexplained more often than not, indicating that a greater understanding of the entire environmental system is still required.

Currently, the key to successful NT-HR-MS is finding “known unknown” chemicals (chemicals documented to exist but unknown upfront to the investigator) of high relevance to the study question in an efficient manner (i.e. via prioritisation). The full identification of unknowns is still extremely time consuming and, while it is easy nowadays to get tentative candidates for many detected masses in NT-HR-MS, candidate selection and validation remains challenging and suspect list/compound database choice has a dramatic influence in the outcome of identification efforts. A well-designed suspect screening approach, i.e. searching for a discrete list of chemicals potentially relevant to the study question, is an ideal way to find masses and thus candidates of particular interest quickly. However, there is a delicate balance in suspect screening. Small, carefully validated lists containing tens to a few hundred entries result in few suspect hits ever being found, but if found, are likely highly relevant. On the other hand, suspect screening using large databases (containing tens of thousands to millions of entries) for all matching candidates rapidly turns into a non-target identification challenge with multiple matching candidates per mass. The current trend in NT-HR-MS is towards compiling very large lists (in the order of tens of thousands of entries) to enable better coverage which is, in effect, closing the gap between suspect screening and typical non-target or unknown identification approaches. A variety of statistical approaches, such as replicates and multiple correction testing, can be used to recognise and reduce the resulting false positives, summarized elsewhere. As validation of any exact mass hit (irrespective of the size of the suspect list) in NT-HR-MS is essential, additional information must also be used to support the candidate structure. This includes orthogonal analytical evidence (e.g. chromatographic retention behaviour, fragmentation information) as well as so-called metadata, i.e. additional information that may indicate that this chemical is relevant to the study question. For many years, literature references have been used to prioritise highly interesting candidates and these have now been built into many identification approaches. In an environmental context, specialised resources such as the CompTox Chemicals Dashboard, the Human Metabolome Database (and related resources such as DrugBank and T3DB), and datasets on the NORMAN Suspect List Exchange offer metadata categories relevant for environmental screening, such as exposure data, activity in in vitro screening bioassays, predictive toxicity values (quantitative structure activity relationships, QSARs), literature counts and occurrence in various matrices of interest. Connecting the metadata associated with mixtures to their individual constituents via the so-called “MS-ready” form now allows screening of mixtures in HR-MS.

While connecting masses detected in environmental samples to “known unknowns” and related metadata is now achievable with NT-HR-MS, connecting chemical exposure to diseases opens up many more challenges. The concept of the “exposome” was first introduced in 2005 by Wild et al. with the statement “at its most complete, the exposome encompasses life-course environmental exposures (including lifestyle factors), from the prenatal period onwards”. This definition comes with numerous challenges, not least the time frame of the life-course, which are not the focus of this current article. Other articles, including (but not limited to) those cited here, contain further discussions and refinements to the concept. The “meet-in-the-middle” approach has been introduced to address the challenge of identifying causal relationships that link exposures and disease outcomes, indicating the need for the connection of exposomics to other “omics” levels such as epigenomics, metabolomics and transcriptomics. While biomarkers have resulted in many impressive epidemiological studies, the vast majority of these are based on targeted studies of well-known pollutants; the field is at the cusp of being able to take full advantage of NT-HR-MS and these are the views we wish to present in this perspective.

In the following sections, we explore how to capture more topic-specific metadata and maintain links back to the associated literature and patient knowledge using the chosen case study of neurotoxicants in the context of chronic neurodegenerative diseases. We will then look at what analytical approaches may be needed to capture broad concepts such as “potential neurotoxicants” and explore what may be missing and remain unseen, before moving onto the challenges associated with relating this information back to the study question (e.g. confirming whether potential candidates are responsible for neurotoxicity and the final disease state). Finally, we reflect on how advances in analytical and data technologies could be leveraged and built into long term cohort studies to build a greater understanding of the environmental influences on complex disease states.

**Cheminformatics to capture the chemical space of “neurotoxicants”**

As mentioned above, the key to successful NT-HR-MS is finding chemicals of high relevance to the study question in an efficient manner (prioritisation). A well-designed suspect screening, i.e. searching for a discrete list of chemicals relevant to the study question, is an ideal way to quickly find masses (features) and thus candidates of particular interest. A number of lists of neurotoxicants were compiled for the purposes of this perspective, summarized in Table 1 and described further in the ESI. These include small, carefully validated lists explained extensively in the source publications (DNTEFFECTS [https://comptox.epa.gov/dashboard/chemical_lists/dnteffects], DNTINVIVO [https://
comptox.epa.gov/dashboard/chemical_lists/dntinvivo) and HUMANNEUROTOX\(^a\) (https://comptox.epa.gov/dashboard/chemical_lists/humanneurotox)), where few suspect hits will ever be found in environmental samples. The DNTPOTNEG\(^b\) (https://comptox.epa.gov/dashboard/chemical_lists/dntpotneg) list contains potential negative controls for neurotoxicity which, if found, should not be associated with neurotoxic effects, enabling a second layer of data quality verification of neurotoxicant findings. Two larger lists (NEUROTOXINS\(^c\) (https://comptox.epa.gov/dashboard/chemical_lists/neurotoxins) and LITMINEDNEURO\(^d\) (https://comptox.epa.gov/dashboard/chemical_lists/litminedneuro) have also been compiled for this perspective. The NEUROTOXINS list was compiled from public resources including ChEBI,\(^55,56\) Wikipedia,\(^57,58\) T3DB,\(^40\) with further details given in the ESL.\(^\dagger\)

LITMINEDNEURO contains chemicals associated with neurotoxicity compiled through systematic literature mining of PubMed\(^59\) using Medical Subject Heading (MeSH) terms\(^60\) and associating these with single chemical substances (where possible) using previously published methods.\(^61\) Articles were identified in which a nervous system disease was annotated with the MeSH node C10 through the MeSH tree with disease subheading “chemically induced” and subheading “toxicity”, “poisoning”, or “adverse effects”. Nerve diseases caused by trauma and manually identified “common English terms” that could not be associated with any specific chemicals (e.g. “particulate matter”, “contrast media”) were omitted. In total 4528 chemicals were identified; all chemicals with 5 or more literature references were registered in the Dashboard and included in the final list. The output of this processing was exported to Microsoft® Excel and is included as ESIF\(^\dagger\) and available on FigShare.\(^53\) The CASRN and the CompTox Chemicals Dashboard substance identifier (DTXSID) were included in the spreadsheet for chemicals for which this relationship was captured and associated with MeSH identifiers. The list code Entries and references Description

<table>
<thead>
<tr>
<th>List code</th>
<th>Entries and references</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNTEFFECTS (<a href="https://comptox.epa.gov/dashboard/chemical_lists/dnteffects">https://comptox.epa.gov/dashboard/chemical_lists/dnteffects</a>)</td>
<td>96 (ref. 49 and 63)</td>
<td>Chemicals with data demonstrating effects on neurodevelopment</td>
</tr>
<tr>
<td>DNTINVIVO (<a href="https://comptox.epa.gov/dashboard/chemical_lists/dntinvivo">https://comptox.epa.gov/dashboard/chemical_lists/dntinvivo</a>)</td>
<td>33 (ref. 50 and 64)</td>
<td>A (non-exhaustive) list of compounds documented to trigger developmental neurotoxicity (DNT) in at least two different laboratories</td>
</tr>
<tr>
<td>DNTPOTNEG (<a href="https://comptox.epa.gov/dashboard/chemical_lists/dntpotneg">https://comptox.epa.gov/dashboard/chemical_lists/dntpotneg</a>)</td>
<td>41 (ref. 50 and 65)</td>
<td>Suggested potential negative controls for developmental neurotoxicity (DNT) assays. Statins can also be used, see <a href="https://comptox.epa.gov/dashboard/chemical_lists/statins">https://comptox.epa.gov/dashboard/chemical_lists/statins</a></td>
</tr>
<tr>
<td>HUMANNEUROTOX (<a href="https://comptox.epa.gov/dashboard/chemical_lists/humanneurotox">https://comptox.epa.gov/dashboard/chemical_lists/humanneurotox</a>)</td>
<td>190 (ref. 51, 66 and 67)</td>
<td>A set of chemicals identified as potential neurotoxidants by the authors using literature searching, not necessarily active neurotoxicants</td>
</tr>
<tr>
<td>NEUROTOXINS (<a href="https://comptox.epa.gov/dashboard/chemical_lists/neurotoxins">https://comptox.epa.gov/dashboard/chemical_lists/neurotoxins</a>)</td>
<td>511 (ref. 52, 62 and 68)</td>
<td>A list of chemicals reported as neurotoxicants, compiled from public resources (source file with details on Zenodo(^\dagger))</td>
</tr>
<tr>
<td>LITMINEDNEURO (<a href="https://comptox.epa.gov/dashboard/chemical_lists/litminedneuro">https://comptox.epa.gov/dashboard/chemical_lists/litminedneuro</a>)</td>
<td>1243 (ref. 53, 54 and 69)</td>
<td>Chemicals associated with neurotoxicity compiled through automated literature mining of PubMed using MeSH terms (node C10, subheadings “chemically induced”, “toxicity”, “poisoning”, or “adverse effects”) and associating these with single chemical substances</td>
</tr>
</tbody>
</table>
Parkinson’s disease. For example, non-smoking carriers of the LRRK2 Gly2385Arg gene have increased risk of developing PD; while another study (published in German) discusses the balance between neurotoxic and neuroprotective effects of nicotine and smoking. Understanding of the context is critical and the interface through the Excel file provided as ESI† is one method to follow-up the literature references for specific chemicals and understand the relevance of the suspect for their particular study question (see examples for nicotine and caffeine above).

### Related structures: mapping “chemicals” and “substances”

Reviewing some of the lists in Table 1 via the Dashboard reveals several interesting cases where new cheminformatics approaches are needed and already under development to capture less well-defined substance information in the form of discrete chemical structures suitable for NT-HR-MS screening studies. Several examples are given in Fig. 1 and more extensive details given in the ESI†.

One of the most interesting challenges associated with capturing chemical information for NT-HR-MS related to neurotoxicity (or any potential health impact) is relating what we as humans consume relative to discrete chemical components. For instance, while toxicity testing is most commonly performed on a chemical such as caffeine, patients will instead report coffee (or tea or energy drink) consumption. Coffee, for instance, is documented to contain over 1000 chemical constituents, which require a variety of analytical techniques for detection.

Nicotine is another common example where the chemical tested is often nicotine (or an associated salt or mixture), but patients will instead report e.g. smoking habits.

---

**Fig. 1** Top: Nicotine (red box) and selected mixtures, salts and components. Middle: Diazinon (red box) and selected mapped transformation products. Bottom row: The PCB class and selected mapped individual PCB compounds. While the generic PCB representation is suboptimal visually, it allows a full, automated enumeration of all members of the PCB class via the ChemAxon Markush Enumeration module to create “Markush child” entries (see label) – an important consideration for upscaling these cheminformatics efforts. All chemicals display the structure, DTXSID, number of assays (active/total) in TOXCAST and the data sources.
is well known that nicotine is not the only chemical in cigarettes that may cause detrimental health effects. Several thousand chemicals have been identified in cigarettes\textsuperscript{78,79} with 599 chemicals listed as additives.\textsuperscript{80} Capturing such knowledge (e.g. via cross-mapping and adding as lists or related substances in databases) will be increasingly important to help reconcile NT-HR-MS results in the future, yet expand suspect lists even further. The connection of individual chemical structures to mixtures\textsuperscript{81} and the associated metadata can enable high throughput screening of NT-HR-MS data, including e.g. toxicity and product information,\textsuperscript{82} which are often associated with mixtures. It also enables better data interpretation downstream. Furthermore, the collection of chemicals into lists combined with well-selected metadata can ensure rapid prioritization by score and can help rapidly pinpoint highly promising candidates amongst hundreds to thousands of possible masses and chemical structures, as demonstrated in Fig. 2. More specific details about how such information can be included in NT-HR-MS studies are given in the ESI.\textsuperscript{†}

Capturing patient and medical knowledge

While the preceding sections discuss capturing and exploring the documented chemical and medical knowledge in the context of neurotoxicity and using this during identification in NT-HR-MS, capturing the patient knowledge is an incredibly important part of connecting chemicals to disease. For instance, neurodegeneration in chronic diseases such as Parkinson’s disease takes place for many years before the first motor manifestations appear that define the clinical diagnosis of this movement disorder.\textsuperscript{83} The time span between the diagnosis of PD and the appearance of the initial non-motor symptoms such as depression, REM sleep behavior disorder, hyposmia or chronic constipation (defining the prodromal phase of PD), can range from approximately 5 up to 30 years. Even before this prodromal phase of the disease, neuronal dysfunction and neuronal cell death are already ongoing and might precede the prodromal phase by many years or even decades.\textsuperscript{84} Thus, in order to analyze the potential neurotoxants and environmental factors that might lead to neurodegenerative diseases, one has to retrace the exposure of patients long before the diagnosis of the disease, better still before the first neuronal dysfunctions appear. Furthermore, as discussed below, there may no longer be any traces of the chemical to which the patient was exposed in cohort samples and thus the patient’s memory of their chemical exposures may be the only documentation that could provide clues to potential causative agents. In the few longitudinal PD patient cohorts that exist, such as the Luxembourg Parkinson’s study,\textsuperscript{85} the first step in analyzing the environmental risk factors leading to neurodegeneration is by using detailed validated patient questionnaires retrospectively trying to trace the patient’s professional or leisure activities throughout their entire life (see Fig. 3 for a screenshot from the REDCap\textsuperscript{86} system (https://www.project-redcap.org/)).

These classical epidemiological strategies apply self-reporting questionnaires that are not without flaws, as there is a non-negligible recall bias, especially considering the fact that patients with neurodegenerative diseases are mainly elderly people and to some degree affected by more or less pronounced cognitive defects. Additional confusion can occur due to misunderstanding of the terminology, for instance weed killers approved for household use are not always understood as a herbicide by patients and thus pesticide/herbicide exposure might be neglected by patients in some questionnaires. Nevertheless, these questionnaires can provide first hints towards potential environmental risk factors using the knowledge of the patients themselves. Anecdotal cases such as the one mentioned above are highly subjective and difficult to avoid with a generic survey form as in Fig. 3. However, the association of chemicals with products (e.g. via synonym linking such as aspirin/acetosalicylic acid (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=aspirin), plus ingredients listing such as the cigarette example above) and products with exposure scores, such as CPDat,\textsuperscript{87} may help connect what patients report and the information required for chemical analysis. Obtaining chemical and approved product information from authorities and linking this to e.g. agricultural land use can support

![Fig. 2 An example of how metadata can help candidate selection in high throughput NT-HR-MS studies, using nicotine in MetFrag.\textsuperscript{82} Further details in the ESI.†](image-url)
epidemiological studies\textsuperscript{87,88} as well as ecological assessments.\textsuperscript{89}

Product information from authorities, while often kept confidential, is essential for improved connections between NT-HR-MS and epidemiology. As such, greater efforts to share this information in the public domain are desperately needed and several authorities are now supportive of information sharing – as evidenced for example by the many lists now available on the CompTox Chemicals Dashboard\textsuperscript{90} and the NORMAN Suspect Exchange.\textsuperscript{41,42}

Using current knowledge, one can hypothesise that neurotoxicant exposure might precede neurodegenerative diseases such as PD by many decades, potentially during young adulthood or childhood (or even \textit{in utero}), launching a cascade of malfunctions progressively leading to apoptosis of dopaminergic neurons. This makes finding a representative sample from patients very challenging, as ideally the sample used should contain the neurotoxicants from that initial exposure period, which will also potentially transform over time. See discussion in the next section. Considering PD specifically, the risk for PD due to the usage of hydrocarbon-based compounds such as pesticides, herbicides or fungicides is quite established by now, especially professional pesticide usage.\textsuperscript{91,92} However, metal-based compounds are also suspected to contribute to an increased risk for PD or atypical parkinsonian syndromes (aPS), such as progressive supranuclear palsy (PSP) – a rapidly progressive aPS with prominent falls and gaze palsy as major clinical hallmarks. Some studies hypothesise a role of toxic metals and metalloids (such as arsenic) in the pathogenesis of PSP.\textsuperscript{93} Exposure to heavy metals (such as lead) is suggested to be associated with an increased risk for PD. However, difficulties in making differential diagnoses between these disorders can confound these conclusions. A further factor to consider is epigenetics, where epigenetic imprints may reflect exposure in the preceding generation and could add some prior knowledge with respect to the effects observed, if not the actual chemical exposure itself.\textsuperscript{94} Thus, while surveys can help capture patient knowledge, in the days of “big data” and with the trend towards personalised medicine, it is important to consider how “on the spot” analysis and digital archiving of small molecule data during cohort sample collection may enable a long term generation of a baseline dataset for greater understanding of disease progression and diagnosis in the future.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{tanner_survey.png}
\caption{Screenshot of the “Tanner survey” used to capture patient information, using REDCap (\url{https://www.project-redcap.org/}).\textsuperscript{86,87}}
\end{figure}
Available sample matrices

When studying disease-related neurotoxicant exposure, human-derived samples are of primary interest as these depict the pathological changes and can be monitored over time. However, not all of these matrices can be obtained easily and depending on the matrix used, the results can vary significantly. For instance, traces of many compounds are very short-lived in complex matrices like blood, as they are quickly metabolised and excreted by the kidneys. If one analyzes the urine, the window of detection can be delayed (the time for the compound to be filtered out from the blood by the kidneys), but is equally short-lived in the context of long-term diseases. The individual’s kidney function, hydration status and metabolism complicate this aspect, as the amount of urine and the frequency of the mictions vary hugely between each individual patient. Not only does the time window of detection vary, but also the concentration of the compound detected in the matrix can vary hugely from one time period to another and if the peak of concentration is missed, the results can become insignificant. Additionally, if one analyzes, for example, the total 24 h diuresis in an attempt not to miss the detection window, the compound can become so diluted that it can falsely render the exposure as insignificant. Furthermore, for those substances where the metabolites are not known or documented, the analyst will have to use additional experiments and data processing approaches to identify potentially relevant metabolites. An impressive recent effort shows the possibility for NT-HR-MS and advanced data processing to provide much broader coverage than the typical targeted urine analysis.95 Yet another very recent effort reports initial developments towards non-invasive drug monitoring via NT-HR-MS analysis of skin.96 However, the transient nature of some compounds and matrices has to be kept in mind while analysing potential neurotoxicants in human samples. Samples containing the compounds for a longer period of time seem to be, in this regard, more valuable. Bone is an interesting matrix for metals exposure and, for example, lead concentration can be measured with 109Cd-excited K-shell X-ray fluorescence, representing a biomarker of cumulative lead exposure.97 Teeth have also been used to estimate long-term exposure to metals and selected organic compounds, with especially recent studies drawing conclusions for various aspects of neurotoxicity.98–101 Hair sampling is a relatively non-invasive longer-term sample, due to the continual release of fat-soluble contaminants that have been proven to provide biomarkers for monitoring human exposures to environmental chemicals such as pesticides and metals.102 In addition, information obtained from hair analysis is more representative of an individual’s level of exposure than other matrices such as urine and blood.103 As many (easily available) biological matrices are so transient, linking chemical exposure to affected pathways with e.g. metabolomics approaches will be increasingly relevant.24 As recent research has indicated a potential connection between the gut microbiome and disease progression (see text below), there is also increasing interest in faecal samples and the biological, as well as chemical, information contained within.

Apart from human-derived samples, environmental samples can also be studied for the presence of neurotoxicants as surrogates for human samples, knowing that biological matrices can be challenging to access. Wastewater samples, for example, may reveal group exposure to specific contaminants as this matrix is a collection from a population and presents an idea of what the bulk population of a specific area is currently (or was) exposed to.11 As wastewater is released back into the environment after treatment, studying local fish, crustacea or even resting eggs104 may show which specific neurotoxicants have been present over a longer time frame and have bio-accumulative potential. Sediments and soil cores, on the other hand, may give an idea of historical exposure that may have happened over the last few years or decades,105 although degradation and transformation must be taken into account. Even groundwater transects can provide interesting historical trends.7 Such samples will generally reveal only population trends. Dust samples are typically more representative of individual indoor exposure, with many targeted studies now showing health-related conclusions.106 Increasing numbers of non-target studies are being published, including several collaborative and ring trials,29,107 showing this to be a matrix of great interest for future efforts, although high quality data processing is still challenging for routine use.

The quality of the samples of interest, and thus downstream chemical and data analysis, is highly impacted by pre-analytical variations. Ideally, according to biobank protocols, tissue samples should be snap-frozen in liquid nitrogen immediately after sampling. However, for logistical reasons, it is often not feasible to respect these recommendations. Thereby, it is crucial to monitor the pre-analytical variations that can occur after sampling, such as the time between the sampling and the processing of the sample, temperature, pH among others. Especially for biomarker discovery/validation studies, it is crucial to monitor these changes. In this regard, biobanks have established a standard pre-analytical coding for biospecimens to enable standardised documentation of the sample collection and processing to limit pre-analytical variations.108 This becomes crucial in multi-center collection studies. In general, biobanks have stringent protocols for sample preservation and storage, together with strict quality controls. Biobanks typically store biospecimens derived from human individuals, such as blood derivatives (plasma, serum), cerebrospinal fluid, urine, saliva, stool and cells,109 and are all potential matrices for neurotoxicant analyses depending on the study interest. Together with the samples, metadata on medical variables, dietary information, as well as environmental data are collected and made available for research. In terms of creating a digital “snapshot” of small molecule conditions, analysis as close to time of receipt will be a distinct advantage to prevent sample degradation, with the downside of creating large variabilities and batch effects in downstream data analysis. However, with sufficient quality controls and strategies to best capture batch effects to allow correction (standard measurements, internal
Analysing the chemical space of neurotoxicants

Chemical properties and analytical methods

Not all analytical methods are equally efficient and, as evidenced by the variety in the physico-chemical properties of potential neurotoxicants (see Fig. 4), it is not realistic to expect a “one size fits all” analytical method to perform a full screen of all potential neurotoxicants in a given matrix. A detailed description of the list and the corresponding diversity in physico-chemical properties can be found in the ESI.† Already the choice of solvents during extraction will have an impact on the chemical nature of the analytes to be found in the data acquisition. Multiple extractions as well as measurement protocols have to be applied to enable a comprehensive picture of the possible neurotoxicant composition of the matrix of interest, especially in NT-HR-MS. To create long term records, a minimum of analytical measurements capturing the maximum amount of information possible within the compromise of time, cost and effort will be needed to provide enough information to direct future, more sophisticated analysis on the frozen and archived (physically and digitally) samples. In the next section we explore some of the analytical options available to cover this extraordinarily broad range of contaminants within the category “potential neurotoxicants”. As there are many excellent reviews on analysis,11 including one specifically on PD,109 the text below covers major analytical aspects and decisions that we consider most likely to affect the quality of NT-HR-MS data processing that can be performed, while the ESI† contains some additional discussion.

When performing potentially long term studies of complex samples, careful sample pre-processing steps are required for the analysis, such as the removal of highly abundant matrix components such as lipids and/or pigments that could interfere with the analysis. However, the use of any sample clean-up technique, as well as the choice of the extraction method, carries the risk of eliminating the actual compounds of interest. Furthermore, careful but simple sample preparation and pretreatment is required that will enable minimum interferences and appropriate normalisation of samples for data processing and subsequent identification efforts over years of data collection. This is especially challenging for unknowns, as the usual techniques applied to targeted compounds (internal standard, recovery and matrix correction) cannot be applied; furthermore as the structure is not known it is challenging to select closely-matching substances. Comparing potential neurotoxicants in samples (especially e.g. faecal matter110 or house dust) from different groups in cohort studies requires the normalisation of the qualitative or quantitative signal. For instance, the concentration measured in an extract of stool is affected by the properties of the sample (water content, percentage of undigested matter, etc.) beyond even the complexity of wastewater samples, which have been “averaged” in a way throughout the journey to the treatment plant. In the metabolomics context, the retrospective normalisation of the measurement signal to the overall metabolite content (total ion chromatogram or TIC normalisation) or the dry weight of the sample are most common. While the TIC normalisation works only under the (theoretical) assumption that the overall metabolic profile is constant and comparable in all samples of the sequence, the retrospective normalisation to the sample dry weight makes it impossible to adjust sample preparation steps, like the amount of extraction fluid, to achieve best comparability between samples. Such issues make high quality non-targeted data processing and the generation of robust statistics extremely challenging, and will require some dedicated efforts in the coming years, as it is easy for such issues to be inadvertently disregarded, especially in automated workflows.

Measuring the chemical diversity of potential neurotoxicants

In terms of measurement, a large percentage of the neurotoxicants have polarities and ionizable groups that make these molecules amenable to analysis by liquid chromatography high resolution mass spectrometry (LC-HRMS). API techniques like electrospray ionisation (ESI) and atmospheric chemical ionisation (APCI) has dramatically improved compound discovery because of enhanced ionisation efficiency, as well as preventing the need for derivatisation (often the case with GC-MS) that may complicate data interpretation. Analytes that easily lose protons, mainly due to their gas phase
Acidity, can be analyzed by negative ESI, while strongly basic compounds in the gas phase are ionized with ease in positive mode.\textsuperscript{111,112} Weaker bases, on the other hand, are better detected with APCI ionization.\textsuperscript{113} Some potential neurotoxicants that can be analyzed by APCI are hexamethyldimethyamine (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4022579), almitrine (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4057899), and diphenylhydantoin (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID8020541), which have functional groups that are known to ionize better using APCI (triazine for the first two and phenylurea for the last). Endosulfan (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID1020560) has also been shown to be APCI amenable, which provides better precursor ion information compared to EI analysis.\textsuperscript{114} Another technique that may find more utility in the future is atmospheric pressure photo-ionisation (APPI) for the analysis of compounds that ionise poorly using the ionisation techniques mentioned thus far.\textsuperscript{115}

Another lesson learned from the neurotoxicant mapping is that many of these compounds are very polar and thus tend to elute at void volume when using reversed phase LC. Because unretracted matrix components like macromolecules and salts also come out at or near the void time, hydrophilic molecules tend to suffer from ionisation suppression. Some specific examples include amygdalin (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID00897159), stavudine (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID1023819), fludarabine (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4039657) and allopurinol (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4022573). Hydrophilic interaction liquid chromatography (HILIC) is often able to retain these compounds better, thus providing better analyte signal integrity. HILIC would potentially open up unique chemical space that has received very little attention in environmental analysis up until recent years.\textsuperscript{1}

Enabling a better separation of compounds of interest will be critical to obtaining sufficient and clean fragmentation information to assist in identification and confirmation, but requires the addition of yet another method to the NT-HR-MS toolbox.

While derivatisation in LC-MS may sometimes be necessary to improve the analytical signal\textsuperscript{116} and can be used to identify compounds with certain toxicophores,\textsuperscript{117} complicated data interpretation often ensues. Measurements made with and without derivatisation, followed by customised data processing, may ease this process in the coming years. Another group of neurotoxins that require special attention are the toxicants with masses above 1000 Da. For these molecules, multiple charging can be expected in NT-HR-MS and care must be taken when doing data analysis. While multiple charges can be detected in many non-target workflows, these are not commonly investigated in great detail and more work needs to be done in the mass spectrometric analysis of these analytes (see also discussion below).

Further inspection of the neurotoxicant list reveals that many entries are either heavy metals or contain heavy metals. Metals like chromium, lead, mercury, nickel, silver, and thallium, organometallics like dimethyl and ethyl mercury, and the metalloid arsenic have all been reported to have neurotoxic effects. However, many of these compounds are neither LC nor GC amenable. For these compounds one needs to perform a spectrophotometric analysis or use an inductively coupled plasma to convert these analytes to the gas phase. Preliminary results in the context of the ENTACT project revealed that some arsenic-containing organometallic compounds can indeed be detected by LC-NT-HR-MS with very informative fragmentation patterns, showing that these should certainly be considered in the interpretation of NT-HR-MS data, although not currently covered by many workflows – this integration will be critical in future years.

Looking beyond small molecules (toxic proteins)
Apart from small molecules, exposure to toxic biomacromolecules (>1000 Da) may also be highly relevant in the context of neurotoxicity. In this context, neurotoxic as well as amyloidogenic (https://en.wikipedia.org/wiki/Amyloid) proteins are of particular relevance as they have been implicated as causal factors in Alzheimer’s and Parkinson’s diseases.\textsuperscript{117} In particular, prions (misfolded proteins with infectious properties) have been implicated in triggering amyloid-beta, alpha-synuclein and tau misfolding. Once the misfolding of endogenous proteins occurs, propagation of misfolding may occur through mechanisms similar to those that underlie prion pathogenesis.\textsuperscript{118,119} Importantly, the causal agent for the initial misfolding may be distant from the brain regions that are ultimately affected by the diseases.\textsuperscript{120,121} In this context, the enteric nervous system has been implicated.\textsuperscript{122} Based on these observations, it has been postulated that molecules produced by the gut microbiota may trigger a slowly ascending pathological process.\textsuperscript{123,124} For instance, the gut microbiota of newly diagnosed Parkinson’s disease patients is enriched in known mucus-foraging bacteria.\textsuperscript{125} Furthermore, apart from other bacterial toxins,\textsuperscript{126} proteins have been posited to trigger the misfolding of proteins in enteric neurons.\textsuperscript{127} Thus, a combination of mucus erosion and a resulting exposure to higher levels of microbial amyloidogenic proteins might at least play a role in Parkinson’s disease. To obtain pointers to causal mechanisms, it is therefore essential to not only consider small molecules as potential causal agents but also proteins. This means that integrated multi-omic analyses will be highly relevant for resolving potential combinatorial mechanisms involved in neurotoxicity.

As mentioned above, it is possible that protein and peptide signals may be observed in small molecule measurements. This is a challenge that computational workflows will have to tackle increasingly in the future, with respect to the multiple charge states, large masses and also dealing with this information in compatible ways. While multi-omic integration is becoming a common buzzword, the reality is that workflows and formats are still largely incompatible. For instance, the protein structures for download in the Database of Bacterial ExoToxins for Human\textsuperscript{128} (DBETH, http://www.hpppi.ibcb.res.in/btox/) are not in a format compatible with compound databases for small
Relating chemical analytical results from NT-HR-MS to biology

Biological confirmation of neurotoxic effects

While a critical part of NT-HR-MS is identification and analytical confirmation of suspected chemicals with reference standards (where possible), the results must be related back to the original study question. In environmental studies this is quite often performed using bioassays, as in effect-directed analysis. A recent article contains extensive material discussing approaches for screening neurotoxicity from an ecological standpoint (“econeurotoxicity”) and we refer readers to Legradi et al.129 rather than reproducing a summarised form of this discussion here. The material below provides additional perspectives.

Specific features that distinguish the central nervous system (CNS) from other organs need to be considered when choosing the biological systems, read-outs and methods to test the potential neurotoxic effects of environmental chemicals. These features include the presence of the blood–brain barrier, the high lipid content of the nervous tissue, the high energy requirement of neurons, the particular intercellular signaling system (synchronous transmission), the neural cell structure (long axonal projections), the presence of specific reactive endogenous molecules (e.g. dopamine) and the post-mitotic nature of neuronal cells (making them more sensitive to age-related accumulation of cellular damage130). In addition, the heterogeneity of the brain tissue and the sometimes highly selective susceptibility of certain neuronal cell types to only one or another neurotoxic substance calls for the use of more than one test system to maximise the chances of identifying hazardous effects of a potentially new neurotoxic compound.

A current tendency in the field of (neuro)toxicity testing is to transition from more observational projects using rodents to more mechanistic studies involving the use of in vitro (cellular) test systems.131 The latter include primary neural cells, immortalised neural cell lines and stem cells of non-human or human origin. Given species-specific responses to neurotoxic substances, even between rodents and humans, working with human derived cells (as opposed to cells from non-human origin) maximises the chances for making predictions on neurotoxicity relevant to the human nervous system. However, access to human primary cells is limited, given that they are derived from aborted foetuses or from brain surgery resections. Although less limited, isolation of primary cells from rodents is laborious and requires animal sacrifices for each new experiment. This explains the widespread use of rodent or human cell lines for neurotoxicity testing, even though the physiological relevance of the results generated in these systems are more questionable due to their tumorigenic origin or the immortalisation process applied (e.g. overexpression of oncogenes or telomerase). A very dynamic and expanding field is the use of mouse or human derived embryonic stem cells (ESCs) and of human induced pluripotent stem cells (iPSCs), which have been derived more recently by reprogramming of human somatic cells.132,133 A panoply of protocols have been developed to differentiate iPSCs into different neural cell types.134 Exposing these cells to potential neurotoxicants during the differentiation process allows testing more specifically for developmental neurotoxicity. Finally, rapid progress is ongoing in the development of 3D organoid models of the brain, which are also derived from iPSCs.135–137 Here the aim is to better mimic the complexity of the human brain, or certain parts of it (e.g. midbrain, relevant for Parkinson’s disease research), by creating in vitro tissue structures including several cell types in a spatial organisation that reflects closely the physiological situation, but that offer easy access for experimental manipulations.

Common endpoints for in vitro neurotoxicity testing and the analytical approaches used for measuring those endpoints have been reviewed in Schmidt et al.138 with a focus on those methods that allow for high-throughput screening. Briefly, the endpoints can be grouped into viability, morphological and functional read-outs. Viability assays allow testing for neurotoxicity by determining whether one or several types of neural cells are more sensitive to compounds of interest than other cell types. At sub-cytotoxic concentrations, compounds can be further tested for their impact on neural cell morphology or function with endpoints including the monitoring of neurite outgrowth, spontaneous electrical activity of neuronal networks, receptor signaling and cell communication, reactive oxygen species formation, cell migration, mitochondrial transport, calcium storage and release, cell membrane potential, gliosis (i.e. proliferation of glial cells), myelination, network formation, and synaptogenesis. A number of these endpoints, as well as cell viability, can be measured by high-throughput screens based on absorbance, fluorescence or luminescence measurements in multi-well plates. An interesting, more recent development are microelectrode arrays (MEA) that allow for the measurement of extracellular electrical fields of neurons that grow on them. The mean firing rate has been proposed to be the most sensitive parameter for neurotoxicity screening via this method, which allows measurement of the spontaneous electrical activity arising in neural networks at relatively high throughput (24-well MEA plates).139 High-content imaging, i.e. automated fluorescent imaging of fixed or live cells in a high-throughput manner, is a technique that continues to undergo important developments, but has already become a central part of neuroscience and related compound screening.140 Transcriptomics, miRNA profiling and metabolomics are used as analytical tools to characterise molecular changes that occur in response to neurotoxicant exposure. These techniques hold promise for defining molecular signatures that are specific for certain chemicals, providing a read-out that can be more sensitive than morphological changes.141,142
Susceptibility to (neuro)toxicants is defined to a certain extent by the genetic background of each individual. Large cohorts of exposed versus non-exposed individuals are necessary to link genes and pathways with neurotoxicants of interest via genome-wide association studies. A way to circumvent this limitation is to work with model organisms in the lab that are easy to manipulate genetically. Budding yeast can be used for toxicogenomic screens in a highly time and cost-efficient way, given the existence of barcoded genome-wide gene deletion collections. Candidate susceptibility genes identified in this simple eukaryotic model organism can subsequently be validated in human cell culture models, if human homologs of the yeast genes exist. Humanised yeast models relevant for Parkinson’s disease research have been created through overexpression in the yeast model of the human protein alpha-synuclein and have already successfully been used for high-throughput genetic and drug screens. Similarly, these models could be used to implement neurotoxicity screens and identify compounds that may trigger or advance the onset of PD. Roundworm (Caenorhabditis elegans) and zebrafish (Danio rerio) have emerged as simplified multicellular (non-mammalian) model organisms to be used in gene–environment interaction studies. Both organisms have well-characterised nervous systems; additionally, neuro-developmental processes can be visualised conveniently in vivo given the transparency of C. elegans and zebrafish larvae. Neurotoxicity can also be probed in these animals via behavioural studies using computerised tracking systems, which enable high-throughput screens. Candidate susceptibility genes and/or pathways emerging from studies on these simpler model organisms have to be validated in more focused studies in human cell culture systems such as the ones described above. The challenges faced in the toxicogenomics and toxicogenetics fields are similar in many ways to the pharmacogenomics and pharmacogenetics fields and closer cooperation and collaboration between the fields would be beneficial; especially the environmental and toxicological fields could benefit from greater mathematical modelling and prediction such as the virtual physiological human and the Avicenna Alliance (Association for Predictive Medicine).

Relating chemicals/environment to disease

While the first instance for confirming potential neurotoxic effects is a biological confirmation with relatively simple models such as those described above, the ultimate aim will be to relate the chemical exposure to the disease state in humans. As mentioned above, complex, long-term neurodegenerative disorders can involve a lifetime of exposure and yield multiple symptoms that may arise from vastly different causes, whereas the same causes may yield different symptoms in patients. Thus, it remains to be determined how compounds detected with NT-HR-MS can be integrated into the pathomechanism of the disease. For instance, rotenone (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID6021248) (a broad-spectrum insecticide) and paraquat (https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID3034799) (a widely used herbicide) are shown to induce oxidative stress and cytotoxicity via the activation of microglia NOX2, thus leading to cell death of dopaminergic neurons. In other cases, there is clear evidence indicating gene–environment interactions. In one case study, a two-fold difference in PD risk was observed for polymorphic forms of CYP2D6, i.e. the combination of a genetic influence and pesticide exposure was needed for increased risk of developing PD. However, for many compounds the biomechanistic role in the disease process is not clear. The world-wide import and export of food, for example, has implications in exposure (and thus source attribution and suspect screening efforts) as the food we consume is no longer typically produced in the local area, but rather internationally in regions with completely different pesticide use and regulations. As it may also be difficult to predict toxicity for some small molecules, it will be important to discover the relevant effects to warn of certain substances and substance classes. All these factors must be considered in screening studies, requiring worldwide data exchange and a far greater (potential) substance coverage than offered by limited targeted chemical analysis.

It is clear that NT-HR-MS has a lot of potential to provide comprehensive information about potential environmental contaminants and metabolite signatures that may influence disease states. Samples taken and measured now can be archived (both digitally and physically) and screened retrospectively such that chemicals identified in the future could be traced back to determine whether these have, in fact, been present in samples for many years, similar to the NORMAN Early Warning System (NormaNEWS) or the daily monitoring of the Rhine river. However, to do this for long term diseases such as PD will require an incredible investment into high quality patient cohorts, biobanking of the samples, appropriate analytical measurements, state-of-the-art data analysis, carefully-designed biological validation and cross-disciplinary interpretation of the myriad of ensuing results. Such research extends far beyond a single few-year research project or a single research group, or even a dedicated institution such as the Luxembourg Centre for Systems Biomedicine, but requires international collaboration and support for the open exchange of information (within legal frameworks such as data protection and respecting patient confidentiality etc.).

Conclusions/perspectives

With this perspective we have started a conversation amongst several research groups at one institute and close collaborators to scope the breadth of experience needed to begin to perform a prospective, high quality, comprehensive assessment of potential small molecule neurotoxicants in environmental and biological samples using cheminformatics and non-target high
resolution mass spectrometry, and what would be required to link this back to the disease state. Screening the thousands of chemicals potentially involved is a huge task that can be eased significantly by using some carefully-considered cheminformatics, literature-based and data processing methods and we hope users find the resources prepared in this article a useful way to start to explore different approaches for data processing. While we have prepared and described the methods using resources we use and develop, the data files and concepts are transferable to the many excellent workflows and approaches already available. Realistically, non-target MS-based analysis will have to evolve towards a few harmonised methods to ensure sufficient coverage of substances of interest in different contexts. However, compounds not covered by small molecule HR-MS analysis also play a defining role in diseases, such as larger molecules (requiring multi-omics approaches) or metals in salt forms (requiring alternative analysis) and these also need consideration in the bigger picture. Despite the compilation of many chemical lists, it is also clear that these chemicals transform and, although not covered in detail here, the candidate space for NT-HR-MS can and should be extended to screen for predicted transformation products and metabolites of potential neurotoxicants. Another area to explore in future work is a better connection to the concept of disease maps, such as the Parkinson’s disease map, which can be queried computationally – and thus potentially integrated into mass spectral workflows – via interfaces such as MINERVA.

While hypothesis generation can begin already from data snapshots on patients and controls at certain time points, in reality the prioritisation and discovery of “driving” neurotoxicants will depend on forming high quality, long term datasets involving cohorts where participants are recruited even before any symptoms are displayed, so that long term trends become clear. Ideally, this would be extended across multiple cohorts to allow hypothesis development in one cohort and replication in another based on chemical-environmental stratification. Alternatively, the increasing development of sensors and portable devices that can be given to patients and the general population open up new opportunities for citizen science, crowd sourcing of data and greater collaborative efforts to form detailed sampling campaigns that would provide complementary information to cohort data. The increasing willingness to share data in open resources such as the Global Natural Products Social Molecular Networking (GNPS) repository will open up many new opportunities in the coming years. This includes greater availability of larger public datasets on which to develop machine learning and artificial intelligence approaches to improve identification and classification of relevant chemicals and exposures. High resolution mass spectrometry of small molecules is certainly a well-placed technique to start building the knowledge to help answer these extremely complicated questions, assuming the long term support is available to create and maintain such extremely valuable datasets. In the end, however, this will provide only a starting point for further investigations. In addition to hands on biological experiments to elucidate possible exposure routes or gene–environment interactions, additional efforts in computational modeling and simulation using mathematical and mechanistic models will be needed to turn identified chemicals into a greater understanding of an extremely complex system and potential environmental causes of chronic diseases.

Author contributions

ELS conceptualised the perspective upon invitation and coordinated the writing. General contributions: cheminformatics: NCB (literature mining), AJW (registration, lists) and ELS (NT-HR-MS); analytical/samples: RS, JPT, PLK, ELS. Domain-specific knowledge: JPT (biobanking, metabolomics), PLK (medical), RK (clinical/molecular neuroscience), PW (metabolomics), CLL (biology), NP (metabolomics), RB (systems biomedicine). All authors made essential contributions to discussions and material presented in this perspective.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

ELS is supported by the Luxembourg National Research Fund (FNR) for the ATTRACT project 12341006 and acknowledges additional discussions with Marek Ostaszewski (Bioinformatics Core, LCSB). The Luxembourg cohort, RK and PLK are supported by grants from the Luxembourg National Research Fund (FNR) within the National Centre of Excellence in Research on Parkinson’s disease (NCER-PD; FNR/ NCER13/BM/11264123) and the PEARL programme (FNR; FNR/P13/6682797). NP is supported by an FNR Core Junior Grant (C16/BM/11339953). We are grateful to the peer reviewers for their thoughtful comments that have helped refine this manuscript. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Notes and references


48 P. Vines, Exposomics: mathematics meets biology: Figure 1, *Mutagenesis*, 2015, gev068.


65 EMBL-EBI, Chemical Entities of Biological Interest (ChEBI), https://www.ebi.ac.uk/chebi/, accessed 9 June 2019.


105 A. C. Chiaia-Hernandez, M. Krauss and J. Hollender, Screening of Lake Sediments for Emerging Contaminants by Liquid Chromatography Atmospheric Pressure Photoionization and Electrospray Ionization Coupled to


